EXHIBIT H-1
Part 1 of 2

1

SOLE INVENTOR

MAIL
NOV
30
1984
PAT. & TRADE MARK OFF

APPLICATION FOR
UNITED STATES LETTERS PATENT

12/03/84 675298	3 201	150.00 CK
12/03/84 675298	3 202	165.00 CK
12/03/84 675298	3 203	250.00 CK
12/03/84 675298	3 204	50.00 CK
12/03/84 675298		
	3 203	50.00 CK

SPECIFICATION

TO ALL WHOM IT MAY CONCERN:

Be it known that I, _____ FU-KUEN LIN _____

a citizen of the United States, residing at __ 438 Thunderhead Street, Thou __

in the County of __ Ventura __ and State of __ California __

have invented a new and useful __ "PRODUCTION OF ERYTHROPOIETIN" __

of which the following is a specification.

4

- 1 -

"PRODUCTION OF ERYTHROPOIETIN"

This is a continuation-in-part of my co-pending
U.S. Patent Application Serial Nos. 561,024, filed
(now abandoned)
5 December 13, 1983, 582,185, filed February 21, 1984, and
(now abandoned)
655,841, filed September 28, 1984.

BACKGROUND

10   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 biolo-
15 gical properties of naturally-occurring erythropoietin.

A. Manipulation Of Genetic Materials

Genetic materials may be broadly defined as
those chemical substances which program for and guide the
20 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
25 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
30 of nucleotides in linear polymeric form is by means of
fusion of the 5' phosphate of one nucleotide 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),

5

- 2 -

which associations occur by means of hydrogen bonding
between purine and pyrimidine bases [i.e.,
"complementary" associations existing either between ade-
nine (A) and thymine (T) or guanine (G) and cytosine
5 (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
10 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 spe-
15 cific 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
20 the operations of small RNA 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
25 orientation of any given one of the twenty amino acids
for polypeptide "expression", is in the form of triplet
"codons" -- sequential groupings of three nucleotide
bases. In one sense, the formation of a protein is the
ultimate form of "expression" of the programmed genetic
30 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 sequen-
ces, also usually "upstream" of (i.e., preceding) a gene
35 in a given DNA polymer, bind proteins that determine the
frequency (or rate) of transcriptional initiation.

                                                6

- 3 -

Collectively referred to as "promoter/regulator" or
"control" DNA sequence, these sequences which precede a
selected gene (or series of genes) in a functional DNA
polymer cooperate to determine whether the transcription
5 (and eventual expression) of a gene will occur. DNA
sequences which "follow" a gene in a DNA polymer and pro-
vide a signal for termination of the transcription into
mRNA are referred to as transcription "terminator"
sequences.

10   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 gene-
tically coded information concerning the desired product
15 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 synthe-
20 sized and then stably introduced into another organism
which is preferably a self-replicating unicellular orga-
nism such as bacteria, yeast or mammalian cells in
culture. Once this is done, the existing machinery for
gene expression in the "transformed" or "transfected"
25 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 publi-
30 cations relating to "recombinant DNA" methodologies for
the isolation, synthesis, purification and amplification
of genetic materials for use in the transformation of
selected host organisms. U.S. Letters Patent No.
No. 4,237,224 to Cohen, et al., for example, relates to
35 transformation of unicellular host organisms with
"hybrid" viral or circular plasmid DNA which includes

7

4

- 4 -

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.

5   Selected foreign ("exogenous" or "heterologous") DNA
strands usually including sequences coding for desired
product are prepared in linear form through use of simi-
lar enzymes. The linear viral or plasmid DNA is incu-
bated with the foreign DNA in the presence of ligating

10   enzymes capable of effecting a restoration process and
"hybrid" vectors are formed which include the selected
exogenous DNA segment "spliced" into the viral or cir-
cular DNA plasmid.

Transformation of compatible unicellular host

15   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 frequently, the goal of

20   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. Letters Patent Nos. 4,264,731 (to

25   Shine), 4,273,875 (to Manis), 4,293,652 (to Cohen), and
European Patent Application No. 093,619, published November
9, 1983.

The development of specific DNA sequences for
splicing into DNA vectors is accomplished by a variety of

30   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)

35   the "isolation" of double-stranded DNA sequence from the
genomic DNA of the donor; (2) the chemical manufacture of

8

5

- 5 -

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 transcrip-
tion" of mRNA isolated from donor cells. The
5   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
10   residues of the desired polypeptide product is known.
DNA manufacturing procedures of co-owned, co-pending
U.S. Patent Application Serial No. 483,451, by Alton, et
al., (filed April 15, 1983 and corresponding to PCT
US83/00605, published November 24, 1983 as WO83/04053),
15   for example, provide a superior means for accomplishing
such highly desirable results as: providing for the pre-
sence 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"
20   codons); avoiding the presence of untranslated "intron"
sequences (commonly present in mammalian genomic DNA
sequences and mRNA transcripts thereof) which are not
readily processed by procaryotic host cells; avoiding
expression of undesired "leader" polypeptide sequences
25   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 vec-
tors in association with desired promoter/regulator and
30   terminator sequences; and providing for ready construc-
tion of genes coding for polypeptide fragments and ana-
logs of the desired polypeptides.

