US 20080081353A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2008/0081353 A1**

Islam et al. (43) **Pub. Date: Apr. 3, 2008**

(54) **PRODUCTION OF RECOMBINANT HUMAN COLLAGEN**

(75) Inventors: **Nazrul Islam**, Ste-Foy (CA); Francine Goulet, legal representative, Ste-Foy (CA); **Francine Goulet**, Ste-Foy (CA); **Ioana Diana Napa**, Mississauga (CA)

Correspondence Address:
GOUDREAU GAGE DUBUC
2000 MCGILL COLLEGE, SUITE 2200
MONTREAL, QC H3A 3H3

(73) Assignee: **UNIVERSITE LAVAL**

(21) Appl. No.: **11/529,478**

(22) Filed: **Sep. 29, 2006**

Publication Classification

(51) **Int. Cl.**

C12P 21/06	(2006.01)
C07H 21/04	(2006.01)
C12N 5/06	(2006.01)
C12N 15/86	(2006.01)

(52) **U.S. Cl.** **435/69.1**; 435/348; 435/320.1; 435/456; 530/356; 536/23.5

(57) **ABSTRACT**

Methods, reagents (e.g. vectors) and host cells for the recombinant production of collagen are described, relating to the recombinant expression of a subunit of a collagen or procollagen and a collagen post-translational enzyme or subunit thereof.

Figure 1

DNA (SEQ ID NO: 1) sequence of human α-1(I) collagen (*Homo sapiens* col1-α1: Accession No. NM_000088; GI: 14719826).

```
   1 tcgtcggagc agacgggagt ttctcctcgg ggtcggagca ggaggcacgc ggagtgtgag
  61 gccacgcatg agcggacgct aacccccctcc ccagccacaa agagtctaca tgtctagggt
 121 ctagacatgt tcagctttgt ggacctccgg ctcctgctcc tcttagcggc caccgccctc
 181 ctgacgcacg gccaagagga aggccaagtc gagggccaag acgaagacat cccaccaatc
 241 acctgcgtac agaacggcct caggtaccat gaccgagacg tgtggaaacc cgagccctgc
 301 cggatctgcg tctgcgacaa cggcaaggtg ttgtgcgatg acgtgatctg tgacgagacc
 361 aagaactgcc ccggcgccga agtccccgag ggcgagtgct gtcccgtctg ccccgacggc
 421 tcagagtcac ccaccgacca agaaaccacc ggcgtcgagg gacccaaggg agacactggc
 481 ccccgaggcc caaggggacc cgcaggcccc cctggccgag atggcatccc tggacagcct
 541 ggacttcccg gacccccggg acccccggga cctcccggac cccctggcct cggaggaaac
 601 tttgctcccc agctgtctta tggctatgat gagaaatcaa ccggaggaat ttccgtgcct
 661 ggccccatgg gtcctctgg tcctcgtggt ctccctggcc cccctggtgc acctggtccc
 721 caaggcttcc aaggtccccc tggtgagcct ggcgagcctg gagcttcagg tcccatgggt
 781 ccccgaggtc ccccaggtcc ccctggaaag aatggagatg atggggaagc tggaaaacct
 841 ggtcgtcctg gtgagcgtgg gcctcctggg cctcagggtg ctcgaggatt gcccggaaca
 901 gctggcctcc ctggaatgaa gggacacaga ggtttcagtg gtttggatgg tgccaaggga
 961 gatgctggtc ctgctggtcc taagggtgag cctggcagcc ctggtgaaaa tggagctcct
1021 ggtcagatgg gcccccgtgg cctgcctggt gagagaggtc gccctggagc ccctggccct
1081 gctggtgctc gtggaaatga tggtgctact ggtgctgccg ggcccccctgg tcccaccggc
1141 cccgctggtc ctcctggctt ccctggtgct gttggtgcta agggtgaagc tggtccccaa
1201 gggccccgag gctctgaagg tccccagggt gtgcgtggtg agcctggccc ccctggccct
1261 gctggtgctg ctggccctgc tggaaaccct ggtgctgatg gacagcctgg tgctaaaggt
1321 gccaatggtg ctcctggtat tgctggtgct cctggcttcc ctggtgcccg aggcccctct
1381 ggaccccagg gccccggcgg ccctcctggt cccaagggta acagcggtga acctggtgct
1441 cctggcagca aaggagacac tggtgctaag ggagagcctg gccctgttgg tgttcaagga
1501 ccccctggcc ctgctggaga ggaaggaaag cgaggagctc gaggtgaacc cggacccact
1561 ggcctgcccg gacccctgg cgagcgtggt ggacctggta gccgtggttt ccctggcgca
1621 gatggtgttg ctggtcccaa gggtcccgct ggtgaacgtg gttctcctgg ccctgctggc
1681 cccaaaggat ctcctggtga agctggtcgt cccggtgaag ctggtctgcc tggtgccaag
1741 ggtctgactg gaagccctgg cagccctggt cctgatggca aaactggccc ccctggtccc
1801 gccggtcaag atggtcgccc cggaccccca ggcccacctg gtgcccgtgg tcaggctggt
1861 gtgatgggat tccctggacc taaaggtgct gctggagagc ccggcaaggc tggagagcga
1921 ggtgttcccg gacccccctgg cgctgtcggt cctgctggca aagatggaga ggctggagct
1981 cagggacccc ctggccctgc tggtcccgct ggcgagagag gtgaacaagg ccctgctggc
2041 tcccccggat tccagggtct ccctggtcct gctggtcctc caggtgaagc aggcaaacct
2101 ggtgaacagg gtgttcctgg agaccttggc gcccctggcc cctctggagc aagaggcgag
2161 agaggtttcc ctggcgagcg tggtgtgcaa ggtccccctg gtcctgctgg tccccgaggg
2221 gccaacggtg ctcccggcaa cgatggtgct aagggtgatg ctggtgcccc tggagctccc
2281 ggtagccagg gcgcccctgg ccttcaggga atgcctggtg aacgtggtgc agctggtctt
2341 ccagggccta agggtgacag aggtgatgct ggtcccaaag gtgctgatgg ctctcctggc
2401 aaagatggcg tccgtggtct gactggcccc attggtcctc ctggccctgc tggtgccccc
2461 ggtgacaagg gtgaaagtgg tcccagcggc cctgctggtc ccactggagc tcgtggtgcc
2521 cccggagacc gtggtgagcc tggtcccccc ggccctgctg gctttgctgg ccccccctggt
2581 gctgacggcc aacctggtgc taaaggcgaa cctggtgatg ctggtgctaa aggcgatgct
2641 ggtcccccctg gccctgccgg acccgctgga ccccctggcc ccattggtaa tgttggtgct
2701 cctggagcca aaggtgctcg cggcagcgct ggtcccccctg gtgctactgg tttccctggt
2761 gctgctggcc gagtcggtcc tcctggcccc tctggaaatg ctggaccccc tggccctcct
2821 ggtcctgctg gcaaagaagg cggcaaaggt ccccgtggtg agactggccc tgctggacgt
2881 cctggtgaag ttggtccccc tggtcccctt ggccctgctg gcgagaaagg atcccctggt
```

Figure 1 (continued)

```
2941 gctgatggtc ctgctggtgc tcctggtact cccgggcctc aaggtattgc tggacagcgt
3001 ggtgtggtcg gcctgcctgg tcagagagga gagagaggct tccctggtct tcctggcccc
3061 tctggtgaac ctggcaaaca aggtccctct ggagcaagtg gtgaacgtgg tccccctggt
3121 cccatgggcc cccctggatt ggctggaccc cctggtgaat ctggacgtga gggggctcct
3181 ggtgccgaag gttcccctgg acgagacggt tctcctggcg ccaagggtga ccgtggtgag
3241 accggccccg ctggaccccc tggtgctcct ggtgctcctg gtgcccctgg ccccgttggc
3301 cctgctggca agagtggtga tcgtggtgag actggtcctg ctggtcccgc cggtcctgtc
3361 ggccctgttg gcgcccgtgg ccccgccgga ccccaaggcc cccgtggtga caagggtgag
3421 acaggcgaac agggcgacag aggcataaag ggtcaccgtg gcttctctgg cctccagggt
3481 ccccctggcc ctcctggctc tcctggtgaa caaggtccct ctggagcctc tggtcctgct
3541 ggtccccgag gtcccctgg ctctgctggt gctcctggca aagatggact caacggtctc
3601 cctggcccca ttgggccccc tggtcctcgc ggtcgcactg gtgatgctgg tcctgttggt
3661 ccccccggcc ctcctggacc tcctggtccc cctggtcctc ccagcgctgg tttcgacttc
3721 agcttcctgc cccagccacc tcaagagaag gctcacgatg gtggccgcta ctaccgggct
3781 gatgatgcca atgtggttcg tgaccgtgac ctcgaggtgg acaccaccct caagagcctg
3841 agccagcaga tcgagaacat ccggagccca gagggcagcc gcaagaaccc cgcccgcacc
3901 tgccgtgacc tcaagatgtg ccactctgac tggaagagtg gagagtactg gattgacccc
3961 aaccaaggct gcaacctgga tgccatcaaa gtcttctgca acatggagac tggtgagacc
4021 tgcgtgtacc ccactcagcc cagtgtggcc cagaagaact ggtacatcag caagaacccc
4081 aaggacaaga ggcatgtctg gttcggcgag agcatgaccg atggattcca gttcgagtat
4141 ggcggccagg gctccgaccc tgccgatgtg gccatccagc tgaccttcct gcgcctgatg
4201 tccaccgagg cctcccagaa catcacctac cactgcaaga acagcgtggc ctacatggac
4261 cagcagactg gcaacctcaa gaaggccctg ctcctccagg gctccaacga gatcgagatc
4321 cgcgccgagg gcaacagccg cttcacctac agcgtcactg tcgatggctg cacgagtcac
4381 accggagcct ggggcaagac agtgattgaa tacaaaacca ccaagacctc ccgcctgccc
4441 atcatcgatg tggcccccct ggacgttggt gccccagacc aggaattcgg cttcgacgtt
4501 ggccctgtct gcttcctgta aactccctcc atcccaacct ggctccctcc cacccaacca
4561 actttccccc caacccggaa acagacaagc aacccaaact gaacccccctc aaaagccaaa
4621 aaatgggaga caatttcaca tggactttgg aaaatatttt tttcctttgc attcatctct
4681 caaacttagt ttttatcttt gaccaaccga acatgaccaa aaaccaaaag tgcattcaac
4741 cttaccaaaa aaaaaaaaaa aaaaagaata aataaataac ttttttaaaaa aggaagcttg
4801 gtccacttgc ttgaagaccc atgcgggggt aagtcccttt ctgcccgttg ggcttatgaa
4861 accccaatgc tgccctttct gctcctttct ccacaccccc cttggggcct cccctccact
4921 ccttcccaaa tctgtctccc cagaagacac aggaaacaat gtattgtctg cccagcaatc
4981 aaaggcaatg ctcaaacacc caagtggccc ccaccctcag cccgctcctg cccgcccagc
5041 acccccaggc cctgggggac ctggggttct cagactgcca aagaagcctt gccatctggc
5101 gctcccatgg ctcttgcaac atctcccctt cgtttttgag ggggtcatgc cgggggagcc
5161 accagccccc cactgggttc ggaggagagt caggaagggc cacgacaaag cagaaacatc
5221 ggatttgggg aacgcgtgtc aatcccttgt gccgcagggc tgggcgggag agactgttct
5281 gttccttgtg taactgtgtt gctgaaagac tacctcgttc ttgtcttgat gtgtcaccgg
5341 ggcaactgcc tgggggcggg gatgggggca gggtggaagc ggctccccat tttataccaa
5401 aggtgctaca tctatgtgat gggtggggtg gggagggaat cactggtgct atagaaattg
5461 agatgccccc ccaggccagc aaatgttcct ttttgttcaa agtctatttt tattccttga
5521 tatttttctt ttttttttt ttttttgtg gatggggact tgtgaatttt tctaaaggtg
5581 ctatttaaca tgggaggaga gcgtgtgcgg ctccagccca gcccgctgct cactttccac
5641 cctctctcca cctgcctctg gcttctcagg cctctgctct ccgacctctc tcctctgaaa
5701 ccctcctcca cagctgcagc ccatcctccc ggctccctcc tagtctgtcc tgcgtcctct
5761 gtccccgggt ttcagagaca acttcccaaa gcacaaagca gtttttcccc ctaggggtgg
5821 gaggaagcaa aagactctgt acctattttg tatgtgtata ataatttgag atgtttttaa
5881 ttattttgat tgctggaata aagcatgtgg aaatgaccca aacataa
```

Figure 1 (continued)

Polypeptide (SEQ ID NO: 2) sequence of human α-1(I) collagen (*Homo sapiens* col1-α1: Accession No. NM_000088; GI: 14719826).

MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNGLRYHDRDVWKPEPCRICV
CDNGKVLCDDVICDETKNCPGAEVPEGECCPVCPDGSESPTDQETTGVEGPKGDTGPRGPRG
PAGPPGRDGIPGQPGLPGPPGPPGPPGPPGLGGNFAPQLSYGYDEKSTGGISVPGPMGPSGP
RGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPG
PQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLPGER
GRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGPRGSEGPQGVRGE
PGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSGPQGPGGPPGPKGNSG
EPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFP
GADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGP
AGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQG
PPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFP
GERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGD
RGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGARGAPGDRGEPG
PPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVGAPGAKGARGSA
GPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGEVGPPGPPGP
AGEKGSPGADGPAGAPGTPGPQGIAGQRGVVGLPGQRGERGFPGLPGPSGEPGKQGPSGASG
ERGPPGPMGPPGLAGPPGESGREGAPAAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAP
GPVGPAGKSGDRGETGPAGPAGPVGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGL
QGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVG
PPGPPGPPGPPGPPSAGFDFSFLPQPPQEKAHDGGRYYRADDANVVRDRDLEVDTTLKSLSQ
QIENIRSPEGSRKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYP
TQPSVAQKNWYISKNPKDKRHVWFGESMTDGFQFEYGGQGSDPADVAIQLTFLRLMSTEASQ
NITYHCKNSVAYMDQQTGNLKKALLLKGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTV
IEYKTTKSSRLPIIDVAPLDVGAPDQEFGFDVGPVCFL

Figure 2

DNA (SEQ ID NO: 3) and polypeptide (SEQ ID NO: 4) sequences of human α-2(I) collagen (*Homo sapiens* col1-α2: Accession No. NM_000089, GI: 48762933), including the 5086 bp sequence coding for pre-pro- col1-α2 set forth in Accession No. Z74616, GI: 1418929.

```
   1 gtgtcccata gtgtttccaa acttggaaag ggcgggggag ggcgggagga tgcggagggc
  61 ggaggtatgc agacaacgag tcagagtttc cccttgaaag cctcaaaagt gtccacgtcc
 121 tcaaaaagaa tggaaccaat ttaagaagcc agccccgtgg ccacgtcccc tcccccattc
 181 gctccctcct ctgcgccccc gcaggctcct cccagctgtg gctgcccggg cccccagccc
 241 cagccctccc attggtggag gccctttgg aggcacccta gggccaggga aacttttgcc
 301 gtataaatag ggcagatccg ggctttatta ttttagcacc acggcagcag gaggtttcgg
 361 ctaagttgga ggtactggcc acgactgcat gcccgcgccc gccaggtgat acctccgccg
 421 gtgacccagg ggctctgcga cacaaggagt ctgcatgtct aagtgctaga catgctcagc
 481 tttgtggata cgcggacttt gttgctgctt gcagtaacct tatgcctagc aacatgccaa
 541 tctttacaag aggaaactgt aagaaagggc ccagccggag atagaggacc acgtggagaa
 601 aggggtccac caggcccccc aggcagagat ggtgaagatg gtcccacagg ccctcctggt
 661 ccacctggtc ctcctggccc ccctggtctc ggtgggaact ttgctgctca gtatgatgga
 721 aaaggagttg gacttggccc tggaccaatg ggcttaatgg gacctagagg cccacctggt
 781 gcagctggag ccccaggccc tcaaggtttc caaggacctg ctggtgagcc tggtgaacct
 841 ggtcaaactg gtcctgcagg tgctcgtggt ccagctggcc ctcctggcaa ggctggtgaa
 901 gatggtcacc ctggaaaacc cggacgacct ggtgagagag gagttgttgg accacagggt
 961 gctcgtggtt tccctggaac tcctggactt cctggcttca aaggcattag gggacacaat
1021 ggtctggatg gattgaaggg acagcccggt gctcctggtg tgaagggtga acctggtgcc
1081 cctggtgaaa atggaactcc aggtcaaaca ggagcccgtg ggcttcctgg tgagagagga
1141 cgtgttggtg cccctggccc agctggtgcc cgtggcagtg atggaagtgt gggtcccgtg
1201 ggtcctgctg gtcccattgg gtctgctggc cctccaggct tcccaggtgc ccctggcccc
1261 aagggtgaaa ttggagctgt tggtaacgct ggtcctgctg gtcccgccgg tccccgtggt
1321 gaagtgggtc ttccaggcct ctccggcccc gttggacctc ctggtaatcc tggagcaaac
1381 ggccttactg gtgccaaggg tgctgctggc cttcccggcg ttgctggggc tcccggcctc
1441 cctggacccc gcggtattcc tggccctgtt ggtgctgccg gtgctactgg tgccagagga
1501 cttgttggtg agcctggtcc agctggctcc aaaggagaga gcggtaacaa gggtgagccc
1561 ggctctgctg ggccccaagg tcctcctggt cccagtggtg aagaaggaaa gagaggccct
1621 aatggggaag ctggatctgc cggccctcca ggacctcctg ggctgagagg tagtcctggt
1681 tctcgtggtc ttcctggagc tgatggcaga gctggcgtca tgggccctcc tggtagtcgt
1741 ggtgcaagtg gccctgctgg agtccgagga cctaatggag atgctggtcg ccctggggag
1801 cctggtctca tgggacccag aggtcttcct ggttcccctg gaaatatcgg ccccgctgga
1861 aaagaaggtc ctgtcggcct ccctggcatc gacggcaggc ctggcccaat tggcccagct
1921 ggagcaagag gagagcctgg caacattgga ttccctggac ccaaaggccc cactggtgat
1981 cctggcaaaa acggtgataa aggtcatgct ggtcttgctg gtgctcgggg tgctccaggt
2041 cctgatggaa acaatggtgc tcagggacct cctggaccac agggtgttca aggtggaaaa
2101 ggtgaacagg gtccccctgg tcctccaggc ttccagggtc tgcctggccc ctcaggtccc
2161 gctggtgaag ttggcaaacc aggagaaagg ggtctccatg gtgagtttgg tctccctggt
2221 cctgctggtc caagagggga acgcggtccc ccaggtgaga gtggtgctgc cggtcctact
2281 ggtcctattg gaagccgagg tccttctgga cccccagggc ctgatggaaa caagggtgaa
2341 cctggtgtgg ttggtgctgt gggcactgct ggtccatctg gtcctagtgg actcccagga
2401 gagaggggtg ctgctggcat acctggaggc aagggagaaa agggtgaacc tggtctcaga
2461 ggtgaaattg gtaaccctgg cagagatggt gctcgtggtg ctcctggtgc tgtaggtgcc
2521 cctggtcctg ctggagccac aggtgaccgg ggcgaagctg gggctgctgg tcctgctggt
2581 cctgctggtc ctcggggaag ccctggtgaa cgtggtgagg tcggtcctgc tggccccaat
2641 ggatttgctg gtcctgctgg tgctgctggt caacctggtg ctaaaggaga aagaggagcc
2701 aaagggccta agggtgaaaa cggtgttgtt ggtcccacag gccccgttgg agctgctggc
```

Figure 2 (continued)

2761 ccagctggtc caaatggtcc ccccggtcct gctggaagtc gtggtgatgg aggcccccct
2821 ggtatgactg gtttccctgg tgctgctgga cggactggtc ccccaggacc ctctggtatt
2881 tctggccctc ctggtccccc tggtcctgct gggaaagaag ggcttcgtgg tcctcgtggt
2941 gaccaaggtc cagttggccg aactggagaa gtaggtgcag ttggtccccc tggcttcgct
3001 ggtgagaagg gtccctctgg agaggctggt actgctggac ctcctggcac tccaggtcct
3061 cagggtcttc ttggtgctcc tggtattctg ggtctccctg gctcgagagg tgaacgtggt
3121 ctaccaggtg ttgctggtgc tgtgggtgaa cctggtcctc ttggcattgc cggccctcct
3181 ggggcccgtg gtcctcctgg tgctgtgggt agtcctggag tcaacggtgc tcctggtgaa
3241 gctggtcgtg atggcaaccc tgggaacgat ggtcccccag gtcgcgatgg tcaacccgga
3301 cacaagggag agcgcggtta ccctggcaat attggtcccg ttggtgctgc aggtgcacct
3361 ggtcctcatg gccccgtggg tcctgctggc aaacatggaa accgtggtga aactggtcct
3421 tctggtcctg ttggtcctgc tggtgctgtt ggcccaagag gtcctagtgg cccacaaggc
3481 attcgtggcg ataagggaga gcccggtgaa aaggggccca gaggtcttcc tggcttaaag
3541 ggacacaatg gattgcaagg tctgcctggt atcgctggtc accatggtga tcaaggtgct
3601 cctggctccg tgggtcctgc tggtcctagg ggccctgctg gtccttctgg ccctgctgga
3661 aaagatggtc gcactggaca tcctggtaca gttggacctg ctggcattcg aggccctcag
3721 ggtcaccaag gccctgctgg ccccccctggt ccccctggcc ctcctggacc tccaggtgta
3781 agcggtggtg gttatgactt tggttacgat ggagacttct acagggctga ccagcctcgc
3841 tcagcacctt ctctcagacc caaggactat gaagttgatg ctactctgaa gtctctcaac
3901 aaccagattg agacccttct tactcctgaa ggctctagaa agaacccagc tcgcacatgc
3961 cgtgacttga gactcagcca cccagagtgg agcagtggtt actactggat tgaccctaac
4021 caaggatgca ctatggatgc tatcaaagta tactgtgatt tctctactgg cgaaacctgt
4081 atccgggccc aacctgaaaa catcccagcc aagaactggt ataggagctc caaggacaag
4141 aaacacgtct ggctaggaga aactatcaat gctggcagcc agtttgaata taatgtagaa
4201 ggagtgactt ccaaggaaat ggctacccaa cttgccttca tgcgcctgct ggccaactat
4261 gcctctcaga acatcaccta ccactgcaag aacagcattg catacatgga tgaggagact
4321 ggcaacctga aaaaggctgt cattctacag ggctctaatg atgttgaact tgttgctgag
4381 ggcaacagca ggttcactta cactgttctt gtagatggct gctctaaaaa gacaaatgaa
4441 tggggaaaga caatcattga atacaaaaca aataagccat cacgcctgcc cttccttgat
4501 attgcacctt tggacatcgg tggtgctgac caggaattct ttgtggacat tggcccagtc
4561 tgtttcaaat aaatgaactc aatctaaatt aaaaaagaaa gaaatttgaa aaaactttct
4621 ctttgccatt tcttcttctt ctttttaac tgaaagctga atccttccat ttcttctgca
4681 catctacttg cttaaattgt gggcaaaaga gaaaaagaag gattgatcag agcattgtgc
4741 aatacagttt cattaactcc ttccccccgct cccccaaaaa tttgaatttt tttttcaaca
4801 ctcttacacc tgttatggaa aatgtcaacc tttgtaagaa aaccaaaata aaaattgaaa
4861 aataaaaacc ataaacattt gcaccacttg tggcttttga atatcttcca cagagggaag
4921 tttaaaaccc aaacttccaa aggtttaaac tacctcaaaa cactttccca tgagtgtgat
4981 ccacattgtt aggtgctgac ctagacagag atgaactgag gtccttgttt tgttttgttc
5041 ataatacaaa ggtgctaatt aatagtattt cagatacttg aagaatgttg atggtgctag
5101 aagaatttga gaagaaatac tcctgtattg agttgtatcg tgtggtgtat tttttaaaaa
5161 atttgattta gcattcatat tttccatctt attcccaatt aaaagtatgc agattatttg
5221 cccaaatctt cttcagattc agcatttgtt ctttgccagt ctcattttca tcttcttcca
5281 tggttccaca gaagctttgt ttcttgggca agcagaaaaa ttaaattgta cctattttgt
5341 atatgtgaga tgtttaaata aattgtgaaa aaaatgaaat aaagcatgtt tggttttcca
5401 aaagaacata t

Figure 2 (continued)

Polypeptide (SEQ ID NO: 4) sequence of human α-2(I) collagen (*Homo sapiens* col1-α2: Accession No. NM_000089, GI: 48762933; coding sequence corresponding to positions 472-4572 of DNA sequence above), including the 5086 bp sequence coding for pre-pro- col1-α2 set forth in Accession No. Z74616, GI: 1418929.

MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGERGPPGPPGRDGED
GPTGPPGPPGPPGPPGLGGNFAAQYDGKGVGLGPGPMGLMGPRGPPGAAGAPGPQG
FQGPAGEPGEPGQTGPAGARGPAGPPGKAGEDGHPGKPGRPGERGVVGPQGARGFP
GTPGLPGFKGIRGHNGLDGLKGQPGAPGVKGEPGAPGENGTPGQTGARGLPGERGR
VGAPGPAGARGSDGSVGPVGPAGPIGSAGPPGFPGAPGPKGEIGAVGNAGPAGPAG
PRGEVGLPGLSGPVGPPGNPGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVGAA
GATGARGLVGEPGPAGSKGESGNKGEPGSAGPQGPPGPSGEEGKRGPNGEAGSAGP
PGPPGLRGSPGSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPGEPGLMG
PRGLPGSPGNIGPAGKEGPVGLPGIDGRPGPIGPAGARGEPGNIGFPGPKGPTGDP
GKNGDKGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKGEQGPPGPPGFQGLPGP
SGPAGEVGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPG
PDGNKGEPGVVGAVGTAGPSGPSGLPGERGAAGIPGGKGEKGEPGLRGEIGNPGRD
GARGAPGAVGAPGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGP
AGAAGQPGAKGERGAKGPKGENGVVGPTGPVGAAGPAGPNGPPGPAGSRGDGGPPG
MTGFPGAAGRTGPPGPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTGEVGAVGPP
GFAGEKGPSGEAGTAGPPGTPGPQGLLGAPGILGLPGSRGERGLPGVAGAVGEPGP
LGIAGPPGARGPPGAVGSPGVNGAPGEAGRDGNPGNDGPPGRDGQPGHKGERGYPG
NIGPVGAAGAPGPHGPVGPAGKHGNRGETGPSGPVGPAGAVGPRGPSGPQGIRGDK
GEPGEKGPRGLPGLKGHNGLQGLPGIAGHHGDQGAPGSVGPAGPRGPAGPSGPAGK
DGRTGHPGTVGPAGIRGPQGHQGPAGPPGPPGPPGPPGVSGGGYDFGYDGDFYRAD
QPRSAPSLRPKDYEVDATLKSLNNQIETLLTPEGSRKNPARTCRDLRLSHPEWSSG
YYWIDPNQGCTMDAIKVYCDFSTGETCIRAQPENIPAKNWYRSSKDKKHVWLGETI
NAGSQFEYNVEGVTSKEMATQLAFMRLLANYASQNITYHCKNSIAYMDEETGNLKK
AVILQGSNDVELVAEGNSRFTYTVLVDGCSKKTNEWGKTIIEYKTNKPSRLPFLDI
APLDIGGADQEFFVDIGPVCFK

Figure 3

DNA (SEQ ID NO: 5) sequence of human prolyl 4-hydroxylase alpha subunit (α-p4H; *Homo sapiens* p4Hα: 2722 bp sequence set forth in Accession No. M24486 (clone PA-II), GI: 190785).

```
   1 gagcgggctg agggtaggaa gtagccgctc cgagtggagg cgactggggg ctgaagagcg
  61 cgccgccctc tcgtcccact ttccaggtgt gtgatcctgt aaaattaaat cttccaagat
 121 gatctggtat atattaatta taggaattct gcttccccag tctttggctc atccaggctt
 181 ttttacttca attggtcaga tgactgattt gatccatact gagaaagatc tggtgacttc
 241 tctgaaagat tatattaagg cagaagagga caagttagaa caaataaaaa aatgggcaga
 301 gaagttagat cggctaacta gtacagcgac aaaagatcca gaaggatttg ttgggcatcc
 361 agtaaatgca ttcaaattaa tgaaacgtct gaatactgag tggagtgagt tggagaatct
 421 ggtccttaag gatatgtcag atggctttat ctctaaccta accattcaga gaccagtact
 481 ttctaatgat gaagatcagg ttggggcagc caaagctctg ttacgtctcc aggataccta
 541 caatttggat acagatacca tctcaaaggg taatcttcca ggagtgaaac acaaatcttt
 601 tctaacggct gaggactgct ttgagttggg caaagtggcc tatacagaag cagattatta
 661 ccatacggaa ctgtggatgg aacaagccct aaggcaactg gatgaaggcg agatttctac
 721 catagataaa gtctctgttc tagattattt gagctatgcg gtatatcagc agggagacct
 781 ggataaggca cttttgctca caaagaagct tcttgaacta gatcctgaac atcagagagc
 841 taatggtaac ttaaaatatt ttgagtatat aatggctaaa gaaaaagatg tcaataagtc
 901 tgcttcagat gaccaatctg atcagaaaac tacaccaaag aaaaaagggg ttgctgtgga
 961 ttacctgcca gagagacaga agtacgaaat gctgtgccgt ggggagggta tcaaaatgac
1021 ccctcggaga cagaaaaaac tcttttgccg ctaccatgat ggaaaccgta atcctaaatt
1081 tattctggct ccagctaaac aggaggatga atgggacaag cctcgtatta ttcgcttcca
1141 tgatattatt tctgatgcag aaattgaaat cgtcaaagac ctagcaaaac caaggctgag
1201 ccgagctaca gtacatgacc ctgagactgg aaaattgacc acagcacagt acagagtatc
1261 taagagtgcc tggctctctg gctatgaaaa tcctgtggtg tctcgaatta atatgagaat
1321 acaagatcta acaggactag atgtttccac agcagaggaa ttacaggtag caaattatgg
1381 agttggagga cagtatgaac cccattttga ctttgcacgg aaagatgagc cagatgcttt
1441 caaagagctg gggacaggaa atagaattgc tacatggctg ttttatatga gtgatgtgtc
1501 tgcaggagga gccactgttt ttcctgaagt tggagctagt gtttggccca aaaaaggaac
1561 tgctgttttc tggtataatc tgtttgccag tggagaagga gattatagta cacggcatgc
1621 agcctgtcca gtgctagttg gcaacaaatg ggtatccaat aaatggctcc atgaacgtgg
1681 acaagaattt cgaagacctt gtacgttgtc agaattggaa tgacaaacag gcttcccttt
1741 ttctcctatt gttgtactct tatgtgtctg atatacacat ttccatagtc ttaactttca
1801 ggagtttaca attgactaac actccatgat tgattcagtc atgaacctca tcccatgttt
1861 catctgtgga caattgctta ctttgtgggt tcttttaaaa gtaacacgaa atcatcatat
1921 tgcataaaac cttaaagttc tgttggtatc acagaagaca aggcagagtt taaagtgagg
1981 aattttatat ttaaagaact ttttggttgg ataaaaacat aatttgagca tccagtttta
2041 gtatttcact acatctcagt tggtgggtgt taagctagaa tgggctgtgt gataggaaac
2101 aaatgcctta cagatgtgcc taggtgttct gtttacctag tgtcttactc tgttttctgg
2161 atctgaagac tagtaataaa ctaggacact aactgggttc catgtgattg ccctttcata
2221 tgatcttcta agttgatttt tttcctccca agtctttttt aaagaaagta tactgtattt
2281 taccaacccc ctctcttttc ttttagctcc tctgtggtga attaaacgta cttgagttaa
2341 aatatttcga ttttttttt ttttttaatg gaaagtcctg cataacaaca ctgggccttc
2401 ttaactaaaa tgctcaccac ttagcctgtt tttttatccc ttttttaaaa tgacagatga
2461 ttttgttcag gaattttgct gtttttctta gtgctaatac cttgcctctt attcctgcta
2521 cagcagggtg gtaatattgg cattctgatt aaatactgtg ccttaggaga ctggaagttt
2581 aaaaatgtac aagtcctttc agtgatgagg gaattgattt tttttaaaag tctttttctt
2641 agaaagccaa aatgtttgtt tttttaagat tctgaaatgt gttgtgacaa caatgaccta
2701 tttatgatct taaatctttt tt
```

Figure 3 (continued)

Polypeptide (SEQ ID NO: 6) sequence of human prolyl 4-hydroxylase alpha subunit (α-p4H; *Homo sapiens* p4Hα: 2722 bp sequence set forth in Accession No. M24486 (clone PA-II), GI: 190785).

MIWYILIIGILLPQSLAHPGFFTSIGQMTDLIHTEKDLVTSLKDYIKAEEDKLEQIKKWAEKLD
RLTSTATKDPEGFVGHPVNAFKLMKRLNTEWSELENLVLKDMSDGFISNLTIQRPVLSNDEDQV
GAAKALLRLQDTYNLDTDTISKGNLPGVKHKSFLTAEDCFELGKVAYTEADYYHTELWMEQALR
QLDEGEISTIDKVSVLDYLSYAVYQQGDLDKALLLTKKLLELDPEHQRANGNLKYFEYIMAKEK
DVNKSASDDQSDQKTTPKKKGVAVDYLPERQKYEMLCRGEGIKMTPRRQKKLFCRYHDGNRNPK
FILAPAKQEDEWDKPRIIRFHDIISDAEIEIVKDLAKPRLSRATVHDPETGKLTTAQYRVSKSA
WLSGYENPVVSRINMRIQDLTGLDVSTAEELQVANYGVGGQYEPHFDFARKDEPDAFKELGTGN
RIATWLFYMSDVSAGGATVFPEVGASVWPKKGTAVFWYNLFASGEGDYSTRHAACPVLVGNKWV
SNKWLHERGQEFRRPCTLSELE

Figure 4

DNA (SEQ ID NO: 7) sequence of human prolyl 4-hydroxylase beta subunit (β-p4H; *Homo sapiens* p4Hβ: 1956 bp sequence set forth in Accession No. X05130, GI: 35654).

```
   1 ccgagcgccc cgcctgctcc gtgtccgaca tgctgcgccg cgctctgctg tgcctgccgt
  61 ggnccgccct ggtgcgcgcc gacgcccccg aggaggagga ccacgtcttg gtgctgcgga
 121 aaagcaactt cgcggaggcg ctggcggccc acaagtaccc gccggtggag ttccatgccc
 181 cctggtgtgg ccactgcaag gctctggccc ctgagtatgc caaagccgct gggaagctga
 241 aggcagaagg ttccgagatc aggttggcca aggtggacgc cacggaggag tctgacctag
 301 cccagcagta cggcgtgcgc ggctatccca ccatcaagtt cttcaggaat ggagacacgg
 361 cttcccccaa ggaatataca gctggcagag aggctgatga catcgtgaac tggctgaaga
 421 agcgcacggg cccggctgcc accaccctgc ctgacggcgc agctgcagag tccttggtgg
 481 agtccagcga ggtggccgtc atcggcttct tcaaggacgt ggagtcggac tctgccaagc
 541 agtttttgca ggcagcagag gccatcgatg acataccatt tgggatcact tccaacagtg
 601 acgtgttctc caaataccag ctcgacaaag atggggttgt cctctttaag aagtttgatg
 661 aaggccggaa caactttgaa ggggaggtca ccaaggagaa cctgctggac tttatcaaac
 721 acaaccagct gccccttgtc atcgagttca ccgagcagac agccccgaag atttttggag
 781 gtgaaatcaa gactcacatc ctgctgttct tgcccaagag tgtgtctgac tatgacggca
 841 aactgagcaa cttcaaaaca gcagccgaga gcttcaaggg caagatcctg ttcatcttca
 901 tcgacagcga ccacaccgac aaccagcgca tcctcgagtt ctttggcctg aagaaggaag
 961 agtgcccggc cgtgcgcctc atcaccttgg aggaggagat gaccaagtac aagcccgaat
1021 cggaggagct gacggcagag aggatcacag agttctgcca ccgcttcctg gagggcaaaa
1081 tcaagcccca cctgatgagc caggagctgc cggaggactg ggacaagcag cctgtcaagg
1141 tgcttgttgg gaagaacttt gaagacgtgg cttttgatga gaaaaaaaac gtctttgtgg
1201 agttctatgc cccatggtgt ggtcactgca aacagttggc tcccatttgg gataaactgg
1261 gagagacgta caaggaccat gagaacatcg tcatcgccaa gatggactcg actgccaacg
1321 aggtggaggc cgtcaaagtg cacggcttcc ccacactcgg gttctttcct gccagtgccg
1381 acaggacggt cattgattac aacggggaac gcacgctgga tggttttaag aaattcctag
1441 agagcggtgg ccaagatggg gcaggggatg ttgacgacct cgaggacctc gaagaagcag
1501 aggagccaga catggaggaa gacgatgacc agaaagctgt gaaagatgaa ctgtaatacg
1561 caaagccgga cccgggcgct gccgagaccc ctcgggggct gcacacccag cagcagcgca
1621 cgcctccgaa gcctgcggcc tcgcttgaag gagggcgtcg ccggaaaccc aaggaacctc
1681 tctgaagtga cacctcaccc ctacacaccg tccgttcacc cccgtctctt ccttctgctt
1741 ttcggttttt ggaaaacccg gatcctactc taggcagccc accttggtgg gcttgtttcc
1801 tgaaaccatg atgtactttt tcatacatga gtctgtccag agtgcttgct accgtgttcg
1861 gagtctcgct gcctccctcc cgcgggaggt tctcctcttt ttgaaaattc cgtccgtggg
1921 atttttagac atttttacga catcagggta tttgtt
```

Figure 4 (continued)

Polypeptide (SEQ ID NO: 8) sequence of human prolyl 4-hydroxylase beta subunit (β-p4H; *Homo sapiens* p4Hβ: 1956 bp sequence set forth in Accession No. X05130, GI: 35654).

MLRRALLCLPWXALVRADAPEEEDHVLVLRKSNFAEALAAHKYPPVEFHAPWCGHCKALAPEYA
KAAGKLKAEGSEIRLAKVDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYTAGREADDIVNW
LKKRTGPAATTLPDGAAAESLVESSEVAVIGFFKDVESDSAKQFLQAAEAIDDIPFGITSNSDV
FSKYQLDKDGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPLVIEFTEQTAPKIFGGEIKTH
ILLFLPKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEECPAVRLITL
EEEMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQELPEDWDKQPVKVLVGKNFEDVAFDE
KKNVFVEFYAPWCGHCKQLAPIWDKLGETYKDHENIVIAKMDSTANEVEAVKVHGFPTLGFFPA
SADRTVIDYNGERTLDGFKKFLESGGQDGAGDVDDLEDLEEAEEPDMEEDDDQKAVKDEL

Figure 5

a)

Autographa californica multicapsid nuclear polyhydrosis virus (AcMNPV)p10 promoter region (SEQ ID NO: 9; derived from Genbank Accession NC_001623 and technical materials for plasmid pBAC4x-1™ [Novagen]):

acaatatattatagttaaataagaattattatcaaatcatttgtatattaattaaaatactatactgtaaattac
attttatttacaatc

b)

Autographa californica multicapsid nuclear polyhydrosis virus (AcMNPV)polH promoter region (SEQ ID NO: 10; derived from Genbank Accession NC_001623 and technical materials for plasmid pBAC4x-1™ [Novagen]):

ataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaat

Figure 6

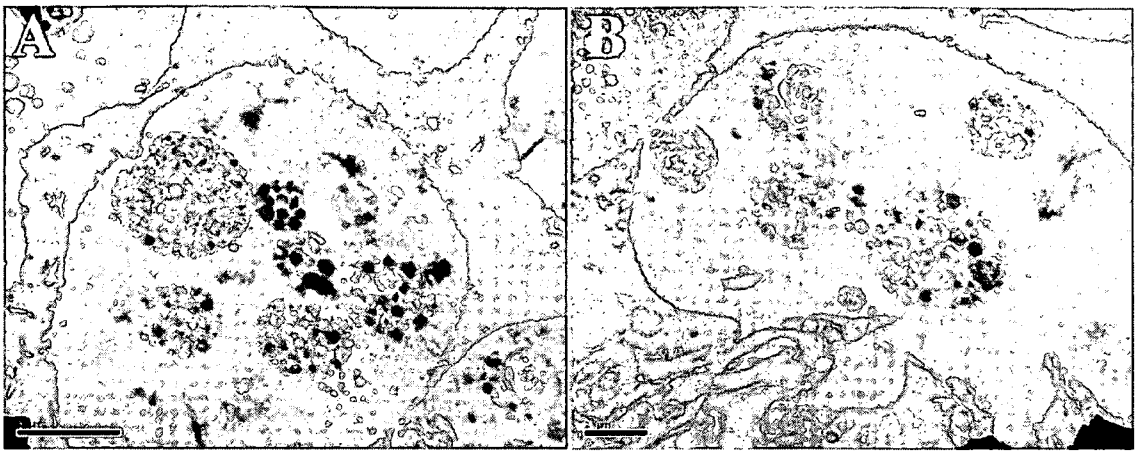


Figure 7

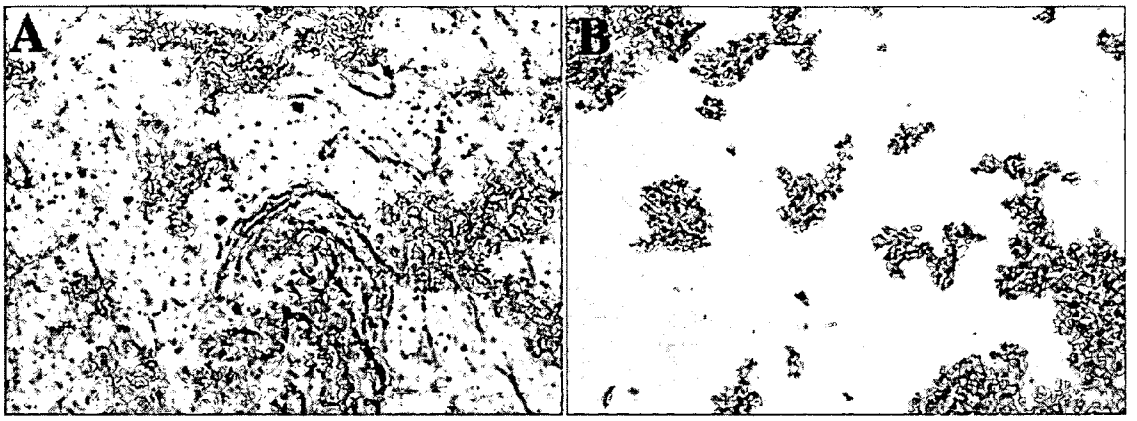


Figure 8

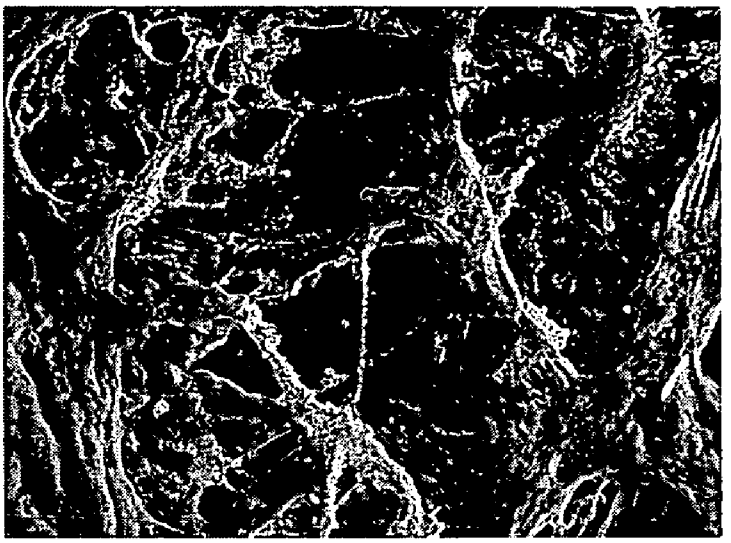


Figure 9

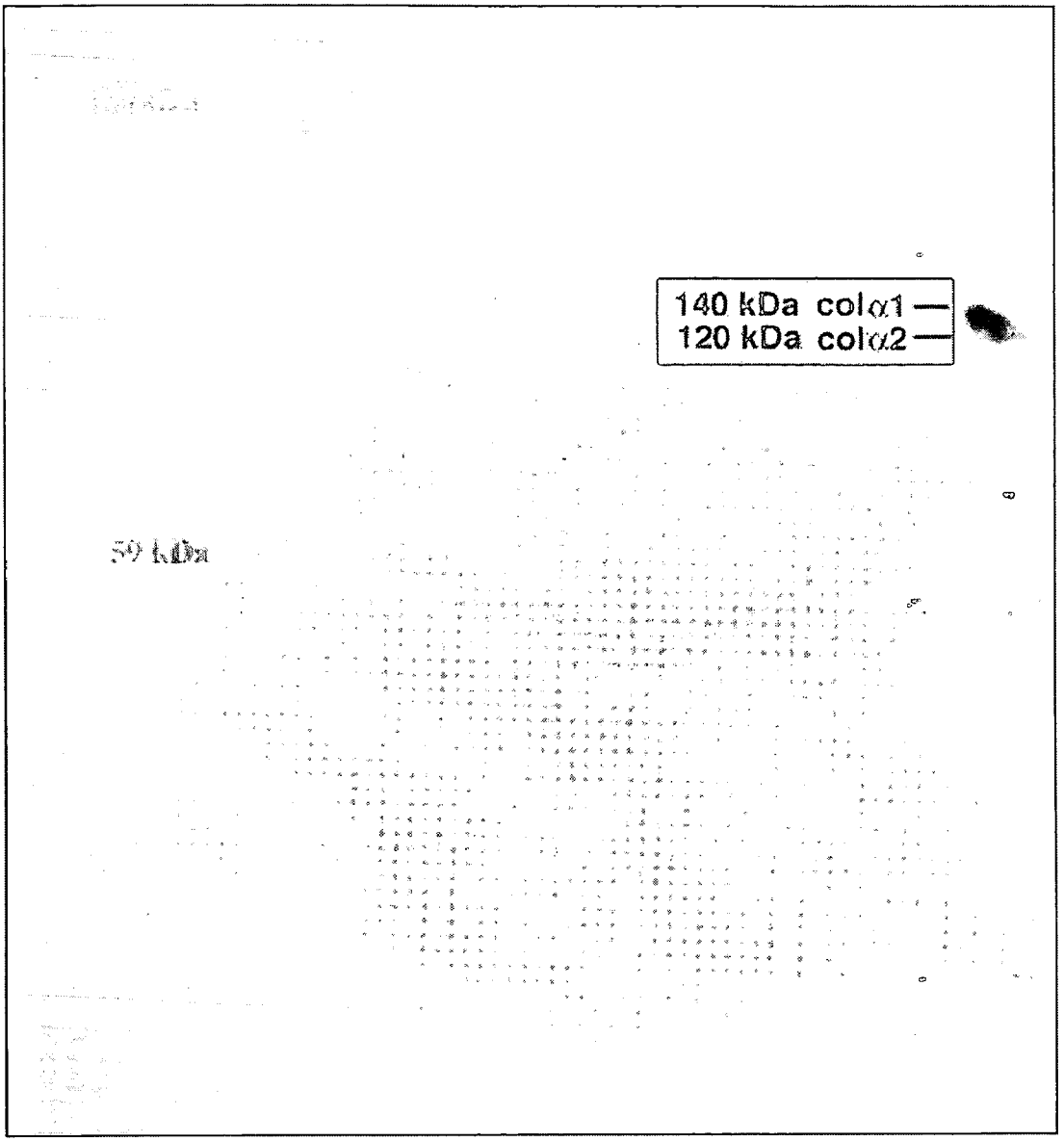
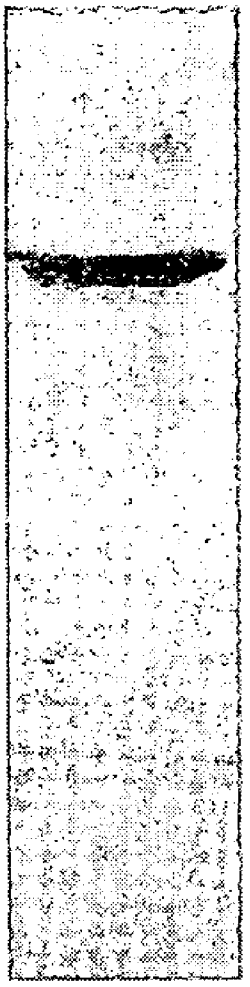


Figure 10



PRODUCTION OF RECOMBINANT HUMAN COLLAGEN

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 60/721,921 filed Sep. 30, 2005, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates to the production of polypeptides using recombinant DNA systems. More specifically, the present invention is concerned with the production of human collagen using such systems.

