EXHIBIT H-15

#15
MM
3/19/87

PATENT

IN THE UNITED STATES PATENT
AND TRADEMARK OFFICE

Application of:	)	"Production of
	)	
FU-KUEN LIN	)	Erythropoietin"
	)	
Serial No: 675,298	)	Group Art Unit 127
	)	
Filed: November 30, 1984	)	Examiner - A. Tanenholtz

AUTHORIZATION TO CHARGE ISSUE FEE TO
DEPOSIT ACCOUNT PURSUANT TO 37 C.F.R. 1.311(b)

RECEIVED
MAR 12 1987
GROUP 120

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

In the event of issuance of a Notice of Allowance under 37 C.F.R. §1.311(a) in the above-identified application, the Commissioner is herewith authorized to charge all issue fees due under 37 C.F.R. §1.18 to Deposit Account No. 13-2855. A duplicate copy of this paper is enclosed.

Respectfully submitted,

MARSHALL, O'TOOLE, GERSTEIN,
MURRAY & BICKNELL

By _____
Michael F. Borun (Reg. No. 25,447)
A Member of the Firm
Attorneys for Applicants
Two First National Plaza
Chicago, Illinois 60603
(312) 346-5750

Chicago, Illinois
March 10, 1987

289. ~~225~~

PATENT

IN THE UNITED STATES PATENT
AND TRADEMARK OFFICE

Application of:	"Production of
FU-KUEN LIN	Erythropoietin"
Serial No: 675,298	Group Art Unit 127
Filed: November 30, 1984	Examiner - A. Tanenholtz

APPLICANT'S AMENDMENT AND REPLY
UNDER 37 C.F.R. §1.111 AND §1.115

RECEIVED
MAR 12 1987
GROUP 120

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

This is in response to the Office Action dated February 5, 1987 in the above-identified application wherein substantially all prior rejections of pending claims 14, 15, 17-36, 58 and 61-72 were withdrawn, but wherein new grounds for rejection were advanced under 35 U.S.C. §§102, 103 and 112, and wherein all claims were "provisionally" rejected under 35 U.S.C. §101. Reconsideration and allowance of all claims is respectfully requested in view of the following amendments and remarks.

AMENDMENT

IN THE CLAIMS

Please cancel claims 14, 15, 17-36, 58 and 61-72 without prejudice and enter the following new claims 73-103.

--73. A purified and isolated DNA sequence for use in securing expression in a procaryotic or eucaryotic

290 ~~269~~

host cell of a polypeptide product having at least a part of the primary structural conformation and having a therapeutic activity of naturally-occurring erythropoietin, said DNA sequence selected from the group consisting of:

(a) the DNA sequences set out in Figures 5 and 6 or their complementary strands;

(b) DNA sequences which hybridize under stringent conditions to the DNA sequences defined in (a); and

(c) DNA sequences which, but for the degeneracy of the genetic code, would hybridize under stringent conditions to the DNA sequences defined in (a) and (b).

74. A procaryotic or eucaryotic host cell transformed or transfected with a DNA sequence according to claim 73 in a manner allowing the host cell to express said polypeptide product.

75. A biologically functional circular plasmid or viral DNA vector including a DNA sequence according to claim 73.

76. A procaryotic or eucaryotic host cell stably transformed or transfected with a DNA vector according to claim 75.

77. A purified and isolated DNA sequence coding for procaryotic or eucaryotic host expression of a polypeptide having part or all of the primary structural conformation and having a therapeutic activity of erythropoietin.

78. A cDNA sequence according to claim 77.

79. A monkey species erythropoietin coding DNA sequence according to claim 78.

80. A DNA sequence according to claim 79 and including the protein coding region set forth in Figure 5.

81. A genomic DNA sequence according to claim 77.

82. A human species erythropoietin coding DNA sequence according to claim 81.

83. A DNA sequence according to claim 82 and including the protein coding region set forth in Figure 6.

84. A manufactured DNA sequence according to claim 77.

85. A manufactured DNA sequence according to claim 84 and including one or more codons preferred for expression in E.coli cells.

86. A manufactured DNA sequence according to claim 85, coding for expression of human species erythropoietin.

87. A manufactured DNA sequence according to claim 86 including the protein coding region set forth in Figure 7.

Cl 1 cont'd

88. A manufactured DNA sequence according to claim 84 and including one or more codons preferred for expression in yeast cells.

89. A manufactured DNA sequence according to claim 88, coding for expression of human species erythropoietin..

90. A manufactured DNA sequence according to claim 89 including the protein coding region set forth in Figure 8.

91. A DNA sequence according to claim 77 covalently associated with a detectable label substance.

92. A DNA sequence according to claim 91 wherein the detectable label is a radiolabel.

93. A single-strand DNA sequence according to claim 91.

94. A biologically functional circular plasmid or viral DNA vector including a DNA sequence according to claim 77.

95. A procaryotic or eucaryotic host cell stably transformed or transfected with a DNA vector according to claim 94.

96. A purified and isolated DNA sequence coding for a polypeptide fragment or polypeptide analog of

naturally-occurring erythropoietin having a therapeutic activity of erythropoietin.

97. A biologically functional circular plasmid or viral DNA vector including a DNA sequence according to claim 96.

98. A procaryotic or eucaryotic host cell stably transformed or transfected with a DNA vector according to claim 97.

99. A DNA sequence according to claim 96 which is a manufactured sequence.

100. A DNA sequence coding for [Phe^{15}]hEPO, [Phe^{49}]hEPO, [Phe^{145}]hEPO, [His^{7}]hEPO, [Asn^{2} des-Pro^{2} through Ile^{6}]hEPO, [des-Thr^{163} through Arg^{166}]hEPO, or [Δ27-55]hEPO.

101. A biologically functional circular plasmid or viral DNA vector including a DNA sequence according to claim 100.

102. A procaryotic or eucaryotic host cell stably transformed or transfected with a DNA vector according to claim 101.

103. A purified and isolated DNA sequence as set out in Figures 5 or 6 or the complementary strand of such a sequence.--

REMARKS

Applicant acknowledges with thanks the interview kindly granted to his counsel, Mr. Borun and Mr. Odre, on March 4, 1987.

Upon entry of the above-requested amendments to the claims, claims 73 through 103 will remain in the application and will be correlated to prior claims 14, 15, 17-36, 58 and 61-72 in the manner indicated in the following Table.

Prior Claim	New Claim	Prior Claim	New Claim
14	73	30	90
15	74	31	91
61	75	32	92
62	76	33	93
69	--	63	94
17	77	64	95
18	78	70	--
19	79	34	96
20	80	65	97
21	81	66	98
22	82	71	--
23	83	36	99
24	84	35	100
25	85	67	101
26	86	68	102
27	87	72	--
28	88	58	103
29	89		

Applicant notes that none of the claims whose entry is sought correspond to prior claims 69-72. Applicant specifically reserves his right to pursue claims of the same or similar scope in a duly filed continuing application.

A. The Claimed Subject Matter

As related in Applicant's communication dated October 3, 1986, the claims remaining in this application relate to DNA sequences, DNA vectors, transformed and transfected host cells useful in the preparation of erythropoietin products including, e.g., polypeptide analogs of erythropoietin.

Independent claim 73 is thus directed to purified and isolated DNA sequences generally defined by reference to the DNA sequences revealed in Figures 5 and 6 (previously Tables V and VI). Dependent claims 74-76 respectively relate to host cells transformed or transfected with DNA of claim 73, vectors including the DNA of claim 73, and hosts transformed with such vectors.

Independent claim 77 is directed generally to DNA sequences which code for procaryotic or eucaryotic host polypeptides having erythropoietin amino acid sequences and having one or more of erythropoietin's biological activities. Dependent claims 78-95 are directed to presently preferred forms of DNA sequences, vectors, and transformed or transfected hosts based on the claim 77 DNA sequences.

Independent claim 96 is generally directed to DNA sequences of the invention which encode polypeptide fragments and analogs of erythropoietin and dependent claims 97-99 are likewise directed to preferred forms of sequences, vectors, and transformed and transfected hosts. Independent claim 100 and dependent claims 101 and 102 relate to specific erythropoietin analog DNA sequences.

Finally, independent claim 103 is directed to the specific human and monkey erythropoietin-encoding purified and isolated DNA sequences as revealed in Figures 5 and 6.

B. The Outstanding Office Action, The Rejections of the Claims and Applicant's Responses Thereto

In a communication dated October 3, 1986 responding to the Action of July 3, 1986, Applicant submitted the full text of the Chirgwin et al. reference (Ref. C8) to complete the Information Disclosure Statement filed on April 24, 1986 and also submitted a Supplemental Information Disclosure Statement directing the Patent Office's attention to references B15, B16, C135 and C136. Applicant respectfully solicits the Examiner's acknowledgement of receipt and consideration of the same and notation of such consideration on the previously submitted Forms PTO-1449.

Applicant understands that the amendments and remarks set out in his communication dated October 3, 1986 have resulted in the reconsideration and withdrawal of the following rejections propounded under Sections 101, 102, 103 and 112 in the Action dated July 3, 1986:

1. The Section 112 (first paragraph) rejection of claims 14, 15, 17-36, 58 and 61-72 on grounds relating to permanence of A.T.C.C. Budapest Treaty deposits;

2. The Section 112 (second paragraph) rejection of claims 14, 15, 17-36, 58 and 61-72 based on,

(a) alleged indefiniteness of the term "procaryotic or eucaryotic",

(b) alleged indefiniteness based on failure to specify a "fragment" size,

(c) alleged indefiniteness of the term "biological properties",

(d) an instance of improper Markush group language, and,

(e) an instance of improper characterization of dependence for one of the claims;

3. The Section 101 rejection of claims 14, 24, 34 and 36 as being drawn to naturally-occurring subject matter;

4. The Section 102(b) and/or 103 rejection of claims 14, 24, 34 and 36 based on the Sugimoto et al. reference;

5. The Section 102(b) and/or 103 rejection of claims 14, 15, 17, 18, 20, 24-27, 33, 34, 61-66, 69, 70 and 71 based on the Lee-Huang (P.N.A.S.) reference;

6. The Section 102(a) and/or 103 rejection of claims 14, 15, 17-20, 24, 33, 34, 36, 58, 61-66, 69-70 and 71 based on Applicant's publication [J.Cell.Bioch., Suppl. 8B, p. 45 (1984)]; and,

7. The Section 103 rejection of all claims based variously on the Sugimoto et al., Cohen et al., Paddock, Farber et al., Bennetzen et al. Gouy et al., and Lewin references.

Due to the number and variety of new objections and rejections set forth in the recently received Action dated February 5, 1987, Applicant once again submits that the issues raised therein are best treated by means of responses which "track" the order of their appearance in the Action.

1. The Objection to the Disclosure Based On Figures 5-8 May Properly Be Withdrawn

The Examiner asserts that new drawing Figures 5-8 did not accompany Applicant's communication of October 3, 1986. Applicant attaches hereto as Exhibit No. 1 an envelope containing another set of Figures 5-8 and requests

that the outstanding objection to the specification be withdrawn.

2. The Rejection of Prior Claims 20, 23, 27 and 30 Under The Second Paragraph of Section 112 Is Inapplicable To Corresponding New Claims 80, 83, 87 and 90

Partially reiterating a prior position taken by the Patent Office, the Examiner has objected to Applicant's reference in certain claims to DNA sequence information contained in Figures of the drawing, stating: "Ordinarily claims should not refer to drawings particularly as here where the material referred to can be easily described without resort to drawings". Applicant respectfully disagrees with the Examiner's position and reiterates the remarks set out at pages 11 and 12 of his prior communication responding to the prior Patent Office notation that the DNA sequences of the Figures could "adequately be expressed in words". Applicant again relies on the authority of the decisions of In re Faust, 86 U.S.P.Q. 114, 115 (1943) and Ex parte Squires, 133 U.S.P.Q. 598, 600 (Bd. App. 1961) in support of his position.

Responding to the prior Patent Office suggestion, Applicant has wholly reconstituted Tables V, VI, XIV and XXI as drawing Figures 5-8. The subject DNA sequences thus appear twice in the application as it is presently constituted. As the Examiner will note, Figure 5 spans three full pages, Figure 6 covers five pages, and Figures 7 and 8 each comprehend a full page of information. Applicant respectfully disagrees that this sequence information "can be easily described without resort to drawings" as suggested by the Examiner. It simply cannot be reasonably suggested that there might be a violation of the standards of definiteness

of Section 112 (second paragraph) in Applicant's refraining from setting out the voluminous DNA sequences in the claims, when this same information already appear twice in the specification. Applicant therefore submits that the outstanding rejection of prior claims 20, 23, 27 and 30 may not properly be applied to corresponding new claims 80, 83, 87 and 90.

3. The "Provisional" Rejection of All Claims Under 35 U.S.C. §101 May Not Properly Be Maintained

Reiterating a prior Patent Office position, the Examiner has lodged a "provisional" rejection of all claims under Section 101 based on the presentation of claims of similar scope in Applicant's "parent" patent applications Serial Nos. 582,185 and 655,841. Applicant previously acknowledged with thanks this notation of potential nonconformity with Section 101 and did not contend (as suggested by the Examiner) that the statute would not bar allowance of the same claims in more than one application. Applicant further notes that Serial No. 582,185 has been expressly abandoned and that a provisional election of prosecution of non-overlapping claim 48 is being concurrently made in Serial No. 655,841, with a corresponding cancellation of claims 1-47 therein. Applicant continues to submit, however, that the provisional notation does not provide a present basis for rejection of the claims. It is thus submitted that the outstanding "provisional" rejection be withdrawn.

4. The Rejection of Prior Claims 14, 15, 17-19, 21, 23, 24, 25, 28, 31-34, 36, 58, 61-66 and 69-71 Under The First Paragraph of Section 112 May Not Properly Be Applied to Corresponding New Claims 73, 74, 77-79, 81, 83, 84, 85, 88, 91, 92, 93, 96 and 103

It was the Examiner's position that Applicant's use of the term "fragments thereof" in reference to claimed DNA sequence portions and/or his use of the term, "having at least a part of the primary structural conformation and one or more of the biological activities of naturally-occurring erythropoietin" in reference to polypeptides encoded by claimed DNA sequences was not "enabled" by the specification. The Examiner argued that either of these terms allows the claims "to read on proteins and peptides completely unrelated to erythropoietin" and that "those unrelated proteins could possess the common biological activity of being an antigen". Based on this argument, a Section 112 rejection was lodged against 14, 15, 17-19, 21, 22, 24, 25, 28, 31-34, 36, 58, 61-66 and 69-71. Applicant respectfully traverses the Examiner's rejection on such grounds.

Consistent with the substance of the discussions with the Examiner at the interview of March 4, 1987, wherein it was noted by the Examiner that the term "fragments" appeared to introduce a redundancy in claim 14, and notwithstanding Applicant's traversal, Applicant has herewith sought amendment to delete reference to "fragments" from prior independent claims 14 and 58. Corresponding new claims 73 and 103 no longer contain this term. Furthermore, Applicant has sought amendment of prior claims 14 and 17 referring to "biological activity" in a manner providing for recitation of "therapeutic activity" and similar recitation was added to prior claim 34. Corresponding new independent

claims 73, 77 and 96 thus refer to "therapeutic activity" rather than "biological activity".

These amendments have been sought for the purpose of advancing prosecution of the application and without waiver of Applicant's right to pursue claims of the form previously advanced in a duly filed continuing application.

Amendment of the claims is believed to moot the issues raised in the rejection and no rejection may properly be lodged against new claims 73, 74, 77-79, 81, 83, 84, 85, 88, 91, 92, 93, 96 and 103.

5. The Rejection of Prior Claim 35 Under 35 U.S.C. §112, Second Paragraph May Not Properly Be Applied to Corresponding Claim 100

The Examiner has rejected claim 35 with the notation that "use of brackets for indicating different species and strains of erythropoietin DNA sequence is improper since brackets in claims designates excluding the bracketed material". Applicant respectfully disagrees with the Examiner's position and proposes that brackets may properly be a part of a claim and are improper only when used for purposes of an amendment designating the deletion of a portion of a claims which already properly contains bracketed subject matter. In support of this position, applicant cites to 37 C.F.R. §1.121 which provides in pertinent part:

> "(d) Where underlining or brackets are intended to appear in the printed patent or are properly part of the claimed material and are not intended as symbolic of changes in the particular claim, amendment by rewriting in accordance with paragraph (b) of this section shall be prohibited."

It should be apparent from the above, that the brackets present in claim 100 are not "intended as symbolic of changes" and their use is thus in keeping with all "definiteness" requirements of 35 U.S.C. §112 and no proper basis for rejection exists.

6. The Rejection of Prior Claims 14, 15, 61, 62 and 69 Under 35 U.S.C. §112, First Paragraph, May Not Properly Be Applied To Corresponding Claims 73-76

It was the Examiner's position, based on the disclosures of Walker et al., Techniques In Mol. Biology, Macmillan Pub. Co., N.Y., p. 280 (1983) and Kennell et al., Progr.Nucl.Acid.Res.Mol.Biol., 11, 259-301 (1971)*, that claims 14, 15, 61 and 69 (which refer to DNA hybridization) may be rejected because "the disclosure is enabling only for claims limited [to] the conditions of hybridization". Applicant respectfully disagrees and submits that a reading of the claims in light of the specification reveals that the reference to hybridization is not at all unduly broad. As noted at specification page 22, lines 5-7 and again at specification page 94, lines 19-24, the context of the hybridization event referred to in the claims is specifically correlated to the hybridization conditions illustrated in the specification with respect to the initial isolation of monkey and human erythropoietin-encoding DNA, or more stringent conditions.

