Amgen Inc. v. F. Hoffmann-LaRoche LTD et al Doc. 502 Att. 9

EXHIBIT E-2
Part 1 of 2

Dockets.Justia.com

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.)

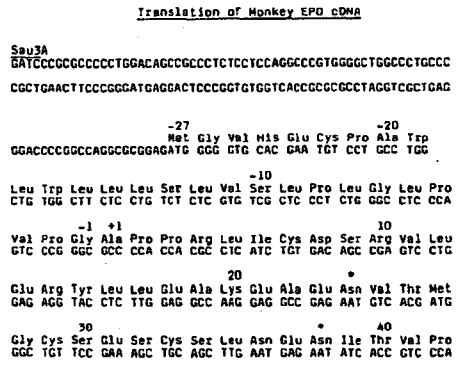
Primary Examiner—Alvin E. Tanenholtz

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



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).*

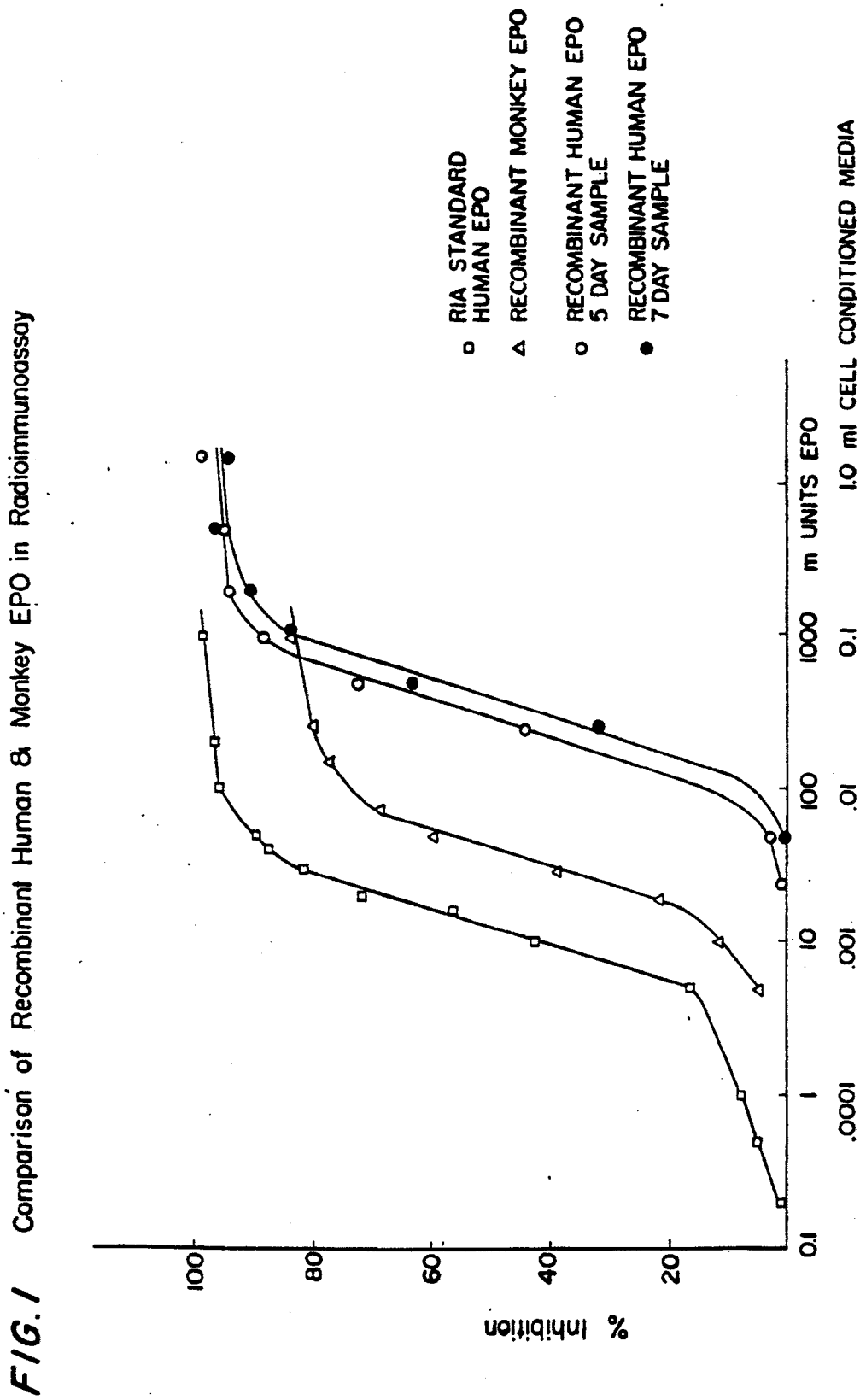


FIG. 1 Comparison of Recombinant Human & Monkey EPO in Radioimmunoassay

FIG. 2

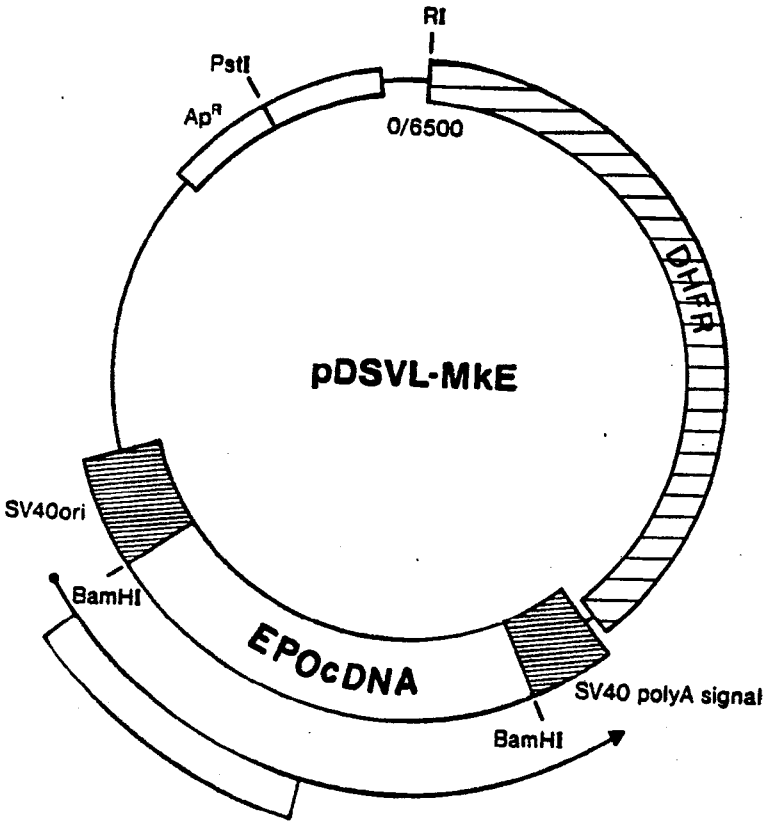

RI
PstI
Ap^R
0/6500
DHFR
pDSVL-MkE
SV40ori
BamHI
EPOcDNA
SV40 polyA signal
BamHI

FIG. 3

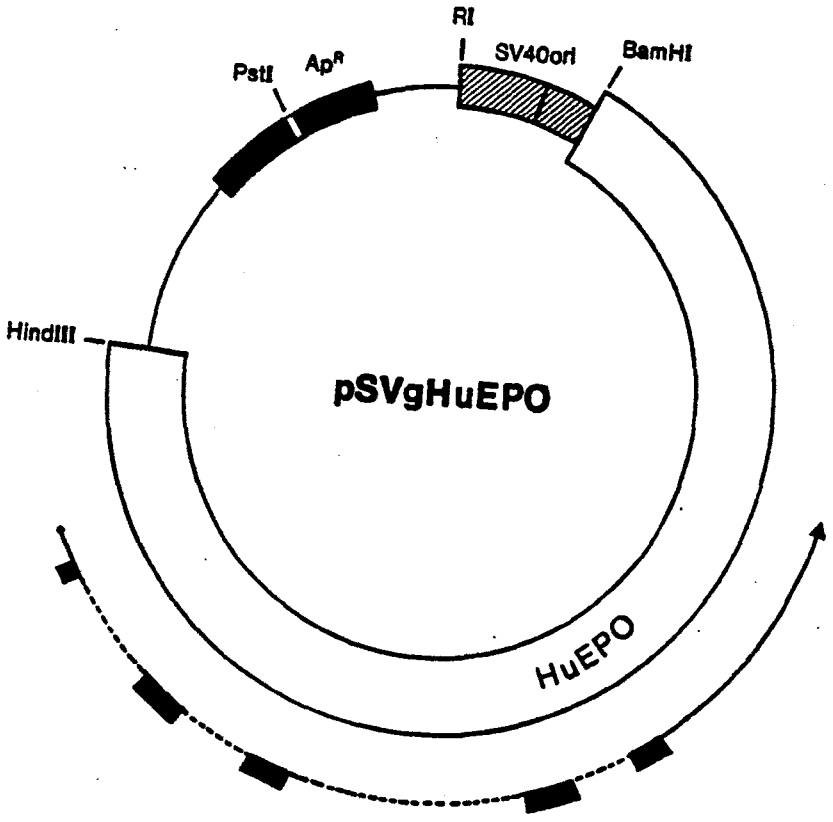
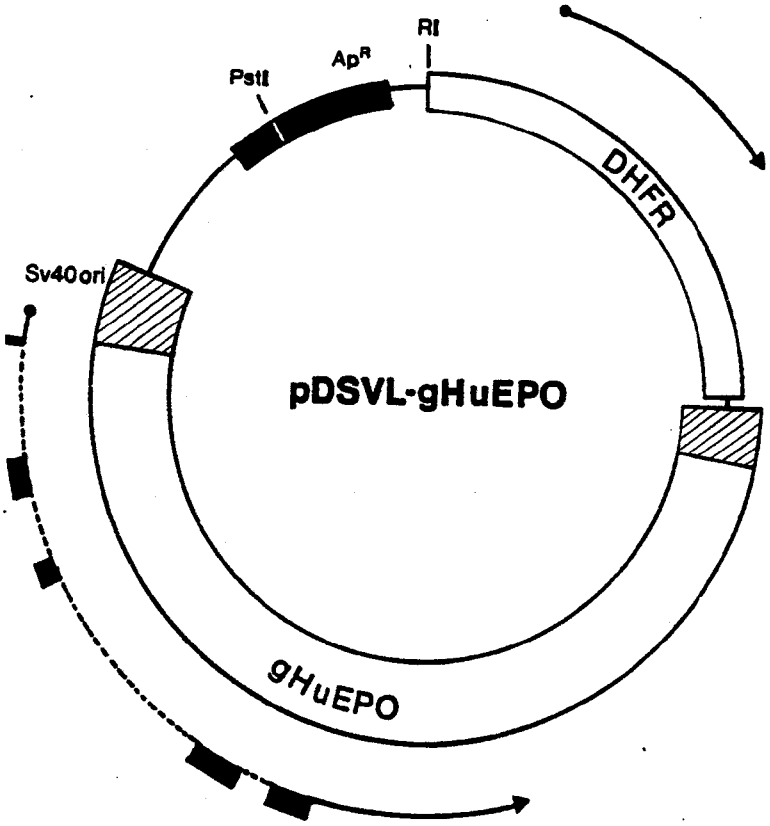


