United States Patent [19]

Lin

[11] **Patent Number: 4,703,008**

[45] **Date of Patent: Oct. 27, 1987**

[54] **DNA SEQUENCES ENCODING ERYTHROPOIETIN**

[75] Inventor: **Fu-Kuen Lin**, Thousand Oaks, Calif.

[73] Assignee: **Kiren-Amgen, Inc.**, Thousand Oaks, Calif.

[21] Appl. No.: **675,298**

[22] Filed: **Nov. 30, 1984**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 561,024, Dec. 13, 1983, abandoned, and a continuation-in-part of Ser. No. 582,185, Feb. 21, 1984, abandoned, and a continuation-in-part of Ser. No. 655,841, Sep. 28, 1984.

[51] **Int. Cl.**[4] **C12N 5/00;** C12N 15/00; C12N 1/20; C12N 1/00; C12Q 1/68; C07H 15/12

[52] **U.S. Cl.** **435/240.2;** 435/172.3; 435/253; 435/6; 435/317; 435/320; 536/27; 935/9; 935/10; 935/13; 935/79; 935/80

[58] **Field of Search** 435/68, 317, 172.3, 435/253, 240; 935/6, 10, 11, 27, 69, 73, 13

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,033,753	5/1962	White et al. .	
3,865,801	2/1975	Chiba et al. .	
4,237,224	12/1980	Cohen et al. .	
4,264,731	4/1981	Shine .	
4,273,875	6/1981	Manis .	
4,293,652	10/1981	Cohen .	
4,338,397	7/1982	Gilbert et al. .	
4,358,535	11/1982	Falkow et al. .	
4,377,513	3/1983	Sugimoto et al.	260/112
4,394,443	7/1983	Weissman et al. .	
4,397,840	8/1983	Takezawa et al. .	
4,399,216	8/1983	Axel et al. .	
4,411,994	10/1983	Gilbert et al. .	
4,442,205	4/1984	Hamer et al. .	
4,465,624	8/1984	Chiba et al. .	
4,468,464	8/1984	Cohen et al.	435/317
4,468,466	8/1984	Cohen et al. .	
4,503,151	3/1985	Paddock	435/68
4,558,005	12/1985	Goldwasser et al. .	
4,558,006	12/1985	Egrie	435/70
4,568,488	2/1986	Lee-Huang .	

FOREIGN PATENT DOCUMENTS

0070685	1/1983	European Pat. Off. .
0070687	1/1983	European Pat. Off. .
0077670	4/1983	European Pat. Off. .
0093619	11/1983	European Pat. Off. .
0123294	4/1984	European Pat. Off. .
0116446	8/1984	European Pat. Off. .

(List continued on next page.)

OTHER PUBLICATIONS

Bennetzen et al., 1982, "Codon Selection in Yeast", *J. Biol Chem,* vol. 257(6), 3026-3031.
Lewin *Genes,* 1983, John Wiley & Sons, p. 307.

(List continued on next page.)

Primary Examiner—Alvin E. Tanenholtz
Attorney, Agent, or Firm—Michael F. Borun; Steven M. Odre

[57] **ABSTRACT**

Disclosed are novel polypeptides possessing part or all of the primary structural conformation and one or more of the biological properties of mammalian erythropoietin ("EPO") which are characterized in preferred forms by being the product of procaryotic or eucaryotic host expression of an exogenous DNA sequence. Illustratively, genomic DNA, cDNA and manufactured DNA sequences coding for part or all of the sequence of amino acid residues of EPO or for analogs thereof are incorporated into autonomously replicating plasmid or viral vectors employed to transform or transfect suitable procaryotic or eucaryotic host cells such as bacteria, yeast or vertebrate cells in culture. Upon isolation from culture media or cellular lysates or fragments, products of expression of the DNA sequences display, e.g., the immunological properties and in vitro and in vivo biological activities of EPO of human or monkey species origins. Disclosed also are chemically synthesized polypeptides sharing the biochemical and immunological properties of EPO. Also disclosed are improved methods for the detection of specific single stranded polynucleotides in a heterologus cellular or viral sample prepared from, e.g., DNA present in a plasmid or viral-borne cDNA or genomic DNA "library".

31 Claims, 21 Drawing Figures

Translation of Monkey EPO cDNA

```
Sau3A
GATCCCCGCGCCCCCTGGACAGCCGCCCTCTCCTCCAGGCCCGTGGGGCTGGCCCTGCCC

CGCTGAACTTCCCGGGATGAGGACTCCCGGTGTGGTCACCGCGCGCCTAGGTCGCTGAG

                        -27                             -20
                    Met Gly Val His Glu Cys Pro Ala Trp
GGACCCCGGCCAGGCGCGGAGATG GGG GTG CAC GAA TGT CCT GCC TGG

                                -10
Leu Trp Leu Leu Leu Ser Leu Val Ser Leu Pro Leu Gly Leu Pro
CTG TGG CTT CTC CTG TCT CTC GTG TCG CTC CCT CTG GGC CTC CCA

          -1  +1                                        10
Val Pro Gly Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val Leu
GTC CCG GGC GCC CCA CCA CGC CTC ATC TGT GAC AGC CGA GTC CTG

                            20                  *
Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Val Thr Met
GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT GTC ACG ATG

                    30                                  *       40
Gly Cys Ser Glu Ser Cys Ser Leu Asn Glu Asn Ile Thr Val Pro
GGC TGT TCC GAA AGC TGC AGC TTG AAT GAG AAT ATC ACC GTC CCA
```

FOREIGN PATENT DOCUMENTS

0117058	8/1984	European Pat. Off. .
0117059	8/1984	European Pat. Off. .
0117060	8/1984	European Pat. Off. .
0136490	4/1985	European Pat. Off. .
85/04419	10/1985	European Pat. Off. .
83/04053	11/1983	PCT Int'l Appl. .
85/03079	7/1985	PCT Int'l Appl. .
2085887	5/1982	United Kingdom .

OTHER PUBLICATIONS

Young et al., PNAS vol. 80, pp. 1194–1198, Mar. 1983.
Broome et al., PNAS, vol 75, pp. 2746–2749, Jun. 1978.
Ullrich et al., Science, vol. 196, pp. 1313–1319, Jun. 17, 1977.
Mantial et al., Science, vol. 205, pp. 602–606, Aug. 10, 1979.
Talmadge et al., PNAS USA 77, pp. 3369–3373, Jun. 1980.
Walker et al., Techniques In Molecular Biology, Macmillan Pub. Co., New York, p. 280 (1983).
Kennell, Prog. Nucl. Acid Res. Mol. Biol. 11, 259–301, p. 293 (1971).
Breslow et al., PNAS USA 79, pp. 6861–6865, Nov. 1982.
Woods et al., PNAS USA 79, pp. 5661–5665, Sep. 1982.
Lee-Huang 1984, "Cloning and Expression of Human Erythropoietin cDNA in *E. Coli*", *PNAS*, vol. 81, pp. 2708–2712.
Lin et al., 1984, "Cloning of the Monkey Erythropoietin Gene" (Abstract), *J Cell Bioch.*, Suppl. 8 B, p. 45.
Farber et al., 1983, "Translation of mRNA from Anemic Baboon Kidney Into Biologically Active Erythropoietin", Exp. Hematol, vol. 11, Suppl. 14, Abstr. 101.
Gouy et al., 1982, "Coden Usage in Bacteria:Correlation with Gene Expressivity", *Nucleic Acids Res.*, vol. 10, 7055–7074.
Lin et al., *Exp. Hematol.*, 12, 357 (1984).
Lee-Huang, *Blood*, 56 620–624 (1980).
Adamson, *Hosp. Practice*, 18(12), 49–57 (1983).
Anderson et al., *P.N.A.S. (USA)*, 80, pp. 6838–6842 (1983).
Baciu et al., *Ann. N.Y. Acad. Sci.*, 414, pp. 66–72 (1983).
Baron et al., *Cell*, 28, 395–404 (1982).
Beaucage et al., *Tetrahedron Letters*, 22, pp. 1859–1862 (1981).
Billat et al., *Expt. Hematol.*, 10(1), 133–140 (1982).
Blattner et al., Science, 196, pp. 161–169 (1977).
Chisholm, *High Technology*, vol. 2, No. 1, pp. 57–63 (1983).
Choo et al., *Nature*, 299, pp. 178–180 (1982).
Choppin et al., *Blood*, 64(2), 341–347 (1984).
Chou et al., *Biochem.*, 13, 222–245 (1974).
Chou et al., *Advances in Enzymology*, 47, 45–47, (1978).
Chou et al., *Ann. Rev. Biochem.*, 47, 251–277, (1978).
Claus-Walker et al., *Arch. Phys. Med. Rehabil.*, 65, 370–374 (1984).
Congote, *Biochem. Biophys. Res. Comm., 115(2), 477–483 (1983).*

