

5,621,080

5

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.)*, 80, pp. 1551-1535 (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 pancreatic 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 libraries of low-to-moderate complexity. Nevertheless, because of their short length and heterogeneity, mixed probes often lack the specificity required 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, but 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. This normal 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 trans-

6

port 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.Hematol.*, 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., *Exp.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); Dessypris, 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 also, Krane, *Henry Ford Hosp.Med.*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.Gynaecol.*, 90(4), 304-311 (1983) (pregnancy, 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.J.Appl.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 Daimiak, et al., *Cancer*, 51(6), 1101-1106 (1983) and Schwartz, et al., *Otolaryngol.*, 109, 269-272 (1983) (various neoplastic disease states accompanied by abnormal erythropoiesis).

Prior attempts to obtain erythropoietin in good yield from plasma or urine have proven relatively unsuccessful. Com-

5,621,080

7

plicated 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 urine. 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.*, 1117, 581-588 (1983); Tarbourin, 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 erythropoietin 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

8

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 molecular 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: NH₂-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 erythropoietin 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 pro-

5,621,080

9

cedure of Hewick, M., et al., *J. Biol. Chem.*, 256, 7990-7997 (1981). See, also, Sue, et al., *Proc. Nat. Acad. Sci. (USA)*, 80, pp. 3651-3655 (1983) relating to development of polyclonal antibodies against a synthetic 26-mar 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 122a(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. While such levels of in vitro translation of exogenous mRNA coding for erythropoietin were acknowledged to be quite 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

10

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*. Briefly 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. (USA)*, 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 characterized 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., *Saccaromyces 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 mammalian 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 which 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 100 U (preferably in excess of 500 U and most preferably in excess of 1,000 to 5,000 U) 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

5,621,080

11

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 biologically 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 polypeptides 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.g., 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 manufacture 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 analogs 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 polypeptidic products having at least a part of the primary structural conformation and one or more of the biological properties of erythropoietin which are comprehended by: (a) the DNA sequences set out in FIGS. 5 and 6 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 DNA 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 encod-

12

ing 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 FIG. 6 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 ¹²⁵I) 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 DNA level and used as gene markers for identifying neighboring 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 transiently 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 ONA/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

5,621,080

13

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 to FIGS. 1 through 21, wherein:

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

FIG. 2 shows vector pDSVL-MkE.

FIG. 3 shows vector pSVgHuEPO.

FIG. 4 shows vector pDSVL-gHuEPO.

FIG. 5A, 5B and 5C (collectively referred to as FIG. 5) show the sequence of monkey EPO cDNA and the encoded EPO.

FIGS. 6A, 6B, 6C, 6D and 6E (collectively referred to as FIG. 6) show the sequence of human genomic EPO DNA and the encoded EPO.

FIG. 7 shows the sequence of the ECEPO gene.

FIG. 8 shows the sequence of the SCEPO gene.

FIG. 9 shows a comparison of the human and monkey EPO polypeptides.

FIG. 10 shows the ECEPO section 1 oligonucleotides.

FIG. 11 shows section 1 of the ECEPO gene.

FIG. 12 shows the ECEPO section 2 oligonucleotides.

FIG. 13 shows section 2 of the ECEPO gene.

FIG. 14 shows the ECEPO section 3 oligonucleotides.

FIG. 15 shows section 3 of the ECEPO gene.

FIG. 16 shows the SCEPO section 1 oligonucleotides.

FIG. 17 shows section 1 of the SCEPO gene.

FIG. 18 shows the SCEPO section 2 oligonucleotides.

FIG. 19 shows section 2 of the SCEPO gene.

FIG. 20 shows the SCEPO section 3 oligonucleotides.

FIG. 21 shows the section 3 of the SCEPO gene.

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 isolat-

14

able quantities of polypeptides displaying biological (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 acid 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 homology to each other and to a partial sequence generated by amino acid analysis of human EPO fragments.

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 immunological verification of EPO expression in such systems.