When the entire sequence of amino acid residues
of the desired polypeptide is not known, direct manufac-
35   ture of DNA sequences is not possible and isolation of
DNA sequences coding for the polypeptide by a cDNA method

9

- 6 -

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

5   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

10   express relatively large quantities of growth hormone
products). Where substantial portions of the polypep-
tide's amino acid sequence are known, labelled, single-
stranded DNA probe sequences duplicating a sequence
putatively present in the "target" cDNA may be employed

15   in DNA/DNA hybridization procedures carried out on cloned
copies of the cDNA which have been denatured to single
stranded form. [See, generally, the disclosure and
discussions of the art provided in U.S. Patent No.
4,394,443 to Weissman, et al. and the recent demonstra-

20   tions 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.

25   Patent 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

30   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

35   Transfer Membrane materials providing instruction manuals
for the transfer and hybridization of DNA and RNA,
Catalog No. NEF-972.]   10

- 7 -

Among the more signficant recent advances in
hybridization procedures for the screening of recombinant
clones is the use of labelled mixed synthetic oligo-
nucleotide probes, each of which is potentially the
5 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
10 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
15 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
20 (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),
25 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
30 a single 11-mer to effect a two-site "positive" confir-
mation of the presence of cDNA 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
35 common of the three above-noted methods for developing

11

- 8 -

specific DNA sequences for use in recombinant procedures. This is especially true in the area of recombinant proce- dures directed to securing microbial expression of mam- malian polypeptides and is due, principally to the

5   complexity of mammalian genomic DNA. Thus, while reliable procedures exist for developing phage-borne libraries of genomic DNA of human and other mammalian species origins [See, e.g., Lawn, et al. Cell, 15, pp. 1157-1174 (1978) relating to procedures for

10   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 endo- nuclease fragmentation procedure; and Blattner, et al.,

15   Science, 196, pp. 161-169 (1977) describing construction of a bovine genomic library] there have been relatively few successful attempts at use of hybridization proce- dures in isolating genomic DNA in the absence of exten- sive foreknowledge of amino acid or DNA sequences. As

20   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 gly- coprotein hormones from the Maniatis Library through use of a "full length" probe including a complete 621 base

25   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. 1531-1535 (1983) report isola- tion of human genomic clones for human HLA-DR using a 175 base pair synthetic oligonucleotide. Finally, Anderson,

30   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

35   determination of poor prospects for isolating mRNA suitable for synthesis of a cDNA library due to apparent

12

- 9 -

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⊖

5   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, repre-
senting every possible codon combination for a small

10 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,

15 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

20 unavailable." (Citations omitted).

There thus continues to exist a need in the art
for improved methods for effecting the rapid and effi-
cient isolation of cDNA clones in instances where little
is known of the amino acid sequence of the polypeptide

25 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 con-

30 cerning 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

35 cells, occurs continuously throughout the human life span
to offset cell destruction. Erythropoiesis is a very

13

/0

- 10 -

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

5 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

10 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 car-
bohydrate (sialic acid) removed. Erythropoietin is pre-
sent in very low concentrations in plasma when the body

15 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.

20 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

25 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

30 of primitive precursor cells in the bone marrow into pro-
erythroblasts which subsequently mature, synthesize
hemoglobin and are released into the circulation as red
blood cells. When the number of red blood cells in cir-
culation is greater than needed for normal tissue oxygen

35 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.,

14

- 11 -

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),
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.,
10 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 pro-
cess of red blood cell formation, the hormone has poten-
tial useful application in both the diagnosis and the
15 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
20 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 correc-
25 tive of anemia of the type associated with chronic renal
failure. See also, Krane, Henry Ford Hosp.Med.J., 31(3),
177-181 (1983).

It has recently been estimated that the availa-
bility of erythropoietin in quantity would allow for
30 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
35 provided a basis for projection of efficacy of erythro-

15

- 12 -

poietin 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)
(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 Dainiak, et al., Cancer, 51(6), 1101-1106 (1983) and
Schwartz, et al., Otolaryngol., 109, 269-272 (1983)
(various neoplastic disease states accompanied by abnor-
mal erythropoiesis).

20   Prior attempts to obtain erythropoietin in good
yield from plasma or urine have proven relatively unsuc-
cessful. Complicated and sophisticated laboratory tech-
niques are necessary and generally result in the
collection of very small amounts of impure and unstable
25 extracts containing erythropoietin.

U.S. Letters Patent 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.

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

16

- 13 -

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 (August
10, 1977), pp. 5558-5564. This seven-step procedure
5   includes ion exchange chromatography, ethanol precipita-
tion, 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. Letters Patent No. 4,397,840 to Takezawa,
10   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.

15   U.K. Patent Application No. 2,085,887 by
Sugimoto, et al., published May 6, 1982, describes a pro-
cess 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
20   (distributed into the cultures after mammalian host propaga-
tion containing up to 10^7 cells per ml. At the highest pro-
duction 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
25   vitro culture following transfer of cells from in vivo
propagation systems. (See also the equivalent U.S.
Letters Patent No. 4,377,513.) Numerous proposals have
been made for isolation of erythropoietin from tissue
sources, including neoplastic cells, but the yields have
30   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),
35   341-347 (1984).

Other isolation techniques utilized to obtain
purified erythropoietin involve immunological procedures.
17

/ 4

- 14 -

A polyclonal, serum-derived antibody directed against
erythropoietin is developed by injecting an animal, pre-
ferably a rat or rabbit, with human erythropoietin. The
injected human erythropoietin is recognized as a foreign

5   antigenic substance by the immune system of the animal
and elicits production of antibodies against the antigen.
Differing cells responding to stimulation by the antige-
nic substance produce and release into circulation anti-
bodies slightly different from those produced by other

10 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,

15 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.

