and a continuation-in-part of application No. 07/483, 913, filed on Feb. 22, 1990, now Pat. No. 5,171,574, which is a continuation-in-part of application No. 07/422,613, filed on Oct. 17, 1989, now Pat. No. 4,975,526, which is a continuation-in-part of application No. 07/315,342, said application No. 07/841, 646 is a continuation-in-part of application No. 07/600,024, filed on Oct. 18, 1990, now abandoned, which is a continuation-in-part of application No. 07/569,920, filed on Aug. 20, 1990, now abandoned, said application No. 07/841,646 is a continuation-in-part of application No. 07/599,543, filed on Oct. 18, 1990, now abandoned, which is a continuation-in-part of application No. 07/569,920, said application No. 07/841,646 is a continuation-in-part of application No. 07/616,374, filed on Nov. 21, 1990, now Pat. No. 5,162,114, which is a division of application No. 07/422,613, said application No. 07/841,646 is a continuation-in-part of application No. 07/483,913.

(51) Int. Cl. C07K 14/51 (2006.01)

(52) U.S. Cl. ........................................ 530/350; 530/351

(58) Field of Classification Search ............... 530/350, 530/300

See application file for complete search history.

(56) References Cited

U.S. PATENT DOCUMENTS

4,120,730 A 10/1978 Trojer et al. ................. 106/39.6
(Continued)

FOREIGN PATENT DOCUMENTS

(Continued)

OTHER PUBLICATIONS

(Continued)

Primary Examiner—Elizabeth C. Kemmerer
(74) Attorney, Agent, or Firm—Kirkpatrick & Lockhart
Nicholson Graham LLP

(57) ABSTRACT

Disclosed are (1) osteogenic devices comprising a matrix containing substantially pure natural-sourced mammalian osteogenic protein; (2) DNA and amino acid sequences for novel polypeptide chains useful as subunits of dimeric osteogenic proteins; (3) vectors carrying sequences encoding these novel polypeptide chains and host cells transfected with these vectors; (4) methods of producing these polypeptide chains using recombinant DNA technology; (5) antibodies specific for these novel polypeptide chains; (6) osteogenic devices comprising these recombinantly produced proteins in association with an appropriate carrier matrix; and (7) methods of using the osteogenic devices to mimic the natural course of endochondral bone formation in mammals.

21 Claims, 49 Drawing Sheets

FOREIGN PATENT DOCUMENTS

EP 128041 12/1984
EP 148155 1/1985
EP 169001 7/1985
EP 182483 10/1985
EP 189016 1/1986
EP 170979 2/1986
EP 212474 8/1986
EP 230647 5/1987
GB 2178447 2/1987
WO 85/05274 12/1985
WO 86/00526 1/1986
WO 88/00205 1/1988
WO 89/09615 10/1989
WO 89/10409 11/1989
WO 90/03733 4/1990
WO 90/11366 10/1990
WO 92/05199 4/1992
WO 92/17584 10/1992
WO 93/00049 1/1993
WO 93/16999 8/1993
WO 94/15949 1/1994
WO 96/14335 5/1996
WO 97/06254 2/1997
WO 99/61611 12/1999
WO 99/66060 12/1999
WO 2001/011041 2/2001

OTHER PUBLICATIONS

Ngo et al., 1994, The Protein Folding Problem and Tertiary Structure Prediction, Merz et al., eds., Birkhäuser, Boston, pp. 491-495.*