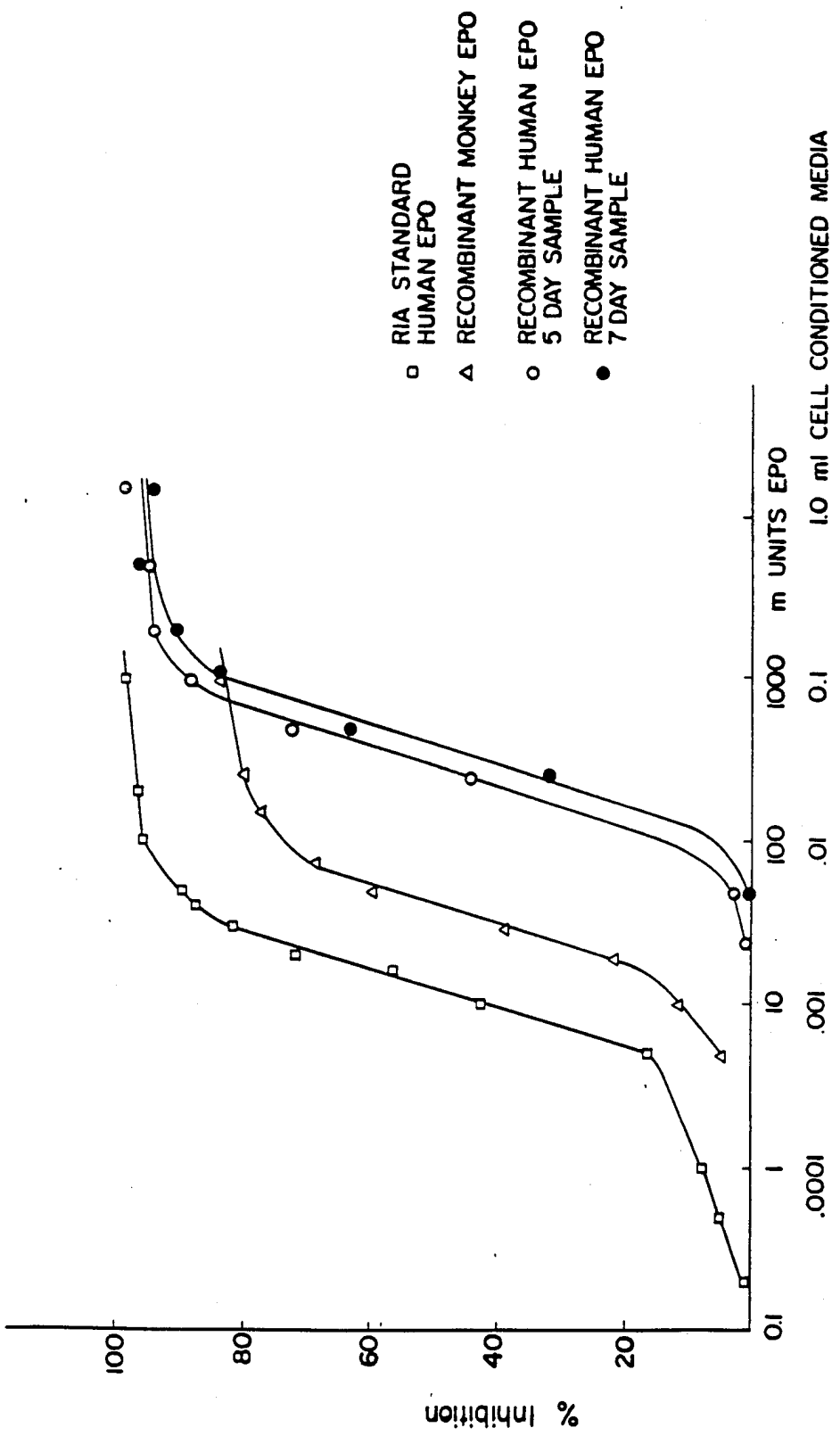

FIG. 1 Comparison of Recombinant Human & Monkey EPO in Radioimmunoassay

FIG. 2

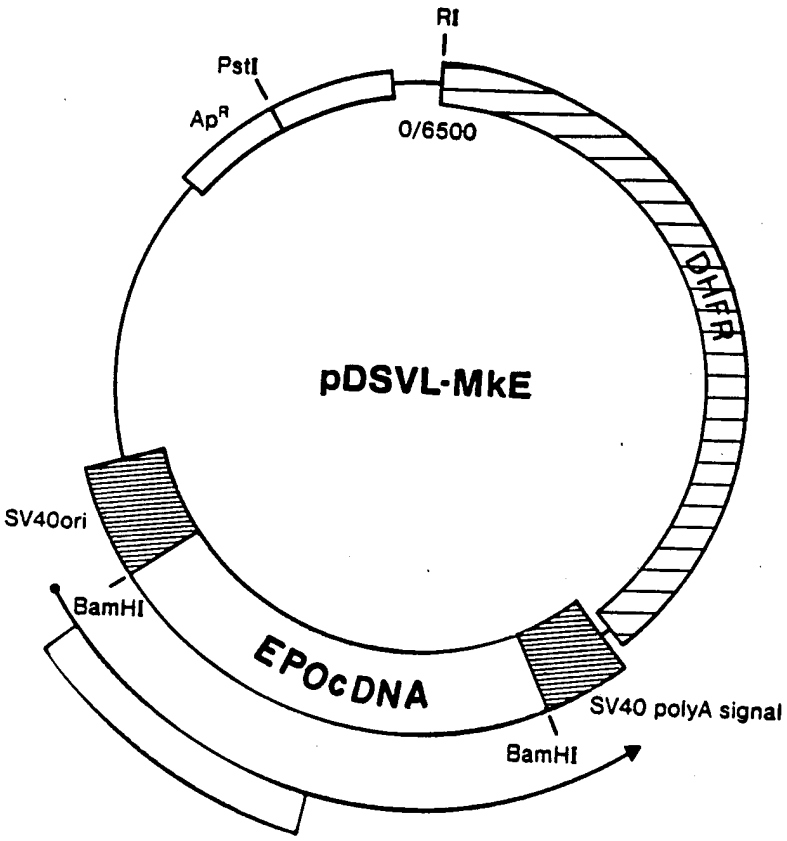
RI
PstI
Ap^R
0/6500
DHFR
pDSVL-MkE
SV40ori
BamHI
EPOcDNA
SV40 polyA signal
BamHI