FIG. 4



Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 8 of 40

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                  70
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                  100
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                 130
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
```

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 10 of 40

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

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 11 of 40

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 12 of 40

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
```

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 13 of 40

FIG.6C

```
TGGTGGCCCCAAACCATACCTGAAACTAGGCAAGGAGCAAAGCCAGCAGATCCTACGCCTGTGGGCCAGGG

                                                          27          30
                                                         Thr Gly Cys Ala Glu
CCAGAGCCTTCAGGGACCCTTGACTCCCCGGGCTGTGTGCATTTCAG          ACG GGC TGT GCT GAA

                                *       40
His Cys Ser Leu Asn Glu Asn Ile Thr Val Pro Asp Thr Lys Val Asn Phe Tyr
CAC TGC AGC TTG AAT GAG AAT ATC ACT GTC CCA GAC ACC AAA GTT AAT TTC TAT

 50                   55
Ala Trp Lys Arg Met Glu
GCC TGG AAG AGG ATG GAG GTGAGTTCCTTTTTTTTTTTTTTTCCTTTCTTTTGGAGAATCTCATT

TGCGAGCCTGATTTTGGATGAAAGGGAGAATGATCGGGGGAAAGGTAAAATGGAGCAGCAGAGATGAGGCT

GCCTGGGCGCAGAGGCTCACGTCTATAATCCCAGGCTGAGATGGCCGAGATGGGAGAATTGCTTGAGCCCT

GGAGTTTCAGACCAACCTAGGCAGCATAGTGAGATCCCCCATCTCTACAAACATTTAAAAAAATTAGTCAG

GTGAAGTGGTGCATGGTGGTAGTCCCAGATATTTGGAAGGCTGAGGCGGGAGGATCGCTTGAGCCCAGGAA

TTTGAGGCTGCAGTGAGCTGTGATCACACCACTGCACTCCAGCCTCAGTGACAGAGTGAGGCCCTGTCTCA
```

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 14 of 40

FIG. 6D

```
AAAAAGAAAAGAAAAAAGAAAAATAATGAGGGCTGTATGGAATACATTCATTATTCATTCACTCACTCACT

CACTCATTCATTCATTCATTCAACAAGTCTTATTGCATACCTTCTGTTTGCTCAGCTTGGTGCTTGG

GGCTGCTGAGGGGCAGGAGGGAGAGGGTGACATGGGTCAGCTCGACTCCCAGAGTCCACTCCCTGTAG

 56                60                                                  70
Val Gly Gln Gln Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu Ala
GTC GGG CAG CAG GCC GTA GAA GTC TGG CAG GGC CTG GCC CTG CTG TCG GAA GCT

                             80           *                           90
Val Leu Arg Gly Gln Ala Leu Leu Val Asn Ser Ser Gln Pro Trp Glu Pro Leu
GTC CTG CGG GGC CAG GCC CTG TTG GTC AAC TCT TCC CAG CCG TGG GAG CCC CTG

                                  100
Gln Leu His Val Asp Lys Ala Val Ser Gly Leu Arg Ser Leu Thr Thr Leu Leu
CAG CTG CAT GTG GAT AAA GCC GTC AGT GGC CTT CGC AGC CTC ACC ACT CTG CTT

110                   115
Arg Ala Leu Gly Ala Gln
CGG GCT CTG GGA GCC CAG GTGAGTAGGAGCGGACACTTCTGCTTGCCCTTTCTGTAAGAAGGGGA

GAAGGGTCTTGCTAAGGAGTACAGGAACTGTCCGTATTCCTTCCCTTTCTGTGGCACTGCAGCGACCTCCT

                      116             120
                      Lys Glu Ala Ile Ser Pro Pro Asp Ala Ala Ser Ala Ala
GTTTTCTCCTTGGCAG.     AAG GAA GCC ATC TCC CCT CCA GAT GCG GCC TCA GCT GCT
```

Case 1:05-cv-12237-WGY Document 502-10 Filed 06/14/2007 Page 15 of 40

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
                                        MetAla
      XbaI                                
      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

SCEPC 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

```
                -20       -10       +1        10        20        30        40
Human   MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTK
        ********************* ********** ********** ********** * ** * **********
Monkey  MGVHECPAWLWLLLSLVSLPLGLPVPGAPPRLICDSRVLERYLLEAKEAENVTMGCSESCSLNENITVPDTK

                50        60        70        80        90        100       110
Human   VNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKE
        ********************** ********** * ***** ****** *** ***** ********** *
Monkey  VNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQAVLANSSQPFEPLQLHMDKAISGLRSITTLLRALGAQ-E

                120       130       140       150       160
Human   AISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR
        *** ********** ********** ********** ********** ***
Monkey  AISLPDAASAAPLRTITADTFCKLFRVYSNFLRGKLKLYTGEACRRGDR
```

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

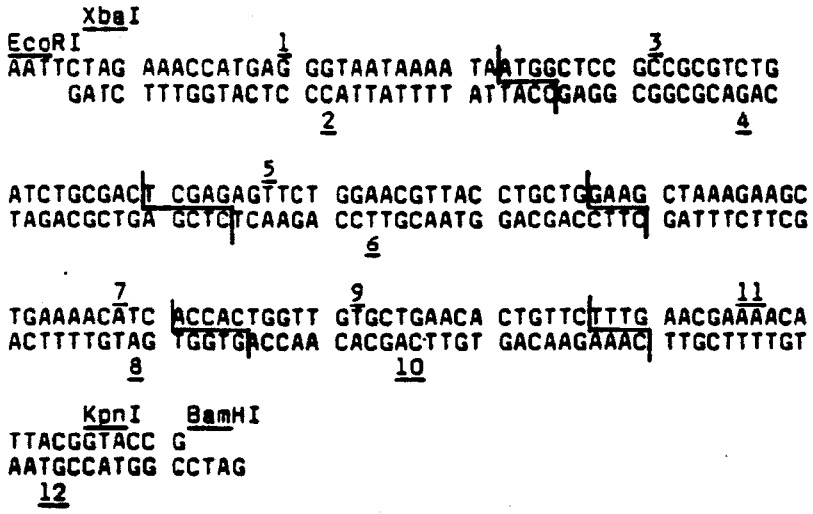


FIG. 11

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

ECEPO SECTION 2

EcoRI KpnI
A ATTCGGTACC AGACACCAAG GTAACTTCT ACGCTTGGAA ACGTATGGAA
GCCATGG TCTGTGGTTC CAATTGAAGA TGCGAACCTT TGCATACCTT
1 2 3 4