More particularly, Example 1 is directed to amino acid sequencing of human EPO fragments 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 informa-

5,621,080

15

tion 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 Examples 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 ("CHO") cells and to the immunological and biological activities of products of these expression systems as well as characterization 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 |
| T21 | K-L-F-R |
| T25 | Y-L-L-E-A-K |
| T26a | L-I-C-D-S-R |
| T26b | L-V-T-G-E-A-C-R |
| T27 | T-I-T-A-D-T-F-R |
| T28 | E-A-I-S-P-P-D-A-A-M-A-A-P-L-R |
| T30 | E-A-E-X-I-T-T-G-X-A-E-H-X-S-L |

16

TABLE I-continued

| Fragment No. | Sequence Analysis Result |
|--------------|--------------------------|
| T31 | N-E-X-I-T-V-P |
| T33 | V-Y-S-N-F-L-R |
| T35 | S-L-T-T-L-L-R |
| T38 | V-N-F-Y-A-W-K |
| | G-Q-A-L-L-V-X-S-S-Q-P-W- |
| | E-P-I-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 phosphoramidite 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 | CCA | ACC | TT |
| | T | A | A | A | T | | |
| | G | | | | G | | |
| | C | | | | C | | |
| | | | | | | | 5' |

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 oligonucleotide 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 |
| | C | T | | C | T | A |
| | | G | | | G | |
| | | C | | | C | |
| | | | | | | 5' |

Oligonucleotide probes were labeled at the 5' end with gamma-³²P-ATP, 7500-8000 Ci/mole (ICN) using T₄ polynucleotide kinase (NEN).

EXAMPLE 2

A. Monkey Treatment Procedures

Female Cynomolgus monkeys *Macaca fascicularis* (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 ketamine hydrochloride.

5,621,080

17

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 trasyol and 0–250 μl of 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, 2 mM 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 (GATCTAAAGACCGTC-CCCCCCC and ACGGTCTTTA) in a three-fold molar excess over the oligo dG tailed vector.

18

B. Colony Hybridization Procedures For Screening Monkey cDNA Library

Transformed *E. coli* were spread out at a density of 9000 colonies per $10\times 10\text{ cm}$ plate on nutrient plates containing 50 micrograms/ml Ampicillin. GeneScreen filters (New England Nuclear Catalog No. NEF-972) were pre-wet on a BIII-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) 0.5M NaOH for ten minutes; and
- (3) 1.0M Tris-HCl (pH 7.5) for three minutes.

The filters were then air dried in a vacuum oven 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 \times SSPE–0.5% SDS–100 micrograms/ml SS *E. coli* DNA–5 \times 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 \times 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 \times SSC–0.1% SDS at room temperature and washed two to three times with 6 \times 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 deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md., under deposit accession No. A.T.C.C. 67545 on Oct. 20, 1987) 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 procedures of Messing, *Methods in Enzymology*, 101,

5,621,080

19

pp. 20-78 (1983). Set out in Table IV is a preliminary restriction map analysis of the approximately 1600 base pair EcoRI/HindIII cloned fragment of clone 83. Approximate locations of restriction endonuclease enzymic 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 (Sau3A 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 | 465 |
| AluI | 546 |
| PstI | 601 |
| PvuII | 604 |
| AluI | 605 |
| AluI | 782 |
| AluI | 788 |
| RsaI | 792 |
| PstI | 807 |
| AluI | 841 |
| AluI | 927 |
| NotI | 946 |
| Sau3A | 1014 |
| AluI | 1072 |
| AluI | 1115 |
| AluI | 1223 |
| PstI | 1301 |
| RsaI | 1343 |
| AluI | 1384 |
| HindIII | 1449 |
| AluI | 1450 |
| HindIII | 1535 |

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, comprising portions 5A, 5B, 5C. 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 presence 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 determined 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.

20

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 Lawli, 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; MgCl₂-6H₂O, 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 with a 1M NaCl-1% SOS buffer for 55° C. for 4 hours or more, after which the buffer was removed. Hybridization and post-hybridization washings were carried out as described in Example 3, Part B. Both the mixture of 128 20-mer probes designated EPV and the mixture of 128 17-mer probes of Table III (designated the EPQ mixture) were employed. Hybridization was carried out at 48° C. using the EPV probe mixture. EPQ probe mixture hybridization was carried out at 46° C.—4 degrees below the lowest calculated Td for members of the mixture. Removal of the hybridized probe for rehybridization was accomplished by boiling with 1xSSC-0.1% SDS for two minutes. Autoradiography of the filters revealed three positive clones (reactive with both probe mixtures) among the 1,500,000 phage plaques screened. Verification of the positive clones as being EPO-encoding was obtained through DNA sequencing and electron micrographic visualization of heteroduplex formation with the monkey cDNA of Example 3. This procedure also gave evidence of multiple introns in the genomic DNA sequence.