BACKGROUND OF THE INVENTION

[0003] Collagen is the most abundant component of the extracellular matrix, and is generally formed by the assembly of three polypeptide chains to create a trimeric structure. Nineteen different types of collagen have been described, numbered as types I-XIX. For example, type I collagen, found in several tissues such as bone, tendons and skin, is a heterotrimeric molecule comprising two α-1 (I) chains and one α-2(I) chain. Type II collagen is a homotrimeric molecule comprising three α-1(II) chains. Type III collagen, found in skin and vascular tissues, is a homotrimeric molecule comprising three α-1(III) chains.

[0004] Various post-translational modifications during collagen biosynthesis have been described, including, for example, the hydroxylation of proline residues to 4-hydroxyproline by the enzyme prolyl 4-hydroxylase, and cleavage of N- and C-propeptides of procollagen by N- and C-proteinase enzymes, respectively. Other reported collagen post-translational enzymes include lysyl oxidase and lysyl hydroxylase.

[0005] Human collagen is desirable for a number of therapeutic applications. Its production by recombinant means is attractive for example to obtain greater amounts where insufficient amounts are available from natural sources, as well as to avoid any adverse immune reactions associated with the use of collagen from non-human sources.

[0006] There is therefore a continued need to develop systems for the recombinant production of collagen.

SUMMARY OF THE INVENTION

[0007] The invention relates to the recombinant production of collagen.

[0008] More specifically, in accordance with the present invention, there is provided a method for producing a recombinant human collagen polypeptide, said method comprising: (a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter different from said first promoter; and (b) recovering said collagen polypeptide from said host insect cell culture.

[0009] The invention further provides a method for producing a recombinant human procollagen polypeptide, said method comprising: (a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter; and (b) recovering the procollagen polypeptide from the host insect cell culture.

[0010] In an embodiment, the above-mentioned infected, transfected or transformed host insect cell comprising the recombinant baculovirus expression vector is obtained by a method comprising: (a) transfecting or transforming a first host insect cell with baculovirus DNA and an expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter; thereby to permit integration of said expression vector into said baculovirus DNA to obtain a recombinant baculovirus expression vector; (b) isolating a nucleic acid molecule comprising the recombinant baculovirus expression vector from said host cell; and (c) transfecting or transforming a second host insect cell with the nucleic acid molecule obtained in (b) thereby to obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector. In an embodiment, the method further comprises (d) culturing the infected, transfected or transformed host insect cell obtained in (c) under conditions suitable for production of recombinant baculovirus; (e) infecting a third host insect cell with the recombinant baculovirus obtained in (d), thereby to obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector.

[0011] The invention further provides a recombinant collagen polypeptide obtained by the above-mentioned method.

[0012] The invention further provides a recombinant procollagen polypeptide obtained by the above-mentioned method.

[0013] The invention further provides a recombinant baculovirus expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter.

[0014] The invention further provides a host insect cell which has been infected, transfected or transformed with the above-mentioned recombinant baculovirus expression vector.

[0015] The invention further provides a method for producing a recombinant human collagen or procollagen polypeptide, the method comprising: (a) culturing a host insect cell, wherein the insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a p10 promoter; and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a polH promoter; and (b) recovering the collagen or procollagen polypeptide from said host insect cell culture.

[0016] In an embodiment, the above-mentioned first promoter is a p10 promoter, e.g., comprising the promoter region set forth in FIG. **5** (SEQ ID NO: 9).

[0017] In an embodiment, the above-mentioned second promoter is a polyhedron (polH) promoter, e.g., comprising the promoter region set forth in FIG. **5** (SEQ ID NO: 10).

[0018] In an embodiment, the above-mentioned collagen subunit is a first collagen subunit and wherein the recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second collagen subunit, operably linked to a first promoter. In a further embodiment, the recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a third collagen subunit, operably linked to a first promoter.

[0019] In an embodiment, the above-mentioned subunit of a collagen post-translational enzyme is a first subunit of a collagen post-translational enzyme, and wherein the recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second subunit of a collagen post-translational enzyme, operably linked to a second promoter.

[0020] In an embodiment, the above-mentioned collagen is selected from collagen types I, II and Ill.

[0021] In an embodiment, the above-mentioned collagen is type II collagen and the collagen subunit is a collagen α1 (II) subunit.

[0022] In an embodiment, the above-mentioned collagen is type III collagen and the collagen subunit is a collagen α1 (III) subunit

[0023] In an embodiment, the above-mentioned collagen is type I collagen, the first collagen subunit is a collagen α1(I) subunit and the second collagen subunit is a collagen α2(I) subunit.

[0024] In an embodiment, the above-mentioned collagen post-translational enzyme is selected from prolyl hydroxylase, lysyl oxidase and lysyl hydroxylase. In a further embodiment, the collagen post-translational enzyme is prolyl 4-hydroxylase

[0025] In an embodiment, the above-mentioned collagen post-translational enzyme is prolyl 4-hydroxylase and wherein the first subunit of a collagen post-translational enzyme is an alpha subunit of prolyl 4-hydroxylase and wherein the second subunit of a collagen post-translational enzyme is a beta subunit of prolyl 4-hydroxylase.

[0026] In a further aspect, the invention provides a method of increasing or enhancing the purity of a collagen preparation, the method comprising incubating the collagen preparation under basic conditions such that the collagen is rendered insoluble in the basic solution, and recovering the insoluble collagen. In an embodiment, the method comprises dialyzing the collagen preparation against a basic solution.

[0027] In a further aspect, the invention provides a method of preparing collagen or processing a procollagen, the method comprising treating a procollagen sample with an elastase.

[0028] Other advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of specific embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] FIG. **1**: DNA (SEQ ID NO: 1) and polypeptide (SEQ ID NO: 2) sequences of human α-1(I) collagen (*Homo sapiens* col1-α1: Accession No. NM_000088; GI: 14719826). SEQ ID NO: 2 corresponds to the coding sequence defined by positions 127-4521 in SEQ ID NO: 1.

[0030] FIG. **2**: DNA (SEQ ID NO: 3) and polypeptide (SEQ ID NO: 4) sequences of human α-2(I) collagen (*Homo sapiens* collI-α2: Accession No. NM_000089, GI: 48762933), including the 5086 bp sequence coding for pre-pro-collI-α2 set forth in Accession No. Z74616, GI: 1418929. SEQ ID NO: 4 corresponds to the coding sequence defined by positions 472-4572 in SEQ ID NO: 3.

[0031] FIG. **3**: DNA (SEQ ID NO: 5) and polypeptide (SEQ ID NO: 6) sequences of human prolyl 4-hydroxylase alpha subunit (α-p4H; *Homo sapiens* p4Hα: 2722 bp sequence set forth in Accession No. M24486 (clone PA-II), GI: 190785). SEQ ID NO: 6 corresponds to the coding sequence defined by positions 119-1723 in SEQ ID NO: 5.

[0032] FIG. **4**: DNA (SEQ ID NO: 7) and polypeptide (SEQ ID NO: 8) sequences of human prolyl 4-hydroxylase beta subunit (β-p4H; *Homo sapiens* p4Hβ: 1956 bp sequence set forth in Accession No. X05130, GI: 35654). SEQ ID NO: 8 corresponds to the coding sequence defined by positions 30-1556 in SEQ ID NO: 7

[0033] FIG. **5**: Sequences of *Autographa californica* multicapsid nuclear polyhedrosis virus (AcMNPV) p10 (SEQ ID NO: 9) and polH (SEQ ID NO: 10) promoter regions (derived from *Autographa californica* nucleopolyhedrovirus complete genome Accession No. NC_001623, GI: 9627742), and also indicated in technical materials for plasmid pBAC4x-1™ (Novagen).

[0034] FIG. **6**. Photomicrographs taken under transmission electron microscope of Sf9 insect cells transfected with recombinant baculovirus, showing (A): viral inclusions within the cytoplasm of the cells and (B): lysis of the cell membrane by the virus.

[0035] FIG. **7**: Photomicrographs taken under phase contrast microscope of the polymerized recombinant human type I collagen fibrils showing alignment and formation of a structured network, (40×). Panel A shows neo-formed recombinant human type I collagen fibrils in formation, whereas panel B shows a sample prepared from the same batch and digested with pepsin, where no fibrils were detected.

[0036] FIG. **8**: Scanning electron microscopy image of polymerized recombinant human type I collagen fibers, assembled naturally.

[0037] FIG. **9**: Results of SDS-PAGE (under reducing conditions) analysis of recombinant human type I collagen chains, showing the purity of recombinant human type I collagen and the expression of the α1- and α2-chains prepared using the method described herein.

[0038] FIG. **10**: Recombinant procollagen type I resolved by SDS-PAGE, transferred on a nitrocellulose paper, followed by Concanavalin-A-biotin and streptavidin-Peroxidase blotting, showing the positive staining of its glycosylated amino acid residues.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0039] Applicants describe herein a method of producing collagen using a recombinant expression system. The method is preferably for the production of human collagen.

[0040] Applicants have found that collagen may be successfully produced in a recombinant expression system via the use of a single expression vector comprising both the nucleic acid(s) encoding a collagen subunit(s) and the nucleic acid(s) encoding a collagen post-translational enzyme or subunit(s) thereof. Prior to applicants' studies

described herein, attempts to produce recombinant collagen entailed the use of multiple expression vectors.

[0041] In an embodiment, four nucleic acids (which encode a corresponding polypeptide) may be inserted into a single vector, preferably a baculovirus vector. For example, these four nucleic acids may encode a first collagen subunit, a second collagen subunit, a first subunit of a collagen post-translational enzyme, and a second subunit of a collagen post-translational enzyme, respectively.

[0042] In a preferred embodiment, the expression system is a baculovirus/insect cell expression system. This system is advantageous because it provides, among other things, high levels of expression of the recombinant protein with the appropriate post-translational modifications and is amenable to scale-up for large-scale production. Various reagents for baculovirus/insect cell expression systems are known in the art and are commercially available. The Baculovirus Expression Vector System (BEVS) is one of the most powerful and versatile eukaryotic expression systems available. This expression system relies on the generation of recombinant baculoviruses in which viral genes, not essential for viral replication in cell culture, are replaced by DNA sequences of interest, (O'Reilly et al., 1992; Kidd and Emery, 1993). The recombinant viral DNA is typically transfected into *Spodoptera frugiperda* (Sf9) insect cells, clonal derivative of the fall armyworm *Spodoptera frugiperda* ovarian cell line, IPLB-Sf21-AE, (Sf21), (Vaughn et al, 1977; Nobiron et al, 2003). The recombinant proteins expressed in the baculovirus system are properly folded, disulphide bonded, oligomerized, and localized in the same subcellular compartment as the authentic protein (Kidd and Emery, 1993). Insect cells are also capable of performing several post-translational modifications such as N- and O-glycosylation, phosphorylation, acylation, amidation, carboximethilation, signal peptide cleavage, and proteolytic cleavage (Matsuura et al., 1987; Nokelainen M., 2000). The sites where these modifications occur are often identical to those of the authentic protein in its native cellular environment (Hoss et al., 1990; Kloc et al., 1991; Kuroda et al., 1990). In addition, insect cells possess a low prolyl-4-hydroxilase activity (Veijola et al., 1994). In this system, expression of the above-noted nucleic acids may be driven by the polyhedron (polH) promoter or the p10 promoter, both of which are known for use in baculovirus expression systems. Multiple copies of these promoters may be used in a vector to express multiple nucleic acids of interest.

[0043] In a preferred embodiment, the expression of the above-noted nucleic acids is driven by two different promoters, e.g., respective first and second promoters, such as the polH promoter for the expression of the collagen subunit and the p10 promoter for the expression of the collagen post-translational enzyme or subunit thereof.

[0044] Accordingly, in a first aspect, the invention provides a method for producing a recombinant collagen or procollagen polypeptide, such as a human collagen or procollagen polypeptide, said method comprising:

[0045] (a) culturing a host cell, such as a host insect cell, wherein the host cell has been infected, transfected or transformed with a recombinant expression vector (e.g., a recombinant baculovirus expression vector) comprising:

[0046] (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter (e.g., a polyhedron (polH) promoter); and

[0047] (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter (e.g. a p10 promoter);

[0048] (b) recovering said collagen or procollagen polypeptide from said host cell culture.

[0049] In embodiments, the host cell is a eukaryotic host cell selected from an insect cell, a fungal (e.g. yeast) cell, a mammalian cell and a plant cell. In a preferred embodiment the host cell is an insect host cell, such as a *Spodoptera frugiperda*-derived cell (e.g., Sf9 or Sf21).

[0050] In embodiments, single or multiple collagen or procollagen subunits can be expressed using the method of the invention. Therefore, in embodiments, the vector further comprises a nucleotide sequence which encodes a second collagen subunit operably linked to a promoter (e.g., a polH or p10 promoter). In a further embodiment, the vector yet further comprises a nucleotide sequence which encodes a third collagen subunit operably linked to a promoter (e.g., a polH or a p10 promoter).

[0051] In an embodiment, the vector further comprises a nucleotide sequence which encodes a further subunit of a collagen post-translational enzyme, operably linked to a promoter (e.g., a polH or a p10 promoter).

[0052] In embodiments, the collagen is selected from collagen types I-XIX. The appropriate collagen subunits for expression in each case are known in the art. For example, relevant subunits in respect of certain types of collagen are set forth in Table 1.

TABLE 1

Structures of certain types of collagen

Type	Structure
I	heterotrimeric: two α 1(I) chains; one α 2(I) chain
II	homotrimeric: three α 1(II) chains
III	homotrimeric: three α 1(III) chains
IV	most common form: heterotrimeric: two α 1(IV) chains; one α 2(IV) chain
V	multiple forms, e.g.: heterotrimeric: two α 1(V) chains; one α 2(V) chain heterotrimeric: one each of α 1(V), α 2(V) and α 3(V) chains homotrimeric: three α 1(V) chains
VI	heterotrimeric: one each of α 1(VI), α 2(VI) and α 3(VI) chains
VII	homotrimeric: three α 1(VII) chains
VIII	heterotrimeric: two α 1(VIII) chains; one α 2(VIII) chain (other structures also described)
IX	heterotrimeric: one each of α 1(IX), α 2(IX) and α 3(IX) chains
X	homotrimeric: three α 1(X) chains
XI	heterotrimeric: one each of α 1(XI), α 2(XI) and α 3(XI) chains
XII	homotrimeric: three α 1(XII) chains
XIV	homotrimeric: three α 1(XIV) chains

[0053] In embodiments, the collagen is selected from types I, II and Ill. In the case of type II collagen, the vector comprises a nucleotide sequence which encodes a collagen α1(II) subunit, operably linked to a promoter (e.g., a polH or a a p10 promoter). In the case of type III collagen, the vector comprises a nucleotide sequence which encodes a collagen α1(III) subunit, operably linked to a promoter (e.g., a polH or a p10 promoter). In the case of type I collagen the vector comprises a nucleotide sequence which encodes a collagen α1(I) subunit, operably linked to a promoter (e.g., a polH or a p10 promoter), and further comprises a a

nucleotide sequence which encodes a collagen α2(I) subunit, operably linked to a promoter (e.g., a polH or a p10 promoter).

[0054] In further embodiments, multiple copies of a a nucleotide sequence which encodes a collagen subunit may be included in the vector. For example, in the case of type II collagen (which has a homotrimeric structure), two (or more) copies of a nucleotide sequence encoding a collagen α1(II) subunit may be included in the vector.

[0055] An example of a human collagen α1(I) subunit corresponds to the polypeptide (SEQ ID NO: 2) set forth in FIG. **1**, which also sets forth the nucleic acid sequence (SEQ ID NO: 1) encoding the polypeptide. Positions 127-192 of SEQ ID NO: 1 correspond to the signal peptide. Positions 193-4518 of SEQ ID NO: 1 correspond to the proprotein.

[0056] An example of a human collagen α2(I) subunit corresponds to the polypeptide (SEQ ID NO: 4) set forth in FIG. **2**, which also sets forth the nucleic acid sequence (SEQ ID NO: 3) encoding the polypeptide.

[0057] In embodiments, the collagen post-translational enzyme is selected from prolyl hydroxylase (e.g. prolyl 4-hydroxylase), lysyl oxidase, lysyl hydroxylase, N-proteinase and C-proteinase.

[0058] Prolyl 4-hydroxylase is classified under enzyme classification EC 1.14.11.2 and catalyzes the hydroxylation of proline residues to 4-hydroxyproline in the synthesis of collagen or procollagen. The human form comprises two subunits, denoted as alpha and beta subunits. The human alpha-I isoform of the alpha subunit corresponds to for example Genbank accession No. M24486. The human alpha-II isoform of the alpha subunit corresponds to for example Genbank accession No. U90441. The human beta subunit corresponds to for example Genbank accession No. X05130. The vertebrate enzyme is a tetramer comprising two alpha and two beta subunits.

[0059] An example of a human prolyl 4-hydroxylase alpha subunit corresponds to the polypeptide (SEQ ID NO: 6) set forth in FIG. **3**, which also sets forth the nucleic acid sequence (SEQ ID NO: 5) encoding the polypeptide.

[0060] An example of a prolyl 4-hydroxylase beta subunit corresponds to the polypeptide (SEQ ID NO: 8) set forth in FIG. **4**, which also sets forth the nucleic acid sequence (SEQ ID NO: 7) encoding the polypeptide. Positions 30-80 of SEQ ID NO: 7 correspond to the signal peptide.

[0061] Lysyl hydroxylase is classified under classification EC 1.14.11.4 and catalyzes the hydroxylation of Lys residues in the -X-Lys-Gly- triplet motif of collagens. The enzyme is a homodimer of two alpha subunits. Various forms of the human enzyme correspond to for example Swiss-Prot accession Nos. Q02809, O00469 and O60568.

[0062] In the studies described herein, applicants have determined that post-translational processing of collagen may be effected by treatment with elastase. Therefore, in a further aspect, the invention provides a method of preparing collagen or processing a procollagen, said method comprising treating said procollagen sample with an elastase enzyme.

[0063] A preferred expression system to be used in the method of the invention is the baculovirus/insect cell expression system, whereby the expression vector comprising the above-noted nucleic acids is introduced into a host insect cell using a baculovirus construct. The host cell is thus cultured under conditions suitable for polypeptide production. An example of such a system utilizes the *Autographica californica* nuclear polyhydrosys virus (AcNPV), which grows in *Spodoptera frugiperda* cells. The nucleic acid(s) encoding the recombinant polypeptide(s) of interest (operably linked to an appropriate promoter for expression in the host cell) can be inserted into a non-essential region of AcNPV such as the polyhedron gene. In an embodiment, recombination of these nucleic acids into the non-essential region can result in the replacement or disruption of a marker gene, such as the lacz gene (β-galactosidase), thus allowing selection of recombinants based on the absence of the marker's activity. Such selection is sometimes referred to as a "plaque assay", as plaques may be selected on the basis of the absence or presence of marker activity. Such recombinant viruses may be used to infect the host insect cell for expression of the polypeptide(s) encoded by the inserted nucleic acid(s).

[0064] Preparation of such recombinant viruses typically entails the co-infection of linear viral DNA and a vector comprising the nucleic acid(s) encoding the recombinant polypeptide(s) of interest (operably linked to an appropriate promoter for expression in the host cell) into a host insect cell, whereby recombination results in the insertion of the nucleic acid(s) into the viral DNA. Recombinant virus produced may be identified by plaque assay, which is typically repeated to perform a second round of plaque purification. Ultimately, a stock of recombinant virus is obtained and used to infect host insect cells for polypeptide expression.

[0065] In an embodiment, the method comprises an isolation or amplification step, whereby the recombinant viral DNA comprising the nucleic acid(s) encoding the polypeptide(s) of interest (operably linked to an appropriate promoter for expression in the host cell) is isolated or obtained by amplification (e.g. by polymerase chain reaction [PCR]). The isolated or amplified recombinant viral DNA may then be directly introduced (e.g. transfected or transformed) into a host insect cell. Applicants have found that the use of such an additional isolation or amplification step further allows for the efficient preparation of host insect cells for the production of recombinant collagen or procollagen, rather than relying on baculovirus infection alone. Recombinant virus obtained from, for example, the culture medium of such host insect cells may be used as a viral stock to infect other host insect cells for further recombinant polypeptide production.

[0066] Accordingly, in an embodiment, the above-mentioned infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector is obtained by a method comprising:

(a) transfecting or transforming a first host insect cell with baculovirus DNA and an expression vector comprising: (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a promoter (e.g., a polH promoter); and (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a promoter (e.g., a p10 promoter); thereby to permit integration of said expression vector into said baculovirus DNA to obtain a recombinant baculovirus DNA expression vector;

(b) isolating a nucleic acid molecule comprising said recombinant baculovirus DNA expression vector from said host cell; and

(c) transfecting or transforming a second host insect cell with said nucleic acid molecule obtained in (b) thereby to

obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector.

[0067] In a further embodiment, the just-noted method further comprises: (d) culturing said infected, transfected or transformed host insect cell obtained in (c) above under conditions suitable for production of recombinant baculovirus; and (e) infecting a third host insect cell with the recombinant baculovirus obtained in (d) above, thereby to obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector.

[0068] The invention further provides a recombinant collagen or procollagen polypeptide obtained by the above-mentioned method.