GTTGGTCAAC AAGCAGTTGA AGTTTGGCAG GGTCTGGCAC TGCTGAGCGA
CAACCAGTTG TTCGTCAACT TCAAACCGTC CCAGACCGTG ACGACTCGCT
5 6 7 8

GGCTGTACTG CGTGGCCAGG CACTGCTGGT AAACTCCTCT CAGCCGTGGG
CCGACATGAC GCACCGGTCC GTGACGACCA TTTGAGGAGA GTCGGCACCC
9 10 11 12

AACCGCTGCA GCTGCATGTT GACAAAGCAG TATCTGGCCT GAGATCTG
TTGGCGACGT CGACGTACAA CTGTTTCGTC ATAGACCGGA CTCTAGACTAC
13 14 15 16

EcoRI BglII BamHI

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

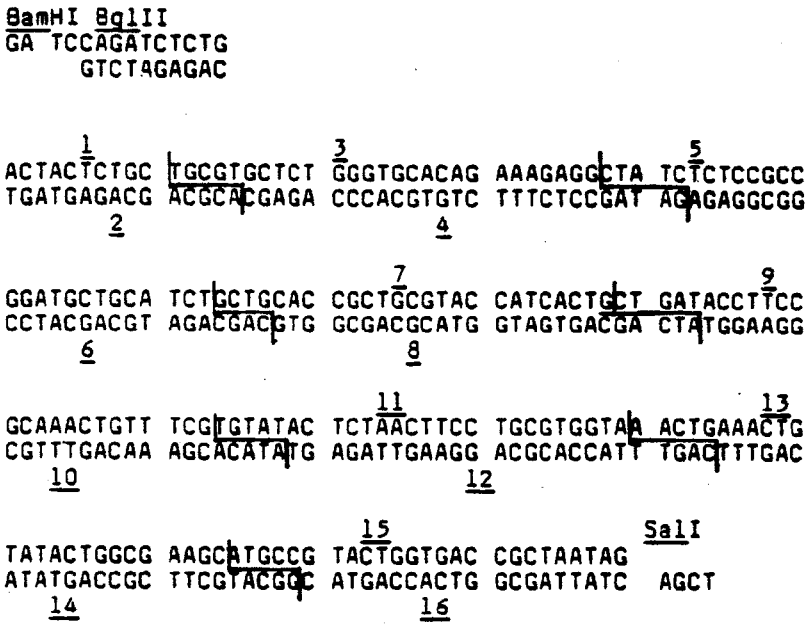


FIG. 15

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

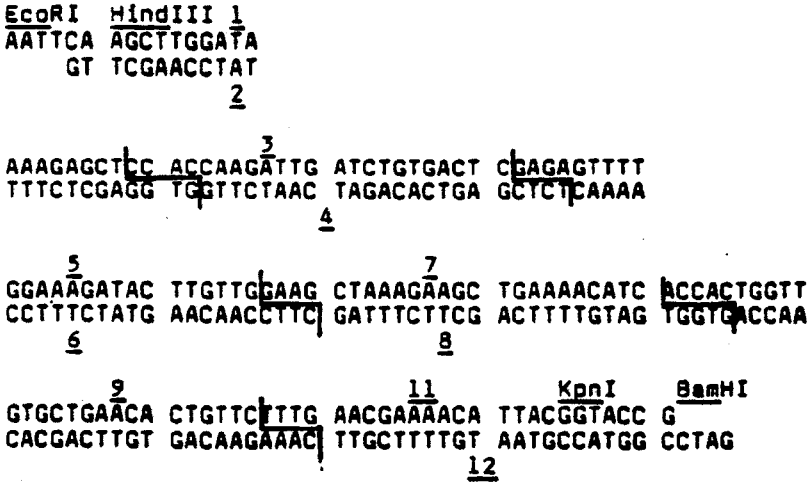


FIG. 17

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

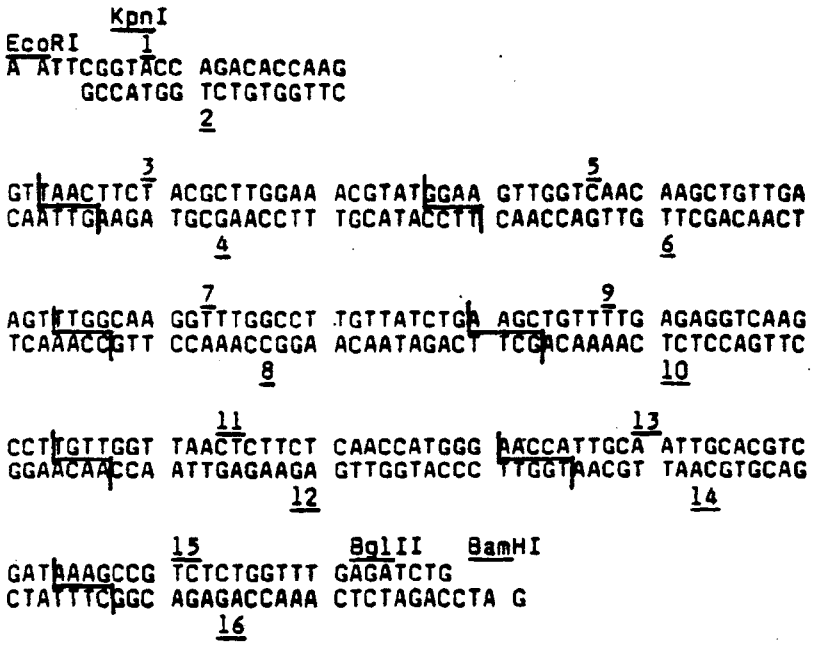


FIG. 19

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

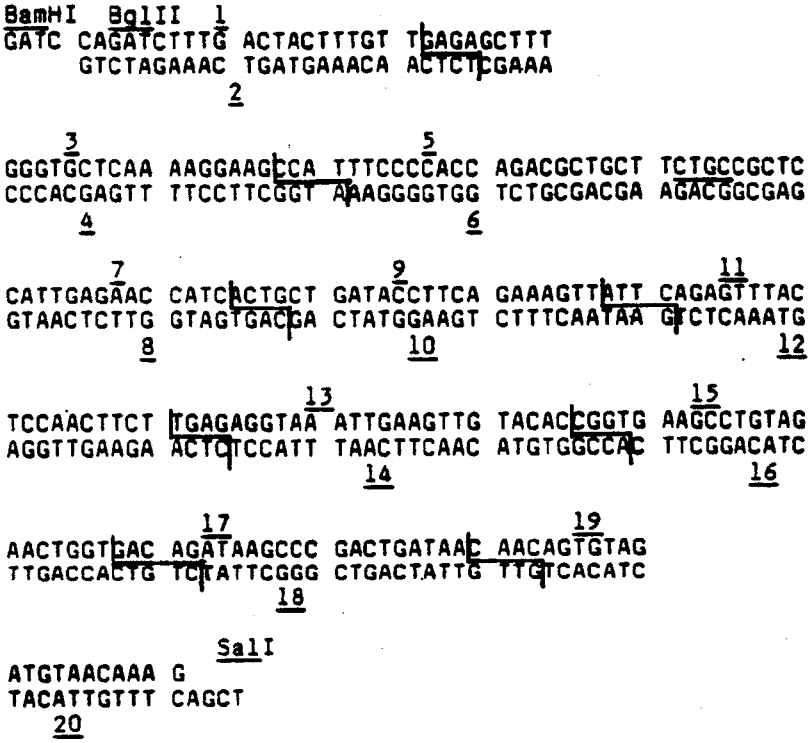


FIG. 21

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. In 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.

The development of specific DNA sequences for splicing into DNA vectors is accomplished by a variety of techniques, depending to a great deal on the degree of "foreignness"of the "donor" to the projected host and the size of the polypeptide to be expressed in the host. At the risk of over-simplification, it can be stated that three alternative principal methods can be employed: (1) the "isolation" of double-stranded DNA sequence from the genomic DNA of the donor; (2) the chemical manufacture of a DNA sequence providing a code for a polypeptide of interest; and (3) the in vitro synthesis of a double-stranded DNA sequence by enzymatic "reverse transcription" of mRNA isolated from donor cells. The last-mentioned methods which involve formation of a DNA "complement" of mRNA are generally referred to as "cDNA" methods.

Manufacture of DNA sequences is frequently the method of choice when the entire sequence of amino acid residues of the desired polypeptide product is known. DNA manufacturing procedures of co-owned, co-pending U.S. patent application Ser. No. 483,451, by Alton, et al., (filed Apr. 15, 1983 and corresponding to PCT US83/00605, published Nov. 24, 1983 as WO83/0405), for example, provide a superior means for accomplishing such highly desirable results as: providing for the presence of alternate codons commonly found in genes which are highly expressed in the host organism selected for expression (e.g., providing yeast or *E.coli* "preference" codons); avoiding the presence of untranslated "intron" sequences (commonly present in mammalian genomic DNA sequenoes and mRNA transcripts thereof) which are not readily processed by procaryotic host cells; avoiding expression of undesired "leader" polypeptide sequences commonly coded for by genomic DNA and cDNA sequences but frequently not readily cleaved from the polypeptide of interest by bacterial or yeast host cells; providing for ready insertion of the DNA in convenient expression vectors in association with desired promoter/regulator and terminator sequences; and providing for ready construction of genes coding for polypeptide fragments and analogs of the desired polypeptides.

When the entire sequence of amino acid residues of the desired polypeptide is not known, direct manufacture of DNA sequences is not possible and isolation of DNA sequences coding for the polypeptide by a cDNA method becomes the method of choice despite the potential drawbacks in ease of assembly of expression vectors capable of providing high levels of microbial expression referred to above. Among the standard procedures for isolating cDNA sequences of interest is the preparation of plasmid-borne cDNA "libraries" derived from reverse transcription of mRNA abundant in donor cells selected as responsible for high level expression of genes (e.g., libraries of cDNA derived from pituitary cells which express relatively large quantities of growth hormone products). Where substantial portions of the polypeptide's amino acid sequence are known, labelled, singlestranded DNA probe sequences duplicating a sequence putatively present in the "target" cDNA may be employed in DNA/DNA hybridization procedures carried out on cloned copies of the cDNA which have been denatured tc single stranded form. [See, generally, the disclosure and discussions of the art provlded in U.S. Pat. No. 4,394,443 to Weissman, et al. and the recent demonstrations of the use of long oligonucleotide hybridization probes reported in Wallace, et al., *Nuc.Acids Res.*, 6, pp. 3543–3557 (1979), and Reyes, et al., *P.N.A.S. (U.S.A.)*, 79, pp. 3270–3274 (1982), and Jaye, et al., *Nuc.Acids Res.*, 11, pp. 2325–2335 (1983). See also, U.S. Pat No. 4,358,535 to Falkow, et al., relating to DNA/DNA hybridization procedures in effecting diagnosis; published European patent application Nos. 0070685 and 0070687 relating to light-emitting labels on single stranded polynucleotide probes; Davis, et al., "A Manual for Genetic Engineering, Advanced Bacterial Genetics", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1980) at pp. 55–58 and 174–176, relating to colony and plaque hybridization techniques: and, New England Nuclear (Boston, Mass.) brochures for "Gene Screen" hybridization Transfer Membrane materials providing instruction manuals for the transfer and hybridization of DNA and RNA, Catalog No. NEF-972.]