20   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 immuno-
logically reactive with a single antigenic determinant of

25 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 tech-
niques to develop "monoclonal" antibodies to erythro-
poietin and to employ these antibodies in the isolation

30 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,

35 520 (1982). As another example, a detailed description

18

- 15 -

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
5 (1983); Yanagawa, et al., Blood, 64(2), 357-364 (1984);
Yanagawa, et al., J.Biol.Chem., 259(5), 2707-2710 (1984);
and U.S. Letters Patent No. 4,465,624.
          Also of interest to the background of the inven-
tion are reports of the immunological activity of synthe-
10 tic 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 simi-
15 lar in duration and extent to the immune reactions of
physiologically significant proteins such as viral anti-
gens, polypeptide hormones, and the like. Included among
the immune reactions of such polypeptides is the provoca-
tion of the formation of specific antibodies in
20 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,
25 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
30 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 pep-
tides which approximately share secondary structures of
peptide hormones but may not share their primary struc-
35 tural conformation. The above studies relate, of course,
to amino acid sequences of proteins other than erythro-

19

- 16 -

poietin, a substance for which no substantial amino acid sequence information has been published. In co-owned, co-pending U.S. Patent Application Serial No. 463,724, filed February 4, 1983, by J. Egrie, published August 22,

5   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

10   the following sequence of amino acids: PS

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

The polypeptide sequence is one assigned to the first twenty amino acid residues of mature human erythropoietin

15   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).

20   See, also, Sue, et al., Proc. Nat. Acad. Sci. (USA), 80, pp. 3651-3655 (1983) relating to development of polyclo-nal antibodies against a synthetic 26-mer based on a dif-fering amino acid sequence, and Sytowski, et al., J.Immunol. Methods, 69, pp.181-186 (1984).

25   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 purifi-cation of erythropoietin, it appears unlikely that these

30   materials can readily provide for the large scale isola-tion of quantities of erythropoietin from mammalian sour-ces sufficient for further analysis, clinical testing and potential wide-ranging therapeutic use of the substance in treatment of, e.g., chronic kidney disease wherein

35   diseased tissues fail to sustain production of erythro-poietin. It is consequently projected in the art that

20

/7

- 17 -

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
5   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
10 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 conven-
tional techniques. Further, so little is known of the
15 continuous sequence of amino acid residues of erythro-
poietin 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
20 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
25 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.
30 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
35 al. Exp.Hematol., 11. Supp. 14, Abstract 101 (1983)
report the extraction of mRNA from kidney tissues of.

21

- 18 -

phenylhydrazine-treated baboons and the injection of the
mRNA into Xenopus laevis oocytes with the rather tran-
sitory result of in vitro production of a mixture of
"translation products" which included among them

5   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

10   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

15   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 iso-

20   lated. [See also, Farber, Clin.Res., 31(4), 769A
(1983).]

Since the filing of U.S. Patent Application
Serial Nos. 561,024 and 582,185, there has appeared a
single report of the cloning and expression of what is

25   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 pro-
ducts were noted to be immunoreactive with a monoclonal
antibody to an unspecified "epitope" of human erythro-

30   poietin. See, Lee-Huang, Proc. Nat. Acad. Sci. (USA),
81, pp. 2708-2712 (1984).

                         BRIEF SUMMARY

35   The present invention provides, for the first
time, novel purified and isolated polypeptide products

                                                        22

- 19 -

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., immunolo-
gical properties and in vivo and in vitro biological
5   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 bac-
terial, yeast and mammalian cells in culture) of exoge-
10   nous 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 asso-
15   ciated 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
20   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
25   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
30   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
35   available which can be propagated in vitro continuously
and which upon growth in culture are capable of producing

23

- 20 -

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.

5   Also provided by the present invention are
synthetic polypeptides wholly or partially duplicative of
continuous sequences of erythropoietin amino acid resi-
dues which are herein for the first time elucidated.
These sequences, by virtue of sharing primary, secondary

10 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 immunolo-

15 gical substitutes for erythropoietin in therapeutic and
immunological processes. Correspondingly provided are
monoclonal and polyclonal antibodies generated by stan-
dard means which are immunoreactive with such polypep-
tides and, preferably, also immunoreactive with

20 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 confor-

25 mation 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

30 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

35 transfected microbial hosts under conditions facilitative
of large scale expression of the exogenous, vector-borne

24

2 1

- 21 -

DNA sequences and isolation of the desired polypeptides
from the growth medium, cellular lysates or cellular
membrane fractions.

Isolation and purification of microbially
5 expressed polypeptides provided by the invention may be
by conventional means including, e.g., preparative chro-
matographic separations and immunological separations
involving monoclonal and/or polyclonal antibody prepara-
tions.

10   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

15 "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. Corres-

20 pondingly, 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

25 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 spe-
cified are replaced by other residues and/or addition

30 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
35 sequences useful in securing expression in procaryotic or
eucaryotic host cells of polypeptide products having at

25

22

- 22 -

least a part of the primary structural conformation and
one or more of the biological properties of erythro-
poietin which are comprehended by: (a) the DNA sequences
set out in Tables V and VI herein or their complementary

5 strands; (b) DNA sequences which hybridize (under hybri-
dization 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

10 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)

15 are manufactured DNA sequences encoding EPO, EPO
fragments and EPO analogs which DNA sequences may incor-
porate codons facilitating translation of messenger RNA
in non-vertebrate hosts.

        Comprehended by the present invention is that

20 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 pro-
teins" as described by Tramontano, et al., Nucleic Acids
Research, 12, pp. 5049-5059 (1984).