Notwithstanding this position, in keeping with the discussions with the Examiner at the interview of March 4, 1987, Applicant has amended claim 73 to refer to hybridiza-

---

* Applicant was provided a copy of pages 259 and 293 of this reference. If other portions are relied upon, advice of same is requested.

tion "under stringent conditions". Applicant therefore submits that claims 73-76 are not properly subject to rejection.

7. The Rejection of Prior Claims 14, 17, 18, 21-24, 26, 27, 31-36, 58 and 61-68 Under 35 U.S.C. §103 Based on Sue et al. Considered with Breslow et al. or Woods et al References May Not Properly be Applied to Corresponding Claims 73, 77, 78, 81-84, 86, 87, 91-93, 96, 99-100, 103, 75, 76, 94, 95, 97, 98, 101 and 102

It was the Examiner's position that the disclosures of the Sue et al. reference (P.N.A.S., 80) taken together with the publications by Breslow et al. [P.N.A.S. (USA), 79, pp. 6861-6865 (1982)] and Woods et al. [P.N.A.S. (USA), 79, pp. 5661-5665 (1982)] render the claimed subject matter obvious. The Examiner noted that the Sue et al. publication discloses what were "believed to be" the first 26 amino terminal residues of human erythropoietin and that the Breslow et al. and Woods et al. references disclose cDNA isolation using mixed probe sequences deduced from known amino acid data of blood protein. Acknowledging that the Sue et al. reference incorrectly designates the presence of an asparagine residue rather than a cysteine residue at position 7, the Examiner nonetheless concludes that:

> "It would be "obvious to isolate the human erythropoietin cDNA sequences by utilizing the Sue et al. erythropoietin amino acid sequence data to devise oligonucleotide probes for use in sequencing a cDNA liver library in the manner taught by Breslow et al. or Woods et al. The fact that the erythropoietin 26 amino acid amino terminal peptide sequence of Sue et al. differs from that of erythropoietin by designating Asn instead of Cys at the seven position is patentably irrelevant since it would not interfere with the preparation of oligonucleotide probes."

Applicant respectfully disagrees with the Examiner's position, submits that the Examiner's conclusions concerning preparation of probes based on the Sue et al. reference are in error, and submits, in turn, that the combination of references falls far short of existing legal standards for support of a conclusion of obviousness.

Briefly stated, Applicant did not do what the Examiner suggests could have been done based on the cited references. More significantly, Applicant would not have been able to do what the Examiner suggests could have been done based on the cited references, i.e., prepare a small number of oligonucleotides and probe for erythropoietin-encoding DNA within a relatively small DNA library.

As the Examiner will recall, Applicant succeeded in his discovery of DNA encoding erythropoietin using screening procedures which are themselves submitted to involve patentable advances in the art of DNA hybridization (as set forth in original claim 60 of the application). More specifically, Applicant employed two distinct sets of mixed probes to find the human genomic sequence. A first set consisted of a mixture of 128 20-mers (see specification Table II). The amino acid sequence which formed the basis for construction of the first set of probes is now known to correspond to residues 46-52 of human erythropoietin. Applicant used both the set of 128 20-mers of Table II and a second set of 128 17-mers (See specification Table III, relating to the sequence now known to correspond to erythropoietin residues 86-91) to jointly probe 1,500,000 phage plaques of human genomic library for the human sequence. Three positive clones were isolated. The set of 128 20-mers was thereafter used to successfully screen a 200,000 colony

Monkey kidney cDNA library, with only seven positive clones being isolated from the 200,000 screened. Applicant's use of mixed probes for screening a DNA library (and especially a mammalian genomic library) where the message sought was present in low abundancy had been projected as being "impractical" shortly before applicant's successful work. See, Anderson et al., Reference C2, and specification page 8, line 29 through page 9, line 20 and page 96, lines 2-13. As noted at specification page 9, the Anderson et al. reference states in pertinent part:

> "More generally, mixed-sequence oligodeoxynucleotide probes have been used to isolate protein genes of unknown sequence from cDNA libraries. Such probes are typtically mixtures of 8-32 oligonucleotides, 14-17 nucleotides in length, representing every possible codon combination for a small stretch (5-6 residues) of amino acid sequence. Under stringent hybridization conditions that discriminate against incorrectly base-paired probes, these mixtures are capable of locating specific gene sequences in clone libraries of low-to-moderate complexity. Nevertheless, because of their short length and heterogeneity, mixed probes often lack the specificity required for probing sequences as complex as a mammalian genome. This makes such a method impractical for the isolation of mammalian protein genes when the corresponding mRNAs are unavailable." (Citations omitted.)

Turning now to the "secondary" references, in both the Woods et al. and Breslow et al. procedures, an opportunity to develop multiple probes suitable for use in screening a cDNA (human liver) library arose as the result of prior knowledge of the certain identity of regions of amino acid residues which were specified by relatively "unambiguous" codons. Thus, in the Breslow reference, the mixture of oligonucleotide probes was synthesized which corresponded to a specific Apo-I protein sequence (Gln-Lys-

Lys-Trp-Gln) known to be present in the polypeptide whose encoding DNA was sought. Only a total of 16 different oligonucleotides was needed in order to develop a complete set of 14-mer probes for use in the cDNA library probing procedures. See Figure 1 on page 6862 of the reference. This low number was due to the relative lack of degeneracy among codons for tryptophane (no degeneracy), lysine (2-fold degeneracy), and glutamine (2-fold degeneracy) residues which made up the known sequence. Screening a cDNA library of only 10,000 colonies (provided by the authors of the Woods, et al. reference) Breslow et al. were able to isolate twenty positive clones. It is noteworthy that Apo-I DNA containing clones were thus conspicuously in relatively large abundance in the library. Nonetheless, the reported screening procedure failed to allow isolation of any clone including the full sequence of the Apo-I gene.

In the "two probe", Woods et al. reference, a total of 32 17-mers were needed to develop one complete set of probes corresponding to amino acids 9-14 of the protein sought, and an additional total of 48 17-mers were needed to ensure complete consonance of at least one probe to the DNA sequence encoding residues 78-83 of the desired protein. See Figure 2 on page 5662 of the reference. The ability of Woods, et al. to isolate 32 positive clones from a total of only 50,000 clones screened when hybridizing with one set of probes, and then to isolate 19 of the 32 using the second set, testifies to the relatively high abundance of the message in the library screened. (Once again, no full sequence clone is stated as having been isolated.)

Turning now to the correct sequence of amino acids within the first 26 residues of human erythropoietin, it is

revealed that no opportunity exists (similar to that presented to Woods et al. and Breslow et al.) to effectively employ such relatively small numbers of probes to isolate genomic encoding human erythropoietin. The correct first 26 residues of human erythropoietin corresponding to the following indicated levels of codon degeneracy: Ala(4), Pro(4), Pro(4), Arg(6), Leu(6), Ile(3), Cys(2), Asp(2), Ser(6), Arg(6), Val(4), Leu(6), Glu(2), Arg(6), Tyr(2), Leu(6), Leu(6), Glu(2), Ala(4), Lys(2), Glu(2), Ala(4), Glu(2), Asn(2), Ile(3) and Thr(4). [A preliminary estimate of the number of oligonucleotides needed for a complete set of mixed probes can be made by simply multiplying the degree(s) of degeneracy within the sequence. As an example, to guarantee that one 17-mer oligonucleotide probe which is an exact replica of the true erythropoietin sequence will be present in a set of mixed probes corresponding to the bases encoding the first five amino acids plus the first two bases of the sixth amino acid, one would need to make 4 X 4 X 4 X 6 X 6, or 2304, probes].

An array of the potential oligonucleotides based on the correct erythropoietin sequence is set forth below in Table I below.

TABLE I

1	2	3	4	5	6	7	8	9	10
Ala	Pro	Pro	Arg	Leu	Ile	Cys	Asp	Ser	Arg
T	G	G	T	T	T		T	T	G
GCC	CCA	CCA	CGC	CTC	ATC	TGT	GAC	AGC	CGA
G	T	T	G	G	A	C			T
A	C	C	A	A					C
								G	
			AGG	TTG				TCA	AGG
			A	A				T	A
								C	

11	12	13	14	15	16	17	18	19	20
Val	Leu	Glu	Arg	Tyr	Leu	Leu	Glu	Ala	Lys
T	C				T			T	
GTC	CTG	GAG	AGG	TAT	CTC	TTG	GAG	GCC	AAG
G	A	A	A	C	G	A	A	G	A
A	T				A			A	
			G			G			
	TTG		CGA		TTG	CTA			
	A		T		A	T			
			C			C			

21	22	23	24	25	26
Glu	Ala	Glu	Asn	Ile	Thr
	T			T	C
GAG	GCC	GAG	AAT	ATC	ACG
A	G	A	C	A	A
	A				T

An array of potential oligonucleotides based on the incorrect sequence of Sue et al. is set out in Table II below. Note well that the sequence includes two errors. Not only is the seventh residue incorrect, the twenty-fourth residue is incorrectly specified as lysine rather than asparagine.

TABLE II

1	2	3	4	5	6	7	8	9	10
Ala	Pro	Pro	Arg	Leu	Ile	Asn	Asp	Ser	Arg
T	G	G	T	T	T		T	T	G
GCC	CCA	CCA	CGC	CTC	ATC	AAT	GAC	AGC	CGA
G	T	T	G	G	A	C			T
G	T	T	G	G					C
								G	
			AGG	TTG				TCA	AGG
			A	A				T	A
								C	

11	12	13	14	15	16	17	18	19	20
Val	Leu	Glu	Arg	Tyr	Leu	Leu	Glu	Ala	Lys
T	C				T			T	
GTC	CTG	GAG	AGG	TAT	CTC	TTG	GAG	GCC	AAG
G	A	A	A	C	G	A	A	G	A
A	T				A			A	
			G			G			
	TTG		CGA		TTG	CTA			
	A		T		A	T			
			C			C			

21	22	23	24	25	26
Glu	Ala	Glu	Lys	Ile	Thr
	T			T	C
GAG	GCC	GAG	AAA	ATC	ACG
A	G	A	G	A	A
	A				T

The practical consequence of the two errors in the Sue et al. sequence is that no probe mixture comprehending either of the two incorrectly projected residues (Asn^{7} and Lys^{24}) could have been employed to detect the subject erythropoietin gene.

To illustrate the problem of incorrect sequence information by Sue et al., one notes that if the Applicant had used the amino acid information in the reference and even gone beyond the notation of Woods et al. by generating a 64-member mixture of 17-mer probes based on the sequence spanning the entire codon for lysine at position 20 through

the first two, non-degenerate bases of the codon for isoleucine at position 25, none of the 64 probes would have been a duplicate of the erythropoietin gene sequence due to the presence of the error in identifying the residue at position 24. Applicant, of course, would have had no way of knowing in advance that such work would be fruitless.

During the interview of March 4, 1987, the Examiner advanced a suggestion that notwithstanding the two errors in the Sue et al. human erythropoietin amino acid sequence, and notwithstanding numerous instances of 4- and 6-fold degeneracy of potential codons within the "correct" portions of the Sue et al. sequences, and notwithstanding the above-quoted notations of the Anderson et al. reference, the "correct" portions of the Sue et al. sequence appeared to provided opportunities to construct relatively small sets of mixed oligonucleotides for use in probing a human genomic library for the human erythropoietin gene. Reference was made to the amino acid residues immediately preceding the incorrectly noted 24th residue and the Examiner suggested that 12-mer or 14-mer mixtures might have been usefully applied in the manner of the 14-mer probes of Breslow et al. to find the human erythropoietin gene. Applicant's counsel acknowledged that mixtures of probes of this length could be made and that such mixtures would include small numbers of oligonucleotides on the order of those used in the Breslow et al. and Woods et al. systems, but traversed the Examiner's suggestion that any substantial likelihood of success would be expected to attend using probes of this size to isolate an erythropoietin gene from a human genomic library. This traversal was based on the relatively straightforward probabilities associated with a hybridiza-

on "hunt" for a DNA sequence that is a specific match to a given probe sequence. The Examiner is invited to consider the following:

1. It is clear that one could construct a set of 32 14-mer probes based on the potential codons for the "correct" 18th through 21st erythropoietin residues (Glu-Ala-Lys-Glu) plus the first two non-degenerate bases of residue 22 (Ala). One could also construct a set of 64 14-mers based on the potential codons for the "correct" 19th through 22nd residues (Ala-Lys-Glu-Ala) plus the first two non-degenerate bases of residue 23 (Glu).

2. The simple random probability of finding any specific string of <u>five</u> amino acid-specifying codons (i.e., <u>15</u> bases in a row) is one in 21^5, or one in 4.08×10^6 amino acid triplet codons.

3. The human genome comprises approximately 3.3×10^9 base pairs, corresponding to 1.1×10^9 triplets.

4. Thus, the simple probabilities are that the human genome will contain approximately 270 copies of any particular string of five amino acid-specifying codons. This means that a single 15-mer sequence probe will likely result in about 270 "positive" hybridizations under rigorously stringent conditions. When 14-mer mixed sequence probes are employed, stringency conditions must be relaxed and many fold more non-specific hybridizations will occur, rendering the mixed probe system essentially useless for identifying an erythropoietin gene in the human genome.

5. As one example of the non-specificity of a mixed probe system as described in paragraph 1 above, Applicant notes that the amino acid sequence Ala-Lys-Glu-Ala-Glu appears as residues 1190-1194 in the beta chain of

an E. coli DNA directed RNA polymerase [See, Ovchinnikov et al., European Journal of Biochemistry, 116, 621-629 (1981) attached as Exhibit No. 2 hereto]. To the extent that E. coli DNA is present in the E. coli cell lawn in which phage DNA plaques of a human genomic library reside, probes for the Ala-Lys-Glu-Ala-Glu sequence would hybridize to E. coli DNA as well as to the human DNA encoding erythropoietin and other duplicate DNA sequences in the human, viral and bacterial genomes.

In hindsight, and with knowledge of the precise errors in the Sue et al. sequence, a "best case" set of 17-mer mixed probes could be designed to include a single oligonucleotide probe exactly corresponding to the true erythropoietin DNA sequence. Such a mixture would span the bases making up the codons beginning with lysine at position 19 through the non-degenerate first two bases of the codon for either asparagine or lysine at position 24. The mixture, however, would have had to have included 128 different oligonucleotides. This number of probes is far in excess of 16 or 32 or 48 component mixtures of the Breslow et al. or Woods et al. references.

To Applicant's knowledge, 128 mixed probes had never before been successfully employed in screening a cDNA library, much less a human genomic library which is approximately one hundred times more complex than a cDNA library. More significantly, in the absence of knowledge of the entire DNA sequence of a monkey kidney cDNA library or of a human genome, one cannot determine whether this sequence is unique to erythropoietin DNA. Indeed, it is only in hindsight (based on sequence data first provided herein) that one sees that in the amino terminal encoding region there is

313 -25- 192

Complete homology between monkey and human DNA sequences. Attached hereto as Exhibit No. 3 is Applicant's paper describing the monkey erythropoietin gene isolation [Lin et al., Gene, 44, 201-209 (1986)]. While Applicant's 20-mer mixture of probes corresponding to human sequence residues 46-52 was useful in isolating the monkey gene, the 17-mer mixture corresponding to residues 86-91 could not be used due to the divergence of homology between the human and monkey amino acid sequences. See Exhibit No. 3 and especially the "Conclusions" spanning pages 206-208.

It should then be noted that, as set forth in Example 4, two sets of probes were needed to isolate 3 positive human erythropoietin genomic clones in a screen of 1,500,000 phage plaques. No useful second set of 17-mer or 20-mer probes comprehending this same amino terminal region could be made without substantial amplification of the number of required oligonucleotides beyond 128 required for the first probe set.

Because Applicant could not have used the Sue et el. reference information to follow the Breslow et al. and Woods et al. procedures to screen a human genomic library or a monkey cDNA library without substantially departing from the quite simple procedures disclosed in the references, it cannot properly be argued that the claimed subject matter would have been obvious to a person of ordinary skill in the art at the time Applicant's invention was made. The outstanding rejection of claims 14, 17, 18, 21-24, 26, 27, 31-36, 58 and 61-68 may not therefore properly be applied to corresponding claims 73, 77, 78, 81-84, 86, 87, 91-93, 96, 99-100, 103 and 75, 76, 94, 95, 97, 98, 101 and 102.

8. The "Subsidary" Rejections of Prior Claims 15, 69-72, 19, 20, 25 and 28-30 Under 35 U.S.C. §103 Over the Sue et al. And Other References May Not Properly Be Applied to Corresponding Claims 74, 79, 80, 85 and 88-90

In three paragraphs spanning pages 5 and 6 of the Office Action dated February 5, 1987, the Examiner lodged rejections of certain claims as being unpatentable over the combination of the Sue et al., Breslow et al. and Woods et al. references (as discussed immediately above) in further view of certain additional secondary references. More specifically, claims 15 and 69 through 72 were rejected in view of Talmadge et al. [P.N.A.S. (USA), 77, pp. 3369-3373 (1980)] which deals with expression of a rat proinsulin in E.coli. Claims 19 and 20 were rejected in further view of Farber et al. (apparently either reference C32 or reference C33) which discloses primate erythropoietin messenger RNA isolation. Finally, claims 25 and 28-30 were rejected in view of the additional references to Gouy et al. (PTO Reference "U") or Bennetzen et al. (PTO Reference "R'") concerning preference codons for E.coli and yeast.