EXAMPLE 5

Nucleotide sequence analysis of one of the positive clones (designated lhE1) deposited with the American Type Culture Collection, 12301 Parklawn drive, Rockville, Md., under deposit accession No. A.T.C.C. 40381 on Oct. 20, 1987 was carried out and results obtained to date are set out in FIG. 6, comprising portions 6A, 6B, 6C, 6D and 6E

In FIG. 6, the initial continuous DNA sequence designates a top strand of 620 bases in what is apparently an untranslated sequence immediately preceding a translated portion of

5,621,080

21

the human EPO gene. More specifically, the sequence appears to comprise the 5' end of the gene which leads up to a translated DNA region coding for the first four amino acids (-27 through -24) of a leader sequence ("presequence"). Four base pairs in the sequence prior to that encoding the beginning of the leader have not yet been unambiguously determined and are therefore designated by an "X". There then follows an intron of about 639 base pairs (439 base pairs of which have been sequenced and the remaining 200 base pairs of which are designated "LS.") and immediately preceding a codon for glutamic acid which has been designated as residue -23 of the translated polypeptide. The exon sequence immediately following is seen to code for amino acid residues through an alanine residue (designated as the +1 residue of the amino acid sequence of mature human EPO) to the codon specifying threonine at position +26, whereupon there follows a second intron consisting of 256 bases as specifically designated. Following this intron is an exon sequence for amino acid residues 27 through 55 and thereafter a third intron comprising 612 base pairs commences. The subsequent exon codes for residues 56 through 115 of human EPO and there then commences a fourth intron of 134 bases as specified. Following the fourth intron is an exon coding for residue Nos. 116 through 166 and a "stop" codon (TGA). Finally, FIG. 6 identifies a sequence of 568 base pairs in what appears to be an untranslated 3' region of the human EPO gene, two base pairs of which ("X") have not yet been unambiguously sequenced.

FIG. 6 thus serves to identify the primary structural conformation (amino acid sequence) of mature human EPO as including 166 specified amino acid residues (estimated M.W.=18,399). Also revealed in the Figure is the DNA sequence coding for a 27 residue leader sequence along with 5' and 3' DNA sequences which may be significant to promoter/operator functions of the human gene operon. Sites for potential glycosylation of the mature human EPO polypeptide are designated in the Figure by asterisks. It is worthy of note that the specific amino acid sequence of FIG. 6 likely constitutes that of a naturally occurring allelic form of human erythropoietin. Support for this position is found in the results of continued efforts at sequencing of urinary isolates of human erythropoietin which provided the finding that a significant number of erythropoietin molecules therein have a methionine at residue 126 as opposed to a serine as shown in the Figure.

FIG. 9, below, illustrates the extent of polypeptide sequence homology between human and monkey EPO. In the upper continuous line of the Figure, single letter designations are employed to represent the deduced translated polypeptide sequences of human EPO commencing with residue -27 and the lower continuous line shows the deduced polypeptide sequence of monkey EPO commencing at assigned residue number -27. Asterisks are employed to highlight the sequence homologies. It should be noted that the deduced human and monkey EPO sequences reveal an "additional" lysine (K) residue at (human) position 116. Cross-reference to FIG. 6 indicates that this residue is at the margin of a putative mRNA splice junction in the genomic sequence. Presence of the lysine residue in the human polypeptide sequence was further verified by sequencing of a cDNA human sequence clone prepared from mRNA isolated from COS-1 cells transformed with the human genomic DNA in Example 7, infra.

EXAMPLE 6

The expression system selected for initial attempts at microbial synthesis of isolatable quantities of EPO polypep-

22

tide material coded for by the monkey cDNA provided by the procedures of Example 3 was one involving mammalian host cells (i.e., COS-1 cells, A.T.C.C. No. CRL-1650). The cells were transfected with a "shuttle" vector capable of autonomous replication in *E.coli* host (by virtue of the presence of pBR322-derived DNA) and the mammalian hosts (by virtue of the presence of SV40 virus-derived DNA).