25   Also comprehended by the invention are phar-
maceutical 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

30 the treatment of anemic disease states and most espe-
cially such anemic states as attend chronic renal
failure.

        Polypeptide products of the invention may be
"labelled" by covalent association with a detectable

35 marker substance (e.g., radiolabelled with ^{125}I) to pro-
vide reagents useful in detection and quantification of

26

- 23 -

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
5 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
10 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 poly-
15 nucleotide of unknown sequence in a heterogeneous cellu-
lar or viral sample including multiple single-stranded
polynucleotides where

(a) a mixture of labelled single-stranded poly-
nucleotide probes is prepared having uniformly varying
20 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,

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

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

(e) the specific polynucleotide is detected by
35 monitoring for the presence of a hybridization reaction
between it and a totally complementary probe within said

24

- 24 -

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
5   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
10 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 iden-
tification of cDNA clones coding for erythropoietin of
monkey species origins within a library prepared from
15 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"
20 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
25 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.

30   The above-noted illustrative procedures consti-
tute 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 mix-
35 ture of more than 32 oligonucleotide probes in the isola-
tion of cDNA clones.

28

- 25 -

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
5 in its presently preferred embodiments.

DETAILED DESCRIPTION

According to the present invention, DNA
10 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
15 expression providing isolatable quantities of polypep-
tides displaying biological (e.g., immunological) proper-
ties 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
20 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 con-
25 taining 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
30 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
35 through DNA/DNA plaque hybridization employing the above-
noted pool of 128 mixed 20-mer oligonucleotide probes and

29

- 26 -

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
5   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.
10 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
15 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 con-
tain as much as 3000 mU of EPO per ml of culture fluid.
20   The following examples are presented by way of
illustration of the invention and are specifically
directed to procedures carried out prior to iden-
tification of EPO encoding monkey cDNA clones and human
genomic clones, to procedures resulting in such iden-
25 tification, 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 con-
30 struction 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 radioim-
35 munoassay (RIA) analysis of animal sera. Example 3 is
directed to the preparation of the cDNA library, colony

30

27

- 27 -

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
5   4 is directed to procedures involved in the iden-
tification of positive human genomic clones and thus pro-
vides information concerning the source of the genomic
library, plaque hybridization procedures and verification
of positive clones. Example 5 is directed to DNA
10   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
15   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
20   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.
25   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 mam-
malian host expression systems for monkey species EPO
30   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
35   genes encoding human species EPO and EPO analogs, which
genes include a number of preference codons for

31

- 28 -

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.

5

EXAMPLE 1

A. Human EPO Fragment Amino Acid Sequencing

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

Fragments were arbitrarily assigned numbers and
were analyzed for amino acid sequence by microsequence
15 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 deter-
mined.

20

25

30

32

T 0300X

- 29 -

<u>TABLE I</u>

<u>Fragment No.</u>   <u>Sequence Analysis Result</u>

5	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
10	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-Y-T-G-E-A-C-R
15	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- N-E-X-I-T-V-P
	T31	V-Y-S-N-F-L-R
20	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

25

30

)

33

35   3

- 30 -

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
5 code for the purpose of ascertaining whether mixed probe
procedures could be applied to DNA/DNA hybridization pro-
cedures on cDNA and/or genomic DNA libraries. This ana-
lysis revealed that within Fragment No. T35 there existed
a series of 7 amino acid residues
10 (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
15 al., Tetrahedron Letters, 22, pp. 1859-1862 (1981) on a
solid support according to the sequence set out in Table
II, below.

T0310Y

20   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			
25		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
30 17-mer probes as set out in Table III, below.

T0311X   TABLE III

Residue	-	Gln	Pro	Trp	Glu	Pro	Leu	
35	3'	GTT	GGA	ACC	CTT	GGA	GA	- 5'
		C	T		C	T	A	
			G			G		
			C			C		

34

- 31 -

Oligonucleotide probes were labelled at the 5'
end with gamma - ^{32}P-ATP, 7500-8000 Ci/mmole (ICN) using
T$_4$ polynucleotide kinase (NEN).

5   EXAMPLE 2

A. Monkey Treatment Procedures ~~and RIA Analysis~~

Female Cynomolgus monkeys Macaca fascicularias
(2.5-3 kg, 1.5-2 years old) were treated subcutaneously
10 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
15 administration of 25 mg/kg doses of ketamine hydroch-
loride. Harvested materials were immediately frozen in
20 liquid nitrogen and stored at -70·C.

B. RIA for EPO
20   Radioimmunoassay procedures applied for quan-
titative detection of EPO in samples were conducted
according to the following procedures:
An erythropoietin standard or unknown sample was
20 incubated together with antiserum for two hours at 37·C.
25 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,
30   10,000 cpm of ^{125}I-erythropoietin, 5 µl trasylol 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

35

- 32 -

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
5 general, this corresponded to a final antiserum dilution
of from 1:50,000 to 1:100,000.

The antibody-bound ^{125}I-erythropoietin was pre-
cipitated by the addition of 150 µl Staph A. After a 40
min. incubation, the samples were centrifuged and the
10 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
15 from all final values to correct for nonspecific precipi-
tation. The erythropoietin content of the unknown
samples was determined by comparison to the standard
curve.