FIG. 3

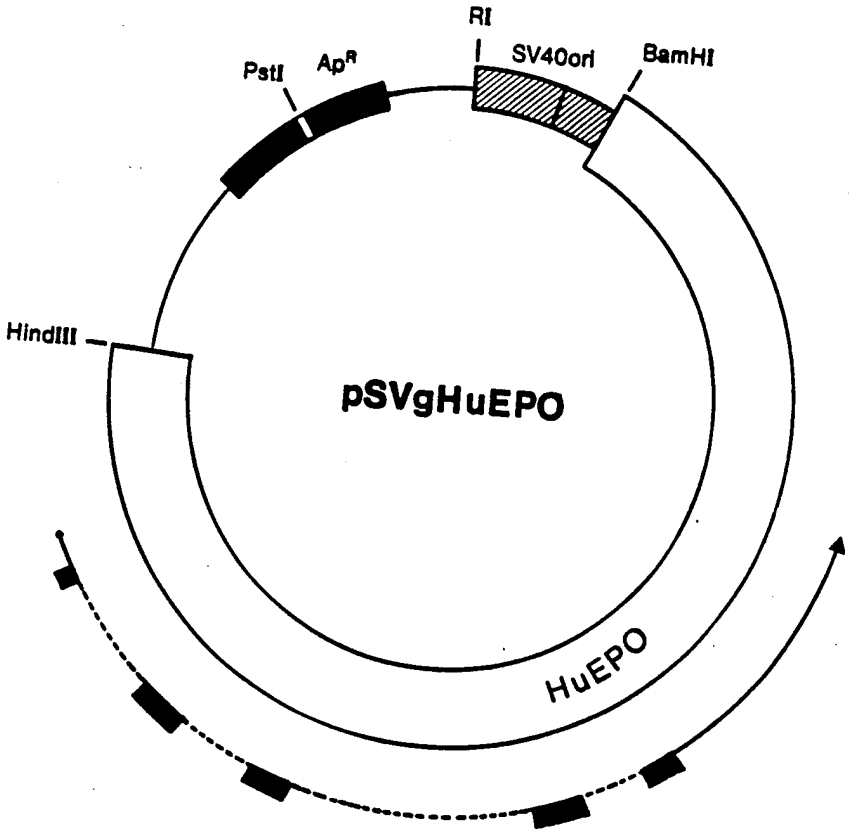

FIG. 4

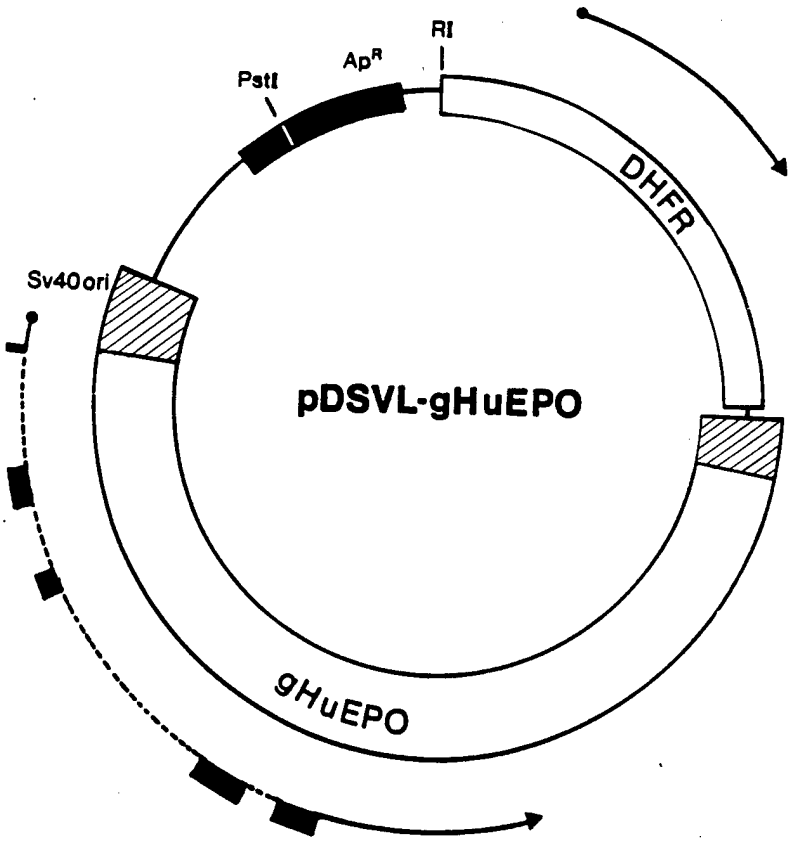

FIG.5A

Translation of Monkey EPO cDNA

```
Sau3A
GATCCCGCGCCCCCTGGACAGCCGCCCTCTCCTCCAGGCCCGTGGGGCTGGCCCTGCCC

CGCTGAACTTCCCGGGATGAGGACTCCCGGTGTGGTCACCGCGCGCCTAGGTCGCTGAG

                        -27                             -20
                        Met Gly Val His Glu Cys Pro Ala Trp
GGACCCCGGCCAGGCGCGGAGATG GGG GTG CAC GAA TGT CCT GCC TGG

                                    -10
Leu Trp Leu Leu Leu Ser Leu Val Ser Leu Pro Leu Gly Leu Pro
CTG TGG CTT CTC CTG TCT CTC GTG TCG CTC CCT CTG GGC CTC CCA

          -1  +1                                    10
Val Pro Gly Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val Leu
GTC CCG GGC GCC CCA CCA CGC CTC ATC TGT GAC AGC CGA GTC CTG

                                20                  *
Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Val Thr Met
GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT GTC ACG ATG

          30                                    *       40
Gly Cys Ser Glu Ser Cys Ser Leu Asn Glu Asn Ile Thr Val Pro
GGC TGT TCC GAA AGC TGC AGC TTG AAT GAG AAT ATC ACC GTC CCA
```

FIG.5B

```
                                                    50
Asp Thr Lys Val Asn Phe Tyr Ala Trp Lys Arg Met Glu Val Gly
GAC ACC AAA GTT AAC TTC TAT GCC TGG AAG AGG ATG GAG GTC GGG

                                                    60
Gln Gln Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu
CAG CAG GCT GTA GAA GTC TGG CAG GGC CTG GCC CTG CTC TCA GAA

                                                    80
                                                    *
Ala Val Leu Arg Gly Gln Ala Val Leu Ala Asn Ser Ser Gln Pro
GCT GTC CTG CGG GGC CAG GCC GTG TTG GCC AAC TCT TCC CAG CCT

                                                    90
Phe Glu Pro Leu Gln Leu His Met Asp Lys Ala Ile Ser Gly Leu
TTC GAG CCC CTG CAG CTG CAC ATG GAT AAA GCC ATC AGT GGC CTT

                                                   110
Arg Ser Ile Thr Thr Leu Leu Arg Ala Leu Gly Ala Gln Glu Ala
CGC AGC ATC ACC ACT CTG CTT CGG GCG CTG GGA GCC CAG GAA GCC

                                                   120
Ile Ser Leu Pro Asp Ala Ala Ser Ala Ala Pro Leu Arg Thr Ile
ATC TCC CTC CCA GAT GCG GCC TCG GCT GCT CCA CTC CGA ACC ATC

                                                   140
Thr Ala Asp Thr Phe Cys Lys Leu Phe Arg Val Tyr Ser Asn Phe
ACT GCT GAC ACT TTC TGC AAA CTC TTC CGA GTC TAC TCC AAT TTC
```