More specifically, an expression vector was constructed according to the following procedures. The plasmid clone 83 provided in Example 3 was amplified in *E.coli* and the approximately 1.4 kb monkey EPO-encoding DNA was isolated by EcoRI and HindIII digestion. Separately isolated was an approximately 4.0 kb, HindIII/SalI fragment from pBR322. An approximately 30 bp, EcoRI/SalI "linker" fragment was obtained from M13mp10 RF DNA (P and L Laboratories). This linker included, in series, an EcoRI sticky end, followed by SstI, SmaI, BamHI and XbaI recognition sites and a SalI sticky end. The above three fragments were ligated to provide an approximately 5.4 kb intermediate plasmid ("pERS") wherein the EPO DNA was flanked on one side by a "bank" of useful restriction endonuclease recognition sites. pERS was then digested with HindIII and SalI to yield the EPO DNA and the EcoRI to SalI (M13mp10) linker. The 1.4 kb fragment was ligated with an approximately 4.0 kb BamHI/SalI of pBR322 and another M13mp10 HindIII/BamHI RF fragment linker also having approximately 30 bp. The M13 linker fragment was characterized by a HindIII sticky end, followed by PstI, SalI, XbaI recognition sites and a BamHI sticky end. The ligation product was, again, a useful intermediate plasmid ("pBR-EPO") including the EPO DNA flanked on both sides by banks of restriction sites.

The vector chosen for expression of the EPO DNA in COS-1 cells ("pDSVLI") had previously been constructed to allow for selection and autonomous replication in *E.coli*. These characteristics are provided by the origin of replication and Ampicillin resistance gene DNA sequences present in the region spanning nucleotides 2448 through 4362 of pBR322. This sequence was structurally modified by the addition of a linker providing a HindIII recognition immediately adjacent nucleotide 2448 prior to incorporation into the vector. Among the selected vector's other useful properties was the capacity to autonomously replicate in COS-1 cells and the presence of a viral promoter sequence functional in mammalian cells. These characteristics are provided by the origin of replication DNA sequence and "late gene" viral promoter DNA sequence present in the 342 bp sequence spanning nucleotide numbers 5171 through 270 of the SV40 genome. A unique restriction site (BamHI) was provided in the vector and immediately adjacent the viral promoter sequence through use of a commercially available linker sequence (Collaborative Research). Also incorporated in the vector was a 237 base pair sequence (derived as nucleotide numbers 2553 through 2770 of SV40) containing the "late gene" viral mRNA polyadenylation signal (commonly referred to as a transcription terminator). This fragment was positioned in the vector in the proper orientation vis-a-vis the "late gene" viral promoter via the unique BamHI site. Also present in the vector was another mammalian gene at a location not material to potential transcription of a gene inserted at the unique BamHI site, between the viral promoter and terminator sequences. [The mammalian gene comprised an approximately 2,500 bp mouse dihydrofolate reductase (DHFR) mini-gene isolated from plasmid pMG-1 as in Gasser, et al., *P.N.A.S. (U.S.A.)*, 79, pp. 6522-6526, (1982).] Again, the major operative component

5,621,080

23

of plasmid pDSVL1 comprise nucleotides 2448 through 4362 of pBR322 along with nucleotides 5171 through 270 (342bp) and 2553 through 2770 (237bp) of SV40 DNA.

Following procedures described, e.g., in Maniatis, et al., supra, the EPO-encoding DNA was isolated from plasmid pBR-EPO as a BamHI fragment and ligated into plasmid pDSVL1 cut with BamHI. Restriction enzyme analysis was employed to confirm insertion of the EPO gene in the correct orientation in two of the resulting cloned vectors (duplicate vectors H and L). See FIG. 2, illustrating plasmid pDSVL-MKE. Vectors with EPO genes in the wrong orientation (vectors F, X and G) were saved for use as negative controls in transfection experiments designed to determine EPO expression levels in hosts transformed with vectors having EPO DNA in the correct orientation.

Vectors H, L, F, X and G were combined with carrier DNA (mouse liver and spleen DNA) were employed to transfect duplicate 60 mm plates by calcium phosphate microprecipitate methods. Duplicate 60 mm plates were also transfected with carrier DNA as a "mock" transformation negative control. After five days all culture media were tested for the presence of polypeptides possessing the immunological properties of naturally-occurring EPO.

EXAMPLE 7

A. Initial EPO Expression System Involving COS-1 Cells

The system selected for initial attempts at microbial synthesis of isolatable quantities of human EPO polypeptide material coded for by the human genomic DNA EPO clone, also involved expression in mammalian host cells (i.e., COS-1 cells, A.T.C.C. No. CRL-1650). The human EPO gene was first sub-cloned into a "shuttle" vector which is capable of autonomous replication in both *E.coli* hosts (by virtue of the presence of pBR322 derived DNA) and in the mammalian cell line COS-1 (by virtue of the presence of SV40 virus derived DNA). The shuttle vector, containing the EPO gene, was then transfected into COS-1 cells. EPO polypeptide material was produced in the transfected cells and secreted into the cell culture media.