[0069] The invention further provides the above-mentioned recombinant expression vector. In an embodiment, the expression vector is a recombinant baculovirus DNA expression vector.

[0070] The invention further provides a host cell, such as an insect host cell, which has been infected, transfected or transformed with the above-mentioned recombinant viral DNA expression vector.

[0071] "p10 promoter" as used herein refers to a nucleic acid sequence derived from the *Autographa californica* multicapsid nuclear polyhedrosis virus (AcMNPV) which can modulate the transcription of the AcMNPV p10 gene. Details of the p10 promoter are set forth in *Autographa californica* nucleopolyhedrovirus complete genome Accession No. NC_001623, GI: 9627742, as well as in the technical materials for plasmid pBAC4x-1™ (Novagen).

[0072] "polH promoter" or "polyhedron promoter" as used herein refers to a nucleic acid sequence derived from the AcMNPV which can modulate the transcription of the AcMNPV polyhedron gene. Details of the polH promoter are set forth in *Autographa californica* nucleopolyhedrovirus complete genome Accession No. NC_001623, GI: 9627742, as well as in the technical materials for plasmid pBAC4x-1™ (Novagen).

[0073] "Collagen" as used herein refers to any of the known collagen types (I-XIX) as well as any variants as described herein, and includes single chain, heterotrimeric and homotrimeric molecules of collagen. "Procollagen" as used herein is similarly defined and refers to any of the known procollagen as well as any variants as described herein, and includes single chain, heterotrimeric and homotrimeric molecules of procollagen, but differs from collagen in that it additionally comprises N-terminal and/or C-terminal peptides which are cleaved off for example by N-proteinase and/or C-proteinase enzymes.

[0074] As noted above, an isolated nucleic acid, for example a nucleic acid sequence encoding a polypeptide of the invention (e.g., a collagen or procollagen subunit; a collagen post-translational enzyme or subunit thereof), or homolog, fragment or variant thereof, may further be incorporated into a vector, such as a recombinant expression vector. In an embodiment, the vector will comprise transcriptional regulatory sequences or a promoter operably linked to a nucleic acid comprising a sequence capable of encoding a peptide compound, polypeptide or domain of the invention. A first nucleic acid sequence is "operably linked" with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequences. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in reading frame. However, since for example enhancers generally function when separated from the promoters by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably-linked but not contiguous. "Transcriptional regulatory sequence/element" is a generic term that refers to DNA sequences, such as initiation and termination signals, enhancers, and promoters, splicing signals, polyadenylation signals which induce or control transcription of protein coding sequences with which they are operably-linked. "Promoter" refers to a DNA regulatory region capable of binding directly or indirectly to RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of the present invention, the promoter is bound at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter will be found a transcription initiation site (conveniently defined by mapping with S1 nuclease), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CCAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the −10 and −35 consensus sequences.

[0075] As noted above, the invention relates to the recombinant production of collagen or procollagen. Thus, various nucleic acid sequences of the invention may be recombinant sequences. The term "recombinant" means that something has been recombined, so that when made in reference to a nucleic acid construct the term refers to a molecule that is comprised of nucleic acid sequences that are joined together or produced by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein or polypeptide molecule which is expressed using a recombinant nucleic acid construct created by means of molecular biological techniques. The term "recombinant" when made in reference to genetic composition refers to a gamete or progeny or cell or genome with new combinations of alleles that did not occur in the parental genomes. Recombinant nucleic acid constructs may include a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Referring to a nucleic acid construct as 'recombinant' therefore indicates that the nucleic acid molecule has been manipulated using genetic engineering, i.e., by human intervention. Recombinant nucleic acid constructs may for example be introduced into a host cell by transformation. Such recombinant nucleic acid constructs may include sequences derived from the same host cell species or from different host cell species, which have been isolated and reintroduced into cells of the host species. Recombinant nucleic acid construct sequences may become integrated into a host cell genome, either as a result of the original transformation of the host cells, or as the result of subsequent recombination and/or repair events.

[0076] The recombinant polypeptides of the invention may also be expressed in the form of a suitable fusion protein, comprising an amino acid sequence of a polypeptide of the invention linked to further polypeptide sequence (e.g.

a heterologous sequence). Such fusion proteins are typically produced by expression of recombinant nucleic acids encoding them. In embodiments, the further polypeptide sequence may confer various functions such as to facilitate cellular localization/secretion, detection and purification (e.g. via affinity methods).

[0077] The terminology "amplification pair" refers herein to a pair of oligonucleotides (oligos), which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

[0078] Oligonucleotide probes or primers of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted sequences employed. In general, the oligonucleotide probes or primers are at least 12 nucleotides in length, preferably between 15 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide probes and primers can be designed by taking into consideration the melting point of hybrizidation thereof with its targeted sequence (see below and in Sambrook et al., 1989, Molecular Cloning—A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

[0079] "Homology" and "homologous" refers to sequence similarity between two peptides or two nucleic acid molecules. Homology can be determined by comparing each position in the aligned sequences. A degree of homology between nucleic acid or between amino acid sequences is a function of the number of identical or matching nucleotides or amino acids at positions shared by the sequences. As the term is used herein, a nucleic acid sequence is "homologous" to another sequence if the two sequences are substantially identical and the functional activity of the sequences is conserved (as used herein, the term 'homologous' does not infer evolutionary relatedness). Two nucleic acid or polypeptide sequences are considered "substantially identical" if, when optimally aligned (with gaps permitted), they share at least about 50% sequence similarity or identity, or if the sequences share defined functional motifs. In alternative embodiments, sequence similarity in optimally aligned substantially identical sequences may be at least 60%, 70%, 75%, 80%, 85%, 90% or 95%. As used herein, a given percentage of homology between sequences denotes the degree of sequence identity in optimally aligned sequences. Similarly, "substantially complementary" nucleic acids are nucleic acids in which the complement of one molecule is "substantially identical" to the other molecule. An "unrelated" or "non-homologous" sequence shares less than 40% identity, though preferably less than about 25% identity, with a nucleic acid or polypeptide of the invention.

[0080] Alignment of sequences for comparisons of identity may be conducted using a variety of algorithms and methods, such as those of Smith and Waterman (1981, *Adv. Appl. Math* 2: 482), Needleman and Wunsch (1970, *J. Mol. Biol.* 48:443), Pearson and Lipman (1988, *Proc. Natl. Acad. Sci. USA* 85: 2444), and the computerised implementations of these algorithms (such as GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, Wis., U.S.A.). Sequence identity may also be determined using the BLAST algorithm, described in Altschul et al., 1990, *J. Mol. Biol.* 215:403-10 (using the published default settings). Software for performing BLAST analysis may be available through the National Center for Biotechnology Information (through the internet at http://www.ncbi.nlm.nih.gov/). The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold. Initial neighbourhood word hits act as seeds for initiating searches to find longer HSPs. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction is halted when the following parameters are met: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program may use as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (Henikoff and Henikoff, 1992, *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10 (or 1 or 0.1 or 0.01 or 0.001 or 0.0001), M=5, N=4, and a comparison of both strands. One measure of the statistical similarity between two sequences using the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. In alternative embodiments of the invention, nucleotide or amino acid sequences are considered substantially identical if the smallest sum probability in a comparison of the test sequences is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0081] An alternative indication that two nucleic acid sequences are substantially complementary is that the two sequences hybridize to each other under moderately stringent, or preferably stringent, conditions. "Nucleic acid hybridization" or "hybridize" generally refer to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples of hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, supra and Ausubel et al., 1989, supra) and are commonly known in the art. In the case of a hybridization to a nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65° C. with a labeled probe in a solution containing 50% formamide, high salt (5×SSC or 5×SSPE), 5× Denhardt's solution, 1% SDS, and 100 μg/ml denatured carrier DNA (i.e. salmon sperm DNA). The non-specifically binding probe can then be washed off the filter by several washes in 0.2×SSC/0.1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42° C. (moderate stringency) or 65° C. (high stringency). The selected tem-

perature is based on the melting temperature (Tm) of the DNA hybrid (Sambrook et al. 1989, supra). Of course, RNA-DNA hybrids can also be formed and detected. In such cases, the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al., 1989, supra).

[0082] As used herein, a "primer" defines an oligonucleotide which is capable of annealing to a target sequence, thereby creating a double stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

[0083] Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14-25. Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the Q-replicase system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, supra). Preferably, amplification will be carried out using PCR.

[0084] Polymerase chain reaction (PCR) is carried out in accordance with known techniques. See, e.g., U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188 (the disclosures of all three U.S. Patents are incorporated herein by reference). In general, PCR involves a treatment of a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer which is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample is analysed to assess whether the sequence or sequences to be detected are present. Detection of the amplified sequence may be carried out by visualization following EtBr staining of the DNA following gel electrophoresis, or using a detectable label in accordance with known techniques, and the like. For a review on PCR techniques (see PCR Protocols, A Guide to Methods and Amplifications, Michael et al. Eds, Acad. Press, 1990).

[0085] Ligase chain reaction (LCR) is carried out in accordance with known techniques (Weiss, 1991, Science 254:1292). Adaptation of the protocol to meet desired needs can be carried out by a person of ordinary skill. Strand displacement amplification (SDA) is also carried out in accordance with known techniques or adaptations thereof to meet the particular needs (Walker et al., 1992, Proc. Natl. Acad. Sci. USA 89:392-396; and ibid., 1992, Nucleic Acids Res. 20:1691-1696).

[0086] The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

[0087] The term "expression" defines the process by which a gene is transcribed into mRNA (transcription), the mRNA is then translated (translation) into one polypeptide (or protein) or more.

[0088] The recombinant expression vector of the present invention can be constructed by standard techniques known to one of ordinary skill in the art and found, for example, in Sambrook et al. (supra). A variety of strategies are available for ligating fragments of DNA, the choice of which depends on the nature of the termini of the DNA fragments and can be readily determined by persons skilled in the art. The vectors of the present invention may also contain other sequence elements to facilitate vector propagation and selection in bacteria and host cells. In addition, the vectors of the present invention may comprise a sequence of nucleotides for one or more restriction endonuclease sites. Coding sequences such as for selectable markers and reporter genes are well known to persons skilled in the art.

[0089] A recombinant expression vector comprising a nucleic acid sequence of the present invention may be introduced into a host cell, which may include a living cell capable of expressing the protein coding region from the defined recombinant expression vector. The living cell may include both a cultured cell and a cell within a living organism. Accordingly, the invention also provides host cells containing the recombinant expression vectors of the invention. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0090] Vector DNA can be introduced into cells via conventional transformation or transfection techniques. The terms "transformation" and "transfection" refer to techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection and viral-mediated transfection. Suitable methods for transforming or transfecting host cells can for example be found in Sambrook et al. (supra), and other laboratory manuals. "Infection" as used herein refers to the introduction of nucleic acids into a cell using a virus or viral vector, such as a baculovirus.

[0091] Polypeptides produced by the recombinant methods described herein can be purified according to standard protocols that take advantage for example of the intrinsic properties thereof, such as size and charge (i.e. SDS gel electrophoresis, gel filtration, dialysis, centrifugation, ion exchange chromatography . . .). In addition, the recombinant polypeptide can be purified via affinity chromatography using polyclonal or monoclonal antibodies or other affinity-based systems (e.g. using a suitable incorporated "tag" in the form of a fusion protein and its corresponding ligand). Its structure can be further modified using one or more enzymes or bioactive compounds.

[0092] Applicants have further demonstrated herein that dialysis under basic conditions allows for the efficient purification of collagen. As described in the Examples below, when dialyzed under basic conditions (e.g., at about pH 8.5 [e.g., in a sodium acetate buffer]), assembled collagen fibrils

polymerize while most contaminant proteins are solubilized. The collagen can then be recovered by appropriate means (e.g., centrifugation).

[0093] Accordingly, in a further aspect, the invention provides a method of enhancing the purity of a collagen preparation, said method comprising incubating the collagen under basic conditions (e.g., dialyzing the preparation against a basic solution), and recovering the collagen by suitable means (e.g., centrifugation, filtration, etc.). "Basic conditions" as used herein refers to conditions exhibiting an average pH greater than ph 7.0. In an embodiment, the pH is greater than or equal to 7.5, in a further embodiment, greater than or equal to 8.0. In an embodiment, the pH is about 8.5. Various buffer systems (e.g. acetate) are known in the art to prepare solutions exhibiting such basic conditions.

[0094] In a further embodiment a product of the invention (e.g., a polypeptide [e.g., a collagen polypeptide]) is substantially pure. A compound is "substantially pure" when it is separated from the components that naturally accompany it. Typically, a compound is substantially pure when it is at least 60%, more generally 75% or over 90%, by weight, of the total material in a sample. Thus, for example, a polypeptide that is chemically synthesised or produced by recombinant technology will generally be substantially free from its naturally associated components. A nucleic acid molecule is substantially pure when it is not immediately contiguous with (i.e., covalently linked to) the coding sequences with which it is normally contiguous in the naturally occurring genome of the organism from which the DNA of the invention is derived. A substantially pure compound can be obtained, for example, by extraction from a natural source; by expression of a recombinant nucleic acid molecule encoding a polypeptide compound; or by chemical synthesis. Purity can be measured using any appropriate method such as column chromatography, gel electrophoresis, HPLC, etc.

[0095] A homolog, variant and/or fragment of a polypeptide of the invention which retains activity, and nucleic acids encoding such a homolog, variant and/or fragment, may also be used in the methods of the invention. Homologs include polypeptide sequences, which are substantially identical to the amino acid sequence of a a polypeptide of the invention, sharing significant structural and functional homology with a polypeptide of the invention. Variants include, but are not limited to, polypeptides, which differ from a a polypeptide of the invention by any modifications, and/or amino acid substitutions, deletions or additions. Modifications can occur anywhere including the polypeptide backbone, (i.e. the amino acid sequence), the amino acid side chains and the amino or carboxy termini. Such substitutions, deletions or additions may involve one or more amino acids. Fragments include a fragment or a portion of a polypeptide of the invention, or a fragment or a portion of a homolog or variant of a polypeptide of the invention.

[0096] The present invention is illustrated in further details by the following non-limiting examples.

EXAMPLES

Example 1

Materials and Methods

[0097] pBAC4x-1™ transfer plasmid and Insect GeneJuice™ transfection reagent were obtained from Novagen (EMD Biosciences/Novagen/VWR CANLAB, Mississaga, Ontario, Canada). Clones containing nucleic acids encoding subunits of collagen and prolyl-4-hydroxylase were obtained from American Type Culture Collection (ATCC) as follows: ATCC # 59480 for P4H beta subunit; ATCC # 138677 for P4H alpha subunit; ATCC # 95501 for coil 1 alpha-2 chain; and ATCC # 95499 for coil 1 alpha-1 chain.

Example 2

Preparation of Baculovirus Expression Construct

[0098] Preparation of the baculovirus expression construct pBacNI-hcoll I was performed by modification of pBAC4x-1™. The β subunit of P4H was inserted first. It was cloned and the cDNA was amplified by PCR with primers tagged to the EcoRI/SpeI ends. The PCR amplified cDNA was ligated to the sites SmaI and SpeI in linearized pBAC4x-1™ Secondly, the coil α-1(I) subunit was inserted via the insertion of an XbaI restriction enzyme fragment containing DNA encoding the coil α-1(I) subunit into the XbaI site of linearized pBAC4x-1™. Thirdly, The coil α-2(I) subunit cDNA was inserted via the insertion of an SphI restriction enzyme fragment containing DNA encoding the coil α-2(I) subunit into the SphI site of the linearized pBAC4x-1™. Subsequently, the α and β subunits of P4H were inserted. The detailed sequence of the final construct, is described in Table 2.

TABLE 2

Description of pBacNI-hcoll I sequence (18,169 bps)

Sequence	Description
1–1242	plasmid pBac4x derived sequence
1243–1248	BgIII site
1249–3004	P4Hβ sequence (oriented counterclockwise)
3148–3152	BgIII site
3153–3401	plasmid pBac4x derived sequence
3402–3406	Xbal site
3408–7207	Col1α1(I) sequence and non coding sequence (oriented counterclockwise)
7208–7213	Xbal site
7214–7453	plasmid pBac4x derived sequence
7454–7459	Smal site
7480–9578	P4Hα sequence
9579–9584	Spel site
9585–9774	plasmid pBac4x derived sequence
9775–9780	HindIII site
14255–14260	Sphl site
14261–18169	plasmid pBac4x derived sequence

Example 3

Expression, Maturation and Purification of Recombinant Human Collagen

[0099] Infected Sf9 cells were grown for 2-6 days in Grace's medium supplemented with ascorbic acid (50 ug/ml) to stimulate collagen synthesis. sf9 cells (suspension of 1 L) were centrifuged to obtain a pellet of cells that contain human recombinant procollagen. The cell pellet was resuspended in about 100 ml of 50 mM Tris-HCl buffer containing 0.2 M NaCl, at pH 7.4. The cells were broken mechanically to liberate pro-collagen (e.g. freezing-thawing twice at −20° C. and 4° C., respectively). Since the extract also contains DNA, coming out of the broken cells, that can

provoke DNA-pro-collagen aggregates, DNAse treatment was used to eliminate the DNA.

[0100] The procoll/coll suspension was then digested with elastase at 4° C., for 2-3 hrs by adding half volume of 50 mM Tris-HCL, pH 8.5 containing elastase (1-2 mg/ml). After this incubation period, the assembled collagen fibrils were dialyzed against a sodium acetate buffer pH 8.5 for 72 hrs at 4° C. The white fibrils polymerize, while most contaminant proteins were solubilized during dialysis.

[0101] After the completion of the dialysis, the collagen was centrifuged for 20 min at 6000 rpm and the pellet was rinsed twice with megapure water, centrifuging each time to recuperate the pellet. A protease inhibitor cocktail was added to the fibrils. The collagen was solubilized in citric acid 0.075M, pH 3.7, overnight, and the residual contaminant proteins that are precipitated were discarded by centrifugation (pellet). The supernatant, containing the acid-solubilized collagen, was dialysed against phosphate buffer 0.02M, pH 9.2 to 9.5, at 4° C. The fibrils slowly precipitated within 2-3 days. The fibrils were centrifuged, washed 3 times and resuspended in megapure water. The suspension was frozen at −86° C. and lyophilized.

[0102] In some experiments, the procoll/coll suspension was then precipitated with ammonium sulfate for about 2 hrs at 4° C. and centrifuged to obtain a pellet of proteins. The pellet was resuspended in Tris-HCl buffer containing 0.2 M NaCl, at pH 7.4 and the suspension was dialyzed against a acetic acid (1:1000 or 0.5M) for 72 hrs at 4° C. Then, the suspension was frozen at −80° C. and lyophilized.

Example 4

Characterization of Recombinant Human Collagen

[0103] The total amino acid composition, including the percentage of the collagen content in proline, hydroxyproline, lysine and hydroxy-lysine, and a partial amino acid sequencing of the final product may be performed. The material for analysis may be cut enzymatically before being analyzed. Electron microscopy analyses can reveal the length, the periodicity and the overall organization of the collagen fibers, and thermostability can be evaluated also (Fertala et al., 1994). For example, FIGS. **7** and **8** show results of microscopic analysis of collagen produced according to the method described herein.

[0104] The gycosylation of procollagen can be assessed by testing its affinity with lectins, such as Concanavalin A, that specifically binds glusose and mannose residues. The purity, the respective molecular weights and amounts of α-1 and α-2 chains of the processed collagen can be analyzed on SDS-PAGE. The confirmation of the nature of the collagen can be tested on Western blots, using antibodies directed specifically against human type I collagen. For example, FIG. **9** shows results of SDS-PAGE analysis of collagen produced according to the method described herein. FIG. **10** shows SDS-PAGE analysis together with Concanavalin A-based staining of collagen produced according to the method described herein.

[0105] The capacity of the collagen to polymerise into a gel can be assessed by solubilizing the collagen in acetic acid 1:1000 and bring the solution to physiological pH (7.2-7.5).

[0106] The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

[0107] Although the present invention has been described hereinabove by way of specific embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES

[0108] Behera A K, Kumar M, Bansal A, Bansal O B, Das R H. Expression of lacZ reporter gene under the control of the polyhedrin promoter of *Spodoptera litura* nuclear polyhedrosis virus. Gene 190: 145-150, 1997.

[0109] Bulleid N J, John D C A, Kadler K E. Recombinant expression systems for the production of collagen. Biochem. Soc. Transact. 28:350-353, 2000.

[0110] Fertala A, Sieron A L, Hojima Y, Ganguly A, Prockop D J. Self-assembly into fibrils of collagen I1 by enzymic cleavage of recombinant procollagen II. J. Biol. Chem. 269: 11584-11589, 1994.

[0111] Hoss A, Moarefi I, Scheidtmann K H, Cisek L J, Corden J L, Dornreiter I, Arthur A K, Fanning E. Altered phosphorylation pattern of simian virus 40 T antigen expressed in insect cells by using a baculovirus vector. J. Virol. 64: 4799-4807, 1990.

[0112] Hojima Y, Behta B, Romanic A M, Prockop D J Cadmium ions inhibit procollagen C-proteinase and cupric ions inhibit procollagen N-proteinase. Matrix Biol. 14:113-120, 1994a.

[0113] Hojima Y, Morgelin M M, Engel J, Boutillon M M, van der Rest M, McKenzie J, Chen G C, Rafi N, Romanic A M, Prockop D J. Characterization of type I procollagen N-proteinase from fetal bovine tendon and skin. Purification of the 500-kilodalton form of the enzyme from bovine tendon. J Biol. Chem. 269:11381-11390, 1994b.

[0114] Kidd I M, Emery V C. The use of baculoviruses as expression vectors. Appl. Biochem. Biotechnol. 42:137-159, 1993.

[0115] Kloc M, Reddy B, Crawford S, Etkin L D. A novel 110-kDa maternal CAAX box-containing protein from *Xenopus* is palmitoylated and isoprenylated when expressed in baculovirus. J. Biol. Chem. 266: 8206-8212, 1991.

[0116] Kuroda K, Geyer H, Geyer R, Doerfler W, Klenk H D. The oligosaccharides of influenza virus hemagglutinin expressed in insect cells by a baculovirus vector. Virology 174: 418-429, 1990.

[0117] Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko K I. Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. J. Biol. Chem. 271: 11988-11995, 1996.

[0118] Matsuura Y, Possee R D, Overton H A, Bishop D H. Baculovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins. J Gen Virol. 68: 1233-1250, 1987.

[0119] Moschcovich L, Bernocco S, Font B, Rivkin H, Eichenberger D, Chejanovsky N, Hulmes D J, Kessler E. Folding and activity of recombinant human procollagen C-proteinase enhancer. Eur. J. Biochem. 268: 2991-2996, 2001.

[0120] Nobiron I, O'Reilly D R, Olszewski J A. *Autographa californica* nucleopolyhedrovirus infection of *Spodoptera frugiperda* cells: a global analysis of host

gene regulation during infection, using a differential display approach. J. Gen. Virol. 84:3029-3039, 2003.

[0121] Nokelainen M. Recombinant human collagens: characterization of type ii collagen expressed in insect cells and production of types I-III collagen in the yeast *pichia pastoris*. Department of medical biochemistry. Faculty of medicine, University of Oulu (Findland), Oulu University press, pp. 70, 2000.

[0122] O'Reilly D, Miller L K, Luckow V A. Baculovirus expression vectors: A Laboratory manual. W.H. Freeman and Co., New York, 1992.

[0123] Possee R D, Howard S C. Analysis of the polyhedrin gene promoter of the *Autographa californica* nuclear polyhedrosis virus. Nucleic Acids Res. 15: 10233-1048, 1987.