Fig. 2A

Fig. 2B
Fig. 8

Fig. 9
DOSE CURVES
PERCENT BONE FORMATION

Fig. 11

Fig. 12A Fig. 12B Fig. 12C
CONSENSUS GENE/PROBE:

```
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCCTAAATGGGCTGTACGGACTTCAGCCAGCCGACGTGGGCTGGGAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPNGLYVDFQRDRDVGD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sau3A</td>
<td>RsaI</td>
<td>AccII</td>
<td>HhaI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>GACTGGATCATCGCCCCCCTCGACTTCAGCCAGCTACTACTGCTCCGGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWIIAPVDFDAYYCSG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sau3A</td>
<td>Accl</td>
<td>TaqI</td>
<td>BspMII</td>
<td></td>
</tr>
<tr>
<td>HincII</td>
<td>AhaII</td>
<td>HpaII</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SalI</td>
<td>HgaI+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TaqI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCTGCCAGTTCCCCCTCTGGCAGATCACTTCACAGCAGCCAACCACGCCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACQFSPADHFSNSTNHA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MnlI+</td>
<td>Sau3A</td>
<td>DraIII</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGTGCAAGGCCCTGGTGAACAACTGAACCCCCGGCAAGGTACCACAGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VVQTVLNNMNPGKVPK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRII</td>
<td>HpaII</td>
<td>BanI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HphI+</td>
<td>NciI</td>
<td>KpnI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ScrFI</td>
<td>ScrFI</td>
<td>RsaI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Fig. 13A
CCTGCTGCCTGCAGTCTGCTCCGACCATCAGCATGCTGTACCTGGA

PCCVPTELSAISMLYLDD

Fnu4HI AluI NspHI EcoRII SphI RsaI ScrFI

CGAGAATTCCACCGTGGTGCTGAAGAATTACCCAGGAGATGACCGTGTT

ENSTVVLKNYQEMTVV

EcoRI MboII+ EcoRII ScrFI

GGGCTGCGGCTGCCGCTAACTGCAG

cccgacgccgacgacgccattgacgt

GCGCR* Fnu4HI Fnu4HI

Fig. 13B
Fig. 15

NON-REDUCIBLE 30K
18K SUBUNIT
16K SUBUNIT
Fig. 19D
Fig. 19F

- pH783
- SV40 splice & polyA
- SV40 e & ori
- DHFR
- SV40 splice & polyA
- SV40 late

Enzymes:
- NheI
- Smal
- SalI
- XhoI
- ClaI
- mMT
- MMTV LTR
- Hind III
- RSV & LTR
- Ndel
- BamH1
- SacI
- EcoRI
- PstI
- AMP
Fig. 20
hOPl-18Ser     Ser Thr Gly Ser Lys Gln Arg Ser Gln
m0Pl-Ser      ... ... ... Gly ... ... ... ...
             1      5

hOPl-18Ser     Asn Arg Ser Lys Thr Pro Lys Asn Gln
m0Pl-Ser      ... ... ... ... ... ... ... ...
             10     15

hOPl-18Ser     Glu Ala Leu Arg Met Ala Asn Val Ala
m0Pl-Ser      ... ... ... ... ... ... ... ...
             20     25

hOPl-18Ser     Glu Asn Ser Ser Ser Asp Gln Arg Gln
m0Pl-Ser      ... ... ... ... ... ... ... ...
             30     35

hOPl-18Ser     Ala Cys Lys Lys His Glu Leu Tyr Val
m0Pl-Ser      ... ... ... ... ... ... ... ...
             40     45

hOPl-18Ser     Ser Phe Arg Asp Leu Gly Trp Gln Asp
m0Pl-Ser      ... ... ... ... ... ... ... ...
             50

hOPl-18Ser     Trp Ile Ile Ala Pro Glu Gly Tyr Ala
m0Pl-Ser      ... ... ... ... ... ... ... ...
             55     60

hOPl-18Ser     Ala Tyr Tyr Cys Glu Gly Glu Cys Ala
m0Pl-Ser      ... ... ... ... ... ... ... ...
             65     70

Fig. 23A
<table>
<thead>
<tr>
<th>Residue</th>
<th>h0Pl-18Ser</th>
<th>m0Pl-Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>Phe Pro Leu Asn Ser Tyr Met Asn Ala</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Thr Asn His Ala Ile Val Gln Thr Leu</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Val His Phe Ile Asn Pro Glu Thr Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asp</td>
</tr>
<tr>
<td>100</td>
<td>Pro Lys Pro Cys Cys Ala Pro Thr Gln</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Leu Asn Ala Ile Ser Val Leu Tyr Phe</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>Asp Asp Ser Ser Asn Val Ile Leu Lys</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Lys Tyr Arg Arg Asn Met Val Val Arg</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>Ala Cys Gly Cys His</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 23B**
Fig. 23C
h0P2-Ala  | Ala  | Tyr  | Tyr  | Cys  | Glu  | Gly  | Glu  | Cys  | Ser  |
          |      |      |      |      |      |      |      |      |      |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | 65   | 70   |

h0P2-Ala  | Phe  | Pro  | Leu  | Asp  | Ser  | Cys  | Met  | Asn  | Ala  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

h0P2-Ala  | Thr  | Asn  | His  | Ala  | Ile  | Leu  | Gln  | Ser  | Leu  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

h0P2-Ala  | Val  | His  | Leu  | Met  | Lys  | Pro  | Asn  | Ala  | Val  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

h0P2-Ala  | Pro  | Lys  | Ala  | Cys  | Cys  | Ala  | Pro  | Thr  | Lys  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

h0P2-Ala  | Leu  | Ser  | Ala  | Thr  | Ser  | Val  | Leu  | Tyr  | Tyr  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

h0P2-Ala  | Asp  | Ser  | Ser  | Asn  | Asn  | Val  | Ile  | Leu  | Arg  |
m0P2-Ala  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  | ...  |

Fig. 23D
h0P2-Ala  Lys His Arg Asn Met Val Val Lys
m0P2-Ala  ......................... 130

h0P2-Ala  Ala Cys Gly Cys His
m0P2-Ala  ......................... 135

Fig. 23E
<table>
<thead>
<tr>
<th></th>
<th>hOP1-18Ser</th>
<th>Ser Thr Gly Ser Lys Gln Arg Ser Gln</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOP1-Ser</td>
<td>... ... ... Gly ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>hOP2-Ala</td>
<td>Ala Val Arg Pro Leu Arg ... Arg ...</td>
<td></td>
</tr>
<tr>
<td>mOP2-Ala</td>
<td>Ala Ala Arg Pro Leu Lys ... Arg ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>hOP1-18Ser</th>
<th>Asn Arg Ser Lys Thr Pro Lys Asn Gln</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOP1-Ser</td>
<td>... ... ... ... ... ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>hOP2-Ala</td>
<td>Pro Lys Lys Ser Asn Glu Leu Pro Gln</td>
<td></td>
</tr>
<tr>
<td>mOP2-Ala</td>
<td>Pro Lys Lys Thr Asn Glu Leu Pro His</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>hOP1-18Ser</th>
<th>Glu Ala Leu Arg Met Ala Asn Val Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOP1-Ser</td>
<td>... ... ... Ser ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>hOP2-Ala</td>
<td>Ala Asn Arg Leu Pro Gly Ile Phe Asp</td>
<td></td>
</tr>
<tr>
<td>mOP2-Ala</td>
<td>Pro Asn Lys Leu Pro Gly Ile Phe Asp</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>hOP1-18Ser</th>
<th>Glu Asn Ser Ser Ser Asp Gln Arg Gln</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOP1-Ser</td>
<td>... ... ... ... ... ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>hOP2-Ala</td>
<td>Asp Val His Gly ... His Gly ... ...</td>
<td></td>
</tr>
<tr>
<td>mOP2-Ala</td>
<td>Asp Gly His Gly ... Arg Gly ... Glu</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>hOP1-18Ser</th>
<th>Ala Cys Lys Lys His Glu Leu Tyr Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>mOP1-Ser</td>
<td>... ... ... ... ... ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>hOP2-Ala</td>
<td>Val ... Arg Arg ... ... ... ... ...</td>
<td></td>
</tr>
<tr>
<td>mOP2-Ala</td>
<td>Val ... Arg Arg ... ... ... ... ...</td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 24A
<table>
<thead>
<tr>
<th></th>
<th>h0P1-18Ser</th>
<th>m0P1-Ser</th>
<th>h0P2-Ala</th>
<th>m0P2-Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Ser</td>
<td>Phe</td>
<td>Arg</td>
<td>Asp</td>
</tr>
<tr>
<td>5</td>
<td>Leu</td>
<td>Gly</td>
<td>Trp</td>
<td>Gln</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>h0P1-18Ser</td>
<td>m0P1-Ser</td>
<td>h0P2-Ala</td>
<td>m0P2-Ala</td>
</tr>
<tr>
<td>0</td>
<td>Trp</td>
<td>Ile</td>
<td>Ile</td>
<td>Ala</td>
</tr>
<tr>
<td>55</td>
<td>Pro</td>
<td>Glu</td>
<td>Gly</td>
<td>Tyr</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>h0P1-18Ser</td>
<td>m0P1-Ser</td>
<td>h0P2-Ala</td>
<td>m0P2-Ala</td>
</tr>
<tr>
<td>0</td>
<td>Ala</td>
<td>Tyr</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td>65</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Cys</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>h0P1-18Ser</td>
<td>m0P1-Ser</td>
<td>h0P2-Ala</td>
<td>m0P2-Ala</td>
</tr>
<tr>
<td>0</td>
<td>Phe</td>
<td>Pro</td>
<td>Leu</td>
<td>Asn</td>
</tr>
<tr>
<td>75</td>
<td>Ser</td>
<td>Tyr</td>
<td>Met</td>
<td>Asn</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>h0P1-18Ser</td>
<td>m0P1-Ser</td>
<td>h0P2-Ala</td>
<td>m0P2-Ala</td>
</tr>
<tr>
<td>0</td>
<td>Thr</td>
<td>Asn</td>
<td>His</td>
<td>Ala</td>
</tr>
<tr>
<td>85</td>
<td>Ile</td>
<td>Val</td>
<td>Gln</td>
<td>Thr</td>
</tr>
<tr>
<td>90</td>
<td>Leu</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h0P1-18Ser</td>
<td>Val His Phe Ile Asn Pro Glu Thr Val</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m0P1-Ser</td>
<td>.....................................</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h0P2-Ala</td>
<td>.....................................</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m0P2-Ala</td>
<td>.....................................</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| h0P1-18Ser | Pro Lys Pro Cys Cys Ala Pro Thr Gln |
| m0P1-Ser  | ..................................... |
| h0P2-Ala  | ..................................... |
| m0P2-Ala  | ..................................... |

| h0P1-18Ser | Leu Asn Ala Ile Ser Val Leu Tyr Phe |
| m0P1-Ser  | ..................................... |
| h0P2-Ala  | ..................................... |
| m0P2-Ala  | ..................................... |

| h0P1-18Ser | Asp Asp Ser Ser Asn Val Ile Leu Lys |
| m0P1-Ser  | ..................................... |
| h0P2-Ala  | ..................................... |
| m0P2-Ala  | ..................................... |

| h0P1-18Ser | Lys Tyr Arg Asn Met Val Val Arg |
| m0P1-Ser  | ..................................... |
| h0P2-Ala  | ..................................... |
| m0P2-Ala  | ..................................... |

**Fig. 24C**
h0P1-18Ser  Ala Cys Gly Cys His
m0P1-Ser    ... ... ... ... ...
H0P2-Ala    ... ... ... ... ...
m0P2-Ala    ... ... ... ... ...

135

Fig. 24D
INFLUENCE OF SOLVENT EXTRACT
ALAKALINE PHOSPHATASE (u/mg PROTEIN)

RAT + OP
RAT + OP + SOLVENT

Fig. 28A

INFLUENCE OF SOLVENT EXTRACT
CALCIUM CONTENT (µg/mg TISSUE)

RAT + OP
RAT + OP + SOLVENT

Fig. 28B
**Figure 32A**

**Alkaline Phosphatase (u/mg Protein)**

- **POOL A**
- **POOL B**
- **POOL C**

<table>
<thead>
<tr>
<th>Concentration (ng)</th>
<th>Pool A</th>
<th>Pool B</th>
<th>Pool C</th>
</tr>
</thead>
<tbody>
<tr>
<td>35ng</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>70ng</td>
<td>1.0</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>140ng</td>
<td>1.5</td>
<td>1.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

**Figure 32B**

**Calcium Content (µg/mg Tissue)**

- **POOL A**
- **POOL B**
- **POOL C**

<table>
<thead>
<tr>
<th>Concentration (ng)</th>
<th>Pool A</th>
<th>Pool B</th>
<th>Pool C</th>
</tr>
</thead>
<tbody>
<tr>
<td>35ng</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>70ng</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>140ng</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>OP1-18Ser</td>
<td>Ser</td>
<td>Thr</td>
<td>Gly</td>
</tr>
<tr>
<td>OP1-16Ser</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP1-16Leu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP1-16Met</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP1-16A1a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP1-16Val</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{N-Termini of Active OP1 Sequences} \]

Fig. 33
OSTEOGENIC PROTEINS

REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Ser. No. 08/957, 425, filed Oct. 24, 1997 now U.S. Pat. No. 6,586,388, which is a continuation of U.S. Ser. No. 08/447,570, filed May 23, 1995 now U.S. Pat. No. 5,714,589, which is a divisional of U.S. Ser. No. 08/147,023, filed Nov. 1, 1993 now U.S. Pat. No. 5,468,845, which is a divisional of U.S. Ser. No. 07/841,646, filed Feb. 21, 1992, now U.S. Pat. No. 5,266,683, which is a continuation-in-part of U.S. application Ser. Nos.: 1) Ser. No. 07/827,052, filed Jan. 28, 1992, now U.S. Pat. No. 5,250,302 and which is a divisional of U.S. Ser. No. 07/179,406, filed Apr. 8, 1988, now U.S. Pat. No. 4,968,590; 2) Ser. No. 07/579,865, filed Sept. 7, 1990, now U.S. Pat. No. 5,108,753 and which is a divisional of U.S. Ser. No. 07/179,406, filed Dec. 4, 1990, now abandoned, that was a divisional of U.S. Ser. No. 07/232,630, filed Aug. 15, 1988, now abandoned, that was a continuation-in-part of Ser. No. 07/179,406; 3) Ser. No. 07/621,849, filed Dec. 4, 1990, abandoned in favor of 07/995,345, now U.S. Pat. No. 5,258,494 and which was a divisional of 07/315,342 filed Feb. 23, 1989, now U.S. Pat. No. 5,011,691 and which is a continuation-in-part of Ser. No. 07/232,630; 5) Ser. No. 07/810,560, filed Dec. 20, 1991, now abandoned, that was a continuation of Ser. No. 07/660,162, filed Feb. 22, 1991, now abandoned, that was a continuation of Ser. No. 07/422,699, filed Oct. 17, 1989, now abandoned, that was a continuation-in-part of Ser. No. 07/315,342; 6) Ser. No. 07/569,920, filed Aug. 20, 1990, now abandoned, that was a continuation-in-part of Ser. No. 07/422,699 and 07/483,913, now U.S. Pat. No. 5,171,574 which is a continuation-in-part of Ser. No. 07/422,613, filed Oct. 17, 1989, now U.S. Pat. No. 4,975,526 and which is a continuation-in-part of Ser. No. 07/315,342; 7) Ser. No. 07/600,024, filed Oct. 18, 1990, now abandoned, that was a continuation-in-part of Ser. No. 07/569,302; 8) Ser. No. 07/590,543, filed Oct. 18, 1990, now abandoned, that was a continuation-in-part of Ser. No. 07/569,920; 9) Ser. No. 07/616,374, filed Nov. 21, 1990, now U.S. Pat. No. 5,162,114 and which is a divisional of Ser. No. 07/422,613; and 10) Ser. No. 07/483,913, filed Feb. 22, 1990, now U.S. Pat. No. 5,171,574, the disclosures of which are incorporated by reference herein.

TECHNICAL FIELD OF THE INVENTION

This invention relates to osteogenic devices, to DNA sequences encoding proteins which can induce new bone formation in mammals, and to methods for the production of these proteins in mammalian cells using recombinant DNA techniques, including host cells capable of expressing these sequences. The invention also relates to the proteins expressed from these DNA sequences, to antibodies capable of binding specifically to these proteins, and to bone and cartilage repair procedures using the osteogenic devices. The invention further relates to matrix materials useful for allogenic or xenogenic implants and which act as a carrier of the osteogenic protein to induce new bone formation in mammals.

BACKGROUND OF THE INVENTION

Mammalian bone tissue is known to contain one or more proteinaceous materials, presumably active during growth and natural bone healing, which can induce a developmental cascade of cellular events resulting in endochondral bone formation. This active factor (or factors) has variously been referred to in the literature as bone morphogenetic or morphogenic protein, bone inductive protein, osteogenic protein, osteogenin, or osteoinductive protein.


Though the precise mechanisms underlying these phenotypic transformations are unclear, it has been shown that the natural endochondral bone differentiation activity of bone matrix can be dissociatively extracted and reconstituted with inactive residual collagenous matrix to restore full bone induction activity (Sampath and Reddi (1981) Proc. Natl. Acad. Sci. USA 78:7599–7603). This provides an experimental method for assaying protein extracts for their ability to induce endochondral bone in vivo. Several species of mammals produce closely related protein as demonstrated by the ability of cross species implants to induce bone formation (Sampath and Reddi (1983) Proc. Natl. Acad. Sci. USA 80:6591–6595).

The potential utility of these proteins has been recognized widely. It is contemplated that the availability of the protein would revolutionize orthopedic medicine, certain types of plastic surgery, dental and various periodontal and craniofacial reconstructive procedures.

The observed properties of these protein fractions have induced an intense research effort in several laboratories directed to isolating and identifying the pure factor or factors responsible for osteogenic activity. The current state of the art of purification of osteogenic protein from mammalian bone is disclosed by Sampath et al. (1987) Proc. Natl. Acad. Sci. USA 84:7109–7113. Urist et al. (1983) Proc. Soc. Exp. Biol. Med. 173:194–199 disclose a human osteogenic protein fraction which was extracted from demineralized cortical bone by means of a calcium chloride-urea inorganic-organic solvent mixture, and retrieved by differential precipitation in guanidine-hydrochloride and preparative gel electrophoresis. The authors report that the protein fraction has an amino acid composition of an acidic polypeptide and a molecular weight in a range of 17–18 kDa. This material was said to be distinct from a protein called “bone derived growth factor” disclosed by Canalis et al. (1980 Science 210:1021–1023, and by Farley et al. (1982) Biochem 21:3508–3513.

Urist et al. (1984) Proc. Natl. Acad. Sci. USA 81:371–375 disclose a bovine bone morphogenetic protein extract having the properties of an acidic polypeptide and a molecular weight of approximately 18 kDa. The authors reported that the protein was present in a fraction separated by hydroxyapatite chromatography, and that it induced bone formation in mouse hindquarter muscle and bone regeneration in trephine defects in rat and dog skulls. Their method of obtaining the extract from bone results in ill-defined and impure preparations.

European Patent Application Serial No. 148,155, published Oct. 7, 1985, purports to disclose osteogenic proteins derived from bovine, porcine, and human origin. One of the proteins, designated by the inventors as a P3 protein having a molecular weight of 22–24 kDa, is said to have been purified to an essentially homogeneous state. This material is reported to induce bone formation when implanted into animals.

US 7,176,284 B2
International Application No. PCT/087/01537, published Jan. 14, 1988 (Int. Pub. No. WO88/00205), discloses an impure fraction from bovine bone which has bone induction qualities. The named applicants also disclose putative “bone inductive factors” produced by recombinant DNA techniques. Four DNA sequences were retrieved from human or bovine genomic or cDNA libraries and expressed in recombinant host cells. While the applicants stated that the expressed proteins may be bone morphogenic proteins, bone induction was not demonstrated. This same group reported subsequently ((1988) Science 242:1528–1534) that three of the four factors induce cartilage formation, and postulate that bone formation activity “is due to a mixture of regulatory molecules” and that “bone formation is most likely controlled . . . by the interaction of these molecules.” Again, no bone induction was attributed to the products of expression of the cDNAs. See also Urist et al., EPO 0,212,474 entitled “Bone Morphogenic Agents”.

Wang et al. (1988) Proc. Nat. Acad. Sci. USA 85:9484–9488, disclose the partial purification of a bovine bone morphogenetic protein from guanidine extracts of demineralized bone having cartilage and bone formation activity as a basic protein corresponding to a molecular weight of 30 kDa determined from gel elution. Separation of the 30 kDa fraction yielded proteins of 30, 18 and 16 kDa which, upon separation, were inactive. In view of this result, the authors acknowledged that the exact identity of the active material had not been determined.

Wang et al. (1990) Proc. Nat. Acad. Sci. USA 87:2220–2224 describe the expression and partial purification of one of the CDNA sequences described in PCT/87/01537. Consistent cartilage and/or bone formation with their protein requires a minimum of 600 ng of 50% pure material.

International Application No. PCT/89/04458 published Apr. 19, 1990 (Int. Pub. No. WO90/003733), describes the purification and analysis of a family of osteogenic factors called “P3 OF 31–34”. The protein family contains at least four proteins, which are characterized by peptide fragment sequences. The impure mixture P3 OF 31–34 is assayed for osteogenic activity. The activity of the individual proteins is neither assessed nor discussed.

It also has been found that successful implantation of the osteogenic factors for endochondral bone formation requires association of the proteins with a suitable carrier material capable of maintaining the proteins at an in vivo site of application. The carrier should be biocompatible, in vivo biodegradable and porous enough to allow cell infiltration. The insoluble collagen particles remaining after guanidine extraction and delipidation of pulverized bone generally have been found effective in allogenic implants in some species. However, studies have shown that while osteoinductive proteins are useful cross species, the collagenous bone matrix generally used for inducing endochondral bone formation is species-specific (Sampath and Reddi (1983) Proc. Nat. Acad. Sci. USA 80: 6591–6594). Demineralized, delipidated, extracted xenogenic bone matrix carriers implanted in vivo invariably fail to induce osteogenesis, presumably due to inhibitory or immunogenic components in the bone matrix. Even the use of allogenic bone matrix in osteogenic devices may not be sufficient for osteoinductive bone formation in many species. For example, allogenic, subcutaneous implants of demineralized, delipidated monkey bone matrix is reported not to induce bone formation in the monkey. (Asperberg et al. (1988) J. Bone Joint Surg. (Br) 70-B: 625–627).

U.S. Pat. No. 4,563,350, issued Jan. 7, 1986, discloses the use of trypsinized bovine bone matrix as a xenogenic matrix to effect osteogenic activity when implanted with extracted, partially purified bone-inducing protein preparations. Bone formation is said to require the presence of at least 5%, and preferably at least 10%, non-fibrillar collagen. The named inventors claim that removal of telopeptides which are responsible in part for the immunogenicity of collagen preparations is more suitable for xenogenic implants.


Dentherage et al. (1987) Collagen Rel. Res. 7: 2225–2231, purport to disclose an apparently xenogenic implantable device comprising a bovine bone matrix extract that has been minimally purified by a one-step ion exchange column and reconstituted, highly purified human Type-I placental collagen.

U.S. Pat. No. 3,394,370, issued Jul. 19, 1983, describes a matrix of reconstituted collagen purportedly useful in xenogenic implants. The collagen fibers are treated enzymatically to remove potentially immunogenic telopeptides (also the primary source of interfibril crosslinks) and are dissolved to remove associated non-collagen components. The matrix is formulated by dispersing the reconstituted collagen in acetic acid to form a disordered matrix of elementary collagen molecules that is then mixed with osteogenic factor and lyophilized to form a “semi-rigid foam or sponge” that is preferably crosslinked. The formulated matrix is not tested in vivo.

U.S. Pat. No. 4,172,128, issued Oct. 23, 1979, describes a method for degrading and regenerating bone-like material of reduced immunogenicity, said to be useful cross-species. Demineralized bone particles are treated with a swelling agent to dissolve any associated mucopolysaccharides (glycosaminoglycans) and the collagen fibers subsequently dissolved to form a homogenous colloidal solution. A gel of reconstituted fibers then can be formed using physiologically inert mucopolysaccharides and an electrolyte to aid in fibril formation.

It is an object of this invention to provide osteogenic devices comprising matrices containing dispersed osteogenic protein, purified from naturally-sourced material or produced from recombinant DNA, and capable of bone induction in allogenic and xenogenic implants. Another object is to provide novel polypeptide chains useful as subunits of dimeric osteogenic proteins, as well as DNA sequences encoding these polypeptide chains and methods for their production using recombinant DNA techniques. Still another object is to provide recombinant osteogenic proteins expressed from procaroytic or eucaryotic cells, preferably mammalian cells, and capable of inducing endochondral bone formation in mammals, including humans, and to provide methods for their production, including host cells capable of producing these proteins. Still another object is to provide antibodies capable of binding specifically to the proteins of this invention. Yet another object is to provide a biocompatible, in vivo biodegradable matrix capable, in combination with an osteoinductive protein, of producing endochondral bone formation in mammals, including humans.

These and other objects and features of the invention will be apparent from the description, drawings, and claims which follow.
This invention provides osteogenic proteins and devices which, when implanted in a mammalian body, can induce at the locus of the implant the full developmental cascade of endochondral bone formation including vascularization, mineralization, and bone marrow differentiation. The devices comprise a carrier material, referred to herein as a matrix, having the characteristics disclosed below, and containing dispersed substantially pure osteogenic protein either purified from naturally sourced material or produced using recombinant DNA techniques. Recombinantly produced osteogenic protein may be expressed from procarvitory or eucaryotic cells, most preferably mammalian cells. As used herein “substantially pure” means substantially free of other contaminating proteins having no endochonial bone formation activity.

The substantially pure osteogenic protein may include forms having varying glycosylation patterns, a family of related proteins having regions of amino acid sequence homology, and active truncated or mutated forms of native proteins, no matter how derived.

Preferred embodiments of the recombinant protein dispersed in the matrix disclosed herein closely mimic the physiological activity of native form protein extracted from natural sources and reconstituted in allogenic demineralized, guanidine-extracted bone powder matrix material. The preferred proteins have a specific activity far higher than any biosynthetic material heretofore reported, an activity which, within the limits of precision of the activity assay, appears essentially identical to the substantially pure material produced as set forth in U.S. Pat. No. 4,968,590. Thus, this application discloses how to make and use osteogenic devices which induce the full developmental cascade of endochondral bone formation essentially as it occurs in natural bone healing.

A key to these developments was the elucidation of amino acid sequence and structure data of native osteogenic protein “OP”. A protocol was developed which results in retrieval of active, substantially pure osteogenic protein from mammalian bone (e.g., bovine or human) having a half-maximum bone forming activity of about 0.8 to 1.0 ng per mg of implant matrix, as compared to implanted rat demineralized bone matrix (see U.S. Pat. No. 4,968,590). The availability of the material enabled the inventors to elucidate all structural details of the protein necessary to achieve bone formation. Knowledge of the protein’s amino acid sequence and other structural features enabled the identification and cloning of genes encoding native osteogenic proteins.

The osteogenic protein in its mature native form is a glycosylated dimer and has an apparent molecular weight of about 30 kDa as determined by SDS-PAGE. When reduced, the 30 kDa protein gives rise to two glycosylated polypeptide chains (subunits) having apparent molecular weights of about 16 kDa and 18 kDa. In the reduced state, the 30 kDa protein has no detectable osteogenic activity. The unglycosylated protein, which has osteogenic activity, has an apparent molecular weight of about 27 kDa. When reduced, the 27 kDa protein gives rise to two unglycosylated polypeptides having molecular weights of about 14 kDa to 16 kDa.

Naturally-sourced osteogenic protein derived from bovine bone, herein referred to as “bOP” and in related applications as “BOP”, is further characterized by the approximate amino acid composition set forth below:

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Rel. no. res./molec.</th>
<th>Amino acid</th>
<th>Rel. no. res./molec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp/Asn</td>
<td>22</td>
<td>Tyr</td>
<td>11</td>
</tr>
<tr>
<td>Glu/Gln</td>
<td>24</td>
<td>Val</td>
<td>14</td>
</tr>
<tr>
<td>Ser</td>
<td>24</td>
<td>Met</td>
<td>3</td>
</tr>
<tr>
<td>Gly</td>
<td>29</td>
<td>Cys</td>
<td>16</td>
</tr>
<tr>
<td>His</td>
<td>5</td>
<td>Ile</td>
<td>15</td>
</tr>
<tr>
<td>Arg</td>
<td>13</td>
<td>Leu</td>
<td>15</td>
</tr>
<tr>
<td>Thr</td>
<td>11</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>Ala</td>
<td>18</td>
<td>Phe</td>
<td>7</td>
</tr>
<tr>
<td>Lys</td>
<td>12</td>
<td>Trp</td>
<td>ND</td>
</tr>
</tbody>
</table>

Analysis of digestion fragments from naturally-sourced material purified from bone indicates that the substantially pure material isolated from bone contains the following amino acid sequences:

1. Ser-Phe-Asp-Ala-Tyr-Tyr-Cys-Ser-Gly-Ala-Cys-Glu-Phe-Pro-Met-Pro-Lys;
2. Ser-Leu-Lys-Pro-Ser-Asn-Tyr-Ala-Thr-Ile-Gln-Ser-Ile-Val;
3. Ala-Cys-Cys-Val-Pro-Thr-Glu-Leu-Ser-Ala-Ile-Ser-Met-Leu-Tyr-Leu-Asp-Glu-Asn-Glu-Lys;
4. Met-Ser-Ser-Leu-Ser-Ile-Leu-Phe-Phe-Asp-Glu-Asn-Lys;
5. Val-Gly-Val-Val-Pro-Gly-Ile-Pro-Glu-Pro-Cys-Cys-Val-Pro-Glu;
6. Val-Asp-Phe-Ala-Asp-Ile-Gly
7. Val-Pro-Lys-Pro;
8. Ala-Pro-Thr.

A consensus DNA gene sequence based in part on these partial amino acid sequence data and on observed homologies with structurally related genes reported in the literature (or the sequences they encode), having a presumed or demonstrated unrelated developmental function, was used as a probe for identifying and isolating genes encoding osteogenic proteins from genomic and cDNA libraries. The consensus sequence probe enabled isolation of a previously unidentified DNA sequence from human genomic and cDNA libraries, portions of which, when appropriately cleaved and ligated, encode a protein comprising a region capable of inducing endochondral bone formation when properly modified, incorporated in a suitable matrix, and implanted as disclosed herein. The predicted amino acid sequence of the encoded protein includes sequences identified in peptide fragments obtained from the substantially pure osteogenic protein (see infra and Kuben Sampath et al. (1990) J. Biol. Chem. 265:13198–13205.) The protein has been expressed from the full length cDNA sequence (referred to herein as “bOP1”), as well as from various truncated DNAs and fusion constructs in both procarvitory (e.g., E. coli) and eucaryotic (various mammalian cells and cell lines) and shown to exhibit osteogenic activity. The OP1 protein in combination with BMP2B also is active (see infra).
Table I lists the various species of the hOPl protein identified to date, including their expression sources and nomenclature and Sequence Listing references. In its native form, hOPl expression yields an immature translation product (“hOPl-PP,” where “PP” refers to “prepro form”) of about 400 amino acids that subsequently is processed to yield a mature sequence of 139 amino acids (“OP1-18Ser”).

The active region (functional domain) of the protein includes the C-terminal 97 amino acids of the OPl sequence (“OPS”). A longer active sequence is OP7 (comprising the C-terminal 102 amino acids).

The consensus sequence probe also retrieved human DNA sequences identified in PCT/087/01537, referenced above, designated therein as BMP2 (Class I and II), and BMP5. The inventors herein discovered that certain subparts of the sequences designated in PCT/087/01537 as BMP-2 Class I and BMP-2 Class II, also referred to in the literature as BMP2 and BMP4, respectively, when properly assembled, encode proteins (referred to herein as “CBMP2A” and “CBMP2B,” respectively) which have true osteogenic activity, i.e., induce the full cascade of events leading to endochondral bone formation when properly folded, dimerized, and implanted in a mammal. Seq. Listing ID Nos. 4 and 6 disclose the cDNA sequences and encoded “prepro” forms of CBMP2A and CBMP2B, respectively. (Nomenclature note: as used herein, “CBMP2(a)” and “CBMP2(b)” refer to the DNA sequence; “CBMP2A” and “CBMP2B” refer to the encoded proteins.) The functional domain (active region) of the CBMP2 proteins comprises essentially amino acid residues 301–396 of Seq. ID No. 4 (designated “CBMP2A”) and residues 313–408 of ID No. 6 (designated “CBMP2B”). Longer active regions are defined by residues 296–396 of Seq. ID No. 4 (“CBMP2AL”) and residues 308–408 of Seq. ID No. 6 (“CBMP2BL”). The CBMP2 proteins share approximately 58–60% amino acid sequence homology with OPl in the active region (e.g., with OP7).

As indicated above, the natural-sourced osteogenic protein is a glycosylated dimer comprising an 18 kDa subunit and a 16 kDa subunit. Protein sequencing data indicate that the larger of the two subunits is mature OPl protein, the other is mature CBMP2A or CBMP2B. CBMP2B differs from CBMP2A at only five residues in the active region. Recombinant versions of both CBMP2A and CBMP2B are active cross species, either as homodimers or in combination with OPl proteins. The recombinant data also indicates that the osteoinductive effect is not dependent on the presence of the entire mature form amino acid sequences of either subunit. Properly folded dimers comprising minimal structure, as short as 96 amino acids, are active. Furthermore, analogs of the active region, e.g., non-native forms never before known in nature, designed based on the observed homologies and known structure and properties of the native protein are capable of inducing bone formation. See, for example, COP5 and COP7 in U.S. Pat. No. 5,011,691. As far as applicants are aware, the biosynthetic constructs disclosed therein constitute the first instance of the design of a functional, active protein without preexisting knowledge of the active region of a native form nucleotide or amino acid sequence.

Further probing of mammalian cDNA libraries with sequences specific to hOPl also have identified, a sequence in mouse sharing almost complete identity with the mature hOPl amino acid sequence (approximately 98% homology with OP1-18). Additional probing in both human and mouse cDNA and genomic libraries also has identified OPl-like sequences herein referred to as “OP2” (“hOP2” or “mOP2”).

The OP2 proteins share significant amino acid sequence homology, approximately 74%, with the active region of the OPl proteins (e.g., OP7), and less homology with the intact mature form (e.g., OP1-18Ser—58% amino acid homology). Table I lists the OP1 and OP2 species identified to date. In its native form, the OP2 proteins share significant homology with various regulatory proteins on which the consensus probe was modeled. In particular, the proteins share significant homology in their C-terminal sequences, which comprise the active region of the osteogenic proteins. (Compare, for example, OP7 with DPP from Drosophila and Vg1 from Xenopus. See, for example, U.S. Pat. No. 5,011,691). In addition, these proteins share a conserved six or seven cysteine skeleton in this region (e.g., the linear arrangement of these C-terminal cysteine residues is conserved in the different proteins.) See, for example, OP7, whose sequence defines the seven cysteine skeleton, or OP2, whose sequence defines the six cysteine skeleton. In addition, the OP2 proteins contain an additional cysteine residue within this region.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OPl, OP2 NOMENCLATURE</strong></td>
</tr>
<tr>
<td><strong>hOPl</strong></td>
</tr>
<tr>
<td><strong>OP1</strong></td>
</tr>
<tr>
<td><strong>OP1-PP</strong></td>
</tr>
<tr>
<td><strong>OP1-18Ser</strong></td>
</tr>
<tr>
<td><strong>OPS</strong></td>
</tr>
<tr>
<td><strong>OP7</strong></td>
</tr>
<tr>
<td><strong>OP1-16Ser</strong></td>
</tr>
<tr>
<td><strong>OP1-16Leu</strong></td>
</tr>
</tbody>
</table>
TABLE 1-continued

| OP1, OP2 NOMENCLATURE | N-terminally truncated mature human OP1 protein species, Seq. ID No. 1, residues 315-431. N-terminal amino acid is methionine; protein migrates at 16 or 15 kDa on SDS-PAGE, depending on glycosylation pattern. Also referred to in related applications as "OP-16M." |
| OP1-16A | N-terminally truncated mature human OP1 protein species, Seq. ID No. 1, residues 315-431. N-terminal amino acid is alanine, protein migrates at 16 or 15 kDa on SDS-PAGE, depending on glycosylation pattern. Also referred to in related applications as "OP-16A." |
| OP1-16Ala | N-terminally truncated mature human OP1 protein species, Seq. ID No. 1, residues 315-431. N-terminal amino acid is alanine, protein migrates at 16 or 15 kDa on SDS-PAGE, depending on glycosylation pattern. Also referred to in related applications as "OP-16A." |
| OP1-16Val | N-terminally truncated mature human OP1 protein species, Seq. ID No. 1, residues 315-431. N-terminal amino acid is valine; protein migrates at 16 or 15 kDa on SDS-PAGE, depending on glycosylation pattern. Also referred to in related applications as "OP-16V." |
| mOP1 | DNA encoding mouse OP1 protein, Seq. ID No. 24. Also referred to in related applications as "mOP-1." |
| mOP1-PP | Prepro form of mouse protein, Seq. ID No. 24, residues 1-390. Also referred to in related applications as "mOP-1-PP." |
| mOP1-Ser | Mature mouse OP1 protein species (Seq. ID No. 24, residues 292-430). N-terminal amino acid is serine. Also referred to in related applications as "mOP-1." |
| mOP2 | DNA encoding mouse OP2 protein, Seq. ID No. 26. Also referred to in related applications as "mOP-2." |
| mOP2-PP | Prepro form of mouse OP2 protein, Seq. ID No. 26, residues 1-399. Also referred to in related applications as "mOP-2-PP." |
| mOP2-Ala | Mature mouse OP2 protein species, Seq. ID No. 26, residues 261-399. N-terminal amino acid is alanine. Also referred to in related applications as "mOP-2" and "mOP-2." |
| hOP2 | DNA encoding human OP2 protein, Seq. ID No. 28. Also referred to in related applications as "hOP-2." |
| hOP2-PP | Prepro form of human OP2 protein, Seq. ID No. 28, res. 1-402. Also referred to in related applications as "hOP-2-PP." |
| hOP2-Ala | Possible mature human OP2 protein species; Seq. ID No. 28, residues 264-402. Also referred to in related applications as "hOP-2." |
| hOP2-Pro | Possible mature human OP2 protein species; Seq. ID NO. 28, residues 267-402. N-terminal amino acid is proline. Also referred to in related applications as "hOP-2." |
| hOP2-Arg | Possible mature human OP2 protein species; Seq. ID No. 28, res. 270-402. N-terminal amino acid is arginine. Also referred to in related applications as "hOP-2." |
| hOP2-Ser | Possible mature human OP2 protein species; Seq. ID No. 28, res. 241-402. N-terminal amino acid is serine. Also referred to in related applications as "hOP-2." |

The invention thus provides recombinant dimeric proteins comprising any of the polypeptide chains described above, as well as allelic variants, and naturally-occurring or biosynthetic mutants thereof, and osteogenic devices comprising any of these proteins. In addition, the invention is not limited to these specific constructs. Thus, the osteogenic proteins of this invention comprising any of these polypeptide chains may include forms having varying glycosylation patterns, varying N-termini, a family of related proteins having regions of amino acid sequence homology which may be naturally occurring or biosynthetically derived, and active truncated or mutated forms of the native amino acid sequence, produced by expression of recombinant DNA in procaryotic or eucaryotic host cells. Active sequences useful in an osteogenic device of this invention are envisioned to include osteogenic proteins having greater than 60% identity, preferably greater than 65% identity, with the amino acid sequence of OPX. This family of proteins includes longer forms of a given protein, as well as allelic variants and biosynthetic mutants, including addition and deletion mutants, such as those which may alter the conserved C-terminal cysteine skeleton, provided that the alteration still allows the protein to form a dimeric species having a conformation capable of inducing bone formation in a mammal when implanted in the mammal in association with a matrix. Particularly envisioned within the family of related proteins are those proteins exhibiting osteogenic activity and wherein the amino acid changes from the OPX sequence include conservative changes, e.g., those as defined by Dayhoff, et al., Atlas of Protein Sequence and Structure; vol.5, Supp.3, pp.345-362, (M. O. Dayhoff,ed. Nat’l Biomed. Research Fdn., Washington, D.C., 1979.)

The novel polypeptide chains and the osteogenic proteins they comprise can be expressed from intact or truncated cDNA or from synthetic DNAs in procaryotic or eucaryotic host cells, and then purified, cleaved, refolded, dimerized, and implanted in experimental animals. Useful host cells include E.coli, Saccharomyces, the insect/baculovirus cell system, myeloma cells and mammalian cells. Currently preferred procaryotic host cells include E. coli. Currently preferred eucaryotic host cells include mammalian cells, such as Chinese hamster ovary (CHO) cells, or simian kidney cells (e.g., COS or BSC cells.) Thus, in view of this disclosure, skilled genetic engineers can isolate genes from cDNA or genomic libraries which encode appropriate amino acid sequences, modify existing sequences, or construct DNAs from oligonucleotides and then can express them in various types of procaryotic or eucaryotic host cells to produce large quantities of active proteins capable of inducing bone formation in mammals, including humans.

In one preferred aspect, the invention comprises dimeric osteogenic proteins and osteogenic devices containing these proteins, wherein the proteins comprise a polypeptide chain having an amino acid sequence sufficiently duplicative of the encoded amino acid sequence of Sequence ID No. 1 (hOP1) or 28 (hOP2) such that a dimeric protein comprising this polypeptide chain has a conformation capable of inducing endochondral bone formation when implanted in a mammal in association with a matrix. As used herein, the term “sufficiently duplicative” is understood to encompass all proteins capable of inducing endochondral bone formation when implanted in a mammal in association with a matrix and whose amino acid sequence comprises at least the conserved six cysteine skeleton and shares greater than 60% amino acid sequence identity in its active region with OPX.

In another preferred aspect, the invention comprises osteogenic proteins comprising species of polypeptide chains having the generic amino acid sequence herein referred to as "OPX" which accommodates the homologies between the various identified species of these osteogenic OP1 and OP2 proteins, and which is described by the amino acid sequence of Sequence ID No. 30.
The identification of mOP2 and hOP2 represents the discovery of osteogenic proteins having an additional cysteine residue in their active region in addition to the conserved six cysteine skeleton defined by OP5, or the conserved seven cysteine skeleton defined by OP7. Thus, in another aspect, the invention comprises species of polypeptide chains herein referred to as “OPX-7C”, comprising the conserved six cysteine skeleton plus the additional cysteine residue identified in the OP2 proteins, and “OPX-8C”, comprising the conserved seven cysteine skeleton plus the additional cysteine residue identified in the OP2 proteins. The OPX-7C and OPX-8C amino acid sequences are described in Seq. ID Nos. 31 and 32, respectively. Each Xaa in Seq. ID Nos. 31 and 32 independently represents one of the 20 naturally occurring L-isomer, α-amino acids or a derivative thereof which, together with the determined cysteine residues, define a polypeptide chain such that dimeric osteogenic proteins comprising this polypeptide chain have a conformation capable of inducing endochondral bone formation when implanted in a mammal in association with a matrix.

In still another preferred aspect, the invention comprises nucleic acids and the osteogenically active polypeptide chains encoded by these nucleic acids which hybridize to DNA or RNA sequences encoding the active region of OP1 or OP2 under stringent hybridization conditions. As used herein, stringent hybridization conditions are defined as hybridization in 40% formamide, 5×SSPE, 5× Denhardt’s Solution, and 0.1% SDS at 37° C. overnight, and washing in 0.1×SSPE, 0.1% SDS at 50° C.

The invention further comprises nucleic acids and the osteogenically active polypeptide chains encoded by these nucleic acids which hybridize to the “pro” region of the OP1 or OP2 proteins under stringent hybridization conditions. As used herein, “osteogenically active polypeptide chains” is understood to mean those polypeptide chains which, when dimerized, produce a protein species having a conformation such that the pair of polypeptide chains is capable of inducing endochondral bone formation in a mammal when implanted in a mammal in association with a matrix.

The proteins of this invention, including fragments thereof, also may be used to raise monoclonal or polyclonal antibodies capable of binding specifically to an epitope of the osteogenic protein. These antibodies may be used, for example, in osteogenic protein purification protocols.

The osteogenic proteins are useful in clinical applications in conjunction with a suitable delivery or support system (matrix). As disclosed herein, the matrix may be combined with osteogenic protein to induce endochondral bone formation reliably and reproducibly in a mammalian body. The matrix is made up of particles of porous materials. The pores must be of a dimension to permit progenitor cell migration into the matrix and subsequent differentiation and proliferation. The particle size should be within the range of 70 μm–850 μm, preferably 70 μm–420 μm, most preferably 150 μm–420 μm. It may be fabricated by close packing particulate material into a shape spanning the bone defect, or by otherwise structuring as desired a material that is biocompatible, and preferably biodegradable in vivo to serve as a “temporary scaffold” and substratum for recruitment of migratory progenitor cells, and as a base for their subsequent anchoring and proliferation. Useful matrix materials comprise, for example, collagen; homopolymers or copolymers of glycolic acid, lactic acid, and butyric acid, including derivatives thereof, and ceramics, such as hydroxyapatite, tricalcium phosphate and other calcium phosphates. Combinations of these matrix materials also may be useful.

Currently preferred carriers include particulate, demineralized, guanidine extracted, species-specific (allogenic) bone, and specially treated particulate, protein extracted, demineralized, xenogenic bone. Optionally, such xenogenic bone powder matrices also may be treated with proteases such as trypsin. Preferably, the xenogenic matrices are treated with one or more fibril modifying agents to increase the intraparticle intrusion volume (porosity) and surface area. Useful agents include solvents such as dichloromethane, trichloroacetic acid, acetonitrile and acids such as trifluoroacetic acid and hydrogen fluoride.

The currently preferred fibril-modifying agent useful in formulating the matrices of this invention is a heated aqueous medium, preferably an acidic aqueous medium having a pH less than about pH 4.5, most preferably having a pH within the range of about pH 2-4. A currently preferred heated acidic aqueous medium is 0.1% acetic acid which has a pH of about 3. Heating demineralized, delipidated, guanidine-extracted bone collagen in an aqueous medium at elevated temperatures (e.g., in the range of about 37° C.–65° C., preferably in the range of about 45° C.–60° C.) for approximately one hour generally is sufficient to achieve the desired surface morphology. Although the mechanism is not clear, it is hypothesized that the heat treatment alters the collagen fibrils, resulting in an increase in the particle surface area. Thus, one aspect of this invention includes osteogenic devices comprising matrices which have been treated to increase the surface area and porosity of matrix collagen particles substantially.

Examination of solvent-treated bone collagenous matrix shows that demineralized guanidine-extracted xenogenic bovine bone comprises a mixture of additional materials and that extracting these materials can improve matrix properties. Chromatographic separation of components in the extract, followed by addition back to active matrix of the various extract fractions corresponding to the chromatogram peaks, indicates that there is a fraction which can inhibit the osteoinductive effect. The identity of the substance or substances in this inhibiting fraction has not as yet been determined. Thus, in one aspect of this invention, a matrix is provided comprising treated Type-I bone collagen particles of the type described above, further characterized in that they are depleted in this inhibiting component.

In still another aspect of this invention, a matrix is provided that is substantially depleted in residual heavy metals. Treated as disclosed herein, individual heavy metal concentrations in the matrix can be reduced to less than about 1 ppm.

In view of this disclosure, one skilled in the art can create a bio compatible matrix of choice having a desired porosity or surface microtexture useful in the production of osteogenic devices, and useful in other implantable contexts, e.g., as a packing to promote bone induction, or as a biodegradable sustained release implant. In addition, synthetically formulated matrices, prepared as disclosed herein, may be used.

The osteogenic proteins and implantable osteogenic devices disclosed herein will permit the physician to obtain predictable bone formation to correct, for example, acquired and congenital craniofacial and other skeletal or dental anomalies (e.g., Glowacki et al. (1981) Lancet 1:959–963). The devices may be used to induce local endochondral bone formation in non-union fractures as demonstrated in animal tests, and in other clinical applications including dental and periodontal applications where bone formation is required. Another potential clinical application is in cartilage repair, for example, in the treatment of osteoarthritis.
BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects of the invention, the various features thereof, as well as the invention itself, may be more fully understood from the following description, when read together with the accompanying drawings, in which:

FIG. 1A-1B are a flow diagram of a purification procedure for isolating osteogenic protein, illustrating purification steps from grinding cortical bone through lyophilization of guanidine-extracted material (1A), and urea solubilization through gel slicing (1B);

FIG. 2A-2D are a collection of plots of protein concentration (as indicated by optical absorption) vs elution volume illustrating the results of BOP fractionation during purification on (2A) heparin-Sepharose-I; (2B) HAP-Ultragel; (2C) TSK 3000; and (2D) heparin-Sepharose-II. Asterisk identifies active peak;

FIG. 3A-3B are a photographic reproduction of a Coomassie blue stained SDS polyacrylamide gel of the osteogenic protein under non-reducing (3A) and reducing (3B) conditions;

FIG. 4A-4B are a photographic reproduction of a Coom A blot of an SDS polyacrylamide gel showing the presence of a carbohydrate component in the oxidized (4A) and reduced (4B) 30 kDa protein;

FIG. 5A-5D are photographic reproductions of autoradiograms of an SDS polyacrylamide gel of peptides produced upon the digestion of the 30 kDa osteogenic protein with V-8 protease (5A), Endo Lys C protease (5C), reducing conditions (5D); deglycosylated and run under non-reducing conditions (5C); glycosylated and run under non-reducing conditions (5B); glycosylated and run under reducing conditions (5C); deglycosylated and run under reducing conditions (5D);