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

25   CL   EXAMPLE 3

A.   Monkey cDNA Library Construction

Messenger RNA was isolated from normal and ane-
mic monkey kidneys by the guanidinium thiocyanate proce-
30 dure 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
35 Springs, Harbor, N.Y., 1982). The cDNA library was con-
structed according to a modification of the general pro-

36

33

- 33 -

cedures of Okayama, et al., <u>Mol.</u> and <u>Cell.Biol.</u>, <u>2</u>,
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 <u>PstI</u> and then tailed with
5 oligo dT of 60-80 bases in length; (2) <u>HincII</u> 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)
<u>BamHI</u> digestion was employed to remove the oligo dG tail
10 from one end of the vector; and (5) replacement of the
RNA strand by DNA was in the presence of two linkers
(GATCTAAAGACCGTCCCCCCCCC and ACGGTCTTTA) in a three-fold
molar excess over the oligo dG tailed vector.

15 B. Colony Hybridization Procedures For
      Screening Monkey cDNA Library

          Transformed <u>E.coli</u> were spread out at a density
of 9000 colonies per 10 x 10 cm plate on nutrient plates
containing 50 micrograms/ml Ampicillin. GeneScreen
20 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
25 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:
30   (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
35 at 80·C for two hours.
          The filters were then subjected to Proteinase K

37

34

- 34 -

digestion through treatment with a solution containing 50
microgramps/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 x SSPE - 0.5% SDS - 100
micrograms/ml SS E.coli DNA - 5 x BFP). The prehybridi-
zation 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 x 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 disso-
ciation temperatures (Td) determined for any of the pro-
bes.

Following hybridization, the filters were washed
three times for ten minutes on a shaker with 6 x SSC
-0.1% SDS at room temperature and washed two to three
times with 6 x SSC - 1% SDS at the hybridization tem-
perature (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 proce-
dure 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

38

35

- 35 -

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

5   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
10 Enzymology, 101, pp. 20-78 (1983). Set out in Table IV
is a preliminary restriction map analysis of the approxi-
mately 1600 base pair EcoRI/HindIII cloned fragment of
clone 83. Approximate locations of restriction endo-
nuclease enzyme recognition sites are provided in terms
15 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
20 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).

25

30

35

39

36

- 36 -

T0370X

TABLE IV

	Restriction Enzyme Recognition Site	Approximate Location(s)
5	EcoRI	1
	Sau3A	111
	SmaI	180
	BstEII	203
	SmaI	324
10	KpnI	371
	RsaI	372
	AluI	424
	PstI	426
	AluI	430
15	HpaI	466
	AluI	546
	PstI	601
	PvuII	604
	AluI	605
20	AluI	782
	AluI	788
	RsaI	792
	PstI	807
	AluI	841
25	AluI	927
	NcoI	946
	Sau3A	1014
	AluI	1072
	AluI	1115
30	AluI	1223
	PstI	1301
	RsaI	1343
	AluI	1384
	HindIII	1449
35	AluI	1450
	HindIII	1585

40

AM-ITC 00952328

- 37 -

Sequencing of approximately 1342 base pairs
(within the region spanning the <u>Sau</u>3A site 3' to the
<u>Eco</u>RI site and the <u>Hind</u>III site) and analysis of all
possible reading frames has allowed for the development

5   of DNA and amino acid sequence information set out in
Table V.   In the table, the putative initial amino acid
residue of the amino terminal of mature EPO (as verified
by correlation to the previously mentioned sequence ana-
lysis of twenty amino terminal residues) is designated by

10   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 pro-
tein is indicative of the likelihood that EPO is ini-

15   tially 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

20   of the translated region was determine to be 21,117
daltons and the M.W. of the 165 residues of the polypep-
tide constituting mature monkey EPO was determined to be
18,236 daltons.

25

30

41

35

- 38 -

TABLE V

Translation of Monkey EPO cDNA

```
Sau3A
GATCCCGCGCCCCTGGACAGCCGCCCTCTCCTCCAGGCCCGTGGGCCTGGCCCTGCCC
CGCTGAACTTCCCGGGATGAGGACTCCCGGTGGTCACCGCGCGCCTAGGTCGCTGAG

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

Leu Trp 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
                             -10                              10

    -1  +1
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
                                                     *

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
                     20

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
        30                            *                  40
```

- 39 -

TABLE V (continued)

```
           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 CTG 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 GCA GCG TCG GCC 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
```

43

- 40 -

TABLE V (continued)

```
               150                                      160
          Leu Arg Gly Lys Leu Lys Tyr Thr Gly Glu Ala Cys Arg Arg
          CTC CGG GGA AAG CTG AAG TAC ACG GGG GAG GCC TGC AGG AGA

                165
          Gly Asp OP
          GGG GAC AGA TGA CCAGGTGCGTCCAGCTGGCACATCCACCACCTCCCTCACCAACA

CTGCCTGTGCCACACCCTCCCTCACCACTCCCGACCACCTCCCATCCGAGGGGCTCTCAGCTAAG

CGCCAGCCTGTCCCATGGACACTCCAGTGCCAGCAATGACATCTCAGGGGCCAGAGGAAC

TGTCCAGAGCACAACTCTGAGATCTTGGGATGTCGCAGGGCCAACTTGAGGGCCCAGAGC

AGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGAGCAGAGACAATGCAGGGAAAACACCT

GAGCTCACTCGGCCACCTGCAAAATTTGATGCAGGACACGCTTTGGAGGCAATTTACCTG

TTTTTGCACCTACCATCAGGGACAGGATGACTGGAGAACTAGGTGGCAAGCTGTGACTT

CTCAAGGCCTCACGGGCACTCCCTTGGTGGCAAGAGCCCCTTGACACTGAGAGAATATT

TTGCAATCTGCAGCAGGAAAAATTACCGGACACAGGTTTTGGAGGTTTGACAGTACTTGACAG

GTGTGTGGGGAAGCAGCAGGGCGGGTAGGGGTGGAGCTGGGATGCGAGTGAGAACCGTGAGAGAC

AGGATGGGGGCTGGCCCTCTGGTTCTCGTGGGGTCCAAGCTT
                                          HindIII
```

44

- 41 -

The polypeptide sequence of Figure 5 ~~Table V~~ may readily
be subjected to analysis for the presence of highly
hydrophilic regions and/or secondary conformational
characteristics indicative of potentially highly immuno-
5 genic 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 ana-
10 lysis 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, California.