[0124] Romanic A M, Adachi E, Kadler K E, Hojima Y, Prockop D J. Copolymerization of pNcollagen III and collagen I. pNcollagen III decreases the rate of incorporation of collagen I into fibrils, the amount of collagen I incorporated, and the diameter of the fibrils formed. J. Biol. Chem. 266:12703-12709, 1991.

[0125] Romanic A M, Adachi E, Hojima Y, Engel J, Prockop D J. Polymerization of pNcollagen I and copolymerization of pNcollagen I with collagen I. A kinetic, thermodynamic, and morphologic study. J Biol. Chem. 267: 22265-71, 1992.

[0126] Vaughn J L, Goodwin R H, Tompkins G J, McCawley P. The establishment of two cell lines from the insect *Spodoptera frugiperda* (*Lepidoptera*; Noctuidae). In Vitro 13:13-17, 1977.

[0127] Veijola J, Koivunen P, Annunen P, Pihlajaniemi T, Kivirikko K I. Cloning, baculovirus expression, and characterization of the alpha subunit of prolyl 4-hydroxylase from the nematode *Caenorhabditis elegans*. This alpha subunit forms an active alpha beta dimer with the human protein disulfide isomerase/beta subunit. J. Biol. Chem. 269: 26746-26753, 1994.

[0128] Vlak J M, Klinkenberg F A, Zaal K J, Usmany M, Klinge-Roode E C, Geervliet J B, Roosien J, van Lent J W. Functional studies on the p10 gene of *Autographa californica* nuclear polyhedrosis virus using a recombinant expressing a p10-beta-galactosidase fusion gene. J. Gen. Virol. 69:765-776, 1988.

[0129] Weyer U, Possee R D. Functional analysis of the p10 gene 5' leader sequence of the *Autographa californica* nuclear polyhedrosis virus. Nucleic Acids Res. 16: 3635-3653, 1988.

[0130] Weyer U, Knight S, Possee R D. Analysis of very late gene expression by *Autographa californica* nuclear polyhedrosis virus and the further development of multiple expression vectors. J Gen Virol. 71: 1525-1534, 1990.

[0131] Weyer U, Possee R D. A baculovirus dual expression vector derived from the *Autographa californica* nuclear polyhedrosis virus polyhedrin and p10 promoters: co-expression of two influenza virus genes in insect cells. J. Gen. Virol. 72: 2967-2974, 1991.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 10

<210> SEQ ID NO 1
<211> LENGTH: 5927
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (127)..(4521)

<400> SEQUENCE: 1

```
tcgtcggagc agacgggagt ttctcctcgg ggtcggagca ggaggcacgc ggagtgtgag     60

gccacgcatg agcggacgct aacccctcc ccagccacaa agagtctaca tgtctagggt    120

ctagac atg ttc agc ttt gtg gac ctc cgg ctc ctg ctc ctc tta gcg       168

             Met Phe Ser Phe Val Asp Leu Arg Leu Leu Leu Leu Ala
             1               5                  10

gcc acc gcc ctc ctg acg cac ggc caa gag gaa ggc caa gtc gag ggc      216
Ala Thr Ala Leu Leu Thr His Gly Gln Glu Glu Gly Gln Val Glu Gly
15                  20                  25                  30

caa gac gaa gac atc cca cca atc acc tgc gta cag aac ggc ctc agg      264
Gln Asp Glu Asp Ile Pro Pro Ile Thr Cys Val Gln Asn Gly Leu Arg
                35                  40                  45

tac cat gac cga gac gtg tgg aaa ccc gag ccc tgc cgg atc tgc gtc      312
Tyr His Asp Arg Asp Val Trp Lys Pro Glu Pro Cys Arg Ile Cys Val
            50                  55                  60

tgc gac aac ggc aag gtg ttg tgc gat gac gtg atc tgt gac gag acc      360
```

-continued

```
Cys Asp Asn Gly Lys Val Leu Cys Asp Asp Val Ile Cys Asp Glu Thr
        65                  70                  75

aag aac tgc ccc ggc gcc gaa gtc ccc gag ggc gag tgc tgt ccc gtc      408
Lys Asn Cys Pro Gly Ala Glu Val Pro Glu Gly Glu Cys Cys Pro Val
    80                  85                  90

tgc ccc gac ggc tca gag tca ccc acc gac caa gaa acc acc ggc gtc      456
Cys Pro Asp Gly Ser Glu Ser Pro Thr Asp Gln Glu Thr Thr Gly Val
95                  100                 105                 110

gag gga ccc aag gga gac act ggc ccc cga ggc cca agg gga ccc gca      504
Glu Gly Pro Lys Gly Asp Thr Gly Pro Arg Gly Pro Arg Gly Pro Ala
                115                 120                 125

ggc ccc cct ggc cga gat ggc atc cct gga cag cct gga ctt ccc gga      552
Gly Pro Pro Gly Arg Asp Gly Ile Pro Gly Gln Pro Gly Leu Pro Gly
            130                 135                 140

ccc ccc gga ccc ccc gga cct ccc gga ccc cct ggc ctc gga gga aac      600
Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Leu Gly Gly Asn
        145                 150                 155

ttt gct ccc cag ctg tct tat ggc tat gat gag aaa tca acc gga gga      648
Phe Ala Pro Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly
    160                 165                 170

att tcc gtg cct ggc ccc atg ggt ccc tct ggt cct cgt ggt ctc cct      696
Ile Ser Val Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro
175                 180                 185                 190

ggc ccc cct ggt gca cct ggt ccc caa ggc ttc caa ggt ccc cct ggt      744
Gly Pro Pro Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly
                195                 200                 205

gag cct ggc gag cct gga gct tca ggt ccc atg ggt ccc cga ggt ccc      792
Glu Pro Gly Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro
            210                 215                 220

cca ggt ccc cct gga aag aat gga gat gat ggg gaa gct gga aaa cct      840
Pro Gly Pro Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro
        225                 230                 235

ggt cgt cct ggt gag cgt ggg cct cct ggg cct cag ggt gct cga gga      888
Gly Arg Pro Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly
    240                 245                 250

ttg ccc gga aca gct ggc ctc cct gga atg aag gga cac aga ggt ttc      936
Leu Pro Gly Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe
255                 260                 265                 270

agt ggt ttg gat ggt gcc aag gga gat gct ggt cct gct ggt cct aag      984
Ser Gly Leu Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys
                275                 280                 285

ggt gag cct ggc agc cct ggt gaa aat gga gct cct ggt cag atg ggc     1032
Gly Glu Pro Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly
            290                 295                 300

ccc cgt ggc ctg cct ggt gag aga ggt cgc cct gga gcc cct ggc cct     1080
Pro Arg Gly Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro
        305                 310                 315

gct ggt gct cgt gga aat gat ggt gct act ggt gct gcc ggg ccc cct     1128
Ala Gly Ala Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro
    320                 325                 330

ggt ccc acc ggc ccc gct ggt cct cct ggc ttc cct ggt gct gtt ggt     1176
Gly Pro Thr Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly
335                 340                 345                 350

gct aag ggt gaa gct ggt ccc caa ggg ccc cga ggc tct gaa ggt ccc     1224
Ala Lys Gly Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro
                355                 360                 365

cag ggt gtg cgt ggt gag cct ggc ccc cct ggc cct gct ggt gct gct     1272
```

-continued

```
Gln Gly Val Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala
            370                 375                 380

ggc cct gct gga aac cct ggt gct gat gga cag cct ggt gct aaa ggt     1320
Gly Pro Ala Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly
        385                 390                 395

gcc aat ggt gct cct ggt att gct ggt gct cct ggc ttc cct ggt gcc     1368
Ala Asn Gly Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala
    400                 405                 410

cga ggc ccc tct gga ccc cag ggc ccc ggc ggc cct cct ggt ccc aag     1416
Arg Gly Pro Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys
415                 420                 425                 430

ggt aac agc ggt gaa cct ggt gct cct ggc agc aaa gga gac act ggt     1464
Gly Asn Ser Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly
                435                 440                 445

gct aag gga gag cct ggc cct gtt ggt gtt caa gga ccc cct ggc cct     1512
Ala Lys Gly Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro
            450                 455                 460

gct gga gag gaa gga aag cga gga gct cga ggt gaa ccc gga ccc act     1560
Ala Gly Glu Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr
        465                 470                 475

ggc ctg ccc gga ccc cct ggc gag cgt ggt gga cct ggt agc cgt ggt     1608
Gly Leu Pro Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly
    480                 485                 490

ttc cct ggc gca gat ggt gtt gct ggt ccc aag ggt ccc gct ggt gaa     1656
Phe Pro Gly Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu
495                 500                 505                 510

cgt ggt tct cct ggc cct gct ggc ccc aaa gga tct cct ggt gaa gct     1704
Arg Gly Ser Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala
                515                 520                 525

ggt cgt ccc ggt gaa gct ggt ctg cct ggt gcc aag ggt ctg act gga     1752
Gly Arg Pro Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly
            530                 535                 540

agc cct ggc agc cct ggt cct gat ggc aaa act ggc ccc cct ggt ccc     1800
Ser Pro Gly Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro
        545                 550                 555

gcc ggt caa gat ggt cgc ccc gga ccc cca ggc cca cct ggt gcc cgt     1848
Ala Gly Gln Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg
    560                 565                 570

ggt cag gct ggt gtg atg gga ttc cct gga cct aaa ggt gct gct gga     1896
Gly Gln Ala Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly
575                 580                 585                 590

gag ccc ggc aag gct gga gag cga ggt gtt ccc gga ccc cct ggc gct     1944
Glu Pro Gly Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala
                595                 600                 605

gtc ggt cct gct ggc aaa gat gga gag gct gga gct cag gga ccc cct     1992
Val Gly Pro Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro
            610                 615                 620

ggc cct gct ggt ccc gct ggc gag aga ggt gaa caa ggc cct gct ggc     2040
Gly Pro Ala Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly
        625                 630                 635

tcc ccc gga ttc cag ggt ctc cct ggt cct gct ggt cct cca ggt gaa     2088
Ser Pro Gly Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu
    640                 645                 650

gca ggc aaa cct ggt gaa cag ggt gtt cct gga gac ctt ggc gcc cct     2136
Ala Gly Lys Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro
655                 660                 665                 670

ggc ccc tct gga gca aga ggc gag aga ggt ttc cct ggc gag cgt ggt     2184
```

-continued

```
Gly Pro Ser Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly
                675                 680                 685

gtg caa ggt ccc cct ggt cct gct ggt ccc cga ggg gcc aac ggt gct     2232
Val Gln Gly Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala
            690                 695                 700

ccc ggc aac gat ggt gct aag ggt gat gct ggt gcc cct gga gct ccc     2280
Pro Gly Asn Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro
        705                 710                 715

ggt agc cag ggc gcc cct ggc ctt cag gga atg cct ggt gaa cgt ggt     2328
Gly Ser Gln Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly
    720                 725                 730

gca gct ggt ctt cca ggg cct aag ggt gac aga ggt gat gct ggt ccc     2376
Ala Ala Gly Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro
735                 740                 745                 750

aaa ggt gct gat ggc tct cct ggc aaa gat ggc gtc cgt ggt ctg act     2424
Lys Gly Ala Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr
                755                 760                 765

ggc ccc att ggt cct cct ggc cct gct ggt gcc cct ggt gac aag ggt     2472
Gly Pro Ile Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly
            770                 775                 780

gaa agt ggt ccc agc ggc cct gct ggt ccc act gga gct cgt ggt gcc     2520
Glu Ser Gly Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala
        785                 790                 795

ccc gga gac cgt ggt gag cct ggt ccc ccc ggc cct gct ggc ttt gct     2568
Pro Gly Asp Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala
    800                 805                 810

ggc ccc cct ggt gct gac ggc caa cct ggt gct aaa ggc gaa cct ggt     2616
Gly Pro Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly
815                 820                 825                 830

gat gct ggt gct aaa ggc gat gct ggt ccc cct ggc cct gcc gga ccc     2664
Asp Ala Gly Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro
                835                 840                 845

gct gga ccc cct ggc ccc att ggt aat gtt ggt gct cct gga gcc aaa     2712
Ala Gly Pro Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys
            850                 855                 860

ggt gct cgc ggc agc gct ggt ccc cct ggt gct act ggt ttc cct ggt     2760
Gly Ala Arg Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly
        865                 870                 875

gct gct ggc cga gtc ggt cct cct ggc ccc tct gga aat gct gga ccc     2808
Ala Ala Gly Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro
    880                 885                 890

cct ggc cct cct ggt cct gct ggc aaa gaa ggc ggc aaa ggt ccc cgt     2856
Pro Gly Pro Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg
895                 900                 905                 910

ggt gag act ggc cct gct gga cgt cct ggt gaa gtt ggt ccc cct ggt     2904
Gly Glu Thr Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly
                915                 920                 925

ccc cct ggc cct gct ggc gag aaa gga tcc cct ggt gct gat ggt cct     2952
Pro Pro Gly Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro
            930                 935                 940

gct ggt gct cct ggt act ccc ggg cct caa ggt att gct gga cag cgt     3000
Ala Gly Ala Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg
        945                 950                 955

ggt gtg gtc ggc ctg cct ggt cag aga gga gag aga ggc ttc cct ggt     3048
Gly Val Val Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly
    960                 965                 970

ctt cct ggc ccc tct ggt gaa cct ggc aaa caa ggt ccc tct gga gca     3096
```

-continued

```
Leu Pro Gly Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala
975                 980                 985                 990

agt ggt gaa cgt ggt ccc cct ggt ccc atg  ggc ccc cct gga ttg  gct   3144
Ser Gly Glu Arg Gly Pro Pro Gly Pro Met  Gly Pro Pro Gly Leu  Ala
                995                 1000                1005

gga ccc cct ggt  gaa tct gga cgt gag  ggg gct cct ggt gcc  gaa      3189
Gly Pro Pro Gly  Glu Ser Gly Arg Glu  Gly Ala Pro Gly Ala  Glu
           1010                1015                1020

ggt tcc cct gga  cga gac ggt tct cct  ggc gcc aag ggt gac  cgt      3234
Gly Ser Pro Gly  Arg Asp Gly Ser Pro  Gly Ala Lys Gly Asp  Arg
           1025                1030                1035

ggt gag acc ggc  ccc gct gga ccc cct  ggt gct cct ggt gct  cct      3279
Gly Glu Thr Gly  Pro Ala Gly Pro Pro  Gly Ala Pro Gly Ala  Pro
           1040                1045                1050

ggt gcc cct ggc  ccc gtt ggc cct gct  ggc aag agt ggt gat  cgt      3324
Gly Ala Pro Gly  Pro Val Gly Pro Ala  Gly Lys Ser Gly Asp  Arg
           1055                1060                1065

ggt gag act ggt  cct gct ggt ccc gcc  ggt cct gtc ggc cct  gtt      3369
Gly Glu Thr Gly  Pro Ala Gly Pro Ala  Gly Pro Val Gly Pro  Val
           1070                1075                1080

ggc gcc cgt ggc  ccc gcc gga ccc caa  ggc ccc cgt ggt gac  aag      3414
Gly Ala Arg Gly  Pro Ala Gly Pro Gln  Gly Pro Arg Gly Asp  Lys
           1085                1090                1095

ggt gag aca ggc  gaa cag ggc gac aga  ggc ata aag ggt cac  cgt      3459
Gly Glu Thr Gly  Glu Gln Gly Asp Arg  Gly Ile Lys Gly His  Arg
           1100                1105                1110

ggc ttc tct ggc  ctc cag ggt ccc cct  ggc cct cct ggc tct  cct      3504
Gly Phe Ser Gly  Leu Gln Gly Pro Pro  Gly Pro Pro Gly Ser  Pro
           1115                1120                1125

ggt gaa caa ggt  ccc tct gga gcc tct  ggt cct gct ggt ccc  cga      3549
Gly Glu Gln Gly  Pro Ser Gly Ala Ser  Gly Pro Ala Gly Pro  Arg
           1130                1135                1140

ggt ccc cct ggc  tct gct ggt gct cct  ggc aaa gat gga ctc  aac      3594
Gly Pro Pro Gly  Ser Ala Gly Ala Pro  Gly Lys Asp Gly Leu  Asn
           1145                1150                1155

ggt ctc cct ggc  ccc att ggg ccc cct  ggt cct cgc ggt cgc  act      3639
Gly Leu Pro Gly  Pro Ile Gly Pro Pro  Gly Pro Arg Gly Arg  Thr
           1160                1165                1170

ggt gat gct ggt  cct gtt ggt ccc ccc  ggc cct cct gga cct  cct      3684
Gly Asp Ala Gly  Pro Val Gly Pro Pro  Gly Pro Pro Gly Pro  Pro
           1175                1180                1185

ggt ccc cct ggt  cct ccc agc gct ggt  ttc gac ttc agc ttc  ctg      3729
Gly Pro Pro Gly  Pro Pro Ser Ala Gly  Phe Asp Phe Ser Phe  Leu
           1190                1195                1200

ccc cag cca cct  caa gag aag gct cac  gat ggt ggc cgc tac  tac      3774
Pro Gln Pro Pro  Gln Glu Lys Ala His  Asp Gly Gly Arg Tyr  Tyr
           1205                1210                1215

cgg gct gat gat  gcc aat gtg gtt cgt  gac cgt gac ctc gag  gtg      3819
Arg Ala Asp Asp  Ala Asn Val Val Arg  Asp Arg Asp Leu Glu  Val
           1220                1225                1230

gac acc acc ctc  aag agc ctg agc cag  cag atc gag aac atc  cgg      3864
Asp Thr Thr Leu  Lys Ser Leu Ser Gln  Gln Ile Glu Asn Ile  Arg
           1235                1240                1245

agc cca gag ggc  agc cgc aag aac ccc  gcc cgc acc tgc cgt  gac      3909
Ser Pro Glu Gly  Ser Arg Lys Asn Pro  Ala Arg Thr Cys Arg  Asp
           1250                1255                1260

ctc aag atg tgc  cac tct gac tgg aag  agt gga gag tac tgg  att      3954
```

-continued

```
Leu Lys Met Cys  His Ser Asp Trp Lys  Ser Gly Glu Tyr Trp  Ile
            1265                1270                1275

gac ccc aac caa  ggc tgc aac ctg gat  gcc atc aaa gtc ttc  tgc      3999
Asp Pro Asn Gln  Gly Cys Asn Leu Asp  Ala Ile Lys Val Phe  Cys
            1280                1285                1290

aac atg gag act  ggt gag acc tgc gtg  tac ccc act cag ccc  agt      4044
Asn Met Glu Thr  Gly Glu Thr Cys Val  Tyr Pro Thr Gln Pro  Ser
            1295                1300                1305

gtg gcc cag aag  aac tgg tac atc agc  aag aac ccc aag gac  aag      4089
Val Ala Gln Lys  Asn Trp Tyr Ile Ser  Lys Asn Pro Lys Asp  Lys
            1310                1315                1320

agg cat gtc tgg  ttc ggc gag agc atg  acc gat gga ttc cag  ttc      4134
Arg His Val Trp  Phe Gly Glu Ser Met  Thr Asp Gly Phe Gln  Phe
            1325                1330                1335

gag tat ggc ggc  cag ggc tcc gac cct  gcc gat gtg gcc atc  cag      4179
Glu Tyr Gly Gly  Gln Gly Ser Asp Pro  Ala Asp Val Ala Ile  Gln
            1340                1345                1350

ctg acc ttc ctg  cgc ctg atg tcc acc  gag gcc tcc cag aac  atc      4224
Leu Thr Phe Leu  Arg Leu Met Ser Thr  Glu Ala Ser Gln Asn  Ile
            1355                1360                1365

acc tac cac tgc  aag aac agc gtg gcc  tac atg gac cag cag  act      4269
Thr Tyr His Cys  Lys Asn Ser Val Ala  Tyr Met Asp Gln Gln  Thr
            1370                1375                1380

ggc aac ctc aag  aag gcc ctg ctc ctc  cag ggc tcc aac gag  atc      4314
Gly Asn Leu Lys  Lys Ala Leu Leu Leu  Gln Gly Ser Asn Glu  Ile
            1385                1390                1395

gag atc cgc gcc  gag ggc aac agc cgc  ttc acc tac agc gtc  act      4359
Glu Ile Arg Ala  Glu Gly Asn Ser Arg  Phe Thr Tyr Ser Val  Thr
            1400                1405                1410

gtc gat ggc tgc  acg agt cac acc gga  gcc tgg ggc aag aca  gtg      4404
Val Asp Gly Cys  Thr Ser His Thr Gly  Ala Trp Gly Lys Thr  Val
            1415                1420                1425

att gaa tac aaa  acc acc aag acc tcc  cgc ctg ccc atc atc  gat      4449
Ile Glu Tyr Lys  Thr Thr Lys Thr Ser  Arg Leu Pro Ile Ile  Asp
            1430                1435                1440

gtg gcc ccc ttg  gac gtt ggt gcc cca  gac cag gaa ttc ggc  ttc      4494
Val Ala Pro Leu  Asp Val Gly Ala Pro  Asp Gln Glu Phe Gly  Phe
            1445                1450                1455

gac gtt ggc cct  gtc tgc ttc ctg taa actccctcca tcccaacctg         4541
Asp Val Gly Pro  Val Cys Phe Leu
            1460

gctccctccc acccaaccaa ctttcccccc aacccggaaa cagacaagca acccaaactg   4601

aacccctca aaagccaaaa aatgggagac aatttcacat ggactttgga aaatattttt   4661

ttcctttgca ttcatctctc aaacttagtt tttatctttg accaaccgaa catgaccaaa   4721

aaccaaaagt gcattcaacc ttaccaaaaa aaaaaaaaaa aaaagaataa ataaataact   4781

ttttaaaaaa ggaagcttgg tccacttgct tgaagaccca tgcgggggta agtccctttc   4841

tgcccgttgg gcttatgaaa ccccaatgct gccctttctg ctcctttctc cacacccccc   4901

ttggggcctc ccctccactc cttcccaaat ctgtctcccc agaagacaca ggaaacaatg   4961

tattgtctgc ccagcaatca aaggcaatgc tcaaacaccc aagtggcccc caccctcagc   5021

ccgctcctgc ccgcccagca cccccaggcc ctgggggacc tggggttctc agactgccaa   5081

agaagccttg ccatctggcg ctcccatggc tcttgcaaca tctccccttc gtttttgagg   5141

gggtcatgcc gggggagcca ccagcccctc actgggttcg gaggagagtc aggaagggcc   5201
```

-continued

```
acgacaaagc agaaacatcg gatttgggga acgcgtgtca atcccttgtg ccgcagggct   5261

gggcgggaga gactgttctg ttccttgtgt aactgtgttg ctgaaagact acctcgttct   5321

tgtcttgatg tgtcaccggg gcaactgcct gggggcgggg atgggggcag ggtggaagcg   5381

gctccccatt ttataccaaa ggtgctacat ctatgtgatg ggtggggtgg ggagggaatc   5441

actggtgcta tagaaattga gatgcccccc caggccagca aatgttcctt tttgttcaaa   5501

gtctattttt attccttgat atttttcttt ttttttttgt ttttttgtgg atggggactt   5561

gtgaattttt ctaaaggtgc tatttaacat gggaggagag cgtgtgcggc tccagcccag   5621

cccgctgctc actttccacc ctctctccac ctgcctctgg cttctcaggc ctctgctctc   5681

cgacctctct cctctgaaac cctcctccac agctgcagcc catcctcccg gctccctcct   5741

agtctgtcct gcgtcctctg tccccgggtt tcagagacaa cttcccaaag cacaaagcag   5801

tttttccccc taggggtggg aggaagcaaa agactctgta cctattttgt atgtgtataa   5861

taatttgaga tgtttttaat tattttgatt gctggaataa agcatgtgga aatgacccaa   5921

acataa                                                             5927
```