FIG. 6A-6E are a photographic reproduction of an autoradiogram of an SDS polyacrylamide gel of peptides produced upon the digestion of the 30 kDa osteogenic protein with trypsin (6D), and trypsin (6E). (6A) is control; protein with V-8 protease (6B), Endo Lys C protease (6C), reducing conditions (5D);

FIG. 7A-7C are a collection of HPLC chromatograms of tryptic peptide digestions of 30kDa BOP (7A), the 16 kDa subunit (7B), and the 18 kDa subunit (7C);

FIG. 8 is an HPLC chromatogram of an elution profile on reverse phase C-18 HPLC of the samples recovered from the second heparin-Sepharose chromatography step (see FIG. 2D). Superimposed is the percent bone formation in each fraction;

FIG. 9 is a gel permeation chromatogram of an elution profile on TSK 3000/2000 gel of the C-18 purified osteogenic peak fraction. Superimposed is the percent bone formation in each fraction;

FIG. 10-10D are a collection of graphs of protein concentration (as indicated by optical absorption) vs elution volume illustrating the results of human osteogenic protein fractionation on heparin-Sepharose I (10A), HAP-Ultragel (10B), TSK 3000/2000 (10C), and heparin-Sepharose II (10D). Arrows indicate buffer changes and asterisk identifies active peak;

FIG. 11 is a graph showing representative dose response curves for bone-inducing activity in samples from various purification steps including reverse phase HPLC on C-18 (A), heparin-Sepharose II (B), TSK 3000 (C), HAP-ultragel (D), and heparin-Sepharose I (E);

FIG. 12 is a bar graph of radiomorphometric analyses of feline bone defect repair after treatment with an osteogenic device (A), carrier control (B), and demineralized bone (C);

FIGS. 13A-13B are a schematic representation of the DNA sequence, restriction sites, and corresponding amino acid sequence of the consensus gene/probe for osteogenic protein, as follows: (13A) nucleotides 1–192; (13B) nucleotides 193–314;

FIG. 14 is a graph of osteogenic activity vs. increasing molecular weight showing peak bone forming activity in the 30 kDa region of an SDS polyacrylamide gel;

FIG. 15 is a photographic representation of a Coomassie blue stained SDS gel showing gel purified subunits of the 30 kDa protein;

FIGS. 16A-16B are a pair of HPLC chromatograms of Endo Asp N proteinase digests of the trypsin-resistant cores from the 18 kDa subunit (16A) and the 16 kDa subunit (16B);

FIGS. 17A-17C are photographic representations of the histological examination of bone implants in the rat model: carrier alone (17A); carrier and glycosylated osteogenic protein (17B); and carrier and deglycosylated osteogenic protein (17C). Arrows indicate osteoblasts;

FIG. 18 is a representation of the hybridization of the consensus gene/probe to the OP1 gene;

FIGS. 19A-19F are restriction maps of various expression vectors designed for the mammalian cell expression of OP1 as follows: (19A) vector pH717; (19B) vector pH731; (19C) vector pH754; (19D) vector pH752; (19E) vector pW24; (19F) vector pH783;

FIGS. 20A-20F are photoreproductions of Western blots (immunoblots) comparing OP1 expressed from pH717/COS cells (20A); pH731/COS cells (20B); pH754/CHO cells (20C); pH752/CHO cells (20D); pH717/BSC cells (20E); and pW24/BSC cells (20F);

FIGS. 21A-21F are elution profiles and photoreproductions of SDS-PAGE gels expressed from BSC cells and purified (in order) on: S-Sepharose—elution profile (21A); SD-PAGE gel (21B); phenyl-Sepharose—elution profile (21C); SDS-PAGE gel (21D); and C-18 columns—elution profile (21E); and SDS-PAGE gel (21F);

FIG. 22 is a photoreproduction of SDS-PAGE gels of OP1 purified from BSC cells, comparing the intact dimer under oxidized conditions (36 kDa, lane 1) and the corresponding monomer, after reduction with dithiothreitol (18kDa, lane 5), with molecular weight standards (lanes 2–4);

FIGS. 23A-23E compare the amino acid sequences of the mature hOP1 and mOP1 polypeptide chains: OP1-18Ser and mOP1-Ser; and mature mOP2 and hOP2 polypeptide chains: hOP2-Ala and mOP2-Ala, as follows: (23A) residues 1–72 of hOP1-Ser, mOP1-Ser; (23B) residues 73–139 of hOP1-Ser, mOP1-Ser; (23C) residues 1–63 of hOP2-Ala, mOP2-Ala; (23D) residues 64–126 of hOP2-Ala, mOP2-Ala; (23E) residues 127–139 of hOP2-Ala, mOP2-Ala;

FIGS. 24A-24D compare the amino acid sequences of the mature OP1 and OP2 polypeptide chains: OP1-18Ser, mOP1-Ser; hOP2-Ala and mOP2-Ala, as follows: (24A) residues 1–45; (24B) residues 46–90; (24C) residues 91–134; (24D) residues 135–139;

FIGS. 25A-25D are scanning electron micrographs (approx. 1000x) of demineralized, delipidated bovine bone matrix heat treated in water at (25A) 37° C., (25B) 45° C., (25C) 55° C., and (25D) 65° C.;

FIGS. 26A and 26B are scanning electron micrographs (5000x) of demineralized, delipidated 26A rat bone collagen particles, and (26B) bovine bone collagen particles;

FIG. 27 is a 214 nm absorbance tracing of the extract isolated from hot water-treated bovine matrix, identifying the inhibitory effect of individual fractions on in vivo bone formation;

FIGS. 28A and 28B are bar graphs showing the inhibitory effect of hot water-treated matrix extract on OP1 activity, as
measured by (28A) alkaline phosphatase activity and (28B) calcium content in day 12 implants, vs. increasing concentration of extract solvents;

FIGS. 29A-29F are photomicrographs (220x) of allogenetic implants of OP1 expressed from COS, BSC and CHO cells, as follows: (29A) control; (29B) 500 ng BSC-produced OP1; (29C) 220 ng COS-produced OP1; (29D) CHO-produced OP1, 220X; (29E) CHO-produced OP1, 440X; (29F) 500 ng BSC-produced OP1;

FIG. 30 is a photomicrograph showing the histology (day 12) of a xenogenic implant of this invention using OP1 expressed from BSC cells and hot water-treated xenogenic bovine matrix;

FIG. 31 describes the dose dependence of osteogenic implants for day 12 implants, as determined by alkaline phosphatase activity and calcium content, for allogenetic implants containing OP1 expressed from COS, BSC and CHO cells;

FIGS. 32A and 32B are bar graphs showing the dose dependence of OP1 expressed in COS and BSC cells, as measured by (32A) alkaline phosphatase activity and (32B) calcium content in xenogenic implants (day 12), vs increasing concentration of protein (dose curve in ng); and

FIG. 33 compares the N-termini of the various forms of human OP1 protein identified to date.

DESCRIPTION

Purification protocols first were developed which enabled isolation of the osteogenic protein present in crude protein extracts from mammalian bone (e.g., from bovine bone, “BOPI,” and human bone. See U.S. Ser. No. 179,406 filed Apr. 8, 1988, now U.S. Pat. No. 4,968,590). Sequence data obtained from the bovine material suggested a probe design which was used to isolate human genes. The human counterpart osteogenic proteins have now been expressed and extensively characterized.

These discoveries have enabled preparation of DNAs encoding totally novel, non-native (e.g., not known to occur in nature) protein constructs which individually as homodimers and combined with other related species are capable of producing true endochondral bone (see U.S. Pat. No. Ser. No. 315,342, filed Feb. 23, 1989, now U.S. Pat. No. 5,011,691). They also permitted expression of the natural protein expression and further characterization. The host known in the art. Vectors containing sequences of interest may be a procaryotic or eucaryotic cell since the former’s ability to glycosylate protein will not destroy the protein’s osteogenic activity. Useful host cells include E. coli, Saccharomyces, the insect/baculovirus cell system, myeloma cells, and various other mammalian cells. The vector additionally may include various sequences to promote correct expression of the recombinant protein, including transcription promoter and termination sequences, enhancer sequences, preferred ribosome binding site sequences, preferred mRNA leader sequences, preferred signal sequences for protein secretion, and the like. The DNA sequence encoding the protein of interest also may be manipulated to remove potentially inhibiting sequences or to minimize...
unwanted secondary structure formation. The recombinant osteogenic protein also may be expressed as a fusion protein. After being translated, the protein may be purified from the cells themselves or recovered from the culture medium. All biologically active protein forms comprise dimeric species linked by disulfide bonds or otherwise associated, produced by oxidizing and refolding one or more of the various recombinant polypeptide chains within an appropriate eucaryotic cell or in vitro after expression of individual subunits. A detailed description of osteogenic protein purified from natural sources or expressed from recombinant DNA in E. coli and numerous different mammalian cells is disclosed below.

In view of this disclosure, and using standard immunology techniques well known in the art, those skilled in the art also may raise polyclonal or monoclonal antibodies against all or part of the polypeptide chains described herein. Useful protocols for antibody production may be found, for example, in Molecular Cloning—A Laboratory Manual (Sambrook et al., eds.) Cold Spring Harbor Press, 2nd ed., 1989). See Book 3, Section 18. The polypeptide chains useful as antigens may be purified from natural-sourced material, synthesized by chemical means, or expressed from recombinant nucleic acid as disclosed herein. Antibodies specific for the osteogenic proteins disclosed herein may be particularly useful in osteogenic protein preparation. For example, when purifying a given osteogenic protein from bone or a cell culture supernatant, the osteogenic protein may be selectively extracted from a mixture by exposing the mixture to the antibody under conditions such that the antibody specifically binds the osteogenic protein to form an antibody-osteogenic protein complex. This complex then may be separated from the mixture by conventional methods, and the complex dissociated to yield substantially purified osteogenic protein.

1. PURIFICATION OF OSTEOCGENIC PROTEIN FROM BONE

A. Bovine Bone

1. Purification

1.1 Preparation of Demineralized Bone

A schematic representation of the general protocol disclosed herein for purifying osteogenic protein from bone is illustrated in FIG. 1. Demineralized bovine bone matrix is prepared by previously published procedures (Sampath and Reddi (1983) Proc. Natl. Acad. Sci. USA 80:6591–6595). Bovine diaphyseal bones (age 1–10 days) are obtained from a local slaughterhouse and used fresh. The bones are stripped of muscle and fat, cleaned of periosteum, demineralized by pressure with cold water, dipped in cold absolute ethanol, and stored at -20° C. They are then dried and fragmented by crushing and pulverized in a large mill. Care is taken to prevent heating by using liquid nitrogen. The suspension is filtered through cheese cloth and stripped of muscle and fat, cleaned of periosteum, demineralized by pressure with cold water, dipped in cold absolute ethanol, and stored at -20° C. The suspension is filtered through cheese cloth and centrifuged at 20,000xg at 4° C. The supernatant is collected and concentrated to one volume using an Amicon ultrafiltration YM-10 hollow fiber membrane. The concentrate is centrifuged (40,000xg for 30 min. at 4° C.), and the supernatant is then subjected to ethanol precipitation. To one volume of concentrate is added seven volumes of cold (-20° C.) absolute ethanol (100%), which is then kept at -20° C. for 30 min. The precipitate is pelleted upon centrifugation at 10,000xg for 10 min. at 4° C. The resulting pellet is resuspended in 250 ml of 85% cold ethanol and recentrifuged. The pellet then is lyophilized.

1.3 Heparin-Sepharose Chromatography I

The ethanol precipitated, lyophilized, extracted crude protein is dissolved in 20 volumes of 6 M urea, 50 mM Tris-HCl, pH 7.0 (Buffer A) containing 0.15 M NaCl, and clarified by centrifugation at 20,000xg for 30 min. The supernatant is stirred for 16 hr. with 50 volumes of hydrated heparin-Sepharose (Pharmacia) equilibrated with Buffer A containing 0.15 M NaCl. The heparin-Sepharose is pre-treated with Buffer A containing 1.0 M NaCl prior to equilibration. The unabsorbed protein is collected by packing the resin into a column. After washing with three column volumes of initial buffer (Buffer A containing 0.15 M NaCl), protein is eluted with Buffer A containing 0.5 M NaCl. The absorption of the eluate is monitored continuously at 280 nm. The pool of protein eluted by 0.5 M NaCl (approximately 20 column volumes) is collected and stored at -20° C.

As shown in FIG. 2A, most of the protein (about 95%) remains unbound. Approximately 5% of the protein is bound to the column. The unbound fraction has no bone inductive activity when bioassayed as a whole or after a partial purification through Sepharose CL-6B.

1.4 Hydroxyapatite-Ultragel Chromatography

The volume of protein eluted by Buffer A containing 0.5 M NaCl from the heparin-Sepharose is applied directly to a column of hydroxyapatite-Ultragel (HAP-Ultragel) (LKB Instruments), and equilibrated with Buffer A containing 0.5 M NaCl. The HAP-Ultragel is treated with Buffer A containing 500 mM Na phosphate prior to equilibration. The unadsorbed protein is collected as an unbound fraction, and the column is washed with three column volumes of Buffer A containing 0.5 M NaCl. The column subsequently is eluted with Buffer A containing 100 mM Na phosphate (FIG. 2B). The approximately 3 column volume pool of the protein peak eluted by 100 mM Na phosphate is concentrated using an Amicon ultrafiltration YM-10 membrane to one volume, dialyzed in a 3.5 kDa molecular weight cut-off bag (Spectrapor) against distilled water, and lyophilized.

The 100 mM Na phosphate eluted component can induce endochondral bone as measured by alkaline phosphatase activity and histology (see section V.5.1, infra). As the biologically active protein is bound to HAP in the presence of 6 M urea and 0.5 M NaCl, it is likely that the protein has an affinity for bone mineral and may be displaced only by phosphate ions.
1.5 TSK 3000 Gel Exclusion Chromatography

Analytical TSK 3000 gel (silica gel), obtained from Bio Rad, is equilibrated with 4 M guanidine-HCl, 50 mm Tris-HCl, pH 7.0. A pre-column (analytical) also is used. A portion of the lyophilized protein from HAP-Ultrogel is dissolved in a known volume of 4 M guanidine-HCl, 50 mM Tris-HCl, pH 7.0, and the solution is clarified by low speed centrifugation. A 200 μl sample containing approximately 10 mg of protein is loaded onto the column and then chromatographed with 4 M guanidine-HCl, 50 mM Tris-HCl, pH 7.0, with a flow rate of 0.3 ml/min. 0.6 ml fractions are collected over 100 min., and the concentration of the protein is measured continuously at A280. Fractions are collected and bioassayed as described below; fractions having a molecular weight less than 35 kDa (30–34 kDa) and osteoinductivity are pooled and stored at 4°C. (FIG. 2C).

1.6 Heparin-Sepharose Chromatography-II

The pooled osteo-inductive fractions from TSK gel exclusion chromatography are dialyzed extensively against distilled water and then against one liter of 6 M urea, 50 mM Tris-HCl, pH 7.0 (Buffer A, also referred to in related applications as “Buffer B”). The dialysate then is dialyzed against centration centrifugation, and the supernatant is stirred for one hr. with 50–100 μl of hydrated heparin-Sepharose (Pharmacia) equilibrated with Buffer A. The heparin-Sepharose is pre-treated with Buffer A containing 1.0 M NaCl prior to equilibration. The unadsorbed protein is collected by packing the resin into a column as an unbound fraction. After washing with three column volumes of initial buffer, the column is developed sequentially with Buffer A containing NaCl, 0.15 M NaCl, and 0.5 M NaCl (see FIG. 2D). The protein eluted by 0.5M NaCl is collected and dialedyzed extensively against distilled water. Then it is dialyzed against one liter of 0.1% trifluoroacetic acid at 4°C.

1.7 Reverse Phase HPLC

The protein further is purified by C-18 Vydac silica-based HPLC column chromatography (particle size 5 μm; pore size 300 Å). The osteoinductive fraction obtained from heparin-Sepharose-II chromatography is concentrated, loaded onto the column, and washed in 0.1% TFA, 10% acetonitrile for five min. The bound proteins are eluted with a linear gradient of 10-30% acetonitrile over 15 min., 30-50% acetonitrile over 60 min, and 50-70% acetonitrile over 15 min. at 22°C with a flow rate of 1.0 ml/min, and 1.0 ml samples are collected in polycarbonate tubes. Protein is monitored by absorbance at 214 nm (see FIG. 8). Column fractions are tested for the presence of concanavalin A (Con A)-blottable 30 kDa protein and then pooled. Pools then are characterized biochemically for the presence of 30 kDa protein by autoradiography, concanavalin A blotting, and Coomassie blue dye staining. They are then assayed for in vivo osteogenic activity. Biological activity is not found in the absence of 30 kDa protein.

1.8 Gel Elution

The glycosylated or unglycosylated protein then is eluted from SDS gels for further characterization. 125I-labelled 30 kDa protein routinely is added to each preparation to monitor yields. TABLE 2 shows the various elution buffers that have been tested and the yields of 125I-labelled protein.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>% Eluted</th>
</tr>
</thead>
<tbody>
<tr>
<td>deionized H2O</td>
<td>22</td>
</tr>
<tr>
<td>Guanidine-HCl, Tris-HCl, pH 7.0</td>
<td>2</td>
</tr>
</tbody>
</table>

TABLE 3 lists the steps used to isolate the 30 kDa or 27 kDa gel-bound protein. The standard protocol uses diffusion elution in Tris-HCl buffer containing 0.1% SDS to achieve greater than 95% elution of the protein from the 27 or 30 kDa region of the gel.

TABLE 3

Preparation of Gel Eluted Protein (C-18 Pool or deglycosylated protein plus 125I-labelled 30 kDa protein)

1. Dry using vacuum centrifugation;
2. Wash pellet with H2O;
3. Dissolve pellet in gel sample buffer (no reducing agent);
4. Electrophorese on pre-electrophoresed 0.5 mm mini gel;
5. Cut out 27 or 30 kDa protein;
6. Elute from gel with 0.1% SDS, 50 mM Tris-HCl, pH 7.0;
7. Filter through Centricon membrane;
8. Concentrate in Centricon tube (10 kDa membrane);
9. Chromatograph on TSK-3000 gel filtration column;

Chromatography in 0.1% SDS on a TSK-3000 gel filtration column is performed to separate gel impurities, such as soluble acrylamide, from the final product. The overall yield of labelled 30 kDa protein from the gel elution protocol is 50–60% of the loaded sample. Most of the loss occurs in the electrophoresis step, due to protein aggregation and/or smearing. In a separate experiment, a sample of gel eluted 30 kDa protein is reduced, electrophoresed on an SDS gel, and transferred to an Immobilon membrane. The membrane is stained with Coomassie blue dye, cut into slices, and the slices are counted. Coomassie blue dye stains the 16 kDa and 18 kDa reduced species almost exclusively. However, the counts showed significant smearing throughout the gel in addition to being concentrated in the 16 kDa and 18 kDa species. This suggests that the 125I-label can exhibit anomalous behavior on SDS gels and cannot be used as an accurate marker for cold protein under such circumstances.

The yield is 0.5 to 1.0 μg substantially pure osteogenic protein per kg of bone.

1.9 Isolation of the 16 kDa and 18 kDa Species

TABLE 4 summarizes the procedures involved in the preparation of the subunits. Gel eluted 30 kDa protein (FIG. 3) is carboxymethylated and electrophoresed on an SDS-gel. The sample contains 125I-label to trace yields and to use as an indicator for slicing the 16 kDa and 18 kDa regions from the gel. FIG. 15 shows a Coomassie stained gel of aliquots of the protein isolated from the different gel slices. The slices corresponding to the 16 kDa, 18 kDa and non-reducible 30 kDa species contained approximately 10 μg, 3–4 μg, and 6–8 μg, of protein respectively, as estimated by staining intensity. Prior to SDS electrophoresis, all of the 30 kDa species can be reduced to the 16 kDa and 18 kDa region.

125I-labelled 30 kDa protein

<table>
<thead>
<tr>
<th>Buffer</th>
<th>% Eluted</th>
</tr>
</thead>
<tbody>
<tr>
<td>deionized H2O</td>
<td>22</td>
</tr>
<tr>
<td>Guanidine-HCl, Tris-HCl, pH 7.0</td>
<td>2</td>
</tr>
</tbody>
</table>
reducible 30 kDa species observed after electrophoresis appears to be an artifact resulting from the electrophoresis procedure.

2. Characterization of Naturally-Sourced bOP

2.1 Molecular Weight and Structure

Electrophoresis of these fractions on non-reducing SDS polyacrylamide gels reveals a single band at about 30 kDa as detected by both Coomassie blue staining (FIG. 3A) and autoradiography.

In order to extend the analysis of bOP, the protein was examined under reducing conditions. FIG. 3B shows an SDS gel of bOP in the presence of dithiothreitol. Upon reduction, 30 kDa bOP yields two species which are stained with Coomassie blue dye: a 16 kDa species and an 18 kDa species. Reduction causes loss of biological activity. Methods for the efficient elution of the proteins from SDS gels have been tested, and a protocol has been developed to achieve purification of both proteins. The two reduced bOP species have been analyzed to determine if they are structurally related. Comparison of the amino acid Composition of the two species (as disclosed below) shows little differences, indicating that the native protein may comprise two chains having some homology.

2.2 Charge Determination

Isoelectric focusing studies are carried out to further evaluate the 30 kDa protein for possible heterogeneity. The oxidized and reduced species migrate as diffuse bands in the basic region of the isoelectric focusing gel, using the iodinated 30 kDa protein for detection. Using two dimensional gel electrophoresis and Con A for detection, the oxidized 30 kDa protein shows a diffuse species migrating in the same basic region as the iodinated 30 kDa protein. The diffuse character of the band may be traced to the presence of carbohydrate attached to the protein.

2.3 Presence of Carbohydrate

The 30 kDa protein has been tested for the presence of carbohydrate by Con A blotting after SDS-PAGE and transfer to nitrocellulose paper. The results demonstrate that the 30 kDa protein has a high affinity for Con A, indicating that the protein is glycosylated (FIG. 4A). In addition, the Con A blots provide evidence for a substructure in the 30 kDa region of the gel, suggesting heterogeneity due to varying degrees of glycosylation. After reduction (FIG. 4B), Con A blots show evidence for two major components at 16 kDa and 18 kDa. In addition, it has been demonstrated that no glycosylated material remains at the 30 kDa regions after reduction.

In order to confirm the presence of carbohydrate and to estimate the amount of carbohydrate attached, the 30 kDa protein is treated with N-glycanase, a deglycosylating enzyme with a broad specificity. Samples of the 125I-labelled 30 kDa protein are incubated with the enzyme in the presence of SDS for 24 hours at 37° C. As observed by SDS-PAGE, the treated samples appear as a prominent species at about 27 kDa (FIG. 5B). Upon reduction, the 27 kDa species is reduced to species having a molecular weight of about 14 kDa-16 kDa (FIG. 5B).

Because the use of N-glycanase for producing deglycosylated protein samples for sequencing or biological activity testing is not advantageous, chemical cleavage of the carbohydrate moieties using hydrogen fluoride (HF) is performed.

Active osteogenic protein fractions pooled from the C-18 chromatography step are derived in vacuo over P2O5 in a polypropylene tube, and 50 μl freshly distilled anhydrous HF at -70° C. is added. After capping the tube tightly, the mixture is kept at 0° C. in an ice-bath with occasional agitation for 1 hr. The HF is then evaporated using a continuous stream of dry nitrogen gas. The tube is removed from the ice bath and the residue dried in vacuo over P2O5 and KOH pellets.

Following drying, the samples are dissolved in 100 μl of 50% acetonitrile/0.1% TFA and aliquoted for SDS gel analysis, Con A binding, and biological assay. Aliquots are dried and dissolved in either SDS gel sample buffer in preparation for SDS gel analysis and Con A blotting, or 4 M guanidine-HCl, 50 mM Tris-HCl, pH 7.0 for biological assay. The deglycosylated protein produces a bone formation response in the in vivo rat model described below as determined by histological examination (FIG. 17C).

The results show that samples are completely deglycosylated by the HF treatment: Con A blots after SDS gel electrophoresis and transfer to Immobilon membrane show no binding of Con A to the treated samples, while untreated controls are strongly positive at 30 kDa. Coomassie gels of treated samples show the presence of a 27 kDa band instead of the 30 kDa band present in the untreated controls.

2.4 Chemical and Enzymatic Cleavage

Cleavage reactions with CNBr are analyzed using Con A binding for detection of fragments associated with carbohydrate. Cleavage reactions are conducted using trifluoroacetic acid (TFA) in the presence and absence of CNBr. Reactions are conducted at 37° C. for 18 hours, and the samples are vacuum dried. The samples are washed with water, dissolved in SDS gel sample buffer with reducing agent, boiled and applied to an SDS gel. After electrophoresis, the protein is transferred to Immobilon membrane and visualized by Con A binding. In low concentrations of acid (1%), CNBr cleaves the majority of 16 kDa and 18 kDa species to one product, a species about 14 kDa. In reactions using 10% TFA, a 14 kDa species is observed both with and without CNBr.

Four proteolytic enzymes are used in these experiments to examine the digestion products of the 30 kDa protein: 1) V-8 protease; 2) Endo Lys C protease; 3) pepsin; and 4) trypsin. Except for pepsin, the digestion buffer for the enzymes is 0.1 M ammonium bicarbonate, pH 8.3. The pepsin reactions are done in 0.1% TFA. The digestion volume is 100 μl and the ratio of enzyme to substrate is 1:10. 125I-labelled 30 kDa bOP is added for detection. After incubation at 37° C. for 16 hr., digestion mixtures are dried down and taken up in gel sample buffer containing dithiothreitol for SDS-PAGE. FIG. 6 shows an autoradiograph of an SDS gel of the digestion products. The results show that under these conditions, only trypsin digests the reduced 16 kDa/18 kDa species com-
pleteness and yields a major species at around 12 kDa. Pepsin digestion yields better defined, lower molecular weight species. However, the 16 kDa/18 kDa fragments were not digested completely. The V-8 digest shows limited digestion with one dominant species at 16 kDa.

2.5 Protein Sequencing

To obtain amino acid sequence data, the protein is cleaved with trypsin. The tryptic digest of reduced and carboxymethylated 30 kDa protein (approximately 10 μg) is fractionated by reverse-phase HPLC using a C-8 narrowbore column (13 cm x 2.1 mm ID) with a TFA/acetonitrile gradient and a flow rate of 150 μl/min. The gradient employs (A) 0.06% TFA in water and (B) 0.04% TFA in water and acetonitrile (1:4; v:v). The procedure is 10% B for five min., followed by a linear gradient for 70 min. to 80% B, followed by a linear gradient for 10 min. to 100% B. Fractions containing fragments as determined from the peaks in the HPLC profile (FIG. 7A) are rechromatographed at least once under the same conditions in order to isolate single components satisfactory for sequence analysis.

The HPLC profile of the similarly digested 16 kDa and 18 kDa subunits are shown in FIGS. 7B and 7C, respectively. These peptide maps are similar, suggesting that the subunits are identical or are closely related.

The trypsin resistant core material of the 16 kDa and 18 kDa subunits is digested with Endo Asp N proteinase. The core protein is treated with 0.5 μg Endo Asp N in 50 mM sodium phosphate buffer, pH 7.8 at 36°C for 2 hr. Subsequently, the samples are centrifuged, and the water soluble peptides injected into the narrow bore HPLC. The water insoluble peptides also are subjected to HPLC fractionation after being dissolved in 50% acetonitrile/0.1% TFA. The conditions for fractionation are the same as those described previously for the 30 kDa, 16 kDa, and 18 kDa digests. The profiles obtained are shown in FIGS. 16A and 16B.

Several of the peptide fragments produced using the foregoing procedures have been analyzed in an automated amino acid sequencer (Applied Biosystems 450A). The following sequence data has been obtained:

(1) Ser-Phe-Asp-Ala-Tyr-Tyr-Cys-Ser-Gly-Ala-Cys-Gln-Phe-Pro-Met-Pro-Lys;
(2) Ser-Leu-Lys-Pro-Ser-Asn-Tyr-Ala-Thr-Ile-Gln-Ser-Ile-Val;
(3) Ala-Cys-Cys-Val-Pro-Thr-Glu-Leu-Ser-Ala-Ile-Ser-Met-Leu-Tyr-Leu-Asp-Glu-An-Ala-Lys;
(4) Met-Ser-Ser-Leu-Ser-Ile-Leu-Phe-Asp-Glu-Aen-Lys;
(5) Val-Gly-Val-Val-Pro-Gly-Ile-Pro-Glu-Pro-Cys-Val-Pro-Glu;
(6) Val-Asp-Asp-Ala-Asp-Ile-Gly
(7) Val-Pro-Lys-Pro;
(8) Ala-Pro-Thr.

Several of the residues in these sequences could not be 5

Sequencing data suggested these residues were both serines, 6

but subsequent experiments identified the residues as cysteines. Accordingly, these data have been eliminated from the sequencing results presented here. Similarly, a spurious glutamic acid residue at the N-terminus of fragment 7, and a spurious lysine at the C-terminus of fragment 8 also have been eliminated from the data presented here (see U.S. Pat. No. 5,011,691, col. 7, fragment 7 for correct sequence).

Fragments 1, 2, and 4-6 are described in the sequences presented in Seq. ID Nos. 20 and 22 (referred to herein as human and marine “CBMP3,” respectively.) Specifically, fragment 1 is described essentially by residues 93–109 of Seq. ID No. 20 and fragment 2 is described essentially by residues 121–134 of Seq. ID No. 22 (note that residue 7 in fragment 2 is identified as a tyrosine. In Seq. ID No. 22 this residue is a histidine. By comparison with the CBMP2 and OP1 sequences, the correct residue likely is a histidine.) Fragment 4 is described essentially by residues 153–165 of Seq. ID No. 22 and fragment 5 is described essentially by residues 137–151 of Seq. ID No. 22 (note that residue 5 in fragment 5 is identified as a proline. In Seq. ID No. 22 this residue is a serine. By comparison with the CBMP2 and OP1 sequences, the correct residue likely is a serine.) Fragment 6 is described essentially by residues 77–83 of Seq. ID No. 22. Fragment 3 is described by residues 359–379 in the sequence presented in Seq. ID No. 4 (referred to herein as “CBMP2A”). Fragments 7 and 8 are described by residues 391–394 and 397–399, respectively, of the sequence presented in Seq. ID No. 4 (referred to herein as “OP1”).

30 Subsequent additional peptide digest experiments performed on each of the two subunits purified from the highest activity fractions and digested with either thermolysin or endoproteinase Asp-N followed by endoproteinase Glu-C unequivocally identifies the 18 kDa subunit as comprising OP1, and the 16 kDa subunit as comprising CBMP2 (see U.S. Pat. No. 5,011,691 and Kuber Sampath et al., (1990) J. Biol. Chem. 265:13198–13205.)

Specifically, pyridylethyllylation of C-18 purified, reduced, bOP fractions showing the highest osteogenic activity, followed by separation by SDS-PAGE, gel slicing, elution, and digestion with endoproteinase Asp-N, then Staph V-8 protease, permitted separation of peptide fragments representative of each of the subunits from natural-sourced bovine material. Sequencing of the peptide fragments from the 18 kDa subunit yielded five sequences unequivocally from OP1. Sequencing of peptide fragments from the 16 kDa subunit yielded six sequences unequivocally from CBMP2A, and three that could have been from either CBMP2A or CBMP2B. The five sequences unequivocally from OP1 correspond to residue Nos. 341–345, 342–346, 346–352, 353–360 and 386–399 of Seq. ID No. 1. The six sequences unequivocally from CBMP2A correspond to residue Nos. 312–324, 312–330, 314–322, 323–330, 335–354 and 366–373 of Seq. ID No. 4. Another peptide, analyzed as Asp-Xaa-Pro-Phe-Pro-Leu, was consistent with the presence of CBMP2B. However, the amino terminal aspartic acid could have been a glutamic acid (Glu), in which case the peptide would have indicated the presence of CBMP2A. The Asp-Xaa-Pro-Phe-Pro-Leu sequence data has been eliminated from the data presented here (see U.S. Pat. No. 5,011,691, col. 7, fragment 7 for correct sequence).

2.6 Amino Acid Analysis

Strategies for obtaining amino acid composition data were developed using gel elution from 15% SDS gels, transfer onto Immobilon, and hydrolysis. Immobilon membrane is a
polymer of vinylidene difluoride and, therefore, is not susceptible to acid cleavage. Samples of oxidized (30 kDa) and reduced (16 kDa and 18 kDa) bOP are electrophoresed on a gel and transferred to Immobilon for hydrolysis and analysis as described below. The composition data generated by amino acid analyses of 30 kDa bOP is reproducible, with some variation in the number of residues for a few amino acids, especially cysteine and isoleucine.

Samples are run on 15% SDS gels, transferred to Immobilon, and stained with Coomassie blue. The bands of interest are excised from the Immobilon, with a razor blade and placed in a Corning 6x50 test tube cleaned by pyrolysis at 55° C. When cysteine is to be determined, the samples are treated with performic acid (PFA), which converts cysteine to cysteic acid. Cysteic acid is stable during hydrolysis with HCl, and can be detected during the HPLC analysis by using a modification of the normal Pico Tag eluents (Millipore) and gradient. The PFA is made by mixing 50 µl 30% hydrogen peroxide with 950 µl 99% formic acid, and allowing this solution to stand at room temperature for 2 hr. The samples then are treated with PFA as follows: 20 µl PFA is pipetted onto each sample and placed in an ice bath at 4° C. for 2.5 hours. After 2.5 hours, the PFA is removed by drying in vacuo, and the samples then are hydrolyzed. A standard protein of known composition and concentration containing cysteine is treated with PFA and hydrolyzed concurrently with the bOP samples.

The hydrolysis of the bOP samples is done in vacuo. The samples, with empty tubes and Immobilon blanks, are placed in a hydrolysis vessel which is placed in a dry ice/ethanol bath to keep the HCl from prematurely evaporating. 200 µl 6 N HCl containing 2% phenol and 0.1% stannous chloride are added to the hydrolysis vessel outside the tubes containing the samples. The hydrolysis vessel is then sealed, flushed with prepurified nitrogen, evacuated, and then held at 115° C. for 24 hours, after which time the HCl is removed by drying in vacuo.

After hydrolysis, each piece of Immobilon is transferred to a fresh tube, where it is rinsed twice with 100 µl 0.1% TFA, 50% acetonitrile. The washings are returned to the original sample tube, which then is redried as below. A similar treatment of amino acid analysis on Immobilon can be found in the literature (LeGendre and Matsudaira (1988) Biotechniques 6:154–159).

The samples are redried twice using 2:2:1 ethanol:water:triethylamine and allowed to dry at least 30 min. after each addition of redry reagent. These redrying steps bring the sample to the proper pH for derivatization.

The samples are derivatized using standard methodology. The solution is added to each sample tube. The tubes are placed in a desiccator which is partially evacuated, and are allowed to stand for 20 min. The desiccator then is fully evacuated, and the samples are dried for at least 3 hr. After this step the samples may be stored under vacuum at −20° C. or immediately diluted for HPLC. The samples are diluted with Pico Tag Sample Diluent (generally 100 µl) and allowed to stand for 20 min., after which they are analyzed on HPLC using the Pico Tag chromatographic system with some minor changes involving gradients, eluents, initial buffer conditions and oven temperature.

After HPLC analysis, the compositions are calculated. The molecular weights are assumed to be 14.4 kDa, 16.2 kDa, and 27 kDa. The number of residues is approximated by dividing the molecular weight by the average molecular weight per amino acid, which is 115. The total picomoles of amino acid recovered is divided by the number of residues, and then the picomoles recovered for each amino acid is divided by the number of picomoles per residue, determined above. This gives an approximate theoretical number of residues of each amino acid in the protein. Glycine content may be overestimated in this type of analysis.

Composition data obtained are shown in TABLE 5.

**TABLE 5**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>30 kDa</th>
<th>16 kDa</th>
<th>18 kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>22</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Asp/Glu</td>
<td>24</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Ser</td>
<td>24</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Gly</td>
<td>29</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>His</td>
<td>5</td>
<td>*</td>
<td>4</td>
</tr>
<tr>
<td>Arg</td>
<td>13</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Thr</td>
<td>11</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Ala</td>
<td>18</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Pro</td>
<td>14</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Tyr</td>
<td>11</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Val</td>
<td>14</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Met</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Cys**</td>
<td>16</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Ile*</td>
<td>15</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Leu</td>
<td>15</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Phe</td>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Trp</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Lys</td>
<td>12</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

* This result is not integrated because histidine is present in low quantities.
** Cysteine is corrected by percent normally recovered from performic acid hydrolysis of the standard protein.

The results obtained from the 16 kDa and 18 kDa subunits, when combined, closely resemble the numbers obtained from the native 30 kDa protein. The high figures obtained for glycine and serine are most likely the result of gel elution.

3. Demonstration that the 30 kDa Protein is Osteogenic Protein

3.1 Gel Slicing

Gel slicing experiments confirm that the isolated 30 kDa protein is the protein responsible for osteogenic activity. Gels from the last step of the purification are sliced. Protein in each fraction is extracted in 15 mM Tris-HCl, pH 7.0 containing 0.1% SDS. The extracted proteins are desalted, concentrated, and assayed for endochondral bone formation activity. The results are set forth in FIG. 14. Activity in higher molecular weight regions apparently is due to protein aggregation. These protein aggregates, when reduced, yield the 16 kDa and 18 kDa species discussed above.

3.2 Con A- Sepharose Chromatography

A sample containing the 30 kDa protein is solubilized using 0.1% SDS, 50 mM Tris-HCl, and is applied to a column of Con A-Sepharose equilibrated with the same buffer. The bound material is eluted in SDS Tris-HCl buffer containing 0.5 M alpha-methyl mannoside. After reverse phase chromatography of both the bound and unbound fractions, Con A-bound materials, when implanted, result in extensive bone formation (see Sections III—V, infra, for assay methodologies). Further characterization of the bound materials show a Con A-blottable 30 kDa protein. Accordingly, the 30 kDa glycosylated protein is responsible for the bone forming activity.

3.3 Gel Permeation Chromatography

TSK-3000/2000 gel permeation chromatography in guanidine-HCl is used to achieve separation of the high
II. NOVEL OSTEOGENIC SEQUENCES

A. OPI

1. DNA Sequence Identification and Characterization

These discoveries enable preparation of DNAs encoding totally novel, non-native (e.g., not known to occur in nature) protein constructs which individually as homodimers and combined with other related species, possibly as heterodimers, are capable of producing true endochondral bone. They also permit expression of the natural material, truncated forms, muteins, analogs, fusion proteins, and various other variants and constructs, from cDNAs and genomic DNAs retrieved from natural sources or from synthetic DNA produced using the techniques disclosed herein and automated, commercially available equipment. The DNAs may be expressed using well established recombinant DNA technologies in procaryotic or eucaryotic host cells, or in cell-free systems, and may be oxidized and refolded in vitro if necessary for biological activity.

More specifically, a synthetic consensus gene shown in Seq. ID No. 33 and FIG. 18, was designed as a hybridization probe (see U.S. Pat. No. 4,968,590, filed Apr. 8, 1988.) The design was based on amino acid sequence data obtained by sequencing digestion fragments of naturally sourced material and on predictions from observed homologies of these sequences with members of the TGF-β gene family. The consensus gene/probe exploited human codon bias as found in human TGF-β. The designed sequence then was constructed using known assembly techniques for oligonucleotides manufactured in a DNA synthesizer. Table 6, below, shows the identified homologies between trypsin peptides derived from bOP and amino acid sequences from Drosophila DPP protein (as inferred from the gene) and the Xenopus Vg1 protein, both of which show strong homology with the bOP peptides, and TGF-beta and inhibin, which share somewhat less homology with the bOP peptides.

### Table 6

<table>
<thead>
<tr>
<th>Protein</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1</td>
<td>SFDAYCGCGAQCFPDE</td>
</tr>
<tr>
<td></td>
<td>* * * * * *</td>
</tr>
<tr>
<td></td>
<td>(9/15 matches)</td>
</tr>
<tr>
<td>OP2</td>
<td>GYDAYCGCGPCFPFL</td>
</tr>
<tr>
<td></td>
<td>* * * * *</td>
</tr>
<tr>
<td></td>
<td>(6/15 matches)</td>
</tr>
<tr>
<td>Vg1</td>
<td>GTHANYCGCPCFPL</td>
</tr>
<tr>
<td></td>
<td>* * * * *</td>
</tr>
<tr>
<td></td>
<td>(5/15 matches)</td>
</tr>
<tr>
<td>Inhibin</td>
<td>GTHANYCGCPSHIH</td>
</tr>
<tr>
<td></td>
<td>* * * * *</td>
</tr>
<tr>
<td></td>
<td>(4/15 matches)</td>
</tr>
<tr>
<td>TGF-β1</td>
<td>GTHANYCGCPCPYW</td>
</tr>
<tr>
<td></td>
<td>* * * * *</td>
</tr>
<tr>
<td></td>
<td>(12/20 matches)</td>
</tr>
<tr>
<td>Vg1</td>
<td>LPCAYFCGCPYCN</td>
</tr>
<tr>
<td></td>
<td>* * * * * *</td>
</tr>
<tr>
<td></td>
<td>(12/20 matches)</td>
</tr>
<tr>
<td>Inhibin</td>
<td>KYCGCYKRLPFYDGD</td>
</tr>
<tr>
<td></td>
<td>* * * * * *</td>
</tr>
<tr>
<td></td>
<td>(12/20 matches)</td>
</tr>
</tbody>
</table>
In addition to its function as a probe, the consensus sequence also was designed to act as a synthetic consensus gene for the expression of a consensus osteogenic protein. In determining the amino acid sequences of a consensus osteogenic protein from which the nucleic acid sequence can be determined, the following points are considered: (1) the amino acid sequence determined by Edman degradation of osteogenic protein tryptic fragments is ranked highest as long as it has a strong signal and shows homology or conservative changes when aligned with the other members of the gene family; (2) where the sequence matches for all four proteins, it is used in the synthetic gene sequence; (3) matching amino acids in DPP and Vgl are used; (4) if Vgl or DPP diverged but either one is matched by TGF-β or by inhibit, this matched amino acid is chosen; (5) where all sequences diverge, the DPP sequence is initially chosen, with a later plan of creating the Vgl sequence by mutagenesis. Finally, as more amino acid sequences of osteogenic proteins become available, the consensus gene can be improved to match, using known methods of site-directed mutagenesis. In the process, a family of analogs can be developed (see, for example, U.S. Pat. No. 5,011,691, filed Feb. 23, 1989).

A human genomic library (Maniatis-library) carried in lambda phage (Charon 4A) was screened using the probe to screen a human placenta S' stretch lambda phage cDNA library (Clontech, Palo Alto, Calif.), and a human hippocampus library (Stratagene, Inc., La Jolla, Calif.), using high stringency hybridization conditions. Positive clones obtained from these libraries yielded a full length cDNA (translated region) for hOP1. This cDNA sequence, and the amino acid sequence it encodes, is set forth in Seq. ID No. 1. The partial genomic DNA sequence for the human OP1 gene is listed in Seq. ID No. 3. The protein coding region is encoded in seven exons separated by six introns in the genomic sequence (see Seq. ID No. 3). It is possible that, as has been found in certain other mammalian genes, one or more of the introns may include sequences having a transcription regulatory function.

The native form protein is expressed originally as an immature translation product referred to herein as a "prepro" form which includes a signal peptide sequence necessary for appropriate secretion of the protein. Removal of the signal peptide yields the "pro" form of the protein, which is processed further to yield the mature secreted protein. Referring to Table 1 and Seq. ID No. 1, the amino acid sequence of the prepro form of OP1 (herein referred to as hOP1-PP) is described by residues 1-431. The amino acid residues 26 to 30 of Seq. ID No. 1 are believed to constitute a cleavage site for the removal of the N-terminal residues, constituting the signal peptide. Residues 289-292 of Seq ID No. 1 represent the pertinent Arg-Xaa-Xaa-Arg sequence where the pro form is believed to be cut to produce the mature form (e.g., cleavage occurs between residues 292 and 293.) Both the pro form and the prepro form, when properly dimerized, folded, adsorbed on a matrix, and implanted, display osteogenic activity, presumably due to proteolytic degradation resulting in cleavage and generation of mature form protein or active truncated analogs. (See Section II.A.2, infra). Mature OP1 contains 3 potential N glycosylation sites; there is an additional site in the precursor region.

The genomic clone lambda #18 DNA was found to contain the complete sequence encoding the protein referred to herein as CBMP2B. The DNA sequence corresponds to the sequence termed human BMP-2 Class II ("BMP4") in PCT US 87/01537. The CBMP2(b) gene consists of two exons. Exon 1 is approximately 0.37 kb and exon 2 (containing the TGF-β domain) is about 0.86 kb. The two exons are interrupted by an approximately 1 kb intron. Following the methodology used to identify the hOP1 cDNA, the coding sequence of the genomic CBMP2(b) clone was used as a probe to clone the full-length CBMP2(b) cDNA from a human placenta S'-stretch cDNA library (Clontech, Palo Alto.) This cDNA sequence, and the predicted amino acid sequence it encodes, are set forth in Seq. ID No. 6.

The cDNA encoding the protein referred to herein as CBMP2A was cloned using the CBMP2(b) cDNA as a probe. The murine homolog first was cloned from a murine cDNA library and a portion of this cDNA sequence then used as a probe to clone the human CBMP2(a) cDNA from a human hippocampus cDNA library. (Stratagene, Inc., La Jolla) and a human fetal lung library. Each of these human cDNA libraries yielded partial length clones which were then fused to yield the full length CBMP2(a) cDNA clone. The cDNA sequence for CBMP2(a), and its predicted
encoded amino acid sequence, are set forth in Seq. ID No. 4. The DNA sequence corresponds to the sequence termed human BMP-2 Class 1 ("BMP2") in PCT US 87/01357.

The amino acid sequence corresponding to the conserved six cysteine skeleton in the active region of CBMP2B is described by amino acid residues 313 to 408 of Seq. ID No. 6 (herein referred to as "CBMP2BS" where "S" refers to "short form.") Similarly, the corresponding amino acid sequence of CBMP2A ("CBMP2AS") is described by amino acid residues 301 to 396 of Seq. ID No. 4. Longer sequences defining the seven cysteine skeleton, are "CBMP2AL" (residues 296 to 396 of ID No. 4), and "CBMP2BL" (residues 308 to 408 of ID No. 6), where "L" refers to "long form.

Seq. ID Nos. 4 and 6 describe the human cDNA sequences for CBMP2a (a) and CBMP2b (b), respectively, as well as the encoded full-length, "prepro" forms of these proteins. Using the prediction methods devised by Von Heijne ("1986 Nucleic Acids Research 14:4683-4691"), residues 20-24 indicate the region for the presumed signal peptide cleavage site for CBMP2a (Seq. ID No. 4), and residues 23-24 of Seq. ID No. 6 indicate the presumed cleavage site for CBMP2B. The cleavage site yielding the mature sequence of CBMP2A is believed to occur within the region described by residues 271-282 of ID No. 4 and within the region described by residues 280-292 of Seq. ID No. 6 for CBMP2B. Although there remains uncertainty regarding where precise cleavage occurs for this protein. Finally, the CBMP2 proteins contain 4 or 5 potential glycosylation sites.

The consensus probe also identified a human genomic clone encoding a protein referred to herein as CBMP3. The DNA sequence corresponds to the sequence termed human BMP3 in PCT US 87/01357. A partial genomic sequence encoding part of the mature region of the CBMP3 protein is set forth in Seq. ID No. 20. Using the same general methodology as described for the cloning of the CBMP2B cDNA sequences, the murine cDNA encoding CBMP3 was cloned ("mCBMP3.") The cDNA encoding the mature region of this protein, and the encoded amino acid sequence, are set forth in Seq. ID No. 22.

Given the foregoing amino acid and DNA sequence information, various DNAs can be constructed which encode at least a minimal sequence encoding the active domain of OP1 and/or CBMP2, and various analogs thereof, as well as fusion proteins, truncated forms of the mature proteins, and similar constructs. Both the pro form and the prepro form are active, presumably because of in situ cleavage events or generation of active products by cleavage during protein processing. These DNA can be produced by those skilled in the art using well known DNA manipulative techniques involving genomic and cDNA isolation, construction of synthetic DNA from synthesized oligonucleotides, and cassette mutagenesis techniques. 15-100 mer oligonucleotides may be synthesized on a Biosense DNA Model 8600 Synthizer, and purified by polyacrylamide gel electrophoresis (PAGE) in Tris-Borate-EDTA buffer. The DNA then is electrophoreted from the gel. Overlapping oligomers may be phosphorylated by T4 polynucleotide kinase and ligated into larger blocks which may also be purified by PAGE.

The cDNA or synthetic DNA then may be integrated into an expression vector and transfected into an appropriate host cell for protein expression. Because both the glycosylated and unglycosylated protein are active, the host may be a procaryotic or eucaryotic cell. Useful host cells include E. coli, Saccharomyces, the insect/baculovirus cell system, myeloma cells, and various other mammalian cells. The proteins of this invention preferably are expressed in mammalian cells, as disclosed herein. The vector additionally may include various sequences to promote correct expression of the recombinant protein, including transcription promoter and termination sequences, enhancer sequences, preferred ribosome binding site sequences, preferred mRNA leader sequences, preferred protein processing sequences, preferred signal sequences for protein secretion, and the like. The DNA sequence encoding the gene of interest also may be manipulated to remove potentially inhibiting sequences or to minimize unwanted secondary structure formation. The recombinant osteogenic protein also may be expressed as a fusion protein. After being translated, the protein may be purified from the cells themselves or recovered from the culture medium. All biologically active protein forms comprise dimeric species joined by disulfide bonds or otherwise associated, produced by oxidizing and refolding one or more of the various recombinant proteins within an appropriate eucaryotic cell or in vitro after expression of individual subunits.