EXAMPLE 4

15

A. Human Genomic Library

A Ch4A phage-borne human fetal liver genomic
library prepared according to the procedures of Lawn, et
20 al., Cell, 18, pp. 533-543 (1979) was obtained and main-
tained for use in a plaque hybridization assay.

B. Plaque Hybridization Procedures For
    Screening Human Genomic Library

25   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-976) and
30 NZYAM plates (NaCl, 5g; MgCl$_2$-6H$_2$O, 2 g; NZ-Amine A, 10g;
yeast extract, 5g; casamino acids, 2 g; maltose; 2g; and
agar, 15g per liter).
         The air-dried filters were baked at 80°C for 1
hour and then digested with Proteinase K as described in
35 Example 3, Part B. Prehybridization was carried out with
a 1M NaCl - 1% SDS buffer for 55°C for 4 hours or more,

45

- 42 -

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 pro-
5 bes 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
10 hybridized probe for rehybridization was accomplished by
boiling with 1 x SSC - 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
15 positive clones as being EPO-encoding was obtained
through DNA sequencing and electron micrographic visuali-
zation of heteroduplex formation with the monkey cDNA of
Example 3. This procedure also gave evidence of multiple
introns in the genomic DNA sequence.
20

EXAMPLE 5

Nucleotide sequence analysis of one of the posi-
tive clones (designated λhE1) was carried out and results
25 obtained to date are set out in Table VI.

30

35

40

46

- 43 -

TABLE VI

AAGCTTCTGGGCTTCCAGACCCAGCTACTTTGCGGAACTCAGCAACCCAGGCATCTCTGAGTCTCCGCCCA

AGACCGGGATGCCCCCCAGGGGAGGTGTCCGGGGAGCCCAGCCTTTCCCAGATAGCACGCTCCGCCAGTCCC

AAGGGTGCGCAACCGGCTGCACTCCCCTCCCGCGGACCCAGGGCCCGGGAGCAGCCCCATGACCCACACGC

ACGTCTGCAGCAGCCCCGCTCACGCCCCGGCGAGCCTCAACCCAGGCGTCCTGCCCCTGCTCTGACCCCGG

GTGGCCCCTACCCCTGGGCGACCCCTGAGGCACACAGCCTCTCCCCACCCCGGCGACGCACACATG

CAGATAACAGCCCCGACCCCGCCAGAGCCGXAGAGTCCCTGGGCCACCCCGGCCTCGCCTGCCGCTG

CGCCGCACCGCGCTGTCCCCGGAGCCGGGGCCCACCGCGCCCXGCTCTGCTCCGACACCGCGCCC

CTTGGACAGCCGCCCTCTCCTCTAGGCCCGTGGGGCTGGCCCTGCACCGCGGAGCTTCCCGGGATGAGGXX

CCCGGTGACCGGCGCCCCAAGTCGCTGAGGGACCCCGGCCAAGCGCGGGAG

GTGAGTACTCGCGGGGCTGGGGGCTCCCGGGCGGGGTTCCTGTTTGAGCGGGGATTTAGGCGCCCCGGGCT

```
-27              -24
Met Gly Val His
ATG GGG GTG CAC G
```

47

- 44 -

TABLE VI (cont'd.)

```
ATTGGGCCAAGAGGTGGCTGGGTTCAAGGACCGGCGACTTGTCAAGGACACCCCGGAGGGGGAGGGGGTGGG

GCAGCCTCCACGTGCCGCGGGGACTTGGGGGACTTCTTGGGGATGGCAAAAACCTGGCCTGTTGAGGGGCA

CAGTTTGGGGTTGGGGAGGAGGTTTGGGGTTCTGCTGCTGCAGTTGTGTCGTTGTCGTGTCTCG[i.s.]

TTGCACACGCACAGATCAATAAGCCAGAGGCAGCACCTGAGTGCTTGCATGGTTGGGACAGGAAGGACGAG

CTGGGGCAGAGACGTGGGGATGAAGGAAGCTGTCCTTCCACAGCCACCCTTCTCCCCCCCGCCTGACTCT

                          -20
                    Glu Cys Pro Ala Trp Leu Trp 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 CGC CTC ATC TGT

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

     26
     Thr
ACG GTGAGACCCCTTCCCCAGCACATTCCACAGAACTCACGCTCAGGGCTTCAGGGAACTCCTCCCAGAT

CCAGGAACCTGCCACTTGGTTTGGGGTGGAGTTGGGAAGCTAGACACTGCCCCCTACATAAGAATAAGTC
```