More specifically, an expression vector was constructed according to the following procedures. DNA isolated from lambda clone λ hE1, containing the human genomic EPO gene, was digested with BamHI and HindIII restriction endonucleases, and a 5.6 Kb DNA fragment known to contain the entire EPO gene was isolated. This fragment was mixed and ligated with the bacterial plasmid pUC8 (Bethesda Research Laboratories, Inc.) which had been similarly digested, creating the intermediate plasmid "pUC8-HuE", providing a convenient source of this restriction fragment.

The vector chosen for expression of the EPO DNA in COS-1 cells (pSV4SEt) had previously been constructed. Plasmid pSV4SEt contained DNA sequences allowing selection and autonomous replication in *E.coli*. These characteristics are provided by the origin of replication and Ampicillin resistance gene DNA sequences present in the region spanning nucleotides 2448 through 4362 of the bacterial plasmid pBR322. This sequence was structurally modified by the addition of a linker providing a HindIII recognition site immediately adjacent to nucleotide 2448. Plasmid pSV4SEt was also capable of autonomous replication in COS-1 cells. This characteristic was provided by a 342 bp fragment containing the SV40 virus origin of repli-

24

cation (nucleotide numbers 5171 through 270). This fragment had been modified by the addition of a linker providing an EcoRI recognition site adjacent to nucleotide 270 and a linker providing a SalI recognition site adjacent nucleotide 5171. A 1061 bp fragment of SV40 was also present in this vector (nucleotide numbers 1711 through 2772 plus a linker providing a SalI recognition site next to nucleotide number 2772). Within this fragment was a unique BamHI recognition sequence. In summary, plasmid pSV4SEt contained unique BamHI and HindIII recognition sites, allowing insertion of the human EPO gene, sequences allowing replication and selection in *E.coli*, and sequences allowing replication in COS-1 cells.

In order to insert the EPO gene into pSV4SEt, plasmid pUCB-HuE was digested with BamHI and HindIII restriction endonucleases and the 5.6 kb EPO encoding DNA fragment isolated. pSV4SEt was also digested with BamHI and HindIII and the major 2513 bp fragment isolated (preserving all necessary functions). These fragments were mixed and ligated, creating the final vector "pSVgHuEPO". (See, FIG. 3.) This vector was propagated in *E.coli* and vector DNA isolated. Restriction enzyme analysis was employed to confirm insertion of the EPO gene.

Plasmid pSVgHuEPO DNA was used to express human EPO polypeptide material in COS-1 cells. More specifically, pSVgHuEPO DNA was combined with carrier DNA and transfected into triplicate 60 mm plates of COS-1 cells. As a control, carrier DNA alone was also transfected into COS-1 cells. Cell culture media were sampled five and seven days later and tested for the presence of polypeptides possessing the immunological properties of naturally occurring human EPO.

B. Second EPO Expression System Involving COS-1 Cells

Still another system was designed to provide improved production of human EPO polypeptide material coded by the human genomic DNA EPO clone in COS-1 cells (A.T.C.C. No. CRL-1650).

In the immediately preceding system, EPO was expressed in COS-1 cells using its own promoter which is within the 5.6 Kb BamHI to HindIII restriction fragment. In the following construction, the EPO gene is altered so that it is expressed using the SV40 late promoter.

More specifically, the cloned 5.6 Kb BamHI to HindIII genomic human EPO restriction fragment was modified by the following procedures. Plasmid pUC8-HuE, as described above, was cleaved with BamHI and with BstEII restriction endonucleases. BstEII cleaves within the 5.6 Kb EPO gene at a position which is 44 base pairs 5' to the initiating ATG coding for the pre-peptide and approximately 680 base pairs 3' to the HindIII restriction site. The approximately 4900 base pair fragment was isolated. A synthetic linker DNA fragment, containing SalI and BstEII sticky ends and an internal BamHI recognition site was synthesized and purified. The two fragments were mixed and ligated with plasmid pBR322 which had been cut with SalI and BamHI to produce the intermediate plasmid pBRgHE. The genomic human EPO gene can be isolated therefrom as a 4900 base pair BamHI digestion fragment carrying the complete structural gene with a single ATG 44 base pairs 3' to BamHI site adjacent the amino terminal coding region.

This fragment was isolated and inserted as a BamHI fragment into BamHI cleaved expression vector plasmid pDSVL1 (described in Example 6). The resulting plasmid,