2. Expression in E. coli

Using such techniques, various fusion genes can be constructed to induce recombinant expression of osteogenic sequences in a procaryotic host such as E. coli. In particular, the following DNAs have been prepared:

<table>
<thead>
<tr>
<th>Fusion DNA Sequences</th>
<th>Osteogenic Fusion Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1(a)</td>
<td>OP1A (Seq. ID No. 8)</td>
</tr>
<tr>
<td>OP1(b)</td>
<td>OP1B (Seq. ID No. 10)</td>
</tr>
<tr>
<td>OP1(c)</td>
<td>OP1C (Seq. ID No. 12)</td>
</tr>
<tr>
<td>OP1(d)</td>
<td>OP1D (Seq. ID No. 14)</td>
</tr>
<tr>
<td>CBMP2b1</td>
<td>CBMP2B1 (Seq. ID No. 16)</td>
</tr>
<tr>
<td>CBMP2b2</td>
<td>CBMP2B2 (Seq. ID No. 18)</td>
</tr>
</tbody>
</table>

Construct OP1(a) is a cDNA sequence encoding substantially all of the mature form of OP1 (residues 326-431, Seq. ID No. 1) linked by an Asp-Pro acid cleavage site to a leader sequence ("MLE leader", amino acid residues 1-60 of Seq. ID No. 8) suitable for promoting expression in E. coli. OP1(b) (Seq. ID No. 10) encodes a truncated "pro" form of OP1. The sequence comprises the MLE leader linked to an OP1 sequence which begins within the precursor ("prepro") sequence (beginning at residue 176 of Seq. ID No. 1). OP1(c) comprises an MLE leader peptide (residues 1-53 of Seq. ID No. 12) linked to the full prepro form of OP1 cDNA including the presumed signal peptide (e.g., residues 1-29 of Seq. ID No. 1). OP1(d) comprises a leader sequence ("short TRP", residues 1-13 of Seq. ID No. 14), an Asp-Pro cleavage site, and the presumed pro form of the OP1 protein (residues 39-431 of Seq. ID No. 1). CBMP2b1 (Seq. ID No. 16) comprises the MLE leader (residues 1-56, Seq. ID No. 16) linked through an Asp-Pro acid cleavage site to substantially all of the mature form of CBMP2B (residues 296-408 of Seq. ID No. 6). Approximately one half of this construct comprised cDNA; the other half was synthesized from oligonucleotides. CBMP2b2 comprises the MLE leader (residues 1-60 of ID No. 18) linked to substantially all of the full length pro form of CBMP2B (residues 52-408 of Seq. ID No. 6).

The genes were expressed in E. coli under the control of a synthetic trp promoter-operator to produce insoluble inclusion bodies. The inclusion bodies were solubilized in 8M urea following lysis, dialyzed against 1% acetic acid, and partly purified by differential solubilization. Constructs containing the Asp-Pro site were cleaved with acid. The result-
Both antisera react with the naturally sourced 30 kDa trityl, and the likelihood that degraded OP1 generated during homology between CBMP2 and OP1 (approx. 60% identity), this was expected in view of the significant reactivity, which was lyophilized at pH 2 so that it remains reduced.

Conditions for refolding were at pH 8.0 using Tris buffer and 6M guanidine-HCl at a protein concentration of several mg/ml. Those solutions were diluted with water to produce a 2M or 3M guanidine concentration and left for 18 hours at 4°C. Air dissolved or entrained in the buffer assures oxidation of the protein in these circumstances.

Samples of the various purified constructs and various mixtures of pairs of the constructs refolded together were applied to SDS polyacrylamide gels, separated by electrophoresis, sliced, incorporated in a matrix as disclosed below, and tested for osteogenic activity. These studies demonstrated that each of the constructs disclosed above have true osteogenic activity. Thus, both the pro form and propro form, when properly dimerized, folded, adsorbed on a matrix, and implanted, display osteogenic activity, presumably due to proteolytic degradation resulting in cleavage and generation of mature form protein or active truncated species. In addition, mixed species also are osteogenically active and may include heterodimers. Specific combinations tested include: OP1A-CBMP2B1, OP1B-CMP2B1, and OP1C-CBMP2B2. Finally, single and mixed species of analogs of the active region, e.g., COP5 and COP7, disclosed in U.S. Pat. No. 5,011,691, also induce osteogenesis, as determined by histological examination.

After N-terminal sequencing of the various constructs to confirm their identity, polyclonal antisera against the recombinant presumed mature form proteins were produced. The human OP1 antisera reacted with both the glycosylated and unglycosylated lower molecular weight subunits of naturally sourced bovine material. Antisera against recombinant mature human CBMP2 reacted with both the glycosylated and unglycosylated lower molecular weight subunit of naturally sourced bovine material. While there was some cross-reactivity, this was expected in view of the significant homology between CBMP2 and OP1 (approx. 60% identity), and the likelihood that degraded OP1 generated during purification contaminates the lower molecular weight subunit. Both antisera react with the naturally sourced 30 kDa dimeric BOP.

3. Mammalian Cell Expression

As stated earlier, it is generally held that recombinant production of mammalian proteins for therapeutic uses are preferably expressed in mammalian cell culture systems in order to produce a protein whose structure is most like that of the natural material. Recombinant protein production in mammalian cells requires the establishment of appropriate cells and cell lines that are easy to transfect, capable of stably maintaining foreign DNA with an unarranged sequence, and which have the necessary cellular components for efficient transcription, translation, post-translation modification, and secretion of the protein. In addition, a suitable vector carrying the gene of interest also is necessary. DNA vector design for transfection into mammalian cells should include appropriate sequences to promote expression of the gene of interest as described supra, including appropriate transcription initiation, termination, and enhancer sequences, as well as sequences that enhance translation efficiency, such as the Kozak consensus sequence. Preferred DNA vectors also include a marker gene and means for amplifying the copy number of the gene of interest.

Substantial progress in the development of mammalian cell expression systems has been made in the last decade and many aspects of the system are well characterized. A detailed review of the state of the art for the production of foreign proteins in mammalian cells, including useful cells, protein expression-promoting sequences, marker genes, and gene amplification methods, is disclosed in Bendig, Mary M., (1988) Genetic Engineering 7:91-127.

Briefly, among the best characterized transcription promoters useful for expressing a foreign gene in a particular mammalian cell are the SV40 early promoter, the adenovirus promoter (AdMLP), the mouse metallothionein-I promoter (mMT-I), the Rous sarcoma virus (RSV) long terminal repeat (LTR), the mouse mammary tumor virus long terminal repeat (MMTV-LTR), and the human cytomegalovirus major intermediate-early promoter (hCMV). The DNA sequences for all of these promoters are known in the art and are available commercially.

One of the better characterized methods of gene amplification in mammalian cell systems is the use of the selectable DHFR gene in a dhfr- cell line. Generally, the DHFR gene is provided on the vector carrying the gene of interest, and association of increasing concentrations of the cytotoxic drug methotrexate leads to amplification of the DHFR gene copy number, as well as that of the associated gene of interest. DHFR as a selectable, amplifiable marker gene in transfected Chinese hamster ovary cell lines (CHO cells) is particularly well characterized in the art. Other useful amplifiable marker genes include the adenosine deaminase (ADA) and glutamine synthetase (GS) genes.

In the currently preferred expression system, gene amplification is further enhanced by modifying marker gene expression regulatory sequences (e.g., enhancer, promoter, and transcription or translation initiation sequences) to reduce the levels of marker protein produced. As disclosed herein, lowering the level of DHFR transcription has the effect of increasing the DHFR gene copy number (and the associated OP1 gene) in order for a transfected cell to adapt to grow in even low levels of MTX (e.g., 0.1 μM MTX). Preferred expression vectors (pH754 and pH752), have been manipulated using standard recombinant DNA technology, to create a weak DHFR promoter (see infra). As will be appreciated by those skilled in the art, other useful weak promoters, different from those disclosed and preferred herein, can be constructed using standard vector construction methodologies. In addition, other, different regulatory sequences also can be modified to achieve the same effect.

The choice of cells/cell lines is also important and depends on the needs of the experimenter. Monkey kidney cells (COS) provide high levels of transient gene expression, providing a useful means for rapidly testing vector construction and the expression of cloned genes. COS cells are transfected with a simian virus 40 (SV40) vector carrying the gene of interest. The transfected COS cells eventually die, thus preventing the long term production of the desired protein product. However, transient expression does not require the time consuming process required for the development of a stable cell line.

Among established cell lines, CHO cell lines may be the best characterized to date, and are the currently preferred cell line for mammalian cell expression of recombinant osteogenic protein. CHO cells are capable of expressing proteins from a broad range of cell types. The general applicability of CHO cells and its successful production for a wide variety of human proteins in unrelated cell types emphasizes the
thereby also amplifying the associated gene of interest. Temperature sensitivity (ts) of BSC40-tsA58 cells transfected with an SV40 vector have been tested. Several Selection marker genes also are available. Other promoters, selectable markers, gene amplification methods and cells also may be used to express the OP1 proteins of this invention, as well as other osteogenic proteins such as OP1 from a variety of mammalian cells, the nature of the xenogenic matrix, and other material aspects concerning the nature, utility, and how to make and how to use the subject matter claimed will be further understood from the following, which constitutes the best method currently known for practicing the invention. The methodology disclosed herein includes the use of COS cells for the rapid evaluation of vector construction and gene expression, and the use of established cell lines for long term protein production. Of the cell lines disclosed, OP1 expression from CHO cell lines currently is most preferred.

3.1 Recombinant Protein Expression In Mammalian Cells

Several different mammalian cell expression systems have been used to express recombinant OP1 proteins of this invention. In particular, COS cells are used for the rapid assessment of vector construction and gene expression, using an SV40 vector to transfect the DNA sequence into COS cells. Stable cell lines are developed using CHO cells (chinese hamster ovary cells) and a temperature-sensitive strain of BSC cells (simian kidney cells, BSC40-tsA58, (1988) Biotechnology 6: 1192-1196) for the long term production of OP1. Two different promoters were found most useful to transcribe hOP1: the CMV promoter and the MMTV promoter, boosted by the enhancer sequence from the Rous sarcoma virus LTR. The mMT promoter (mouse metallothionein promoter) and the SV40 late promoter also are useful for mediating plasmid replication in primate cells (e.g., COS and BSC cells). In addition, the early SV40 promoter is used to drive transcription of marker genes on the vector (e.g., neo and DHFR). It will be appreciated by those skilled in the art that DNA sequences encoding truncated forms of the osteogenic protein also may be used, provided that the expression vector or host cell then provides the sequences necessary to direct processing and secretion of the expressed protein.

The pH717 expression vector (FIG. 19A) contains the neomycin (neo) gene as a selection marker. This marker gene is well characterized in the art and is available commercially. Alternatively, other selectable markers may be used. The particular vector used to provide the neo gene DNA fragment for pH717 may be obtained from Clontech, Inc., Palo Alto, Calif. (pMAM-neo-blue). This vector also may be used as the backbone. In pH717, hOP1 DNA transcription is driven by the CMV promoter, boosted by the RSV-LTR and MMTV-LTR (mouse mammary tumor virus) enhancer sequences. These sequences are known in the art, and are available commercially. For example, vectors containing the CMV promoter sequence may be obtained from Invitrogen Inc., San Diego, Calif., (e.g., pCDM8).

Expression vector pH731 (FIG. 19B), utilizes the SV40 late promoter to drive hOP1 transcription. As indicated above, the sequence and characteristics of this promoter also are well known in the art. For example, pH731 may be generated by inserting the Small-BamHI fragment of hOP1 into pEUK-C1 (Clontech, Inc., Palo Alto, Calif.).

The pH752 and pH754 expression vectors contain the DHFR gene, under SV40 early promoter control, as both a selection marker and as an inducible gene amplifier. The DNA sequence for DHFR is well characterized in the art, and is available commercially. For example, pH754 may be generated from pMAM-neo (Clontech, Inc., Palo Alto, Calif.) by replacing the neo gene (BamHI digest) with an SpSl-BamHI fragment from pSV5-DHFR (ATCC #37148), which contains the DHFR gene under SV40 early promoter control. A BamHI site can be engineered at the SphI or PvuII site using standard techniques (e.g., by linker insertion or site-directed mutagenesis) to allow insertion of the fragment into the vector backbone.

The hOP1 DNA can be inserted into the polylinker site downstream of the MMTV-LTR sequence (mouse mammary tumor virus LTR), yielding pH752 (FIG. 19D). The CMV promoter sequence then may be inserted into pH752 (e.g., from pCDM8, Invitrogen, Inc.), yielding pH754 (FIG. 19C.).

The SV40 early promoter, which drives DHFR expression, is modified in these vectors to reduce the level of DHFR mRNA produced. Specifically, the enhancer sequences and part of the promoter sequence have been deleted, leaving only about 200 bases of the promoter sequence upstream of the DHFR gene. Host cells transfected with these vectors are adapted to grow in 0.1 µM MTX, and can increase OP1 production significantly (see Table 8).

The pHW24 vector (FIG. 19E), is essentially identical in sequence to pH754, except that neo is used as the marker gene (see pH717), in place of DHFR.
37

Similarly, pH783 (FIG. 19F) contains the amplifiable marker DHFR, but here OP1 is under mMT (mouse metallothionein promoter) control. The mMT promoter is well characterized in the art and is available commercially.

All vectors tested are stable in the various cells used to express OP1, and provide a range of OP1 expression levels.

3.3 Exemplary Mammalian Cells

Recombinant OP1 has been expressed in three different cell expression systems: COS cells for rapidly screening the functionality of the various expression vector constructs, CHO cells for the establishment of stable cell lines, and BSC40-tsA58 cells as an alternative means of producing OP1 protein. The CHO cell expression system disclosed herein is contemplated to be the best mode currently known for long term recombinant OP1 production in mammalian cells.

a) COS Cells

COS cells (simian kidney cells) are used for rapid screening of vector constructs and for immediate, small scale production of OP1 protein. COS cells are well known in the art and are available commercially. The particular cell line described herein may be obtained through the American Type Culture Collection (ATCC #COS-1, CRL-1650).

OP1 expression levels from different vectors, analyzed by Northern and Western blot assays, are compared in Table 7 below:

<table>
<thead>
<tr>
<th>Vector</th>
<th>mRNA</th>
<th>OP1 Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH717</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>pH731</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pH752</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>pH754</td>
<td>+++</td>
<td>+++</td>
</tr>
</tbody>
</table>

pH752- and pH754-transfected COS cells appear to produce the highest yield of OP1 to date. However, because transfected COS cells do not divide and die several days post-transfection, large amounts of plasmid DNA are required for each scaled up transformation.

Large scale preparations of OP1 from transfected COS cells may be produced using conventional roller bottle technology. Briefly, 14×10⁶ cells are used to seed each bottle. After 24 hrs of growth, the cells are transfected with 10 μg of vector DNA (e.g., pH717) per 10⁶ cells, using the DEAE-dextran method. Cells are then conditioned in serum-free media for 120 hr before harvesting the media for protein analysis. Following this protocol, OP1 yield is approximately 2–6 ng/ml.

b) BSC Cells

The BSC40-tsA58 cell line ("BSC cells") is a temperature-sensitive strain of simian kidney cells ([1988], Biotechnology 6: 1192–1196) which overcomes some of the problems associated with COS cells. These BSC cells have the advantage of being able to amplify gene sequences rapidly on a large scale with temperature downshift, without requiring the addition of exogenous, potentially toxic drugs. In addition, the cells may be recycled. That is, after induction and stimulation of OP1 expression, the cells may be transferred to new growth medium, grown to confluence at 39.5° C, and induced a second time by downshifting the temperature to 33° C. BSC cells may be used to establish stable cell lines rapidly for protein production.

OP1 expression in transfected BSC cells may be induced by shifting the temperature down to 33° C., in media containing 10% FCS, and harvesting the conditioned media after 96 hrs of incubation. Comparable amounts of OP1 mRNA and protein are obtained, as compared with CHO cells (e.g., 100–150 ng OP1/ml conditioned media from BSC clones transfected with pH717, see infra).

c) CHO Cells

CHO cells (chinese hamster ovary cells) may be used for long term OP1 production and are the currently preferred cell line for mammalian cell expression of OP1. CHO cell lines are well characterized for the small and large scale production of foreign genes and are available commercially. The particular cell line described herein is CHO-DXB11, (Lawrence Chasin, Columbia University, NY). Table 8, below, shows exemplary OP1 yields obtained with a variety of expression vectors.

<table>
<thead>
<tr>
<th>CHO Cells</th>
<th>Plasmid</th>
<th>Selection Marker</th>
<th>OP1 Production ng/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO Cells</td>
<td>pH717/neo</td>
<td>DHFR</td>
<td>2-5</td>
</tr>
</tbody>
</table>

CHO cells may be transfected by conventional calcium phosphate technique. CHO cells preferably are transfected with pH754 or pH752 and are conditioned in media containing serum proteins, as this appears to enhance OP1 yields. Useful media includes media containing 0.1–0.5% dialyzed fetal calf serum (FCS).

The currently preferred best mode for establishing a stable OP1 production cell line with high hOP1 expression levels comprises transfecting a stable CHO cell line, preferably CHO-DXB11, with the pH752 OP1 expression vector, isolating clones with high OP1 expression levels, and subjecting these clones to cycles of subcloning using a limited dilution method described below to obtain a population of high expression clones. Subcloning preferably is performed in the absence of MTX to identify stable high expression clones which do not require addition of MTX to the growth media for OP1 production.

In the subcloning protocol cells are seeded on ten 100 mm petri dishes at a cell density of either 50 or 100 cells per plate, with or preferably without MTX in the culture media. After 14 days of growth, clones are isolated using cloning cylinders and standard procedures, and cultured in 24-well plates. Clones then are screened for OP1 expression by Western immunoblots using standard procedures, and OP1 expression levels compared to parental lines. Cell line stability of high expression subclones then is determined by monitoring OP1 expression levels over multiple cell passages (e.g., four or five passages).

3.4 Evaluation of OP1 Transfected Cells

Expression levels of transfected OP1 sequences can be measured in the different systems by analyzing mRNA levels (Northern blots), using total cellular RNA and conventional hybridization methodology. Generally, about 1x10⁶ cells are needed for mRNA analysis. Data between individual cell lines can be compared if the total number of cells and the total amount of mRNA is normalized, using rRNA as an internal standard. Ribosomal RNA is visualized in the agarose gel by ethidium bromide stain prior to transfer.
of the RNA to nitrocellulose sheets for hybridization. Ribosomal RNA also provides an indicator of the integrity of the RNA preparation.

OP1 protein levels also may be measured by Western blots (immunoblots) using rabbit antiserum against human OP1. FIG. 20 is an immunoblot showing OP1 production in: COS cells—(20A) pH717; (20B) pH731; CHO cells—(20C) pH754, (20D) pH752; and BSC cells—(20E) pH717 and (20F) pW24.

Southern blots may be used to assess the state of integrated OP1 sequences and the extent of their copy number amplification. The copy number of excised plasmids in temperature-shifted BSC cells also can be determined using Southern blot analysis.

3.5 Protein Purification

The purification scheme developed to purify the recombinant osteogenic proteins of this invention is rapid and highly effective. The protocol involves three chromatographic steps (S-Sepharose, phenyl-Sepharose and C-18 HPLC), and produces OP1 of about 90% purity.

For a typical 2L preparation of transfected BSC cells conditioned in 0.5% FCS, the total protein is 700 mg. The amount of OP1 in the media, estimated by Western blot, is about 80 µg. OP1 media is diluted to 6M urea, 0.05M NaCl, 13 mM HEPES, pH 7.0 and loaded onto an S-Sepharose column, which acts as a strong cation exchanger. OP1 binds to the column in low salt, and serum proteins are removed. The column is subsequently developed with two step salt elutions. The first elution (0.1M NaCl) removes contaminants and approximately 10% of the bound OP1. The remaining 90% of OP1 then is eluted in 6M urea, 0.3M NaCl, 20 mM HEPES, pH 7.0.

Ammonium sulfate is added to the 0.3M NaCl fraction to obtain final solution conditions of 6M urea, 1M (NH₄)₂SO₄, 0.3M NaCl, 20 mM HEPES, pH 7.0. The sample then is loaded onto a phenyl-Sepharose column (hydrophobic interaction chromatography). OP1 binds phenyl-Sepharose in the presence of high concentrations of a weak chaotropic salt (e.g., 1M (NH₄)₂SO₄). Once OP1 is bound, the column is developed with two step elutions using decreasing concentrations of ammonium sulfate. The first elution (containing 0.6M (NH₄)₂SO₄) primarily removes contaminants. The bound OP1 then is eluted with a 6M urea, 0.3M NaCl, 20 mM HEPES, pH 7.0 buffer containing no ammonium sulfate.

The OP1 eluted from the phenyl-Sepharose column is dialyzed against water, followed by 30% acetonitrile (0.1% TFA), and then applied to a C-18 reverse phase HPLC column. FIGS. 21A, 21C, and 21E are chromatograms and PAGE gels of fractions after reduction with dithiothreitol (DTT) eluting from the (21A, 21D) S-Sepharose, (21B, 21E) phenyl-Sepharose, and (21C, 21F) C-18 columns. Gel separation of oxidized and reduced OP1 samples show that the reduced subunit has an apparent molecular weight of about 18 kDa, and the dimer has an apparent molecular weight of about 36 kDa, as illustrated in FIG. 22. The subunit size appears to be identical to that purified from COS cells, as well as that of the naturally-sourced OP purified from bone. This purification protocol yields about 30 µg of OP1 for 2 L of conditioned media, a recovery of about 25% of the total OP1 in the conditioned media, as estimated by gel scanning.

An alternative chromatography protocol is to perform the S-Sepharose chromatography in the absence of 6 M urea. The bound proteins then are eluted with salt step elutions (e.g., 100-400 mM NaCl. Most of the OP1 is eluted with about 300 mM NaCl. Additional OP1 then can be eluted with 300 mM NaCl in the presence of 6M urea. The 6M urea elution also may be used in place of the non-urea elution to achieve maximum recovery in one step. In addition, OP1 may be eluted from the phenyl-Sepharose column in 38% ethanol-0.01% TFA, thereby eliminating the need to dialyze the eluent before applying it to the C-18 column. Finally, multiple C-18 columns may be used (e.g., three), to further enhance purification and concentration of the protein.

OP1 also will bind hydroxyapatite efficiently, but only in the absence of 6 M urea and at low phosphate concentrations (less than 5 mM phosphate). Bound OP1 can be removed from the column with a step elution of 1 mM to 0.5M phosphate (in 0.5 M NaCl, 50 mM Tris, pH 7.0). OP1 elutes at about 250 mM phosphate. Additionally, urea (6M) may be added during the elution step.

Other related chromatography methods also may be useful in purifying OP1 from eucaryotic cell culture systems. For example, heparin-Sepharose may be used in combination with the S-Sepharose column. Alternatively, Cu⁺⁺-immobilized metal-ion affinity chromatography (IMAC) will bind OP1 in a phosphate buffer (pH 7.0) containing 6M urea.

3.6 Protein Characterization

Recombinant osteogenic protein expression in COS cells yields essentially a single species having an apparent molecular weight of 18 kDa, as determined by SDS-PAGE analysis. Subsequent N-terminal sequencing data indicates that this species contains the intact mature OP1 sequence, referred to herein as “OPl-18Ser” (“Ser Thr Gly . . . ”, beginning at residue 293 of Seq. ID No.1.) Both the BSC and CHO preparations, by contrast, contain both the intact mature sequence and one or more active degraded species.

BSC cell-derived preparations yield two major species having an apparent molecular weight of about 18 kDa and 16 kDa, and a minor species of about 23 kDa as determined by SDS-PAGE analysis. N-terminal sequencing of the two major species using standard techniques reveals that the 18 kDa species, like the COS-derived OP1 protein, contains the intact mature form of OP1 (OPl-18Ser). The 16 kDa fraction appears to contain five species of the mature sequence, having different N-termini. One form, “OPl-16Ser,” has its N-terminus at +8 of the mature sequence (“Ser Gln Asn . . . “), beginning at residue 300 of Seq. ID No.1.) A second species, referred to herein as “OPl-16Leu”, has its N-terminus at +21 of the mature sequence (“Leu Arg Met . . . “), beginning at residue 313 of Seq. ID No. 1). A third and fourth species, referred to herein as OPl-16Met and OPl-16Ala, love their N-termini at +23 and +24, respectively, of the mature OP1 sequence. (See Seq. ID No.1: OPl-16Met begins at residue 315, “Met Ala Asn . . . “, and OPl-16Ala begins at residue 316, “Ala Asn Val . . . “.) Finally, a fifth degraded species has its N-terminus at +26 of the mature sequence (“Val Ala Gln . . . “), beginning at residue 318 of Seq. ID No. 1 and is referred to herein as “OPl-16Val.” The various species are listed in Table 1 and their N-termini are presented in FIG. 33. Biochemical analyses and in vivo bioassays indicate all species are active (see infra). Preliminary sequencing data of the minor species migrating at 23 kDa suggests that this species also contains the mature active sequence. Accordingly, the protein’s altered mobility on an electrophoresis gel may be due to an altered glycosylation pattern.

Similarly, CHO-derived OP1 preparations generally produce three species having an apparent molecular weight within the range of 15–20 kDa, as determined by SDS-
PAGE (specifically, 19 kDa, 17 kDa, and 15 kDa). A minor species also migrates at about 23 kDa. N-terminal and C-terminal sequencing (by CNBr analysis) of proteins in the different fractions reveals that CHO expression produces the same species of OP1 proteins as produced by BSC cell expression, but having different electrophoretic mobility on an SDS polyacrylamide gel. Both the 19 kDa and the 17 kDa protein fractions contain the intact mature form of OP1 (OP1-18Ser) and the OP1-16Ser degraded form. Preliminary sequencing data of the 23 kDa species suggest that this species also contains the intact mature form of OP1. Finally, N-terminal sequencing of the protein species migrating at 15 kDa indicates that proteins in this fraction contain the other four degraded forms of OP1 identified in the BSC cell system: OP1-16Leu, OP1-16Met, OP1-16Ala and OP1-16Val. These data suggest that the apparent molecular weight differences among the various OP1 species detected may be due primarily to variations in their glycosylation patterns. In addition, protein glycosylation pattern variations are a known characteristic of CHO expression systems. In vivo biosays of all OP1 species detected indicate that all truncated forms are active (see infra).

The glycosylation patterns of the proteins in the various OP1 preparations can be investigated by measuring their reactivity with different lectins, using standard methodologies. Here, reactivity with Concanavalin A (Con A), which binds to the mannose core region, and Wheat Germ Agglutinin (WGA), which binds to N-acetyl glucosamine (GlcNAc) and sialic acid (SA) residues, was measured. Results indicate that there may be substantial variation among the glycosylation patterns of the various OP1 species. Con A reacts strongly with both the CHO-derived 17 kDa species and the BSC-derived 16 and 18 kDa species, but only weakly with the other species. Conversely, WGA reacts strongly only with the 19 kDa and 23 kDa CHO-derived species and the 18 and 23 kDa BSC-derived proteins. These results further suggest that variations in the electrophoretic migration patterns of the various OP1 preparations reflect variations in protein glycosylation patterns, which appear to be host cell-specific characteristics.

The various different OP1 preparations also have been analyzed by standard HPLC chromatography. Preparations of OP1 from both CHO and BSC cells have very similar characteristics by HPLC analysis in oxidized, reduced, pyridylethylated or degraded forms. Although distinct by SDS-PAGE analysis, the differences between the different cell type preparations appear insufficient to influence the binding to HPLC C-18 columns.

Accordingly, as will be appreciated by those skilled in the art, it is anticipated that active mature OP1 sequences can be expressed from any of different procarcic and eucaryotic cell expression systems as disclosed herein. The proteins produced may have varying N-termini, and those expressed from eucaryotic cells may have varying glycosylation patterns. Finally, it will also be appreciated that these variations in the recombinant osteogenic protein produced will be characteristic of the host cell expression system used rather than of the protein itself.

B. Identification of Additional, Novel Osteogenic Sequences

In an effort to identify additional DNA sequences encoding osteogenic proteins, a hybridization probe specific to the DNA sequence encoding the C-terminus of the mature OP1 protein was prepared using a StuI-EcoRl digest fragment of hOP1 (base pairs 1034–1354 in Seq. ID No. 1), and labelled with 32P by nick translation, as described in the art. As disclosed supra, applicants have previously shown that the OP1 C-terminus encodes a key functional domain e.g., the “active region” for osteogenic activity (OPS or OP7). The C-terminus also is the region of the protein whose amino acid sequence shares specific amino acid sequence homology with particular proteins in the TGF-β super-family of regulatory proteins and which includes the conserved cysteine skeleton.

Approximately 7×10⁹ phages of an oligo (d1) primed 17.5 days p.c. mouse embryo 5' stretch cDNA (gt10) library (Clontech, Inc., Palo Alto, Calif.) was screened with the labelled probe. The screen was performed using the following hybridization conditions: 40% formamide, 5×SSPE, 5× Denhardt’s Solution, 0.1% SDS, at 37° C. overnight, and washing in 0.1×SSPE, 0.1% SDS at 50° C. Where only partial clones were obtained, the complete gene sequence was subsequently determined by screening either a second cDNA library (e.g., mouse PCC4 cDNA (ZAP) library, Stratagene, Inc., La Jolla, Calif.), or a mouse genomic library (e.g., Clontech, Inc., Palo, Alto, Calif.).

Five recombinant phages were purified over three rounds of screening. Phage DNA was prepared from all five phages, subjected to an EcoRl digest, subcloned into the EcoRl site of a common pUC-type plasmid modified to allow single strand sequencing, and sequenced using means well known in the art.

Two different mouse DNA sequences, referred to herein as mOPl and mOP2, were identified by this procedure. The characteristics of the proteins encoded by these sequences are described below.

1. mOPl

mOPl is the murine homolog of hOPl. The cDNA and encoded amino acid sequence for the full length mOPl protein is depicted in Sequence ID No. 24. The full-length form of the protein is referred to as the prepro form of mOPl (“mOPl-PP”), and includes a signal peptide sequence at its N-terminus. The amino acid sequence Ser-Ala-Leu-Ala-Asp (amino acid residues 26–30 in Seq. ID No. 24) is believed to constitute the cleavage site for the removal of the signal peptide sequence, leaving an intermediate form of the protein, the “pro” form, to be secreted from the expressing cell. The amino acid sequence Arg-Ser-Ile-Arg-Ser (amino acid residue nos. 288–292 in Sequence ID No. 24) is believed to constitute the cleavage site that produces the mature form of the protein, herein referred to as “mOPl-Ser” and described by amino acid residues 292–430 of Seq. ID No. 24. The amino acid sequence defining the conserved 6 cysteine skeleton of the mOPl active region is defined by residues 334–430 of Seq. ID No. 24.

FIG. 23A and 23B compare the amino acid sequence homology of the mature hOPl and mOPl proteins (OP1-18Ser and mOPl-Ser). Amino acid identity is indicated by three dots (...). As can be seen in this figure, the mature form of mOPl, mOPl-Ser, shares significant amino acid sequence homology with OP1-18Ser (98% identity), differing at only three positions in this region. Like OP1-18Ser, mOPl-Ser has a seven cysteine functional domain. In addition, the prepro form of the mOPl protein shows substantially the same homology with that of OP1. The high degree of amino acid sequence homology shared by the mature proteins is not surprising as the amino acid sequences of the mature forms of other TGF-β-like proteins generally also have been found to be highly conserved across different animal species (e.g., compare Vgr and Vgl, two related genes isolated from mouse and Xenopus, respectively). The high degree of amino acid sequence homology exhibited between the mature forms of the two animal species of OP1 proteins suggests that the mOPl protein will purify essen-
kDa. As a glycosylated oxidized homodimer, and about 18 kDa as a reduced single subunit, as determined by comparison with molecular weight standards on an SDS-polyacrylamide electrophoresis gel. There appear to be three potential N glycosylation sites in the mature protein. The unglycosylated homodimer (e.g., one expressed from E. coli) is predicted to have a molecular weight of about 27 kDa.

1. OP1
2. OP2
3. mOP2

The cDNA encoding the C-terminus of mOP2 protein first was identified following the procedure for retrieving mOP1 DNA. The 5' end of the gene was identified subsequently by screening a second mouse cDNA library (Mouse PCC4 cDNA (ZAP) library, Stratagene, Inc., La Jolla, Calif.).

Mouse OP2 (mOP2) protein shares significant amino acid sequence homology with the amino acid sequence of the OP1 active region, e.g., OPS or OP7, about 74% identity, and less homology with the intact mature form, e.g., OP1-18Ser, about 58% identity. The mOP2 protein differs from the OP1 protein by only one non-conservative amino acid change in the active region. The cDNA sequence, and the encoded amino acid sequence, for the full length mOP2 protein are depicted in Sequence ID No. 26. The full-length form of the protein is referred to as the pro-form of mOP2 (“mOP2-Pro”), and includes a signal peptide sequence at its N-terminus. The amino acid sequence Leu-Ala-Leu (amino acid residues 13–18 of Sequence ID No. 26) is believed to constitute the cleavage site for the removal of the signal peptide sequence, leaving an intermediate form of the protein, the “pro” form, to be secreted from the expressing cell. The amino acid sequence Arg-Ala-Pro-Arg-Ala (amino acid residues 257–261 of Sequence ID No. 26) is believed to constitute the cleavage site that produces what is believed to be the mature form of the protein, herein referred to as “mOP2-Ala”, and described by residues 261–399 of Seq. ID No. 26. The amino acid sequence defining the conserved 6 cysteine skeleton of the mOP2 active region is defined by residues 303–399 of Seq. ID No. 26.

2.2 hOP2

Using a probe prepared from the pro region of mOP2 (an EcoRI-BamHI digest fragment, bp 467–771 of Sequence ID No. 26), a human hippocampus library was screened (human hippocampus cDNA lambda ZAP II library, Stratagene, Inc., La Jolla, Calif.) following essentially the same procedure as for the mouse library screens. The procedure identified the N-terminus of a novel DNA encoding an amino acid sequence having substantial homology with the mOP2 protein. The C-terminus of the gene subsequently was identified by probing a human genomic library (in lambda phage EMBL-3, clontech, Inc., Palo Alto, Calif.) with a labelled fragment from the novel human DNA in hand. The novel polypeptide chain encoded by this DNA is referred to herein as hOP2 protein, and the mature form of which shares almost complete amino acid sequence identity (about 92%) with mOP2-A (see FIGS. 23C–23E and infra).

The cDNA sequence, and the encoded amino acid sequence, for the pro-form of hOP2 (“hOP2-Pro”) is depicted in Seq. ID No. 28. This full-length form of the protein also includes a signal peptide sequence at its N-terminus. The amino acid sequence Leu-Ala-Leu-Cys-Ala-Leu (amino acid residues 13–18 of Seq. ID No. 28) is believed to constitute the cleavage site for the removal of the signal peptide sequence, leaving an intermediate form of the protein, the “pro” form, to be secreted from the expressing cell. The amino acid sequence Arg-Thr-Pro-Arg-Ala (amino acid residues 260–264 of Seq. ID No. 28) is believed to constitute the cleavage site that produces what is believed to be the mature form of the protein, herein referred to as “hOP2-Ala”, and described by residues 264 to 402 of Seq. ID No. 28. The amino acid sequence defining the conserved 6 cysteine skeleton of the hOP2 active region is defined by residues 306–402 of Seq. ID No. 28.

Additional mature species of hOP2 thought to be active include truncated short sequences, “hOP2-Pro” (described by residues 267 to 402, Seq. ID No. 28) and “hOP2-Arg” (described by residues 270 to 402, Seq. ID No. 28), and a slightly longer sequence (“hOP2-Ser”, described by residues 243 to 402, Seq. ID No. 28). It should be noted that the nucleic acid sequence encoding the N-terminus of the pro-form of both mOP2 and hOP2 is rich in guanine and cytosine base pairs. As will be appreciated by those skilled in the art, sequencing such a “G-C rich” region can be problematic, due to stutter and/or band compression. Accordingly, the possibility of sequencing errors in this region can not be ruled out. However, the definitive amino acid sequence for these and other, similarly identified proteins can be determined readily by expressing the protein from recombinant DNA using, for example, any of the means disclosed herein, and sequencing the polypeptide chain by conventional peptide sequencing methods well known in the art.

The genomic sequences of both the murine and human OP2 genes also have been cloned. Like the human OP1 gene, the protein coding region of the OP2 gene is contained on seven exons.

FIGS. 23C–23E compare the amino acid sequences of the mature mOP2 and hOP2 proteins, mOP2-A and hOP2-Ala. Identity is indicated by three dots (...). Like the mOP2 protein, the hOP2 protein shares significant homology (about 74% identity) with the amino acid sequence defining the OP1 active region (OPS or OP7, residues 43–139 and 38–139, respectively), and less homology with OP1-1BSer (about 58% identity). Both OP2 proteins share the conserved seven cysteine skeleton seen in the OP1 proteins. In addition, the OP2 proteins comprise an eighth cysteine residue within this region (see position 78 in FIG. 24B).

The greatest homology between sequences (about 74% identity, indicated by dots) occurs within the C-terminal active region defined by OPS and OP7. The OP1 and OP2 proteins share less amino acid sequence homology with the active regions of the CBMP2A and CBMP2B proteins. The OP1 proteins share only about 60% sequence identity with the CBMP2 proteins in this region; the OP2 proteins share only about 58% identity with the CBMP2 protein in this region. The CBMP2 proteins are most easily distinguished from the OP1/OP2 proteins in this active region by at least 9 nonconservative amino acid changes, in addition to numerous conservative amino acid changes which may have smaller effects on activity.
A preferred generic amino acid sequence useful as a subunit of a dimeric osteogenic protein capable of inducing endochondral bone or cartilage formation when implanted in a mammal in association with a matrix, and which incorporates the maximum homology between the identified OP1 and OP2 proteins (see FIG. 24) can be described by the sequence referred to herein as “OPX”, described below and in Seq. ID No. 30. OPX is a composite sequence designed from the four sequences presented in FIG. 24 (beginning at residue 38), and includes both the specific amino acid sequence created by the amino acid identity shared by the four OP1, OP2 species, as well as alternative residues for the variable positions within the sequence.

Cys Xaa Xaa His Glu Leu Tyr Val Ser Phe
1 5 10
Xaa Asp Leu Gly Thr Xaa Asp Trp Xaa Ile
15 20
Xaa Ala Pro Xaa Gly Tyr Xaa Ala Tyr Tyr Cys
25 30
Xaa Glu Gly Glu Cys Xaa Phe Pro Leu Xaa Ser
35 40
Xaa Met Asn Ala Thr Asn His Ala Ile Xaa
45 50
Xaa Gln Xaa Leu Val His Xaa Xaa Xaa Pro Xaa
55 60
Xaa Val Pro Lys Xaa Cys Cys Ala Pro Thr
65 70
Xaa Leu Xaa Ala Xaa Ser Val Leu Tyr Xaa
75 80
Xaa Asp Xaa Ser Xaa Asn Val Ile Leu Xaa Lys
85 90
Xaa Arg Xaa Asn Met Val Val Xaa Ala Cys Gly
95 100
Cys His,

and wherein Xaa at res. 2=(Lys or Arg); Xaa at res. 3=(Lys or Arg); Xaa at res. 11=(Arg or Glu); Xaa at res. 16=(Glu or Leu); Xaa at res. 19=(Ile or Val); Xaa at res. 23=(Glu or Glu); Xaa at res. 26=(Ala or Ser); Xaa at res. 35=(Ala or Ser); Xaa at res. 39=(Asn or Asp); Xaa at res. 41=(Tyr or Cys); Xaa at res. 50=(Val or Leu); Xaa at res. 52=(Ser or Thr); Xaa at res. 56=(Phe or Leu); Xaa at res. 57=(Ile or Met); Xaa at res. 58=(Asn or Lys); Xaa at res. 60=(Glu, Asp or Asn); Xaa at res. 61=(Thr, Ala or Val); Xaa at res. 65=(Pro or Ala); Xaa at res. 71=(Gln or Lys); Xaa at res. 73=(Asn or Ser); Xaa at res. 75=(Ile or Thr); Xaa at res. 80=(Phe or Tyr); Xaa at res. 82=(Asp or Ser); Xaa at res. 84=(Ser or Asn); Xaa at res. 85=(Lys or Arg); Xaa at res. 91=(Tyr or His); and Xaa at res. 97=(Arg or Lys).

The high degree of homology exhibited between the various OP1 and OP2 proteins suggests that the novel osteogenic proteins identified herein will purify essentially as OP1 does, or with only minor modifications of the protocols disclosed for OP1. Similarly, the purified mOP1, mOP2; and hOP2 proteins are predicted to have an apparent molecular weight of about 18 kDa as reduced single subunits, and an apparent molecular weight of about 36 kDa as oxidized dimers, as determined by comparison with molecular weight standards on an SDS-polyacrylamide electrophoresis gel. Unglycosylated dimers (e.g., proteins produced by recombinant expression in E. coli) are predicted to have an apparent molecular weight of about 27 kDa. There appears to be one potential N glycosylation site in the mature forms of the mOP2 and hOP2 proteins.

The identification of osteogenic proteins having an active region comprising eight cysteine residues also allows one to construct osteogenic polypeptide chains patterned after either of the following template amino acid sequences, or to identify additional osteogenic proteins having this sequence. The template sequences contemplated are “OPX-7C”, comprising the conserved six cysteine skeleton plus the additional cysteine residue identified in the OP2 proteins, and “OPX-8C”, comprising the conserved seven cysteine skeleton plus the additional cysteine residue identified in the OP2 proteins. The OPX-7C and OPX-8C sequences are described below and in Seq. ID Nos. 31 and 32, respectively. Each Xaa in these template sequences independently represents one of the 20 naturally-occurring L-isomer, α-amino acids, or a derivative thereof. Biosynthetic constructs patterned after this template readily are constructed using conventional DNA synthesis or peptide synthesis techniques well known in the art. Once constructed, osteogenic proteins comprising these polypeptide chains can be tested as disclosed herein.

“OPX-7C” (Sequence ID No. 31): Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 1 5 10 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 15 20 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 25 30 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 35 40 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 45 50 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 55 60 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 75 80 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 85 90 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 95 100

“OPX-8C” (Sequence ID No. 32 comprising additional five residues at the N-terminus, including a conserved cysteine residue):

Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 1 5 10 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 15 20 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 25 30 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 35 40 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa 45 50 55
A. General Consideration of Matrix Properties

The currently preferred carrier material is a xenogenic bone-derived particulate matrix treated as disclosed herein. This carrier may be replaced by either a biodegradable-synthetic or synthetic-inorganic matrix (e.g., HAP, collagen, tricalcium phosphate or polyactic acid, polyglycolic acid, polybutyric acid and various copolymers thereof.)

Studies have shown that surface charge, particle size, the presence of mineral, and the methodology for combining matrix and osteogenic protein all play a role in achieving successful bone induction. Perturbation of the charge by chemical modification abolishes the inductive response. Particle size influences the quantitative response of new bone; particles between 70 μm and 420 μm elicit the maximum response. Contamination of the matrix with bone mineral will inhibit bone formation. Most importantly, the procedures used to formulate osteogenic protein onto the matrix are extremely sensitive to the physical and chemical state of both the osteogenic protein and the matrix.

The sequential cellular reactions in the interface of the bone matrix/osteogenic protein implants are complex. The multitstep cascade includes: binding of fibrin and fibronectin to implanted matrix, migration and proliferation of mesenchymal cells, differentiation of the progenitor cells into chondroblasts, cartilage formation, cartilage calcification, vascular invasion, bone formation, remodeling, and bone marrow differentiation.

A successful carrier for osteogenic protein should perform several important functions. It should carry osteogenic protein and act as a slow release delivery system, accommodate each step of the cellular response during bone development, and protect the osteogenic protein from nonspecific proteolysis. In addition, selected materials must be biocompatible in vivo and preferably biodegradable; the carrier must act as a temporary scaffold until replaced completely by new bone. Polyactic acid (PLA), polyglycolic acid (PGA), and various combinations have different dissolution rates in vivo. In bones, the dissolution rates can vary according to whether the implant is placed in cortical or trabecular bone.

Matrix geometry, particle size, the presence of surface charge, and the degree of both intra-and inter-particle porosity are all important to successful matrix performance. It is preferred to shape the matrix to the desired form of the new bone and to have dimensions which span non-union defects. Rat studies show that the new bone is formed essentially by implanted matrix, migration and proliferation of mesenchymal cells, differentiation of the progenitor cells into chondroblasts, cartilage formation, cartilage calcification, vascular invasion, bone formation, remodeling, and bone marrow differentiation.

The matrix may comprise a shape-retaining solid made of loosely adhered particulate material, e.g., with collagen. It may also comprise a molded, porous solid, or simply an aggregation of close-packed particles held in place by surrounding tissue. Masticated muscle or other tissue may also be used. Large allogenic bone implants can act as a carrier for the matrix if their marrow cavities are cleaned and packed with particles and the dispersed osteogenic protein.

The preferred matrix material, prepared from xenogenic bone and treated as disclosed herein, produces an implantable material useful in a variety of clinical settings. In addition to its use as a matrix for bone formation in various orthopedic, periodontal, and reconstructive procedures, the matrix also may be used as a sustained release carrier, or as a collagenous coating for implants. The matrix may be shaped as desired in anticipation of surgery or shaped by the physician or technician during surgery. Thus, the material may be used for topical, subcutaneous, intraperitoneal, or intramuscular implants; it may be shaped to span a nonunion fracture or to fill a bone defect. In bone formation procedures, the material is slowly absorbed by the body and is replaced by bone in the shape of or very nearly the shape of the implant.

Various growth factors, hormones, enzymes, therapeutic compositions, antibiotics, and other bioactive agents also may be adsorbed onto the carrier material and will be released over time when implanted as the matrix material is slowly absorbed. Thus, various known growth factors such as EGF, PDGF, IGF, FGF, TGF-α, and TGF-β may be released in vivo. The material can be used to release chemotherapeutic agents, insulin, enzymes, or enzyme inhibitors.

B. Bone-Derived Matrices

1. Preparation of Demineralized Bone

Demineralized bone matrix, preferably bovine bone matrix, is prepared by previously published procedures (Sampath and Reddi (1983) Proc. Natl. Acad. Sci. USA 80:6591–6595). Bovine diaphyseal bones (age 1–10 days) are obtained from a local slaughterhouse and used fresh. The bones are stripped of muscle and fat, cleaned of periosteum, demarrowed by pressure with cold water, dipped in cold absolute ethanol, and stored at −20° C. They are then dried and fragmented by crushing and pulverized in a large mill. Care is taken to prevent heating by using liquid nitrogen. The pulverized bone is milled to a particle size in the range of 70–850 μm, preferably 150–420 μm, and is defatted by two washes of approximately two hours duration with three volumes of chloroform and methanol (3:1). The particulate bone is then washed with one volume of absolute ethanol and dried over one volume of anhydrous ether yielding defatted bone powder. The defatted bone powder is then demineralized by four successive treatments with 10 volumes of 0.5 N HCl at 4° C. for 40 min. Finally, neutralizing washes are done on the demineralized bone powder with a large volume of water.