Among the more signficant recent advances in hybridization procedures for the screening of recombinant clones is the use of labelled mixed synthetic oligonucleotide probes, each of which is potentially the complete complement of a specific DNA sequence in the hybridization sample including a heterogenous mixture of single stranded DNAs or RNAs. These procedures are acknowledged to be especially useful in the detection of cDNA clones derived from sources which provide extremely low amounts of mRNA sequences for the polypeptide of interest. Briefly put, use of stringent hybridization conditions directed toward avoidance of non-specific binding can allow, e.g., for the autoradiographic visualization of a specific cDNA clone upon the event of hybridization of the target DNA to that single probe within the mixture which is its complete complement. See generally, Wallace, et al., *Nuc.Acids Res.*, 9, pp. 879–897 (1981); Suggs, et al. *P.N.A.S. (U.S.A.)*, 78, pp. 6613–6617 (1981); Choo, et al., *Nature*, 299, pp. 178–180 (1982); Kurachi, et al., *P.N.A.S. (U.S.A.)*, 79, pp. 6461–6464 (1982); Ohkubo, et al., *P.N.A.S. (U.S.A.)*, 80, pp. 2196–2200 (1983): and Kornblihtt, et al. *P.N.A.S. (U.S.A.)*, 80, pp. 3218–3222 (1983). In general, the mixed probe procedures of Wallace, et al. (1981), supra, have been expanded upon by various workers to the point where reliable results have reportedly been obtained in a cDNA clone isolation using a 32 member mixed "pool" of 16-base-long (16-mer) oligonucleotide probes of uniformly, varying DNA sequences together with a single 11-mer to effect a two-site "positive" confirmation of the presence of cDMA of interest. See, Singer-Sam, et al., *P.N.A.S. (U.S.A.)*, 80, pp. 802–806 (1983).

The use of genomic DNA isolates is the least common of the three above-noted methods for developing specific DNA sequences for use in recombinant procedures. This is especially true in the area of recombinant procedures directed to securing microbial expression of mammalian polypeptides and is due, principally to the complexity of mammalian genomic DNA. Thus, while reliable procedures exist for developing phage-borne libraries of genomic DMA of human and other mammalian species origins [See, e.g., Lawn, et al. *Cell*, 15, pp. 1157–1174 (1978) relating to procedures for generating a human genomic library commonly referred to as the "Maniatis Library"; Karn et al. *P.N.A.S. (U.S.A.)*, 77, pp. 5172–5176 (1980) relating to a human genomic library based on alternative restriction endonuclease fragmentation procedure; and Blattner, et al., *Science*, 196 pp. 161–169 (1977) describing construction of a bovine genomic library] there have been relatively few successful attempts at use of hybridization procedures in isolating genomic DNA in the absence of extensive

foreknowledge of amino acid or DNA sequences. As one example, Fiddes, et al., *J.Mol. and App.Genetics*, 1, pp. 3–18 (1981) report the successful isolation of a gene coding for the alpha subunit of the human pituitary glycoprotein hormones from the Maniatis Library through use of a "full length" probe including a complete 621 base pair fragment of a previously-isolated cDNA sequence for the alpha subunit. As another example, Das, et al., *P N.A.S. (U.S.A.)* 80pp. 1531–1535 (1983) report isolation of human genomic clones for human HLA-DR using a 175 base pair synthetic oligonucleotide. Finally, Anderson, et al., *P.N.A.S. (U.S.A.)*, 80, pp. 6838–6842 (1983) report the isolation of genomic clone for bovine panoreatic trypsin inhibitor (BPTI) using a single probe 86 base pairs in length and constructed according to the known amino acid sequence of BPTI. The authors note a determination of poor prospects for isolating mRNA suitable for synthesis of a cDNA library due to apparent low levels of mRNA in initially targeted parotid gland and lung tissue sources and then address the prospects of success in probing a genomic library using a mixture of labelled probes, stating: "More generally, mixed-sequence oligodeoxynucleotide probes have been used to isolate protein genes of unknown sequence from cDNA libraries. Such probes are typically mixtures of 8–32 oligonucleotides, 14–17 nucleotides in length, representing every possible codon combination for a small stretch (5–6 residues) of amino acid sequence. Under stringent hybridization conditions that discriminate against incorrectly base-paired probes, these mixtures are capable of locating specific gene sequences in clone librarles of low-to-moderate complexity. Nevertheless, because of their short length and heterogeneity, mixed probes often lack the specificity reguired for probing sequences as complex as a mammalian genome. This makes such a method impractical for the isolation of mammalian protein genes when the corresponding mRNAs are unavailable." (Citations omitted).

There thus continues to exist a need in the art for improved methods for effecting the rapid and efficient isolation of cDNA clones in instances where little is known of the amino acid sequence of the polypeptide coded for and where "enriched" tissue sources of mRNA are not readily available for use in constructing cDNA libraries. Such improved methods would be especially useful if they were applicable to isolating mammalian genomic clones where sparse information is available concerning amino acid sequences of the polypeptide coded for by the gene sought.

B. Erythropoietin As A polypeptide Of Interest

Erythropoiesis, the production of red blood cells, occurs continuously throughout the human life span to offset cell destruction. Erythropoiesis is a very precisely controlled physiological mechanism enabling sufficient numbers of red blood cells to be available in the blood for proper tissue oxygenation, but not so many that the cells would impede circulation. The formation of red blood cells occurs in the bone marrow and is under the control of the hormone, erythropoietin.

Erythropoietin, an acidic glycoprotein of approximately 34,000 dalton molecular weight, may occur in three forms: α, β and asialo. The α and β forms differ slightly in carbohydrate components have the same potency, biological activity and molecular weight. The asialo form is an α or β form with the terminal Carbohydrate (sialic acid) removed. Erythropoietin is present in very low concentrations in plasma when the body is in a healthy state wherein tissues receive sufficient oxygenation from the existing number of erythrocytes. Tbis normai low concentration is enough to stimulate replacement of red blood cells which are lost normally through aging.

The amount of erythropoietin in the circulation is increased under conditions of hypoxia when oxygen transport by blood cells in the circulation is reduced. Hypoxia may be caused by loss of large amounts of blood through hemorrhage, destruction of red blood cells by over-exposure to radiation, reduction in oxygen intake due to high altitudes or prolonged unconsciousness, or various forms of anemia. In response to tissues undergoing hypoxic stress, erythropoietin will increase red blood cell production by stimulating the conversion of primitive precursor cells in the bone marrow into proerythroblasts which subsequently mature, synthesize hemoglobin and are released into the circulation as red blood cells. When the number of red blood cells in circulation is greater than needed for normal tissue oxygen requirements, erythropoietin in circulation is decreased.

See generally, Testa, et al., *Exp.Mematol.*, 8(*Supp* 8), 144–152 (1980); Tong, et al., *J.Biol.Chem.*, 256(24), 12666–12672 (1981); Goldwasser, *J.Cell.Physiol.*, 110(Supp. 1), 133–135 (1982); Finch, Blood, 60(6), 1241–1246 (1982); Sytowski, et al., *Expt.Hematol.*, 8(Supp. 8), 52–64 (1980: Naughton, *Ann.Clin.Lab.Sci.*, 13(5), 432–438 (1983); Weiss, et al., *Am.J.Vet.Res.*, 44(10),1832–1835 (1983); Lappin, et al., *Exp.Hematol.*, 11(7), 661–666 (1983); Baciu, et al., *Ann.N.Y.Acad.Sci.*, 414, 66–72 (1983); Murphy, et al., *Acta.Haematologica Japonica*, 46(7), 1380–1396 (1983); Desspyris, et al., *Brit.J.Haematol.*, 56, 295–306 (1984); and, Emmanouel, et al., *Am.J.Physiol.*, 247 (1 Pt 2), F168–76 (1984).

Because erythropoietin is essential in the process of red blood cell formation, the hormone has potential useful application in both the diagnosis and the treatment of blood disorders characterized by low or defective red blood cell production. See, generally, Pennathur-Das, et al., *Blood*, 63(5), 1168–71 (1984) and Haddy, *Am.Jour.Ped.Hematol./Oncol.*, 4, 191–196, (1982) relating to erythropoietin in possible therapies for sickle cell disease, and Eschbach, et al. *J.Clin.Invest.*, 74(2), pp. 434–441, (1984), describing a therapeutic regimen for uremic sheep based on in vivo response to erythropoietin-rich plasma infusions and proposing a dosage of 10 U EPO/kg per day for 15–40 days as corrective of anemia of the type associated with chronic renal failure. See alo, Krane, *Henry Ford Hosp.Med.J.*, 31(3), 177–181 (1983).

It has recently been estimated that the availability of erythropoietin in quantity would allow for treatment each year of anemias of 1,600,000 persons in the United States alone. See, e.g., Morrison, "Bioprocessing in Space—an Overview", pp. 557–571 in The World Biotech Report 1984, Volume 2: USA, (Online Publications, New York, N.Y. 1984). Recent studies have provided a basis for projection of efficacy of erythropoietin therapy in a variety of disease states, disorders and states of hematologic irregularity: Vedovato, et al., *Acta.Haematol*, 71, 211–213 (1984) (beta-thalassemia); Vichinsky, et al., *J.Pediatr.*, 105(1), 15–21 (1984) (cystic fibrosis); Cotes, et al., *Brit.J.Obstet.Gyneacol.*, 90(4), 304–311 (1983) (pregancy, menstrual disorders); Haga, et al., *Acta.Pediatr.Scand.*, 72, 827–831 (1983) (early anemia of prematurity); Claus-Walker, et al., *Arch.-Phys.Med.Rehabil.*, 65, 370–374 (1984), (spinal cord

injury); Dunn, et al., *Eur.JAppl.Physiol.*, 52, 178–182 (1984) (space flight); Miller, et al., *Brit.J.Haematol.*, 52, 545–590 (1982), (acute blood loss); Udupa, et al., *J.Lab.-Clin.Med.*, 103(4), 574–580 and 581–588 (1984); and Lipschitz, et al., *Blood*, 63(3), 502–509 (1983) (aging); and Dainiak, et al., *Cancer*, 51(6), 1101–1106 1 (1983) and Schwartz et al., *Otolaryngol.*, 109, 269–272 (1983) (various neoplastic disease states accompanied by abnormal erythropoiesis).