FIG.5C

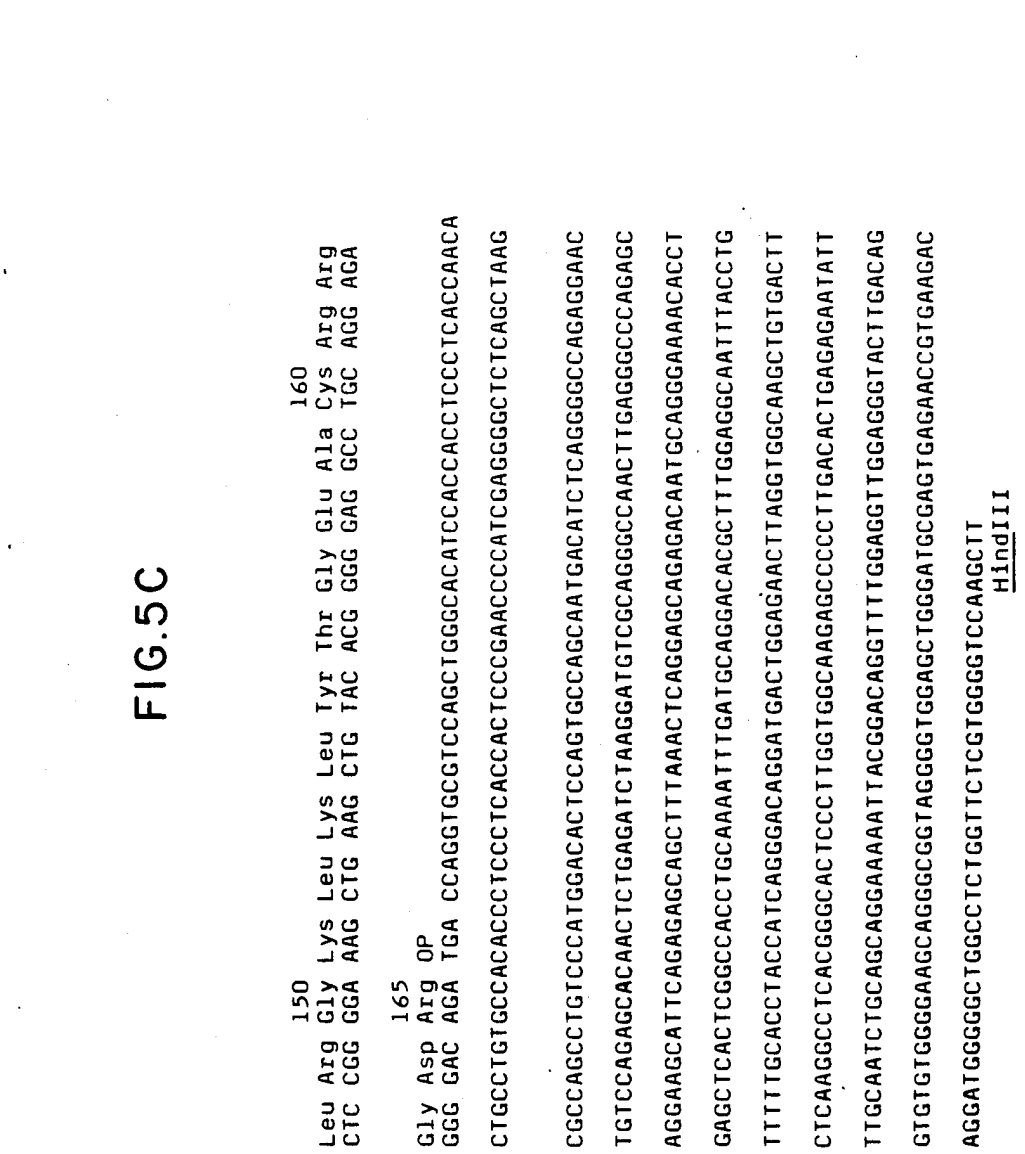
150
160
Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg Arg
CTC CGG GGA AAG CTG AAG CTG TAC ACG GGG GAG GCC TGC AGG AGA
165
Gly Asp Arg OP
GGG GAC AGA TGA CCAGGTGCGTCCAGCTGGGCACATCCACCACCTCCCTCACCAACA
CTGCCTGTGCCACACCCTCCCTCACCACTCCCGAACCCCATCGAGGGGCTCTCAGCTAAG
CGCCAGCCTGTCCCATGGACACTCCAGTGCCAGCAATGACATCTCAGGGGCCAGAGGAAC
TGTCCAGAGCACAACTCTGAGATCTAAGGATGTCGCAGGGCCAACTTGAGGGCCCAGAGC
AGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGAGCAGAGACAATGCAGGGAAAACACCT
GAGCTCACTCGGCCACCTGCAAAATTTGATGCAGGACACGCTTTGGAGGCAATTTACCTG
TTTTTGCACCTACCATCAGGGACAGGATGACTGGAGAACTTAGGTGGCAAGCTGTGACTT
CTCAAGGCCTCACGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACTGAGAGAATATT
TTGCAATCTGCAGCAGGAAAAATTACGGACAGGTTTTGGAGGTTGGAGGGTACTTGACAG
GTGTGTGGGGAAGCAGGGCGGTAGGGGTGGAGCTGGGATGCGAGTGAGAACCGTGAAGAC
AGGATGGGGGCTGGCCTCTGGTTCTCGTGGGGTCCAAGCTT
HindIII

FIG.6A


AAGCTTCTGGGCTTCCAGACCCAGCTACTTTGCGGAACTCAGCAACCCAGGCATCTCTGAGTCTCCGCCCA
AGACCGGGATGCCCCCCAGGGGAGGTGTCCGGGAGCCCAGCCTTTCCCAGATAGCACGCTCCGCCAGTCCC
AAGGGTGCGCAACCGGCTGCACTCCCCTCCCGCGACCCAGGGCCCGGGAGCAGCCCCCATGACCCACACGC
ACGTCTGCAGCAGCCCCGCTCACGCCCCGGCGAGCCTCAACCCAGGCGTCCTGCCCCTGCTCTGACCCCGG
GTGGCCCCTACCCCTGGCGACCCCTCACGCACACAGCCTCTCCCCCACCCCCACCCGCGCACGCACACATG
CAGATAACAGCCCCGACCCCCGGCCAGAGCCGXAGAGTCCCTGGGCCACCCCGGCCGCTCGCCTGCCGCTG
CGCCGCACCGCGCTGTCCTCCCGGAGCCGGACCGGGGCCACCGCGCCCXGCTCTGCTCCGACACCGCGCCC
CTTGGACAGCCGCCCTCTCCTCTAGGCCCGTGGGGCTGGCCCTGCACCGCCGAGCTTCCCGGGATGAGGXX
-27 -24
Met Gly Val His
CCCGGTGACCGGCGCGCCCCAAGTCGCTGAGGGACCCCGGCCAAGCGCGGAG ATG GGG GTG CAC G
GTGAGTACTCGCGGGCTGGGCGCTCCCGGCGGCCGGGTTCCTGTTTGAGCGGGGATTTAGCGCCCCGGCT

FIG.6B

```
ATTGGCCAAGAGGTGGCTGGGTTCAAGGACCGGCGACTTGTCAAGGACCCCGGAAGGGGGAGGGGGGTGGG

GCAGCCTCCACGTGCCGCGGGGACTTGGGGGAGTTCTTGGGGATGGCAAAAACCTGGCCTGTTGAGGGGCA

CAGTTTGGGGTTGGGGAGGAGGTTTGGGGTTCTGCTGTGCAGTTGTGTCGTTGTCAGTGTCTCG[I.S.]

TTGCACACGCACAGATCAATAAGCCAGAGGCAGCACCTGAGTGCTTGCATGGTTGGGACAGGAAGGACGAG


CTGGGGCAGAGACGTGGGGATGAAGGAAGCTGTCCTTCCACAGCCACCCTTCTCCCCCCCCGCCTGACTCT

                        -23         -20
                        Glu Cys Pro Ala Trp Leu Trp Leu Leu Leu Ser Leu
CAGCCTGGCTATCTGTTCTAG    AA TGT CCT GCC TGG CTG TGG CTT CTC CTG TCC CTG

    -10                                 -1  +1
Leu Ser Leu Pro Leu Gly Leu Pro Val Leu Gly Ala Pro Pro Arg Leu Ile Cys
CTG TCG CTC CCT CTG GGC CTC CCA GTC CTG GGC GCC CCA CCA CGC CTC ATC TGT

         10                                        20              *
Asp Ser Arg Val Leu Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Ile
GAC AGC CGA GTC CTG GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT ATC

 26
Thr
ACG GTGAGACCCCTTCCCCAGCACATTCCACAGAACTCACGCTCAGGGCTTCAGGGAACTCCTCCCAGAT


CCAGGAACCTGGCACTTGGTTTGGGGTGGAGTTGGGAAGCTAGACACTGCCCCCCTACATAAGAATAAGTC
```