2. Guanidine Extraction

Demineralized bone matrix thus prepared is extracted with 5 volumes of 4 M guanidine-HCl, 50 mM Tris-HCl, pH 7.0 for 16 hr. at 4° C. The suspension is filtered. The insoluble material is collected and used to fabricate the matrix. The material is mostly collagenous in nature. It is devoid of osteogenic or chondrogenic activity.

3. Matrix Treatments

“The major component of all bone matrices is Type-I collagen. In addition to collagen, demineralized bone extracted as disclosed above includes non-collagenous proteins which may account for 5% of its mass. In a xenogenic matrix, these noncollagenous components may present themselves a potent antigens, and may constitute immuno-
genic and/or inhibitory components. These components also may inhibit osteogenesis in allogenic implants by interfering with the developmental cascade of bone differentiation. It has been discovered that treatment of the matrix particles with a collagen fibril-modifying agent extracts potentially unwanted components from the matrix, and alters the surface structure of the matrix material. Useful agents include acids, organic solvents or heated aqueous media. Various treatments are described below. A detailed physical analysis of the effect these fibril-modifying agents have on demineralized, guanidine-extracted bone collagen particles is disclosed in copending U.S. patent application Ser. No. 483,913, filed Feb. 22, 1990, now U.S. Pat. No. 5,171,574, issued Dec. 15, 1992."

After contact with the fibril-modifying agent, the treated matrix is washed to remove any extracted components, following a form of the procedure set forth below:

1. Suspend in TBS (Tris-buffered saline) 1g/200 ml and stir at 4°C for 2 hrs; or in 6 M urea, 50 mM Tris-HCl, 500 mM NaCl, pH 7.0 (UTBS) or water and stir at room temperature (RT) for 30 minutes (sufficient time to neutralize the pFl);
2. Centrifuge and repeat wash step; and
3. Centrifuge; discard supernatant; water wash residue; and then lyophilize.

3.1 Acid Treatments

1. Trifluoroacetic Acid.
Trifluoroacetic acid is a strong non-oxidizing acid that is a known swelling agent for proteins, and which modifies collagen fibrils.

Bovine bone residue prepared as described above is sieved, and particles of the appropriate size are collected. These particles are extracted with various percentages (1.0% to 100%) of trifluoroacetic acid and water (v/v) at 0°C or room temperature for 1–2 hours with constant stirring. The treated matrix is filtered, lyophilized, or washed with water and then lyophilized.

Like trifluoroacetic acid, hydrogen fluoride is a strong acid and swelling agent, and also is capable of altering intra-particle surface structure. Hydrogen fluoride is also a known deglycosylating agent. As such, HF may function to increase the osteogenic activity of these matrices by removing the antigenic carbohydrate content of any glycoproteins still associated with the matrix after guanidine extraction.

Bovine bone residue prepared as described above is sieved, and particles of the appropriate size are collected. The sample is dried in vacuo over P2 Os, transferred to the reaction vessel and exposed to anhydrous hydrogen fluoride (10–20 ml/g of matrix) by distillation onto the sample at −70°C. The vessel is allowed to warm to 0°C and the reaction mixture is stirred at this temperature for 120 minutes. After evaporation of the hydrogen fluoride in vacuo, the residue is dried thoroughly in vacuo over KOH pellets to remove any remaining traces of acid. Extent of deglycosylation can be determined from carbohydrate analysis of matrix samples taken before and after treatment with hydrogen fluoride, after washing the samples appropriately to remove non-covalently bound carbohydrates. SDS-extracted protein from HF-treated material is negative for carbohydrate as determined by Con A blotting.

The deglycosylated bone matrix is next washed twice in TBS (Tris-buffered saline) or UTBS, water-washed, and then lyophilized.

Other acid treatments are envisioned in addition to HF and TFA. TFA is a currently preferred acidifying reagent in these treatments because of its volatility. However, it is understood that other, potentially less caustic acids may be used, such as acetic or formic acid.

3.2 Solvent Treatment

1. Dichloromethane.

Dichloromethane (DCM) is an organic solvent capable of denaturing proteins without affecting their primary structure. This swelling agent is a common reagent in automated peptide synthesis, and is used in washing steps to remove components.

Bovine bone residue, prepared as described above, is sieved, and particles of the appropriate size are incubated in 100% DCM or, preferably, 99.9% DCM/0.1% TFA. The matrix is incubated with the swelling agent for one or two hours at 0°C or at room temperature. Alternatively, the matrix is treated with the agent at least three times with short washes (20 minutes each) with no incubation.

2. Acetonitrile.

Acetonitrile (ACN) is an organic solvent, capable of denaturing proteins without affecting their primary structure. It is a common reagent used in high-performance liquid chromatography, and is used to elute proteins from silica-based columns by perturbing hydrophobic interactions.

Bovine bone residue particles of the appropriate size, prepared as described above, are treated with 100% ACN (1.0 g/30 ml) or, preferably, 99.9% ACN/0.1% TFA at room temperature for 1–2 hours with constant stirring. The treated matrix is then water-washed, or washed with urea buffer, or 4 M NaCl and lyophilized. Alternatively, the ACN or ACN/TFA treated matrix may be lyophilized without wash.

3. Isopropanol.

Isopropanol is also an organic solvent capable of denaturing proteins without affecting their primary structure. It is a common reagent used to elute proteins from silica HPLC columns.

Bovine bone residue particles of the appropriate size prepared as described above are treated with 100% isopropanol (1.0 g/30 ml) or, preferably, in the presence of 0.1% TFA, at room temperature for 1–2 hours with constant stirring. The matrix is then water-washed or washed with urea buffer or 4 M NaCl before being lyophilized.

4. Chloroform

Chloroform also may be used to increase surface area of bone matrix like the reagents set forth above, either alone or acidified.

Treatment as set forth above is effective to assure that the material is free of pathogens prior to implantation.

3.3 Heat Treatment

The currently most preferred agent is a heated aqueous fibril-modifying medium such as water, to increase the matrix particle surface area and porosity. The currently most preferred aqueous medium is an acidic aqueous medium having a pH of less than about 4.5, e.g., within the range of about pH 2–pH 4 which may help to “swell” the collagen before heating. 0.1% acetic acid, which has a pH of about 3, currently is most preferred. 0.1 M acetic acid also may be used.

Various amounts of delipidated, demineralized guanidine-extracted bone collagen are heated in the aqueous medium (1 g matrix/30 ml aqueous medium) under constant stirring in a water-jacketed glass flask, and maintained at a given temperature for a predetermined period of time. Preferred treatment times are about one hour, although exposure times of between about 0.5 to two hours appear acceptable. The temperature employed is held constant at a temperature.
within the range of about 37°C to 65°C. The currently preferred heat treatment temperature is within the range of about 45°C to 60°C.

After the heat treatment, the matrix is filtered, washed, lyophilized and used for implant. Where an acidic aqueous medium is used, the matrix also is preferably neutralized prior to washing and lyophilization. A currently preferred neutralization buffer is a 200 mM sodium phosphate buffer, pH 7.0. To neutralize the matrix, the matrix preferably first is allowed to cool following thermal treatment, the acidic aqueous medium (e.g., 0.1% acetic acid) then is removed and replaced with the neutralization buffer and the matrix agitated for about 30 minutes. The neutralization buffer then may be removed and the matrix washed and lyophilized (see infra). The effects of heat treatment on morphology of the matrix material is apparent from a comparison of the photomicrographs in FIG. 25 with those of FIG. 26. FIG. 25 illustrates the morphology of the successfully altered collagen surface treated with water heated to (25A) 37°C, (25B) 45°C, (25C) 55°C, and (25D) 65°C. The photomicrograph of FIG. 26 describes the morphology of untreated rat and bovine bone matrix (26A and 26B, respectively). As is evident from the micrographs, the hot aqueous treatment can increase the degree of micropitting on the particle surface (e.g., about 10-fold), as well as also substantially increasing the particle's porosity (compare FIGS. 26B and 25C, 25D). This alteration of the matrix particle's morphology substantially increases the particle surface area. Careful measurement of the pore and micropit sizes reveals that hot aqueous medium treatment of the matrix particles yields particle pore and micropit diameters within the range of 1 μm to 100 μm.

Characterization of the extract produced by the hot aqueous treatment reveals that the treatment also may be removing component(s) whose association with the matrix may interfere with new bone formation in vivo. FIG. 27 is a 214 nm absorbance tracing of the extract isolated from hot water-treated bovine matrix, and indicates the effect of each peak (or fraction) on in vivo bone formation.

The extract from a large scale preparative run (100 g bovine matrix, hot water-treated) was collected, acidified with 0.1% TFA, and run on a C-18 HPLC column, using a Millipore Delta Prep Cartridge. Fractions were collected at 50 ml intervals at a flow rate of 25 ml/min. and pooled appropriately to isolate the individual peaks in the tracing. Each of these fractions then was implanted with recombinant OP1 and an appropriate rat matrix carrier (see infra), and its effect on bone formation activity measured. Fraction 12 alone appears to inhibit bone formation in allogenic implants. The inhibitory activity appears to be dose dependent. It is possible that the removal of the inhibitory component(s) present in this peak may be necessary to support osteogenic activity in xenogenic implants.

FIG. 28 describes the influence of complete solvent extract from hot water-treated matrix on osteogenic activity as measured in 12-day implants by alkaline phosphatase activity (28A) and calcium content (28B). Rat carrier matrix and OP1 implanted without any extract is used as a positive control. The solvent extract obtained from 100 grams of hot water-treated bovine matrix was evaporated and taken up in 6 M of 50% acetonitrile/0.1% TFA. 100–300 μl aliquots then were combined with known amounts of recombinant OP1, and 25 mg of rat matrix carrier, and assayed (see infra). The results clearly show the extract inhibits new bone formation in a dose dependent manner.

The matrix also may be treated to remove contaminating heavy metals, such as by exposing the matrix to a metal ion chelator. For example, following thermal treatment with 0.1% acetic acid, the matrix may be neutralized in a neutralization buffer containing EDTA (sodium ethylenediaminetetraacetic acid), e.g., 200 mM sodium phosphate, 5 mM EDTA, pH 7.0. 5 mM EDTA provides about a 100-fold molar excess of chelator to residual heavy metals present in the most contaminated matrix tested to date. Subsequent washing of the matrix following neutralization appears to remove the bulk of the EDTA. EDTA treatment of matrix particles reduces the residual heavy metal content of all metals tested (Sb, As, Be, Cd, Cr, Cu, Co, Pb, Hg, Ni, Se, Ag, Zn, Tl) to less than about 1 ppm. Bioassays with EDTA-treated matrices indicate that treatment with the metal ion chelator does not inhibit bone inducing activity.

The collagen matrix materials preferably take the form of a fine powder, insoluble in water, comprising nonadherent particles. It may be used simply by packing into the volume where new bone growth or sustained release is desired, held in place by surrounding tissue. Alternatively, the powder may be encapsulated in, e.g., a gelatin or polyacrylic acid coating, which is absorbed readily by the body. The powder may be shaped to a volume of given dimensions and held in that shape by interadhering the particles using, for example, soluble, species-biocompatible collagen. The material may also be produced in sheet, rod, bead, or other macroscopic shapes.

Demineralized rat bone matrix used as an allogenic matrix in certain of the experiments disclosed herein, is prepared from several of the dehydrated diaphyseal shafts of rat femur and tibia as described herein to produce a bone particle size which passes through a 420 μm sieve. The bone particles are subjected to dissociative extraction with 4 M guanidine-HCl. Such treatment results in a complete loss of the inherent ability of the bone matrix to induce endochondral bone differentiation. The remaining insoluble material is used to fabricate the matrix. The material is mostly collagenous in nature, and upon implantation, does not induce cartilage and bone formation. All new preparations are tested for mineral content and osteogenic activity before use. The total loss of biological activity of bone matrix is restored when an active osteoinductive protein fraction or a substantially pure osteoinductive protein preparation is reconstituted with the biologically inactive insoluble collagenous matrix.

C. Synthetic Tissue-Specific Matrices

"In addition to the naturally-derived bone matrices described above, useful matrices also may be formulated synthetically if appropriately modified. One example of such a synthetic matrix is the porous, biocompatible, in vivo biodegradable synthetic matrix disclosed in copending U.S. Ser. No. 529,852, filed May 30, 1990, now U.S. Pat. No. 5,645,591, issued Jul. 8, 1997, the disclosure of which is hereby incorporated by reference. Briefly, the matrix comprises a porous crosslinked structural polymer of biocompatible, biodegradable collagen, most preferably tissue-specific collagen, and appropriate, tissue-specific glycosaminoglycans as tissue-specific cell attachment factors. Bone tissue-specific collagen (e.g., Type I collagen) derived from a number of sources may be suitable for use in these synthetic matrices, including soluble collagen, acid-soluble collagen, collagen soluble in neutral or basic aqueous solutions, as well as those collagen which are commercially available. In addition, Type II collagen, as found in cartilage, also may be used in combination with Type I collagen."
Glycosaminoglycans (GAGs) or mucopolysaccharides are polysaccharides made up of residues of hexosamines glycosidically bound and alternating in a more-or-less regular manner with either hexuronic acid or hexose moieties. GAGs are of animal origin and have a tissue specific distribution (see, e.g., Dodgson et al. in Carbohydrate Metabolism and its Disorders (Dickens et al., eds.) Vol. 1, Academic Press (1968)). Reaction with the GAGs also provides collagen with another valuable property, i.e., the ability to provoke an immune reaction (foreign body reaction) from an animal host.

Useful GAGs include those containing sulfate groups, such as hyaluronic acid, heparin, heparin sulfate, chondroitin 6-sulfate, chondroitin 4-sulfate, dermatan sulfate, and keratin sulfate. For osteogenic devices chondroitin 6-sulfate currently is preferred. Other GAGs also may be suitable for forming the matrix described herein, and those skilled in the art will either know or be able to ascertain other suitable GAGs using no more than routine experimentation. For a more detailed description of mucopolysaccharides, see Aspinall, Polysaccharides, Penguin Press, Oxford (1970).

Collagen can be reacted with a GAG in aqueous acidic solutions, preferably in diluted acetic acid solutions. By adding the GAG dropwise into the aqueous collagen dispersion, coprecipitates of tangled collagen fibrils coated with GAG results. This tangled mass of fibers then can be homogenized to form a homogeneous dispersion of fine fibers and then filtered and dried.

Insolubility of the collagen-GAG products can be raised to the desired degree by covalently cross-linking these materials, which also serves to raise the resistance to resorption of these materials. In general, any covalent cross-linking method suitable for cross-linking collagen also is suitable for cross-linking these composite materials, although cross-linking by a dehydrothermal process is preferred.

When dry, the cross-linked particles are essentially spherical with diameters of about 500 μm. Scanning electron microscopy shows pores of about 20 μm on the surface and 40 μm on the interior. The interior is made up of both fibrous and sheet-like structures, providing surfaces for cell attachment. The voids interconnect, providing access to the cells in the matrix. No activity is observed if the matrix is implanted in rats alone. In vivo bioassays performed in the rat model and bioassayed as described herein (e.g., as determined by calcium content, alkaline phosphatase levels and histology of 12-day implants, see Section V, infra).

### IV. FABRICATION OF OSTEGENIC DEVICE

The naturally sourced and recombinant proteins as set forth above, as well as other constructs, can be combined and dispersed, in a suitable matrix preparation using any of the methods described below. In general, 50–100 ng of active protein is combined with the inactive carrier matrix (e.g., 25 mg matrix for rat bioassays). Greater amounts may be used for large implants.

1. Ethanol Trifluoroacetic Acid Lyophilization

   In this procedure, osteogenic protein is solubilized in an ethanol trifluoroacetic acid solution (47.5% EtOH/0.01% TFA) and added to the carrier material. Samples are vortexed vigorously and then lyophilized. This method currently is preferred.

2. Acetonitrile Trifluoroacetic Acid Lyophilization

   This is a variation of the above procedure, using an acetonitrile trifluoroacetic acid (ACN/TFA) solution to solubilize the osteogenic protein that then is added to the carrier material. Samples are vigorously vortexed many times and then lyophilized.

3. Ethanol Precipitation

   Matrix is added to osteogenic protein dissolved in guanidine-HCl. Samples are vortexed and incubated at a low temperature (e.g., 4° C.). Samples are then further vortexed. Cold absolute ethanol (5 volumes) is added to the mixture which is then stirred and incubated, preferably for 30 minutes at −20° C. After centrifugation (microfuge, high speed) the supernatant is discarded. The reconstituted matrix is washed twice with cold concentrated ethanol in water (85% EtOH) and then lyophilized.

4. Urea Lyophilization

   For those osteogenic proteins that are prepared in urea buffer, the protein is mixed with the matrix material, vortexed many times, and then lyophilized. The lyophilized material may be used “as is” for implants.

5. Buffered Saline Lyophilization

   Osteogenic protein preparations in physiological saline may also be vortexed with thematrix and lyophilized to produce osteogenically active material.

   These procedures also can be used to adsorb other active therapeutic drugs, hormones, and various bioactive species to the matrix for sustained release purposes.

### V. BIOASSAY

The functioning of the various proteins and devices of this invention can be evaluated with an in vivo bioassay. Studies in rats show the osteogenic effect in an appropriate matrix to be dependent on the dose of osteogenic protein dispersed in the matrix. No activity is observed if the matrix is implanted alone. In vivo bioassays performed in the rat model also have shown that demineralized, guanidine-extracted xenogenic bone matrix materials of the type described in the literature generally are ineffective as a carrier, can fail to induce bone, and can produce an inflammatory and immunological response when implanted unless treated as disclosed above. In certain species (e.g., monkey) allogenic matrix materials also apparently are ineffective as carriers (Aspenberg, et al. (1988) J. Bone Joint Surgery 70;
The following sets forth various procedures for preparing osteogenic devices from the proteins and matrix materials as set forth above, and for evaluating their osteogenic utility.

A. Rat Model

1. Implantation

The bioassay for bone induction as described by Sampath and Reddi (1983) Proc. Natl. Acad. Sci. USA 80 6591–6595, herein incorporated by reference, may be used to monitor endochondral bone differentiation activity. This assay consists of implanting test samples in subcutaneous sites in recipient rats under ether anesthesia. Male Long-Evans rats, aged 28–32 days, were used. A vertical incision (1 cm) is made under sterile conditions in the skin over the thoracic region, and a pocket is prepared by blunt dissection. Approximately 25 mg of the test sample is implanted deep into the pocket and the incision is closed with a metallic skin clip. The day of implantation is designated as day one of the experiment. Implants were removed on day 12. The heterotropic site allows for the study of bone induction without the possible ambiguities resulting from the use of orthotropic sites. As disclosed herein, both allogenic (rat bone matrix) and xenogenic (bovine bone matrix) implants were assayed for bone forming activity. Allogenic implants were used in experiments performed to determine the specific activity of bone purified and recombinant osteogenic protein.

Bone inducing activity is determined biochemically by the specific activity of alkaline phosphatase and calcium content of the day 12 implant. An increase in the specific activity of alkaline phosphatase indicates the onset of bone formation. Calcium content, on the other hand, is proportional to the amount of bone formed in the implant. Bone formation therefore is calculated by determining the calcium content of the implant on day 12 in rats and is expressed as “bone forming units,” where one bone forming unit represents the amount of protein that is needed for half maximal bone forming activity of the implant on day 12. Bone induction exhibited by intact demineralized rat bone matrix is considered to be the maximal bone differentiation activity for comparison purposes in this assay.

2. Cellular Events

Successful implants exhibit a controlled progression through the stages of protein-induced endochondral bone development, including: (1) transient infiltration by polymorphonuclear leukocytes on day one; (2) mesenchymal cell migration and proliferation on days two and three; (3) chondrocyte appearance on days five and six; (4) cartilage matrix formation on day seven; (5) cartilage calcification on day eight; (6) vascular invasion, appearance of osteoblasts, and formation of new bone on days nine and ten; (7) appearance of osteoclasts, bone remodeling and dissolution of the implanted matrix on days twelve to eighteen; and (8) hematopoietic bone marrow differentiation in the ossicles on day twenty-one. The results show that the shape of the new bone conforms to the shape of the implanted matrix.

3. Histological Evaluation

Histological sectioning and staining is preferred to determine the extent of osteogenesis in the implants. Implants are fixed in Bouins Solution, embedded in paraffin, and cut into 6–8 μm sections. Staining with toluidine blue or hematoxylin/eosin demonstrates clearly the ultimate development of endochondral bone. Twelve day implants are usually sufficient to determine whether the implants contain newly induced bone.

4. Biological Markers

Alkaline phosphatase activity may be used as a marker for osteogenesis. The enzyme activity may be determined spectrophotometrically after homogenization of the implant. The activity peaks at 9–10 days in vitro and thereafter slowly declines. Implants showing no bone development by histology have little or no alkaline phosphatase activity under these assay conditions. The assay is useful for quantification and obtaining an estimate of bone formation quickly after the implants are removed from the rat. Alternatively, the amount of bone formation can be determined by measuring the calcium content of the implant.

5. Results

Histological examination of implants indicates that osteogenic devices containing the natural-sourced osteogenic protein or recombinant osteogenic protein have true osteogenic activity. Moreover, the osteogenic specific activity of recombinant OP1 homodimers matches that of the substantially pure natural-sourced material.

5.1 Bone Purified Osteogenic Protein

Implants containing osteogenic protein at several levels of purity have been tested to determine the most effective dose/purity level, in order to seek a formulation which could be produced on an industrial scale. As described supra, the results were measured by alkaline phosphatase activity level, calcium content, and histological examination and represented as bone forming units. Also as described supra, one bone forming unit represents the amount of protein that is needed for half maximal bone forming activity of the implant on day 12. The bone forming activity elicited by intact rat demineralized bone matrix is considered to be the maximal bone differentiation activity for comparison purposes in this assay.

Dose curves were constructed for bone inducing activity in vivo by assaysing various concentrations of protein purified from bone at each step of the purification scheme. FIG. 11 shows representative dose curves in rats. Approximately 10–12 μg of the TSK-fraction (FIG. 11C), 3–4 μg of heparin-Sepharose-II fraction (FIG. 11B), 0.5–1 μg of the C-18 column fraction (FIG. 11A), and 25–50 ng of gel eluted highly purified 30 kDa protein is needed for unequivocal bone formation (half maximum activity) when implanted with 25 mg of matrix. Subsequent additional experiments have measured a half maximum activity of about 21–25 ng protein per 25 mg matrix for the highly purified, gel eluted 30 kDa osteogenic protein (see U.S. Pat. No. 5,011,691.

Thus, 0.8–1.0 ng of highly purified osteogenic protein per mg of implant matrix is sufficient to exhibit half maximal bone differentiation in vivo. 50 to 100 ng per 25 mg of implant normally is sufficient to produce maximum endochondral bone. Thus, 2 to 4 ng osteogenic protein per mg of implant matrix is a reasonable dosage, although higher dosages may be used.

As shown in FIG. 17, osteogenic devices comprising undeglycosylated osteogenic protein are osteogenically active. Compare FIGS. 17B (showing carrier and glycosylated protein) and 17C (showing carrier and deglycosylated protein). Arrows indicate osteoblasts. FIG. 17A is a control where carrier alone was implanted. No bone formation is evident in this control implant.

5.2 Recombinant Osteogenic Protein

Homodimers of the various fusion constructs disclosed herein and expressed in E. coli all are osteogenically active. In addition, osteogenic activity is present with OP1A-CBMP2B1, OP1B-CBMP2B1, and OP1C-CBMP2B2 pro-
tein combinations. In addition, dimeric species of the truncated analog active regions (COP5 and COP7, disclosed in U.S. Pat. No. 5,011,691), alone or in combination, also induce osteogenesis as determined by histological examination.

Recombinant OP1 expressed from different mammalian cell sources and purified to different extents (1-5% pure to 30-90% pure) were tested for osteogenic activity in vivo as set forth above using 25 mg matrix. Table 9 below shows the histology score for OP1 expressed in all three cell types.

**TABLE 9**

<table>
<thead>
<tr>
<th>Mammalian Cells</th>
<th>OP1 Subunit</th>
<th>OP1 Protein Concentration* (ng/implant)</th>
<th>Histology Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSC40-tsA58</td>
<td>18 kDa</td>
<td>32.5</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>(70-90% pure)</td>
<td>65.0</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>130.0</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>260.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>16 kDa</td>
<td>12.5</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>(30-40% pure)</td>
<td>25.0</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50.0</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200.0</td>
<td>100</td>
</tr>
<tr>
<td>CHO</td>
<td>16-20 kDa</td>
<td>50.0</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>(less than 5% pure)</td>
<td>100.0</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>18 kDa</td>
<td>25.0</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>(less than 5% pure)</td>
<td>50.0</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200.0</td>
<td>90</td>
</tr>
<tr>
<td>COS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

10-30%: moderate bone formation
30-80%: extensive bone formation
above 80%: evidence of hematopoietic bone marrow recruitment.
*estimated by immunoblots or gel scanning

The histology scores detailed in Table 9 show that OP1 is active regardless of cell source, and that the activity mimics that of natural-sourced bovine osteogenic protein (see discussion of FIG. 31 and 32, infra.) Moreover, the bone-inducing activity is highly reproducible and dose dependent.

Additional bioassays, performed using highly purified OP1 (90% pure), and formulated with rat matrix by the acetonitrile/TEA method, suggest that CHO-produced OP1 shows slightly more bone-inducing activity when compared to BSC-derived OP1 preparations (at lower protein concentrations). Finally, numerous bioassays have been conducted with the various degraded species identified in the different OP1 preparations (e.g., OP1-16Ala, OP1-16Val, OP1-16Ser, OP1-16Leu and OP1-16Met). Significant variations in bone inducing activity, as measured by calcium content or histology, could not be detected among these different OP1 species.

Further evidence of the bone-forming activity of recombinant OP1 is provided in the photomicrographs of FIGS. 29 and 30. FIGS. 29A-F are photomicrographs recording the histology of allogenic implants using recombinant OP1 expressed from COS, BSC, and COS cells. The micrographs (magnified 220x), provide graphic evidence of the full developmental cascade induced by the osteogenic proteins of this invention, confirming that recombinantly produced OP1 alone is sufficient to induce endochondral bone formation, when implanted in association with a matrix. As evidenced in FIG. 29A, allogenic implants that do not contain OP1 show no new bone formation at 12 days post implant. Only the implanted bone matrix (m) and surrounding mesenchyme are seen. Conversely, implants containing OP1 already show evidence of extensive chondrogenesis by 7 days post implant (FIG. 29B, 500 ng BSC-produced protein, 30% pure). Here, newly formed cartilage cells, chondroblasts (Cb) and chondrocytes (Cy) are in close contact with the matrix (m). By 9 days post implant endochondral bone differentiation, cartilage calcification, hypertrophy of chondrocytes, vascular invasion, and the onset of new bone formation are all evident (FIG. 29C, 220 ng COS-produced protein, approx. 5% pure). Invading capillaries (e) and the appearance of basophilic osteoblasts (indicated by arrows) near the vascular endothelium are particularly evident. By 12 days post implant extensive bone formation and remodeling has occurred (FIG. 29D (220x), and 29E (400x), GHO-produced protein, approx. 60% pure). The newly formed bone laid down by osteoblasts is being remodeled by multinucleated osteoclasts (Oc), and the implanted matrix is being reabsorbed and replaced by remodeled bone. Bone marrow recruitment in the newly formed ossicles is also evident. Finally, hematopoietic bone marrow differentiation within the ossicles can be seen by 22 days post implant (FIG. 29F, 500 ng BSC-produced protein, 30% pure). By this time most of the implanted matrix (m) has been resorbed and is occupied by newly-formed bone containing ossicles filled with bone marrow elements including erythrocytic and granulocytic series and megakaryocytes. Similar histological observations have been made for implants incorporating greater than 90% pure OP1 preparations.

FIG. 30 is a photomicrograph showing the histology at 12 days post implant for a xenogenic implant using hot water-treated bovine matrix and OP1 (BSC-produced). The recruitment of hematopoietic bone marrow elements is evident in the photomicrograph, showing that the bone forming activity of xenogenic implants with OP1 parallels that of allogenic implants (compare FIG. 30 with FIGS. 29D and 29E).

The cellular events exhibited by the OP1 matrix implants and evidenced in FIGS. 29 and 30 truly mimic the endochondral bone differentiation that occurs during the foetal development. Although endochondral bone differentiation has been the predominant route, there is also evidence for intra-membraneous bone formation at the outer surface of the implant.

FIGS. 31 and 32 describe the dose dependence of osteogenic activity for 12-day implants comprising recombinant OP1 expressed from different mammalian cell sources, as determined by specific activity of alkaline phosphatase and calcium content of allogenic implants (FIG. 31) and xenogenic implants of this invention (FIGS. 32A and 32B, respectively). In all cases, OP1 protein concentration (quantitated by immuno blot staining or by gel scanning), is represented in nanograms. In each case, bone inducing activity is specific to OP1 in a dose dependent manner in all cells. Moreover, osteogenic activity of the mammalian cell-produced protein mimics that of the natural-sourced material. Highly purified gel-eluted osteogenic bovine protein, purified as disclosed herein and in U.S. Pat. Nos. 4,968,590 and 5,011,691, has a half maximal activity of at least about 0.8-1 ng/mg matrix (20-25 ng protein/25 mg matrix). As can be seen in Table 9 and FIGS. 31 and 32, even partially purified recombinantly produced OP1 falls within this range of osteogenic activity (about 20-30 ng/25 mg matrix).

**B. Feline Model**

The purpose of this study is to establish a large animal efficacy model for the testing of the osteogenic devices of the invention, and to characterize repair of massive bone defects and simulated fracture non-union encountered fre-
quent in the practice of orthopedic surgery. The study is designed to evaluate whether implants of osteogenic protein with a carrier can enhance the regeneration of bone following injury and major reconstructive surgery by use of this large mammal model. The first step in this study design consists of the surgical preparation of a femoral osteotomy defect which, without further intervention, would consistently progress to non-union of the simulated fracture defect. The effects of implants of osteogenic devices into the created bone defects are evaluated by the study protocol described below. While this and the rabbit study, described infra, use allometric matrices as carrier material, appropriate treatment as described herein of any bone-derived matrix material is anticipated to render the matrix suitable for xenogenic implants. Similarly, while the osteogenic protein used in this and the rabbit study is BOP; it is anticipated that any of the osteogenic proteins disclosed herein may be substituted.

1. Procedure

Sixteen adult cats each weighing less than 10 lbs. undergo unilateral preparation of a 1 cm bone defect in the right femur through a lateral surgical approach. In other experiments, a 2 cm bone defect was created. The femur is immediately internally fixed by lateral placement of an 8-hole plate to preserve the exact dimensions of the defect. There are three different types of materials implanted in the surgically created cat femoral defects: group I (n=3) is a control group which undergoes the same plate fixation with implants of 4 M guanidine-HCl-treated (inactivated) cat demineralized bone matrix powder (GuHCl-DBM) (360 mg); group II (n=3) is a positive control group implanted with biologically active demineralized bone matrix powder (DBM) (360 mg); and group III (n=10) undergoes a procedure identical to groups I—II, with the addition of osteogenic protein onto each of the GuHCl-DBM carrier samples. To summarize, the group III osteogenic protein-treated animals are implanted with exactly the same material as the group I animals, but with the singular addition of osteogenic protein.

All animals are allowed to ambulate ad libitum within their cages post-operatively. All cats are injected with tetracycline (25 mg/kg, subcutaneously (SQ) each week for four weeks) to aid in bone laboratory reaction to the implanted device. In this bone study and showed no evidence of clinical or histological evidence of any general systemic effects or any effects on the musculature at mid-femur (adjacent to defect site in experimental femur). No unusual or pathological lesions are seen in any of the tissues. Mild lesions seen in the quadriceps muscles are compatible with healing responses to the surgical manipulation at the defect site. Pulmonary edema is attributable to the euthanasia procedure. There is no evidence of any general systemic effects or any effects on the specific organs examined.

2. Radiomorphometrics

In vivo radiomorphometric studies are carried out immediately post-op at 4, 8, 12 and 16 weeks by taking a standardized X-ray of the lightly anesthetized animal positioned in a cushioned X-ray jig designed to consistently produce a true antero-posterior view of the femur and the osteotomy site. All X-rays are taken in exactly the same fashion and in exactly the same position on each animal. Bone repair is calculated as a function of mineralization by means of random point analysis. A final specimen radiographic study of the excised bone is taken in two planes after sacrifice. X-ray results are shown in FIG. 12, and displaced as percent of bone defect repair. To summarize, at 16 weeks, 60% of the group III femurs are united with average 80% bone defect regeneration (FIG. 12, sec. A). By contrast, the group I GuHCl-DBM negative-control implants exhibit no bone growth at four weeks, less than 10% at eight and 12 weeks, and 16% (±10%) at 16 weeks with one of the five exhibiting a small amount of bridging bone (FIG. 12, sec. B). The group II DBM positive-control implants exhibited 18% (±3%) repair at four weeks, 35% at eight weeks, 50% (±10%) at 12 weeks and 70% (±12%) by 16 weeks, a statistical difference of p<0.01 compared to osteogenic protein at every month. One of the three (33%) is united at 16 weeks (FIG. 12, sec. C.)

3. Biomechanics

Excised test and normal femurs are immediately studied by bone densitometry, or wrapped in two layers of saline-soaked towels, placed into sealed plastic bags, and stored at -20° C. until further study. Bone repair strength, load to failure, and work to failure are tested by loading to failure on a specially designed steel 4-point bending jig attached to an Instron testing machine to quantitate bone strength, stiffness, energy absorbed and deformation to failure. The study of test femurs and normal femurs yield the bone strength (load) in pounds and work to failure in joules. Normal femurs exhibit a strength of 96 (±12) pounds. Osteogenic protein-implanted femurs exhibit 35 (±4) pounds, but when corrected for surface area at the site of fracture (due to the "hourglass" shape of the bone defect repair) this correlated closely with normal bone strength. Only one demineralized bone specimen was available for testing with a strength of 25 pounds, but, again, the strength correlated closely with normal bone when corrected for fracture surface area.

4. Histomorphometry/Histology

Following biomechanical testing the bones are immediately sliced into two longitudinal sections at the defect site, weighed, and the volume measured. One-half is fixed for standard calcified bone histomorphometrics with fluorescent stain incorporation evaluation, and one-half is fixed for decalcified hemotoxylin/eosin stain histology preparation.

5. Biochemistry

Selected specimens from the bone repair site (n=6) are homogenized in cold 0.15 M NaCl, 3 mM NaHCO3, pH 9.0 by a Spex freezer mill. The alkaline phosphatase activity of the supernatant and total calcium content of the acid soluble fraction of sediment are then determined.

6. Histopathology

The final autopsy reports reveal no unusual or pathologic findings noted at necropsy of any of the animals studied. A portion of all major organs are preserved for further study. A histopathological evaluation is performed on samples of the following organs: heart, lung, liver, both kidneys, spleen, both adrenals, lymph nodes, left and right quadriceps muscles at mid-femur (adjacent to defect site in experimental femur). No unusual or pathological lesions are seen in any of the tissues. Mild lesions seen in the quadriceps muscles are compatible with healing responses to the surgical manipulation at the defect site. Pulmonary edema is attributable to the euthanasia procedure. There is no evidence of any general systemic effects or any effects on the specific organs examined.

7. Feline Study Summary

The 1 cm and 2 cm femoral defect cat studies demonstrate that devices comprising a matrix containing disposed osteogenic protein can: (1) repair a weight-bearing bone defect in a large animal; (2) consistently induce bone formation shortly following (less than two weeks) implantation; and (3) induce bone by endochondral ossification, with a strength equal to normal bone, on a volume for volume basis. Furthermore, all animals remained healthy during the study and showed no evidence of clinical or histological laboratory reaction to the implanted device. In this bone defect model, there was little or no healing at control bone
implant sites. The results provide evidence for the successful use of osteogenic devices to repair large, non-union bone defects.

C. Rabbit Model:

1. Procedure and Results

Eight mature (less than 10 lbs) New Zealand White rabbits with epiphyseal closure documented by X-ray were studied. The purpose of this study is to establish a model in which there is minimal or no bone growth in the control animals, so that when bone induction is tested, only a strongly inductive substance will yield a positive result. Defects of 1.5 cm are created in the rabbits, with implantation of: osteogenic protein (n=5), DBM (n=8), GuHCl-DBM (n=6), and no implant (n=10). Six osteogenic protein implants are supplied and all control defects have no implant placed.

Of the eight animals (one animal each was sacrificed at one and two weeks), 11 ulnae defects are followed for the full course of the eight week study. In all cases (n=7) following osteo-periosteal bone resection, the no implant animals establish no radiographic union by eight weeks. All no implant animals develop a thin “shell” of bone growing from surrounding bone present at four weeks and, to a slightly greater degree, by eight weeks. In all cases (n=4), radiographic union with marked bone induction is established in the osteogenic protein-implanted animals by eight weeks. As opposed to the no implant repairs, this bone is in the site of the removed bone.

Radiomorphometric analysis reveal 90% osteogenic protein-implant bone repair and 18% no-implant bone repair at sacrifice at eight weeks. At autopsy, the osteogenic protein bone appears normal, while “no implant” bone sites have only a soft fibrous tissue with no evidence of cartilage or bone repair in the defect site.

2. Allograft Device

In another experiment, the marrow cavity of the 1.5 cm ulnar defect is packed with activated osteogenic protein rabbit bone powder and the bones are allografted in an intercalary fashion. The two control ulnae are not healed by eight weeks and reveal the classic “ivory” appearance. In distinct contrast, the osteogenic protein-treated implants “disappear” radiographically by four weeks with the start of remineralization by six to eight weeks. These allografts heal at each end with mild proliferative bone formation by eight weeks.

This type of device serves to accelerate allograft repair.

3. Summary

These studies of 1.5 cm osteo-periosteal defects in the ulnae of mature rabbits show that: (1) it is a suitable model for the study of bone growth; (2) “no implant” or GuHCl negative control implants yield a small amount of periosteal-type bone, but not medullary or cortical bone growth; (3) osteogenic protein-implanted rabbits exhibited proliferative bone growth in a fashion highly different from the control groups; (4) initial studies show that the bones exhibit 50% of normal bone strength (100% of normal correlated vol:vol) at only eight weeks after creation of the surgical defect; and (5) osteogenic protein-allograft studies reveal a marked effect upon both the allograft and bone healing.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 33

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1822 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: HOMO SAPIENS
(F) TISSUE TYPE: HIPPOCAMPUS

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 49 .. 1341
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= “OP1”
/evidence= EXPERIMENTAL
/standard_name= “OP1”
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GCTCAGAGCTGCGGGGCA CGCGCCGGCGC GCGCGGCCAG CTGCTGTTG CGCGTGCA GAG
57

GCC TCA CTG CGA GCT GCG GCG CAC AGC TCC GTC GCG CTC TGG GCA
Ser Leu Arg Arg Arg Thr Gln Arg Gln Ser Tyr Gin Leu Trp Ala
Arg Ser Leu Arg Ala Ala Pro His Ser Phe Val Ala Leu Trp Ala
Arg Ser Leu Arg Ala Ala Pro His Ser Phe Val Ala Leu Trp Ala
ARG Ser Leu Arg Ala Ala Pro His Ser Phe Val Ala Leu Trp Ala

105

CCC CTG TCC CTG CGC TCC GCC CTC GAC TGC GAC GAC AAC
Pro Leu Phe Leu Leu Ala Leu Ala Asp Phe Ser Leu Asp Asn
Pro Leu Phe Leu Leu Ala Leu Ala Asp Phe Ser Leu Asp Asn

153

GAG GTG CAC TCG AGC TTC ATC CAC CGC CTC CAC GCG GAC GAG CCG
Glu Val His Ser Ser Phe Ile His Arg Arg Leu Ser Gin Glu Arg
Glu Val His Ser Ser Phe Ile His Arg Arg Leu Ser Gin Glu Arg

201

253

GCC GCG CAC CTC CAG GCC AAC AAG CAC TCG GCC CCC ATG TTC ATG
Thr Met Arg His Arg Ser Leu Met Gin Ser Leu Ser Met Arg His Arg
Thr Met Arg His Arg Ser Leu Met Gin Ser Leu Ser Met Arg His Arg

249

305

CTG GAC TCT TAC AAC GCC ATG GCG GTG GAG GGC GCG GCC GGC GGC
Leu Asp Leu Tyr Asn Leu Met Arg Gin Gin Met Gin Gin Gin Gin
Leu Asp Leu Tyr Asn Leu Met Arg Gin Gin Met Gin Gin Gin Gin

345

393

GCC GAG GCC TGC TCC TAC CCC TAC GCC AGC GCC ACG GCC CTC AAC
Gln Gly Phe Ser Tyr Pro Tyr Lys Ala Val Phe Ser Thr Gin
Gln Gly Phe Ser Tyr Pro Tyr Lys Ala Val Phe Ser Thr Gin

441

489

CCC CCT TCG GCC CAG CAT ACC CAT TCC ACC AAC CAC GCG GCC GCA
Pro Pro Leu Ala Leu Ala Gin Ser Ser His Pro His Met Phe Met
Pro Pro Leu Ala Leu Ala Gin Ser Ser His Pro His Met Phe Met

537

585

CCA GAA GGG GAA GCT CAC AGC GCC GAA TCG GCG ACG ATC TAC AAC
Pro Gly Gin Ala Val Thr Ala Ala Gin Leu Thr Ala Ala Gin Leu
Pro Gly Gin Ala Val Thr Ala Ala Gin Leu Thr Ala Ala Gin Leu

633

681

CAG GTG CTC CAG GAG CTC TGG GCC AGA TCG CAT CTG TCC GCC AGC
Gln Val Leu Gin Leu Gin Arg Leu Ser Asp Leu Phe Leu Leu
Gln Val Leu Gin Leu Gin Arg Leu Ser Asp Leu Phe Leu Leu

729

777

GAC AAC COT ACC CTG TCG GCC TCG GAG GGC TGG CTG TCT CTT GAC
Asp Ser Arg Thr Leu Thr Ala Ser Gin Gin Leu Leu Gin Leu Gin
Asp Ser Arg Thr Leu Thr Ala Ser Gin Gin Leu Leu Gin Leu Gin

825

873

ATT AGG GCC CTG ATT GCG CGC CAC GCG GCC GAC ACG GCC CAC AAC
His Leu Ala Leu Leu Gin Arg His Gin Phe Gin Gin Gin Gin
His Leu Ala Leu Leu Gin Arg His Gin Phe Gin Gin Gin Gin

921

977
```
CGG TCC AGC GGG AGC AAA CAG CGC AGC AAC CAG CGC TCC AAG AGC CCC
Arg Ser Thr Gly Ser Lys Gln Arg Ser Gln Arg Ser Lys Thr Pro 969
Arg Ser Thr Gly Ser Lys Gln Arg Ser Gln Arg Ser Lys Thr Pro 969
AAG AAC CAG GAA GCC TCG CGG ATG GCC AAC GVG GCA GAG AAC AGC
Lys Asn Glu Glu Ala Leu Arg Met Ala Asn Val Ala Glu Asn Ser Ser 1017
Lys Asn Glu Glu Ala Leu Arg Met Ala Asn Val Ala Glu Asn Ser Ser 1017
AGC GAC CAG AGG GCC TGT AAG AAC CAG CGG TAT GCC AAC AGC
Ser Asp Gln Arg Glu Ala Cys Lys His Glu Leu Tyr Val Ser Phe 1065
Ser Asp Gln Arg Glu Ala Cys Lys His Glu Leu Tyr Val Ser Phe 1065
CGA GAC TCG GCC TGG CAG GAC TGC ATC ATC GCG CCT GAA GGC TAC GCC
Arg Asp Leu Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala 1113
Arg Asp Leu Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala 1113
GCC TAC TAC TGT GAG GGG GAG TGC GCC AAG GTG ACG GAA AAC
Ala Tyr Tyr Cys Gly Glu Ala Pro Leu Asn Ser Thr Met 1161
Ala Tyr Tyr Cys Gly Glu Ala Pro Leu Asn Ser Thr Met 1161
AAG GCC ACC AAC CAC GCC ATC GTC CAG ACG CGT TGC CAT ACG AAC
Asn Ala Thr Asn His Ala Ile Val Glu Thr Leu His Phe Ile Asn 1209
Asn Ala Thr Asn His Ala Ile Val Glu Thr Leu His Phe Ile Asn 1209
CGG GAA ACG TGG CCC AAG CCC TGC TGT GCG CCC AAG CAG CTC ACT GCC
Pro Glu Thr Val Pro Lys Pro Cys Ala Pro Thr Glu Leu Asn Ala 1257
Pro Glu Thr Val Pro Lys Pro Cys Ala Pro Thr Glu Leu Asn Ala 1257
ATC TCC TGT CTC TAC GAT GAC GCC TCG GCC AGC TCC AAC TCC TAC
Ile Ser Val Leu Tyr Phe Asp Asp Ser Ser Asn Val Ile Leu Lys Lys 1305
Ile Ser Val Leu Tyr Phe Asp Asp Ser Ser Asn Val Ile Leu Lys Lys 1305
TAC AGA AAC ATG GTG GTC CCC GCC GCT GTC TGC TAC TCT CTC GC
Tyr Arg Asn Met Val Arg Val Val Ala Val Val Ala Ala Asp 1351
Tyr Arg Asn Met Val Arg Val Val Ala Val Val Ala Ala Asp 1351
GAACATCACC ACCCTTTGG GCGAAGTTTT TTGTGTACCT CCATGGTCG CCTGGCCAG
1411
GAACACGAG ACCAAGCGTT TTTGTGTGAA CCTTCCCCCT CCTATCCCCCA ACTTTAAGGG 1471
TGTGAACTTA TTAGAACA TGACAGCACT ATGCTTTTGT ATCACTTTTAGT CATGGAACG
1531
ATCCAGAAG CACAGCTCTA CAAGCTGGCC AGGCAAACAC TACCAAGAAA AAAAAACAC 1591
GCACATAGAA AATGTGGGCG GCCGTCATC TGCTGGGAAA GCTCAACCT AGTACGAGACT
1651
GCTTCCAGA GCTGTAATAG AGGCCTGATC ACCAGCGGCC CCCCAGGCTG GAGGAAGGG 1711
GCGTGAGCA GCCGCTGGC TCTGGGAAAG GAAATTTGAC CCCAGAGTTC 1771
CTGTAATAAAA TGTCAACATA AAACGATAGA ATGAAAAAAA AAAAAAAA A 1822
(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 431 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
Met His Val Arg Ser Leu Arg Ala Ala Pro His Ser Phe Val Ala
1  5  10  15
Leu Trp Ala Pro Leu Phe Leu Leu Arg Ser Ala Leu Ala Asp Phe Ser
20  25  30
Leu Asp Asn Glu Val His Ser Ser Phe Ile His Arg Arg Leu Arg Ser
35  40  45
Gln Glu Arg Arg Glu Met Glu Arg Glu Ile Leu Ser Ile Leu Gly Leu
50  55  60
Pro His Arg Pro Arg Pro His Leu Glu Gly Lys His Asn Ser Ala Pro
65  70  75  80
Met Phe Met Leu Asp Leu Tyr Asn Ala Met Ala Val Glu Gly Gly
85  90  95
Gly Pro Gly Gly Gln Gly Phe Ser Tyr Pro Tyr Lys Ala Val Phe Ser
100 105 110
Thr Gln Gly Pro Pro Leu Ala Ser Leu Gln Aep Ser His Phe Leu Thr
115 120 125
Asp Ala Asp Met Val Met Ser Phe Val Asn Leu Val Glu His Aep Lys
130 135 140
Glu Phe Phe His Pro Arg Tyr His His Arg Glu Phe Arg Phe Aep Leu
145 150 155 160
Ser Lys Ile Pro Glu Gly Glu Ala Val Thr Ala Ala Glu Phe Arg Ile
165 170 175
Tyr Lys Asp Tyr Ile Arg Glu Arg Phe Asp Asn Glu Thr Phe Arg Ile
180 185 190
Ser Val Tyr Gln Val Leu Gln His Leu Gly Arg Glu Ser Aep Leu
195 200 205
Phe Leu Leu Aep Ser Arg Thr Leu Trp Ala Ser Glu Gly Trp Leu
210 215 220
Val Phe Asp Ile Thr Ala Thr Ser Asn His Trp Val Val Asn Pro Arg
225 230 235 240
His Asn Leu Gly Leu Gln Leu Ser Val Glu Thr Leu Aep Gly Gln Ser
245 250 255
Ile Asn Pro Lys Leu Ala Gly Leu Ile Gly Arg His Gly Pro Glu Asn
260 265 270
Lys Glu Pro Phe Met Val Ala Phe Phe Lys Ala Thr Glu Val His Phe
275 280 285
Arg Ser Ile Arg Ser Thr Gly Ser Lys Gln Arg Ser Glu Asn Arg Ser
290 295 300
Lys Thr Pro Lys Asn Gln Glu Ala Leu Arg Met Ala Asn Val Ala Glu
305 310 315 320
Asn Ser Ser Ser Asp Gln Arg Glu Ala Cys Lys Lys His Glu Leu Tyr
325 330 335
Val Ser Phe Arg Asp Leu Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu
340 345 350
Gly Tyr Ala Ala Tyr Tyr Cys Glu Gly Glu Cys Ala Phe Pro Leu Asn
355 360 365
Ser Tyr Met Asn Ala Thr Asn His Ala Ile Val Glu Thr Leu Val His
370 375 380
Phe Ile Asn Pro Glu Thr Val Pro Lys Pro Cys Cys Ala Pro Thr Gln
385 390 395 400
Leu Asn Ala Ile Ser Val Leu Tyr Phe Asp Asp Ser Ser Asn Val Ile
405 410 415
Leu Lys Lys Tyr Arg Asn Met Val Val Arg Ala Cys Gly Cys His
420 425 430