Because the primary rejection of claims based on the Sue et al. reference taken together with Breslow et al. or Woods et al. has been demonstrated to be without proper foundation, Applicant submits that the "subsidiary" rejections of claims 15, 69-72, 19, 20, 25, and 28-30 may not properly be applied to corresponding claims.

9. The Rejection of Prio Claims 14, 15, 17-19, 21, 22, 24, 25, 28, 31-34, 36, 58 and 61-72 Under 35 U.S.C. §102(b) in View Of Talmadge et al. May Not Properly Be Applied to Corresponding Claims 73, 74, 77-79, 81-82, 84-85, 88, 91-93, 96, 99, 103 75, 76, 94, 95, 96, 98, 101 and 102

Cross-referencing to the outstanding Section 112 rejection of claims 14, 15, 17-19, 21, 22, 24, 25, 28, 31-34, 36, 58 and 61-72 based on alleged undue breadth of the terms "fragments thereof" and "having at least...activities", the Examiner has interposed a Section 102(b) rejection of these claims based on the Talmadge et al. reference. More specifically, the Examiner argues that the claims are "deemed to embrace the DNA sequences and protein expression method of Talmadge et al.".

Applicant respectfully disagrees with the Examiner's position but submits that the issue is mooted by the amendments to the claims discussed, infra, with respect to Section 112 issues.

10. The Rejection of Claims 69-72 Under 35 U.S. §103 Based on Talmadge et al. Is Mooted By Cancellation of These Claims

Citing to the authority of the decisions of In re Durden and In re Albertson*, the Examiner rejected claims 69-72 as being directed to obvious methods in view of the DNA expression systems described in Talmadge, et al.

As previously noted Applicant has sought cancellation without prejudice of prior claims 69-72 and has not inserted claims corresponding thereto in this amendment. The issues raised by the rejection are no longer present in the application.

---

* The Examiner appears to have inadvertently mis-cited "In re Larsen" rather than In re Albertson.

CONCLUSION

The foregoing remarks are believed to establish that claims 73-103 are in condition for allowance and an early notice under 37 C.F.R. §1.311 is solicited.

Consistent with Applicant's position, enclosed herewith pursuant to 37 C.F.R. §1.311(b) is an "Authorization to Charge Issue Fee to Deposit Account".

Applicant further notes that this case has been made special by the Commissioner (see Notice dated May 2, 1986) and, consistent with MPEP §1309, Applicant requests that this application be suitably "tagged" upon allowance of the claims to allow for priority in printing.

Respectfully submitted,

MARSHALL, O'TOOLE, GERSTEIN,
MURRAY & BICKNELL

By _____
Michael F. Boruh (Reg. No. 25,447)
A Member of the Firm
Attorneys for Applicants
Two First National Plaza
Chicago, Illinois 60603
(312) 346-5750

Chicago, Illinois
March 11, 1987

FIG.5

TABLE V

Translation of Monkey EPO cDNA

Sau3A
GATCCCCGCGCCCCCCTGGACAGCCGCCCTCTCCTCCAGGCCCGTGGGGCTGGCCCTGCCC

CGCTGAACTTCCCGGGATGAGGACTCCCGGTGTGGTCACCGCGCGCCTAGGTCGCTGAG

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

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

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

20 * 40
Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Val Thr Met
GAG AGG TAC CTC TTG GAG GCC AAG GAG GCC GAG AAT GTC ACG ATG

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

318
297

FIG.5

TABLE v (continuted)

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

                  60                                70
Gln Gln Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu
CAG CAG GCT GTA GAA GTC TGG CAG GGC CTG GCC CTG CTC TCA GAA

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

                  90                               100
Phe Glu Pro Leu Gln Leu His Met Asp Lys Ala Ile Ser Gly Leu
TTC GAG CCC CTG CAG CTG CAC ATG GAT AAA GCC ATC AGT GGC CTT

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

                 120                               130
Ile Ser Leu Pro Asp Ala Ala Ser Ala Ala Pro Leu Arg Thr Ile
ATC TCC CTC CCA GAT GCG GCC TCG GCT GCT CCA CTC CGA ACC ATC

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

319

~~298~~

FIG.5

TABLE V (continued)

```
                                                    160
Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg Arg
CTC CGG GGA AAG CTG AAG CTG TAC ACG GGG GAG GCC TGC AGG AGA
                150
                                165
Gly Asp Arg OP
GGG GAC AGA TGA CCAGGTGCGTCCAGCTGGGCACATCCACCACCTCCCTCACCAACA

CTGCCTGTGCCACACCCTCCCTCACCACTCCCGAACCCCATCGAGGGGCTCTCAGCTAAG

CGCCAGCCTGTCCCATGGACACTCCAGTGCCAGCAATGACATCTCAGGGGCCAGAGGAAC

TGTCCAGAGCACAACTCTGAGATCTAAGGATGTCGCAGGGCCAACTTGAGGGCCCAGAGC

AGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGAGCAGACAATGCAGGGAAAACACCT

GAGCTCACTCGGCCACCTGCAAAATTTGATGCAGGACACGCTTTGGAGGCAATTTACCTG

TTTTTGCACCTACCATCAGGGACAGGATGACTGGAGAACTTAGGTGGCAAGCTGTGACTT

CTCAAGGCCTCACGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACTGAGAGAATATT

TTGCAATCTGCAGCAGGAAAAATTACGGACAGGTTTTGGAGGTTGGAGGGTACTTGACAG

GTGTGTGGGGAAGCAGGGCGGTAGGGGTGGAGCTGGCATGCGAGTGAGAACCGTGAAGAC

AGGATGGGGGCTGGCCTCTGGTTCTCGTGGGGTCCAAGCTT
                                                    HindIII
```

320

~~299~~

FIG.6

TABLE VI

```
AAGCTTCTGGGCTTCCAGACCCAGCTACTTTGCGGAACTCAGCAACCCAGGCATCTCTGAGTCTCCGCCCA

AGACCGGGATGCCCCCCCAGGGGAGGTGTCCGGGAGCCCAGCCTTTCCCAGATAGCACGCTCCGCCAGTCCC

AAGGGTGCGCAACCGGCTGCACTCCCCTCCCGCGACCCAGGGCCCGGGAGCAGCCCCCATGACCCACACGC

ACGTCTGCAGCAGCCCCGCTCACGCCCCGGCGAGCCTCAACCCAGGCGTCCTGCCCCTGCTCTGACCCCGG

GTGGCCCCTACCCCTGGCGACCCCTCACGCACACAGCCCTCTCCCCCACCCCCACCCGCGCACGCACACATG

CAGATAACAGCCCCGACCCCCGGCCAGAGCCCGXAGAGTCCCTGGGCCCACCCCGGCCGCTCGCCTGCCGCTG

CGCCGCACCGCGCTGTCCTCCCGGAGCCGGACCGGGGCCCACCGCGCCCXGCTCTGCTCCGACACCGCGCCC

CTTGGACAGCCGCCCTCTCCTCTAGGCCCGTGGGGCTGGCCCTGCACCGCCGAGCTTCCCCGGGATGAGGXX

                                    -27      -24
CCCGGTGACCGGCGCGCCCCCAAGTCGCTGAGGGGACCCCGGCCAAGCGCGGAG  ATG GGG GTG CAC G

GTGAGTACTCGCGGGCTGGGCGCTCCCGGCGCCGGGTTCCTGTTTGAGCGGGGGATTTAGCGCCCCCGGCT
```

321

300

FIG.6

TABLE VI (cont d.)

```
ATTGGCCAAGAGGTGGCTGGGTTCAAGGACCGGCGACTTGTCAAGGACCCCGGAAGGGGGAGGGGGGTGGG

GCAGCCTCCACGTGCCCGCGGGACTTCTTGGGGATGGCAAAAACCTGGCCTGTTGAGGGGCA

CAGTTTGGGGTTGGGGAGGAGGTTTGGGGTTCTGCTGTGCAGTTGTGTCGTTGTCAGTGTCTCG[I.S.]

TTGCACACGCACAGATCAATAAGCCAGAGGCACGACCTGAGTGCTTGCATGGTTGGGACAGGAAGGACGAG

CTGGGGCAGAGACGTGGGGATGAAGGAAGCTGTCCTTCCACAGCCACCCTTCTCCCCCCCGCCTGACTCT

                    -23                 -20
CAGCCTGGCTATCTGTTCTAG    Glu Cys Pro Ala Trp Leu Trp Leu Leu Leu Ser Leu
                          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 Leu Gly Ala Pro Pro Arg Leu Ile Cys
CTG TCG CTC CCT CTG GGC CTC CCA GTC CTG GGC GCC CCA CCA CGC CTC ATC TGT

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

  26
Thr
ACG GTGAGACCCCTTCCCCAGCACATTCCACAGAACTCACGCTCAGGGCTTCAGGGAACTCCTCCCCAGAT

CCAGGAACCTGGCACTTGGTTTGGGGTGGAGTTGCGGAAGCTAGACACTGCCCCCCCTACATAAGAATAAGTC
```

322

304

330

FIG.6

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 GTGAGTTCCTTTTTTTTTTTTTTTTTTCCTTTCTTTTGGAGAATCTCATT

TGCGAGCCTGATTTTGGATGAAAGGGAGAATGATCGGGGGAAAGGTAAAATGGAGCAGCAGAGATGAGGCT

GCCTGGGCGCAGAGGCTCACGTCTATAATCCCAGGCTGAGATGGCCGAGATGGGAGAATTGCTTGAGCCCT

GGAGTTTCAGACCAACCTAGGCAGCATAGTGAGATCCCCCATCTCTACAAACATTTAAAAAAATTAGTCAG

GTGAAGTGGTGCATGGTGGTAGTCCCAGATATTTGGAAGGCTGAGGCGGGAGGATCGCTTGAGCCCAGGAA

TTTGAGGCTGCAGTGAGCTGTGATCACACCACTGCACTCCAGCCTCAGTGACAGAGTGAGGCCCTGTCTCA
```

323

305

FIG.6

TABLE VI (cont'd.)

```
AAAAAGAAAAGAAAAAAGAAAAATAATGAGGGCTGTATGGAATACATTCATTATTCATTCACTCACT

CACTCATTCATTCATTCATTCAACAAGTCTTATTGCATACCTTCTGTTTGCTCAGCTTGGTGCTTGG

GGCTGCTGAGGGCCAGGAGGGAGAGGGTGACATGGGTCAGCTCGACTCCCAGAGTCCACTCCCTGTAG

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

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

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

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

GAAGGGTCTTGCTAAGGAGTACAGGAACTGTCCGTATTCCTTCCCTTTCTGTGGCACTGCAGCGACCTCCT

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

324

303

FIG.6

TABLE VI (cont'd.)

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

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

    166
Asp Arg OP
GAC AGA TGA CCAGGTGTGTCCACCTGGGCATATCCACCACCTCCCTCACCAACATTGCTTGTGCCACA

CCCTCCCCCGCCACTCCTGAACCCCGTCGAGGGGCTCTCAGCTCAGCGCCAGCCTGTCCCATGGACACTCC

AGTGCCAGCAATGACATCTCAGGGGCCAGAGGAACTGTCCAGAGAGCAACTCTGAGATCTAAGGATGTCAC

AGGGCCAACTTGAAGGGCCCAGAGCAGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGGACAGAGCCATGC

TGGGAAGACGCCTGAGCTCACTCGGCACCCTGCAAAATTTGATGCCAGGACACGCTTTGGAGGCGATTTAC

CTGTTTTCGCACCTACCATCAGGGACAGGATGACCTGGAGAACTTAGGTGGCAAGCTGTGACTTCTCCAGG

TCTCACGGGCATGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACCGGGGTGGTGGGAACCATGAAGAC

AXGATXGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTATTCTCAACCTATTGACAGACTGAA

ACACAATATGAC
```

325
~~304~~

FIG.7

TABLE XIV

ECEPO GENE

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

ATCTGCGACT CGAGAGTTCT GGAACGTTAC CTGCTGGAAG CTAAAGAAGC
TAGACGCTGA GCTCTCAAGA CCTTGCAATG GACGACCTTC GATTTCTTCG

TGAAAACATC ACCACTGGTT GTGCTGAACA CTGTTCTTTG AACGAAAACA
ACTTTTGTAG TGGTGACCAA CACGACTTGT GACAAGAAAC TTGCTTTTGT

TTACGGTACC AGACACCAAG GTTAACTTCT ACGCTTGGAA ACGTATGGAA
AATGCCATGG TCTGTGGTTC CAATTGAAGA TGCGAACCTT TGCATACCTT

GTTGGTCAAC AAGCAGTTGA AGTTTGGCAG GGTCTGGCAC TGCTGAGCGA
CAACCAGTTG TTCGTCAACT TCAAACCGTC CCAGACCGTG ACGACTCGCT

GGCTGTACTG CGTGGCCAGG CACTGCTGGT AAACTCCTCT CAGCCGTGGG
CCGACATGAC GCACCGGTCC GTGACGACCA TTTGAGGAGA GTCGGCACCC

AACCGCTGCA GCTGCATGTT GACAAAGCAG TATCTGGCCT GAGATCTCTG
TTGGCGACGT CGACGTACAA CTGTTTCGTC ATAGACCGGA CTCTAGAGAC

ACTACTCTGC TGCGTGCTCT GGGTGCACAG AAAGAGGCTA TCTCTCCGCC
TGATGAGACG ACGCACGAGA CCCACGTGTC TTTCTCCGAT AGAGAGGCGG

GGATGCTGCA TCTGCTGCAC CGCTGCGTAC CATCACTGCT GATACCTTCC
CCTACGACGT AGACGACGTG GCGACGCATG GTAGTGACGA CTATGGAAGG

GCAAACTGTT TCGTGTATAC TCTAACTTCC TGCGTGGTAA ACTGAAACTG
CGTTTGACAA AGCACATATG AGATTGAAGG ACGCACCATT TGACTTTGAC

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

326
305

FIG.8

TABLE XXI

SCEPO GENE

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

GGAAAGATAC TTGTTGGAAG CTAAAGAAGC TGAAAACATC ACCACTGGTT
CCTTTCTATG AACAACCTTC GATTTCTTCG ACTTTTGTAG TGGTGACCAA

GTGCTGAACA CTGTTCTTTG AACGAAAACA TTACGGTACC AGACACCAAG
CACGACTTGT GACAAGAAAC TTGCTTTTGT AATGCCATGG TCTGTGGTTC

GTTAACTTCT ACGCTTGGAA ACGTATGGAA GTTGGTCAAC AAGCTGTTGA
CAATTGAAGA TGCGAACCTT TGCATACCTT CAACCAGTTG TTCGACAACT

AGTTTGGCAA GGTTTGGCCT TGTTATCTGA AGCTGTTTTG AGAGGTCAAG
TCAAACCGTT CCAAACCGGA ACAATAGACT TCGACAAAAC TCTCCAGTTC

CCTTGTTGGT TAACTCTTCT CAACCATGGG AACCATTGCA ATTGCACGTC
GGAACAACCA ATTGAGAAGA GTTGGTACCC TTGGTAACGT TAACGTGCAG

GATAAAGCCG TCTCTGGTTT GAGATCTTTG ACTACTTTGT TGAGAGCTTT
CTATTTCGGC AGAGACCAAA CTCTAGAAAC TGATGAAACA ACTCTCGAAA

GGGTGCTCAA AAGGAAGCCA TTTCCCCACC AGACGCTGCT TCTGCCGCTC
CCCACGAGTT TTCCTTCGGT AAAGGGGTGG TCTGCGACGA AGACGGCGAG

CATTGAGAAC CATCACTGCT GATACCTTCA GAAAGTTATT CAGAGTTTAC
GTAACTCTTG GTAGTGACGA CTATGGAAGT CTTTCAATAA GTCTCAAATG

TCCAACTTCT TGAGAGGTAA ATTGAAGTTG TACACCGGTG AAGCCTGTAG
AGGTTGAAGA ACTCTCCATT TAACTTCAAC ATGTGGCCAC TTCGGACATC

AACTGGTGAC AGATAAGCCC GACTGATAAC AACAGTGTAG
TTGACCACTG TCTATTCGGG CTGACTATTG TTGTCACATC

            SalI
ATGTAACAAA G
TACATTGTTT CAGCT
```

327 ~~304~~

IN THE UNITED STATES PATENT
AND TRADEMARK OFFICE

Application of:	)	"PRODUCTION OF ERYTHROPOIETIN"
FU-KUEN LIN	)	
Serial No: 675,298	)	Group Art Unit
Filed: November 30, 1984	)	Examiners - J. Giesser T. G. Wiseman

DECLARATION AS TO DEPOSIT OF MICROORGANISM

GEORGE B. RATHMANN, DECLARES AND SAYS:

1. That he is President of Kirin-Amgen, Inc., the owner of the above-identified patent application,

2. That certain cultures have been deposited in the permanent culture collection of the American Type Culture Collection, 12031 Parklawn Drive, Rockville, Maryland, and the culture deposits meet the requirements of the Budapest Treaty for the purposes of patent procedure,

3. That said cultures have been accorded the accession numbers 39881, 39882, 39932, 39933, 39934, 20733 and 20734, prior to the filing date of the above-identified patent application;

4. That said cultures have been deposited under conditions which ensure that access thereto will be available during the pendency of the above-identified patent application to one determined by the Commissioner to be entitled thereto under 37 C.F.R. 1.14 and 35 U.S.C. 122,

5. That upon issuance of a patent on the above-identified patent application Kirin-Amgen, Inc. will remove any restrictions as to public availability of the culture deposit, and will replace the same culture deposit should it become nonviable, during the period that extends thirty years from the date of the deposit, or the period of the enforceable life of the patent, or the period of five years after the last public request for the deposit, whichever period is longest; and

309378

336

6. That he declares further that all statements made herein of his own knowledge are true and that all statements made on information and belief and believed to true, and further that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 10 of the United States Code and that such willful false statements may jeopardize the validity of the instant patent application or any patent issuing thereon.