48

AM670167669   AM-ITC 00952336

- 45 -

TABLE VI (cont'd.)

```
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 GTGAGTTCGTTTTTTTTTTTTTCCTTTCTTTTGGAGAATCTCATT

TGCCAGCCTGATTTGGATGAAAGGGAGAATGATCGGGGGAAAGGTAAAATGGAGCAGCAGAGATGAGGCT

GCCTGGGCGCAGAGGCTCACGTCTATAATCCCAGCTGAGATGGCCGAGATGGGAGAATTGCTTGAGCCCT

GGAGTTTCAGACCAACCTAGGCAGCATAGTGAGATCCCCCATCTCTACAAACATTTAAAAAAATTAGTCAG

GTGAAGTGGTGCATGGTGGTAGTCCCAGATATTTGGAAGGCTGAGGCGGGAGGATCGCTTGAGCCCAGGAA

TTTGAGGCTGCAGTGAGCTGTGATCACACCACTGCACTCCAGCCTCAGTGACAGAGTGAGGCCCTGTCTCA
```

49

- 46 -

TABLE VI (cont'd.)

```
AAAAAGAAAAGAGAAAAAGAGAAAAATAATGAGGGCTGTATGGAATACATTCATTATTCATTCACTCACTCACT

CACTCATTCATTCATTCATTCAACAAGTCTTATTGCATACCTTCTGTTGCTCAGCTTGGTGCTTGG

GGCTGCTGAGGGGCCAGGAGGAGGAGAGGGTGACATGGGTCAGCTCGACTCCCAGAGTCCACTCCCTGTAG

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

                                        *    80                    90
         Val Leu Arg Gly Gln Ala Leu Leu Val 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   GTGAGTAGGAGCGGACACTTCTGCTTGCCCTTTCTGTAAGAAGGGGA
         CGG GCT CTG GGA GCC CAG

GAAGGGGTCTTGCTAAGGAGTACAGGAACTGTCCGTATTCCTTCCCTTTCTGTGGCACTGCAGCGACCTCCT

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

GTTTCTCCTTGGCAG
```

50

AM670167671   AM-ITC 00952338

- 47 -

TABLE VI (cont'd.)

```
        130
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
                                              140
                                                            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
        150

166
Asp Arg OP
GAC AGA TGA CCAGGTGTCCACCTGGGCATATCCACCACCTCCCTCACCAACATTGCTTGTGCCACA

CCCTCCCCCGCCACTCCTGAACCCGTCGAGCCGCTCTCAGCTCAGCGCCAGCCTGTCCCATGGACACTCC

AGTGCCAGCAATGACATCTCAGGGCCCAGAGGAAC GTCCAGAGAGCAACTCTGAGATCTAAGGATGTCAC

AGGGCCAACTTGAAGGGCCCAGAGAGCAGGAAGCATTCAGAGAGCAGCTTAAAACTCAGGGACAGAGCCATGC

TGGGAAGACGCCTGAGCTCACTCGGCACCCTGCAAAAATTTGATGCCAGGACACGCTTTGGAGGCGATTTAC

CTGTTTTCGCCACCTACCATCAGGGACAGGATGACCTGGAGAACTTAGGTGGCAAGCTGTGACTTCTCCAGG

TCTCACGGGCATGGGCACTCCCTTGGTGGCAAGAGCCCCTTGACACCGGGGTGGTGGGAACCATGAAGAC

AXGATXGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTATTCTCAACCTATTGACAGACTGAA

ACACAATATGAC
```

AM670167672   AM-ITC 00952339

51

- 48 -

Figure 6

In ~~Table VI~~, 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 the human EPO gene. More speci-
fically, 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
10   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 "I.S.") and immediately preceding a
15   codon for glutamine 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)
20   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
25   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, Table VI
30   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 unam-
biguously sequenced.

Figure 6
~~Table VI~~ thus serves to identify the primary
35   structural conformation (amino acid sequence) of mature
human EPO as including 166 specified amino acid residues

52

41

- 49 -

(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 signifi-
cant to promoter/operator functions of the human gene

5   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 Figure 6 likely constitutes that of a
naturally occurring allelic form of human erythropoietin.

10 Support for this position is found in the results of con-
tinued efforts at sequencing of urinary isolates of human
erythropoietin which provided the finding that a signifi-
cant number of erythropoietin molecules therin have a
methionine at residue 126 as opposed to a serine as shown

15 in the Figure.
    Figure 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

20 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

25 that the deduced human and monkey EPO sequences reveal an
"additional" lysine (K) residue at (human) position 116.
Cross-reference to Figure 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

30 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
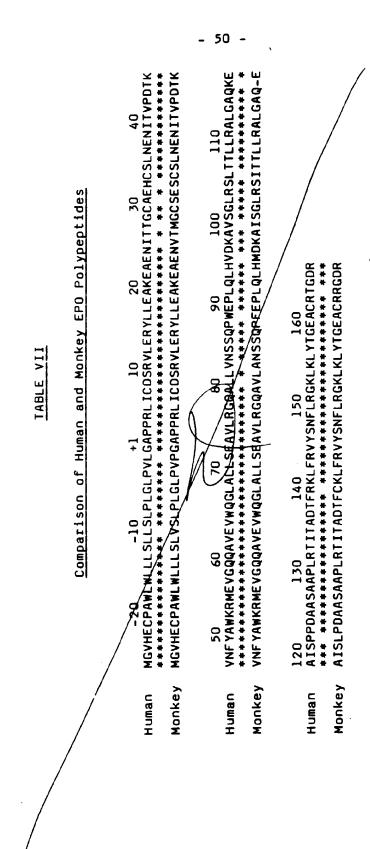genomic DNA in Example 7, infra.