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)
(vi) ORIGINAL SOURCE:
(A) ORGANISM: homo sapiens

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3192..3730
(D) OTHER INFORMATION: /label= EXON-1
   /note= "START CODON BEGINS AT POSITION 3313"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 10413..10414
(D) OTHER INFORMATION: /label= GAP-1
   /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
THIS SEQUENCE."

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 10696..10891
(D) OTHER INFORMATION: /label= EXON-2

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 10960..10961
(D) OTHER INFORMATION: /label= GAP-2
   /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN
THIS SEQUENCE."

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 11059..11211
(D) OTHER INFORMATION: /label= EXON-3

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 11351..11352
(D) OTHER INFORMATION: /label= GAP-3
   /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
THIS SEQUENCE."

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 11420..11617
(D) OTHER INFORMATION: /label= EXON-4

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 11721..11722
(D) OTHER INFORMATION: /label= GAP-4
   /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
THIS SEQUENCE."

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 13354 .. 13436
(D) OTHER INFORMATION: /label= EXON-5

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 15044..15160
(D) OTHER INFORMATION: /label= EXON-6

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 17245..17410
(D) OTHER INFORMATION: /label= EXON-7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TCAACCGGTC TCTTTAGGTT TTGGCTGTGC TTATTACTAT TCATTCAACA GGTACTAATT
ACTAGAATCC TTGCTTAGAC ACTGACGGAT CAGTTGTTTC ATATGTAAAT TGTAGCACCA
GAGGACTGC TGTGTCAGCA GCTCAGAATA GGCTCAGGTG AGATGCACAA AGAAGGGTAA
ACTAGATGC TTGGTCAGCA ACTGACGGAT CAGTGTTTTC ATATGTAAAT TGTGACCA
AGACCTGCTG CCCCCTCCCC CAGCTCTACC TGCTTTGAGA GATCCCTCCA AAAGATTTGA 240
AGATGCTAAA AAGAGGAGGACTACTAGCTA AGACAGGGCC TGGTTTTGCT CTTATATTAT 300
TGGATCATTG GAGAAAGAAC ACCCTCTTTT TCTCTCTAGGG CAGCTCTGAC ATAGCTGACC 360
TAGAAGGCCC GGGCTGGAGGG GGGTTGCGAG GGGCTGGGAC GGGCTGGGCT TGCTTTTCCC 420
TGCTCTCTTC TGCTCTCTCA GAGCTCAGTT CAGTGCACAG TGCTCAGAAA CTCTCCCCCT 540
CCACCTGACA TCACTCTACCC CTGGTTATGAC GCACCTCTCC CCCAAGCTGG TTGTACTATT 600
TGCAAGGCTGA AGTGACCCAG GCTCTCCAGT TGTACACTCT CAGATGGACC CTGGAGCACT 660
TGCCACCTGG TTTGATATGAC CTGGTTGCTAC TTTGCTGCTTG CTTTGCTGCTG 720
CAAGCTTCTG GTCAGCCCTGC CTGAGTGGTA CAGTGGCATC TCCTCCAAAA AGCCCTACCT 780
ACGCTGCTGA AGTGACCCAG GCTCTCCAGT TGTACACTCT CAGATGGACC CTGGAGCACT 840
TGGCTTCCTGG GCTGGAGGCC GCTCTTGATGC GAGTCTGCCAT GTCTGGGATT CTGGATGAG 900
TGCTCTTCGT GATGAGGATG GCTCTCCAGT TGTACACTCT CAGATGGACC CTGGAGCACT 960
AGAATGATAG ACAAGATCTG CCTCTCCTTC AGTCCCTGAA TCTTATTTAA GGCTCTTGCT 1020
TTGCTTCCTGG GCTGGAGGCC GCTCTTGATGC GAGTCTGCCAT GTCTGGGATT CTGGATGAG 1080
TGGCTTCCTG TGGAGGAGGG GCTCTCCAGT TGTACACTCT CAGATGGACC CTGGAGCACT 1140
GGAGAAAGGA GTGAATGGATG GAAGAGCAGG TGGTGGGGGG TGGTTTGAGG GCCTTGCTGG 1200
TGGGTGGGTA GAGGCCCCTC CCTGGCATGG GGGCTCAAGAC CTGTTCCATC CCACAGCCTG 1260
GGGCCCTGTGT GTAATGGCCAG GGGCTGGCATT TTTCTGCTCC TTGCCTGGCT 1320
CTGGCTCCCC TTTCCTACCC CATGGGGGCG CTCAGGCTGC CATCTACTGC AAAATGCCCC 1380
AAGCGCTTTG AGAGCCCACT TGCCAAACTT CTTCTGCTCC AGAAGAATCT AGAAGACCACAT 1440
AATCTCTTCG CCCCCCTCC CGTCTCAGGG AGGAGCCGCC CTAAAGTGGG CGTCAAGAGA 1500
GCCCTCTCCA GACAAGGCGG CCCCTTCCCG CAACTCAAC GCCTCCCTCC TCAATCCCCA 1560
GGCCTCTGGG GACCAACACT GCCACACACC CCCAGGAGG GAGAGAAGAT SGATGGAAG 1620
TTACCACCGG CCGCTGCGCC CACTCTCCAT CGTCTTGGAG TAAATCTCTTT GAGATCCAC 1680
ACCGCGAGAG AGGACGTGAG GTCTCTGAGG GCTCTGGGAC TCGTCTGAGT ACGCTGATG 1740
GCTCGGATATT GATTGAGGGA TGAATGGATG AGGAGAGACA GGAGAGAGAG CCGATGGGGA 1800
GTCAAGGCGG AGAAGCCCTC TTGGAGGAGG AGGAGAGACA GGAGAGAGAG CCGATGGGGA 1860
GGCAAGCTCCT GGATGGGGAAT CCAATGAGAT GCATGCTGAT 1920
CAGCTGCTGG GGGAGAGAGA AAGATGAGG CCCCAGGAGG GAGAGAAGAT SGATGGAAG 1980
GAGCGCTTTG AGAGCACTGC ACGTCACGGC AGGCTCTCCCG GAAATCTACG CACTATGCAA 2040
GCTACCCAGG CTGCTTTGCA AGGAGAGAGA AAGATGAGG CCCCAGGAGG GAGAGAAGAT SGATGGAAG 2100
CGGCGCCGCG GCTGCTCAGGG TGGATGGGGAAT CCAATGAGAT GCATGCTGAT 2160
TGAGCTCAGG AGAAGCCCTC TTGGAGGAGG AGGAGAGACA GGAGAGAGAG CCGATGGGGA 2220
GGGACCGGCC GGAATTCCAA CTCACGGCTT GCAGGCCTGG CAGGCTCTGG GAGAGAGAAGAT SGATGGAAG 2280
GAGCGCTTTG AGAGCACTGC ACGTCACGGC AGGCTCTCCCG GAAATCTACG CACTATGCAA 2340
GAAGAGAGAGA AAGATGAGG CCCCAGGAGG GAGAGAAGAT SGATGGAAG 2400
TAAGAGCAGT GGGGAGAGAGA AAGATGAGG CCCCAGGAGG GAGAGAAGAT SGATGGAAG 2460
TAAAAGAATA GAGGAGACAG CTGGCAAGAG CTCAGCAGTT CAGGCTCTGC CAGAGAGACAG 2520
TAAATTCGCG GTTGATTTCC AGCTTCCCGG CTCTCCCGG ATTTCCCGG GCCTCCGGCG 2580
CGGCTGACGAT TCCTCCAGCC GCATCACCGC GACCTCCTAG GACCCCGTGG 2640
CACCGCTCTG AGGAGGGGGA AGCCCGGGCG CGCAGGGGCG CCGAGGGGGG 2700
CTCCCGGCTG CAGCCCGGGC CAGCCCGGGC CAGCCCGGGG 2760
GCAGGCGGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 2820
AGGAGGGGAG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 2880
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 2940
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3000
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3060
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3120
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3180
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3240
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3300
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3360
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3420
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3480
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3540
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3600
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3660
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3720
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3780
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3840
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3900
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 3960
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4020
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4080
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4140
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4200
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4260
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4320
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4380
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4440
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4500
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4560
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4620
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4680
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4740
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4800
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4860
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4920
GGGGGGAGGG GAGGGGAGGG GAGGGGAGGG GAGGGGAGGG 4980
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGATCAAATA TTTGTTTCC GAGATAACAC ACCCGGATAG CGCTGGTTTCC TGAAGCCGCTT</td>
<td>5040</td>
</tr>
<tr>
<td>TCATTCCTACT TGTGGTTCTC GCTGCCAAA ACCUGACCRA GTCAAGACAG CAAATCTGACG</td>
<td>5100</td>
</tr>
<tr>
<td>CCCCCGGGCCC TGGTCAACA TGGAAATAT GAGACTGAGC CCCCGCTGCG GCAGCTGGG</td>
<td>5160</td>
</tr>
<tr>
<td>GGGGCGATG GGGGGCGGCT GGAAGCGGAG ATCGGCGCTC ATGCTCCCTGC TCTCCCTGGAT</td>
<td>5220</td>
</tr>
<tr>
<td>AGGCGCCGCT GAGTCTGCCG GAGCCATGTT CGTACGTCAAA TTAGTACTA ATGCCAAGAT</td>
<td>5280</td>
</tr>
<tr>
<td>TGACACCAAGC TCCACGACG CTAGCGAGG AGAGTTGAGA TGGAAAATT AATCCTGAAAA</td>
<td>5340</td>
</tr>
<tr>
<td>AGATTAAATT TATACTAGAC AGCGAGTCAA GAGAAGAAC GAAAGATTTA ATGCAGATTATA</td>
<td>5400</td>
</tr>
<tr>
<td>GAACTACGC TCAAGATGAT GAGTTGGGAA GAACTGAGG GTGAGTGGAT GGGCTGGATG</td>
<td>5460</td>
</tr>
<tr>
<td>TTTCGTCTGG TCTCCACCTC TGATTTGCGG ATTTTTTTTTA AACAATGCTG ATTTTATAC</td>
<td>5520</td>
</tr>
<tr>
<td>TATACAAAATA AATGAAAGAA AAATCCAAAT ACATACATAT AAAATAAGGA GCCGAGAAGCT</td>
<td>5580</td>
</tr>
<tr>
<td>CCCCCGGCCT TGGTATACGC CCGGGTACGG AGCCGCTTTT CGTCTGGGAGAT GAGATTTAG</td>
<td>5640</td>
</tr>
<tr>
<td>CATCCTCCCT CTTGTTGAGT ACCCGGACT GCCCACTGAA GTGACGCTGC CAGTGTCGTC</td>
<td>5700</td>
</tr>
<tr>
<td>CTTAGACGTC TGGCCAGCTC GTATTTTGGG ATTTTTTTTTA CATAAGAAGCT CTTTTGCTGa</td>
<td>5760</td>
</tr>
<tr>
<td>CCGGGCCGC CTCCACCAAG CAGAAGACGC AGGGTCCGTT GGGGCTTTGG AMTGGAGTTTC</td>
<td>5820</td>
</tr>
<tr>
<td>TCATACCAAA CTGTCGCCCT CCTGAAGCCT GGGGTTAGCC AGGGCCCTTT CTCTGGTGGG</td>
<td>5880</td>
</tr>
<tr>
<td>CATCTTCCCT TCTGTTGGGT ACCCGGACT CAGAACTTAA GTGAGACGTC TGGAGTTTGGG</td>
<td>5940</td>
</tr>
<tr>
<td>CAGAAGCTTG CTGGGTGTTCC CGGAGACAGC GTGAGTCTCC GTCTCTGCTG TCTCTAAACT</td>
<td>6000</td>
</tr>
<tr>
<td>CCGGCTCTA TTATACGCTG AGAAGCCTTG TGGAGCTGTC TGAGAGCGTT TTCAAGAGCC</td>
<td>6060</td>
</tr>
<tr>
<td>CCCTGCAAAAC CAAGAAGCAAA AAAGCCCCCA GTGCTTATGG GCGGGCAGTG TGGCTAGGCG</td>
<td>6120</td>
</tr>
<tr>
<td>TGGCTGGCGC CTTCCACCCA GAAAGAACGC AGTGGTCTGG GAGGCTGCCCT TGCTTACTGG</td>
<td>6180</td>
</tr>
<tr>
<td>ATCTACCAAG CAGAAGACGC GCGCCTGGTC AGCTCTTAA CTGACTGGGC TGGCCTTTGGG</td>
<td>6240</td>
</tr>
<tr>
<td>CTTTCGCTGC TGGCCGCCTC TGATTTTGAG AGTTTTTCTG TCTCTCTGCT ATCTTAAACT</td>
<td>6300</td>
</tr>
<tr>
<td>CAGCTAACTC TTTTCAACAG CGCACTGACC TTTTTTTCCT TTCTCTTCTG TCTCTAAACT</td>
<td>6360</td>
</tr>
<tr>
<td>CGGCGGATT CAGTTGAGTA AGACGCTTGG TCAATTTTGT GGCTAGAATT CTCTCTTGGG</td>
<td>6420</td>
</tr>
<tr>
<td>CTTCCTGCCG TTTCCCTTAA TTTCCTCGAA GCAAGACGAC ACAATTAATAT GCCTCTGCCC</td>
<td>6480</td>
</tr>
<tr>
<td>TCTCATTGCT TTTTTAGCTG CTTTTTGTCC ATAGGAGGAA CTGGGGGCTG GCAGAGATTT</td>
<td>6540</td>
</tr>
<tr>
<td>TGGCAGGTTC TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>6600</td>
</tr>
<tr>
<td>TTGTTATAGC CAGTTGAGTA AGGAGGTTCC AACGAGCTGC ACCGAGCTGCT TTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td></td>
</tr>
<tr>
<td>TTGTTATAGC CAGTTGAGTA AGGAGGTTCC AACGAGCTGC ACCGAGCTGCT TTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>6660</td>
</tr>
<tr>
<td>GGGAAAAGA TACAGAGGGGT GTTTGAGCAG ACTGAGCTGC GCAGAAAGCTGC TTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TACAGAGGGGT GTTTGAGCAG ACTGAGCTGC GCAGAAAGCTGC TTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGTTTCAGT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGTTTCAGT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGTTTCAGT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td>6720</td>
</tr>
<tr>
<td>TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td></td>
</tr>
<tr>
<td>TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td>6780</td>
</tr>
<tr>
<td>TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT</td>
<td>6840</td>
</tr>
<tr>
<td>GAGCGCTCCG GTTCTGGGAAG TTTTTCATCC GGGGATGATGG AAAGAGTCAAG AGAAGTGTGA</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td></td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGCGCTCCG GTTCTGGGAAG TTTTTCATCC GGGGATGATGG AAAGAGTCAAG AGAAGTGTGA</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td></td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>6900</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7020</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7080</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7140</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7200</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7260</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7320</td>
</tr>
<tr>
<td>GAGGGTCCGGG TGGGCTTGCC GAGCTGAGGC CGTACACGCT CAGACGCTGC GCTAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>TGGGCAAGG GTGATGTCAG AGTGTGATGG GGAGGGTCCG GTAAGAGCTGC TGGGCTTGCC</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGAGG TGGGGGAGG TGGAGAGAGC TTTTCTGCCA TACGGATGTC TGAGAAAGCT</td>
<td>7380</td>
</tr>
</tbody>
</table>
GAGTTAGGCA TCTGGCTTGC AGCAAGACSG GCCCTTCCCTG TCTGGCTGCT CTCCCTGACG 7440
GGAGGATTGT TCACTTTCAT TATGAGGAA ATGGGACAGG AGAACCCTAT TCCTTTCCTG 7500
ATCACAATAC TCAGGCTGGA GTGGTCAAGC TCTGCTGCTA CTGGCCACCT TCCCAAACGC 7560
TCACTGAGC TCAGGAGAGA TCTTAAGACT GAAGGGAGGA GTCAAATCC TAGCTGGCTG 7620
AGTGCGCCAAA TCTGCGTCTC GAGACCTTTC TGGCTTCCGG GTCAAAGGGA TCCCTGACG 7680
TCACTCTCC AAATGAGCTGG GACCATAGGC ATGACCTGCT ATGCCCTGCT AATTTTTGTA 7740
TTTTTCTAAT TTTTTGTAAT GACAGAGTTT CAACGATTTG GCCAGCCAG TCTGGAACCT 7800
CTGACTCTCA GCAACTTCCC CACCTTGGCC TACCAGATGC CGGGGATAC AGGTGTCGAG 7860
CATCACTGCTA GTGCGGCAAC GTGCGGGGAA ACTGACAGAT GAAAGACGG AGCACTGAGC 7920
GTCACTCTCA TAAAGAGACT CTCTACTTCT TTTGAGGTCT TGGTTTCTCA TCAMTCTGG 7980
GCAGAATTAC AAATGAGGAG CCAAAACGTT GAAATATGAA ATTCTATGCA 8040
TAAAGAAGAC AGATATTTC TCAAATGCTA GCTTATGGGA GTCAATGAGG ACACCTGACA 8100
TTAACCTTCT TTTGTAGGTA AATGGAAGAC AAATACCAAT CAACCTTACG ACCAAATGCT 8160
GACCATAGGC ATGTTAAGCT GGTGCTTTGA ATGGGAGTTT GCACAAAGAT ACAGACCCAA 8220
CAACCTAAGA CACTTCAATG CTGGAATGAA CGTCATTCTC ACAGACCCAA 8280
ACTCAAGCAAG TGTTCTCAAG CCCCTTTGGA AAATACCATG GCACAAAGAT ACAGACCCAA 8340
TGAAGCTTCA TTATCTTATG AAATGAGGAG CCAAAACGTT GAAATATGAA ATTCTATGCA 8400
AAATCATTCT TAAAGAGACT CTCTACTTCT TTTGAGGTCT TGGTTTCTCA TCAMTCTGG 8460
TCACTGAGC TCAGGCTGGA GTGGTCAAGC TCTGCTGCTA CTGGCCACCT TCCCAAACGC 8520
AGTTTATGCT GCTGGTCTTG AGGCCACCT TGGGGAACAA ATGGAATGAG ACAGACCCAA 8580
AGATCAGCAA GGGTGGAATA ATTTTTAAAA TTTGGTAAAT ATCCCAATAA ACAGAAAGAG 8640
TACAAATTTTA ACCCACTTGA GAGAGGGAGG ATCTAAGAC AGAATTGTAT ATGGACATCA 8700
AGGGTATGTC ATGATCGTAA TATATGAGCT TAAACCATGA ATGGCAGTAA GGGCCCTTGTG 8760
GACCAACCCCA ATGCTGGGAT GGCGTCCAGG TGGCTATAGG TTTGAGATGG GGGTCCTCCT 8820
ACAACCAACAG ATGGGAAAT TTAATGGCAA GCTGACATAT ATGGAAGGAT TATGACCTTT 8880
TAAGGGCAT ATGCTGGTAC AGGGATCCAC CTGCGGGAAT TACTGCAAAT TGGCGGAGGT 8940
GAGTGATGGC CCAATCTGAA GGCGTGGGAT TACGATTCAT CAGGTCGTTG TATATAAAGT 9000
AGGCTGCTTC TGCTTTTTAA TCCTTTGGCA GCCACTTCTT TCCCTTCCCA TCTCTGACG 9060
AGGGTATGTC ATGATCGTAA TATATGAGCT TAAACCATGA ATGGCAGTAA GGGCCCTTGTG 9120
GAACCTCCTGA GCGAGGAGGA GCCCTGACCC TGGGTTTGGG TCACTCTCTT CTTCCTCCTG 9180
CTCAGATGTC GTTGATCTAA GCAACTACTG AGATGTGGGA AGATGTGGGA TCACTCTCTT 9240
CTCTACTTCT CTCCTACTCT CTGCACTTATG GACAGAGAAG AGACAGCTGA AAGAGGAAT 9300
ATCGGGCAGC TCGAGGACAT TCCCTCTGTC CTCCTTCTCC TCTCCTGAGC AGTCTCTGAGT 9360
GACTGCAATC TAAGTCGATC ATAGAACCCA CTTTCCTCCT GACTGTTATG GACAGGAAGC 9420
AGGCTTCTGA CCAACATTGG TCGAGGGGAA ATGGGAGGTG GGAAGAAGAC ATAGGGAATG 9480
TAGGAAGATG GCCCTTGGAG TGGAGTGCTG GACAGACAGC AGAAGGAATG AAGAGGAAT 9540
GTACCAGAGA GCTTATTATTG TTATAGTAA GTCAAGTCAA AGAAGGAATG AAGAGGAATG 9600
GAAGGCTGGG GCAGAGAGGA GCCCTGACCC TGGGTTTGGG TCACTCTCTT CTTCCTCCTG 9660
GACCTGCTAC GCAGAGAGGA GCCCTGACCC TGGGTTTGGG TCACTCTCTT CTTCCTCCTG 9720
AGCTGAGGCC AAATCTTCTT TATGCTCCTA AGAGGCACCC TTGGTGCGAA CCTCTCCTCT 9780
TTCTTGGGTC AATGGGGGAT GCTCGCAAGG GACCATCAGT AGGAAGGCAT GGTACACTAA 9840

CCCAGTCTGG GGTGGGCTTT TAGACTAGTC TTCCTCCCAT GCTCCTCCTC CCATTGGAAC 9900

CCCGGACTTT CAAGACTGCT ACCTAGCACA CCAGTGCACC AGATGTCACT CAAAACCTCT 9960

TCAGCAATGG CCACTCACC TTCAAAAAGG CTGAAGAGCA GACTGGCTGG GTTTCTTCATG 10020

GTGGAGGGGC AGTCTGGGAG GTTTTAAGGT TGAAGATGAA AACTTTCACT TTTTGGCTCAA 10080

TGGTCTGAAA AAGAGAAGGA CCAGCAAGTG AACTGAAGCC TCCTGGAAAG CATCTTGATA 10140

ACAGGGCAG ATTTTCAAGA TGAGAGCCTG TGGCCTTACG GAAATGACCT 10200

CTAAGTATCT CAGTATAATTA AAGGAGTCAA ACTCTAGACT CGAAGGAGAA GATCTACAAT 10260

TTTCAATAAC ATAGTCTACC CTCCCCTCCT TCCCCCACCT TCACCTCTTC TTTCATCACA 10320

GGCTTACAGG GCACCTCTTA GAGCCAGGCA CGGTGTTGGG ATCAGGAACA AGGCCACTGC 10380

ACAGGGGCAG AGTTTCAAGA TGAGAAGCTG TGGCACTTAC TCTGGCTTTG GAAATGACCT 10440

CTAAGTATCT CAGTATAATTA AAGGAGTCAA ACTCTAGACT CGAAGGAGAA GATCTACAAT 10500

ATGGGGACCA TGGGTCTGAG CCTGTGACCC TCCAGTCTAC AGTGTGTTGG TGAGGAAGGA 10560

GCAGTTGTCA CTGGGGTCAC TGGCAATGGG CATGCCTCCA TCTAGCTTAG GCAAGATGCT 10620

TAGACTCAGA GCCAGAGAGT GAAACCCAGA CACTAATGAG CTGTCGGTGT TGGTGTGTGT 10680

TCTCTTCCCTC TTCCAGTGGA ACATGACAAG GAATTCTTCC ACCCCAGCTA CCACCATCGA 10740

GAGTTCCGGT TTGATCTTTC CAAGATCCCA GAAGGGGAAG CTGTCACGGC AGCCGAATTC 10800

CGGATCTACA AGGACTACAT CCGGGAACGC TTCGACAATG AGACGTTCCG GATCAGCGTT 10860

TATCAGGTGC TCCAGGAGCA CTTGGGCAGG TGGGTGCTAT ACGGGTATCT GGGAGAGGTG 10920

CTGAGTTTCC TCTGGGGGCA GAGGAAGAAG GTGGTGAGGG TTTCCCTCCC CTCCCACCCC 10980

ATGAGCTCTG CTTCCCATCT GTTGGGGTAG TGGAGCTGTG ACCTGCTAAC GCGAAGCCCG 11040

TAGACTCAGA GCCAGAGAGT GAAACCCAGA CACTAATGAG CTGTCGGTGT TGGTGTGTGT 11100

CGGCACTGCC AGTCCTAATG CAGCCTGTGC TCCTGGACTT CAGGAGGGTC TCAGCAGTGC 11160

TCATGCTTGC TTCACTACAA ACAGGCTTCC CGCCCCCTCC CAACCAGTAC TCATGCTTCA 11220

CTGCTTGGAT CTTCACTTCC TCTCCCTCTC AGGGAATCGG ATCTCTTCCT GCTCGACAGC 11280

GGCCTCGGAG GAGGGCTGGC TGGTGTTTGA CATCACAGCC ACCAGCAACC ACTGGGTGGT 11340

CAATCCGCGG CACAACCTGG GCCTGCAGCT CTCGGTGAGA CAGGCTGGATG GTGAGTCCCC 11400

CGCCACTGCC AGTCCTAATG CAGCCTGTGC TCCTGGACTT CAGGAGGGTC TCAGCAGTGC 11460

CTGCTTGGAT CTTCACTTCC TCTCCCTCTC AGGGAATCGG ATCTCTTCCT GCTCGACAGC 11520

GCACGGCCG CAGAACAAGC AGACCCTCAT GGTGCTTTCG TCAAGACCCA CGIAAGTCCA 11580

CCAGAACCAG ATCCGGTCCA CGGGGACCAA ACAGCCGACG CACAACCGCT CCAAGACGCC 11640

CAGACCAACC ACAGGACCAC AGCAAGGCCCC GGGGACCGCA GGAGACTGAC TAAAATCATT 11700

CAGTGCTCAC CAAGATGCTC TGAGCTCTCT TAGTTTTAAGG CAAAAGCCGA GTCGAAGAGT 11760

CTAGGGAGG GCTGGGTTTT GACTCCGAG GCTCGAAGG TCCCAAAAGG CTGGCTTCTG 11820

CTACAGCTT CAGATCCTAC GCTCGCCCTG TGAGTCTTCC CATCTACAGC TCTACAGCGG 11880

ACATGGGGCT CAAGCAAGCT GCCGTTTTCT ATATGGGAAG AGCCTCTCAA GGCAGACCG 11940

TGGAGAGG TGGATTTGCT CGTGAGACAA TGAGACACCC AAGAAGAAGG CGCTGACACC 12000

AGTAGAAAAT TCACCTGCCG CCAGAAGCTA GGTGCGGTG ACATTAGGG TGGTTGATCC 12060

AGGAAGCGA TGATCTTGCT TGAAGCCCCG TCTCCCTCCT CCCCACCACG CATCTGCAGT 12120

GCTGCTTCC TCTAGGCGCA CAGCTGTAAT TGGCCAGGGG GCCTTCAATA TTTTTGCTC 12180
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTGGCCAGTA GGAGGAAGAS AATGAAAGTC TCTCCATGGG TCTTCTTAG GAAAGGAGGA</td>
<td>12240</td>
</tr>
<tr>
<td>ACTTTTACCAGAAGCTCTCA TGTTTCTTGG TTTGTGTGGG GCACCTTCGC CTTCCCTGAC</td>
<td>12300</td>
</tr>
<tr>
<td>CACTCTCGAGA GAGCTGACCA TGAGTAGCCTT AGCTGTTGGG ACTCATTAGT</td>
<td>12360</td>
</tr>
<tr>
<td>GACTTTACCAA AGCCCTTGGGA AGAGGAACGT ATTGAACCA AGGATGAGC AGACACACCA</td>
<td>12420</td>
</tr>
<tr>
<td>AGATTTACAG GAGGGGCACT TCCAGTGGGA GAGACACCGCA CACTGAGAG CACTCTCTGG</td>
<td>12480</td>
</tr>
<tr>
<td>GACTTGAAGCTT TGTTTGTTGG CACTGAGTGG TCAGAGACCA CGGGCGGTTG TCAGAGGAGC</td>
<td>12540</td>
</tr>
<tr>
<td>CAGCTCGGCC TCTCTCTGAG CAGAGACAGAG GAGGAGGGAG TGAGAAGAGG</td>
<td>12600</td>
</tr>
<tr>
<td>AGACCTGAT AGTACCAAAAAA GAGCACAAGGT TCTGTTGTGG GGTGAAGAGGA AGGTTGACGT</td>
<td>12660</td>
</tr>
<tr>
<td>GTGCTGAGA GAGGAGACAG AGACACAGAC AGACACCTGC CACTGATTACAG GAGGACGTAC</td>
<td>12720</td>
</tr>
<tr>
<td>GGCCTGACCC GAACTCTCTCC AAATTTAGCT ATGTTGGAAG AAACCCCTGG ATGACCTTTCA</td>
<td>12780</td>
</tr>
<tr>
<td>CTACTCAAGA AAACCTCAGG AGCTGTCTCT TGGAAGAGGT CATGAGATTT TGACTCTCTG</td>
<td>12840</td>
</tr>
<tr>
<td>CTCTCTCATCT TCTCTGCTTG GGTGTGGTTG ATGTGTCTTT GTCTGGAGCCGT CTTCTCCGACA</td>
<td>12900</td>
</tr>
<tr>
<td>TCCTCTCTGCT TCTCTGAGG GAGATGAGTGT TGGGGCGTCTT CGAGATGTTG OTTGTTGCTA</td>
<td>12960</td>
</tr>
<tr>
<td>AGCGATCTC TCTGAGCAGG GGCGCTGAG TGCGCTTGGGT TGGCTGGGTA GGGTTTTGAGG</td>
<td>13020</td>
</tr>
<tr>
<td>CTCTCTCTAGA GAACTGACACAG TGGTGTCTT GCAGATGTTG ATGACCTCTCTG</td>
<td>13080</td>
</tr>
<tr>
<td>AGTGCCCGCA GAGGTCAAGGA GAAAGTATTAG CGACATGAGA GAGGAAGAGGA</td>
<td>13140</td>
</tr>
<tr>
<td>GCATTGACCT CTTCCCTGAGG AGGCGGCTTCG CTGGGCTGATG AATAAGTGGT TAAGCGAGGT</td>
<td>13200</td>
</tr>
<tr>
<td>TAGACCCCTCG AACTGGCTTT TGAAGCGGTA AATTTACACC CGACGGGCTC TGCTGTCTCAT</td>
<td>13260</td>
</tr>
<tr>
<td>TGCTGTCTAC ATGATCTGGGC AAGGAAAGGG ATGTGTGTGT TGTCTGTCTT AGTCTCTCTC</td>
<td>13320</td>
</tr>
<tr>
<td>CACTGGGCTG CTGCTCTTCA CTGCTTGGCTC AATGAGACAAG AAGCAAGCAG GCACTGGAGG</td>
<td>13380</td>
</tr>
<tr>
<td>CGACCTCTGCT GAAAGACAGG AGCTGTATGT CAGCTCTGAGG GCAGTGAGT GCGAGTTTAGG</td>
<td>13440</td>
</tr>
<tr>
<td>GGCCTCTCCGT GGTCTGTCTT GGTGTGGGCG CCGTCTGCGG GTCTGCTCGC GAAAGCAGAGG</td>
<td>13500</td>
</tr>
<tr>
<td>TCCTCTCTGG TCACCTCTTG CACAGTCTGT GCAAGTTAAGA GACTGATCTGC AGTGGCCTCC</td>
<td>13560</td>
</tr>
<tr>
<td>GTGTGTGTGA TGTCTGTTT GTGGAAAGCC GCACACGAGGCA GTCAGCTGGAC GACATGGAGG</td>
<td>13620</td>
</tr>
<tr>
<td>CTGGCGGCTG CAAATCCTCA AACTGGAGCT CAGGAAAGCG ACTAGAATCC AGAAGCAGAGG</td>
<td>13680</td>
</tr>
<tr>
<td>CTCTCGCATGC ACCTGATCAG CACCCCTTGG CGTCTGCTTT CAGAGACAGA TCGGACGGACG</td>
<td>13740</td>
</tr>
<tr>
<td>TGCTGTGTCA TGGGAGACGC TCTCCTCTGC TGTCTGCTTG TGTCTGCTTG TGTCTGCTTG</td>
<td>13800</td>
</tr>
<tr>
<td>CCACTCCGA ATCCGCAAGG AGGAGCAGAT CGGAGCAGGAT CTTGGTGTG CAGCTGGGAGG</td>
<td>13860</td>
</tr>
<tr>
<td>GAAAGTTGCA GACTGTCTGAGGGTGCTTCG ATGTTACTG CATGAGCTGG CCCCCCTCTG</td>
<td>13920</td>
</tr>
<tr>
<td>AGCATTCTGC ATGAGAAGAG GCAAGAGCGCA GGGAGATGAC CTAGGAGACAGG ATGAGGAGATC</td>
<td>13980</td>
</tr>
<tr>
<td>AGTCTGATCCA CACCTGGTGAG GCTGCTGAGCT GTGAGTCTCA TCACGGAGG ACCGACATCT</td>
<td>14040</td>
</tr>
<tr>
<td>CGACCTGTTCT CGGCCCTGCT CTGCTGCTGC GATGAGGCTG GCACGAGCATA ACACCTGAC</td>
<td>14100</td>
</tr>
<tr>
<td>CTCTGCTGCTG TGAGGCTCTCC AACTACCC TCGAGAGGATG ACTGAGACG ATATTTGTTCTC</td>
<td>14160</td>
</tr>
<tr>
<td>CAACTGGGGTT ATATTCACTC TCGAGGAGCT TGTATGTCAG TGGGAAGAGG CAGAATCTCG</td>
<td>14220</td>
</tr>
<tr>
<td>CTTGCGGATCT ATGAGAGGAC GACCGACAGG GGGAGATGG ACTAGAAGATC GAGTTATGCTG</td>
<td>14280</td>
</tr>
<tr>
<td>ACTCTGCTTG GTCCTGTGGG AAGATTCTAC GCTCAGAGGC ATCAGGATG GTGACGGTCTC</td>
<td>14340</td>
</tr>
<tr>
<td>ACCTACAGGC AGCTGCTGGG AGCAGAGCT ACGAGACAGA CTGAGAGACCA CTGAGGCTGCTC</td>
<td>14400</td>
</tr>
<tr>
<td>TTAATGGGCA ACAATTCTGT GCAACCGAGC CCCACAGAAGG AAGGAGGACATG ACTAGGATGCT</td>
<td>14460</td>
</tr>
<tr>
<td>TTAACGACCA GTCACCGCC AGGCCAGCTG GCGAAGATCG GACACTGACTA TTAGACACAC TTAGACCTTC</td>
<td>14520</td>
</tr>
<tr>
<td>TGCCAGTGCG GCAACCCGAG GCATGGGACT TGGTTTGAAGG TGACCTGCGA GTGCTGAGG</td>
<td>14580</td>
</tr>
</tbody>
</table>
ACCTAGTGGG GTAAGAGCTT GACACGGGAG GCTCCCTCCT GTGCCGCGTT 14640
CTGCCAAGAA GACACGCA GATATGGATAA AAGCCAAGCC GTGGGCGTTTT 14700
CCTCTTTTCTT TTTTTTTTTA AAAGGTGCCCT GTGCTCAGC CCAAGTGGG 14760
GTGAAAGGTTT GCTCATACAG GTGAGGCGATATTGTTGCTT CTGTTTCCACT GAGTCTACTA 14820
AGACCGGAC CAACTCTGCTT GGCCTCGTCT GTACGAGGAC AGGTCTGAG 14880
GCTCCTCAGC TGGTCTGAG CCCTGCTGCTT CCAAGTGGG 15000
TGCTCAGGCT TCTGCTAGGG TTCCAGACCT CTCTTTCCT GCTCAGGACT GAGTACGTG 15060
GCCGTAAGCC TACCTGTGCA GGGGAGTGT GCCCTCCCCT TGAACCTCTA 15120
CATAGACCAGG CCACTCTGCA GACGCTCTGAG GGGGTACAGC TAGCTTTGAG 15180
TGGGTCTGCC GGGCGGATAG CCAGGAGAGT TGGGAGCATG 15240
GCCCTAGTGA ATCTCAAGGG ATTCCAACAT GTAGCTTTTT GGGGCCCCCT TGGGCCTTGA 15300
AAAAAATGCA AAAATCGGAT TTTTACTGATT GGGTGGTTT ATATAGGAGT AATAGGCTCG 15360
ACCCGTGATG TATGTTATGT TTCTCAATTT TTAAAGTAAAC TAAAAGTTGT ATGGGCTCCT 15420
TTTACCAAGGC TCTGATGGAG TTCAAGCTTATT GATCCTTTTG GCTGTGCTGG 15480
CTTCAAGATTC CAGTCAGAGC AAGGAGCGA AATTCAGTGA GGGCGCATG 15540
GATCAGGCG GCCCGCGCAT GATAGTGGGA TAGAGCGGAC CATCGGCGAT 15600
TTCAAGGGAA ATTCCAGCTG GGCCTCACAG AAGATCCCAT GTGGATAAGT GAGTACCTTG 15660
CCCTTGATCG GGGGACCTCA TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 15720
ATACAGGCTG TTGGTGATTT GTGCTTTTTC GTCAGAACCC AACCTGAGGT 15780
ATGGACGCTG ATCAGAGGCG CACAGAGCC TCTCCAGTCT TCTTCAGTCT 15840
CCTGAGGCTC TCTCAGGAGC CATCATCCAG GCTCCAGGG TCTCCAGTC 15900
TACTCTCTCT CCTCCTCTCT CTTGAGCTGT AAGTCTCGGA GAAAAATCTC 15960
ACGCCCGGTT CACTTCTGTC ACCCTGTTGTTT GCTGATTTCT GCTGCTTCTT 16020
TTCCTTTTAA GTGCTGTTTCTT GAGCCAGGGA GGGTGTGGTT TCTGGTTTGT 16080
CTCAGGCGA AAGGAGCAGC TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16140
CCCTTGATCG GGGGACCTCA TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16200
CTCCAGGGGA AACAGCTTCT ATCCCTGCTC AACCCTGGGC AAGAACAGG 16260
ACTGCTCGCT GCTGACGAGC AAGCAGACGG AACGCTTCTA ATCCCTGCTC AACCCTGGGC AAGAACAGG 16320
AAACAAGTCT CTCAGGCGAA ATGGAAAGGA AACGAGTAAT GACATGTTTT AAGAAGGCGA 16380
ACGCCTCGCT AGAAGCTGCT CAGGAAGGTAC TCCAGGGGAA ATGAGGCTCT GCTGCTTCTT 16440
TCTTCTTTTAA ATCATGCAAG GTGGAGGCGA TGGGGAAAGG TTTGTCTGGG GCAGAGTTTT 16500
AAATGGCCTA CTCTAAAAAG TCTAAAAAGC ATGCGGCTCA GGCAGAGGAC TACCAAGCTG 16560
GAGGGCTGAA CAGAGTGCA AATCAGCGAT TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16620
CCCTGATCG GGGGACCTCA TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16680
GCTCGAGATCC TCTTACTTCTG AAAGACCGCTT GCTGCTTCTT GCTGCTTCTT 16740
CTCCAGGTTT GCTGCTTCTT CCCCCTTATT TCACTGTGGT CTATTAGGT 16800
TATGACTGGA CTCAGGCGAA CAGAGTCCA AATCAGCGAT TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16860
CTCGAGGTTT GCTGCTTCTT CCCCCTTATT TCACTGTGGT CTATTAGGT 16920
GAGGGAGCTT TACAGCGGAG CAGAGTCCA AATCAGCGAT TCTAGAGTGA TTGAGTGCTG CTAAGAGGAC AGGTCTGAG 16980
TCCTGAATGT CCGAATCTGC CCCCCTCTTC CCCCACCTAT AATGTGAAA CCTCTATGCT

TTGTTCCCCT TGTTCTGCAAA ACGGAGAATAA TCCCCAGAAT CAGGTGTCCA GTTTCTATT ATTTCCCCAC

AGGATGTTT TTTTATAGAAA CCGGAGGCAAG GTTGAAGGCG ACTCAGCCTG TTCACTCAGT

CGCCATGTTT TGTTGTTTTG TTTTATAGAAA CCGGAGGCAAG GGTTGCTGCC TTTTATAGAAA CCGGAGGCAAG

TCGAGAATTC

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: HOMO SAPIENS

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 9..1196
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "CBMP2A"
/note= "CBMP2A (CDNA)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGTCGACC ATG GTG GCC GGG ACC CGC TGT CTT CTA GCG TTG CTG CTT CCC

50

CAG GTC CTC TCG GCC GCG GGG ACT GTG CCG TCT GAG TGC TGC GGC GCG

98

Gln Val Leu Leu Gly Ala Ala Leu Leu Leu Leu Val Pro 15

100

Arg Lys Phe Ala Ala Ser Gly Arg Ser Pro Ser Ser Ser Ser 45

146

GAG GAG GTC TCG AGC GAG TCG GAG TCG CAG AGC ATG TAC GCC

194

Asp Glu Val Leu Ser Glu Phe Leu Arg Leu Leu Leu Met Phe Gly

50

CTG AAA CAG AGA CCC ACC CCC ACC AGG GAC GCC GGT GTG CCC CCC TAC

242

Leu Lys Gln Arg Pro Thr Pro Ser Arg Asp Ala Val Val Pro Pro Tyr

65

AVG CTA GAC TCG TAT CCG AGG AGC TCG GCT CAG CCG GCC TCG CCC GGC

290

Met Leu Asp Leu Tyr Arg Arg His Ser Gly Glu Phe Gly Ser Ala

80

CGA GAC CAC CGG TCG GAG AGG GAA GCC AGG CCG AAC ACT GTG GCC

338

Pro Asp His Arg Leu Glu Arg Ala Ser Arg Ala Asn Thr Val Arg

95

AGC TTC CAC CAT GAA GAA TCT TGG TGG TGA GAA CTA CCA GAA AGC AGT GGG

386

Ser Phe His His Glu Glu Ser Leu Glu Leu Leu Pro Thr Ser Gly

115

AAA ACA ACC CGG AGA TCC TCC TTT AAT TTA AGT TCT ATC CCC ACC AGG

434

Lys Thr Thr Arg Arg Phe Phe Phe Asn Leu Ser Ser Ile Pro Thr Glu

130
(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Val Ala Gly Thr Arg Cys Leu Leu Ala Leu Leu Leu Pro Gln Val
1 5 10 15
Leu Leu Gly Gly Ala Ala Gly Leu Val Pro Glu Leu Gly Arg Arg Lys
20 25 30
Phe Ala Ala Ala Ser Ser Gly Arg Pro Ser Ser Gin Pro Ser Asp Glu
35 40 45
Val Leu Ser Glu Phe Glu Leu Arg Leu Leu Ser Met Phe Gly Leu Lys
50 55 60
Gln Arg Pro Thr Pro Ser Arg Asp Ala Val Val Pro Tyr Met Leu
65 70 75 80
Asp Leu Tyr Arg Arg His Ser Gly Gin Pro Gly Ser Pro Ala Pro Asp
85 90 95
His Arg Leu Glu Arg Ala Ala Ser Arg Ala Asn Thr Val Arg Ser Phe
100 105
His His Glu Ser Leu Glu Pro Glu Thr Ser Gly Lys Thr
115 120 125
Thr Arg Arg Phe Phe Asn Leu Ser Ser Ile Pro Thr Glu Glu Phe
130 135 140
Ile Thr Ser Ala Glu Leu Gin Val Phe Arg Glu Gin Met Gin Asp Ala
145 150 155 160
Leu Gly Asn Aan Ser Ser Phe His His Ile Aan Ile Tyr Glu Ile
165 170 175
Ile Lys Pro Ala Thr Ala Aan Ser Lys Phe Pro Val Thr Ser Leu Leu
180 185 190
Asp Thr Arg Leu Val Aan Gin Ala Ser Ser Arg Trp Glu Ser Phe Asp
195 200 205
Val Thr Pro Ala Val Met Arg Trp Thr Ala Gin Gly His Ala Aen His
210 215 220
Gly Phe Val Val Glu Val Ala His Leu Glu Lys Gin Gly Val Ser
225 230 235 240
Lys Arg His Val Arg Ile Ser Arg Ser Leu His Gin Asp Gin His Ser
245 250 255
Trp Ser Gin Ile Arg Pro Leu Leu Val Thr Phe Gly His Asp Gly Lys
260 265 270
Gly His Pro Leu His Lys Arg Gin Glu Gin Ala Lys His Lys Gin
275 280 285
Arg Lys Arg Leu Lys Ser Ser Cys Lys Arg His Pro Leu Tyr Val Asp
290 295 300
Phe Ser Asp Val Gly Trp Asn Asp Trp Ile Val Ala Pro Pro Gly Tyr
305 310 315 320
His Ala Phe Tyr Cys His Gly Glu Cys Pro Phe Pro Leu Ala Asp His
325 330 335
Leu Aan Ser Thr Aan His Ala Ile Val Gin Thr Leu Val Aan Ser Val
340 345 350
Asn Ser Lys Ile Pro Lys Ala Cys Cys Val Pro Thr Glu Leu Ser Ala
355 360 365
Ile Ser Met Leu Tyr Leu Aep Glu Aep Glu Lys Val Val Leu Lys Aan
370 375 380
Tyr Gln Aep Met Val Val Glu Gly Cys Gly Cys Arg
385 390 395
(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1788 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: HOMO SAPIENS
(F) TISSUE TYPE: HIPPOCAMPUS