KIRIN-AMGEN, INC.

By : George B. Rathmann
George B. Rathmann
President

Date: September 26, 1986

EpoLtr2

329
~~328~~

337

Exhibit No. 2

Eur. J. Biochem. 122, 621–629 (1982)
FEBS 1982

The Primary Structure of *Escherichia coli* RNA Polymerase

Nucleotide Sequence of the *rpoB* Gene and Amino-Acid Sequence of the β-Subunit

Yury A. OVCHINNIKOV, Galina S. MONASTYRSKAYA, Valentin V. GUBANOV, Sergey O. GURYEV, Oleg Yu. CHERTOV, Nikolay N. MODYANOV, Vladimir A. GRINKEVICH, Irina A. MAKAROVA, Tatjana V. MARCHENKO, Irina N. POLOVNIKOVA, Valery M. LIPKIN, and Eugene D. SVERDLOV

Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, Moscow

Received November 5, 1980 March 2, 1981

The combined structural study of proteins and of their corresponding genes utilizing the methods of both protein and nucleotide chemistry greatly accelerates and considerably simplifies both the nucleotide and protein structure determination and, in particular, enhances the reliability of the analysis. This approach has been successfully applied in the primary structure determination of the β and β′ subunits of *Escherichia coli* DNA-dependent RNA polymerase and of their structural genes, yielding a continuous nucleotide sequence (4714 base pairs) that embraces the entire *rpoB* gene, the initial part of the *rpoC* gene and the intercistronic region, together with the total amino acid sequence of the β subunit, comprising 1342 residues, and the N-terminal sequence of the β′ subunit (176 residues).

Elucidation of the transcription mechanism requires detailed knowledge of the active-center organization of RNA polymerase at the various stages of the RNA synthesis. This, in turn, can be obtained only after determining the primary and spatial structure of the enzyme.

Earlier we had established the amino acid sequence of the β subunit of *Escherichia coli* DNA-dependent RNA polymerase by resorting solely to the ordinary methods of protein chemistry [5]. In the case of the β and β′ subunits with their much higher molecular weights (≈ 155000 and ≈ 165000, respectively) [6], such an approach could no longer suffice, in view of the difficulties in isolating and purifying the resulting fragments and in reconstituting the amino acid sequence via overlapping peptides [7].

The progress in DNA sequencing methods and the possibility of using the genetic code to obtain information on the primary protein structure from the nucleotide sequences is an attractive way to circumvent such difficulties; although here, too, there are many pitfalls, requiring considerable caution to avoid possible sources of error.

In the first place the mRNA can undergo processing, leading to erroneous deduction of the protein structure. Secondly, the protein itself can be processed. Thirdly, it is often difficult to recognize in the overall DNA structure the beginning of a structural gene. The criterion for this purpose is the presence of an initiating codon together with the adjacent sequences complementary to the 3′ end of 16-S RNA [8, 9]. Frequently more than one such combination can be found for one and the same protein. Such cases have been observed both by us and by other authors [10, 11]. Moreover, one has to bear in mind that a single error (deletion or insertion) in the DNA sequence determination could lead to a completely erroneous amino acid sequence of the protein. Thus, primary structure determination of DNA cannot serve as a substitute for the direct sequencing of the protein.

In view of this, we decided to utilize the methods of both protein and nucleotide chemistries, performing the parallel sequencing of the structural genes *rpoB* (β subunit) and *rpoC* (β′ subunit) and of the corresponding proteins. Knowledge of the nucleotide sequence of the pertinent DNA segments would permit aligning of the peptide fragments from the protein analysis into an uninterrupted polypeptide chain. Such an approach provides the key to the most complicated problem in the primary structure analysis of high-molecular-weight proteins, namely, reconstitution of the amino acid sequence. In the classical methods, this requires the isolation and sequencing of a large number of relatively big overlapping peptides, whereas in this approach one can utilize small (e.g. 5–20 amino acid residues) and non-overlapping fragments with the only proviso that they cover the entire length of the polypeptide chain. Methods for their separation are routine and their structural determination requires only small (10–30 nmol) amounts of peptide.

At present, methods are available for the complete [12] or partial [13] insertion of the *E. coli rpoBC* operon containing the β and β′ subunit genes into the DNA of transducing phages. This permitted its functional study and simplified the structural problem. The *Eco*RI restriction endonuclease cleavage map of the *E. coli* DNA region containing the structural genes of the β and β′ subunits of RNA polymerase has been plotted by several workers [14–16] (see Fig. 1). The present paper reports the total sequencing of the *Eco*RI-C and *Eco*RI-F fragments and partial sequencing of the *Eco*RI-G fragment carrying the beginning of the *rpoB* gene. The structure of the larger part of the *Eco*RI-G segments adjacent to the *Eco*RI-E segment had been elucidated earlier by Post et al. [17]. The sequence of 33 N-terminal amino acids of the β subunit is coded in this structure. Gurevich et al. [18] have also recently published a paper on the struc-

Preliminary papers to this report appeared elsewhere [1–4].

Enzymes. RNA polymerase or nucleosidetriphosphate : RNA nucleotidyltransferase (EC 2.7.7.6); polynucleotide kinase or ATP : 5′-dephosphopolynucleotide 5′-phosphotransferase (EC 2.7.1.78); restriction endonucleases *Eco*RI (EC 3.1.23.13), *Taq*I (EC 3.1.23.39), *Hin*fI (EC 3.1.23.22), *Sau*3AI (EC 3.1.23.27) and *Sal*I (EC 3.1.23.37); trypsin (EC 3.4.21.4); *Staphylococcus aureus* protease (EC 3.4.21.19).

331 350

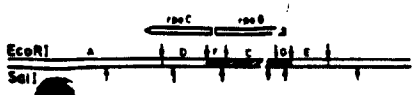


Fig. 1. *EcoRI* [redacted]*all restriction cleavage map of the* E. coli *DNA region including the structural genes (rpoB and rpoC) of the β and β′ RNA polymerase subunits.* Hatched segments correspond to those in [16]

ture of this segment. There are a number of divergences between our results and those in the latter paper (see discussion in [3]).

In general, the continuous nucleotide chain we have sequenced (4714 base pairs) contains the entire *rpoB* gene, the beginning of the *rpoC* gene and the corresponding intercistronic region. Comparing this with the amino acid sequences of the peptides from the β subunit, we have deduced its complete amino acid sequence (1342 amino acid residues). With the aid of an automatic sequencer the N-terminal sequence (8 amino acid residues) of the β′ subunit was determined, permitting location of the beginning of its translation and from this deduction of the sequence of 176 initial β′ subunit amino acid residues.

MATERIALS AND METHODS

*Eco*RI fragments were obtained from the DNA of pOD-162 and pJB-35 plasmids [19] and of λ rif^{s} 47 phage by *Eco*RI restriction endonuclease digestion. The fragments were separated either by centrifugation in a sucrose gradient, similar to the procedure described in [20], or by preparative horizontal electrophoresis in 1% agarose blocks [21].

The *Eco*RI [22], *Sal*I [23] and *Taq*I [24] restriction endonucleases were isolated as described in the cited references. Restriction endonucleases *Hin*fI and *Sau*3AI were the generous gifts of Prof. Zachau and Prof. Müller-Hill (FRG).

The phage T4 polynucleotide kinase was isolated according to [25]. The enzymes did not display phosphatase or nuclease activity. [γ-^{32}P]ATP (2000 Ci/mmol) was from Amersham (England).

*Hin*fI, *Sau*3AI and *Sal*I restriction endonuclease digestion of the *Eco*RI-C and *Eco*RI-F fragments was carried out for 2 h at 37 °C in a buffer solution (pH 7.5) containing 0.01 M Tris/HCl, 0.01 M $MgCl_2$ and 0.01 M mercaptoethanol, utilizing 1 unit enzyme activity/μg DNA fragment. The *Eco*RI-G fragment was cleaved with *Sau*3AI + *Sal*I and *Hin*fI + *Sal*I mixtures under the same conditions.

The *Eco*RI-C and *Eco*RI-F fragments were digested similarly with *Taq*I, but at 70 °C. For their cleavage with *Hpa*II restriction endonuclease use was made of a buffer containing 0.01 M Tris/HCl, 0.01 M $MgCl_2$, 0.006 M KCl, 0.001 M dithiothreitol (pH 7.4), the hydrolysis being carried out under the same conditions (37 °C).

Sequence determination of the nucleotide subfragments was carried out by our modification of the Maxam and Gilbert procedure [26]. Complementary strands were obtained according to [27]. The purine base positioning was achieved by spurinization of the DNA with 60% HCOOH [28,29]. The positions of the thymidine units were determined by treatment with OsO_4 [30–32]; after incubation at 0 °C for 5 min, 200 μl of 0.3 M sodium acetate (pH 5.5) and 750 μl of alcohol cooled to 0 °C were added, and the precipitate was treated with piperidine as described in [26]. For the chemical modification use was made of the isolated complementary chains.

Tryptic and Staphylococcus aureus Protease Digestion of the β Subunit

The β subunit (2 μmol) was dissolved in 50 ml of 6 M guanidine · HCl + 0.34 ml mercaptoethanol buffer (pH 8.6) and carboxymethylated with iodoacetic acid according to [33]. The protein solution was desalted on a Sephadex G-50 column (2.5 × 100 cm) equilibrated with 6 M urea, 0.01 M Tris/HCl buffer (pH 8.5). To the solution thermostated at 2–4 °C, was added 1 ml of freshly distilled citroconic anhydride, maintaining the pH constant at 8.5 by NaOH additions with a TT1/BP2 (Radiometer, Denmark) titrator. The protein was desalted on a Sephadex G-50 column equilibrated with ammonia solution (pH 9–10) and was then lyophilized. The lyophilized preparation was taken up in 50 ml 0.1 M ammonium bicarbonate buffer (pH 8.0) and 5 mg of *Staphylococcus aureus* proteinase (Miles, England) or of trypsin (Worthington, USA) was added. In the former case the digestion was carried out for 20 h at 37 °C; in the latter for 4 h at the same temperature. After hydrolysis the mixture was lyophilized.

The separation, purification and sequencing of the peptides was carried out by methods described earlier [34–37].

RESULTS AND DISCUSSION

The Strategy of the DNA Primary Structure Elucidation

The general strategy for determining the nucleotide sequence of the *Eco*RI-G, *Eco*RI-C and *Eco*RI-F fragments is shown in Fig. 2. The fragments were consecutively digested with one of the restriction endonuclease (*Sau*3AI, *Hin*fI, *Hpa*II and *Taq*I) cleaving the DNA into relatively small blocks. The *Eco*RI-G fragment was digested with *Hpa*II and a mixture of *Sal*I + *Sau*3AI and also of *Sal*I + *Hin*fI. The resulting subfragments were phosphorylated by means of [γ-^{32}P]ATP and phage T4 polynucleotide kinase and the mixture was separated by electrophoresis on polyacrylamide gel. As a rule both complementary chains obtained after denaturation of each subfragment and separation on polyacrylamide gel were analyzed. Altogether about 97% of the total sequence was determined from both strands. Whenever the subfragment chains could not be separated, the cleavage was repeated with other restriction endonucleases (for instance *Eco*RI-C–*Sau*-F was digested with *Hpa*II). The *Eco*RI fragments were reassembled via the overlapping subfragment sequences. The total sequence, presented in Fig. 3, was reconstituted from the sequences of three *Eco*RI fragments based on the continuity of the template chain (see below).

Determination of the Subfragment Sequences

Sequencing of the complementary strands was performed by the Maxam-Gilbert procedure [26], but instead of the usual reaction for determining dA > dG, we applied partial apurinization by formic acid that we had proposed earlier [28,29] for locating purine units in oligonucleotides and polynucleotides. We utilized this reaction in previous work [10] and it has also been employed by other workers [38]. The method is simple and always gives reliable results.

In a number of cases we utilized reaction with osmium tetraoxide in the presence of amines for locating thymidine units in oligonucleotides and polynucleotides [30,31]. Soon after, this procedure was employed by Brown [32]. In general

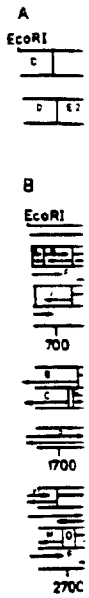


Fig. 2. *Res*
*(B) Eco*R
sequence.
operon in
correspon
The restri
designate
in the *Ec*
tion of the

it gives g
in the s
degree
upon th
In t
5-methy
modifie
agreem
lysis ba
weakly
the san
quencir
such c

311 332

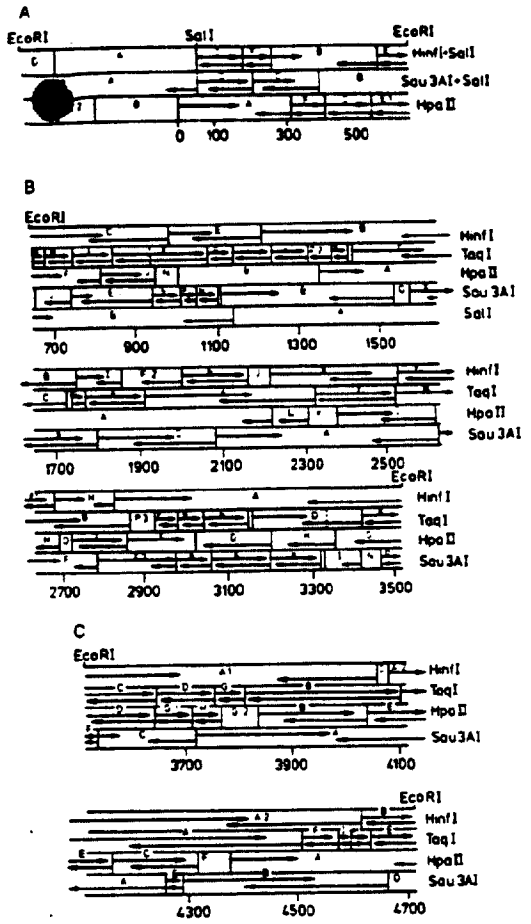


Fig. 2. *Restriction endonuclease cleavage map for fragments (A)* Eco*RI-G, (B)* Eco*RI-C and (C)* Eco*RI-F and the scheme for determining their sequence.* The aligment of fragments is opposite to that of the *rpoBC* operon in Fig. 1. The numbering in the lower part of the fragments corresponds to sequence of the *rpoBC* operon fragment shown in Fig. 3. The restriction subfragments are represented by rectangles. The arrows designate lengths of the subfragments determined. Vertical dotted lines in the *Eco*RI-C – *Taq*-I and *Eco*RI-G – *Taq*-E segments mark the position of the d(T-C-G-A) sequence not cleavable by *Taq*I

it gives good results, but is apparently sensitive to peculiarities in the secondary structure. Consequently, differences in the degree of modification were sometimes noted, depending upon the position of the bases in the polynucleotide chain.

In the process of the structural elucidation we found 5-methylcytidine (C*) in 16 sites (Fig. 3). In all cases the modified base was in the sequence 5'-d(C-C-$\frac{A}{T}$-G-G)-3', in agreement with the results of Maxam [39]. After hydrazinolysis bands corresponding to 5-methylcytidine appear only weakly or are completely absent from the structural gel. At the same time, when osmium tetraoxide is used for the sequencing, 5-methylcytidine is modified like thymidine. In such cases possible errors were eliminated by analyzing the complementary ... Moreover, the correctness of the derived sequence was often confirmed by comparison with the amino acid sequence of the corresponding peptides.

Another example of digression from the regular band pattern in the structural gel is compression. It was most clearly manifested during our analysis of the template chains of the *Hpa*II fragment for the nucleotides 364 – 366 and 398 – 401. In the latter case inversion as well as compression was observed, so that the partial cleavage products corresponding to dT-398 displayed a higher electrophoretic mobility than the shorter dC-399 cleavage products. Inversion was also observed in analyzing the template chain of the *Hin*fI fragment for the nucleosides dC-2481 and dG-2482. In addition note should be made of compression in the nontemplate chains of the fragments: *Hin*fI (311, 312), *Hin*fI (4267 – 4269) and in the template chains of the *Hin*fI (2444, 2445) and *Taq*I (3753, 3754) fragments. Often such compression cannot be explained by 'hairpin' formation. The correctness of the proposed structure for all these segments was confirmed by the sequence of the complementary chain and the amino acid sequence of the peptides.