<210> SEQ ID NO 2
<211> LENGTH: 1464
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

```
Met Phe Ser Phe Val Asp Leu Arg Leu Leu Leu Leu Leu Ala Ala Thr
1               5                   10                  15

Ala Leu Leu Thr His Gly Gln Glu Glu Gly Gln Val Glu Gly Gln Asp
            20                  25                  30

Glu Asp Ile Pro Pro Ile Thr Cys Val Gln Asn Gly Leu Arg Tyr His
        35                  40                  45

Asp Arg Asp Val Trp Lys Pro Glu Pro Cys Arg Ile Cys Val Cys Asp
    50                  55                  60

Asn Gly Lys Val Leu Cys Asp Asp Val Ile Cys Asp Glu Thr Lys Asn
65                  70                  75                  80

Cys Pro Gly Ala Glu Val Pro Glu Gly Glu Cys Cys Pro Val Cys Pro
                85                  90                  95

Asp Gly Ser Glu Ser Pro Thr Asp Gln Glu Thr Thr Gly Val Glu Gly
            100                 105                 110

Pro Lys Gly Asp Thr Gly Pro Arg Gly Pro Arg Gly Pro Ala Gly Pro
        115                 120                 125

Pro Gly Arg Asp Gly Ile Pro Gly Gln Pro Gly Leu Pro Gly Pro Pro
    130                 135                 140

Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Leu Gly Gly Asn Phe Ala
145                 150                 155                 160

Pro Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser
                165                 170                 175

Val Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro
            180                 185                 190

Pro Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro
        195                 200                 205

Gly Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly
    210                 215                 220
```

-continued

```
Pro Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg
225                 230                 235                 240

Pro Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro
                245                 250                 255

Gly Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly
            260                 265                 270

Leu Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu
        275                 280                 285

Pro Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg
    290                 295                 300

Gly Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly
305                 310                 315                 320

Ala Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro
                325                 330                 335

Thr Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys
            340                 345                 350

Gly Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly
        355                 360                 365

Val Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro
    370                 375                 380

Ala Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn
385                 390                 395                 400

Gly Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly
                405                 410                 415

Pro Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn
            420                 425                 430

Ser Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys
        435                 440                 445

Gly Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly
    450                 455                 460

Glu Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu
465                 470                 475                 480

Pro Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro
                485                 490                 495

Gly Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly
            500                 505                 510

Ser Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg
        515                 520                 525

Pro Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro
    530                 535                 540

Gly Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly
545                 550                 555                 560

Gln Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln
                565                 570                 575

Ala Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro
            580                 585                 590

Gly Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly
        595                 600                 605

Pro Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro
    610                 615                 620
```

-continued

```
Ala Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro
625                 630                 635                 640

Gly Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly
                645                 650                 655

Lys Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro
            660                 665                 670

Ser Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln
        675                 680                 685

Gly Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly
    690                 695                 700

Asn Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser
705                 710                 715                 720

Gln Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala
                725                 730                 735

Gly Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly
            740                 745                 750

Ala Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro
        755                 760                 765

Ile Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser
    770                 775                 780

Gly Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly
785                 790                 795                 800

Asp Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro
                805                 810                 815

Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala
            820                 825                 830

Gly Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly
        835                 840                 845

Pro Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala
    850                 855                 860

Arg Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala
865                 870                 875                 880

Gly Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly
                885                 890                 895

Pro Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu
            900                 905                 910

Thr Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro
        915                 920                 925

Gly Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly
    930                 935                 940

Ala Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val
945                 950                 955                 960

Val Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro
                965                 970                 975

Gly Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly
            980                 985                 990

Glu Arg Gly Pro Pro Gly Pro Met  Gly Pro Pro Gly Leu  Ala Gly Pro
        995                 1000                1005

Pro Gly  Glu Ser Gly Arg Glu  Gly Ala Pro Gly Ala  Glu Gly Ser
    1010                1015                1020

Pro Gly  Arg Asp Gly Ser Pro  Gly Ala Lys Gly Asp  Arg Gly Glu
```

-continued

```
    1025                1030                1035

Thr Gly  Pro Ala Gly Pro Pro  Gly Ala Pro Gly Ala  Pro Gly Ala
    1040                1045                1050

Pro Gly  Pro Val Gly Pro Ala  Gly Lys Ser Gly Asp  Arg Gly Glu
    1055                1060                1065

Thr Gly  Pro Ala Gly Pro Ala  Gly Pro Val Gly Pro  Val Gly Ala
    1070                1075                1080

Arg Gly  Pro Ala Gly Pro Gln  Gly Pro Arg Gly Asp  Lys Gly Glu
    1085                1090                1095

Thr Gly  Glu Gln Gly Asp Arg  Gly Ile Lys Gly His  Arg Gly Phe
    1100                1105                1110

Ser Gly  Leu Gln Gly Pro Pro  Gly Pro Pro Gly Ser  Pro Gly Glu
    1115                1120                1125

Gln Gly  Pro Ser Gly Ala Ser  Gly Pro Ala Gly Pro  Arg Gly Pro
    1130                1135                1140

Pro Gly  Ser Ala Gly Ala Pro  Gly Lys Asp Gly Leu  Asn Gly Leu
    1145                1150                1155

Pro Gly  Pro Ile Gly Pro Pro  Gly Pro Arg Gly Arg  Thr Gly Asp
    1160                1165                1170

Ala Gly  Pro Val Gly Pro Pro  Gly Pro Pro Gly Pro  Pro Gly Pro
    1175                1180                1185

Pro Gly  Pro Pro Ser Ala Gly  Phe Asp Phe Ser Phe  Leu Pro Gln
    1190                1195                1200

Pro Pro  Gln Glu Lys Ala His  Asp Gly Gly Arg Tyr  Tyr Arg Ala
    1205                1210                1215

Asp Asp  Ala Asn Val Val Arg  Asp Arg Asp Leu Glu  Val Asp Thr
    1220                1225                1230

Thr Leu  Lys Ser Leu Ser Gln  Gln Ile Glu Asn Ile  Arg Ser Pro
    1235                1240                1245

Glu Gly  Ser Arg Lys Asn Pro  Ala Arg Thr Cys Arg  Asp Leu Lys
    1250                1255                1260

Met Cys  His Ser Asp Trp Lys  Ser Gly Glu Tyr Trp  Ile Asp Pro
    1265                1270                1275

Asn Gln  Gly Cys Asn Leu Asp  Ala Ile Lys Val Phe  Cys Asn Met
    1280                1285                1290

Glu Thr  Gly Glu Thr Cys Val  Tyr Pro Thr Gln Pro  Ser Val Ala
    1295                1300                1305

Gln Lys  Asn Trp Tyr Ile Ser  Lys Asn Pro Lys Asp  Lys Arg His
    1310                1315                1320

Val Trp  Phe Gly Glu Ser Met  Thr Asp Gly Phe Gln  Phe Glu Tyr
    1325                1330                1335

Gly Gly  Gln Gly Ser Asp Pro  Ala Asp Val Ala Ile  Gln Leu Thr
    1340                1345                1350

Phe Leu  Arg Leu Met Ser Thr  Glu Ala Ser Gln Asn  Ile Thr Tyr
    1355                1360                1365

His Cys  Lys Asn Ser Val Ala  Tyr Met Asp Gln Gln  Thr Gly Asn
    1370                1375                1380

Leu Lys  Lys Ala Leu Leu Leu  Gln Gly Ser Asn Glu  Ile Glu Ile
    1385                1390                1395

Arg Ala  Glu Gly Asn Ser Arg  Phe Thr Tyr Ser Val  Thr Val Asp
    1400                1405                1410
```

-continued

```
Gly Cys  Thr Ser His Thr Gly  Ala Trp Gly Lys Thr  Val Ile Glu
    1415                 1420                 1425

Tyr Lys  Thr Thr Lys Thr Ser  Arg Leu Pro Ile Ile  Asp Val Ala
    1430                 1435                 1440

Pro Leu  Asp Val Gly Ala Pro  Asp Gln Glu Phe Gly  Phe Asp Val
    1445                 1450                 1455

Gly Pro  Val Cys Phe Leu
    1460
```

<210> SEQ ID NO 3
<211> LENGTH: 5411
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (472)..(4572)

<400> SEQUENCE: 3

```
gtgtcccata gtgtttccaa acttggaaag ggcgggggag ggcgggagga tgcggagggc     60

ggaggtatgc agacaacgag tcagagtttc cccttgaaag cctcaaaagt gtccacgtcc    120

tcaaaaagaa tggaaccaat ttaagaagcc agccccgtgg ccacgtccct tcccccattc    180

gctccctcct ctgcgccccc gcaggctcct cccagctgtg gctgcccggg cccccagccc    240

cagccctccc attggtggag gccctttttgg aggcacccta gggccaggga aacttttgcc    300

gtataaatag ggcagatccg ggctttatta ttttagcacc acggcagcag gaggtttcgg    360

ctaagttgga ggtactggcc acgactgcat gcccgcgccc gccaggtgat acctccgccg    420

gtgacccagg ggctctgcga cacaaggagt ctgcatgtct aagtgctaga c atg ctc     477
                                                          Met Leu
                                                          1

agc ttt gtg gat acg cgg act ttg ttg ctg ctt gca gta acc tta tgc      525
Ser Phe Val Asp Thr Arg Thr Leu Leu Leu Leu Ala Val Thr Leu Cys
        5                   10                  15

cta gca aca tgc caa tct tta caa gag gaa act gta aga aag ggc cca      573
Leu Ala Thr Cys Gln Ser Leu Gln Glu Glu Thr Val Arg Lys Gly Pro
    20                  25                  30

gcc gga gat aga gga cca cgt gga gaa agg ggt cca cca ggc ccc cca      621
Ala Gly Asp Arg Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Pro Pro
35                  40                  45                  50

ggc aga gat ggt gaa gat ggt ccc aca ggc cct cct ggt cca cct ggt      669
Gly Arg Asp Gly Glu Asp Gly Pro Thr Gly Pro Pro Gly Pro Pro Gly
                55                  60                  65

cct cct ggc ccc cct ggt ctc ggt ggg aac ttt gct gct cag tat gat      717
Pro Pro Gly Pro Pro Gly Leu Gly Gly Asn Phe Ala Ala Gln Tyr Asp
            70                  75                  80

gga aaa gga gtt gga ctt ggc cct gga cca atg ggc tta atg gga cct      765
Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met Gly Pro
        85                  90                  95

aga ggc cca cct ggt gca gct gga gcc cca ggc cct caa ggt ttc caa      813
Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln
    100                 105                 110

gga cct gct ggt gag cct ggt gaa cct ggt caa act ggt cct gca ggt      861
Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro Ala Gly
115                 120                 125                 130

gct cgt ggt cca gct ggc cct cct ggc aag gct ggt gaa gat ggt cac      909
```

-continued

```
Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp Gly His
               135                 140                 145

cct gga aaa ccc gga cga cct ggt gag aga gga gtt gtt gga cca cag      957
Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val Gly Pro Gln
           150                 155                 160

ggt gct cgt ggt ttc cct gga act cct gga ctt cct ggc ttc aaa ggc     1005
Gly Ala Arg Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly Phe Lys Gly
       165                 170                 175

att agg gga cac aat ggt ctg gat gga ttg aag gga cag ccc ggt gct     1053
Ile Arg Gly His Asn Gly Leu Asp Gly Leu Lys Gly Gln Pro Gly Ala
   180                 185                 190

cct ggt gtg aag ggt gaa cct ggt gcc cct ggt gaa aat gga act cca     1101
Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly Glu Asn Gly Thr Pro
195                 200                 205                 210

ggt caa aca gga gcc cgt ggg ctt cct ggt gag aga gga cgt gtt ggt     1149
Gly Gln Thr Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly Arg Val Gly
               215                 220                 225

gcc cct ggc cca gct ggt gcc cgt ggc agt gat gga agt gtg ggt ccc     1197
Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser Val Gly Pro
           230                 235                 240

gtg ggt cct gct ggt ccc att ggg tct gct ggc cct cca ggc ttc cca     1245
Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro Gly Phe Pro
       245                 250                 255

ggt gcc cct ggc ccc aag ggt gaa att gga gct gtt ggt aac gct ggt     1293
Gly Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly Asn Ala Gly
   260                 265                 270

cct gct ggt ccc gcc ggt ccc cgt ggt gaa gtg ggt ctt cca ggc ctc     1341
Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu Pro Gly Leu
275                 280                 285                 290

tcc ggc ccc gtt gga cct cct ggt aat cct gga gca aac ggc ctt act     1389
Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn Gly Leu Thr
               295                 300                 305

ggt gcc aag ggt gct gct ggc ctt ccc ggc gtt gct ggg gct ccc ggc     1437
Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly Ala Pro Gly
           310                 315                 320

ctc cct gga ccc cgc ggt att cct ggc cct gtt ggt gct gcc ggt gct     1485
Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala Ala Gly Ala
       325                 330                 335

act ggt gcc aga gga ctt gtt ggt gag cct ggt cca gct ggc tcc aaa     1533
Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala Gly Ser Lys
   340                 345                 350

gga gag agc ggt aac aag ggt gag ccc ggc tct gct ggg ccc caa ggt     1581
Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly Pro Gln Gly
355                 360                 365                 370

cct cct ggt ccc agt ggt gaa gaa gga aag aga ggc cct aat ggg gaa     1629
Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro Asn Gly Glu
               375                 380                 385

gct gga tct gcc ggc cct cca gga cct cct ggg ctg aga ggt agt cct     1677
Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg Gly Ser Pro
           390                 395                 400

ggt tct cgt ggt ctt cct gga gct gat ggc aga gct ggc gtc atg ggc     1725
Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly Val Met Gly
       405                 410                 415

cct cct ggt agt cgt ggt gca agt ggc cct gct gga gtc cga gga cct     1773
Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val Arg Gly Pro
   420                 425                 430

aat gga gat gct ggt cgc cct ggg gag cct ggt ctc atg gga ccc aga     1821
```

-continued

```
Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met Gly Pro Arg
435                 440                 445                 450

ggt ctt cct ggt tcc cct gga aat atc ggc ccc gct gga aaa gaa ggt     1869
Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly Lys Glu Gly
                455                 460                 465

cct gtc ggc ctc cct ggc atc gac ggc agg cct ggc cca att ggc cca     1917
Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro Ile Gly Pro
            470                 475                 480

gct gga gca aga gga gag cct ggc aac att gga ttc cct gga ccc aaa     1965
Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro Gly Pro Lys
        485                 490                 495

ggc ccc act ggt gat cct ggc aaa aac ggt gat aaa ggt cat gct ggt     2013
Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly His Ala Gly
    500                 505                 510

ctt gct ggt gct cgg ggt gct cca ggt cct gat gga aac aat ggt gct     2061
Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn Asn Gly Ala
515                 520                 525                 530

cag gga cct cct gga cca cag ggt gtt caa ggt gga aaa ggt gaa cag     2109
Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly Glu Gln
                535                 540                 545

ggt ccc cct ggt cct cca ggc ttc cag ggt ctg cct ggc ccc tca ggt     2157
Gly Pro Pro Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro Ser Gly
            550                 555                 560

ccc gct ggt gaa gtt ggc aaa cca gga gaa agg ggt ctc cat ggt gag     2205
Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His Gly Glu
        565                 570                 575

ttt ggt ctc cct ggt cct gct ggt cca aga ggg gaa cgc ggt ccc cca     2253
Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg Gly Pro Pro
    580                 585                 590

ggt gag agt ggt gct gcc ggt cct act ggt cct att gga agc cga ggt     2301
Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly Ser Arg Gly
595                 600                 605                 610

cct tct gga ccc cca ggg cct gat gga aac aag ggt gaa cct ggt gtg     2349
Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu Pro Gly Val
                615                 620                 625

gtt ggt gct gtg ggc act gct ggt cca tct ggt cct agt gga ctc cca     2397
Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser Gly Leu Pro
            630                 635                 640

gga gag agg ggt gct gct ggc ata cct gga ggc aag gga gaa aag ggt     2445
Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly Glu Lys Gly
        645                 650                 655

gaa cct ggt ctc aga ggt gaa att ggt aac cct ggc aga gat ggt gct     2493
Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp Gly Ala
    660                 665                 670

cgt ggt gct cct ggt gct gta ggt gcc cct ggt cct gct gga gcc aca     2541
Arg Gly Ala Pro Gly Ala Val Gly Ala Pro Gly Pro Ala Gly Ala Thr
675                 680                 685                 690

ggt gac cgg ggc gaa gct ggg gct gct ggt cct gct ggt cct gct ggt     2589
Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro Ala Gly
                695                 700                 705

cct cgg gga agc cct ggt gaa cgt ggt gag gtc ggt cct gct ggc ccc     2637
Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala Gly Pro
            710                 715                 720

aat gga ttt gct ggt cct gct ggt gct gct ggt caa cct ggt gct aaa     2685
Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro Gly Ala Lys
        725                 730                 735

gga gaa aga gga gcc aaa ggg cct aag ggt gaa aac ggt gtt gtt ggt     2733
```

-continued

```
Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly Val Val Gly
    740                 745                 750

ccc aca ggc ccc gtt gga gct gct ggc cca gct ggt cca aat ggt ccc     2781
Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro Asn Gly Pro
755                 760                 765                 770

ccc ggt cct gct gga agt cgt ggt gat gga ggc ccc cct ggt atg act     2829
Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro Gly Met Thr
                775                 780                 785

ggt ttc cct ggt gct gct gga cgg act ggt ccc cca gga ccc tct ggt     2877
Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly Pro Ser Gly
            790                 795                 800

att tct ggc cct cct ggt ccc cct ggt cct gct ggg aaa gaa ggg ctt     2925
Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys Glu Gly Leu
        805                 810                 815

cgt ggt cct cgt ggt gac caa ggt cca gtt ggc cga act gga gaa gta     2973
Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr Gly Glu Val
    820                 825                 830

ggt gca gtt ggt ccc cct ggc ttc gct ggt gag aag ggt ccc tct gga     3021
Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly Pro Ser Gly
835                 840                 845                 850

gag gct ggt act gct gga cct cct ggc act cca ggt cct cag ggt ctt     3069
Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro Gln Gly Leu
                855                 860                 865

ctt ggt gct cct ggt att ctg ggt ctc cct ggc tcg aga ggt gaa cgt     3117
Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg Gly Glu Arg
            870                 875                 880

ggt cta cca ggt gtt gct ggt gct gtg ggt gaa cct ggt cct ctt ggc     3165
Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly Pro Leu Gly
        885                 890                 895

att gcc ggc cct cct ggg gcc cgt ggt cct cct ggt gct gtg ggt agt     3213
Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala Val Gly Ser
    900                 905                 910

cct gga gtc aac ggt gct cct ggt gaa gct ggt cgt gat ggc aac cct     3261
Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp Gly Asn Pro
915                 920                 925                 930

ggg aac gat ggt ccc cca ggt cgc gat ggt caa ccc gga cac aag gga     3309
Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly His Lys Gly
                935                 940                 945

gag cgc ggt tac cct ggc aat att ggt ccc gtt ggt gct gca ggt gca     3357
Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala Ala Gly Ala
            950                 955                 960

cct ggt cct cat ggc ccc gtg ggt cct gct ggc aaa cat gga aac cgt     3405
Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His Gly Asn Arg
        965                 970                 975

ggt gaa act ggt cct tct ggt cct gtt ggt cct gct ggt gct gtt ggc     3453
Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly Ala Val Gly
    980                 985                 990

cca aga ggt cct agt ggc  cca caa ggc att cgt  ggc gat aag gga       3498
Pro Arg Gly Pro Ser Gly  Pro Gln Gly Ile Arg  Gly Asp Lys Gly
995                 1000                1005

gag  ccc ggt gaa aag ggg  ccc aga ggt ctt cct  ggc tta aag gga      3543
Glu  Pro Gly Glu Lys Gly  Pro Arg Gly Leu Pro  Gly Leu Lys Gly
1010                1015                1020

cac  aat gga ttg caa ggt  ctg cct ggt atc gct  ggt cac cat ggt      3588
His  Asn Gly Leu Gln Gly  Leu Pro Gly Ile Ala  Gly His His Gly
1025                1030                1035

gat  caa ggt gct cct ggc  tcc gtg ggt cct gct  ggt cct agg ggc      3633
```

-continued

```
Asp  Gln Gly Ala Pro Gly  Ser Val Gly Pro Ala  Gly Pro Arg Gly
1040                1045                1050

cct  gct ggt cct tct ggc  cct gct gga aaa gat  ggt cgc act gga      3678
Pro  Ala Gly Pro Ser Gly  Pro Ala Gly Lys Asp  Gly Arg Thr Gly
1055                1060                1065

cat  cct ggt aca gtt gga  cct gct ggc att cga  ggc cct cag ggt      3723
His  Pro Gly Thr Val Gly  Pro Ala Gly Ile Arg  Gly Pro Gln Gly
1070                1075                1080

cac  caa ggc cct gct ggc  ccc cct ggt ccc cct  ggc cct cct gga      3768
His  Gln Gly Pro Ala Gly  Pro Pro Gly Pro Pro  Gly Pro Pro Gly
1085                1090                1095

cct  cca ggt gta agc ggt  ggt ggt tat gac ttt  ggt tac gat gga      3813
Pro  Pro Gly Val Ser Gly  Gly Gly Tyr Asp Phe  Gly Tyr Asp Gly
1100                1105                1110

gac  ttc tac agg gct gac  cag cct cgc tca gca  cct tct ctc aga      3858
Asp  Phe Tyr Arg Ala Asp  Gln Pro Arg Ser Ala  Pro Ser Leu Arg
1115                1120                1125

ccc  aag gac tat gaa gtt  gat gct act ctg aag  tct ctc aac aac      3903
Pro  Lys Asp Tyr Glu Val  Asp Ala Thr Leu Lys  Ser Leu Asn Asn
1130                1135                1140

cag  att gag acc ctt ctt  act cct gaa ggc tct  aga aag aac cca      3948
Gln  Ile Glu Thr Leu Leu  Thr Pro Glu Gly Ser  Arg Lys Asn Pro
1145                1150                1155

gct  cgc aca tgc cgt gac  ttg aga ctc agc cac  cca gag tgg agc      3993
Ala  Arg Thr Cys Arg Asp  Leu Arg Leu Ser His  Pro Glu Trp Ser
1160                1165                1170

agt  ggt tac tac tgg att  gac cct aac caa gga  tgc act atg gat      4038
Ser  Gly Tyr Tyr Trp Ile  Asp Pro Asn Gln Gly  Cys Thr Met Asp
1175                1180                1185

gct  atc aaa gta tac tgt  gat ttc tct act ggc  gaa acc tgt atc      4083
Ala  Ile Lys Val Tyr Cys  Asp Phe Ser Thr Gly  Glu Thr Cys Ile
1190                1195                1200

cgg  gcc caa cct gaa aac  atc cca gcc aag aac  tgg tat agg agc      4128
Arg  Ala Gln Pro Glu Asn  Ile Pro Ala Lys Asn  Trp Tyr Arg Ser
1205                1210                1215

tcc  aag gac aag aaa cac  gtc tgg cta gga gaa  act atc aat gct      4173
Ser  Lys Asp Lys Lys His  Val Trp Leu Gly Glu  Thr Ile Asn Ala
1220                1225                1230

ggc  agc cag ttt gaa tat  aat gta gaa gga gtg  act tcc aag gaa      4218
Gly  Ser Gln Phe Glu Tyr  Asn Val Glu Gly Val  Thr Ser Lys Glu
1235                1240                1245

atg  gct acc caa ctt gcc  ttc atg cgc ctg ctg  gcc aac tat gcc      4263
Met  Ala Thr Gln Leu Ala  Phe Met Arg Leu Leu  Ala Asn Tyr Ala
1250                1255                1260

tct  cag aac atc acc tac  cac tgc aag aac agc  att gca tac atg      4308
Ser  Gln Asn Ile Thr Tyr  His Cys Lys Asn Ser  Ile Ala Tyr Met
1265                1270                1275

gat  gag gag act ggc aac  ctg aaa aag gct gtc  att cta cag ggc      4353
Asp  Glu Glu Thr Gly Asn  Leu Lys Lys Ala Val  Ile Leu Gln Gly
1280                1285                1290

tct  aat gat gtt gaa ctt  gtt gct gag ggc aac  agc agg ttc act      4398
Ser  Asn Asp Val Glu Leu  Val Ala Glu Gly Asn  Ser Arg Phe Thr
1295                1300                1305

tac  act gtt ctt gta gat  ggc tgc tct aaa aag  aca aat gaa tgg      4443
Tyr  Thr Val Leu Val Asp  Gly Cys Ser Lys Lys  Thr Asn Glu Trp
1310                1315                1320

gga  aag aca atc att gaa  tac aaa aca aat aag  cca tca cgc ctg      4488
```