Prior attempts to obtain erythropoletin in good yield from plasma or urine have proven relatively unsuccessful. Complicated and sophisticated laboratory techniques are necessary and generally result in the collection of very small amounts of impure and unstable extracts containing erythropoietin.

U.S. Pat. No. 3,033,753 describes a method for partially purifying erythropoietin from sheep blood plasma which provides low yields of a crude solid extract containing erythropoietin.

Initial attempts to isolate erythropoietin from urine yielded unstable, biologically inactive preparations of the hormone. U.S. Pat. No. 3,865,801 describes a method of stabilizing the biological activity of a crude substance containing erythropoietin recovered from urin. The resulting crude preparation containing erythropoietin purportedly retains 90% of erythropoietin activity, and is stable.

Another method of purifying human erythropoietin from urine of patients with aplastic anemia is described in Miyake, et al., *J. Biol.Chem.*, Vol. 252, No. 15 Aug. 10, 1977), pp. 5558–5564. This seven-step procedure includes ion exchange chromatography, ethanol precipitation, gel filtration, and adsorption chromatography, and yields a pure erythropoietin preparation with a potency of 70,400 units/mg of protein in 21% yield.

U.S. Pat. No. 4,397,840 to Takezawa, et al. describes methods for preparing "an erythropoietin product" from healthy human urine specimens with weakly basic ion exchangers and proposes that the low molecular weight products obtained "have no inhibitory effects against erythropoietin.

U.K. patent application No. 2,085,887 by Sugimoto, et al., published May 6, 1982, describes a process for the production of hybrid human lymphoblastoid cells, reporting production levels ranging from 3 to 420 Units of erythropoietin per ml of suspension of cells (distributed into the cultures after mammalian host propagation containing up to 10^7 cells per ml. At the highest production levels asserted to have been obtained, the rate of erythropoietin production could be calculated to be from 40 to about 4,000 units/10^6 cells/48 hours in in vitro culture following transfer of cells from in vivo propagation systems. (See also the equivalent U.S. Pat. No. 4,377,513.) Numerous proposals have been made for isolation of erythropoietin from tissue sources, including neoplastic cells, but the yields have been quite low. See, e.g., Jelkman, et al., *Expt.Hematol.*, 11(7), 581–588 (1983); Tambourin, et al., *P.N.A.S. (U.S.A.)*, 80, 6269–6273 (1983); Katsuoka, et al., *Gann*, 74, 534–541 (1983); Hagiwara, et al., *Blood*, 63(4), 828–835 (1984); and Choppin, et al., *Blood*, 64(2), 341–347 (1984).

Other isolation techniques utilized to obtain purified erythropoietin involve immunological procedures. A polyclonal, serum-derived antibody directed against erythlopoietin is developed by injecting an animal, preferably a rat or rabbit, with human erythropoietin. The injected human erythropoietin is recognized as a foreign antigenic substance by the immune system of the animal and elicits production of antibodies against the antigen. Differing cells responding to stimulation by the antigenic substance produce and release into circulation antibodies slightly different from those produced by other responding cells. The antibody activity remains in the serum of the animal when its blood is extracted. While unpurified serum or antibody preparations purified as a serum immunoglobulin G fraction may then be used in assays to detect and complex with human erythropoietin, the materials suffer from a major disadvantage. This serum antibody, composed of all the different antibodies produced by individual cells, is polyclonal in nature and will complex with components in crude extracts other than erythropoietin alone.

Of interest to the background of the present invention are recent advances in the art of developing continuous cultures of cells capable of producing a single species of antibody which is specifically immunologically reactive with a single antigenic determinant of a selected antigen. See, generally, Chisholm, *High Technology*, Vol. 3, No. 1, 57–63 (1983). Attempts have been made to employ cell fusion and hybridization techniques to develop "monoclonal" antibodies to erythropoietin and to employ these antibodies in the isolation and quantitative detection of human erythropoietin. As one example, a report of the successful development of mouse-mouse hybridoma cell lines secreting monoclonal antibodies to human erythropoietin appeared in abstract form in Lee-Huang, Abstract No. 1463 of *Fed.Proc.*, 41, 520 (1982). As another example, a detailed description of the preparation and use of a monoclonal, anti-erythropoietin antibody appears in Weiss, et al., *P.N.A.S. (U.S.A.)*, 79, 5465–5469 (1982). See also, Sasaki, *Biomed.Biochim Acta.*, 42(11/12), S202–S206 (1983) Yanagawa, et al., *Blood*, 64(2), 357–364 (1984); Yanagawa, et al., *J.Biol.-Chem.*, 259(5), 2707–2710 (1984); and U.S. Pat. No. 4,465,624.

Also of interest to the background of the invention are reports of the immunological activity of synthetic peptides which substantially duplicate the amino acid sequence extant in naturally-occurring proteins, glycoproteins and nucleoproteins. More specifically, relatively low moiecular weight polypeptides have been shown to participate in immune reactions which are similar in duration and extent to the immune reactions of physiologically significant proteins such as viral antigens, polypeptide hormones, and the like. Included among the immune reactions of such polypeptides is the provocation of the formation of specific antibodies in immunologically active animals. See, e.g., Lerner, et al., *Cell*, 23, 309–310 (1981); Ross, et al., *Nature*, 294, 654–656 (1981); Walter, et al., *P.N.A.S. (U.S.A.)*, 77, 5197–5200 (1980); Lerner, et al., *P.N.A.S. (U.S.A.)*, 78, 3403–3407 (1981); Walter, et al., *P.N.A.S. (U.S.A.)*, 78, 4882–4886 (1981); Wong, et al., *P.N.A.S. (U.S.A.)*, 78, 7412–7416 (1981); Green, et al. *Cell*, 28, 477–487 (1982); Nigg, et al., *P.N.A.S. (U.S.A.)*, 79, 5322–5326 (1982); Baron, et al., *Cell*, 28, 395–404 (1982); Dreesman, et al., *Nature*, 295, 158–160 (1982); and Lerner, *Scientific American*, 248, No. 2, 66–74 (1983). See, also, Kaiser, et al., *Science*, 223 pp. 249–255 (1984) relating to biological and immunological activities of synthetic peptides which approximately share secondary structures of peptide hormones but may not share their primary structural conformation. The above studies relate, of course, to amino acid sequences of proteins other than erythropoietin, a substance for which no substantial amino acid sequence information has been published. In

co-owned, co-pending U.S. patent application Ser. No. 463,724, filed Feb. 4, 1983, by J. Egrie, published Aug. 22, 1984 as European Patent Application No. 0 116 446, there is described a mouse-mouse hybridoma cell line (A.T.C.C. No. HB8209) which produces a highly specific monoclonal, anti-erythropoietin antibody which is also specifically immunoreactive with a polypeptide comprising the following sequence of amino acids:

NH_2-Ala-Pro-Pro-Arg-Leu-Ile-Cys-Asp-Ser-Arg-Val-Leu-Glu-Arg-Tyr-Leu-Leu-Glu-Ala-Lys-COOH.

The polypeptide sequence is one assigned to the first twenty amino acid residues of mature human erythlopoietin isolated according to the method of Miyake, et al., *J.Biol.Chem.*, 252, 5558–5564 (1977) and upon which amino acid analysis was performed by the gas phase sequencer (Applied Biosystems, Inc.) according to the procedure of Hewick, M., et al., *J.Biol.Chem.*, 256, 7990–7997 (1981). See, also, Sue, et al., *Proc.Nat.Acad.Sci. (U.S.A.)*, 80, pp. 3651–3655 (1983) relating to development of polyclonal antibodies against a synthetic 26-mer based on a differing amino acid sequence, and Sytowski, et al., *J.Immunol. Methods*, 69, pp. 181–186 (1984).

While polyclonal and monoclonal antibodies as described above provide highly useful materials for use in immunoassays for detection and quantification of erythropoietin and can be useful in the affinity purification of erythropoietin, it appears unlikely that these materials can readily provide for the large scale isolation of quantities of erythropoietin from mammalian sources sufficient for further analysis, clinical testing and potential wide-ranging therapeutic use of the substance in treatment of, e.g., chronic kidney disease wherein diseased tissues fail to sustain production of erythropoietin. It is consequently projected in the art that the best prospects for fully characterizing mammalian erythropoietin and providing large quantities of it for potential diagnostic and clinical use involve successful application of recombinant procedures to effect large scale microbial synthesis of the compound.

While substantial efforts appear to have been made in attempted isolation of DNA sequences coding for human and other mammalian species erythropoietin, none appear to have been successful. This is due principally to the scarcity of tissue sources, especially human tissue sources, enriched in mRNA such as would allow for construction of a cDNA library from which a DNA sequence coding for erythropoietin might be isolated by conventional techniques. Further, so little is known of the continuous sequence of amino acid residues of erythropoietin that it is not possible to construct, e.g., long polynucleotide probes readily capable of reliable use in DNA/DNA hybridization screening of cDNA and especially genomic DNA libraries. Illustratively the twenty amino acid sequence employed to generate the above-named monoclonal antibody produced by A.T.C.C. No. HB8209 does not admit to the construction of an unambiguous, 60 base oligonucleotide probe in the manner described by Anderson, et al., supra. It is estimated that the human gene for erythropoietin may appear as a "single copy gene" within the human genome and, in any event, the genetic material coding for human erythropoietin is likely to constitute less than 0.00005% of total human genomic DNA which would be present in a genomic library.