Determination of the Template Chains and Establishment of the Translation Frames in Individual Fragments

In the case of *Eco*RI fragments the template chain is easily defined due to the asymmetric position of the sole *Sal*I site. According to the physical genetic maps of the *rpoBC* operon region [14 – 16], in each fragment this site is adjacent to the operon promoter. Consequently, in every case where the DNA chain is the template chain, the 3' end is located closer to the point of *Sal*I cleavage. The complementary chain possesses a sequence corresponding to that of the mRNA

The results of the nucleotide sequence determinations were constantly compared with those of the amino acid sequencing of the β-subunit peptides, facilitating determination of the template chain and the translation frame in the individual fragments and also helping to establish their mutual orientation.

Isolation of the Peptides and Determination of Their Amino Acid Sequence

Investigation of the primary structure of the β subunit began with its limited tryptic hydrolysis. The analytical data showed that in this case the β subunit splits into five large fragments (M_r 62000, 52000, 37000, 24000 and 10000). However in the preparative experiment their yields were quite low and the mixture contained peptides of smaller size [7]. From the hydrolyzate 53 low-molecular-weight peptides were isolated, in all containing 380 amino acid residues. The complete sequence was established for 46 of these peptides and partial sequences for the remaining 7 [34]. Isolation of the high-molecular-weight peptides proved difficult because of both the limited hydrolytic specificity and the low yield of most products.

As an alternative method, splitting the polypeptide chain by digestion with *Staphylococcus aureus* protease was chosen, a method we had effectively used in our primary structure study of the α subunit [5]. The enzyme mainly cleaves glutamic carboxyl bonds. The digestion was carried out with an enzyme : substrate ratio of 1:60 in 0.1 M ammonium bicarbonate buffer (pH 8). In order to improve the solubility of the β subunit in this buffer, before the hydrolysis the carboxymethylated protein was treated with citraconic anhydride.

3FF 333

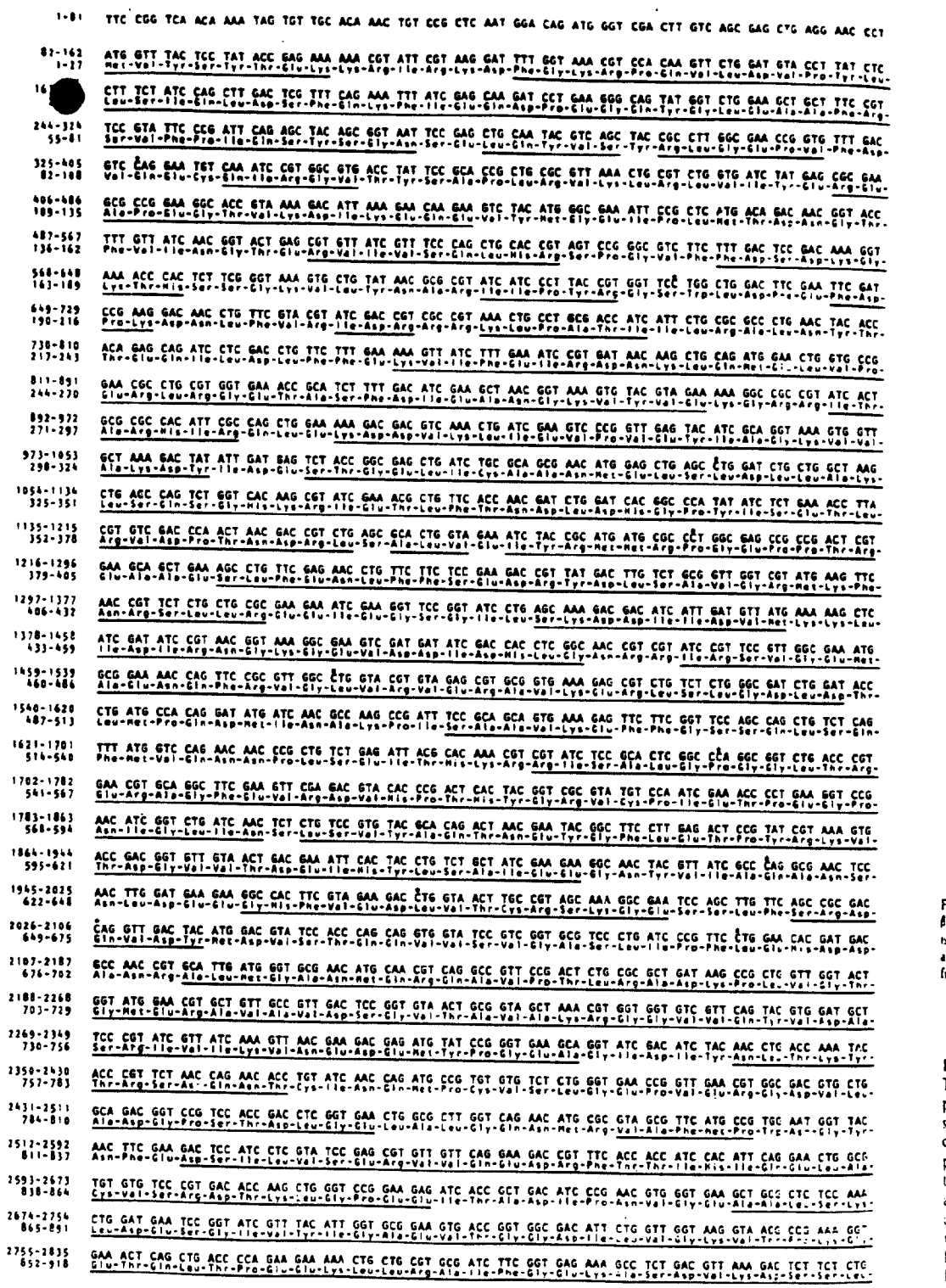


Fig 3
β sub
n give
are th
fragm

150 p
This
pepti
smal
chro
pape
the
anal
addi
was
frag
II –

213 334

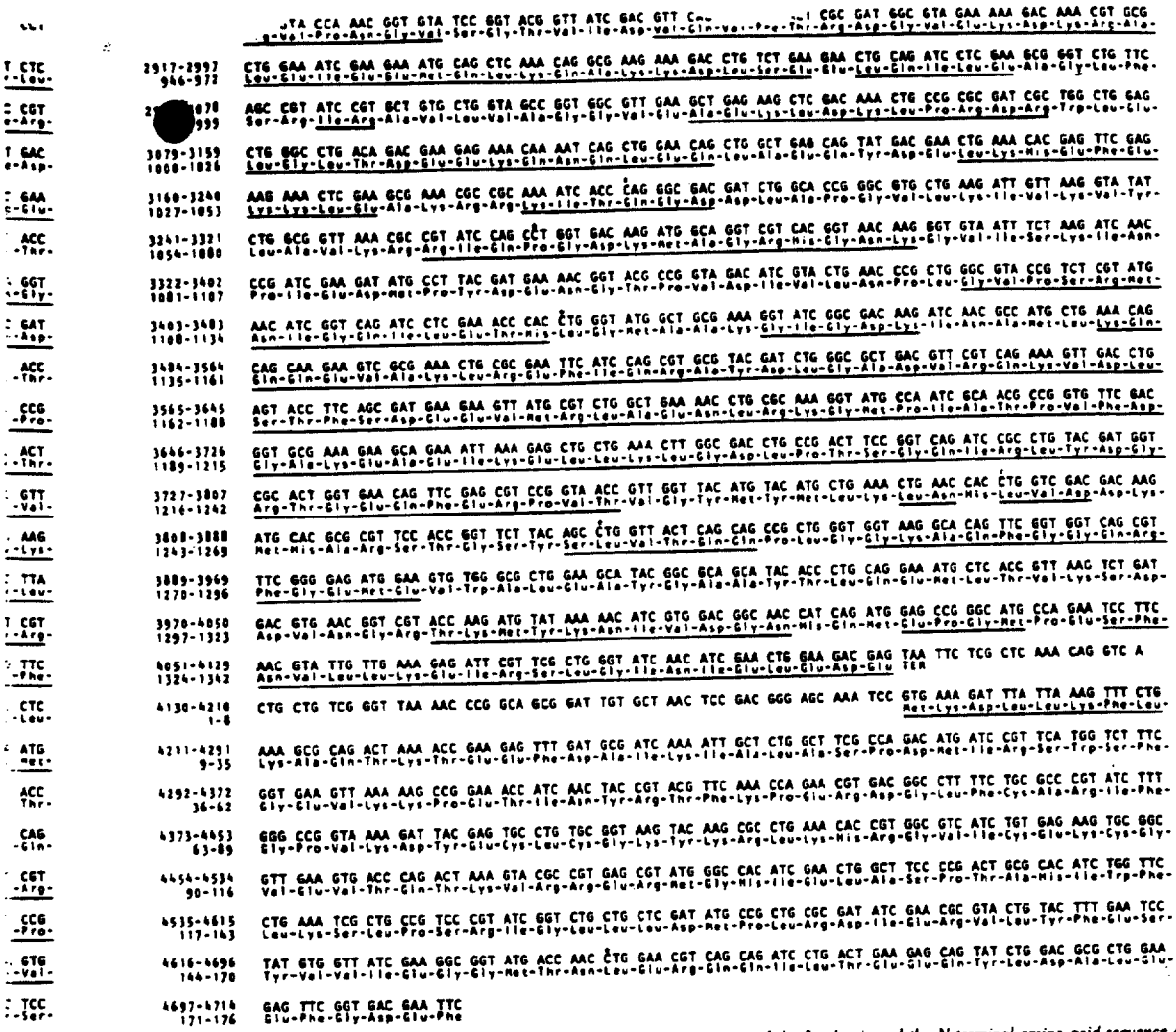


Fig. 3. *The nucleotide sequence of the* rpoBC *segment, the total amino acid sequence of the β subunit and the N-terminal amino acid sequence of the β′ subunit of* E. coli *RNA polymerase.* Here the nucleotide sequence of the complementary DNA chain, corresponding to the sequence of mRNA, is given but the d representing deoxy and the hyphens representing phosphodiester linkages have been omitted. The underlined amino acid sequences are those of which the structure has been determined from analysis of corresponding peptides. The restriction *Eco*RI cleavage sites dividing the fragments *Eco*RI-G, *Eco*RI-C and *Eco*RI-F are situated between nucleotides 640 – 641 and 3508 – 3509. C* = 5-methylcytidine

Initial fractionation of the hydrolysate containing about 150 peptides was performed by gel filtration on BioGel P-4. This yielded four fractions. Subsequent separation of the peptides into fractions II, III and IV containing relatively small peptides (2 – 20 amino acid residues) was achieved by chromatography on the cation exchanger AG-50Wx4 and paper chromatography. Fraction I constituted a mixture of the largest peptides. In order to facilitate separation and analysis of the peptides in this fraction the mixture was additionally digested with chymotrypsin. The hydrolyzate was desalted on Sephadex G-10 and the resulting peptide fragments were separated similarly to those of fractions II – IV.

Altogether three peptide fragments were isolated from fraction IV, 73 from fraction III, 48 from fraction II and 60 from fraction I. The N-terminal sequence of the peptides was determined by the Edman degradation, the amino acids being identified as their phenylthiohydantoin or dansyl derivatives. The C-terminal sequence was established with the aid of carboxypeptidase Y. At present, the total sequences of 122 and partial sequences of 22 peptides, containing 1027 and 269 amino acid residues, respectively, have been determined.

Obviously, not all peptides could be isolated from such a complex mixture as formed in the staphylococcal protease hydrolysis of the *β* subunit. In order to obtain the missing fragments, exhaustive tryptic digestion of the subunit at the

314/335

arginine residues was carried out after modification of the lysine residues with citraconic anhydride. The tryptic peptides were separated according to the same procedure as that used for isolating peptides from the staphylococcal protease hydrolysate. After removal of the citraconic protection, the high-molecular weight peptides were subjected to additional tryptic cleavage at the lysine residues.

Combining Determination of the DNA Primary Structure with Structural Study of the Peptides

The amino acid sequences established from analysis of the tryptic and staphylococcal protease peptides of the RNA polymerase β subunit have been underlined in Fig. 3. The search for correspondence between the amino acid sequences of the peptides and the nucleotide sequences of the DNA fragments was carried out by means of a computer. As a rule, the peptide sequence was compared with three possible amino acid sequences. When the template chain was not identified six possible amino acid sequences were investigated. The selection program was drawn up such that during the comparison process, possible errors in either the DNA or the peptide structures could be detected. As shown by the results of this study the combined amino acid sequence determination of the protein and the nucleotide sequence determination of its structural gene enables one with the aid of a computer readily to find the position of each peptide fragment in the polypeptide chain of the β subunit (Fig. 3).

The known amino acid sequences of the peptides simplified the elucidation of the nucleotide sequence. For instance the structures of fragments *Eco*RI-G–*Hpa*II-C and *Eco*RI-G–*Hpa*II-F were joined together into a continuous sequence by virtue of the Arg-Glu-Ala-Pro-Glu (400–414) peptide overlapping their terminal sequences. As noted earlier, the known structure of the peptides also helped to define the template chain and the translation frame in the DNA subfragments.

Sequence of the Escherichia coli rpoB *Gene and* E. coli *RNA Polymerase β Subunit*

The N-terminal amino acid sequence of the β subunit was determined by means of an automatic sequencer. As well as the major sequence Met-Val-Tyr-Ser-Tyr-Thr-Glu-Lys-, a minor one, Val-Tyr-Ser-Tyr-Thr-Glu-Lys- was also revealed [40], apparently generated from cleavage of the N-terminal methionine in the post-translational modification of the β subunit. The corresponding nucleotide sequence is revealed at the coordinates 82–105 base pairs (Fig. 3). The sequence corresponding to the initiating codon for the β subunit is apparently d(A-T-G), which is preceded by the ribosome binding site d(G-A-G-G) (72–75). The d(A-T-G) triplet is followed by a purine nucleoside dG which, according to [41], increases the probability of initiation at just that site. Yet another argument in favor of this codon being the initiating one is the absence of the d(T-A-G) sequence for a stop codon in the preceding sequence for a distance of at least 50 nucleotides [42].

A peptide Ser-Leu-Gly-Ile-Asx-Ile-Glx(Leu,Glx,Asx,Glx) was isolated from the tryptic hydrolysate of the β subunit; that it did not contain basic amino acid residues suggested that it might be the C terminus [34]. After staphylococcal protease digestion, the peptides Ile-Arg-Ser-Leu-Gly-Ile-Asn-Ile-Glu and Leu-Glu-Asp-Glu were isolated from the same region of the polypeptide chain. The nucleotide sequence corresponding to these peptides (4069–4107) is found immediately before the sequence d(T-A-A) for the terminating codon. Hence these peptides are in fact the C-terminal peptides of the β subunit.

Table 1. *Amino acid composition of the β subunit of DNA-dependent RNA polymerase from* E. coli

Amino acid	Data from primary structure		Data from amino acid analysis		
	no. of residues	proportion	[37]	[43]	[6]
		mol 100 mol			
Asp	91	6.78			
Asn	51	3.80			
Asp + Asn	142	10.58	10.7	10.25	11.3
Thr	60	4.47	4.4	4.90	4.9
Ser	74	5.51	5.7	5.95	5.4
Glu	122	9.09			
Gln	58	4.32			
Glu + Gln	180	13.41	13.5	13.75	14.4
Pro	56	4.17	4.3	4.38	5.6
Gly	106	7.90	8.2	7.88	8.4
Ala	79	5.89	6.7	6.48	7.9
½Cys	7	0.52	1.0	0.70	–
Val	110	8.20	8.4	7.88	6.2
Met	37	2.76	2.7	2.27	–
Ile	84	6.26	5.9	5.69	5.1
Leu	127	9.46	9.5	9.72	9.6
Tyr	43	3.20	3.1	2.80	3.0
Phe	44	3.28	2.9	2.89	3.1
His	19	1.42	1.3	1.57	1.5
Lys	80	5.96	5.3	5.60	6.3
Arg	90	6.71	6.3	6.83	7.3
Trp	4	0.30	0.5	0.35	–

Consequently the polypeptide chain of the β subunit consists of 1342 amino acid residues (Fig. 3) (M_r 150618.6). Table 1 gives the amino acid composition of the protein both deduced from its primary structure and as calculated earlier by us [37] and by others [6, 43] on the basis of its amino acid analysis.

The β subunit is an acid protein: there are 213 dicarboxylic acid residues but only 170 basic amino acid residues.

The amino acid sequence determined up to now constitutes more than 70% of the total β-subunit sequence as deduced from the structure of the *rpoB* gene. When comparing the two sequences it should be born in mind that the β subunit itself came from the RNA polymerase of a wild strain of *E. coli*, whereas the DNA for the nucleotide sequence determination had undergone mutations in the *rpoB* gene leading to dominant rifamycin stability. According to Nomura [15] and Collins [16], *rif*^{d} mutation (or mutations) are localized in the *Eco*RI-C segment. Available data as yet do not define its exact location.