FIG.6C

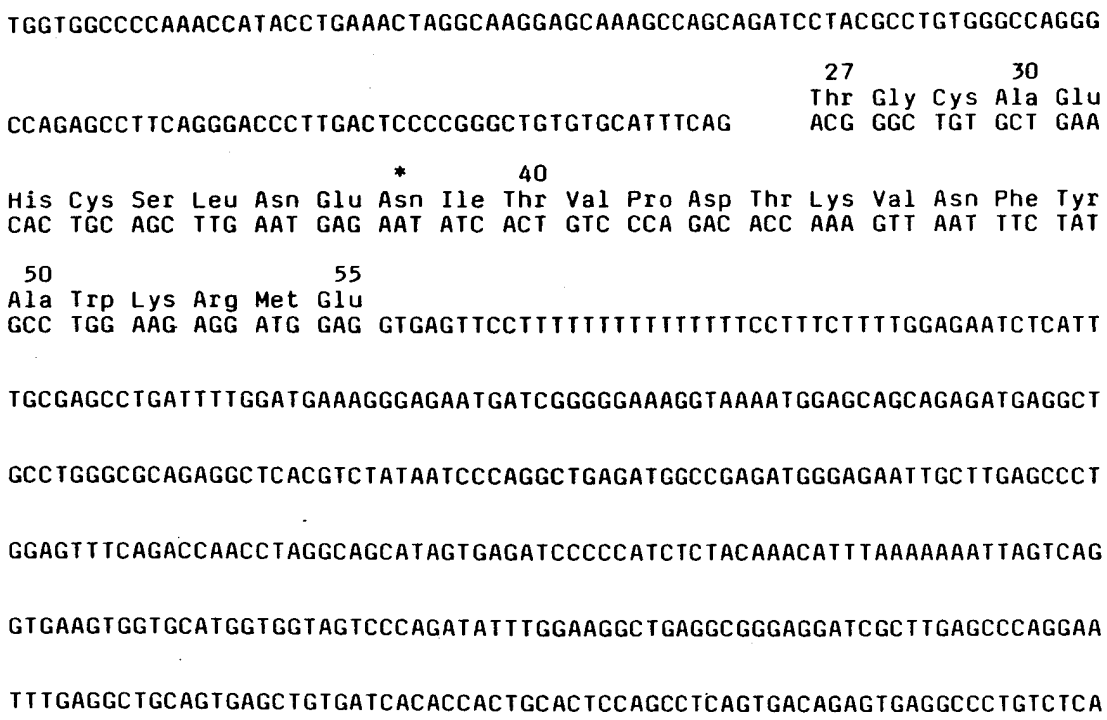

FIG. 6D

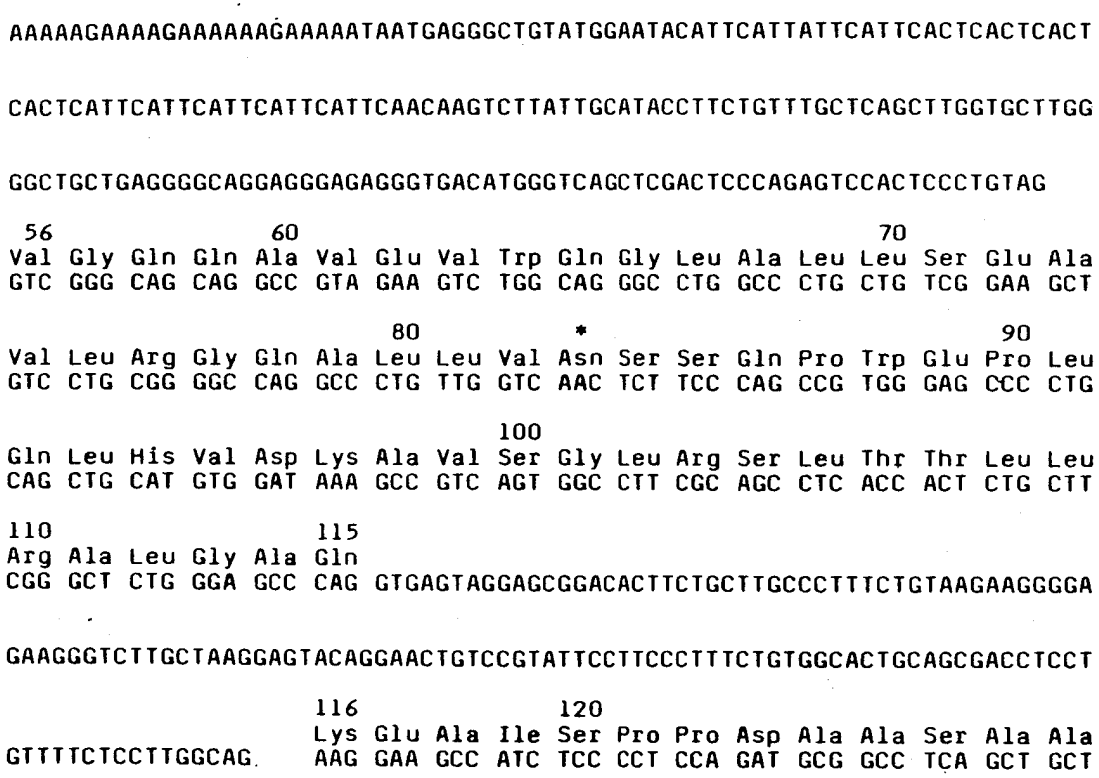

FIG. 6E

```
    130                                         140
Pro Leu Arg Thr Ile Thr Ala Asp Thr Phe Arg Lys Leu Phe Arg Val Tyr Ser
CCA CTC CGA ACA ATC ACT GCT GAC ACT TTC CGC AAA CTC TTC CGA GTC TAC TCC

            150                                         160
Asn Phe Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg Thr Gly
AAT TTC CTC CGG GGA AAG CTG AAG CTG TAC ACA GGG GAG GCC TGC AGG ACA GGG

    166
Asp Arg OP
GAC AGA TGA CCAGGTGTGTCCACCTGGGCATATCCACCACCTCCCTCACCAACATTGCTTGTGCCACA

CCCTCCCCCGCCACTCCTGAACCCCGTCGAGGGGCTCTCAGCTCAGCGCCAGCCTGTCCCATGGACACTCC

AGTGCCAGCAATGACATCTCAGGGGCCAGAGGAACTGTCCAGAGAGCAACTCTGAGATCTAAGGATGTCAC

AGGGCCAACTTGAAGGGCCCAGAGCAGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGGACAGAGCCATGC

TGGGAAGACGCCTGAGCTCACTCGGCACCCTGCAAAATTTGATGCCAGGACACGCTTTGGAGGCGATTTAC

CTGTTTTCGCACCTACCATCAGGGACAGGATGACCTGGAGAACTTAGGTGGCAAGCTGTGACTTCTCCAGG

TCTCACGGGCATGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACCGGGGTGGTGGGAACCATGAAGAC

AXGATXGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTATTCTCAACCTATTGACAGACTGAA

ACACAATATGAC
```

FIG.7

ECEPO GENE

```
                                     -1  1
      XbaI                           MetAla
      CTAG AAACCATGAG GGTAATAAAA TAATGGCTCC GCCGCGTCTG
           TTTGGTACTC CCATTATTTT ATTACCGAGG CGGCGCAGAC

ATCTGCGACT CGAGAGTTCT GGAACGTTAC CTGCTGGAAG CTAAAGAAGC
TAGACGCTGA GCTCTCAAGA CCTTGCAATG GACGACCTTC GATTTCTTCG

TGAAAACATC ACCACTGGTT GTGCTGAACA CTGTTCTTTG AACGAAAACA
ACTTTTGTAG TGGTGACCAA CACGACTTGT GACAAGAAAC TTGCTTTTGT

TTACGGTACC AGACACCAAG GTTAACTTCT ACGCTTGGAA ACGTATGGAA
AATGCCATGG TCTGTGGTTC CAATTGAAGA TGCGAACCTT TGCATACCTT

GTTGGTCAAC AAGCAGTTGA AGTTTGGCAG GGTCTGGCAC TGCTGAGCGA
CAACCAGTTG TTCGTCAACT TCAAACCGTC CCAGACCGTG ACGACTCGCT

GGCTGTACTG CGTGGCCAGG CACTGCTGGT AAACTCCTCT CAGCCGTGGG
CCGACATGAC GCACCGGTCC GTGACGACCA TTTGAGGAGA GTCGGCACCC

AACCGCTGCA GCTGCATGTT GACAAAGCAG TATCTGGCCT GAGATCTCTG
TTGGCGACGT CGACGTACAA CTGTTTCGTC ATAGACCGGA CTCTAGAGAC

ACTACTCTGC TGCGTGCTCT GGGTGCACAG AAAGAGGCTA TCTCTCCGCC
TGATGAGACG ACGCACGAGA CCCACGTGTC TTTCTCCGAT AGAGAGGCGG

GGATGCTGCA TCTGCTGCAC CGCTGCGTAC CATCACTGCT GATACCTTCC
CCTACGACGT AGACGACGTG GCGACGCATG GTAGTGACGA CTATGGAAGG

GCAAACTGTT TCGTGTATAC TCTAACTTCC TGCGTGGTAA ACTGAAACTG
CGTTTGACAA AGCACATATG AGATTGAAGG ACGCACCATT TGACTTTGAC

                                           SalI
TATACTGGCG AAGCATGCCG TACTGGTGAC CGCTAATAG
ATATGACCGC TTCGTACGGC ATGACCACTG GCGATTATCA GCT
```