35

42

- 50 -

TABLE VII

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  VNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQAVLANSSQPEEPLQLHMDKAISGLRSITTLLRALGAQ-E

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

54

AM670167675

AM-ITC 00952342

- 51 -

EXAMPLE 6

The expression system selected for initial
attempts at microbial synthesis of isolatable quantities
5 of EPO polypeptide 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
10 (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
15 plasmid clone 83 provided in Example 3 was amplified in
E.coli and the approximately 1.4kb 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
20 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
25 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.
30 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
35 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 site.

55

- 52 -

       The vector chosen for expression of the EPO DNA in COS-1 cells ("pDSVL1") had previously been constructed to allow for selection and autonomous replication in E.coli. These characteristics are provided by the origin

5 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

10 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

15 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

20 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

25 (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

30 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) minigene isolated from plasmid pMG-1 as in Gasser, et al.,

35 P.N.A.S. (U.S.A.), 79, pp. 6522-6526, (1982).] Again,

56

44

- 53 -

the major operative components 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.

5   Following procedures described, e.g., in
Maniatis, et al., _supra_, the EPO-encoding DNA was iso-
lated 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
10 EPO gene in the correct orientation in two of the
resulting cloned vectors (duplicate vectors H and L).
See Figure 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
15 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
20 transfect duplicate 60mm plates by calcium phosphate
microprecipitate methods. Duplicate 60 mm plates were
also transfected with carrier DNA as a "mock" transfor-
mation negative control. After five days all culture
media were tested for the presence of polypeptides
25 possessing the immunological properties of naturally-
occurring EPO.

EXAMPLE 7

30 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
35 EPO clone, also involved expression in mammalian host
cells (i.e., COS-1 cells, A.T.C.C. No. CRL-1650). The

57

45

- 54 -

human EPO gene was first sub-cloned into a "shuttle" vec-
tor 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

5   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.

10   More specifically, an expression vector was
constructed according to the following procedures. DNA
isolated from lambda clone λhE1, containing the human
genomic EPO gene, was digested with BamHI and HindIII
restriction endonucleases, and a 5.6 Kb DNA fragment

15 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

20 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 selec-
tion and autonomous replication in E.coli. These charac-

25 teristics are provided by the origin of replication and
Ampicillin resistance gene DNA sequences present in the
region spanning nucleotides 2448 through 4362 of the bac-
terial plasmid pBR322. This sequence was structurally
modified by the addition of a linker providing a HindIII

30 recognition site immediately adjacent to nucleotide 2448.
Plasmid pSV4SEt was also capable of autonomous replica-
tion in COS-1 cells. This characteristic was provided by
a 342 bp fragment containing the SV40 virus origin of
replication (nucleotide numbers 5171 through 270). This

35 fragment had been modified by the addition of a linker
providing an EcoR1 recognition site adjacent to

58

- 55 -

nucleotide 270 and a linker providing a Sall 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 Sall recogni-
5   tion site next to nucleotide number 2772). Within this
fragment was an 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 selec-
10 tion in E.coli, and sequences allowing replication in
COS-1 cells.

In order to insert the EPO gene into pSV4SEt,
plasmid pUC8-HuE was digested with BamHl and HindIII
restriction endonucleases and the 5.6 kb EPO encoding DNA
15 fragment isolated. pSV4SEt was also digested with BamHl
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, Figure 3.) This vector was propa-
20 gated 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 specifi-
25 cally, 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 polypep-
30 tides possessing the immunological properties of
naturally occurring human EPO.

B. Second EPO Expression System
    Involving COS-1 Cells

35 Still another system was designed to provide
improved production of human EPO polypeptide material

59

47

- 56 -

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
5   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
10   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
15   approximately 680 base pairs 3' to the HindIII restric-
tion 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
20   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
25   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
30   plasmid, pSVLgHuEPO, as illustrated in Figure 4, was used
to express EPO polypeptide material from COS-1 cells, as
described in Examples 6 and 7A.

EXAMPLE 8

35

Culture media from growth of the six transfected
COS-1 cultures of Example 6 were analyzed by radioim-

60

- 57 -

munoassay according to the procedures set forth in
Example 2, Part B. Each sample was assayed at 250, 125,
50, and 25 microliter aliquot levels. Supernatants from
growth of cells mock transfected or transfected with vec-
5 tors having incorrect EPO gene orientation were unam-
biguously negative for EPO immunoreactivity. For each
sample of the two supernatants derived from growth of
COS-1 cells transfected with vectors (H and L) having the
EPO DNA in the correct orientation, the % inhibition of
10 ^{125}I-EPO binding to antibody ranged from 72 to 88%, which
places all values at the top of the standard curve. The
exact concentration of EPO in the culture supernatant
could not then reliably be estimated. A quite conser-
vative estimate of 300 mU/ml was made, however, from the
15 value calculation of the largest aliquot size (250
microliter).

A representative culture fluid according to
Example 6 and five and seven day culture fluids obtained
according to Example 7A were tested in the RIA in order
20 to compare activity of recombinant monkey and human EPO
materials to a naturally-occurring human EPO standard and
the results are set out in graphic form in Figure 1.
Briefly, the results expectedly revealed that the recom-
binant monkey EPO significantly competed for anti-human
25 EPO antibody although it was not able to completely inhi-
bit binding under the test conditions. The maximum per-
cent inhibition values for recombinant human EPO,
however, closely approximated those of the human EPO
standard. The parallel nature of the dose response
30 curves suggests immunological identity of the sequences
(epitopes) in common. Prior estimates of monkey EPO in
culture fluids were re-evaluated at these higher dilution
levels and were found to range from 2.91 to 3.12 U/ml.
Estimated human EPO production levels were correspon-
35 dingly set at 392 mU/ml for the five-day growth sample

61

4 9