To date, the most successful of known reported attempts at recombinant-related methods to provide DNA sequences suitable for use in microbial expression of isolatable quantities of mammalian erythropoietin have fallen far short of the goal. As an example, Farber, et al. *Exp.Hematol.*, 11. Supp. 14, Abstract 101 (1983) report the extraction of mRNA from kidney tissues of phenylhydrazine-treated baboons and the injection of the mRNA into *Xenopus laevis* oocytes with the rather transitory result of in vitro production of a mixture of "translation products" which included among them displaying biological properties of erythropoietin. More recently, Farber, et al., *Blood*, 62, No. 5, Supp. No. 1, Abstract 392, at page 122*a* (1983) reported the in vitro translation of human kidney mRNA by frog oocytes. The resultant translation product mixture was estimated to include on the order of 220 mU of a translation product having the activity of erythropoietin per microgram of injected mRNA. Whlle such levels of in vitro translation of exogenous mRNA coding for erythropoietin were acknowledged to be guite low (compared even to the prior reported levels of baboon mRNA translation into the sought-for product) it was held that the results confirm the human kidney as a site of erythropoietin expression, allowing for the construction of an enriched human kidney cDNA library from which the desired gene might be isolated. [See also, Farber, Clin.Res., 31(4), 769A (1983).]

Since the filing of U.S. patent application Ser. Nos. 561,024 and 582,185, there has appeared a single report of the cloning and expression of what is asserted to have been human erythropoietin cDNA in *E coli.* Brlefly put, a number of cDNA clones were inserted into *E.coli* plasmids and β-lactamase fusion products were noted to be immunoreactive with a monoclonal antibody to an unspecified "epitope" of human erythropoietin. See, Lee-Huang, *Proc. Nat. Acad. Sci.* (*U.S.A*), 81, pp. 2708–2712 (1984).

BRIEF SUMMARY

The present invention provides, for the first time, novel purified and isolated polypeptide products having part or all of the primary structural conformation (i.e., continuous sequence of amino acid residues) and one or more of the biological properties (e.g., immunological properties and in vivo and in vitro biological activity) of naturally-occurring erythropoietin, including allelic variants thereof. These polypeptides are also uniquely characterired by being the product of procaryotic or eucaryotic host expression (e.g., by bacterial, yeast and mammalian cells in culture) of exogenous DNA sequences obtained by genomic or cDNA cloning or by gene synthesis. Products of microbial expression in vertebrate (e.g., mammalian and avian) cells may be further characterized by freedom from association with human proteins or other contaminants which may be associated with erythropoietin in its natural mammalian cellular environment or in extracellular fluids such as plasma or urine. The products of typical yeast (e.g., *Saccharomyces cerevisiae*) or procaryote (e.g., *E.coli*) host cells are free of association with any mammalian proteins. Depending upon the host employed, polypeptides of the invention may be glycosylated with mammallan or other eucaryotic carbohydrates or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue (at position −1).

Novel glycoprotein products of the invention include those having a primary structural conformation sufficiently duplicative of that of a naturally-occurring (e.g., human) erythropoietin to allow possession of one or more of the biological properties thereof and having an average carbohydrate composition which differs from that of naturally-occurring (e.g., human) erythropoietin.

Vertebrate (e.g., COS-1 and CHO) cells provided by the present invention comprise the first cells ever available whioh can be propagated in vitro continuously and which upon growth in culture are capable of producing in the medium of their growth in excess of 100U (preferably in excess of 500U and most preferably in excess of 1,000 to 5,000U) of erythropoietin per 10^6 cells in 48 hours as determined by radioimmunoassay.

Also provided by the present invention are synthetic polypeptides wholly or partially duplicative of continuous sequences of erythropoietin amino acid residues which are herein for the first time elucidated. These sequences, by virtue of sharing primary, secondary or tertiary structural and conformational characteristics with naturally-occurring erythropoietin may possess biological activity and/or immunological properties in common with the naturally-occurring product such that they may be employed as biologically active or immunological substitutes for erythropoietin in therapeutic and immunological processes. Correspondingly provided are monoclonal and polyclonal antibodies generated by standard means which are immunoreactive with such polypeptides and, preferably, also immunoreactive with naturally-occurring erythropoietin.

Illustrating the present invention are cloned DNA sequences of monkey and human species origins and polypeptide sequences suitably deduced therefrom which represent, respectively, the primary structural conformation of erythropoietins of monkey and human species origins.

Also provided by the present invention are novel blologically functional viral and circular plasmid DNA vectors incorporating DNA sequences of the invention and microbial (e.g., bacterial, yeast and mammalian cell) host organisms stably transformed or transfected with such vectors. Correspondingly provided by the invention are novel methods for the production of useful polypep-tides comprising cultured growth of such transformed or transfected microbial hosts under conditions facilitative of large scale expression of the exogenous, vector-borne DNA sequences and isolation of the desired polypeptides from the growth medium, cellular lysates or cellular membrane fractions.

Isolation and purification of microbially expressed polypeptides provided by the invention may be by conventional means including, e.q., preparative chromatographic separations and immunological separations involving monoclonal and/or polyclonal antibody preparations.

Having herein elucidated the sequence of amino acid residues of erythropoietin, the present invention provides for the total and/or partial manfucture of DNA sequences coding for erythropoietin and including such advantageous characteristics as incorporation of codons "preferred" for expression by selected non-mammalian hosts, provision of sites for cleavage by restriction endonuclease enzymes and provision of additional initial, terminal or intermediate DNA sequences which facilitate construction of readily expressed vectors. Correspondingly, the present invention provides for manufacture (and development by site specific mutagenesis of cDNA and genomic DNA) of DNA sequences coding for microbial expression of polypeptide analogs or derivatives of erythropoietin which differ from naturally-occurring forms in terms of the identity or location of one or more amino acid residues (i.e., deletion analogs containing less than all of the residues specified for EPO and/or substitution analogs wherein one or more residues specified are replaced by other residues and/or addition analoqs wherein one or more amino acid residues is added to a terminal or medial portion of the polypeptide); and which share some or all the properties of naturally-occurring forms.

Novel DNA sequences of the invention include all sequences useful in securing expression in procaryotic or eucaryotic host cells of polypeptide products having at least a part of the primary structural ccnformation and one or more of the biological propezties of erythropoietin which are comprehended by: (a) the DNA sequences set out in Tables V and VI herein or their complementary strands; (b) DNA sequences which hybridize (under hybridization conditions such as illustrated herein or more stringent conditions) to DNA sequences defined in (a) or fragments thereof; and (c) DNA sequences which, but for the degeneracy of the genetic code, would hybridize to DMA sequences defined in (a) and (b) above. Specifically comprehended in part (b) are genomic DNA sequences encoding allelic variant forms of monkey and human erythropoietin and/or encoding other mammalian species of erythropoietin. Specifically comprehended by part (c) are manufactured DNA sequences encoding EPO, EPO fragments and EPO analogs which DNA sequences may incorporate codons facilitating translation of messenger RNA in non-vertebrate hosts.

Comprehended by the present invention is that class of polypeptides coded for by portions of the DNA complement to the top strand human genomic DNA sequence of Table VI herein, i.e., "complementary inverted proteins" as described by Tramontano, et al., *Nucleic Acids Research*, 12, pp. 5049–5059 (1984).

Also comprehended by the invention are pharmaceutical compositions comprising effective amounts of polypeptide products of the invention together with suitable diluents, adjuvants and/or carriers which allow for provision of erythropoietin therapy, especially in the treatment of anemic disease states and most especially such anemic states as attend chronic renal failure.

Polypeptide products of the invention may be "labelled" by covalent association with a detectable marker substance (e.g., radiolabelled with ^{125}I) to provide reagents useful in detection and quantification of erythropoietin in solid tissue and fluid samples such as blood or urine. DNA products of the invention may also be labelled with detectable markers (such as radiolabels and non-isotopic labels such as biotin) and employed in DNA hybridization processes to locate the erythropoietin gene position and/or the position of any related gene family in the human, monkey and other mammalian species chromosomal map. They can also be used for identifying the erythropoietin gene disorders at the DMA level and used as gene markers for identifying neiqhboring genes and their disorders.

As hereinafter described in detail, the present invention further provides significant improvements in methods for detection of a specific single stranded polynucleotide of unknown sequence in a heterogeneous cellular

or viral sample including multiple single-stranded polynucleotides where

(a) a mixture of labelled single-stranded polynucleotide probes is prepared having uniformly varying sequences of bases, each of said probes being potentially specifically complementary to a sequence of bases which is putatively unique to the polynucleotide to be detected,

(b) the sample is fixed to a solid substrate,

(c) the substrate having the sample fixed thereto is treated to diminish further binding of polynucleotides thereto except by way of hybridization to polynucleotides in said sample,

(d) the treated substrate having the sample fixed thereto is transitorily contacted with said mixture of labelled probes under conditions facilitative of hybridization only between totally complementary polynucleotides, and,

(e) The specific polynucleotide is detected by monitoring for the presence of a hybridization reaction between it and a totally complementary probe within said mixture of labelled probes, as evidenced by the presence of a higher density of labelled material on the substrate at the locus of the specific polynucleotide in comparison to a background density of labelled material resulting from non-specific binding of labelled probes to the substrate.

The procedures are especially effective in situations dictating use of 64, 128, 256, 512, 1024 or more mixed polynucleotide probes having a length of 17 to 20 bases in DNA/DNA or RNA/RNA or DNA/RNA hybridizations.

As described infra, the above-noted improved procedures have illustratively allowed for the identification of cDNA clones coding for erythropoietin of monkey species origins within a library prepared from anemic monkey kidney cell mRNA. More specifically, a mixture of 128 uniformly varying 20-mer probes based on amino acid sequence information derived from sequencing fractions of human erythropoietin was employed in colony hybridization procedures to identify seven "positive" erythropoietin cDNA clones within a total of 200,000 colonies. Even more remarkably, practice of the improved procedures of the invention have allowed for the rapid isolation of three positive clones from within a screening of 1,500,000 phage plaques constituting a human genomic library. This was accomplished through use of the above-noted mixture of 128 20-mer probes together with a second set of 128 17-mer probes based on amino acid analysis of a different continuous sequence of human erythropoietin.