FIG. 8

SCEPO GENE

```
                -1 +1
HindIII         ArgAla
AGCTTGGATA AAAGAGCTCC ACCAAGATTG ATCTGTGACT CGAGAGTTTT
    ACCTAT TTTCTCGAGG TGGTTCTAAC TAGACACTGA GCTCTCAAAA

GGAAAGATAC TTGTTGGAAG CTAAAGAAGC TGAAAACATC ACCACTGGTT
CCTTTCTATG AACAACCTTC GATTTCTTCG ACTTTTGTAG TGGTGACCAA

GTGCTGAACA CTGTTCTTTG AACGAAAACA TTACGGTACC AGACACCAAG
CACGACTTGT GACAAGAAAC TTGCTTTTGT AATGCCATGG TCTGTGGTTC

GTTAACTTCT ACGCTTGGAA ACGTATGGAA GTTGGTCAAC AAGCTGTTGA
CAATTGAAGA TGCGAACCTT TGCATACCTT CAACCAGTTG TTCGACAACT

AGTTTGGCAA GGTTTGGCCT TGTTATCTGA AGCTGTTTTG AGAGGTCAAG
TCAAACCGTT CCAAACCGGA ACAATAGACT TCGACAAAAC TCTCCAGTTC

CCTTGTTGGT TAACTCTTCT CAACCATGGG AACCATTGCA ATTGCACGTC
GGAACAACCA ATTGAGAAGA GTTGGTACCC TTGGTAACGT TAACGTGCAG

GATAAAGCCG TCTCTGGTTT GAGATCTTTG ACTACTTTGT TGAGAGCTTT
CTATTTCGGC AGAGACCAAA CTCTAGAAAC TGATGAAACA ACTCTCGAAA

GGGTGCTCAA AAGGAAGCCA TTTCCCCACC AGACGCTGCT TCTGCCGCTC
CCCACGAGTT TTCCTTCGGT AAAGGGGTGG TCTGCGACGA AGACGGCGAG

CATTGAGAAC CATCACTGCT GATACCTTCA GAAAGTTATT CAGAGTTTAC
GTAACTCTTG GTAGTGACGA CTATGGAAGT CTTTCAATAA GTCTCAAATG

TCCAACTTCT TGAGAGGTAA ATTGAAGTTG TACACCGGTG AAGCCTGTAG
AGGTTGAAGA ACTCTCCATT TAACTTCAAC ATGTGGCCAC TTCGGACATC

AACTGGTGAC AGATAAGCCC GACTGATAAC AACAGTGTAG
TTGACCACTG TCTATTCGGG CTGACTATTG TTGTCACATC

            SalI
ATGTAACAAA G
TACATTGTTT CAGCT
```

Comparison of Human and Monkey EPO Polypeptides

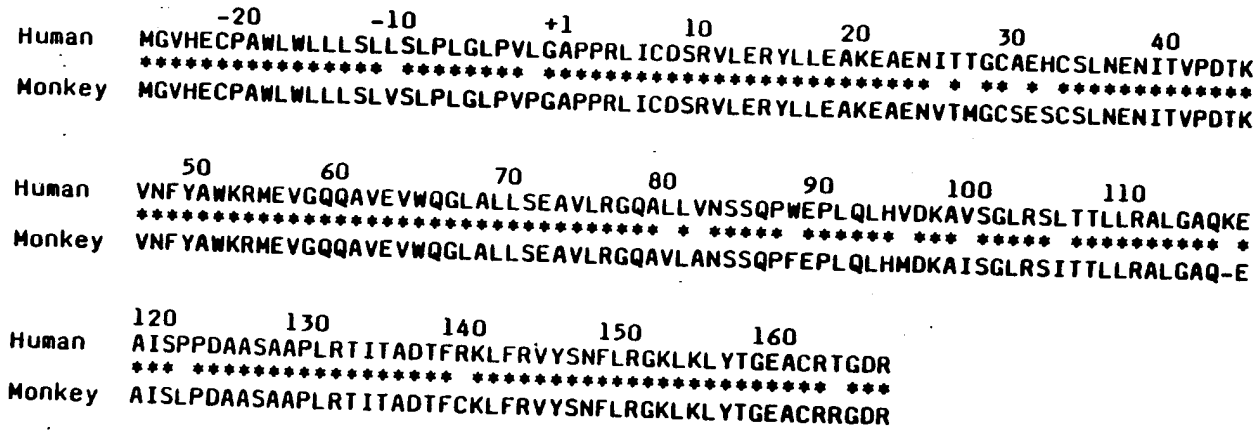

FIG. 9

ECEPO SECTION 1 OLIGONUCLEOTIDES

1. AATTCTAGAAACCATGAGGGTAATAAAATA
2. CCATTATTTTATTACCCTCATGGTTTCTAG
3. ATGGCTCCGCCGCGTCTGATCTGCGAC
4. CTCGAGTCGCAGATCAGACGCGGCGGAG
5. TCGAGAGTTCTGGAACGTTACCTGCTG
6. CTTCCAGCAGGTAACGTTCCAGAACT
7. GAAGCTAAAGAAGCTGAAAACATC
8. GTGGTGATGTTTTCAGCTTCTTTAG
9. ACCACTGGTTGTGCTGAACACTGTTC
10. CAAAGAACAGTGTTCAGCACAACCA
11. TTTGAACGAAAACATTACGGTACCG
12. GATCCGGTACCGTAATGTTTTCGTT

FIG. 10

ECEPO SECTION 1

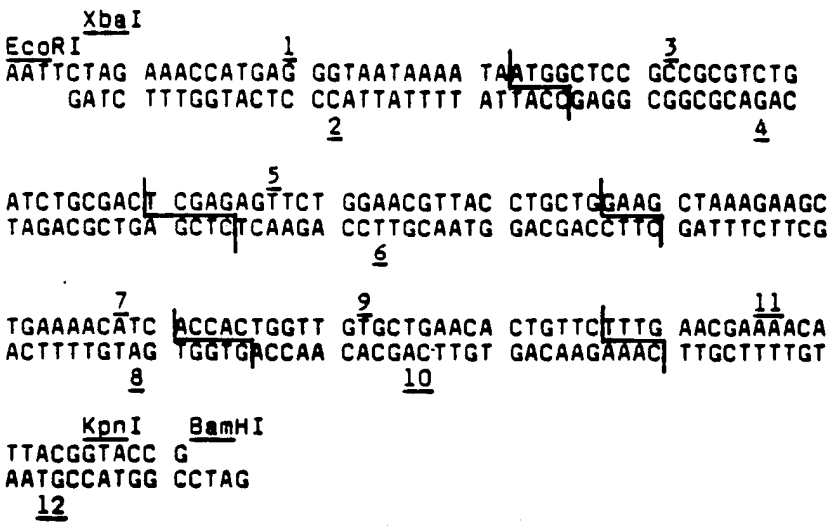

FIG. 11

ECEPO SECTION 2 OLIGONUCLEOTIDES

1. AATTCGGTACCAGACACCAAGGT
2. GTTAACCTTGGTGTCTGGTACCG
3. TAACTTCTACGCTTGGAAACGTAT
4. TTCCATACGTTTCCAAGCGTAGAA
5. GGAAGTTGGTCAACAAGCAGTTGAAGT
6. CCAAACTTCAACTGCTTGTTGACCAAC
7. TTGGCAGGGTCTGGCACTGCTGAGCG
8. GCCTCGCTCAGCAGTGCCAGACCCTG
9. AGGCTGTACTGCGTGGCCAGGCA
10. GCAGTGCCTGGCCACGCAGTACA
11. CTGCTGGTAAACTCCTCTCAGCCGT
12. TTCCCACGGCTGAGAGGAGTTTACCA
13. GGGAACCGCTGCAGCTGCATGTTGAC
14. GCTTTGTCAACATGCAGCTGCAGCGG
15. AAAGCAGTATCTGGCCTGAGATCTG
16. GATCCAGATCTCAGGCCAGATACT

FIG. 12

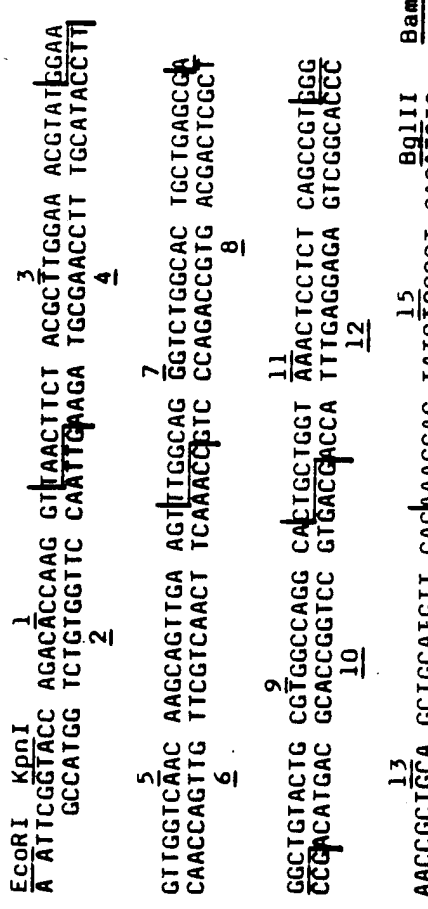

FIG. 13

ECEPO SECTION 3

1. GATCCAGATCTCTGACTACTCTGC
2. ACGCAGCAGAGTAGTCAGAGATCTG
3. TGCGTGCTCTGGGTGCACAGAAAGAGG
4. GATAGCCTCTTTCTGTGCACCCAGAGC
5. CTATCTCTCCGCCGGATGCTGCATCT
6. CAGCAGATGCAGCATCCGGCGGAGA
7. GCTGCACCGCTGCGTACCATCACTG
8. ATCAGCAGTGATGGTACGCAGCGGTG
9. CTGATACCTTCCGCAAACTGTTTCG
10. ATACACGAAACAGTTTGCGGAAGGT
11. TGTATACTCTAACTTCCTGCGTGGTA
12. CAGTTTACCACGCAGGAAGTTAGAGT
13. AACTGAAACTGTATACTGGCGAAGC
14. GGCATGCTTCGCCAGTATACAGTTT
15. ATGCCGTACTGGTGACCGCTAATAG
16. TCGACTATTAGCGGTCACCAGTAC