Utilization of Codons in Translating the rpoB *Gene mRNA*

Table 2 presents codon usage data in the translation of *rpoB* gene as compared with the average for the genomes of bacteria, mitochondria and DNA-containing bacteriophages. The comparison revealed *rpoB* to be a typical bacterial gene in this respect; for all amino acids, those codons which are most frequently utilized are those that are predominant in the bacterial genome. An exception is Ala in the *rpoB* gene, the most frequent codons for which are GCA and

395 336

Table 2. *Codon usage in rpoB gene translation as compared with the average usage frequencies in other systems*

Amino acid	Codon	Frequency of use in the gene or genome				
		rpoB	bacteria[a]	mitochondria[a]	DNA-containing phage[a]	[17] ribosomal protein genes
Arg	CGA	1	3	0	16	0
	CGC	28	20	0	27	12
	CGG	0	2	0	0	0
	CGU	61	20	0	15	16
	AGA	0	2	13	6	0
	AGG	0	0	0	9	0
Leu	CUA	0	3	13	5	0
	CUC	15	3	0	12	1
	CUG	99	59	0	19	43
	CUU	6	7	0	25	1
	UUA	1	6	147	13	1
	UUG	6	6	0	12	1
Ser	UCA	0	3	67	10	0
	UCC	31	14	0	4	9
	UCG	3	4	0	5	1
	UCU	23	18	0	14	12
	AGC	15	9	0	13	5
	AGU	2	6	0	8	1
Thr	ACA	3	4	27	17	1
	ACC	34	21	0	22	15
	ACG	6	6	0	6	0
	ACU	17	24	0	13	18
Pro	CCA	10	6	13	10	2
	CCC	0	5	0	3	0
	CCG	38	26	0	14	21
	CCU	8	1	13	5	0
Ala	GCA	22	40	53	31	31
	GCC	9	18	13	15	7
	GCG	29	28	0	22	16
	GCU	19	70	67	42	60
Gly	GGA	0	4	27	13	0
	GGC	35	31	0	12	24
	GGG	2	5	0	10	0
	GGU	69	31	107	16	25
Val	GUA	31	34	67	7	28
	GUC	14	9	0	6	3
	GUG	24	14	0	9	7
	GUU	41	42	13	26	35
Lys	AAA	56	65	27	48	55
	AAG	24	11	0	39	8
Asn	AAC	48	21	13	35	18
	AAU	3	7	13	11	2
Gln	CAA	8	14	13	24	5
	CAG	50	23	0	22	15
His	CAC	18	6	0	0	3
	CAU	1	4	0	8	0
Glu	GAA	89	57	13	38	41
	GAG	33	14	0	41	7
Asp	GAC	61	30	13	26	24
	GAU	30	19	0	22	8
Tyr	UAC	29	10	0	10	7
	UAU	14	4	13	11	1
Cys	UGC	2	4	0	3	2
	UGU	5	2	13	3	0
Phe	UUC	33	12	80	22	10
	UUU	11	12	13	16	6
Ile	AUA	0	3	0	5	0
	AUC	66	28	27	29	22
	AUU	18	18	93	26	9
Met	AUG	37	29	27	32	19
Trp	UGG	4	3	0	17	0

[a] Average frequency per 1000 codons [44].

GCG, whereas in bacteria it is GCU. There are differences from the typical codon spectrum for bacteria also with respect to histidine and phenylalanine.

It can be seen that considerably greater deviation from the typical picture occurs in the bacteriophage genome codon usage. A completely different spectrum is found in mitochondria. Large differences can be observed also when compared with genes of animal origin [44]. A close similarity is observed with codon usage for the expression of the ribosomal proteins L11, L1, L10 and L7/L12 [17], the genes of which precede the *rpoB* gene in the *E. coli* genome. The non-random selection of codons often is found to correlate with the amount of the corresponding isoacceptor tRNA in the cell (for a review see [17]) and is possibly due to necessity of the most effective and correct translation.

The N-Terminal Sequence of the rpo-C *Gene and the Intercistronic Region*

With the aid of an automatic sequencer we established the N-terminal sequence for the β' subunit of RNA polymerase as Met-Lys-Asp-Leu-Leu-Lys-Phe-Leu- which is in agreement with that found earlier [43]. The presence of the terminal methionine gives grounds for the belief that the subunit did not undergo post-translational processing. The corresponding nucleotide sequence is found in the *Eco*RI-F segment (4187–4210). Here the N-terminal methionine residue is coded by the rare initiating triplet d(G-T-G) (4187–4189). Further evidence for the initiating role of this triplet is: (a) directly adjacent to its 5′ end is the sequence d(G-G-A-G) (4176–4179) complementary to the 3′ terminus of 16-S RNA; (b) adjacent to it from the 3′ end is the purine nucleoside dA, enhancing the initiation probability; (c) no sequence d(T-A-G) for the nonsense codon is to be found for a length of 22 base pairs from d(G-T-G) [42].

337

Assuming the codon sequence d(G-T-G) to thus be the initiating one, we deduced a hypothetical N-terminal 176 amino acid sequence for the β' subunit. This sequence extends without a termination signal up to the end of the *Eco*RI-[redacted]gment. The two other possible reading frames contain [redacted]derable numbers of terminating codons.

The intercistronic (4111–4186) region between the genes of the β and β' subunits contains 76 base pairs. This is significantly less than the intercistronic region between the *rplL* gene of L7/L12 protein and the *rpoB* gene of the RNA polymerase β subunit [17]. It is presently believed that in the *rplL-rpoB* intercistronic region there occurs an additional regulatory element, possibly an attenuator [15,45,46]. On the other hand, the intercistronic region between the *rpoA* gene coding for the α subunit and the nearest ribosomal gene contains only 25 base pairs [17] making it unlikely that there are transcription regulatory elements in this region.

On analyzing the sequence for the *rpoB-rpoC* intercistronic area the region 4130–4142 and 4145–4157 was found to have a pseudo-symmetric axis. There is at present no indication of the functional properties of the *rpoB-rpoC* intercistronic region although its size suggests the existence of some regulatory translational or transcriptional element.

Peculiarities of the Nucleotide Sequence Determined

The structure we have determined contains two segments wherein the tetranucleotide d(T-C-G-A), usually recognized and cleaved by restriction endonuclease *Taq*I, does not undergo cleavage (or does so very slowly). The coordinates of the segments are 935–938 and 3326–3329. In both cases the sequences containing this tetranucleotide are identical: d(G-A-T-C-G-A). It can be seen that the site recognizing *Taq*I in this structure overlaps the recognition site of another restriction endonuclease, *Sau*3AI: d(G-A-T-C) which here readily cleaves its specific tetranucleotide. We do not know why *Taq*I does not attack this sequence.

According to [16], in the 639–1143 base pairs region there is a segment that binds firmly to *E. coli* RNA polymerase. We have reported earlier on the structure of this segment and have discussed the possibility of its complexing with RNA polymerase [47]. While our work was in progress, a paper appeared by Burgess [48] which also describes studies of the binding by RNA polymerase to various restriction fragments generated in the hydrolysis of phage λ *rif*^{d} 18 DNA. Contrary to [16], Burgess found no formation of RNA polymerase complexes with *Eco*RI-C fragments that were firmly retained by nitrocellulose filters. We have also investigated, by means of adsorption on nitrocellulose filters, the possibility of formation of strong RNA polymerase complexes with the *Eco*RI-C fragment and its subfragments generated by *Taq*I, *Hpa*II and *Sau*3AI. We could not reproduce the data of Collins [16]. Apparently if complexes with these fragments are formed, their stability is very low. Recent electron microscopic studies of the possibility of forming complexes of *Eco*RI-C with RNA polymerases, carried out by A. A. Alexandrov and D. I. Chernyi (Institute of Molecular Genetics, USSR Academy of Sciences), also did not reveal the formation of stable complexes, although small above-background binding of RNA polymerase was noted in two sites of this fragment, at 6% and 30% of the total length from one of its ends.

We are at present using the structural data for ascertaining the part played by the RNA polymerase β subunit in the transcription process.

The authors are grateful to Prof. Zachau and Prof. Müller-Hill (FRG) for the restriction endonucleases, and to Prof. R. B. Hesin and O. D. Danilevskaya (Institute of Molecular Genetics USSR Academy of Sciences) for the plasmid pOD 162. They also wish to acknowledge the able assistance of Z. I. Tsarkova and G. M. Arinushkina in the course of the work and the help of P. V. Kostetskii and G. V. Vasil'ev in the treatment of the mathematical data.

REFERENCES

1. Ovchinnikov, Yu. A., Sverdlov, E. D., Lipkin, V. M., Monastyrskaya, G. S., Chertov, O. Yu., Gubanov, V. V., Guryev, S. O., Modyanov, N. N., Grinkevich, V. A., Makarova, I. A., Marchenko, T. V. & Polovnikova, I. N. (1980) *Bioorg. Khim.* 6, 655–665.
2. Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Lipkin, V. M. & Sverdlov, E. D. (1980) *Bioorg. Khim.* 6, 1106–1109.
3. Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Lipkin, V. M. & Sverdlov, E. D. (1980) *Bioorg. Khim.* 6, 1423–1426.
4. Ovchinnikov, Yu. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Chertov, O. Yu., Modyanov, N. N., Grinkevich, V. A., Makarova, I. A., Marchenko, T. V., Polovnikova, I. N., Lipkin, V. M. & Sverdlov, E. D. (1980) *Dokl. Akad. Nauk SSSR*, 253, 994–999.
5. Ovchinnikov, Yu. A., Lipkin, V. M., Modyanov, N. N., Chertov, O. Yu. & Smirnov, Yu. V. (1977) *FEBS Lett.* 76, 108–111.
6. Burgess, R. R. (1976) in *RNA Polymerase* (Chamberlin, M. & Losick, R., eds) pp. 68–100, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
7. Marchenko, T. V., Modyanov, N. N., Lipkin, V. M. & Ovchinnikov, Yu. A. (1980) *Bioorg. Khim.* 6, 325–331.
8. Shine, J. & Dalgarno, L. (1974) *Proc. Natl Acad. Sci. USA*, 71, 1342–1346.
9. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA*, 72, 4734–4738.
10. Ovchinnikov, Yu. A., Guryev, S. O., Krayev, A. S., Monastyrskaya, G. S., Skryabin, K. G., Sverdlov, E. D., Zakharyev, B. M. & Bayev, A. A. (1979) *Gene*, 6, 235–249.
11. Schwarz, E., Scherer, G., Hobom, G. & Kössel, H. (1978) *Nature (Lond.)* 272, 410–414.
12. Kirschbaum, J. B. & Konrad, B. E. (1973) *J. Bacteriol.* 116, 517–526.
13. Mindlin, S. S., Ilyina, T. S., Gorlenko, Ch. M., Hachikyan, N. A. & Kovalev, Yu. N. (1976) *Genetika*, 12, 116–130.
14. Lindahl, L., Yamamoto, M. & Nomura, M. (1977) *J. Mol. Biol.* 109, 23–47.
15. Yamamoto, M. & Nomura, M. (1978) *Proc. Natl Acad. Sci. USA*, 75, 3891–3895.
16. Collins, J. (1979) *Mol. Gen. Genet.* 173, 217–220.
17. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. & Dennis, P. P. (1979) *Proc. Natl Acad. Sci. USA*, 76, 1697–1701.
18. Gurevich, A. I., Avakov, A. E. & Kolosov, M. N. (1979) *Bioorg. Khim.* 5, 1735–1739.
19. Bass, T. A., Danilevskaya, O. N., Mekhedov, S. L., Fedoseeva, V. B. & Gorlenko, Zh. M. (1979) *Mol. Gen. Genet.* 173, 101–107.
20. Sverdlov, E. D., Monastyrskaya, G. S. & Rostopshov, V. M. (1978) *Bioorg. Khim.* 4, 894–900.
21. McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* 110, 119–146.
22. Sümegi, J., Breedveld, D., Hossenlepp, P. & Chambon, P. (1977) *Biochem. Biophys. Res. Commun.* 76, 78–85.
23. Greene, P. J., Heyneker, H. L., Bolivar, F., Rodriquez, R. L., Betlach, M. C., Covarrubias, A. A., Backman, K., Russel, D., Tait, R. & Boyer, H. W. (1978) *Nucleic Acids Res.* 5, 2373–2380.
24. Bickle, T. A., Pirotta, V. & Imber, R. (1977) *Nucleic Acids Res.* 5, 2561–2572.
25. Richardson, C. C. (1971) in *Procedures in Nucleic Acids Research* (Cantoni, G. L. & Davies, D. R., eds) vol. 2, pp. 815–828, Harper and Row, New York.
26. Maxam, A. & Gilbert, W. (1977) *Proc. Natl Acad. Sci. USA*, 74, 560–564.

314 338

27. Szalay, A. A., Grohmann, K. & Sutcheimer, R. L. (1977) *Nucleic Acids Res. 4*, 1569–1578.
28. Sverdlov, E. D., Monastyrskaya, G. S., Chestukhin, A. V. & Budovsky, E. I. (1973) *FEBS Lett. 33*, 15–17.
29. Sverdlov, E. D. & Levitan, T. L. (1976) *Bioorg. Khim. 2*, 370–375.
30. Bu[illegible] K. & Riley, W. T. (1966) *Biochem. J. 98*, 70–77.
31. Sverdlov, E. D., Monastyrskaya, G. S. & Budovski, E. I. (1977) *Mol. Biol. 11*, 116–123.
32. Brown, D. M. (1978) *Nucleic Acids Res. 5*, 615–621.
33. Hirs, C., Moore, S. & Stein, W. H. (1956) *J. Biol. Chem. 219*, 623–642.
34. Lipkin, V. M., Marchenko, T. V., Khokhryakov, V. S., Polovnikova, I. N., Potapenko, N. A., Modyanov, N. N. & Ovchinnikov, Yu A. (1980) *Bioorg. Khim. 6*, 332–342.
35. Ovchinnikov, Yu. A., Kiryushkin, A. A., Egorov, S. A., Abdulaev, N. G., Kiselev, A. P. & Modyanov, N. N. (1972) *Biokhimiya. 37*, 451–460.
36. Vinogradova, E. I., Feigina, M. Yu., Aldanova, N. A., Lipkin, V. M., Smirnov, Yu. V., Potapenko, N. A., Abdulaev, N. G., Kiselev, A. P., Egorov, S. A. & Ovchinnikov, Yu. A. (1973) *Biokhimiya. 38*, 3–21.
37. Modyanov, N. N., Lipkin, V. M., Smirnov, Yu. V., Chertov, O. Yu & Potapenko, N. A. (1978) *Bioorg. Khim. 4*, 158–179.
38. Korobko, V. G. & Grachev, S. A. (1977) *Bioorg. Khim. 3*, 1419–1422.
39. Olimon, H., Tomizawa, J. & Maxam, A. M. (1978) *Nucleic Acids Res. 5*, 1479–1485.
40. Lipkin, V. M., Modyanov, N. N., Marchenko, T. V., Chertov, O. Yu. & Ovchinnikov, Yu. A. (1980) in *Methods in Peptide and Protein Sequence Analysis* (Birr, Chr., ed.) pp. 453–459, Elsevier North-Holland, Amsterdam.
41. Manderschied, U., Bertram, S. & Gassen, H. G. (1980) *FEBS Lett. 90*, 162–166.
42. Atkins, J. F. (1979) *Nucleic Acids Res. 7*, 1035–1041.
43. Fujiki, H. & Zurek, G. (1975) *FEBS Lett. 55*, 242–244.
44. Granthum, R., Gautier, C., Gony, M., Mercier, R. & Pave, A. (1980) *Nucleic Acids Res. 8*, r49–r62.
45. Linn, T. & Scaife, J. (1978) *Nature (Lond.) 276*, 33–37.
46. Dennis, P. P. & Fill, N. P. (1979) *J. Biol. Chem. 254*, 7540–7547.
47. Sverdlov, E. D., Lipkin, V. M., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O. & Chertov, O. Yu. (1980) *Bioorg. Khim. 6*, 309–312.
48. Taylor, W. E. & Burgess, R. R. (1979) *Gene. 6*, 331–365.

Y. A. Ovchinnikov, G. S. Monastyrskaya, V. V. Gubanov, S. O. Guryev, O. Y. Chertov, N. N. Modyanov, V. A. Grinkevich, I. A. Makarova, T. V. Marchenko, I. N. Polovnikova, V. M. Lipkin, and E. D. Sverdlov,
Institut Bioorganicheskoj Khimii imeni M. M. Shemyakina, Akademiya Nauk S.S.S.R., Vavilova ulitsa 32, Moskva, U.S.S.R. 117312

Note Added in Proof. Recently we have established the primary structure of the *Eco*RI-C fragment of the wild type *rpoB* gene. The only difference which we were able to detect is the transversion of dT·dA (1628, Fig 3) in the *rif*^{s} mutant to dA·dT in the wild type gene. This change causes replacement of Val (516) in the mutant β subunit of the RNA polymerase to Asp in the wild protein.

339

Exhibit No. 3

Gene, 44 (1986) 201–209
Elsevier

GENE 1646

Monkey erythropoietin gene: cloning, expression and comparison with the human erythropoietin gene

(Recombinant DNA; oligodeoxyribonucleotide probe; kidney cDNA library; nucleotide sequence; transfection; signal peptide)

Fu-Kuen Lin*, Chi-Hwei Lin, Por-Hsiung Lai, Jeffrey K. Browne, Joan C. Egrie, Ralph Snailing, Gary M. Fox, Kenneth K. Chen, Miguel Castro and Sidney Suggs

Amgen, 1900 Oak Terrace Lane, Thousand Oaks, CA 91320 (U.S.A.) Tel. (805)499-5725

(Received January 6th, 1986)
(Revision received and accepted April 25th, 1986)

SUMMARY

The erythropoietin (Epo) gene from Cynomolgus monkeys has been isolated from a kidney cDNA library using mixed 20-mer oligodeoxynucleotide probes. The gene encodes a 168 amino acid (aa) mature protein with a calculated M_r of 18490 and a presumptive signal peptide of 24 aa. The Epo gene, when transfected into Chinese hamster ovary (CHO) cells, produces a glycosylated protein with an apparent M_r of 34000. The expressed product is biologically active in vivo. The monkey gene exhibits 92% and 94% homology to the human gene at the aa and nucleotide sequence levels, respectively. When compared with the human Epo, monkey Epo has an additional 3-aa residue at the N terminus of the mature protein and a deletion of an internal lysine residue.