The above-noted illustrative procedures constitute the first known instance of the use of multiple mixed oligonucleotide probes in DNA/DNA hybridization processes directed toward isolation of mammalian genomic clones and the first known instance of the use of a mixture of more than 32 oligonucleotide probes in the isolation of cDNA clones.

Numerous aspects and advantages of the invention will be apparent to those skilled in the art upon consideration of the following detailed description which provides illustrations of the practice of the invention in its presently preferred embodiments.

Reference is made in FIGS. 1 through 21, wherein:

FIG. 1 is a graphic representation of a radioimmunoassay analysis of products of the invention;

FIGS. 2 through 4 illustrate vector construction according to the invention; and,

FIGS. 5 through 21 are DNA and polypeptide sequences according to the invention.

DETAILED DESCRIPTION

According to the present invention, DNA sequences encoding part or all of the polypeptide sequence of human and monkey species erythropoietin (hereafter, at times, "EPO") have been isolated and characterized. Further, the monkey and human origin DNA has been made the subject of eucaryotic and procaryotic expression providing isolatable quantities of polypeptides displaying bioloqical (e.g., immunological) properties of naturally-occurring EPO as well as both in vivo and in vitro biological activities of EPO.

The DNA of monkey species origins was isolated from a cDNA library constructed with mRNA derived from kidney tissue of a monkey in a chemically induced anemic state and whose serum was immunologically determined to include high levels of EPO compared to normal monkey serum. The isolation of the desired cDNA clones containing EPO encoding DNA was accomplished through use of DNA/DNA colony hybridization employing a pool of 128 mixed, radiolabelled, 20-mer oligonucleotide probes and involved the rapid screening of 200,000 colonies. Design of the oligonucleotide probes was based on amino acid sequence information provided by enzymatic fragmentation and sequencing a small sample of human EPO.

The DNA of human species origins was isolated from a human genomic DNA library. The isolation of clones containing EPO-encoding DNA was accomplished through DNA/DNA plaque hybridization employing the above-noted pool of 128 mixed 20-mer oligonucleotide probes and a second pool of 128 radiolabelled 17-mer probes whose sequences were based on amino acids sequence information obtained from a different enzymatic human EPO fragment.

Positive colonies and plaques were verified by means of dideoxy sequencing of clonal DNA using a subset of 16 sequences within the pool of 20-mer probes and selected clones were subjected to nucleotide sequence analysis resulting in deduction of primary structural conformation of the EPO polypeptides encoded thereby. The deduced polypeptide sequences displayed a high degree of homoloqy to each other and to a partial sequence generated by amino acid analysis of human EPO fraqments.

A selected positive monkey cDNA clone and a selected positive human genomic clone were each inserted in a "shuttle" DNA vector which was amplified in *E.coli* and employed to transfect mammalian cells in culture. Cultured growth of transfected host cells resulted in culture medium supernatant preparations estimated to contain as much as 3000 mU of EPO per ml of culture fluid.

The following examples are presented by way of illustration of the invention and are specifically directed to procedures carried out prior to identification of EPO encoding monkey cDNA clones and human genomic clones, to procedures resulting in such identification, and to the sequencing, development of expression systems and immunoloqical verification of EPO expression in such systems.

More particularly, Example 1 lis directed to amino acid sequencing of human EPO fraqments and construction of mixtures of radiolabelled probes based on the results of this sequencing. Example 2 is generally directed to procedures involved in the identification of

positive monkey cDNA clones and thus provides information concerning animal treatment and preliminary radioimmunoassay (RIA) analysis of animal sera. Example 3 is directed to the preparation of the cDNA library, colony hybridization screening and verification of positive clones, DNA sequencing of a positive cDNA clone and the generation of monkey EPO polypeptide primary structural conformation (amino acid sequence) information. Example 4 is directed to procedures involved in the identification of positive human genomic clones and thus provides information concerning the source of the genomic library, plaque hybridization procedures and verification of positive clones. Example 5 is directed to DNA sequencing of a positive genomic clone and the generation of human EPO polypeptide amino acid sequence information including a comparison thereof to the monkey EPO sequence information. Example 6 is directed to procedures for construction of a vector incorporating EPO-encoding DNA derived from a positive monkey cDNA clone, the use of the vector for transfection of COS-1 cells and cultured growth of the transfected cells. Example 7 is directed to procedures for construction of a vector incorporating EPO-encoding DNA derived from a positive human genomic clone, the use of the vector for transfection of COS-1 cells and the cultured growth of the transfected cells. Example 8 is directed to immunoassay procedures performed on media supernatants obtained from the cultured growth of transfected cells according to Example 6 and 7. Example 9 is directed to in vitro and in vivo biological activity of microbially expressed EPO of Examples 6 and 7.

Example 10 is directed to a development of mammalian host expression systems for monkey species EPO cDNA and human species genomic DNA involving Chinese hamster ovary ("CMO") cells and to the immunological and biological activities of products of these expression systems as well as charactezization of such products. Example 11 is directed to the preparation of manufactured genes encoding human species EPO and EPO analogs, which genes include a number of preference codons for expression in *E.coli* and yeast host cells, and to expression systems based thereon. Example 12 relates to the immunological and biological activity profiles of expression products of the systems of Example 11.

EXAMPLE 1

A. Human EPO Fragment Amino Acid Sequencing

Human EPO was isolated from urine and subjected to tryptic digestion resulting in the development and isolation of 17 discrete fragments in quantities approximating 100–150 picomoles.

Fragments were arbitrarily assigned numbers and were analyzed for amino acid sequence by microsequence analysis using a gas phase sequencer (Applied Biosystems) to provide the sequence information set out in Table I, below, wherein single letter codes are employed and "X" designates a residue which was not unambiguously determined.

TABLE I

Fragment No.	Sequence Analysis Result
T4a	A-P-P-R
T4b	G-K-L-K
T9	A-L-G-A-Q-K
T13	V-L-E-R
T16	A-V-S-G-L-R
T18	L-F-R

TABLE I-continued

Fragment No.	Sequence Analysis Result
T21	K-L-F-R
T25	Y-L-L-E-A-K
T26a	L-I-C-D-S-R
T26b	L-Y-T-G-E-A-C-R
T27	T-I-T-A-D-T-F-R
T28	E-A-I-S-P-P-O-A-A-M-A-A-P-L-R
T30	E-A-E-X-I-T-T-G-X-A-E-H-X-S-L-N-E-X-I-T-V-P
T31	V-Y-S-N-F-L-R
T33	S-L-T-T-L-L-R
T35	V-N-F-Y-A-W-K
T38	G-Q-A-L-L-V-X-S-S-Q-P-W-E-P-L-Q-L-H-V-D-K

B. Design and Construction of Oligonucleotide Probe Mixtures

The amino acid sequences set out in Table I were reviewed in the context of the degeneracy of the genetic code for the purpose of ascertaining whether mixed probe procedures could be applied to DNA/DNA hybridization procedures on cDNA and/or genomic DNA libraries. This analysis revealed that within Fragment No. T35 there existed a series of 7 amino acid residues (Val-Asn-Phe-Tyr-Ala-Trp-Lys) which could be uniquely characterized as encoded for by one of 128 possible DNA sequences spanning 20 base pairs. A first set of 128 20-mer oligonucleotides was therefore synthesized by standard phosphoamidite methods (See, e.g., Beaucage, et al., *Tetrahedron Letters*, 22, pp. 1859–1862 (1981) on a solid support according to the sequence set out in Table II, below.

TABLE II

Residue	Val	Asn	Phe	Tyr	Ala	Trp	Lys	
3'	CAA	TTG	AAG	ATG	CGA	ACC	TT	5'
	T	A	A	A	T			
	G				G			
	C				C			

Further analysis revealed that within fragment No. T38 there existed a series of 6 amino acid residues Gln-Pro-Trp-Glu-Pro-Leu) on the basis of which there could be prepared a pool of 128 mixed olignucleotide 17-mer probes as set out in Table III, below.

TABLE III

Residue	Gln	Pro	Trp	Glu	Pro	Leu	
3'	GTT	GGA	ACC	CTT	GGA	GA	5'
	C	T		C	T	A	
		G			G		
		C			C		

Oligonucleotide probes were labelled at the 5' end with qamma - ^{32}P-ATP, 7500-8000 Ci/mmole (ICN) using T_4 polynucleotide kinase (NEN).

EXAMPLE 2

A. Monkey Treatment Procedures

Female Cynomolgus monkeys *Macaca fascicularias* (2.5–3 kg, 1.5–2 years old) were treated subcutaneously with a pH 7.0 solution of phenylhydrazine hydrochloride at a dosage level of 12.5 mg/kg on days 1, 3 and 5. The hematocrit was monitored prior to each injection. On day 7, or whenever the hematocrit level fell below 25% of the initial level, serum and kidneys were harvested after administration of 25 mg/kg doses of keta-

mine hydrochloride. Harvested materials were immediately frozen in liquid nitrogen and stored at −70° C.

B. RIA for EPO

Radioimmunoassay procedures applied for quantitative detection of EPO in samples were conducted according to the following procedures:

An erythropoietin standard or unknown sample was incubated together with antiserum for two hours at 37° C. After the two hour incubation, the sample tubes were cooled on ice, ^{125}I-labelled erythropoietin was added, and the tubes were incubated at 0° C. for at least 15 more hours. Each assay tube contained 500 μl of incubation mixture consisting of 50 μl of diluted immune sera, 10,000 cpm of ^{125}I-erythropoietin, 5 μl trasylol and 0–250 μl either EPO standard or unknown sample, with PBS containing 0.1% BSA making up the remaining volume. The antiserum used was the second test bleed of a rabbit immunized with a 1% pure preparation of human urinary erythropoietin. The final antiserum dilution on the assay was adjusted so that the antibody-bound ^{125}I-EPO did not exceed 10–20% of the input total counts. In general, this corresponded to a final antiserum dilution of from 1:50,000 to 1:100,000.