FIG. 14

ECEPO SECTION 3

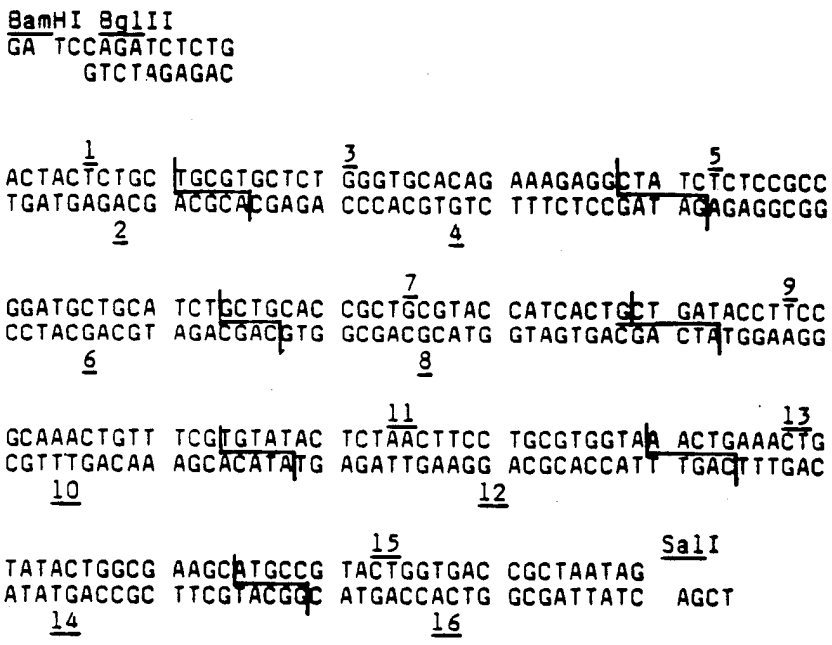

FIG. 15

SCEPO SECTION 1 OLIGONUCLEOTIDES

1. AATTCAAGCTTGGATAAAAGAGCT
2. GTGGAGCTCTTTTATCCAAGCTTG
3. CCACCAAGATTGATCTGTGACTC
4. TCTCGAGTCACAGATCAATCTTG
5. GAGAGTTTTGGAAAGATACTTGTTG
6. CTTCCAACAAGTATCTTTCCAAAAC
7. GAAGCTAAAGAAGCTGAAAACATC
8. GTGGTGATGTTTTCAGCTTCTTTAG
9. ACCACTGGTTGTGCTGAACACTGTTC
10. CAAAGAACAGTGTTCAGCACAACCA
11. TTTGAACGAAAACATTACGGTACCG
12. GATCCGGTACCGTAATGTTTTCGTT

FIG. 16

SCEPO SECTION 1

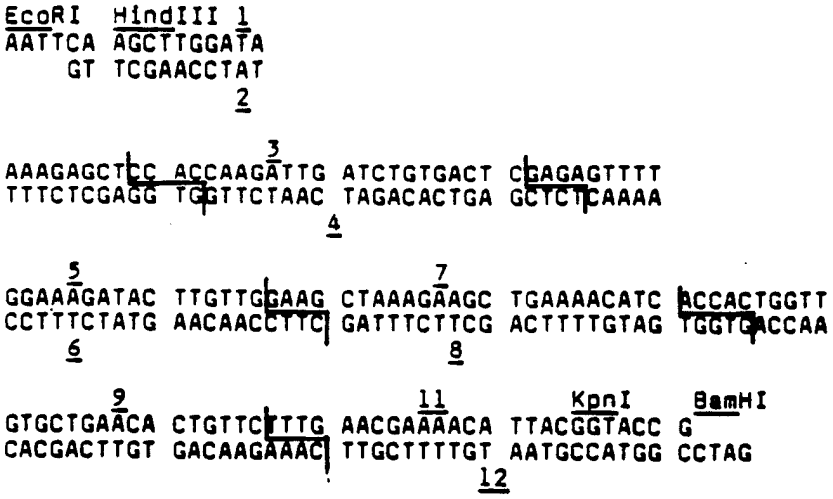

FIG. 17

SCEPO SECTION 2 OLIGONUCLEOTIDES

1. AATTCGGTACCAGACACCAAGGT
2. GTTAACCTTGGTGTCTGGTACCG
3. TAACTTCTACGCTTGGAAACGTAT
4. TTCCATACGTTTCCAAGCGTAGAA
5. GGAAGTTGGTCAACAAGCAGTTGAAGT
6. CCAAACTTCAACTGCTTGTTGACCAAC
7. TTGGCAAGGTTTGGCCTTGTTATCTG
8. GCTTCAGATAACAAGGCCAAACCTTG
9. AAGCTGTTTTGAGAGGTCAAGCCT
10. AACAAGGCTTGACCTCTCAAAACA
11. TGTTGGTTAACTCTTCTCAACCATGGG
12. TGGTTCCCATGGTTGAGAAGAGTTAACC
13. AACCATTGCAATTGCACGTCGAT
14. CTTTATCGACGTGCAATTGCAA
15. AAAGCCGTCTCTGGTTTGAGATCTG
16. GATCCAGATCTCAAACCAGAGACGG

FIG. 18

SCEPO SECTION 2

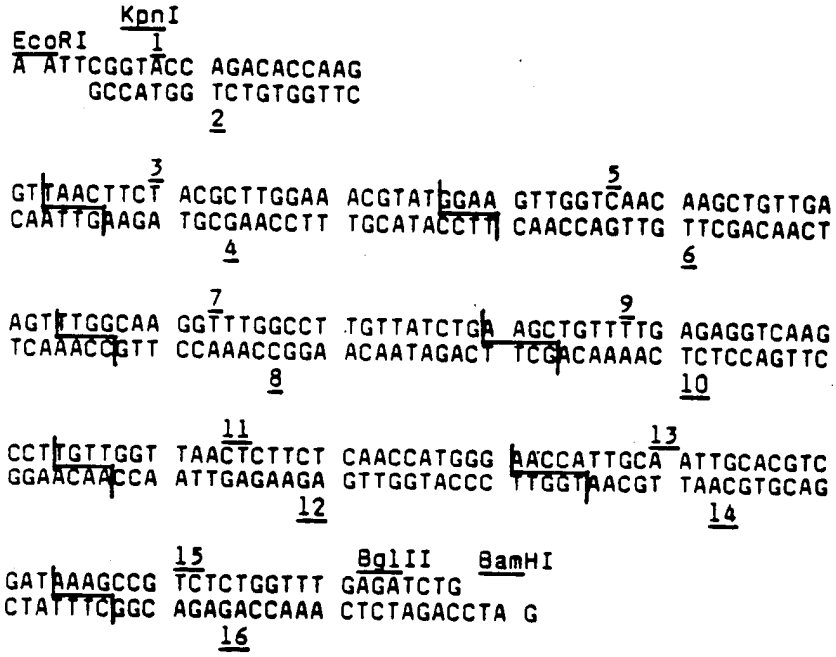

FIG. 19

SCEPO SECTION 3 OLIGONUCLEOTIDES

1. GATCCAGATCTTTGACTACTTTGTT
2. TCTCAACAAAGTAGTCAAAGATCTG
3. GAGAGCTTTGGGTGCTCAAAAGGAAG
4. ATGGCTTCCTTTTGAGCACCCAAAGC
5. CCATTTCCCCACCAGACGCTGCTT
6. GCAGAAGCAGCGTCTGGTGGGGAA
7. CTGCCGCTCCATTGAGAACCATC
8. CAGTGATGGTTCTCAATGGAGCG
9. ACTGCTGATACCTTCAGAAAGTT
10. GAATAACTTTCTGAAGGTATCAG
11. ATTCAGAGTTTACTCCAACTTCT
12. CTCAAGAAGTTGGAGTAAACTCT
13. TGAGAGGTAAATTGAAGTTGTACAC
14. ACCGGTGTACAACTTCAATTTACCT
15. CGGTGAAGCCTGTAGAACTGGT
16. CTGTCACCAGTTCTACAGGCTTC
17. GACAGATAAGCCCGACTGATAA
18. GTTGTTATCAGTCGGGCTTAT
19. CAACAGTGTAGATGTAACAAAG
20. TCGACTTTGTTACATCTACACT