INTRODUCTION

Epo is the principal hormone involved in the regulation of erythrocyte differentiation and the maintenance of a physiological level of circulating erythrocyte mass (Goldwasser, 1975). The hormone purified from urine of patients with aplastic anemia (Miyake et al., 1977) is a glycoprotein with an apparent M_r of 34000 (Wang et al., 1985). It is produced primarily by the kidney in the adult and by the liver during fetal life (Jacobson et al., 1957; Fried, 1972; Zanjani et al., 1981). Production of Epo is stimulated under conditions of hypoxia (Erslev, 1955). Epo has been proposed to exert its biological effect by attaching to specific binding sites on erythroid progenitor cells to stimulate their differentiation into mature red blood cells (Krantz and Goldwasser, 1984).

In the course of cloning the human Epo gene, due to the lack of a human source of mRNA, we adopted a heterologous approach to first isolate the gene from a monkey cDNA library. Based on the limited aa sequence information obtained for human Epo, a

* To whom correspondence and reprint requests should be addressed.

Abbreviations: aa, amino acid(s); bp, base pair(s); CHO, Chinese hamster ovary; Epo, erythropoietin; EpV, mixed oligo probe specific for Epo; kb, kilobase(s) or 1000 bp; LM-Ap, 1% (w/v) Bacto tryptone/0.5% (w/v) yeast extract – 10 mM NaCl – 10 mM $MgSO_4$ 1.5% (w/v) Bacto agar containing 50 μg ampicillin per ml; nt, nucleotide(s); oligo, oligodeoxynucleotide; PA, polyacrylamide; RIA, radioimmunoassay; SV40, simian virus 40; SDS, sodium dodecyl sulfate; U, unit.

341 220

mixture of oligos was designed which included all possible coding sequences for a short region of the protein. The mixture of oligos, 128 sequences each 20 nt long, was used as a hybridization probe to identify and isolate the cloned cDNA for monkey Epo. The methodologies for the isolation of specific genes using mixed oligo probes was first developed and used successfully by Wallace et al. (1981) and Suggs et al. (1981). This approach has subsequently been used to isolate many cloned genes (for a review, see Itakura et al., 1984).

The gene for human Epo has recently been isolated directly from a genomic DNA library (Lin et al., 1985) and from a fetal liver cDNA library (Jacobs et al., 1985). In the present report, the detailed process of isolating a monkey Epo cDNA clone using a large mixture of short oligo probes is described and the human and monkey Epo nt and aa sequences are compared. A preliminary report of this research has been published (Lin et al., 1984).

MATERIALS AND METHODS

(a) Design of oligo probe

Purified human urinary Epo (Miyake et al., 1977) was subjected to tryptic digestion. One of the tryptic fragments, Val-Asn-Phe-Tyr-Ala-Trp-Lys, was selected for probe synthesis as described (Lin et al., 1985). The probe, EpV, contains a mixture of 128 oligo 20-mers corresponding to all possible codon combinations:

```
3' CAA TTG AAG ATG CGA ACC TT 5'
    T   A   A   A   T
    G                G
    C                C
```

The probe mixture was labeled with [γ-^{32}P]ATP, as described (Lin et al., 1985).

(b) Induction of anemia in monkeys

Female Cynomolgus monkeys *Macaca fascicularis* (2.5–3 kg, 1.5–2 years old) were made anemic by subcutaneous injection of phenylhydrazine hydrochloride (pH 7.0) at 12.5 mg/kg on days 1, 3 and 5. On day 7, or whenever the hematocrit level fell below 25% of the initial level, serum and kidneys were harvested after administration of 25 mg/kg doses of ketamine hydrochloride. Harvested materials were immediately frozen in liquid nitrogen and stored at −70°C.

(c) Poly(A)$^+$mRNA isolation and Northern blotting

Total RNA was prepared from normal and phenylhydrazine-treated monkey kidneys by the guanidinium thiocyanate procedure of Chirgwin et al. (1979) and poly(A)$^+$mRNA was isolated by binding to an oligo(dT)-cellulose column (Aviv and Leder, 1972). 30 μg of poly(A)$^+$mRNA was separated by electrophoresis through a denaturing 1.5% agarose gel containing 2.2 M formaldehyde and transferred to nitrocellulose filters (Maniatis et al., 1982). The Northern blot was hybridized with ^{32}P-labeled EpV probe at 48°C. Post-hybridization wash was as described previously (Lin et al., 1985).

(d) Synthesis of double-stranded cDNA

A cDNA library was constructed from poly(A)$^+$mRNA of anemic monkeys using modifications of the general procedures of Okayama and Berg (1982). The key features of the present procedures are as follows: pUC8 vector was cut with *Pst*I and then tailed with oligo(dT), followed by *Hin*cII digestion. After first-strand synthesis, oligo(dG) tailing and *Bam*HI digestion, a synthetic linker (CGGTCTTTA and GATCTAAAGACCGTCC-CCCCCCC) in three-fold molar excess over vector was added before the replacement synthesis of the RNA strand by DNA.

(e) Bacterial transformation

Transformation of DNA into *E. coli* strain DH1 was performed and transformants were selected on LM-Ap agar as described by Hanahan (1983). Transformants were obtained at a level of 1.5×10^5 per μg of poly(A)$^+$RNA.

(f) Colony hybridization procedures

The in situ colony screening procedure employed modifications of the procedures described by

342 ~~331~~

Hanahan and Meselson (1980) and Taub and Thom[redacted] (1982). The screening procedures are detail[redacted]low.

The colonies on the LM-Ap plates were transferred to GeneScreen filters (New England Nuclear) and the plasmids were amplified on brain heart infusion plates containing 500 μg chloramphenicol per ml. The filters (colony side up) were treated serially with (a) 50 mM glucose, 25 mM Tris · HCl, 10 mM EDTA (pH 8.0) for 5 min; (b) 0.5 N NaOH for 10 min; and (c) 1.0 M Tris · HCl (pH 7.5) for 3 min. The filters were dried in a vacuum oven and subjected to proteinase K digestion as described (Lin et al., 1985).

The pre-hybridization, hybridization (at 48°C with the EpV mixture at 0.025 nM of each sequence) and post-hybridization wash were as previously described (Lin et al., 1985).

(g) Expression of the Epo gene

An approx. 1600-bp *Eco*RI-*Hin*dIII fragment containing the entire monkey Epo gene from clone pMkE83 (see RESULTS, section e) was inserted into an expression vector containing the SV40 late promoter (Lin et al., 1985). The resulting plasmid, pDSVL-MkE, was then used to transfect CHO cells and the transformants were selected as described previously (Lin et al., 1985).

(h) Epo assays

The methods for RIA and for determining in vivo biological activity of the expressed product were those previously described (Lin et al., 1985). The International Reference Preparation No. 2 was used as a calibration standard for Epo units (Cotes and Bangham, 1966).

(i) Western blotting procedures

The Epo present in serum of control or phenylhydrazine-treated monkeys was concentrated by immunoprecipitation using a rabbit anti-Epo serum. The immunoprecipitates were collected with protein-A Sepharose (Kessler, 1981), subjected to SDS–12.5% PA gel electrophoresis (Laemmli, 1970), and the resolved protein mixture was then transferred to nitrocellulose (Burnette, 1981). The nitrocellulose filter was probed with mouse anti-Epo monoclonal antibody (Egrie, 1985) using the Vectastain ABC kit (Vector Labs, CA).

RESULTS AND DISCUSSION

(a) Quantitation and characterization of Epo in normal and anemic monkey sera

Under normal physiological conditions, Epo mRNA is present in the kidney in very low amounts. To isolate the monkey Epo cDNA clone, it was necessary to increase the abundance of Epo-specific mRNA by inducing anemia in animals with a hemolytic agent, phenylhydrazine (Buetler, 1978).

To verify that the condition of anemia resulted in increased circulating Epo levels, sera from normal and phenylhydrazine-treated monkeys were tested for Epo by a RIA. The RIA was based on competition between serum Epo and ^{125}I-labeled, pure human urinary Epo (Miyake et al., 1977) for binding to antiserum against human Epo (Lin et al., 1985). Normal serum levels were measured to contain 36 mU/ml, while anemic monkey serum contained 1700 mU/ml. This represented an approximately 50-fold increase in Epo level under conditions of induced anemia.

Epo in the monkey sera was characterized by Western blot analysis using an anti-Epo monoclonal antibody as a probe, as described in MATERIALS AND METHODS, section I. When equal aliquots (2 ml) of normal and anemic monkey sera were subjected to this analysis, the monoclonal antibody identified a specific band with an apparent M_r of approx. 34 000 in the anemic but not in the control serum. The monkey serum Epo migrated identically to the human urinary Epo standard (not shown). The fact that the Epo band could be visualized from the anemic but not the control serum samples agrees with the increased Epo titer previously measured by the RIA.

(b) Northern analysis of mRNA

Polyadenylated RNA isolated from normal and phenylhydrazine-induced anemic monkey kidneys was subjected to RNA blotting analysis. There was

343 ~~322~~

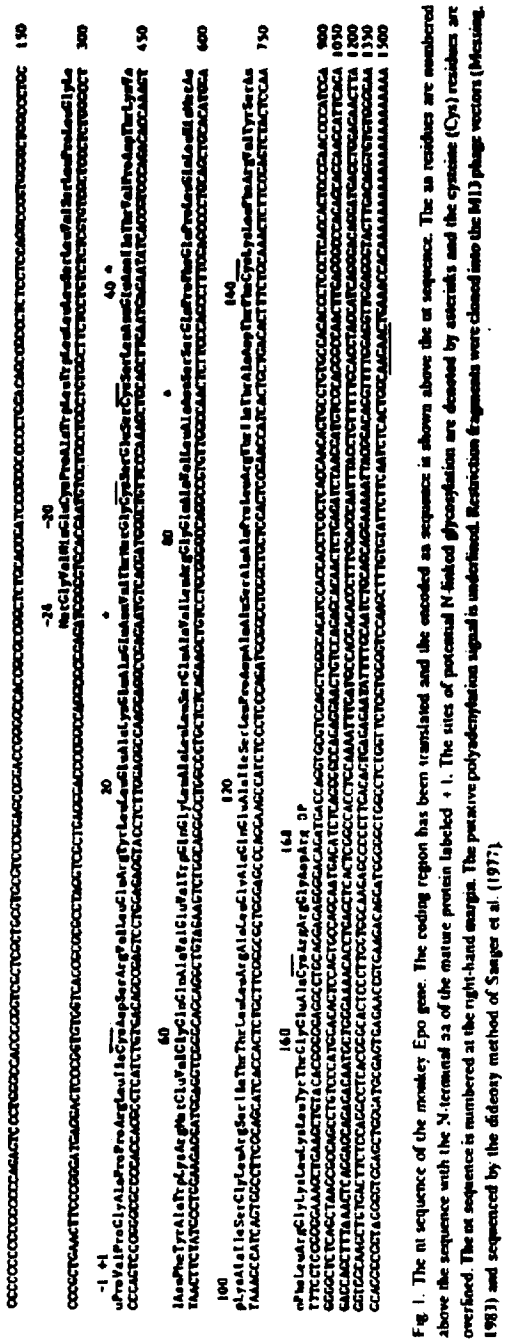


Fig 1. The nt sequence of the monkey Epo gene. The coding region has been translated and the encoded aa sequence is shown above the nt sequence. The aa residues are numbered above the sequence with the N-terminal aa of the mature protein labeled +1. The sites of potential N-linked glycosylation are denoted by asterisks and the cysteine (Cys) residues are overlined. The nt sequence is numbered at the right-hand margin. The putative polyadenylation signal is underlined. Restriction fragments were cloned into the M13 phage vectors (Messing, 1983) and sequenced by the dideoxy method of Sanger et al. (1977).

344 325

no detectable hybridization of the EpV probe to the RNA from normal monkey kidneys; however, there was a single band of hybridization observed with the RNA from the kidneys of phenylhydrazine-treated monkeys. The size of the mRNA which hybridized was approx. 1600 nt in length (not shown). This band was later shown to correspond to Epo mRNA by its hybridization to an Epo cDNA clone. This result indicates an anemia-induced increase in Epo mRNA levels, which is in agreement with the elevated level of Epo in anemic serum described above.

(c) Isolation of the monkey Epo gene

Using poly(A)$^+$ RNA isolated from phenylhydrazine-induced anemic monkey kidneys, a cDNA library was constructed as described in MATERIALS AND METHODS, section d. The cDNA library was screened by in situ colony hybridization using the EpV probe. Out of a total of 200 000 colonies screened, seven positive clones were obtained. Subsequent Southern blot and DNA sequence analysis confirmed that all seven isolates were authentic Epo clones (not shown). This result shows that even a complex mixture of oligonucleotides such as the one used here can detect specific cloned sequences in a cDNA library without obtaining false positives. One of the positive clones, pMkE83, was chosen for further analysis.

(d) Expression of the Epo gene

A conditioned medium (3 days) from CHO cells, which had been stably transformed with an expression vector containing the monkey Epo gene, pDSVL-MkE, contained 42.0 ± 5.0 units of Epo per ml when measured by the in vivo assay. Control medium from CHO cells transfected with the expression vector alone contained no detectable Epo (< 50 mU). When analyzed by SDS–PA gel electrophoresis, the secreted Epo has an apparent M_r of 34 000 (not shown), which is similar in size to the Epo present in phenylhydrazine-treated monkey serum. These results show that the cloned cDNA present in pDSVL-MkE encodes the production of biologically active Epo and that the recombinant Epo is glycosylated to an extent similar to Epo found in the circulation.

(e) Characterization of the monkey Epo gene

The nt sequence of the monkey Epo gene from clone pMkE83 is shown in Fig. 1. The size of the Epo gene insert in pMkE83 is 1462 bp without the poly(A) and in vitro added oligo(dC) tails. This size is in close agreement with the 1600 bp for the poly(A) containing Epo mRNA as estimated from the Northern blot, indicating that the cDNA clone is full length or almost full length.

The monkey Epo gene encodes a protein of 192 aa. Based on the N-terminal aa sequence of Epo expressed in CHO cells, position + 1 corresponds to the sequence of the N-terminus (not shown). The last 168 aa corresponds to the mature protein with a calculated M_r of 18 490 in an unglycosylated form. The first 24 aa of the coding region are predominantly hydrophobic residues, which is consistent with this region encoding a signal sequence (Watson, 1984). The mature protein has three potential sites for N-linked glycosylation at residues 27, 41 and 86, according to the rule of Asn-X-Ser/Thr (Marshall, 1974) and five cysteine residues at positions 10, 32, 36, 141 and 163.

The nt sequence of the monkey Epo gene has several unusual features. (1) The 5'-untranslated region is longer than normal. The length of the 5'-untranslated region in the Epo gene is at least 218 nt. In the survey of 211 eukaryotic mRNAs compiled by Kozak (1984), only six mRNAs have 5' leader sequences longer than 200 nt. The exact function of the unusually long 5' leader region in Epo mRNA is not known; however, it has been proposed (Kozak, 1984) that long leader regions may be involved in the regulation of expression of inducible genes. (2) The translation of the Epo gene initiates at the second AUG in the mRNA. In 95% of the mRNAs tabulated by Kozak (1984), the 5'-proximal AUG is used to initiate protein synthesis. As shown in Fig. 1, in the monkey Epo gene as well as in the human Epo gene, the second ATG from the 5' end (position 233) must be used to initiate translation because this ATG is the only one upstream from the coding region of the mature protein that is in the proper reading frame. The presence of in-phase termination codons at positions 156 and 200 rules out the possibility that translation of the Epo mRNA could have initiated upstream from these locations. (3) The normal polyadenylation signal sequence is

345 324

not present in the Epo mRNA. Usually the consensus signal sequence AATAAA is located 10–30 nt upstream from the site at which poly(A) is added (Nevins, 1983). In the case of both monkey and human Epo genes, the only sequence resembling AATAAA near the site of poly(A) addition is the sequence AAGAAC which is located 10 nt upstream from this site. Further experiments are necessary to determine if this sequence is involved in the polyadenylation process.

(f) Comparison of monkey and human aa sequences

The aa sequence of monkey Epo is shown in Fig. 1. Human Epo differs at only 16 positions: −8(Leu), 2(Leu), 28(Ile), 30(Thr), 33(Ala), 35(His), 83(Leu), 85(Val), 91(Trp), 98(Val), 102(Val), 108(Leu), 123(Pro), 141(Arg) and 165(Thr) and there is an additional lysine between aa 118 and 119. This represents a 92% sequence homology between the two species. Other significant features of the comparison are: (1) the N-terminus of mature monkey Epo contains an additional three aa, Val-Pro-Gly; (2) due to the change in the N-terminal sequence of the mature protein, there are 24 aa in the monkey putative signal peptide instead of the 27 aa in the human one; (3) the monkey Epo is missing aa 116, lysine, of the human, which corresponds to the first aa of the last exon of the human Epo gene (see Lin et al., 1985; Fig. 3); (4) monkey Epo contains five Cys as compared to four Cys in the human, with the additional Cys at position 141; (5) three potential N-linked glycosylation sites are conserved in both species.