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 403 .. 1626
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/evidence= EXPERIMENTAL
/note* "CBMP2B (CDNA)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
GAATTCGGGGGAGAGGAGGA
GGGAGGGAGGGAGGAGCGC
GGAGCCCGGCGCGGAAGCTA
60
GGTGAGTGTGGCATCCGAGC
TGAGGGACGCGAGCCTGAGA
120
CGCCGCTGCTGCTCCGGCTG
180
AGTATCTAGCTTTGCTCCCC
GATGGGATTCCCGTCCAAGC
240
CTGCAGCGCCACAGTCCCCG
GCCCTCGCCCAGGTTCACTG
300
CAACCGTTCAAGGTCCCCAG
GGAGCTGCTGCTGGCGAGCC
360
CGCTACTGCAAGGACCTATG
GAGCCATTCCGTAGTGCCAT
420
CCCGAGCAACGCACTGCTGC
AGCTTCCCTGAGCCTTTCCA
480
GCAAGTTTGTTCATGGACTG
TTATTATATGCCTTGTTTTC
540
TCATGATTGCCTGAGAGCC
AGCAGAGGAGG
60
```
GAA GAA GAA Pro Ala Asp Leu Asn Leu Ser Ser Ile Pro 135 140 145
GAG AAG GAG GAG ATC TCC TCT GCA GAG CTT CCG CTC TCC CGG GAG CAG 150 155 160
Glu Aen Glu Val Ile Ser Ser Ala Glu Leu Arg Phe Arg Glu Gln Val Aep Gln Gly Pro Trp Gln Asp Gly Arg His Arg Ile Aen Ile 165 170 175 180
GTG CAC GAG GCC CCT CAT TGG GAA AGG GCC TTC CAC CCT ATC AAC ATT 942
Val Aep Gln Gly Pro Aep Trp Gln Asp Gly His Arg Ile Aen Ile
TAT GAG GTG ARG AAG CCC GCA GAA GTG GTG CCT GGG CAC CTC AEC 990
Tyr Glu Val Met Lys Pro Pro Ala Glu Val Met Gly His Ile 185 190 195
ACA CGA CTA CGC GAC AAC AGA CTG CAG CAC AAT GTG GCA CGA CGG TGG 1038
Thr Arg Leu Leu Aep Thr Arg Leu Val Leu Arg Trp Thr Arg Glu Lye
GAA ACT TTT GAT GTG AGC CCT GCG GTC CTT GCC TGG ACC CGG GAG 1086
Glu Thr Phe Asp Val Ser Pro Ala Val Leu Arg Trp Thr Arg Glu Lye
GAG CGC TAA AAC TAT GGA CTA GCC ACC ATT AGC AGA TGG CAG 1134
Arg Thr His Gly Glu His Val Arg Ile Ser Arg Ser Leu Pro Glu
CGG ACC CAC CAG CGC CAT GTG AGG ATT ACC CGA TGG TTA CCT CAA 1182
Arg Thr His Gly Glu His Val Arg Ile Ser Arg Ser Leu Pro Glu
GAG CAC CAC GTG CAT GTC AGG ATT ACC CGA TGG TTA CCT CAA 1134
Arg Thr His Gly Glu His Val Arg Ile Ser Arg Ser Leu Pro Glu
GCG ACC CAC CAG CGC CAT GTG AGG ATT ACC CGA TGG TTA CCT CAA 1134
Arg Thr His Gly Glu His Val Arg Ile Ser Arg Ser Leu Pro Glu
CAT GTG GGC CGG GCC CAT GTC TGG ACC CGA CGC CGG AGG GCC AAC CTT 1230
His Asp Gly Arg Gly His Ala Leu Thr Arg Arg Arg Ala Lys Arg
GCG ACC CAC TCG TCA CAG CGG AGG AAG AAT AAG AAC TGG 1326
Ser Pro Lys His Thr Ser Glu Ala Ala Lys Asn Lys Asn Aen Thc
CGG ACC CAC TCG TCG TAT GTG GAC TTC AGC CAT GTG GAG TGG AAT GAC 1374
Arg Arg His Ser Leu Tyr Val Aep Phe Ser Aep Val Gly Trp Aep Aep
TGG ATT GTC GCC CCA GCC GCC TAC CAG GCC TTC TAC TGCC CAT GGG GAC 1422
Trp Ile Val Ala Pro Pro Gly Tyr Glu Ala Phe Tyr Cys His Gly Asp
TGC CCC TTT CCA CTC GGC ACC CTC ACC AEC TCA ACC AAC CAT GCC ATT 1470
Cys Pro Phe Pro Leu Ala Asp His Leu Aen Thr Aen Aen Ile 345 350 355
GTG CAG ACC CTC GTG AAT GTC TAT GCT AAT TCC AGT ATC CCC AAG GCC TGT 1518
Val Glu Thr Val Aen Ser Val Asn Ser Ser Ile Pro Lys Ala Cys
TGT GCG CCT ACT GAA CTC GGC ATC TCC AAC TCC GAC TCC AAG GCC TTC 1566
Cys Val Pro Thr Glu Leu Ser Val Aen Ser Met Leu Tyr Leu Aep Glu
TAT GAT AAG GTC GAT CAG AAA AAT TAT CAG GAG AAG GTA GTA GGA GCA 1614
Tyr Aep Lye Val Val Leu Lye Aen Tyr Glu Met Val Val Glu Gly
TGT GGG TGC GCG TGAGCATCGG CAGTCTTGA GGATAGACAG ATATACACAC 1666
Cys Gly Cys Arg 405
ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC 1726
CAGACTGC TTCTTATAG CGGGACTTCTT TTATAAAAA AAAAAAAAA AAACCGGA 1786

US 7,176,284 B2

-continued
(2) INFORMATION FOR SEQ ID NO: 7:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 408 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Ile Pro Gly Asn Arg Met Leu Met Val Val Leu Leu Cys Gln Val  
1   5   10   15

Leu Leu Gly Gly Ala Ser His Ala Ser Leu Ile Pro Glu Thr Gly Lys  
20  25  30

Lys Lys Val Ala Glu Ile Gln Gly His Ala Gly Gly Arg Arg Ser Gly  
35  40  45

Gln Ser His Glu Leu Leu Arg Asp Phe Glu Ala Thr Leu Leu Gln Met  
50  55  60

Phe Gly Leu Arg Arg Pro Gln Pro Ser Lys Ser Ser Ala Val Ile Pro  
65  70  75  80

Asp Tyr Met Arg Asp Leu Tyr Leu Gln Ser Gly Glu Glu Glu Glu  
85  90  95

Glu Glu Ile His Ser Thr Gly Leu Glu Tyr Pro Glu Arg Pro Ala Ser  
100 105 110

Arg Ala Asn Thr Val Arg Ser Phe His His Glu Glu His Leu Glu Asn  
115 120 125

Ile Pro Gly Thr Ser Glu Asn Ser Ala Phe Arg Phe Leu Phe Asn Leu  
130 135 140

Ser Ser Ile Pro Glu Asn Val Ile Ser Ser Ala Glu Leu Arg Leu  
145 150 155 160

Phe Arg Glu Gln Val Asp Gln Gly Pro Asp Trp Glu Arg Gly Phe His  
165 170 175

Arg Ile Asn Ile Tyr Glu Val Met Lys Pro Pro Ala Glu Val Val Pro  
180 185 190

Gly His Leu Ile Thr Arg Leu Leu Aep Thr Arg Val Leu His His Asn  
195 200 205

Val Thr Arg Trp Glu Thr Phe Aep Val Ser Pro Ala Val Leu Arg Trp  
210 215 220

Thr Glu Gly Lys Gin Pro Asn Tyr Gly Leu Ala Ile Glu Val Thr His  
225 230 235 240

Leu His Gln Thr Arg Thr His Gln Gly Gin His Val Arg Ile Ser Arg  
245 250 255

Ser Leu Pro Gln Gly Ser Gly Asn Trp Ala Gin Leu Arg Pro Leu Leu  
260 265 270

Val Thr Phe Gly His Aep Gly Arg Gly His Ala Leu Thr Arg Arg Arg  
275 280 285

Arg Ala Lys Arg Ser Pro Lys His His Ser Gln Arg Ala Arg Lys Lys  
290 295 300

Asn Lys Asn Cys Arg Arg His Ser Leu Tyr Val Asp Phe Ser Aep Val  
305 310 315 320

Gly Trp Asn Arg Trp Ile Val Ala Pro Pro Gly Tyr Gin Ala Phe Tyr  
325 330 335

Cys His Gly Asp Cys Pro Phe Pro Leu Ala Asp His Leu Asn Ser Thr  
340 345 350

Asn His Ala Ile Val Gin Thr Leu Val Asn Ser Val Asn Ser Ser Ile  
355 360 365
Pro Lys Ala Cys Cys Val Pro Thr Leu Ser Ala Ile Ser Met Leu
370 375 380
Tyr Leu Asp Glu Tyr Asp Lys Val Val Leu Lys Tyr Gln Glu Met
385 390 395 400
Val Val Glu Gly Cys Gly Cys Arg
405

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..507
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product* "OP1A"
/note* "OP1A FUSION"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATG AAA GCA ATT TTC GTA CTG AAA GGT TCA CTG GAC
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Asp Leu Asp
 1   5  10  15

TCT CGT CTG GAT CTG GAC GTC
Ser Arg Leu Asp Leu Asp Val Tyr His Lys Asp Leu Ser Asp
 20  25  30

CAC CTG GAT CTG GAC CAC AAA GAC CTG TCT
His Leu Val Leu Val Asp Leu Ala Arg Asn Ala Arg Ile Val
 35  40  45

ACT CCC GGC TCT CTG TAC GTC GAT CGG CAT GAA TAT
Thr Pro Ser Arg Tyr Val Ala Asp Leu Asp Pro His Gln
 50  55  60

AGG CAG GCC TGC GAC CAG TAT CTG CGA GTC GAT
Arg Glu Ala Ala Tyr Asp Tyr Asp Val Leu Asp Asp Leu
 65  70  75  80

GCC TGG CAG GAC TGG TAC ATC GTC GAA TAC TAC
Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala Ala Tyr
 85  90  95

TGT GAG GAT TCT GGC CCT AAC TGC AAC ATC CAC TAC TAC
Cys Gly Glu Glu Cys Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr
100 105 110

AAT CAC GCC ATG GAT GCT ACG CTG AAC TAC AAT TGT
Asn His Ala Ile Val Gln Thr Leu Val His Phe Ile Asp Pro Glu Thr
115 120 125

GCG CCC AAG CCC TGC TGT GGC CCC ACG CAG CTC AAC GAC TCC GTG
Val Pro Lys Pro Cys Ala Pro Thr Gln Leu Asn Ala Ile Ser Val
130 135 140

CTC TGC GAT GAC AGC TCC TAC AAC GTC ATC CTG AAA TAC AAC
Leu Tyr Phe Asp Asp Ser Ser Asn Val Ile Leu Lys Tyr Arg Asn
145 150 155 160

ATG GTC GTG GCC TGC TGC CAC TACT CAG
Met Val Val Arg Ala Cys Gly Cys His
165
(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 169 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
1 5 10 15
Ser Arg Leu Asp Leu Arg Thr Asp His Lys Asp Leu Ser Asp
20 25 30
His Leu Val Leu Val Asp Leu Ala Arg Asn Leu Ala Arg Ile Val
35 40 45
Thr Pro Gly Ser Arg Tyr Val Ala Leu Asp Phe Asp Pro His Gln
50 55 60
Arg Gln Ala Cys Lys His Gln Leu Tyr Val Ser Phe Arg Asp Leu
65 70 75 80
Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala Ala Tyr Tyr
85 90 95
Cys Glu Gly Glu Cys Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr
100 105 110
Asn His Ala Ile Val Gln Thr Leu Val His Phe Ile Asn Pro Glu Thr
115 120 125
Val Pro Lys Pro Cys Ala Pro Thr Gln Leu Asn Ala Ile Ser Val
130 135 140
Leu Tyr Phe Asp Asp Ser Ser Asn Val Ile Leu Lys Tyr Arg Asn
145 150 155 160
Met Val Val Arg Ala Cys Gly Cys His
165

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1004 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..951
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/produces= "OP1B"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
ATG AAA GCA ATT TTC CTA CTG AAA GAT TCA CTG GAC AGA GAT CTG GAC
48
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
1 5 10 15
TCT GGT CTG CAT CTG GAC GAT GCT CAC ACC GAC CAC AAA GAC CTG CTG TCT GAT
96
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
20 25 30
CAC CTG GGT CTG GAC CTG GCT COT ACC GAC CAC AAA GAC CTG GCT COT ATC GTT
144
His Leu Val Leu Val Asp Leu Ala Arg Asn Asp Leu Ala Arg Ile Val
35 40 45

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
Met Lys Ala Ile Phe Val Leu Lys Ser Leu Asp Arg Asp Leu Asp
1 5 10 15
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
20 25 30
His Leu Val Leu Val Asp Leu Ala Arg Asn Asp Leu Ala Arg Ile Val
35 40 45
Thr Pro Gly Ser Arg Tyr Val Ala Asp Leu Glu Phe Arg Ile Tyr Lys
50 55 60
Asp Tyr Ile Arg Glu Arg Phe Asp Asp Glu Thr Phe Arg Ile Ser Val
65 70 75 80
Tyr Gln Val Leu Gln Glu His Leu Gly Arg Ser Leu Phe Leu
85 90 95
Leu Asp Ser Arg Thr Leu Trp Ala Ser Glu Gly Trp Leu Val Phe
100 105 110
Asp Ile Thr Ala Thr Ser Asn His Thr Val Val Asn Pro Arg His Asn
115 120 125
Leu Gly Leu Gln Leu Ser Val Glu Thr Leu Asp Gly Gin Ser Ile Asn
130 135 140
Pro Lys Leu Ala Gly Leu Ile Gly Arg His Gln Pro Gin Lys Gin
145 150 155 160
Pro Phe Met Val Ala Phe Phe Ala Thr Glu Val His Phe Arg Ser
165 170 175
Ile Arg Ser Thr Gly Ser Lys Gin Arg Ser Gin Asn Arg Ser Lys Thr
180 185 190
Pro Lys Asn Gin Glu Ala Leu Arg Met Ala Asn Val Ala Glu Asn Ser
195 200 205
Ser Ser Asp Gln Arg Gin Ala Cys Lys His Gin Leu Tyr Val Ser
210 215 220
Phe Arg Asp Leu Gly Trp Gin Asp Trp Ile Ile Ala Pro Glu Gly Tyr
225 230 235 240
Ala Ala Tyr Tyr Cys Glu Gly Glu Cys Ala Phe Pro Leu Asn Ser Tyr
245 250 255
Met Asn Ala Thr Asn His Ala Ile Val Gin Thr Leu Val His Phe Ile
260 265 270
Asn Pro Glu Thr Val Pro Lys Pro Asn Cys Ala Pro Thr Gin Leu Asn
275 280 285
Ala Ile Ser Val Leu Tyr Phe Asp Ser Ser Asn Val Ile Leu Lys
290 295 300
Lys Tyr Arg Asn Met Val Val Arg Ala Cys Gly Cys His
305 310 315

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO
```
<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Lys</td>
<td>Ala</td>
<td>Ile</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Arg</td>
<td>Leu</td>
<td>Asp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>31</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>Leu</td>
<td>Val</td>
<td>Val</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>46</th>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>61</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>Ser</td>
<td>Phe</td>
<td>Val</td>
<td>Ala</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>81</th>
<th>85</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>Asp</td>
<td>Phe</td>
<td>Val</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>96</th>
<th>100</th>
<th>105</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
<td>Ser</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>111</th>
<th>115</th>
<th>120</th>
<th>125</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Ile</td>
<td>Leu</td>
<td>Gly</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>126</th>
<th>130</th>
<th>135</th>
<th>140</th>
</tr>
</thead>
<tbody>
<tr>
<td>His</td>
<td>Asn</td>
<td>Ser</td>
<td>Ala</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>141</th>
<th>145</th>
<th>150</th>
<th>155</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>Gln</td>
<td>Gln</td>
<td>Gly</td>
<td>Arg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>161</th>
<th>165</th>
<th>170</th>
<th>175</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>Ala</td>
<td>Val</td>
<td>Phe</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>176</th>
<th>180</th>
<th>185</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Cys</td>
<td>CAT</td>
<td>TTC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>191</th>
<th>195</th>
<th>200</th>
<th>205</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val</td>
<td>Glu</td>
<td>His</td>
<td>Asp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>206</th>
<th>210</th>
<th>215</th>
<th>220</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe</td>
<td>Arg</td>
<td>Phe</td>
<td>Asp</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>221</th>
<th>225</th>
<th>230</th>
<th>235</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Phe</td>
<td>Asp</td>
<td>Ile</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>241</th>
<th>245</th>
<th>250</th>
<th>255</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Thr</td>
<td>Phe</td>
<td>Arg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>256</th>
<th>260</th>
<th>265</th>
<th>270</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Glu</td>
<td>Ser</td>
<td>Asp</td>
</tr>
</tbody>
</table>
GAG GAG GGC TGG CTT GTC TTT GAC ACA GCC ACC AGC AAC CAC TGG
Glu Glu Gly Trp Leu Val Phe Asp Ile Thr Ala Thr Ser Asn His Trp
 275  280  285

GTC GTC AAT CCG CAG CAC ACG CTC GGG CTG GTC CAG GAG
Val Val Asn Pro Arg His Asn Leu Gly Leu Gln Leu Ser Val Glu Thr
 290  295  300

CTG GAT GGG CAG AGC ATC AAC CCC AAG TTG GCG GCC CAG ATT GGG
Leu Asp Gly Glu Ser Asn Pro Leu Ala Gly Leu Ile Gly Arg
 305  310  315  320

CAC GGG CCC CAG AAC AAG CAC CTC ATG GTC GCT GTC TAC AAG GCC
His Gly Pro Glu Asn Lys Gln Pro Phe Met Val Ala Phe Lys Ala
 325  330  335

ACG GAC GTC CAC TTC GCC AGC ATC TCC AGC ACG GGC AAC AAA CAC GCC
Thr Glu Val His Phe Arg Ser Ile Arg Ser Thr Gly Ser Lys Gln Arg
 340  345  350

ACG CAC AAG CCC TCC AAC AAG CAG GAC AAG CAC ACG AGG CAG GCC TGT AAG
Ser Glu Asn Arg Ser Lys Thr Pro Lys Asn Glu Ala Leu Arg Met
 355  360  365

GCC AAC GTC CAG TCC AGC AGC AAG ACG GAC AAG CAG CCC TGT TAT
Ala Asn Ala Glu Ala Glu Ser Asn Arg Glu Ala Cys Lys
 370  375  380

AAC CAC GAC CTG TAT GTC AGC CTC GAG CTC GGG CTG CAG GCC TGG
Ile Asn Ala Leu Tyr Val Ser Phe Arg Asp Leu Gly Trp Glu Asp Trp
 385  390  395  400

ATC ATC CGC CCT GAA GCC TAC GCC TAG TAT GAG GGG GAG TGT
Ile Ile Ala Pro Glu Tyr Ala Tyr Tyr Cys Glu Gly Glu Cys
 405  410  415

GCC TCC CCT GAC TCC ATC TAG ACG GCC ACC AAC CAC GCC ATC GTG
Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr Asn His Ala Ile Val
 420  425  430

CAG CGG CGC TCT CAC ATC AAC CGC TCT GGC AAG CAG CCC AGC TGC
Gln Thr Leu Val His Phe Asn Pro Glu Thr Val Pro Lys Pro Cys
 435  440  445

TCT CGG CCC ACG CAG CTG AAT GCC ATC TCC GTC CTC TAC TCC GAT
Cys Ala Pro Thr Glu Asn Ala Ile Ser Leu Tyr Phe Asp Asp
 450  455  460

AAC TCC AAG GTC ATC CAG AAA TAC AGA AAC ATG GTC CTG CCG GCC
Ser Ser Asn Val Ile Leu Lys Tyr Arg Asn Met Val Val Arg Ala
 465  470  475  480

TCT GCC TTC CAC TAG CTC TTC CGA AAG TGC CGA AGA CTT TCG GCC CAA
Cys Gly Cys His

TTT TTCTGGA TTC

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
 1    5    10    15
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
 20   25   30
His Leu Val Leu Val Asp Leu Ala Arg Asn Glu Asn Ser Arg Val Ala
 35   40   45

(392) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
 1    5    10    15
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
 20   25   30
His Leu Val Leu Val Asp Leu Ala Arg Asn Glu Asn Ser Arg Val Ala
 35   40   45
Arg Arg Ala Gly Ala Met His Val Arg Ser Leu Arg Ala Ala Ala Pro
50 55 60
His Ser Phe Val Ala Leu Trp Ala Pro Leu Phe Leu Leu Arg Ser Ala
65 70 75 80
Leu Ala Asp Phe Ser Leu Asp Asn Glu Val His Ser Ser Phe Ile His
85 90 95
Arg Arg Leu Arg Ser Glu Glu Arg Glu Met Gin Arg Glu Ile Leu
100 105 110
Ser Ile Leu Gly Leu Pro His Arg Pro Arg Pro His Leu Gin Gly Lys
115 120 125
His Asn Ser Ala Pro Met Phe Met Leu Asp Ser Tyr Ala His Gly
130 135 140
Gly Gly Gly Arg Arg Gly Gly Gin Gly Phe Ser Tyr Pro Tyr
145 150 155 160
Lys Ala Val Phe Ser Thr Gin Gin Pro Leu Ala Ser Leu Gin Asp
165 170 175
Ser His Phe Leu Thr Asp Ala Asp Met Val Met Ser Phe Val Asn Leu
180 185 190
Val Glu His Asp Lys Glu Phe Phe His Pro Arg Tyr His His Arg Glu
195 200 205
Phe Arg Phe Asp Leu Ser Lys Ile Pro Glu Gly Glu Ala Val Thr Ala
210 215 220
Ala Glu Phe Arg Ile Tyr Lys Asp Tyr Ile Arg Glu Arg Phe Asp Asn
225 230 235 240
Glu Thr Phe Arg Ile Ser Val Tyr Gin Val Leu Gin Glu His Leu Gly
245 250 255
Arg Glu Ser Asp Leu Phe Leu Asp Ser Arg Thr Leu Trp Ala Ser
260 265 270
Glu Gly Trp Leu Val Phe Asp Ile Thr Ala Thr Ser Asn His Trp
275 280 285
Val Val Asn Pro Arg His Asn Leu Gly Leu Gin Leu Ser Val Glu Thr
290 295 300
Leu Asp Gly Gin Ser Ile Asn Pro Lys Leu Ala Gly Leu Ile Gly Arg
305 310 315 320
His Gly Pro Gin Asn Lys Gin Pro Phe Met Val Ala Phe Phe Lys Ala
325 330 335
Thr Glu Val His Phe Arg Ser Ile Arg Ser Thr Gly Ser Lys Gin Arg
340 345 350
Ser Gin Asn Arg Ser Lys Thr Pro Lys Asn Gin Glu Ala Leu Arg Met
355 360 365
Ala Asn Val Ala Glu Asn Ser Ser Ser Asp Gin Arg Gin Ala Cys Lys
370 375 380
Lys His Glu Leu Tyr Val Ser Phe Arg Asp Leu Gly Trp Gin Asp Trp
385 390 395 400
Ile Ile Ala Pro Glu Gly Tyr Ala Ala Tyr Tyr Cys Glu Gly Glu Cys
405 410 415
Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr Asn His Ala Ile Val
420 425 430
Gln Thr Leu Val His Phe Ile Asn Pro Glu Thr Val Pro Lys Pro Cys
435 440 445
Cys Ala Pro Thr Gin Leu Asn Ala Ile Ser Val Leu Tyr Phe Asp Asp
450 455 460
(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1277 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1224
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "OP1D"
/note* "OP1D - FUSION"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
AGT AAA GCA ATT TTC GCA CTG AAA GGT TCA CTG GAC AGA GAT CCG TCG
Ser Ile Phe Val Leu Lys Tyr Arg Asp Arg Ala
1  5 10 15

AGC TTC ATC CAC CGG GCA ATG GAG CGG GAG ATG CAG
Ser Phe Ile His Arg Leu Arg Ser Glu Arg Arg Met Glu
20 25 30

CGG GAG ATC TCC ATT TAC CAC GCA GCA CTG CAC GAC
Arg Glu Ile Leu Ser Ile Leu Gly Leu Pro His Arg Arg Pro His
35 40 45

CTC GAC GCG AAG CAC AAC ATG GCA CCC ACG TAC ATG CTC ACG CAT
Leu Glu Gly His Asn Ser Ala Pro Met Phe Met Leu Asp Tyr
50 55 60

AAC GCC CAT GGC GGT GGA GGG CGG CGG CCC GCC GGC GAC TTC
Asn Ala His Gly Gly Gly Gly Gly Arg Arg Pro Gly Gly Gly Phe
65 70 75 80

TCC TAC CCC TAC ACC GGC GTG AAT ACC CAG GCC CCC CCT CTG GCC
Ser Tyr Pro Tyr Lys Ala Val Ser Thr Glu Pro Pro Ala Leu
85 90 95

AGC CTG CAA GAT AGC CAT TCT TTC ACC GAC GCC GAC ATG ATG AAC
Ser Leu Ala Glu Asp Ser His Phe Thr Asp Ala Asp Met Val Ser
100 105 110

TTC GTC AAC TTC GCG GAA CAT GAC AAG GAA TTC TCC AAC CCA GCC TAC
Phe Val Asn Leu Val Glu His Lys Glu Arg Asp Tyr
115 120 125

CAC CAT GCA GAG TTC GAT CTT TCC AAG ATC CCA GCA GGA GAA
His His Arg Glu Phe Arg Asp Leu Ser Lys Ile Pro Glu Glu
130 135 140

GCT GTC AGC GCA GCC GAA TTC CGG ATC TAC AAC GAC TAC ATG CGG
Ala Val Thr Ala Ala Glu Phe Arg Ile Tyr Lys Asp Tyr Ile Arg Glu
145 150 155 160

GGC TTC GAC AAT GAG ACG TTC CGG ATC AGC GAT TAT CAG GTC CTC CAG
Arg Phe Asp Asn Glu Thr Phe Arg Ile Ser Val Tyr Glu Val Leu Glu
165 170 175

GAG CAC TTG GCC AGG GAA TTC GAT CTC TCT CTC GAC AGC CTT ACC
Glu His Leu Arg Glu Ser Asp Leu Phe Leu Leu Asp Ser Arg Thr
180 185 190
```
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>5</td>
</tr>
<tr>
<td>Trp</td>
<td>10</td>
</tr>
<tr>
<td>Val</td>
<td>15</td>
</tr>
<tr>
<td>Phe</td>
<td>20</td>
</tr>
<tr>
<td>Asp</td>
<td>25</td>
</tr>
<tr>
<td>Pro</td>
<td>30</td>
</tr>
<tr>
<td>Ser</td>
<td>35</td>
</tr>
</tbody>
</table>

**Sequence Information**

**Sequence:** CTC TGG GCC TCG GAG GAG GCC TGG TGT TTT GAC ATC ACA GCC ACC

**Characteristics:**
- **Length:** 408 amino acids
- **Type:** Amino acid
- **Topology:** Linear

**Molecule Type:** Protein

**Sequence Description:**
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Pro Ser
Ser Phe Ile His Arg Arg Leu Arg Ser Glu Arg Arg Gly Met Glu
Arg Glu Ile Leu Ser Ile Leu Gly Leu Pro His Arg Pro Arg Pro His

**Sequence Details:**

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>624</td>
</tr>
<tr>
<td>Ser</td>
<td>672</td>
</tr>
<tr>
<td>Trp</td>
<td>720</td>
</tr>
<tr>
<td>Val</td>
<td>768</td>
</tr>
<tr>
<td>Phe</td>
<td>816</td>
</tr>
<tr>
<td>Asp</td>
<td>864</td>
</tr>
<tr>
<td>Pro</td>
<td>912</td>
</tr>
<tr>
<td>Gin</td>
<td>960</td>
</tr>
<tr>
<td>Lys</td>
<td>1008</td>
</tr>
<tr>
<td>Ser</td>
<td>1056</td>
</tr>
<tr>
<td>Gly</td>
<td>1104</td>
</tr>
<tr>
<td>Phe</td>
<td>1152</td>
</tr>
<tr>
<td>Arg</td>
<td>1200</td>
</tr>
<tr>
<td>Gly</td>
<td>1254</td>
</tr>
<tr>
<td>Leu</td>
<td>1277</td>
</tr>
</tbody>
</table>
Leu Gln Gly Lys His Asn Ser Ala Pro Met Phe Met Leu Asp Leu Tyr
50       55       60
Asn Ala His Gly Gln Arg Gly Arg Pro Gly Gly Gln Gly Phe
65       70       75       90
Ser Tyr Pro Tyr Lys Ala Val Phe Ser Thr Gln Gly Pro Pro Leu Ala
85       90
Ser Leu Gln Asp Ser His Phe Leu Thr Asp Ala Asp Met Val Met Ser
100      105      110
Phe Val Asn Leu Val Glu His Asp Lys Glu Phe Phe His Pro Arg Tyr
115      120      125
His His Arg Glu Phe Arg Phe Asp Leu Ser Lys Ile Pro Glu Gly Glu
130      135      140
Ala Val Thr Ala Ala Glu Phe Arg Ile Tyr Lys Asp Tyr Ile Arg Glu
145      150      155      160
Arg Phe Asp Asn Glu Thr Phe Arg Ile Ser Val Tyr Gln Val Leu Gln
165      170      175
Glu His Leu Gly Arg Glu Ser Asp Leu Phe Leu Leu Asp Ser Arg Thr
180      185      190
Leu Trp Ala Ser Glu Gly Trp Leu Val Phe Asp Ile Thr Ala Thr
195      200      205
Ser Asn His Trp Val Val Asn Pro Arg His Asn Leu Gly Leu Gln Leu
210      215      220
Ser Val Glu Thr Leu Asp Gly Gln Ser Ile Asn Pro Lys Leu Ala Gly
225      230      235      240
Leu Ile Gly Arg His Gly Pro Gln Asn Lys Gln Pro Phe Met Val Ala
245      250      255
Phe Phe Lys Ala Thr Glu Val His Phe Arg Ser Ile Arg Ser Thr Gly
260      265      270
Ser Lys Gln Arg Ser Glu Asn Arg Ser Lys Thr Pro Lys Asn Gln Glu
275      280      285
Ala Leu Arg Met Ala Asn Val Ala Glu Asn Ser Ser Ser Asp Gln Arg
290      295      300
Gln Ala Cys Lys Lys His Glu Leu Tyr Val Ser Phe Arg Asp Leu Gly
305      310      315      320
Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly Tyr Ala Ala Tyr Tyr Cys
325      330      335
Glu Gly Glu Cys Ala Phe Pro Leu Asn Ser Tyr Met Asn Ala Thr Asn
340      345      350
His Ala Ile Val Gln Thr Leu Val His Phe Ile Asn Pro Glu Thr Val
355      360      365
Pro Lys Pro Cys Ala Pro Thr Gln Leu Asn Ala Ile Ser Val Leu
370      375      380
Tyr Phe Asp Asp Ser Ser Asn Val Ile Leu Lys Tyr Arg Asn Met
385      390      395      400
Val Val Arg Ala Cys Gly Cys His
405

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 525 base pairs
(B) TYPE: nucleic acid
(C) STRANDINESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: YES

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..516
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "CBMP2B-1"
/note= "CBMP2B-1 - FUSION"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

ATG AAA GCA ATT TTC CTA CTG AAA GGT TCA CTG GAC AGA GAT CTG GAC
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp 48
1  5  10  15

TCT CGT CTG GAT CTG GAC GYT GCT GTT ACC GAG CAC AAA GAC CTG TCT GAT
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp 96
20 25 30

CAC CTG GTC GTC GAC CTG GCT GCT AAC GAC CGT CCT OTC GTT
His Leu Val Leu Val Ala Arg Asn Asp Leu Ala Arg Ile Val 144
35 40 45

ACT CCC GGC TCT GGT TAC GGT GGC CCT AAG CAT CAC TCA CAG GCG
Thr Pro Gly Ser Arg Tyr Val Ala Asp Pro Lys His Ser Gln Arg 192
50 55 60

GCC AGG AAG AAT AAG AAC TSC CGG CGG CAC TCG CTC TAT GAC
 Ala Arg Lys Lys Asn Lys Ala Cys Arg Arg His Ser Leu Tyr Val Asp 240
65 70 75 80

TTC ACC GAT GTG GCC ATT GAC TGG GTG CCT CCA CCA GCA GCG TAC
Phe Ser Asp Val Gly Trp Asn Asp Trp Ile Val Ala Pro Pro Gly Tyr 288
85 90 95

CAG GCC TTC TAC TAC CAT GGC CAA TGC CCT CCC CTA CGG GAT CAC
 Gln Ala Phe Tyr Cys His Gly Cys Pro Phe Pro Leu Ala Asp His 336
100 105 110

TTC ACC ACC ACC GAC GCC GTG GTG ACC CTG GTG AAC TCT GTC
Phe Asn Ser Thr Asn His Val Val Gln Thr Leu Val Asn Ser Val 384
115 120 125

AAC TCC AAG ATC CCT AAG GCT TGC TGC CCC ACC GAG CTG TCC GCC
Aan Ser Lys Ile Pro Aas Cys Val Pro Thr Glu Leu Ser Ala 432
130 135 140

ATC ACC GTG TAC CTG GAC GAG AAT GAG AAG GTG GTG CGT AAG AAC
Ile Ser Met Leu Tyr Leu Asp Glu Asn Aas Leu Leu Val Lys Asn 480
145 150 155 160

TAC CAG GAG ATG GTA GAG GCC TGC GCC GCC TCG TAC GCG TAA CTG CAG
Tyr Gln Glu Met Val Val Glu Gly Cys Gly Cys Arg 525
165 170

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 172 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
1  5  10  15

Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
20 25 30

His Leu Val Leu Val Ala Arg Asn Asp Leu Ala Arg Ile Val
35 40 45

Thr Pro Gly Ser Arg Tyr Val Ala Asp Pro Lys His Ser Gln Arg
50 55 60

Ala Arg Lys Lys Asn Lys Asn Cys Arg Arg His Ser Leu Tyr Val Asp
65 70 75 80
Phe Ser Asp Val Gly Trp Asn Trp Ile Val Ala Pro Pro Gly Tyr
85 90 95
Gln Ala Phe Tyr Cys His Gly Cys Pro Phe Pro Leu Ala Asp His
100 105 110
Phe Asn Ser Thr Asn His Ala Val Gly Thr Leu Val Asn Ser Val
115 120 125
Asn Ser Lys Ile Pro Lys Ala Cys Cys Pro Thr Glu Leu Ser Ala
130 135 140
Ile Ser Met Leu Tyr Leu Asp Glu Asn Glu Lys Val Leu Lys Asn
145 150 155 160
Tyr Gln Glu Met Val Val Gly Cys Gly Cys Arg
165 170

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1586 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1257
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "CBMP2B-2"
/note= "CBMP2B - FUSION"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ATG AAA GCA ATT TTC GCA GTA GGC AGA GAT
Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
1 5 10 15
TCT CGT TCT CGT GAC GAA
Ser Arg Leu Asp Leu Asp Asp Lys Leu Ser Ser Leu
20 25 30
CAC CCG GTT GTC GAC GCT GAT
His Leu Met Leu Leu Asp Leu Ala Arg Asn Asp Leu Ala Arg Ile Val
35 40 45
ACT CCC GGG TCT GTT GCG GAT
Thr Pro Gly Ser Arg Tyr Val Asp Leu Glu Phe Pro Gly Leu Leu
50 55 60
CTG CCG TGC GAG GGC ACA CTT CAG ARG TTT GCG CTC GCC
Leu Arg Asp Phe Glu Ala Thr Leu Leu Gln Met Phe Leu Arg Arg
65 70 75 80
GCC CCG CAG CCT AGG AGT GCC GTC ATT CGC CAT
Arg Pro Glu Pro Ser Lys Ser Ala Val Ile Pro Asp Tyr Met Arg Asp
85 90 95
CTT TAC CCG GTC TTT GAG GAA GAG GAG GAG CAG GAT
Leu Tyr Arg Leu Glu Ser Gly Glu Glu Glu Glu Ile His Ser
100 105 110
ACT GGT CTT GAT TAT CTT GAC GCC CCG AGC CAG GCC GCC ACC ACC GTC
Thr Gly Leu Glu Tyr Pro Glu Arg Pro Ala Ser Arg Ala Asn Thr Val
115 120 125
AGG AGC TTC CAC CAC GAA GAA CAT CTG GAG AAC ATC CCA GGG ACC AGT
  Arg Ser Phe His His Glu Glu His Leu Glu Asn Ile Pro Gly Thr Ser
  130 135

GAA AAC TCT GCT TTT CGT TGC TTC ACG AGG ACC ATC CCA TTC GAG
  Glu Asn Ser Ala Phe Arg Phe Leu Phe Leu Ser Ser Ile Pro Glu
  145 150 155 160

AAC GAG GCC ATG TCC CTC GCA GAT CTT CGC CTC TTC CGG GAG CAG
  Asn Glu Ala Ile Ser Ala Leu Arg Phe Arg Gly Glu Val
  165 170 175

GAC CAG GCC CCT GAT TGG GAA AGG GCC TTC CAC CGT ATA AAC ATT TAT
  Asp Glu Gly Pro Asp Trp Glu Arg Gly Phe His Arg Ile Asn Ile Tyr
  180 185 190

GAG ATT ATG AAG CCC CCA CAA GTA GTG CCT CGG CAC CTC ATC ACA
  Glu Val Met Lys Pro Pro Ala Glu Val Val Pro His Leu Ile Thr
  195 200 205

CGA CTA CGG GAC AGG AGA CTG CAC CCT AAC ATT ACA CGA GGA
  Arg Leu Leu Asp Thr Arg Leu Val His His Val Thr Arg Trp Glu
  210 215 220

ACT TTG GAT GCG ACC CCT CGG ACC CTG ACC CGG GAG AAC
  Thr Phe Asp Val Ser Pro Ala Val Arg Thr Arg Gly Lys Gin
  225 230 235 240

CCA AAC TAT GGG CTA CAC CAC GAG GTG ACT CTC CAT CAG ACT CGG
  Pro Asn Tyr Gly Leu Ala Ile Glu Val Thr His Leu His Leu Thr Arg
  245 250 255

ACC CAC CAG GCC GAG CAT TGC AGG ATT AGC CTA CGG TTC TTA CCT CAA GGG
  Thr His Gly Gly Gin His Ile Ser Arg Ser Leu Met Pro Gly Gin
  260 265 270

AGT GGG AAT TGG GCC CAG CCT CGG CCC CTC CGG GAC ACC ATC
  Ser Gly Asn Trp Ala Gin Leu Arg Pro Leu Leu Val Phe Gly His
  275 280 285

GAT GGC CGG GCC CAT TTC ACC CGA CCC CGG AGG AAG CAA
  Asp Arg Gly Gly Ala Thr Arg Arg Ala Lys Arg Ser
  290 295 300

CCT AAG CAT CAC TCA CAG CGG GCC AGG AAG AAG AAT AAC TGC COG
  Pro Lys His His Ser Gin Arg Arg Arg Asn Asp Arg COG
  305 310 315 320

CCG CAC TGG CTC TAT GGG CAC CTC CTG ACC GTG GCC AAG TGG
  Arg His Ser Leu Tyr Val Asp Phe Ser Pro Val Gly Trp Asn Asp Trp
  325 330 335

ATT GGG GCC CCA CCA GAC CTG CAC GCC TTC TAC GAC CCC TAC GGC TGC
  Ile Val Ala Pro Pro Gly Tyr Gin Ala Phe Tyr Cys His Gly Aep Cys
  340 345 350

CCC TTT CCA CTG GCT GAC CAC CTC AAC TCA ACC AAC CAT GCC ATT GTG
  Pro Pro Pro Leu Ala Pro His Leu Ser Thr Asn His Ala Ile Val
  355 360 365

CAG ACC CTG GTC AAT TCT GTT AAT TCC AGT ATC CCC AAA GCC TGG TG
  Gln Thr Leu Val Asn Ser Val Asn Ser Ser Ile Pro Lys Ala Cys Cys
  370 375 380

GGG ATG GCC GAC ATC TTC ATC AGG GCC TGG CTG
  Val Pro Thr Glu Asp Ala Ile Ser Met Leu Tyr Leu Asp Gly Tyr
  385 390 395 400

ACCAACACACACACACACACACACAGTATGAGAGATGATCAGAGATGACGACGACGACGAC 1357

CAGACGTGC TTCTTTATAG ATGGACTTTTT ATTAA AAAAAAA AATGGAAAA 1417
(2) INFORMATION FOR SEQ ID NO: 19:
  (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 419 amino acids
  (B) TYPE: amino acid
  (D) TOPOLOGY: linear
  (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Met Lys Ala Ile Phe Val Leu Lys Gly Ser Leu Asp Arg Asp Leu Asp
  1     5       10      15
Ser Arg Leu Asp Leu Asp Val Arg Thr Asp His Lys Asp Leu Ser Asp
  20    25       30
His Leu Val Leu Val Asp Ala Arg Asn Ala Asp Leu Ala Arg Ile Val
  35    40       45
Thr Pro Gly Ser Arg Tyr Val Ala Asp Leu Phe Pro Gly Glu Leu
  50    55       60
Leu Arg Asp Phe Glu Ala Thr Leu Leu Gln Met Phe Gly Leu Arg Arg
  65    70       75      80
Arg Pro Gln Pro Ser Lys Ser Ala Val Ile Pro Asp Tyr Met Arg Asp
  85    90      95
Leu Tyr Arg Leu Gln Ser Gly Glu Glu Glu Glu Glu Ile His Ser
 100   105     110
Thr Gly Leu Glu Tyr Pro Glu Arg Pro Ala Ser Arg Ala Asn Thr Val
 115   120     125
Arg Ser Phe His His Glu His Leu Glu Asn Ile Pro Gly Thr Ser
 126   135     140
Glu Asn Ser Ala Phe Arg Phe Leu Phe Asn Ser Ile Pro Glu
 145   150     155    160
Asn Glu Ala Ile Ser Ser Ala Glu Leu Arg Phe Glu Glu Gln Val
 165   170     175
Asp Gln Gly Pro Asp Trp Glu Arg Gly Phe His Arg Ile Asn Ile Tyr
 180   185     190
Glu Val Met Lys Pro Pro Ala Glu Val Val Pro Gly His Leu Ile Thr
 195   200     205
Arg Leu Leu Asp Thr Arg Leu Val His Asn Val Thr Arg Trp Glu
 210   215     220
Thr Phe Asp Val Ser Pro Ala Val Leu Arg Thr Thr Arg Glu Lys Gln
 225   230     235    240
Pro Asn Tyr Gly Leu Ala Ile Glu Val Thr His Leu His Gln Thr Arg
 245   250     255
Thr His Gln Gly Gln His Val Arg Ile Ser Arg Ser Leu Pro Gln Gly
 260   265     270
Ser Gly Asn Trp Ala Gln Leu Arg Pro Leu Leu Val Thr Phe Gly His
 275   280     285
Asp Gly Arg Gly His Ala Thr Arg Arg Arg Ala Lys Arg Ser
 290   295     300
Pro Lys His His Ser Glu Arg Ala Arg Lys Lys Asn Lys Asn Cys Arg
 305   310     315    320
Arg His Ser Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asp Val Trp 325 330 335
Ile Val Ala Pro Pro Gly Tyr Gin Ala Phe Tyr Cys His Gly Asp Cys 340 345 350
Pro Phe Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala Ile Val 355 360 365
Gln Thr Leu Val Asn Ser Val Asn Ser Ser Ile Pro Lys Ala Cys Cye 370 375 380
Val Pro Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu Asp Glu Tyr 385 390 395 400
Asp Lys Val Val Lys Tyr Gin Glu Met Val Val Glu Gly Cys Gly Cys Arg 405 410 415

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 574 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:
(A) ORGANISM: HOMO SAPIENS

(iv) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..327
(D) OTHER INFORMATION: /product= "MATURE hCBMP3 (PARTIAL)"
/feature= "THIS PARTIAL SEQUENCE OF THE MATURE HUMAN CBMP3 PROTEIN INCLUDE THE FIRST THREE CYSTEINES OF THE CONSERVED 7 CYSTEINE SKELETON."

(v) FEATURE:
(A) NAME/KEY: inttron
(B) LOCATION: 328..574

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 20:
CGA GCT TCT AAA ATA GAA TAC CAG TAT AAA AAG GAT GAG GTG TGG GAG 48
Arg Ala Ser Lys Ile Glu Tyr Gin Tyr Lys Lys Asp Glu Val Trp Glu 1 5 10 15
GAG AGA AAG CCT TAC AAG ACC CTT CAG GGC TCA GGC TGA AAG AGT 96
Glu Arg Lys Pro Tyr Lys Thr Leu Gin Ser Gly Pro Glu Lys Ser 20 25 30
AAG AAT AAA AAG AAA CAG AGA AAG GGC CCT CAT CCG AAG AGC CAG ACG 144
Lys Aen Lys Lys Lys Gin Arg Lys Gin Arg Lys Pro His Arg Lys Ser Gin Thr 35 40 45
CTC CAA TTT GAT GAG CAC CCG AAA AAG GCA AGG AGA AAG CAG TGG 192
Leu Gin Phe Asp Glu Gin Thr Leu Lys Ala Arg Arg Arg Trp Thr 50 55 60
ATT CCT CGG AAT TGC GGC AGG AGA TAC CTC AAG GTA GAC TTG GCA 240
Ile Glu Pro Asn Ser Cys Ala Arg Tyr Leu Lys Val Asp Phe Ala 65 70 75 80
GAT ATT GCC TGG AGT GGA TGG ATT ATC TTC CCA AAG TCC TTC GTG GCC 288
Asp Ile Gin Trp Ser Glu Trp Ile Ile Ser Pro Ser Phe Asp Ala 85 90 95
TAT TAT TCG TGG GCA GCC TGC CAG TTC CCC ATG CCA AAG GTG CATTG 337
Tyr Tyr Cys Ser Gly Ala Cys Gin Glu Phe Pro Met Pro Lys 100 105
TTCTCTGCC TGTACCTACT TCCTATTTCC ATTATAGAA AGACATTG ACTAAGTTAG
397
TGTCATATA GGGGGTTTGT GTAAGTTT GGTTTCCAT TGGCAAAATC CATTGGACC 457
(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Arg Ala Ser Lys Ile Glu Tyr Gln Tyr Lys Lys Asp Glu Val Trp Glu
Glu Arg Lys Pro Tyr Lys Thr Leu Gln Gly Ser Gly Pro Glu Lys Ser
Lys Asn Lys Lys Gln Arg Gly Pro His Arg Lys Ser Gln Thr
Leu Gln Phe Asp Glu Gln Thr Leu Lys Ala Arg Arg Lys Glu Trp
Ile Glu Pro Arg Asn Cys Ala Arg Arg Tyr Leu Lys Val Asp Phe Ala
Asp Ile Gly Trp Ser Glu Trp Ile Ile Ser Pro Lys Ser Phe Asp Ala
Tyr Tyr Cys Ser Gly Ala Cys Gin Phe Pro Met Pro Lys

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 779 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: MURINE
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..549
(D) OTHER INFORMATION: /product= "MATURE mBMP3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TCT ACG GGG GTC CTT CTC CCC TTG CAG AAC AAT GAG CTA CCT GGG GCA
Ser Thr Gly Val Leu Leu Pro Leu Gln Asn Asn Glu Leu Pro Gly Ala
Glu Tat CAG TAC AAG GAG GAG GGC TGG GAG GAG AAG CTT TAT
Glu Tyr Gln Tyr Lys Glu Gly Ala Trp Glu Arg Lys Pro Tyr
Lys Ser Leu Gln Thr Gln Pro Pro Glu Lys Ser Arg Asn Lys Lys Lys
35 40
CAG AGG AAG GGG TCC CAT CAG AAG GGA CAG CTC CAA TTT CAT GAG
Gln Arg Lys Gly Ser His Gln Lys Gly Gln Thr Leu Gln Phe Asp Glu
50 55 60
CAG ACC CTC AAG AAG GCA AGG CCA AAG CAG TGG TTC GAA CCT CGG AAC
Gln Thr Leu Lys Ala Arg Arg Lys Gln Trp Val Glu Pro Arg Asn
65 70 75 80
```
US 7,176,284 B2