-continued

```
Gly  Lys Thr Ile Ile Glu  Tyr Lys Thr Asn Lys  Pro Ser Arg Leu
1325                1330                1335

ccc  ttc ctt gat att gca  cct ttg gac atc ggt  ggt gct gac cag      4533
Pro  Phe Leu Asp Ile Ala  Pro Leu Asp Ile Gly  Gly Ala Asp Gln
1340                1345                1350

gaa  ttc ttt gtg gac att  ggc cca gtc tgt ttc  aaa taa atgaactcaa   4582
Glu  Phe Phe Val Asp Ile  Gly Pro Val Cys Phe  Lys
1355                1360                1365

tctaaattaa aaaagaaaga aatttgaaaa aactttctct ttgccatttc ttcttcttct   4642

tttttaactg aaagctgaat ccttccattt cttctgcaca tctacttgct taaattgtgg   4702

gcaaaagaga aaaagaagga ttgatcagag cattgtgcaa tacagtttca ttaactcctt   4762

cccccgctcc cccaaaaatt tgaatttttt tttcaacact cttacacctg ttatggaaaa   4822

tgtcaacctt tgtaagaaaa ccaaaataaa aattgaaaaa taaaaaccat aaacatttgc   4882

accacttgtg gcttttgaat atcttccaca gagggaagtt taaaacccaa acttccaaag   4942

gtttaaacta cctcaaaaca ctttcccatg agtgtgatcc acattgttag gtgctgacct   5002

agacagagat gaactgaggt ccttgttttg ttttgttcat aatacaaagg tgctaattaa   5062

tagtatttca gatacttgaa gaatgttgat ggtgctagaa gaatttgaga agaaatactc   5122

ctgtattgag ttgtatcgtg tggtgtattt tttaaaaaat ttgatttagc attcatattt   5182

tccatcttat tcccaattaa aagtatgcag attatttgcc caaatcttct tcagattcag   5242

catttgttct ttgccagtct cattttcatc ttcttccatg gttccacaga agctttgttt   5302

cttgggcaag cagaaaaatt aaattgtacc tattttgtat atgtgagatg tttaaataaa   5362

ttgtgaaaaa aatgaaataa agcatgtttg gttttccaaa agaacatat               5411


<210> SEQ ID NO 4
<211> LENGTH: 1366
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

Met Leu Ser Phe Val Asp Thr Arg Thr Leu Leu Leu Leu Ala Val Thr
1               5                   10                  15

Leu Cys Leu Ala Thr Cys Gln Ser Leu Gln Glu Glu Thr Val Arg Lys
            20                  25                  30

Gly Pro Ala Gly Asp Arg Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly
        35                  40                  45

Pro Pro Gly Arg Asp Gly Glu Asp Gly Pro Thr Gly Pro Pro Gly Pro
    50                  55                  60

Pro Gly Pro Pro Gly Pro Pro Gly Leu Gly Gly Asn Phe Ala Ala Gln
65                  70                  75                  80

Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met
                85                  90                  95

Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly
            100                 105                 110

Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro
        115                 120                 125

Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp
    130                 135                 140

Gly His Pro Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Val Val Gly
145                 150                 155                 160
```

-continued

```
Pro Gln Gly Ala Arg Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly Phe
                165                 170                 175

Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly Leu Lys Gly Gln Pro
            180                 185                 190

Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly Glu Asn Gly
        195                 200                 205

Thr Pro Gly Gln Thr Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly Arg
    210                 215                 220

Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser Val
225                 230                 235                 240

Gly Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro Gly
                245                 250                 255

Phe Pro Gly Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly Asn
            260                 265                 270

Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Glu Val Gly Leu Pro
        275                 280                 285

Gly Leu Ser Gly Pro Val Gly Pro Pro Gly Asn Pro Gly Ala Asn Gly
    290                 295                 300

Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu Pro Gly Val Ala Gly Ala
305                 310                 315                 320

Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro Gly Pro Val Gly Ala Ala
                325                 330                 335

Gly Ala Thr Gly Ala Arg Gly Leu Val Gly Glu Pro Gly Pro Ala Gly
            340                 345                 350

Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu Pro Gly Ser Ala Gly Pro
        355                 360                 365

Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu Gly Lys Arg Gly Pro Asn
    370                 375                 380

Gly Glu Ala Gly Ser Ala Gly Pro Pro Gly Pro Pro Gly Leu Arg Gly
385                 390                 395                 400

Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala Asp Gly Arg Ala Gly Val
                405                 410                 415

Met Gly Pro Pro Gly Ser Arg Gly Ala Ser Gly Pro Ala Gly Val Arg
            420                 425                 430

Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly Glu Pro Gly Leu Met Gly
        435                 440                 445

Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn Ile Gly Pro Ala Gly Lys
    450                 455                 460

Glu Gly Pro Val Gly Leu Pro Gly Ile Asp Gly Arg Pro Gly Pro Ile
465                 470                 475                 480

Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly Asn Ile Gly Phe Pro Gly
                485                 490                 495

Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys Asn Gly Asp Lys Gly His
            500                 505                 510

Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro Gly Pro Asp Gly Asn Asn
        515                 520                 525

Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly
    530                 535                 540

Glu Gln Gly Pro Pro Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro
545                 550                 555                 560
```

-continued

```
Ser Gly Pro Ala Gly Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His
               565                 570                 575

Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly Pro Arg Gly Glu Arg Gly
            580                 585                 590

Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro Thr Gly Pro Ile Gly Ser
        595                 600                 605

Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly Asn Lys Gly Glu Pro
    610                 615                 620

Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly Pro Ser Gly
625                 630                 635                 640

Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly Glu
               645                 650                 655

Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp
            660                 665                 670

Gly Ala Arg Gly Ala Pro Gly Ala Val Gly Ala Pro Gly Pro Ala Gly
        675                 680                 685

Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro
    690                 695                 700

Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala
705                 710                 715                 720

Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly Ala Ala Gly Gln Pro Gly
               725                 730                 735

Ala Lys Gly Glu Arg Gly Ala Lys Gly Pro Lys Gly Glu Asn Gly Val
            740                 745                 750

Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly Pro Ala Gly Pro Asn
        755                 760                 765

Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly Asp Gly Gly Pro Pro Gly
    770                 775                 780

Met Thr Gly Phe Pro Gly Ala Ala Gly Arg Thr Gly Pro Pro Gly Pro
785                 790                 795                 800

Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys Glu
               805                 810                 815

Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly Pro Val Gly Arg Thr Gly
            820                 825                 830

Glu Val Gly Ala Val Gly Pro Pro Gly Phe Ala Gly Glu Lys Gly Pro
        835                 840                 845

Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro Gly Thr Pro Gly Pro Gln
    850                 855                 860

Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly Leu Pro Gly Ser Arg Gly
865                 870                 875                 880

Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly Pro
               885                 890                 895

Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala Val
            900                 905                 910

Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp Gly
        915                 920                 925

Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly His
    930                 935                 940

Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala Ala
945                 950                 955                 960

Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His Gly
```

-continued

```
               965                 970                 975

Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly Ala
            980                 985                 990

Val Gly Pro Arg Gly Pro Ser Gly  Pro Gln Gly Ile Arg  Gly Asp Lys
        995                 1000                1005

Gly Glu  Pro Gly Glu Lys Gly  Pro Arg Gly Leu Pro  Gly Leu Lys
    1010                1015                1020

Gly His  Asn Gly Leu Gln Gly  Leu Pro Gly Ile Ala  Gly His His
    1025                1030                1035

Gly Asp  Gln Gly Ala Pro Gly  Ser Val Gly Pro Ala  Gly Pro Arg
    1040                1045                1050

Gly Pro  Ala Gly Pro Ser Gly  Pro Ala Gly Lys Asp  Gly Arg Thr
    1055                1060                1065

Gly His  Pro Gly Thr Val Gly  Pro Ala Gly Ile Arg  Gly Pro Gln
    1070                1075                1080

Gly His  Gln Gly Pro Ala Gly  Pro Pro Gly Pro Pro  Gly Pro Pro
    1085                1090                1095

Gly Pro  Pro Gly Val Ser Gly  Gly Gly Tyr Asp Phe  Gly Tyr Asp
    1100                1105                1110

Gly Asp  Phe Tyr Arg Ala Asp  Gln Pro Arg Ser Ala  Pro Ser Leu
    1115                1120                1125

Arg Pro  Lys Asp Tyr Glu Val  Asp Ala Thr Leu Lys  Ser Leu Asn
    1130                1135                1140

Asn Gln  Ile Glu Thr Leu Leu  Thr Pro Glu Gly Ser  Arg Lys Asn
    1145                1150                1155

Pro Ala  Arg Thr Cys Arg Asp  Leu Arg Leu Ser His  Pro Glu Trp
    1160                1165                1170

Ser Ser  Gly Tyr Tyr Trp Ile  Asp Pro Asn Gln Gly  Cys Thr Met
    1175                1180                1185

Asp Ala  Ile Lys Val Tyr Cys  Asp Phe Ser Thr Gly  Glu Thr Cys
    1190                1195                1200

Ile Arg  Ala Gln Pro Glu Asn  Ile Pro Ala Lys Asn  Trp Tyr Arg
    1205                1210                1215

Ser Ser  Lys Asp Lys Lys His  Val Trp Leu Gly Glu  Thr Ile Asn
    1220                1225                1230

Ala Gly  Ser Gln Phe Glu Tyr  Asn Val Glu Gly Val  Thr Ser Lys
    1235                1240                1245

Glu Met  Ala Thr Gln Leu Ala  Phe Met Arg Leu Leu  Ala Asn Tyr
    1250                1255                1260

Ala Ser  Gln Asn Ile Thr Tyr  His Cys Lys Asn Ser  Ile Ala Tyr
    1265                1270                1275

Met Asp  Glu Glu Thr Gly Asn  Leu Lys Lys Ala Val  Ile Leu Gln
    1280                1285                1290

Gly Ser  Asn Asp Val Glu Leu  Val Ala Glu Gly Asn  Ser Arg Phe
    1295                1300                1305

Thr Tyr  Thr Val Leu Val Asp  Gly Cys Ser Lys Lys  Thr Asn Glu
    1310                1315                1320

Trp Gly  Lys Thr Ile Ile Glu  Tyr Lys Thr Asn Lys  Pro Ser Arg
    1325                1330                1335

Leu Pro  Phe Leu Asp Ile Ala  Pro Leu Asp Ile Gly  Gly Ala Asp
    1340                1345                1350
```

-continued

```
Gln Glu  Phe Phe Val Asp Ile  Gly Pro Val Cys Phe  Lys
    1355                1360                1365


<210> SEQ ID NO 5
<211> LENGTH: 2722
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (119)..(1723)

<400> SEQUENCE: 5

gagcgggctg agggtaggaa gtagccgctc cgagtggagg cgactggggg ctgaagagcg     60

cgccgccctc tcgtcccact ttccaggtgt gtgatcctgt aaaattaaat cttccaag      118

atg atc tgg tat ata tta att ata gga att ctg ctt ccc cag tct ttg      166
Met Ile Trp Tyr Ile Leu Ile Ile Gly Ile Leu Leu Pro Gln Ser Leu
1               5                  10                  15

gct cat cca ggc ttt ttt act tca att ggt cag atg act gat ttg atc      214
Ala His Pro Gly Phe Phe Thr Ser Ile Gly Gln Met Thr Asp Leu Ile
            20                  25                  30

cat act gag aaa gat ctg gtg act tct ctg aaa gat tat att aag gca      262
His Thr Glu Lys Asp Leu Val Thr Ser Leu Lys Asp Tyr Ile Lys Ala
        35                  40                  45

gaa gag gac aag tta gaa caa ata aaa aaa tgg gca gag aag tta gat      310
Glu Glu Asp Lys Leu Glu Gln Ile Lys Lys Trp Ala Glu Lys Leu Asp
    50                  55                  60

cgg cta act agt aca gcg aca aaa gat cca gaa gga ttt gtt ggg cat      358
Arg Leu Thr Ser Thr Ala Thr Lys Asp Pro Glu Gly Phe Val Gly His
65                  70                  75                  80

cca gta aat gca ttc aaa tta atg aaa cgt ctg aat act gag tgg agt      406
Pro Val Asn Ala Phe Lys Leu Met Lys Arg Leu Asn Thr Glu Trp Ser
                85                  90                  95

gag ttg gag aat ctg gtc ctt aag gat atg tca gat ggc ttt atc tct      454
Glu Leu Glu Asn Leu Val Leu Lys Asp Met Ser Asp Gly Phe Ile Ser
            100                 105                 110

aac cta acc att cag aga cca gta ctt tct aat gat gaa gat cag gtt      502
Asn Leu Thr Ile Gln Arg Pro Val Leu Ser Asn Asp Glu Asp Gln Val
        115                 120                 125

ggg gca gcc aaa gct ctg tta cgt ctc cag gat acc tac aat ttg gat      550
Gly Ala Ala Lys Ala Leu Leu Arg Leu Gln Asp Thr Tyr Asn Leu Asp
    130                 135                 140

aca gat acc atc tca aag ggt aat ctt cca gga gtg aaa cac aaa tct      598
Thr Asp Thr Ile Ser Lys Gly Asn Leu Pro Gly Val Lys His Lys Ser
145                 150                 155                 160

ttt cta acg gct gag gac tgc ttt gag ttg ggc aaa gtg gcc tat aca      646
Phe Leu Thr Ala Glu Asp Cys Phe Glu Leu Gly Lys Val Ala Tyr Thr
                165                 170                 175

gaa gca gat tat tac cat acg gaa ctg tgg atg gaa caa gcc cta agg      694
Glu Ala Asp Tyr Tyr His Thr Glu Leu Trp Met Glu Gln Ala Leu Arg
            180                 185                 190

caa ctg gat gaa ggc gag att tct acc ata gat aaa gtc tct gtt cta      742
Gln Leu Asp Glu Gly Glu Ile Ser Thr Ile Asp Lys Val Ser Val Leu
        195                 200                 205

gat tat ttg agc tat gcg gta tat cag cag gga gac ctg gat aag gca      790
Asp Tyr Leu Ser Tyr Ala Val Tyr Gln Gln Gly Asp Leu Asp Lys Ala
    210                 215                 220

ctt ttg ctc aca aag aag ctt ctt gaa cta gat cct gaa cat cag aga      838
```

-continued

```
Leu Leu Leu Thr Lys Lys Leu Leu Glu Leu Asp Pro Glu His Gln Arg
225                 230                 235                 240

gct aat ggt aac tta aaa tat ttt gag tat ata atg gct aaa gaa aaa      886
Ala Asn Gly Asn Leu Lys Tyr Phe Glu Tyr Ile Met Ala Lys Glu Lys
                245                 250                 255

gat gtc aat aag tct gct tca gat gac caa tct gat cag aaa act aca      934
Asp Val Asn Lys Ser Ala Ser Asp Asp Gln Ser Asp Gln Lys Thr Thr
            260                 265                 270

cca aag aaa aaa ggg gtt gct gtg gat tac ctg cca gag aga cag aag      982
Pro Lys Lys Lys Gly Val Ala Val Asp Tyr Leu Pro Glu Arg Gln Lys
        275                 280                 285

tac gaa atg ctg tgc cgt ggg gag ggt atc aaa atg acc cct cgg aga     1030
Tyr Glu Met Leu Cys Arg Gly Glu Gly Ile Lys Met Thr Pro Arg Arg
    290                 295                 300

cag aaa aaa ctc ttt tgc cgc tac cat gat gga aac cgt aat cct aaa     1078
Gln Lys Lys Leu Phe Cys Arg Tyr His Asp Gly Asn Arg Asn Pro Lys
305                 310                 315                 320

ttt att ctg gct cca gct aaa cag gag gat gaa tgg gac aag cct cgt     1126
Phe Ile Leu Ala Pro Ala Lys Gln Glu Asp Glu Trp Asp Lys Pro Arg
                325                 330                 335

att att cgc ttc cat gat att att tct gat gca gaa att gaa atc gtc     1174
Ile Ile Arg Phe His Asp Ile Ile Ser Asp Ala Glu Ile Glu Ile Val
            340                 345                 350

aaa gac cta gca aaa cca agg ctg agc cga gct aca gta cat gac cct     1222
Lys Asp Leu Ala Lys Pro Arg Leu Ser Arg Ala Thr Val His Asp Pro
        355                 360                 365

gag act gga aaa ttg acc aca gca cag tac aga gta tct aag agt gcc     1270
Glu Thr Gly Lys Leu Thr Thr Ala Gln Tyr Arg Val Ser Lys Ser Ala
    370                 375                 380

tgg ctc tct ggc tat gaa aat cct gtg gtg tct cga att aat atg aga     1318
Trp Leu Ser Gly Tyr Glu Asn Pro Val Val Ser Arg Ile Asn Met Arg
385                 390                 395                 400

ata caa gat cta aca gga cta gat gtt tcc aca gca gag gaa tta cag     1366
Ile Gln Asp Leu Thr Gly Leu Asp Val Ser Thr Ala Glu Glu Leu Gln
                405                 410                 415

gta gca aat tat gga gtt gga gga cag tat gaa ccc cat ttt gac ttt     1414
Val Ala Asn Tyr Gly Val Gly Gly Gln Tyr Glu Pro His Phe Asp Phe
            420                 425                 430

gca cgg aaa gat gag cca gat gct ttc aaa gag ctg ggg aca gga aat     1462
Ala Arg Lys Asp Glu Pro Asp Ala Phe Lys Glu Leu Gly Thr Gly Asn
        435                 440                 445

aga att gct aca tgg ctg ttt tat atg agt gat gtg tct gca gga gga     1510
Arg Ile Ala Thr Trp Leu Phe Tyr Met Ser Asp Val Ser Ala Gly Gly
    450                 455                 460

gcc act gtt ttt cct gaa gtt gga gct agt gtt tgg ccc aaa aaa gga     1558
Ala Thr Val Phe Pro Glu Val Gly Ala Ser Val Trp Pro Lys Lys Gly
465                 470                 475                 480

act gct gtt ttc tgg tat aat ctg ttt gcc agt gga gaa gga gat tat     1606
Thr Ala Val Phe Trp Tyr Asn Leu Phe Ala Ser Gly Glu Gly Asp Tyr
                485                 490                 495

agt aca cgg cat gca gcc tgt cca gtg cta gtt ggc aac aaa tgg gta     1654
Ser Thr Arg His Ala Ala Cys Pro Val Leu Val Gly Asn Lys Trp Val
            500                 505                 510

tcc aat aaa tgg ctc cat gaa cgt gga caa gaa ttt cga aga cct tgt     1702
Ser Asn Lys Trp Leu His Glu Arg Gly Gln Glu Phe Arg Arg Pro Cys
        515                 520                 525

acg ttg tca gaa ttg gaa tga caaacaggct tccctttttc tcctattgtt        1753
```

-continued

```
Thr Leu Ser Glu Leu Glu
    530

gtactcttat gtgtctgata tacacatttc catagtctta actttcagga gtttacaatt   1813

gactaacact ccatgattga ttcagtcatg aacctcatcc catgtttcat ctgtggacaa   1873

ttgcttactt tgtgggttct tttaaaagta acacgaaatc atcatattgc ataaaacctt   1933

aaagttctgt tggtatcaca gaagacaagg cagagtttaa agtgaggaat tttatattta   1993

aagaactttt tggttggata aaaacataat ttgagcatcc agttttagta tttcactaca   2053

tctcagttgg tgggtgttaa gctagaatgg gctgtgtgat aggaaacaaa tgccttacag   2113

atgtgcctag gtgttctgtt tacctagtgt cttactctgt tttctggatc tgaagactag   2173

taataaacta ggacactaac tgggttccat gtgattgccc tttcatatga tcttctaagt   2233

tgatttttt cctcccaagt ctttttaaa gaaagtatac tgtattttac caacccccctc   2293

tcttttcttt tagctcctct gtggtgaatt aaacgtactt gagttaaaat atttcgattt   2353

ttttttttt tttaatggaa agtcctgcat aacaacactg ggccttctta actaaaatgc   2413

tcaccactta gcctgttttt ttatcccttt tttaaaatga cagatgattt tgttcaggaa   2473

ttttgctgtt tttcttagtg ctaatacctt gcctcttatt cctgctacag cagggtggta   2533

atattggcat tctgattaaa tactgtgcct taggagactg gaagtttaaa aatgtacaag   2593

tcctttcagt gatgagggaa ttgatttttt ttaaaagtct ttttcttaga aagccaaaat   2653

gtttgttttt ttaagattct gaaatgtgtt gtgacaacaa tgacctattt atgatcttaa   2713

atctttttt                                                          2722


<210> SEQ ID NO 6
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

Met Ile Trp Tyr Ile Leu Ile Ile Gly Ile Leu Leu Pro Gln Ser Leu
1               5                   10                  15

Ala His Pro Gly Phe Phe Thr Ser Ile Gly Gln Met Thr Asp Leu Ile
            20                  25                  30

His Thr Glu Lys Asp Leu Val Thr Ser Leu Lys Asp Tyr Ile Lys Ala
        35                  40                  45

Glu Glu Asp Lys Leu Glu Gln Ile Lys Lys Trp Ala Glu Lys Leu Asp
    50                  55                  60

Arg Leu Thr Ser Thr Ala Thr Lys Asp Pro Glu Gly Phe Val Gly His
65                  70                  75                  80

Pro Val Asn Ala Phe Lys Leu Met Lys Arg Leu Asn Thr Glu Trp Ser
                85                  90                  95

Glu Leu Glu Asn Leu Val Leu Lys Asp Met Ser Asp Gly Phe Ile Ser
            100                 105                 110

Asn Leu Thr Ile Gln Arg Pro Val Leu Ser Asn Asp Glu Asp Gln Val
        115                 120                 125

Gly Ala Ala Lys Ala Leu Leu Arg Leu Gln Asp Thr Tyr Asn Leu Asp
    130                 135                 140

Thr Asp Thr Ile Ser Lys Gly Asn Leu Pro Gly Val Lys His Lys Ser
145                 150                 155                 160

Phe Leu Thr Ala Glu Asp Cys Phe Glu Leu Gly Lys Val Ala Tyr Thr
```