The antibody-bound ^{125}I-erythropoietin was precipitated by the addition of 150 μl Staph A. After a 40 min. incubation, the samples were centrifuged and the pellets were washed two times with 0.75 ml 10 mM Tris-HCl pH 8.2 containing 0.15M NaCl, 2mM EDTA, and 0.05% Triton X-100. The washed pellets were counted in a gamma counter to determine the percent of ^{125}I-erythropoietin bound. Counts bound by pre-immune sera were subtracted from all final values to correct for nonspecific precipitation. The erythropoietin content of the unknown samples was determined by comparison to the standard curve.

The above procedure was applied to monkey serum obtained in Part A, above, as well as to the untreated monkey serum. Normal serum levels were assayed to contain approximately 36 mU/ml while treated monkey serum contained from 1000 to 1700 mU/ml.

EXAMPLE 3

A. Monkey cDNA Library Construction

Messenger RNA was isolated from normal and anemic monkey kidneys by the guanidinium thiocyanate procedure of Chirgwin, et al., *Biochemistry,* 18, p. 5294 (1979) and poly (A)+mRNA was purified by two runs of oligo(dT)-cellulose column chromatography as described at pp. 197–198 in Maniatis, et al., "Molecular Cloning, A Laboratory Manual" (Cold Springs Harbor Laboratory, Cold Springs, Harbor, N.Y., 1982). The cDNA library was constructed according to a modification of the general procedures of Okayama, et al., *Mol. and Cell. Biol.,* 2, pp. 161–170 (1982). The key features of the presently preferred procedures were as follows: (1) pUC8 was used as the sole vector, cut with PstI and then tailed with oligo dT of 60–80 bases in length; (2) HincII digestion was used to remove the oligo dT tail from one end of the vector; (3) first strand synthesis and oligo dG tailing was carried out according to the published procedure; (4) BamHI digestion was employed to remove the oligo dG tail from one end of the vector; and (5) replacement of the RNA strand by DNA was in the presence of two linkers (GATC-TAAAGACCGTCCCCCCCCC and ACGGTCTTTA) in a three-fold molar excess over the oligo dG tailed vector.

B. Colony Hybridization Procedures For Screening Monkey cDNA Library

Transformed *E.coli* were spread out at a density of 9000 colonies per 10×10 cm plate on nutrient plates containing 50 micrograms/ml Ampicillin. GeneScreen filters (New England Nuclear Catalog No. NEF-972) were pre-wet on a BHI-CAM plate (Bacto brain heart infusion 37 g/L, Casamino acids 2 g/L and agar 15 g/L, containing 500 micrograms/ml Chloramphenicol) and were used to lift the colonies off the plate. The colonies were grown in the same medium for 12 hours or longer to amplify the plasmid copy numbers. The amplified colonies (colony side up) were treated by serially placing the filters over 2 pieces of Whatman 3 MM paper saturated with each of the following solutions:

(1) 50 mM glucose—25 mM Tris-HCl (pH 8.0)—10 mM EDTA (pH 8.0) for five minutes;

(2) 0.5 M NaOH for ten minutes; and

(3) 1.0 M Tris-HCl (pH 7.5) for three minutes.

The filters were then air dried in a vacuum over at 80° C. for two hours.

The filters were then subjected to Proteinase K digestion through treatment with a solution containing 50 micrograms/ml of the protease enzyme in Buffer K [0.1M Tris-HCl (pH 8.0) —0.15M NaCl —10 mM EDTA (pH 8.2) —0.2% SDS]. Specifically, 5 ml of the solution was added to each filter and the digestion was allowed to proceed at 55° C. for 30 minutes, after which the solution was removed.

The filters were then treated with 4 ml of a prehybridization buffer (5×SSPE —0.5% SDS —100 micrograms/ml SS *E.coli* DNA - 5×BFP). The prehybridization treatment was carried out at 55° C., generally for 4 hours or longer, after which the prehybridization buffer was removed.

The hybridization process was carried out in the following manner. To each filter was added 3 ml of hybridization buffer (5×SSPE —0.5% SDS —100 micrograms/ml yeast tRNA) containing 0.025 picomoles of each of the 128 probe sequences of Table II (the total mixture being designated the EPV mixture) and the filters were maintained at 48° C. for 20 hours. This temperature was 2° C. less than the lowest of the calculated dissociation temperatures (Td) determined for any of the probes.

Following hybridization, the filters were washed three times for ten minutes on a shaker with 6×SSC —0.1% SDS at room temperature and washed two to three times with 6×SSC —1% SDS at the hybridization temperature (48° C.).

Autoradiography of the filters revealed seven positive clones among the 200,000 colonies screened.

Initial sequence analysis of one of the putative monkey cDNA clones (designated clone 83) was performed for verification purposes by a modification of the procedure of Wallace, et al., *Gene,* 16, pp. 21–26 (1981). Briefly, plasmid DNA from monkey cDNA clone 83 was linearized by digestion with EcoRI and denatured by heating in a boiling water bath. The nucleotide sequence was determined by the dideoxy method of Sanger, et al., *P.N.A.S.* (*U.S.A.*), 74, pp. 5463–5467 (1977). A subset of the EPV mixture of probes consisting of 16 sequences was used as a primer for the sequencing reactions.

C. Monkey EPO cDNA Sequencing

Nucleotide sequence analysis of clone 83 was carried out by the procedure of Messing, *Methods in Enzymology,* 101, pp. 20–78 (1983). Set out in Table IV is a pre-

liminary restriction map analysis of the approximately 1600 base pair EcoRI/HindIII cloned fragment of clone 83. Approximate locations of restriction endonuclease enzyme recognition sites are provided in terms of number of bases 3′ to the EcoRI site at the 5′ end of the fragment. Nucleotide sequencing was carried out by sequencing individual restriction fragments with the intent of matching overlapping fragments. For example, an overlap of sequence information provided by analysis of nucleotides in a restriction fragment designated C113 (Sau 3A at ~111/SmaI at ~324) and the reverse order sequencing of a fragment designated C73 (AluI at ~424/BstEII at ~203).

TABLE IV

Restriction Enzyme Recognition Site	Approximate Location(s)
EcoRI	1
Sau3A	111
SmaI	180
BstEII	203
SmaI	324
KpnI	371
RsaI	372
AluI	424
PstI	426
AluI	430
HpaI	466
AluI	546
PstI	601
PvuII	604
AluI	605
AluI	782
AluI	788
RsaI	792
PstI	807
AluI	841
AluI	927
NcoI	946
Sau3A	1014
AluI	1072
AluI	1115
AluI	1223
PstI	1301
RsaI	1343

TABLE IV-continued

Restriction Enzyme Recognition Site	Approximate Location(s)
AluI	1384
HindIII	1449
AluI	1450
HindIII	1585

Sequencing of approximately 1342 base pairs (within the region spanning the Sau3A site 3′ to the EcoRI site and the HindIII site) and analysis of all possible reading frames has allowed for the development of DNA and amino acid sequence information set out in FIG. 5. In the Figure, the putative initial amino acid residue of the amino terminal of mature EPO (as verified by correlation to the previously mentioned sequence analysis of twenty amino terminal residues) is designated by the numeral +1. The pressure of a methionine-specifying ATG codon (designated −27) "upstream" of the initial amino terminal alanine residue as the first residue designated for the amino acid sequence of the mature protein is indicative of the likelihood that EPO is initially expressed in the cytoplasm in a precursor form including a 27 amino acid "leader" region which is excised prior to entry of mature EPO into circulation. Potential glycosylation sites within the polypeptide are designated by asterisks. The estimated molecular weight of the translated region was determine to be 21,117 daltons and the M.W. of the 165 residues of the polypeptide constituting mature monkey EPO was determined to be 18,236 daltons.

The polypeptide sequence of FIG. 5 may readily be subjected to analysis for the presence of highly hydrophilic regions and/or secondary conformational characteristics indicative of potentially highly immunogenic regions by, e.g., the methods of Hopp, et al., *P.N.A.S. (U.S.A.)*, 78, pp. 3824–3828 (1981) and Kyte et al., *J.Mol.Biol.*, 157, pp. 105–132 (1982) and/or Chou, et al., *Biochem.*, 13, pp. 222–245 (1974) and *Advances in Enzymology*, 47, pp. 45–47 (1978). Computer-assisted analysis according to the Hopp, et al. method is available by means of a program designated PEP Reference Section 6.7 made available by Intelligenetics, Inc. 124 University Avenue, Palo Alto, Calif.

EXAMPLE 4

A. Human Genomic Library

A Ch4A phage-borne human fetal liver genomic library prepared according to the procedures of Lawn, et al., *Cell*, supra was obtained and maintained for use in a plaque hybridization assay.

B. Plaque Hybridization Procedures For Screening Human Genomic Library

Phage particles were lysed and the DNAs were fixed on filters (50,000 plaques per filter) according to the procedures of Woo, *Methods In Enzymology*, 68, pp. 389–395 (1979) except for the use of GeneScreen Plus filters (New England Nuclear Catalog No. NEF-972) and NZYAM plates (NaCl, 5 g; $MgCl_2$-$6H_2O$, 2 g; NZ-Amine A, 10 g; yeast extract, 5 g; casamino acids, 2 g; maltose; 2 g; and agar, 15 g per liter).

The air-dried filters were baked at 80° C. for 1 hour and then digested with Proteinase K as described in Example 3, Part B. Prehybridization was carried out