-continued

TGT GCC AGG AGG TAC CTT AAA GTC GAC TTT GCT GAT ATC GGC TGG AGC
Cys Ala Arg Arg Tyr Leu Lys Val Asp Phe Ala Asp Ile Gly Trp Ser
  85           90

GAA TGG ATT ATC TCT CCC AAG TCA TTT GAT TCT TAC TGC TCT GGA
Glu Trp Ile Ile Ser Pro Lys Ser Phe Ala Phe Tyr Cys Ser Gly
  100          105 110

GCC TGC CAG TTC CCC ATG CCA AAG TCT TGG AAA CCA TCA AAT CAC GCC
Ala Cys Gln Phe Pro Met Pro Lys Ser Leu Lys Pro Ser Asn His Ala
  115         120 125

ACC ATC CAG AGC ATA GTC GCA GGG GTC GTC GCC TGG ATT CCC
Thr Ile Gln Ser Ile Val Arg Ala Val Gly Val Ser Gly Ile Pro
  130         135 140

GAG CTC TGC TGT GGG GAA AAG ATG TCA GCT ACC ATC ATC TTC
Glu Pro Cys Cys Val Pro Glu Lys Met Ser Leu Ser Ile Leu Phe
  145         150 155 160

TTT GAT GAA AAC AAG AAT GTA GTC CTA AAA GCC TCC GGC TCC GGC ATT CCC
Phe Asp Glu Asn Val Lys Leu Asp Leu Tyr Phe Asn Pro Met Thr
  165         170 175

GTC GCC TGC TGT CTA ACT TTC TCT CAG TGT GCT TGT AGA TAACCTCTTC
Val Asp Ser Cys Ala Cys Arg
  180

TCCAATCAC AGTGGTTT TATGCTTT TTTTTTTT TTTTTGCCC AAAAGATT
  185

TGATAAGGG AAGAAAATGA ACAATGAT TGAAGTTTC CACCAAAAC AACCAGACTG
  190

TATTTCCCT CAAAGTAC TAAAAATGAG ATTTTAGA AAATGACCT CTTTTT
  195

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
  200

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:
Ser Thr Gly Val Leu Leu Pro Leu Gln Asn Asn Glu Leu Pro Gly Ala
  1      5      10     15
Glu Tyr Gln Tyr Lys Glu Gly Ala Trp Glu Arg Lys Pro Tyr
  20     25     30
Lys Ser Leu Gln Thr Gln Pro Glu Lys Ser Arg Asn Lys Lys Lys
  35     40     45
Gln Arg Lys Gly Ser His Gln Gly Lys Thr Leu Gln Phe Asp Glu
  50     55     60
Gln Thr Leu Lys Lys Ala Arg Arg Lys Gln Trp Val Glu Pro Arg Asn
  65     70     75     80
Cys Ala Arg Arg Tyr Leu Lys Val Asp Phe Ala Asp Ile Gly Trp Ser
  85     90     95
Glu Trp Ile Ile Ser Pro Lys Ser Phe Asp Ala Phe Tyr Cys Ser Gly
 100    105    110
Ala Cys Gln Phe Pro Met Pro Lys Ser Leu Lys Pro Ser Asn His Ala
 115    120    125
Thr Ile Gln Ser Ile Val Arg Ala Val Gly Val Ser Gly Ile Pro
 130    135    140
Glu Pro Cys Cys Val Pro Glu Lys Met Ser Ser Leu Ser Ile Leu Phe
 145    150    155    160
```
Val Asp Ser Cys Ala Cys Arg

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1873 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 104 .. 1393
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "MOP1"
/note= "MOP1 CDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
CTG CAG CAAG TGAC CGGG
CGT GTG ACCG CTGCC CCGC
CCCT CGCTGC CCAC CCGGG
CGG CGCGGG GC CGGTGCC
GGAT CGCGCG TAGAG CGGC
G CG ATG CAC GTG CGC

Met His Val Arg

TCG CTG CGC GCT CGG CCAC
CAC AGC TTC GTG GCG CTC
TGG CGC CTGCC CCGCGG GCG
GCG ATG CAC GTC GCC

Met His Val Arg

CTG TTC TTG CTG CGC TCC
GCC GCC CGT GCC TC CCAG
AGC AA GAG CAG CAG GAG
GAG ATG CAG CGG GAG ATC

CTC AAC TTC AAC TAC CGC
GCC TGG CAG GAC TGG CTC
GAG GAG GAG GAG GTG

Glu Met Gln Arg Glu Ile Leu Gly Leu Pro His Arg Pro

GCC CGC CAC TCC GAG GCA
AAC CAT TCG CGC CCC ATG TTC
ATG TGG GAG ATG CAG

Val His Ser Pro Phe Ile His Arg Arg Arg Arg Arg

GCC CGC GAC TAC AAC GCC
ATG GCG GAG ATG GCG
GAG CAG AAC GAG

Arg Pro His Leu Val Ala Leu Gly Ser Gly Pro Asp Gly Gln

GAC CGC TAG TAC GCC ATG
GTG GCG GAG ATG TGG
GAG CAG AAC GAG

Asp Leu Tyr Aan Ala Leu Val Glu Glu Ser Gly Pro Asp Gly Gln

GTC TCC TAC CCC TAC AAG GCC
GCT TCC ACC GAT GCC ACC
GAC GCA GCC CCT

Gly Phe Ser Tyr Pro Phe Ile Leu Gly Leu Pro His Arg Pro

TTA GCC AGC TCT CAG GCC
AGC ATG GCC GTC TCC
ACC GAC GCC GCC ATG GTC

Leu Ala Ser Leu Gln Asp Ser His Phe Leu Thr Asp Ala Asp Met Val

AYG AGC TTC GTC AAC CTA GTG
GAA CAT GAC AAA GAA TTC TTC
TCC CAC CCT

Met Ser Phe Val Asn Leu Val Glu His Asp Lys Glu Phe Phe His Pro

CGA TAC CAC CGG GAG TTC CGG
TCT CTT CAT AAG GCC CCC GAG

Arg Tyr His His Arg Glu Phe Arg Phe Leu Ser Lys Ile Pro Glu

GCC GAA CGG GTG ACC GCA GCC
GAA TTC AGC ATC TAT AAC AAG GAC TAC ACC

Gly Glu Arg Val Thr Ala Ala Glu Phe Arg Ile Tyr Lys Asp Tyr Ile

CGG GAG CGA TTT GAC AAC GAG TTC CAG ATC ACA GTC

Arg Glu Arg Phe Asn Asp Glu Thr Phe Glu Ile Thr Val Tyr Glu Val
CTC CAG GAG CAC TCA GCC AGG GAG TCG GAC CTC TCT TCG CTG GAC AGC
Leu Gln Gln His Ser Gly Arg Arg Ser Asp Leu Phe Leu Leu Asp Ser
200 205 210
CAG ACC ATC TGG GCT CTC GAG GAG TGG TGG GTC TTT GAT ATC ACA
Arg Thr Ile Trp Ala Ser Gly Gly Trp Leu Val Phe Asp Ile Thr
215 220 225
GCC ACC AGC AAC TGG TGG GTC CTC CCT CGG CAC AAC CTG GGC TTA
Ala Thr Ser Asn His Trp Val Val Asn Pro Arg His Aen Leu Gly Leu
230 235 240
GCA GCC CAG ATT GGA CGG CAT GGA CCC CAG AAC TGC ATG
Ala Gly Leu Ile Gly Arg His Gly Pro Gln Aen Lys Gly Pro Phe Met
739 744 749
787
835
883
931
979
1027
1075
1123
1171
1219
1267
1315
1363
1413
1463
1513
1563
1613
1663
1713
1763
1813
1863
1913
1963
2013
(2) INFORMATION FOR SEQ ID NO: 25:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 430 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met His Val Arg Ser Leu Arg Ala Ala Ala Pro His Ser Phe Val Ala
1 5 10 15

Leu Trp Ala Pro Leu Phe Leu Leu Arg Ser Ala Leu Ala Asp Phe Ser
20 40

Leu Asp Asn Glu Val His Ser Ser Phe Ile His Arg Arg Leu Arg Ser
35 40

Gln Glu Arg Arg Glu Met Gln Arg Glu Ile Leu Ser Ile Leu Gly Leu
50 55 60

Pro His Arg Pro Arg Pro His Leu Gln Gly Lys His Asn Ser Ala Pro
65 70 75 80

Met Phe Met Leu Asp Leu Tyr Asn Ala Met Ala Val Gly Glu Ser Gly
85 90 95

Pro Asp Gly Gln Gly Phe Ser Tyr Pro Tyr Lys Ala Val Phe Ser Thr
100 105 110

Gln Gly Pro Pro Leu Ala Ser Leu Gln Asp Ser His Phe Leu Thr Asp
115 120 125

Ala Asp Met Val Met Ser Phe Val Asn Leu Val Glu His Asp Lys Glu
130 135 140

Phe Phe His Pro Arg Tyr His His Arg Glu Phe Arg Phe Asp Leu Ser
145 150 155 160

Lys Ile Pro Glu Gly Glu Arg Val Thr Ala Ala Glu Phe Arg Ile Tyr
165 170 175

Lys Asp Tyr Ile Arg Glu Arg Phe Asp Asn Glu Thr Phe Gln Ile Thr
180 185 190

Val Tyr Gln Val Leu Gln Glu His Ser Gly Arg Glu Ser Asp Leu Phe
195 200 205

Leu Leu Asp Ser Arg Thr Ile Trp Ala Ser Glu Gly Trp Leu Val
210 215 220

Phe Asp Ile Thr Ala Thr Ser Asn His Thr Val Val Asn Pro Arg His
225 230 235 240

Asn Leu Gly Leu Gln Leu Ser Val Glu Thr Leu Asp Gly Gln Ser Ile
245 250 255

Asn Pro Lys Leu Ala Gly Leu Ile Gly Arg His Gly Pro Gln Asn Lys
260 265 270

Gln Pro Phe Met Val Ala Phe Phe Lys Ala Thr Glu Val His Leu Arg
275 280 285

Ser Ile Arg Ser Thr Gly Lys Gln Arg Ser Glu Ser Asp Leu Glu
290 295 300

Thr Pro Lys Aan Gln Glu Ala Leu Arg Met Ala Ser Val Ala Glu Asn
305 310 315 320

Ser Ser Ser Asp Gln Arg Gln Ala Cys Lys His Glu Leu Tyr Val
325 330 335

Ser Phe Arg Asp Leu Gly Trp Gln Asp Trp Ile Ile Ala Pro Glu Gly
340 345 350

Tyr Ala Ala Tyr Tyr Cys Glu Gly Glu Cys Ala Phe Pro Leu Asn Ser
355 360 365
<table>
<thead>
<tr>
<th>Tyr Met Asn Ala Thr Asn His Ala Ile Val Gln Thr Leu Val His Phe</th>
</tr>
</thead>
<tbody>
<tr>
<td>370 375 380</td>
</tr>
<tr>
<td>Ile Asn Pro Asp Thr Val Pro Lys Pro Cys Cys Ala Pro Thr Gln Leu</td>
</tr>
<tr>
<td>385 390 395 400</td>
</tr>
<tr>
<td>Asn Ala Ile Ser Val Leu Tyr Phe Asp Ser Ser Asn Val Ile Leu</td>
</tr>
<tr>
<td>405 410 415</td>
</tr>
<tr>
<td>Lys Lys Tyr Arg Asn Met Val Val Arg Ala Cys Gly Cys His</td>
</tr>
<tr>
<td>420 425 430</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1926 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:
(A) ORGANISM: MURIDAE
(F) TISSUE TYPE: EMBRYO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 93..1289
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product= "mOP2-PP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GCCAGGCACA GCCGCGCCGGT CGTGCTCTCC CCGTCTGGCG TCAGCCGAGC CCGACCAGCT 60
ACCAGCAGT GGGGGCCGCG TGAAGTCGCG AGATG GCT ATG GCT CCC GGG CCA 113
     Met Ala Met Arg Pro Gly Pro
CTC TGG CTA TGG GCC CTT GCT CTG GCG CTG GGA GCC GCC CAC GGT  161
     Leu Trp Leu Leu Gly Leu Ala Leu Cys Ala Leu Gly Gly Gly Gly
     10  15  20
CCG COT CCC CGG CAC ACC TGT CCC CAG COT CCC CTG GGA GCC GCC CAG  209
     Pro Arg Pro Pro His Thr Cys Pro Gln Arg Arg Leu Arg Glu
     25  30  35
CCG GCC GAC ATG CAG COT GAA ATC CTG GCC CTG GCC GTA CCA CCG GGA  257
     Arg Arg Asp Met Gln Arg Ile Leu Leu Val Leu Gly Leu Pro Gly
     40  45  50  55
CGG CCC CGA CCC GTG CAA CCC CCC GCT GCC CGS CAG CAA GCC TCC  305
     Arg Pro Arg Pro Arg Ala Gln Pro Ala Ala Arg Glu Ser
     60  65  70
GGG CCC CTC TGC ATG TGG GAC CTA TAC CAC GCC ATG ACC GAT GAC GAC  353
     Ala Pro Leu Phe Met Leu Asp Leu Tyr His Ala Met Thr Asp Asp
     75  80  85
GAC GCC GGG CCA CCA CAG CAC TAT CAC GTA CGC CGC CAG CTG ATG  401
     Asp Gly Gly Pro Pro Gln Ala His Leu Gly Arg Ala Asp Leu Val Met
     95  100
AAG TGC GTC AAC ATG CAG GAA CCG GAC GCT TCA CAC TAC CCG CAG CAG  449
     Ser Phe Val Ala Gly Arg Asp Thr Leu Gly Tyr Glu Gln
     105 110 115
CCA CAC TGG AAG GAA TCC CAC TAC GAC CTA ACC CAG ATC CCT GCT GGG  497
     Pro His Trp Lys Glu Phe His Phe Asp Leu Thr Gln Ile Pro Ala Gly
     120 125 130 135
GAG GCT GTC ACA GCT GAG TGC CAG ATC TAC AAA GAA CCC ACG ACC  545
     Glu Ala Val Thr Ala Ala Glu Phe Arg Ile Tyr Lys Glu Pro Ser Thr
     140 145 150
CAC CGG CTC AAC ACA ACC CTC CAC ATC AGC ATG TGC GAA GTC GTC CAA  593
     His Pro Leu Asn Thr Thr Leu His Ile Ser Met Phe Glu Val Glu
GAG CAC TCC AAC AGG GAG TCT GCAC TCC TTT TCG GAT CTT CAG AGG
641
Glu His Ser Asn Arg Glu Ser Asp Leu Phe Leu Leu Asp Leu Gln Thr
170 175 180

CTC GGA TCT GGG GAC GAC GAG TGG CTC GTC GAG ATC ACA GCA GCC
689
Leu Arg Ser Gly Asp Gly Trp Leu Val Leu Asp Ile Thr Ala Ala
165 190 195

AGT GAC CGA TGG CTC AAG CAT AAC AAG GAC CTC GGA TCT GCC GTC
737
Ser Asp Arg Trp Leu Leu His His Asp Leu Gly Arg Leu
200 205 210 215

TAG GAT GAA ACC GCG GAT GAT CCT GGC TCG GCT GCT
785
Tyr Val Glu Thr Ala Asp Gly His Ser Met Asp Pro Gly Leu Ala Gly
220 225 230

CTG CTT GGA CCA GCA CCA GCG TCC AGA CAG CCT TGC GAT GTA ACC
833
Leu Leu Gly Arg Gln Ser Ser Pro Val Arg Ala Pro Arg Ala Ala Arg
235 240 245

TTC TTC AGG GCC AGC CAG ACT CCT GTG CGG GCC CTC GGA AGA
881
Phe Phe Arg Ala Ser Gln Ser Pro Val Arg Ala Pro Arg Ala Ala Arg
250 255 260

CCA CTA GGC AGG AGG CAG CCA AAG AAA AGG AAC GAG CTT CGG CCC
929
Pro Leu Lys Arg Arg Asp CAA CAA Gln Ser His Pro Ser Pro Ser Leu
265 270 275

AAG AAA CTC CCA GGG ATC TTT GAT GAT GCC CAC GGT TCC GGC GGC
977
Asn Lys Leu Pro Gly Ile Phe Asp Asp His Gly Gly Gly Gly Ala
280 285 290 295

GAG GTT GGC AGG CAT GAG CTC TAC GTC AGC TCC GGC GCC TCT GCC
1025
Glu Val Cys Arg Arg Cys Ala Ser Ser Pro Val Arg Ala Pro Arg Ala Ala Arg
300 305 310

TGC GTG GAC TCG TGG GTC ACT GCC CCC CAG GCC TAC TCT GCC TAT TAC TGT
1073
Trp Leu Asp Trp Val Ile Ala Pro Gly Lys Tyr Ser Ala Tyr Tyr Cys
315 320 325

GAG GGG GAG TGT GCT TCC CTA CGC GAC TCC TGT ATG AAC GCC ACC AAC
1121
Glu Gly Gly Cys Ala Phe Pro Leu Asp Ser Cys Met Asn Ala Thr Asn
330 335 340

CAT GCC ATC TGG CAG TCT CTG CGC ACT ATG AAG CCA GAT GTT GTC
1169
His Ala Ile Leu Gln Ser Leu Val His Leu Met Asp Asp Val Val
345 350 355

CCC AAG GCA TGC TGT GCC CCC ACC AAA CTC ATG GCC ACC TTT TGT CTG
1217
Pro Lys Ala Cys Ala Pro Thr Lys Ser Ala Thr Ser Val Leu
360 365 370 375

TAC TAT GAC ACC AAC AAT ATC TGC CTC CTT AAA AAC GCT AAC ACG
1265
Tyr Tyr Asp Ser Ser Asn Asn Val Ile Leu Arg Lys His Arg Asn Met
380 385 390 395

GTG GTC AAG GCC TGT GCC TGC CAC TGGGCCCCG CCCAGCATCC TCTCCTACT
1319
Val Val Lys Ala Cys Gly Cys His
395

ACCTAACCAT CTGGCGGGGC CCCCTTCGAC AGGGAGAAGC CTTCTCATGT TATCTGACT
1379

CAGACAGGGG CAATAGGAGG CCTTCTACCT CCCCTGGCCA CTTCTGCTA AAMTTCGTG
1439

CTCCCTCCAG TCCCTGCTCC TCCAGGGGTC ATCCGGGCTA TACCCGCC CTCTCCTCCT
1499

TCTTACCCCA AGCATGACTT GAAAGCAGAC AGCAGCAAGC AGCATGACTAT ACTGAGACT
1559

CTGGGCTCAC CACGGAGG CTGAGAGAC AAGACTGATC CTGGGCGATC CTGGGCGAGC
1619

AAAGCCAAAT TCTGAGGTT CTAGAGGGC CTTGGAATC TAAAGTAAT GATCTGGGCT
1679

CTCTACCACT TCCATGCGG CATTGGAGG ATTTCAGGG GTAACAGGAC AGAGAGAC
1739

GCTACCATC TGCCTGACTG CCTGGAATC AGAGACTACT TTTGAGAAA AAGATCCAGG
1799

CCAGCAGTAG CGGGAGCTAT CATTATGCC AGGCTAGAG AGGAGAAGTCT
1859
CTGTGAGTTC AGAAGGACAG AGCAAGGACG TGCTCGGAG AGGGAAAAA AAAAAAACGGAATTC

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:
Met Ala Met Arg Pro Gly Pro Leu Trp Leu Leu Gly Leu Ala Leu Cys
1 5 10 15
Ala Leu Gly Gly Gly His Gly Pro Arg Pro Pro His Thr Cys Pro Gln
20 25 30
Arg Arg Leu Gly Ala Arg Glu Arg Asp Met Gln Arg Glu Ile Leu
35 40 45
Ala Val Leu Gly Leu Gly Arg Pro Arg Pro Arg Ala Glu Pro Ala
50 55 60
Ala Ala Arg Gln Pro Ala Ser Ala Pro Leu Phe Met Leu Asp Leu Tyr
65 70 75 80
His Ala Met Thr Asp Asp Asp Gly Gly Pro Pro Gln Ala His Leu
85 90 95
Gly Arg Ala Asp Leu Val Met Ser Phe Val Asn Met Val Gln Arg Asp
100 105 110
Arg Thr Leu Gly Tyr Gln Glu Pro His Trp Lys Glu Phe His Phe Asp
115 120 125
Leu Thr Gln Ile Pro Ala Gly Glu Ala Val Thr Ala Ala Glu Phe Arg
130 135 140
Ile Tyr Lys Glu Pro Ser Thr His Pro Leu Asn Thr Thr Leu His Ile
145 150 155 160
Ser Met Phe Glu Val Val Gln Glu His Ser Asn Arg Glu Ser Asp Leu
165 170 175
Phe Phe Leu Asp Leu Gln Thr Leu Arg Ser Gly Asp Glu Gly Trp Leu
180 185 190
Val Leu Asp Ile Thr Ala Ala Ser Asp Arg Trp Leu Leu Asn His His
195 200 205
Lys Asp Leu Gly Leu Arg Leu Tyr Val Glu Thr Ala Asp Gly His Ser
210 215 220
Met Asp Pro Gly Leu Ala Gly Leu Leu Gly Arg Gln Ala Pro Arg Ser
225 230 235 240
Arg Glu Pro Phe Met Val Thr Phe Phe Arg Ala Ser Gin Ser Pro Val
245 250 255
Arg Ala Pro Arg Ala Ala Arg Pro Leu Lys Arg Arg Glu Gin Pro Lys Lys
260 265 270
Thr Asn Glu Leu Pro His Pro Asn Lys Leu Pro Gly Ile Phe Asp Asp
275 280 285
Gly His Gly Ser Arg Gly Arg Glu Val Cys Arg Arg His Glu Leu Tyr
290 295 300
Val Ser Phe Arg Asp Leu Gly Trp Leu Asp Trp Val Ile Ala Pro Gln
305 310 315 320
Gly Tyr Ser Ala Tyr Tyr Cys Glu Gly Glu Cys Ala Phe Pro Leu Asp
325 330 335
(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
(F) TISSUE TYPE: HIPPOCAMPUS

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 490 .. 1696
(D) OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
/product* "hOP2-PP"
/note* "hOP2 (cDNA)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGCGCCGGCA
GAGCAGGAGT
GGCTGGAGGA
GCTGTGGTTG
GAGCAGGAGG
TGGCACGGCA

60
GGGCTGGAGG
GCTCCCTATG
AGTGGCGGAG
ACGGCCCAGG
AGGCGCTGGA
GCAACAGCTC

120
CCACACCGCA
CCAAGCGGTG
GCTGCAGGAG
CTCGCCCATC
GCCCCTGCGC
TGCTCGGACC

180
GCGGCCACAG
CCGGACTGGC
GGGTACGGCG
GCGACAGAGG
CATTGGCCGA
GAGTCCCAGT

240
CCGCAGAGTA
GCCCGGCTCT
CGAGGCGGTG
GCGTCCCGGT
CCTCTCCGTC
CAGGAGCCAG

300
GACAGGTGTC
GCGCGGCGGG
GCTCCAGGGA
CCGCGCCTGA
GGCCGGCTGC
CCGCCCGTCC

360
CGCCCCGCC
CGCCGCCCGC
CGCCCGCGGA
GCCCAGCTC
CTTGCCGTCG
GGGCGTCCCC

420
AGGCCCTGGG
TCGGCCGCGG
AGCCGATGCG
CGCCCGCTGA
GCGCCCCAGC
TGAGCGCCCC

480
CGGCCTGCC
ATG ACC GGC
CTC CCC GGC
CCG GCC CTC
TGG CTC GTC
GCG CTC CTG

528
Met Thr Ala Leu Pro Gly Pro Leu Trp Leu Leu Gly Leu

1  5  10

GCG CTA TGC GGC CTG GGC GAG GCC GGC GCC CTG CTA GCA
Ala Leu Cys Ala Leu Gly Pro Gly Pro Leu Arg Pro Pro Pro

15  20  25

GGC TGT CCC CAG CAG GCT CTC GCC GGC CCG CAG CAG CAG
Gly Cys Pro Gln Arg Arg Leu Gly Ala Arg Arg Asp Val Gln

30  35  40  45

CAG GAG ATC CTC GGC CTC CTC GGT CTC CTC GCC GCG
Arg Glu Ile Leu Ala Val Leu Gly Pro Gly Pro Arg Pro Arg

50  55  60

GCG CCA CCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC
Ala Pro Pro Ala Ala Ser Arg Leu Pro Ala Leu Pro Leu Phe Met

65  70  75

CTG GAC CTC TAC CAC GCC ATG GCC GCC GAC GAC GAC GAC
Leu Asp Leu Tyr His Ala Met Ala Gly Asp Asp Asp Gly Ala

80  85  90

CCC GCG GAG CCG GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC
Pro Ala Glu Arg Arg Ala Asp Val Met Ser Phe Val

95  100  105
(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Met Thr Ala Leu Pro Gly Pro Leu Leu Leu Gly Leu Ala Leu Cys 1 5 10 15
Ala Leu Gly Gly Gly Pro Gly Leu Arg Pro Pro Gly Cys Pro 20 25 30
Gln Arg Arg Leu Gly Ala Arg Glu Arg Arg Asp Val Gln Arg Glu Ile 35 40 45
Leu Ala Val Leu Gly Pro Gly Arg Pro Arg Pro Ala Pro Pro 50 55 60
Ala Ala Ser Arg Leu Pro Ala Ser Ala Pro Leu Phe Met Leu Asp Leu 65 70 75 80
Tyr His Ala Met Ala Gly Asp Aep Glu Asp Gly Ala Pro Ala Glu 85 90 95
Arg Arg Leu Gly Arg Ala Leu Val Met Ser Phe Val Aan Met Val 100 105 110
Glu Arg Aep Arg Ala Leu Gly His Gln Glu Pro His Trp Lys Glu Phe 115 120 125
Arg Phe Aep Leu Thr Gln Ile Pro Ala Gly Glu Ala Val Thr Ala Ala 130 135 140
Glu Phe Arg Ile Tyr Lys Val Pro Ser Ile His Leu Leu Aan Arg Thr 145 150 155 160
Leu His Val Ser Met Phe Gln Val Val Gln Glu Gin Ser Aan Arg Glu 165 170 175
Ser Asp Leu Phe Phe Leu Asp Leu Gln Thr Leu Arg Ala Gly Asp Glu 180 185 190
Gly Trp Leu Val Leu Aep Val Thr Ala Ala Ser Aep Cys Trp Leu Leu 195 200 205
Lys Arg His Lys Aep Leu Gly Leu Arg Leu Tyr Val Glu Thr Glu Asp 210 215 220
Gly His Ser Val Asp Pro Gly Leu Ala Gly Leu Leu Gly Gin Arg Ala 225 230 235 240
Pro Arg Ser Gin Gln Pro Phe Val Val Thr Phe Phe Arg Ala Ser Pro 245 250 255
Ser Pro Ile Arg Thr Pro Arg Ala Val Arg Pro Leu Arg Arg Gln 260 265 270
Pro Lys Lys Ser Aan Glu Leu Pro Gln Ala Aan Arg Leu Pro Gly Ile 280 285
Phe Asp Aep Val His Gly Ser His Gly Arg Glu Val Cys Arg Arg His 295 295 300
Glu Leu Tyr Val Ser Phe Gln Aep Leu Gly Gly Trp Leu Aep Trp Val Ile 305 310 315 320
Ala Pro Gin Gly Tyr Ser Ala Tyr Tyr Cys Glu Gly Glu Cys Ser Phe 325 330 335
Pro Leu Aep Ser Cys Met Aan Ala Thr Aan His Ala Ile Leu Gln Ser 340 345 350
Leu Val His Leu Met Lys Pro Aan Ala Val Pro Lys Ala Cys Cys Ala 355 360 365
Pro Thr Lys Leu Ser Ala Thr Ser Val Leu Tyr Asp Ser Ser Asn
370 375 380
Asn Val Ile Leu Arg Lys Ala Arg Asn Met Val Val Lys Ala Cys Gly
385 390 395 400
Cys His

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 102 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..102
(D) OTHER INFORMATION: /label= OPX

/(note= “WHEREIN EACH XAA IS INDEPENDENTLY SELECTED
FROM A GROUP OF ONE OR MORE SPECIFIED AMINO ACIDS
AS DEFINED IN THE SPECIFICATION (SECTION II.B.2.)”

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:
Cys Xaa Xaa His Glu Leu Tyr Val Xaa Phe Xaa Asp Leu Gly Trp Xa
1 5 10 15
Asp Trp Xaa Ile Ala Pro Xaa Gly Tyr Xaa Ala Tyr Tyr Cys Glu Gl
20 25 30
Glu Cys Xaa Phe Pro Leu Xaa Ser Xaa Met Asn Ala Thr Asn Al
35 40 45
Ile Xaa Gin Xaa Leu Val His Xaa Xaa Xaa Pro Xaa Xaa Val Pro Ly
50 55 60
Xaa Cys Cys Ala Pro Thr Xaa Leu Xaa Ala Ala Ser Val Leu Tyr Xa
65 70 75 80
Asp Xaa Ser Xaa Asn Val Xaa Leu Xaa Lys Xaa Arg Asn Met Val Va
85 90 95
Xaa Ala Cys Gly Cys His
100

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 97 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..97
(D) OTHER INFORMATION: /label= OPX-7C

/(note= “WHEREIN EACH XAA INDEPENDENTLY INDICATES
ONE OF THE 20 NATURALLY-OCcurring L-ISOMER,
A-AMINO ACIDS, OR A DERIVATIVE THEREOF.”

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xa
1 5 10 15
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xa
20 25 30
Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xa
35 40 45
(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 102 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
    (A) NAME/KEY: Protein
    (B) LOCATION: 1..102
    (D) OTHER INFORMATION: /label* OPX-8C
    /note* “WHEREIN EACH XAA INDEPENDENTLY INDICATES ONE OF THE 20 NATURALLY-OCURRENCE L-ISOMER A-AMINO ACIDS, OR A DERIVATIVE THEREOF.”

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Cy
85 90 95

Xaa

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 314 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
    (A) NAME/KEY: misc_feature
    (B) LOCATION: 1..314
    (D) OTHER INFORMATION: /note* “CONSENSUS PROBE”

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GATCCTAATG GGCTGTACGT GGACTTCCAG CGCGACGTGG GCTGGGACGA CTGGATCATC
60
GCCCCCGTCG ACTTCGACGC CTACTACTGC TCCGGAGCCT GCCAGTTCCC CTCTGCGGAT
120
CACTTCAACA GCACCAACCA CGCCGTGGTG CAGACCCTGG TGAACAACAT GAACCCCGGC
180
What is claimed is:

1. An isolated osteogenic protein comprising two polypeptides, not joined by a cysteine disulfide bond, each polypeptide having an amino acid sequence sufficiently duplicative of a sequence comprising residues 335 to 431 of SEQ ID NO:1, wherein each polypeptide has an alteration in the conserved cysteine skeleton depicted in amino acid residues 335 to 431 of SEQ ID NO:1 such that the osteogenic protein is capable of inducing local endochondral bone and cartilage when implanted in a mammal in association with a matrix.

2. The isolated osteogenic protein of claim 1, wherein the polypeptide has at least 96 amino acids.

3. The isolated osteogenic protein of claim 2, wherein the polypeptide has no more than 139 amino acids.

4. The isolated osteogenic protein of claim 1, wherein the amino acid sequence is sufficiently duplicative of a sequence comprising residues 335 to 431 of SEQ ID NO:1 such that the osteogenic protein is capable of inducing local endochondral bone and cartilage formation in a rat subcutaneous assay.

5. The isolated osteogenic protein of claim 4, wherein each polypeptide has a molecular weight of approximately 14–16 kD in an unglycosylated form or a molecular weight of 16–18 kD in a glycosylated form.

6. The isolated osteogenic protein of claim 5, wherein the protein is unglycosylated.

7. The isolated osteogenic protein of claim 5, wherein the protein is glycosylated.

8. An isolated osteogenic protein capable of inducing endochondral bone formation when implanted in a mammal in association with a matrix, the osteogenic protein having at least half maximum activity at about 25 to 50 ng per 25 mg of matrix and comprising two polypeptides not joined by a cysteine disulfide bond, wherein each of said polypeptides has an alteration in the conserved cysteine skeleton depicted in amino acid residues 335 to 431 of SEQ ID NO:1 and is encoded by a DNA, one strand of which hybridizes to a DNA consensus probe of SEQ ID NO:33 in 5xSSPE, 10xDenhardt’s solution, and 0.5% SDS at 50° C.

9. The isolated osteogenic protein of claim 8, wherein the polypeptide has at least 96 amino acids.

10. The isolated osteogenic protein of claim 9, wherein the polypeptide has no more than 139 amino acids.

11. The isolated osteogenic protein of claim 8, wherein the osteogenic protein is capable of inducing local endochondral bone and cartilage formation in association with a matrix in a rat subcutaneous assay.

12. The isolated osteogenic protein of claim 11, wherein each polypeptide has a molecular weight of approximately 14–16 kD in an unglycosylated form or a molecular weight of 16–18 kD in a glycosylated form.

13. The isolated osteogenic protein of claim 12, wherein the protein is unglycosylated.

14. The isolated osteogenic protein of claim 12, wherein the protein is glycosylated.

15. An isolated osteogenic protein capable of inducing endochondral bone formation when implanted in a mammal in association with a matrix, the osteogenic protein comprising two polypeptides, not joined by a cysteine disulfide bond, each polypeptide comprising an amino acid sequence having an alteration in the conserved cysteine skeleton depicted in residues 335 to 431 of SEQ ID NO:1 and comprising a conservative amino acid variant of residues 397 to 399 of SEQ ID NO:1, the conservative amino acid variant having at least 60% identity with residues 397 to 399 of SEQ ID NO:1, wherein each of said polypeptides is encoded by a DNA, one strand of which hybridizes to a DNA consensus probe of SEQ ID NO:33 in 5xSSPE, 10xDenhardt’s solution, and 0.5% SDS at 50° C.

16. The isolated osteogenic protein of claim 15, wherein the polypeptide has at least 96 amino acids.

17. The isolated osteogenic protein of claim 15, wherein the polypeptide has no more than 139 amino acids.

18. The isolated osteogenic protein of claim 15, wherein the osteogenic protein is capable of inducing local endochondral bone and cartilage formation in association with a matrix in a rat subcutaneous assay.

19. The isolated osteogenic protein of claim 18, wherein each polypeptide has a molecular weight of approximately 14–16 kD in an unglycosylated form or a molecular weight of 16–18 kD in a glycosylated form.

20. The isolated osteogenic protein of claim 19, wherein the protein is unglycosylated.

21. The isolated osteogenic protein of claim 19, wherein the protein is glycosylated.

* * * * *