-continued

```
                165                 170                 175

Glu Ala Asp Tyr Tyr His Thr Glu Leu Trp Met Glu Gln Ala Leu Arg
            180                 185                 190

Gln Leu Asp Glu Gly Glu Ile Ser Thr Ile Asp Lys Val Ser Val Leu
        195                 200                 205

Asp Tyr Leu Ser Tyr Ala Val Tyr Gln Gln Gly Asp Leu Asp Lys Ala
    210                 215                 220

Leu Leu Leu Thr Lys Lys Leu Leu Glu Leu Asp Pro Glu His Gln Arg
225                 230                 235                 240

Ala Asn Gly Asn Leu Lys Tyr Phe Glu Tyr Ile Met Ala Lys Glu Lys
                245                 250                 255

Asp Val Asn Lys Ser Ala Ser Asp Asp Gln Ser Asp Gln Lys Thr Thr
            260                 265                 270

Pro Lys Lys Lys Gly Val Ala Val Asp Tyr Leu Pro Glu Arg Gln Lys
        275                 280                 285

Tyr Glu Met Leu Cys Arg Gly Glu Gly Ile Lys Met Thr Pro Arg Arg
    290                 295                 300

Gln Lys Lys Leu Phe Cys Arg Tyr His Asp Gly Asn Arg Asn Pro Lys
305                 310                 315                 320

Phe Ile Leu Ala Pro Ala Lys Gln Glu Asp Glu Trp Asp Lys Pro Arg
                325                 330                 335

Ile Ile Arg Phe His Asp Ile Ile Ser Asp Ala Glu Ile Glu Ile Val
            340                 345                 350

Lys Asp Leu Ala Lys Pro Arg Leu Ser Arg Ala Thr Val His Asp Pro
        355                 360                 365

Glu Thr Gly Lys Leu Thr Thr Ala Gln Tyr Arg Val Ser Lys Ser Ala
    370                 375                 380

Trp Leu Ser Gly Tyr Glu Asn Pro Val Val Ser Arg Ile Asn Met Arg
385                 390                 395                 400

Ile Gln Asp Leu Thr Gly Leu Asp Val Ser Thr Ala Glu Glu Leu Gln
                405                 410                 415

Val Ala Asn Tyr Gly Val Gly Gly Gln Tyr Glu Pro His Phe Asp Phe
            420                 425                 430

Ala Arg Lys Asp Glu Pro Asp Ala Phe Lys Glu Leu Gly Thr Gly Asn
        435                 440                 445

Arg Ile Ala Thr Trp Leu Phe Tyr Met Ser Asp Val Ser Ala Gly Gly
    450                 455                 460

Ala Thr Val Phe Pro Glu Val Gly Ala Ser Val Trp Pro Lys Lys Gly
465                 470                 475                 480

Thr Ala Val Phe Trp Tyr Asn Leu Phe Ala Ser Gly Glu Gly Asp Tyr
                485                 490                 495

Ser Thr Arg His Ala Ala Cys Pro Val Leu Val Gly Asn Lys Trp Val
            500                 505                 510

Ser Asn Lys Trp Leu His Glu Arg Gly Gln Glu Phe Arg Arg Pro Cys
        515                 520                 525

Thr Leu Ser Glu Leu Glu
    530
```

<210> SEQ ID NO 7
<211> LENGTH: 1956
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

-continued

```
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (30)..(1556)
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (63)..(63)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 7

ccgagcgccc cgcctgctcc gtgtccgac atg ctg cgc cgc gct ctg ctg tgc       53

                                    Met Leu Arg Arg Ala Leu Leu Cys
                                    1               5

ctg ccg tgg ncc gcc ctg gtg cgc gcc gac gcc ccc gag gag gag gac      101
Leu Pro Trp Xaa Ala Leu Val Arg Ala Asp Ala Pro Glu Glu Glu Asp
    10                  15                  20

cac gtc ttg gtg ctg cgg aaa agc aac ttc gcg gag gcg ctg gcg gcc      149
His Val Leu Val Leu Arg Lys Ser Asn Phe Ala Glu Ala Leu Ala Ala
25                  30                  35                  40

cac aag tac ccg ccg gtg gag ttc cat gcc ccc tgg tgt ggc cac tgc      197
His Lys Tyr Pro Pro Val Glu Phe His Ala Pro Trp Cys Gly His Cys
                45                  50                  55

aag gct ctg gcc cct gag tat gcc aaa gcc gct ggg aag ctg aag gca      245
Lys Ala Leu Ala Pro Glu Tyr Ala Lys Ala Ala Gly Lys Leu Lys Ala
            60                  65                  70

gaa ggt tcc gag atc agg ttg gcc aag gtg gac gcc acg gag gag tct      293
Glu Gly Ser Glu Ile Arg Leu Ala Lys Val Asp Ala Thr Glu Glu Ser
        75                  80                  85

gac cta gcc cag cag tac ggc gtg cgc ggc tat ccc acc atc aag ttc      341
Asp Leu Ala Gln Gln Tyr Gly Val Arg Gly Tyr Pro Thr Ile Lys Phe
    90                  95                  100

ttc agg aat gga gac acg gct tcc ccc aag gaa tat aca gct ggc aga      389
Phe Arg Asn Gly Asp Thr Ala Ser Pro Lys Glu Tyr Thr Ala Gly Arg
105                 110                 115                 120

gag gct gat gac atc gtg aac tgg ctg aag aag cgc acg ggc ccg gct      437
Glu Ala Asp Asp Ile Val Asn Trp Leu Lys Lys Arg Thr Gly Pro Ala
                125                 130                 135

gcc acc acc ctg cct gac ggc gca gct gca gag tcc ttg gtg gag tcc      485
Ala Thr Thr Leu Pro Asp Gly Ala Ala Ala Glu Ser Leu Val Glu Ser
            140                 145                 150

agc gag gtg gcc gtc atc ggc ttc ttc aag gac gtg gag tcg gac tct      533
Ser Glu Val Ala Val Ile Gly Phe Phe Lys Asp Val Glu Ser Asp Ser
        155                 160                 165

gcc aag cag ttt ttg cag gca gca gag gcc atc gat gac ata cca ttt      581
Ala Lys Gln Phe Leu Gln Ala Ala Glu Ala Ile Asp Asp Ile Pro Phe
    170                 175                 180

ggg atc act tcc aac agt gac gtg ttc tcc aaa tac cag ctc gac aaa      629
Gly Ile Thr Ser Asn Ser Asp Val Phe Ser Lys Tyr Gln Leu Asp Lys
185                 190                 195                 200

gat ggg gtt gtc ctc ttt aag aag ttt gat gaa ggc cgg aac aac ttt      677
Asp Gly Val Val Leu Phe Lys Lys Phe Asp Glu Gly Arg Asn Asn Phe
                205                 210                 215

gaa ggg gag gtc acc aag gag aac ctg ctg gac ttt atc aaa cac aac      725
Glu Gly Glu Val Thr Lys Glu Asn Leu Leu Asp Phe Ile Lys His Asn
            220                 225                 230

cag ctg ccc ctt gtc atc gag ttc acc gag cag aca gcc ccg aag att      773
Gln Leu Pro Leu Val Ile Glu Phe Thr Glu Gln Thr Ala Pro Lys Ile
        235                 240                 245

ttt gga ggt gaa atc aag act cac atc ctg ctg ttc ttg ccc aag agt      821
```

-continued

```
Phe Gly Gly Glu Ile Lys Thr His Ile Leu Leu Phe Leu Pro Lys Ser
    250                 255                 260

gtg tct gac tat gac ggc aaa ctg agc aac ttc aaa aca gca gcc gag      869
Val Ser Asp Tyr Asp Gly Lys Leu Ser Asn Phe Lys Thr Ala Ala Glu
265                 270                 275                 280

agc ttc aag ggc aag atc ctg ttc atc ttc atc gac agc gac cac acc      917
Ser Phe Lys Gly Lys Ile Leu Phe Ile Phe Ile Asp Ser Asp His Thr
                285                 290                 295

gac aac cag cgc atc ctc gag ttc ttt ggc ctg aag aag gaa gag tgc      965
Asp Asn Gln Arg Ile Leu Glu Phe Phe Gly Leu Lys Lys Glu Glu Cys
            300                 305                 310

ccg gcc gtg cgc ctc atc acc ttg gag gag gag atg acc aag tac aag     1013
Pro Ala Val Arg Leu Ile Thr Leu Glu Glu Glu Met Thr Lys Tyr Lys
        315                 320                 325

ccc gaa tcg gag gag ctg acg gca gag agg atc aca gag ttc tgc cac     1061
Pro Glu Ser Glu Glu Leu Thr Ala Glu Arg Ile Thr Glu Phe Cys His
    330                 335                 340

cgc ttc ctg gag ggc aaa atc aag ccc cac ctg atg agc cag gag ctg     1109
Arg Phe Leu Glu Gly Lys Ile Lys Pro His Leu Met Ser Gln Glu Leu
345                 350                 355                 360

ccg gag gac tgg gac aag cag cct gtc aag gtg ctt gtt ggg aag aac     1157
Pro Glu Asp Trp Asp Lys Gln Pro Val Lys Val Leu Val Gly Lys Asn
                365                 370                 375

ttt gaa gac gtg gct ttt gat gag aaa aaa aac gtc ttt gtg gag ttc     1205
Phe Glu Asp Val Ala Phe Asp Glu Lys Lys Asn Val Phe Val Glu Phe
            380                 385                 390

tat gcc cca tgg tgt ggt cac tgc aaa cag ttg gct ccc att tgg gat     1253
Tyr Ala Pro Trp Cys Gly His Cys Lys Gln Leu Ala Pro Ile Trp Asp
        395                 400                 405

aaa ctg gga gag acg tac aag gac cat gag aac atc gtc atc gcc aag     1301
Lys Leu Gly Glu Thr Tyr Lys Asp His Glu Asn Ile Val Ile Ala Lys
    410                 415                 420

atg gac tcg act gcc aac gag gtg gag gcc gtc aaa gtg cac ggc ttc     1349
Met Asp Ser Thr Ala Asn Glu Val Glu Ala Val Lys Val His Gly Phe
425                 430                 435                 440

ccc aca ctc ggg ttc ttt cct gcc agt gcc gac agg acg gtc att gat     1397
Pro Thr Leu Gly Phe Phe Pro Ala Ser Ala Asp Arg Thr Val Ile Asp
                445                 450                 455

tac aac ggg gaa cgc acg ctg gat ggt ttt aag aaa ttc cta gag agc     1445
Tyr Asn Gly Glu Arg Thr Leu Asp Gly Phe Lys Lys Phe Leu Glu Ser
            460                 465                 470

ggt ggc caa gat ggg gca ggg gat gtt gac gac ctc gag gac ctc gaa     1493
Gly Gly Gln Asp Gly Ala Gly Asp Val Asp Asp Leu Glu Asp Leu Glu
        475                 480                 485

gaa gca gag gag cca gac atg gag gaa gac gat gac cag aaa gct gtg     1541
Glu Ala Glu Glu Pro Asp Met Glu Glu Asp Asp Asp Gln Lys Ala Val
    490                 495                 500

aaa gat gaa ctg taa tacgcaaagc cggacccggg cgctgccgag acccctcggg     1596
Lys Asp Glu Leu
505

ggctgcacac ccagcagcag cgcacgcctc cgaagcctgc ggcctcgctt gaaggagggc   1656

gtcgccggaa acccaaggaa cctctctgaa gtgacacctc acccctacac accgtccgtt   1716

caccccccgtc tcttccttct gcttttcggt ttttggaaaa cccggatcct actctaggca   1776

gcccaccttg gtgggcttgt ttcctgaaac catgatgtac ttttttcatac atgagtctgt   1836

ccagagtgct tgctaccgtg ttcggagtct cgctgcctcc ctcccgcggg aggttctcct   1896
```

-continued

cttttttgaaa attccgtccg tgggattttt agacattttt acgacatcag ggtatttgtt 1956

<210> SEQ ID NO 8
<211> LENGTH: 508
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: The 'Xaa' at location 12 stands for Thr, Ala, Pro, or Ser.

<400> SEQUENCE: 8

Met Leu Arg Arg Ala Leu Leu Cys Leu Pro Trp Xaa Ala Leu Val Arg
1 5 10 15

Ala Asp Ala Pro Glu Glu Glu Asp His Val Leu Val Leu Arg Lys Ser
20 25 30

Asn Phe Ala Glu Ala Leu Ala Ala His Lys Tyr Pro Pro Val Glu Phe
35 40 45

His Ala Pro Trp Cys Gly His Cys Lys Ala Leu Ala Pro Glu Tyr Ala
50 55 60

Lys Ala Ala Gly Lys Leu Lys Ala Glu Gly Ser Glu Ile Arg Leu Ala
65 70 75 80

Lys Val Asp Ala Thr Glu Glu Ser Asp Leu Ala Gln Gln Tyr Gly Val
85 90 95

Arg Gly Tyr Pro Thr Ile Lys Phe Phe Arg Asn Gly Asp Thr Ala Ser
100 105 110

Pro Lys Glu Tyr Thr Ala Gly Arg Glu Ala Asp Asp Ile Val Asn Trp
115 120 125

Leu Lys Lys Arg Thr Gly Pro Ala Ala Thr Thr Leu Pro Asp Gly Ala
130 135 140

Ala Ala Glu Ser Leu Val Glu Ser Ser Glu Val Ala Val Ile Gly Phe
145 150 155 160

Phe Lys Asp Val Glu Ser Asp Ser Ala Lys Gln Phe Leu Gln Ala Ala
165 170 175

Glu Ala Ile Asp Asp Ile Pro Phe Gly Ile Thr Ser Asn Ser Asp Val
180 185 190

Phe Ser Lys Tyr Gln Leu Asp Lys Asp Gly Val Val Leu Phe Lys Lys
195 200 205

Phe Asp Glu Gly Arg Asn Asn Phe Glu Gly Glu Val Thr Lys Glu Asn
210 215 220

Leu Leu Asp Phe Ile Lys His Asn Gln Leu Pro Leu Val Ile Glu Phe
225 230 235 240

Thr Glu Gln Thr Ala Pro Lys Ile Phe Gly Gly Glu Ile Lys Thr His
245 250 255

Ile Leu Leu Phe Leu Pro Lys Ser Val Ser Asp Tyr Asp Gly Lys Leu
260 265 270

Ser Asn Phe Lys Thr Ala Ala Glu Ser Phe Lys Gly Lys Ile Leu Phe
275 280 285

Ile Phe Ile Asp Ser Asp His Thr Asp Asn Gln Arg Ile Leu Glu Phe
290 295 300

Phe Gly Leu Lys Lys Glu Glu Cys Pro Ala Val Arg Leu Ile Thr Leu
305 310 315 320

-continued

```
Glu Glu Glu Met Thr Lys Tyr Lys Pro Glu Ser Glu Glu Leu Thr Ala
                325                 330                 335

Glu Arg Ile Thr Glu Phe Cys His Arg Phe Leu Glu Gly Lys Ile Lys
            340                 345                 350

Pro His Leu Met Ser Gln Glu Leu Pro Glu Asp Trp Asp Lys Gln Pro
        355                 360                 365

Val Lys Val Leu Val Gly Lys Asn Phe Glu Asp Val Ala Phe Asp Glu
    370                 375                 380

Lys Lys Asn Val Phe Val Glu Phe Tyr Ala Pro Trp Cys Gly His Cys
385                 390                 395                 400

Lys Gln Leu Ala Pro Ile Trp Asp Lys Leu Gly Glu Thr Tyr Lys Asp
                405                 410                 415

His Glu Asn Ile Val Ile Ala Lys Met Asp Ser Thr Ala Asn Glu Val
            420                 425                 430

Glu Ala Val Lys Val His Gly Phe Pro Thr Leu Gly Phe Phe Pro Ala
        435                 440                 445

Ser Ala Asp Arg Thr Val Ile Asp Tyr Asn Gly Glu Arg Thr Leu Asp
    450                 455                 460

Gly Phe Lys Lys Phe Leu Glu Ser Gly Gly Gln Asp Gly Ala Gly Asp
465                 470                 475                 480

Val Asp Asp Leu Glu Asp Leu Glu Glu Ala Glu Glu Pro Asp Met Glu
                485                 490                 495

Glu Asp Asp Asp Gln Lys Ala Val Lys Asp Glu Leu
            500                 505


<210> SEQ ID NO 9
<211> LENGTH: 90
<212> TYPE: DNA
<213> ORGANISM: Autographa californica nucleopolyhedrovirus

<400> SEQUENCE: 9

acaatatatt atagttaaat aagaattatt atcaaatcat ttgtatatta attaaaatac     60

tatactgtaa attacatttt atttacaatc                                      90


<210> SEQ ID NO 10
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Autographa californica nucleopolyhedrovirus

<400> SEQUENCE: 10

ataaccatct cgcaaataaa taagtatttt actgttttcg taacagtttt gtaataaaaa     60

aacctataaa t                                                          71
```

What is claimed is:

1. A method for producing a recombinant human collagen polypeptide, said method comprising:

(a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising:

(i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and

(ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter different from said first promoter; and

(b) recovering said collagen polypeptide from said host insect cell culture.

2. The method of claim **1**, wherein said first promoter is a p10 promoter.

3. The method of claim **1**, wherein said second promoter is a polyhedron (polH) promoter.

4. The method of claim **1**, wherein said collagen subunit is a first collagen subunit and wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second collagen subunit, operably linked to a first promoter.

5. The method of claim 4, wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a third collagen subunit, operably linked to a first promoter.

6. The method of claim 1, wherein said subunit of a collagen post-translational enzyme is a first subunit of a collagen post-translational enzyme, and wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second subunit of a collagen post-translational enzyme, operably linked to a second promoter.

7. The method of claim 1, wherein said collagen is selected from collagen types I, II and III.

8. The method of claim 4, wherein said collagen is selected from collagen types I, II and III.

9. The method of claim 7 wherein said collagen is type II collagen and said collagen subunit is a collagen α1(II) subunit.

10. The method of claim 7 wherein said collagen is type III collagen and said collagen subunit is a collagen α1(III) subunit

11. The method of claim 8 wherein said collagen is type I collagen, said first collagen subunit is a collagen α1(I) subunit and said second collagen subunit is a collagen α2(I) subunit.

12. The method of claim 1, wherein said collagen post-translational enzyme is selected from prolyl hydroxylase, lysyl oxidase and lysyl hydroxylase.

13. The method of claim 12, wherein said collagen post-translational enzyme is prolyl 4-hydroxylase

14. The method of claim 6, wherein said collagen post-translational enzyme is prolyl 4-hydroxylase and wherein said first subunit of a collagen post-translational enzyme is an alpha subunit of prolyl 4-hydroxylase and wherein said second subunit of a collagen post-translational enzyme is a beta subunit of prolyl 4-hydroxylase.

15. A method for producing a recombinant human procollagen polypeptide, said method comprising:

(a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising:
  (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and
  (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter; and

(b) recovering said procollagen polypeptide from said host insect cell culture.

16. The method of claim 15, wherein said first promoter is a p10 promoter.

17. The method of claim 15, wherein said second promoter is a polyhedron (polH) promoter.

18. The method of claim 15, wherein said collagen subunit is a first collagen subunit and wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second collagen subunit, operably linked to a first promoter.

19. The method of claim 18, wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a third collagen subunit, operably linked to a first promoter.

20. The method of claim 15, wherein said subunit of a collagen post-translational enzyme is a first subunit of a collagen post-translational enzyme, and wherein said recombinant baculovirus expression vector further comprises a nucleotide sequence which encodes a second subunit of a collagen post-translational enzyme, operably linked to a polH promoter.

21. The method of claim 15, wherein said collagen is selected from collagen types I, II and III.

22. The method of claim 18, wherein said collagen is selected from collagen types I, II and III.

23. The method of claim 21 wherein said collagen is type II collagen and said collagen subunit is a collagen α1(II) subunit.

24. The method of claim 21 wherein said collagen is type III collagen and said collagen subunit is a collagen α1(III) subunit.

25. The method of claim 22 wherein said collagen is type I collagen, said first collagen subunit is a collagen α1(I) subunit and said second collagen subunit is a collagen α2(I) subunit.

26. The method of claim 15, wherein said collagen post-translational enzyme is selected from prolyl hydroxylase, lysyl oxidase and lysyl hydroxylase.

27. The method of claim 26, wherein said collagen post-translational enzyme is prolyl 4-hydroxylase

28. The method of claim 20, wherein said collagen post-translational enzyme is prolyl 4-hydroxylase and wherein said first subunit of a collagen post-translational enzyme is an alpha subunit of prolyl 4-hydroxylase and wherein said second subunit of a collagen post-translational enzyme is a beta subunit of prolyl 4-hydroxylase.

29. The method of claim 1, wherein said infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector is obtained by a method comprising:

(a) transfecting or transforming a first host insect cell with baculovirus DNA and an expression vector comprising:
  (i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and
  (ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter;
  thereby to permit integration of said expression vector into said baculovirus DNA to obtain a recombinant baculovirus expression vector;

(b) isolating a nucleic acid molecule comprising said recombinant baculovirus expression vector from said host cell; and

(c) transfecting or transforming a second host insect cell with said nucleic acid molecule obtained in (b) thereby to obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector.

30. The method of claim 29, further comprising:

(d) culturing said infected, transfected or transformed host insect cell obtained in (c) under conditions suitable for production of recombinant baculovirus; and

(e) infecting a third host insect cell with the recombinant baculovirus obtained in (d), thereby to obtain an infected, transfected or transformed host insect cell comprising said recombinant baculovirus expression vector.

31. The method of claim 29, wherein said first promoter is a p10 promoter.

32. The method of claim 29, wherein said second promoter is a polyhedron (polH) promoter.

33. A recombinant collagen polypeptide obtained by the method of claim **1**.

34. A recombinant procollagen polypeptide obtained by the method of claim **15**.

35. A recombinant baculovirus expression vector comprising:

(i) a nucleotide sequence which encodes a collagen subunit, operably linked to a first promoter; and

(ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a second promoter.

36. The vector of claim **31**, wherein said first promoter is a p10 promoter.

37. The vector of claim **31**, wherein said second promoter is a polyhedron (polH) promoter.

38. A host insect cell which has been infected, transfected or transformed with the recombinant baculovirus expression vector of claim **31**.

39. A method for producing a recombinant human collagen polypeptide, said method comprising:

(a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising:

(i) a nucleotide sequence which encodes a collagen subunit, operably linked to a p10 promoter;

(ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a polH promoter; and

(b) recovering said collagen polypeptide from said host insect cell culture.

40. A method for producing a recombinant human procollagen polypeptide, said method comprising:

(a) culturing a host insect cell, wherein said insect cell has been infected, transfected or transformed with a recombinant baculovirus expression vector comprising:

(i) a nucleotide sequence which encodes a collagen subunit, operably linked to a p10 promoter;

(ii) a nucleotide sequence which encodes a collagen post-translational enzyme or subunit thereof, operably linked to a polH promoter; and

(b) recovering said procollagen polypeptide from said host insect cell culture.

41. A method of enhancing the purity of a collagen preparation, said method comprising incubating the collagen preparation under basic conditions such that the collagen is rendered insoluble in the basic solution, and recovering the insoluble collagen.

42. The method of claim **37**, wherein said method comprises dialyzing the collagen preparation against a basic solution.

43. A method of preparing collagen or processing a procollagen, said method comprising treating a procollagen sample with an elastase.

* * * * *