FIG. 20

SCEPO SECTION 3

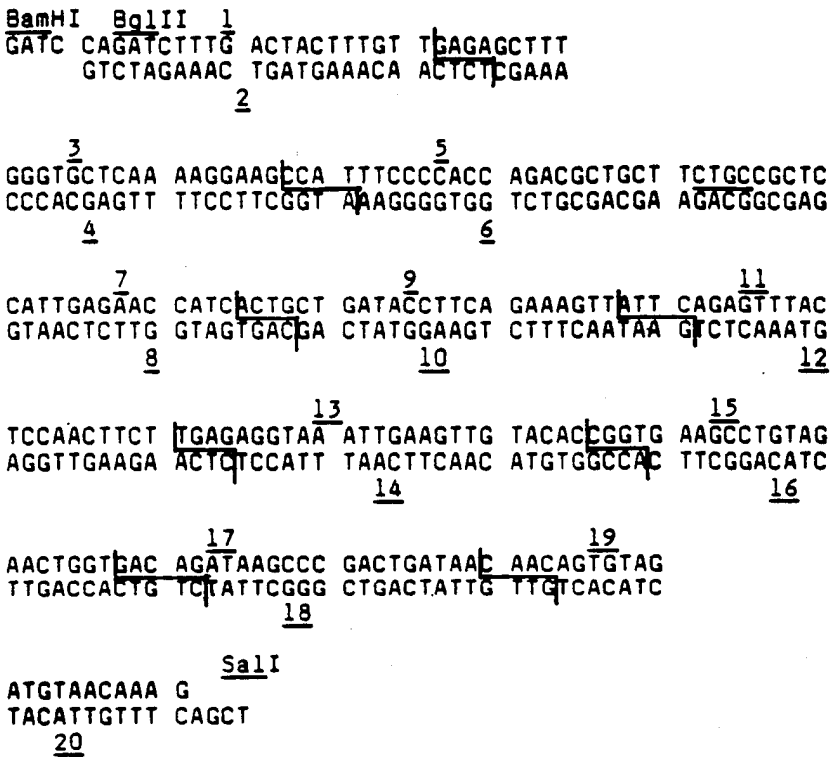

FIG. 21

DNA SEQUENCES ENCODING ERYTHROPOIETIN

This is a continuation-ln-part of my co-pendlng U.S. patent application Ser. Nos. 561,024, filed Dec. 13, 1983, (now abandoned) 582,185, filed Feb. 21, 1984, (now abandoned) and 655,841, filed Sept. 28, 1984.

BACKGROUND

The present invention relates generally to the manipulation of genetic materials and, more particularly, to recombinant procedures making possible the production of polypeptides possessing part or all of the primary structural conformation and/or one or more of the biological properties of naturally-occurring erythropoletin.

A. Manipulation of Genetic Materials

Genetic materials may be broadly defined as those chemical substances which program for and guide the manufacture of constituents of cells and viruses and direct the responses of cells and viruses. A long chain polymeric substance known as deoxyribonucleic acid (DNA) comprises the genetic material of all living cells and viruses except for certain viruses which are programmed by ribonucleic acids (RNA). The repeating units in DNA polymers are four different nucleotides, each of which consists of either a purine (adenine or guanine) or a pyrimidine (thymine or cytosine) bound to a deoxyribose sugar to which a phosphate group is attached. Attachment of nucleotides in linear polymeric form is by means of fusion of the 5′ phosphate of one nucleotlde to the 3′ hydroxyl group of another. Functional DNA occurs in the form of stable double stranded associations of single strands of nucleotides (known as deoxyoligonucleotides), which associations occur by means of hydrogen bonding between purine and pyrimidine bases [i.e., "complementary" associations existing either between adenine (A) and thymine (T) or guanine (G) and cytosine (C)]. By convention, nucleotides are referred to by the names of their constituent purine or pyrimidine bases, and the complementary associations of nucleotides in double stranded DNA (i.e., A-T and G-C) are referred to as "base pairs". Ribonucleic acid is a polynucleotide comprising adenine, guanine, cytosine and uracil (U), rather than thymine, bound to ribose and a phosphate group.

Most briefly put, the programming function of DNA is generally effected through a process wherein specific DNA nucleotide sequences (genes) are "transcribed" into relatively unstable messenger RNA (mRNA) polymers. The mRNA, in turn, serves as a template for the formation of structural, regulatory and catalytic proteins from amino acids. This mRNA "translation" process involves the operations of smallRNA strands (tRNA) which transport and align individual amino acids along the mRNA strand to allow for formation of polypeptides in proper amino acid sequences. The mRNA "message", derived from DNA and providing the basis for the tRNA supply and orientation of any given one of the twenty amino acids for polypeptide "expression", is in the form of triplet "codons"—sequential grouplngs of three nucleotide bases. ln one sense, the formation of a protein is the ultimate form of "expression" of the programmed genetic message provided by the nucleotide sequence of a gene.

"promoter" DNA sequences usually "precede" a gene in a DNA polymer and provide a site for initiation of the transcription into mRNA. "Regulator" DNA sequences, also usually "upstream" of (i.e., preceding) a gene in a given DNA polymer bind proteins that determine the frequency (or rate) of transcriptional initiation. Collectively referred to as "promoter/regulator" or "control " DNA sequence, these sequenoes which precede a selected gene (or series of genes) in a functional DNA polymer cooperate to determine whether the transcription (and eventual expression) of a gene will occur. DNA sequences which "follow" a gene in a DNA polymer and provide a signal for termination of the transcription into mRNA are referred to as transcription "terminator" sequences.

A focus of microbiological processing for the last decade has been the attempt to manufacture industrially and pharmaceutically significant substances using organisms which either do not initially have genetically coded information concerning the desired product included in their DNA, or (in the case of mammalian cells in culture) do not ordinarily express a chromosomal gene at appreciable levels. Simply put, a gene that specifies the structure of a desired polypeptide product is either isolated from a "donor" organism or chemically synthesized and then stably introduced into another organism which is preferably a self-replicating unicellular organism such as bacteria, yeast or mammalian cells in culture. Once this is done, the existing machinery for gene expression in the "transformed" or "transfected" microbial host cells operates to construct the desired product, using the exogenous DNA as a template for transcription of mRNA which is then translated into a continuous sequence of amino acid residues.

The art is rich in patent and literature publications relating to "recombinant DNA" methodologies for the isolation, synthesis, purification and amolification of genetic materials for use in transformation of selected host organisms. U.S. Pat. No. 4,237,224 to Cohen, et al., for example, relates to transformation of unicellular host organisms with "hybrid" viral or circular plasmid DNA whioh includes selected exogenous DNA sequences. The procedures of the Cohen, et al. patent first involve manufacture of a transformation vector by enzymatically cleaving viral or circular plasmid DNA to form linear DNA strands. Selected foreign ("exogenous" or "heterologous" DNA strands usually including sequences coding for desired product are prepared in linear form through use of similar enzymes. The linear viral or plasmid DNA is incubated with the foreign DNA in the presence of ligating enzymes capable of effecting a restoratlon process and "hybrid" vectors are formed which include the selected exogenous DNA segment "spliced" into the viral or circular DNA plasmid.

Transformation of compatible unicellular host organisms with the hybrid vector results in the formation of multiple copies of the exogenous DNA in the host cell population. In some instances, the desired result is simply the amplification of the foreign DNA and the "product" harvested is DNA. More frequentiy, the goal of transformation is the expression by the host cells of the exogenous DNA in the form of large scale synthesis of isolatable quantities of commercially significant protein or polypeptide fragments coded for by the foreign DNA. See also, e.g., U.S. Pat. Nos. 4,264,731 (to Shine), 4,273,875 (to Manis), 4,293,652 (to Cohen), and European patent application No. 093,619, published Nov. 9, 1983.