Two features arising from the aa sequence comparison are particularly intriguing. First, monkey Epo contains one aa residue less which corresponds to lysine 116 in the human sequence. Since lysine 116 is the first aa of an exon (Lin et al., 1985; Jacobs et al., 1985), the deletion of this aa in the monkey sequence may be due to an alteration in splicing of the mRNA. Second, there is a difference in the N-terminal residues of the two species. Upon expression in CHO cells, the Epo molecules from the two species have different N-termini, indicating that the site of cleavage by the signal peptide processing enzyme is different. The alteration in the cleavage of the signal peptide may be due to changes near the N-terminus of the protein. Of the first 50 aa residues of the Epo precursor, there are only two aa at which the sequences for the two species diverge: (a) Val (monkey aa −8) vs. Leu (human aa −11) and (b) Pro (monkey aa 2) vs. Leu (human aa −2). Either or both of these two changes may be responsible for the difference in processing of the N-terminus as a result of a change in conformation or primary sequence.

(g) Comparison of nt sequence between the monkey and human Epo gene

At the nt level, monkey and human Epo are 94% homologous, as shown in Fig. 2. In the protein-coding region, besides the absence of the lysine codon corresponding to position 116 in the human Epo sequence, out of a total of 15 aa differences, 4 of these differences involve 2 nt substitutions and 11 differences involve single nt substitutions. Furthermore, there are 11 silent nt substitutions at the third position of codons. As compared to the human gene, in the 5′-untranslated region, the monkey gene contains nt deletions at three locations. In the 3′-untranslated region, the monkey gene contains three nt insertions and one nt deletion relative to the human sequence. Also in the 3′-untranslated region, a 12-bp sequence, CCTCCCTCACCA, is exactly repeated in the monkey but is imperfectly repeated in man. Both genes have the same putative polyadenylation signal sequence AAGAAC.

(h) Conclusions

The Epo gene was isolated from monkey kidneys enriched for Epo-specific mRNA by induced anemia. The cloned cDNA directed the synthesis of Epo that is biologically active in vivo.

To identify cloned cDNA for monkey Epo, a mixture of 128 oligo 20-mers was used as a hybridization probe. The sequences in the mixture were designed from the region of aa sequence of human urinary Epo corresponding to residues 46–52. The success of this approach was dependent on absolute homology between the monkey and human proteins in the region from which the probe was designed. Attempts to isolate the monkey cDNA clone using a second mixed probe corresponding to aa 86–91 of human Epo were unsuccessful (unpublished data) due to divergence at a single aa residue (Trp 88 in

~~345~~ 346

Fig. 2 Comparison of the nt sequences of the monkey and human Epo genes. The nt sequence of the monkey gene is shown with the coding region divided into codons. Above the monkey sequence are indicated only those nt of human Epo sequence which differ from the monkey Epo sequence. A dash indicates the nt deletion. The direct repeat mentioned in RESULTS AND DISCUSSION, section g is underlined

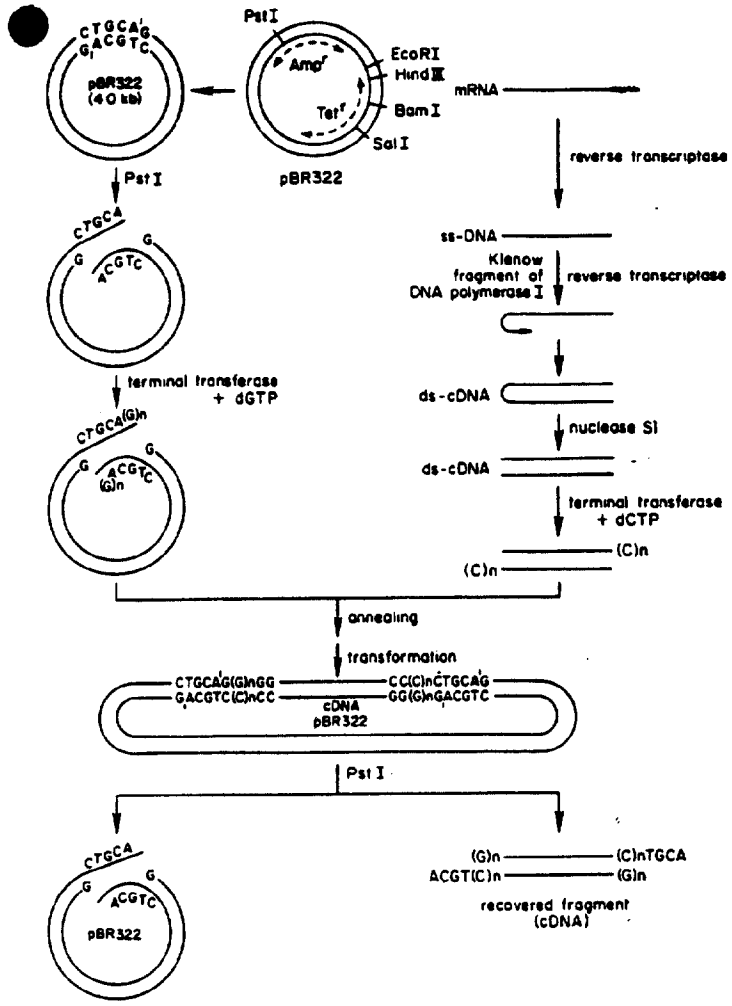


Figure 7.1

The conditions first used to achieve full-length, second-strand cDNA synthesis by DNA polymerase I (Efstratiadis et al. 1976) are still widely used (Wickens et al. 1978). In brief, the reaction is carried out at pH 6.9 to minimize the 5′ → 3′ exonuclease activity of DNA polymerase I and at 15°C to minimize the possibility of synthesizing "snapback" DNA. The Klenow fragment of DNA polymerase I, which lacks the 5′ → 3′ exonuclease activity, has also been successfully employed to synthesize the second cDNA strand.

~~329~~ 348

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of:	)	"PRODUCTION OF ERYTHROPOIETIN"
FU-KUEN LIN	)	
Serial No. 675,298	)	Group Art Unit 127
Filed: November 30, 1984	)	Examiner: J. M. Giesser

Declaration Under 37 CFR 1.132

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

I, Fu-Kuen Lin, declare as follows:

1. THAT I am the sole inventor of the subject matter described and claimed in the United States Patent Application Serial No. 678,298 entitled "Production of Erythropoietin".

2. THAT, to the best of my knowledge, Dr. Sylvia Lee-Huang (hereinafter Dr. Lee-Huang), Department of Biochemistry, New York University School of Medicine, New York, New York 10016, is the author of the Lee-Huang reference (Proc. Natl. Acad. Sci. USA, Vol. 81, pp. 2708-2712, May, 1984) cited by the Examiner in the above-identified patent application.

3. THAT on August 5, 1986, I contacted Dr. Lee-Huang by telephone and requested a sample of monoclonal antibody 7A7 from Dr. Lee Huang in order to determine whether or not the monoclonal antibody asserted to exist in the Lee-Huang reference recognizes epitope sites of the erythropoietin molecule;

4. THAT Dr. Lee-Huang stated that she would consider my request and would give me her reply on August 18, 1986;

5. THAT on August 19, 1986, I again contacted Dr. Lee-Huang by telephone and repeated my earlier request and then Dr. Lee-Huang asked me if the samples would be utilized for my personal use or in conjunction with my work at Amgen, my current employer;

6. THAT I told Dr. Lee-Huang that all my research is conducted on behalf of Amgen and at that time Dr. Lee-Huang stated that she would not provide any samples to me;

349 ~~329~~

357

7. THAT I do not know of any other source for the monoclonal antibody of the Lee-Huang reference;

8. THAT all statements herein of my own knowledge are true and that all statements made on information and belief are believed to be true and further; that these statements were made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: 8/29/86

Fu-Kuen Lin

Fu-Kuen Lin

Lin-1(08.29.86)

350

330

Exhibit No. 5

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of:	)	"PRODUCTION OF ERYTHROPOIETIN"
FU-KUEN LIN	)	
Serial No. 675,298	)	Group Art Unit 127
Filed: November 30, 1984	)	Examiner: J.M. Giesser

DECLARATION PURSUANT TO 37 C.F.R. SECTION 1.132

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

I, Fu-Kuen Lin, solemnly declare:

1. That since August 6, 1981 I have been employed by Amgen, Thousand Oaks, California and currently hold the position of Research Scientist;

2. That from May, 1979 to July, 1981 I held the position of Postdoctoral Research Associate at the Medical University of South Carolina, Charleston, S.C.;

3. That from October, 1977 to May, 1979 I held the position of Postdoctoral Research Associate at Louisiana State University, Baton Rouge, Louisiana;

4. That from January, 1977 to to June, 1977 I held the position of Associate Professor at the National Taiwan University, Taiwan, China;

5. That from October, 1975 to September, 1977 I held the position of Associate Research Fellow at the Institute of Botany Academia Sinica, Taipei, Taiwan, China;

6. That from July, 1976 to October, 1976 I held the position of Visiting Scientist at the Tumor Biology Lab, University of Nebraska, Lincoln, Nebraska;

7. That from January, 1974 to September, 1975 I held the position of Postdoctoral Research Associate at the Tumor Biology Lab at the University of Nebraska, Lincoln, Nebraska;

352

360

8. That from October, 1971 to December, 1973 I held the position of Postdoctoral Research Associate at the Department of Biological Sciences, Purdue University, West Lafayette, Indiana;

9. That I obtained a Ph.D. from the University of Illinois, Urbana, Illinois in 1971 in Plant Pathology/Fungi Physiology, a M.S. from the National Taiwan University, Taiwan, China in 1967 and a B.S. from the National Taiwan University, Taiwan, China in 1964;

10. That I am the sole inventor of the subject matter described and claimed in the United States Patent Application Serial No. 675,298, entitled "Production of Erythropoietin";

11. That I am a coauthor of an abstract entitled "Cloning of the Monkey Erythropoietin Gene" of an abstract/published in the Journal of Cell Biochemistry, Suppl. 8 B., page 45 (1984).

12. That Chi-Hwei Lin, also a coauthor of the publication referred to in Paragraph 11, was a research technician working under my direction and supervision to assist in cloning erythropoietin and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

13. That Michael Castro, also a coauthor of the publication referred to in Paragraph 11, was a research technician who prepared the oligonucleotide probes and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

14. That Por H. Lai, also a coauthor of the publication referred to in Paragraph 11, was a research scientist who assisted in sequencing fragments of the erythropoietin protein, and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

15. That Joan Egrie, also a coauthor of the publication referred to in Paragraph 11, was a research scientist who provided analytical support and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

16. That Eugene Goldwasser, also a coauthor of the publication referred to in Paragraph 11, is a member of the Department of Biochemistry and Molecular Biology, The University of Chicago, is a

consultant to Amgen regarding erythropoietin and has provided Amgen with natural erythropoietin and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

17. That Fung-Fang Wang, also a coauthor of the publication referred to in paragraph 11, was a graduate student working under the direction and supervision of Eugene Goldwasser at the University of Chicago, and, that while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

18. That all statements herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further; that these statements were made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: 8/29/86

Fu-Kuen Lin

Fu-Kuen Lin

LinDec(08.20.86)

Exhibit No. 6

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of: FU-KUEN LIN	)	"PRODUCTION OF ERYTHROPOIETIN"
Serial No. 675,298	)	Group Art Unit 127
Filed: November 30, 1984	)	Examiner: J.M. Giesser

DECLARATION PURSUANT TO 37 C.F.R. SECTION 1.132

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

I, Fu-Kuen Lin, solemnly declare:

1. That since August 6, 1981 I have been employed by Amgen, Thousand Oaks, California and currently hold the position of Research Scientist;

2. That from May, 1979 to July, 1981 I held the position of Postdoctoral Research Associate at the Medical University of South Carolina, Charleston, S.C.;

3. That from October, 1977 to May, 1979 I held the position of Postdoctoral Research Associate at Louisiana State University, Baton Rouge, Louisiana;

4. That from January, 1977 to June, 1977 I held the position of Associate Professor at the National Taiwan University, Taiwan, China;

5. That from October, 1975 to September, 1977, I held the position of Associate Research Fellow at the Institute of Botany Academia Sinica, Taipei, Taiwan, China;

6. That from July, 1976 to October, 1976 I held the position of Visiting Scientist at the Tumor Biology Lab, University of Nebraska, Lincoln, Nebraska;

7. That from January, 1974 to September, 1975 I held the position of Postdoctoral Research Associate at the Tumor Biology Lab at the University of Nebraska, Lincoln, Nebraska;

8. That from October, 1971 to December, 1973 I held the position of Postdoctoral Research Associate at the Department of Biological Sciences, Purdue University, West Lafayette, Indiana;

9. That I obtained a Ph.D. from the University of Illinois, Urbana, Illinois in 1971 in Plant Pathology/Fungi Physiology, a M.S. from the National Taiwan University, Taiwan, China in 1967 and a B.S. from the National Taiwan University, Taiwan, China in 1964;

10. That I am the sole inventor of the subject matter described and claimed in the United States Patent Application Serial No. 675,298, entitled "Production of Erythropoietin";

11. That I am a coauthor of an abstract entitled "Cloning and Expression of Monkey Human Erythropoietin Gene" of an abstract published in the Exp. Hematol.12, 357 (1984).

12. That Por H. Lai, also a coauthor of the publication referred to in Paragraph 11, was a research scientist who assisted in sequencing fragments of the erythropoietin protein, and while co-authoring said publication, is not a coinventor of the subject matter claimed in said patent application.

13. That Jeffrey Browne, also a coauthor of the publication referred to in Paragraph 11, was a research scientist who expressed the erythropoietin clone in the SV40 promoter expression vector, and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

14. That Joan Egrie, also a coauthor of the publication referred to in Paragraph 11, was a research scientist who provided analytical support and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

15. That Chi-Hwei Lin, also a coauthor of the publication referred to in Paragraph 11, was a research technician working under my direction and supervision to assist in cloning erythropoietin and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

16. That Eugene Goldwasser, also a coauthor of the publication referred to in Paragraph 11, is a member of the Department of Biochemistry and Molecular Biology, The University of Chicago, is a consultant to Amgen regarding erythropoietin and has provided Amgen with

-1- 357

natural erythropoietin and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application;

17. That Fung-Fang Wang, also a coauthor of the publication referred to in paragraph 11, was a graduate student working under the direction and supervision of Eugene Goldwasser at the University of Chicago, and, that while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application;

18. That Ralph Smalling, also a coauthor of the publication referred to in Paragraph 11, is a research associate working under the direction and supervision of Jeffrey Browne and was involved in expressing erythropoietin clone in the SV40 promoter expression vector and while coauthoring said publication, is not a coinventor of the subject matter claimed in said patent application.

19. That all statements herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further; that these statements were made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: 8/29/86

Fu-Kuen Lin

HCP2(08.29.86)

-6-
358

Exhibit No. 7

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of: FU-KUEN LIN	)	"PRODUCTION OF ERYTHROPOIETIN"
Serial No. 675,298	)	Group Art Unit 127
Filed: November 30, 1984	)	Examiner: J.M. Giesser

DECLARATION PURSUANT TO 37 C.F.R. SECTION 1.132

Hon. Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

I, Fu-Kuen Lin, solemnly declare:

1. That since August 6, 1981 I have been employed by Amgen, Thousand Oaks, California and currently hold the position of Research Scientist;

2. That from May, 1979 to July, 1981 I held the position of Postdoctoral Research Associate at the Medical University of South Carolina, Charleston, S.C.;

3. That from October, 1977 to May, 1979 I held the position of Postdoctoral Research Associate at Louisiana State University, Baton Rouge, Louisiana;

4. That from January, 1977 to to June, 1977 I held the position of Associate Professor at the National Taiwan University, Taiwan, China;

5. That from October, 1975 to September, 1977 I held the position of Associate Research Fellow at the Institute of Botany Academia Sinica, Taipei, Taiwan, China;

6. That from July, 1976 to October, 1976 I held the position of Visiting Scientist at the Tumor Biology Lab, University of Nebraska, Lincoln, Nebraska;

7. That from January, 1974 to September, 1975 I held the position of Postdoctoral Research Associate at the Tumor Biology Lab at the University of Nebraska, Lincoln, Nebraska;

359

368

8. That from October, 1971 to December, 1973 I held the position of Postdoctoral Research Associate at the Department of Biological Sciences, Purdue University, West Lafayette, Indiana;

9. That I obtained a Ph.D. from the University of Illinois, Urbana, Illinois in 1971 in Plant Pathology/Fungi Physiology, a M.S. from the National Taiwan University, Taiwan, China in 1967 and a B.S. from the National Taiwan University, Taiwan China in 1964;

10. That I am the sole inventor of the subject matter described and claimed in the United States Patent Application Serial No. 675,298, entitled "Production of Erythropoietin";

11. That on page 49, lines 29-33 of the said application, there is described the preparation of cDNA sequences encoding human erythropoietin from messenger RNA isolated from COS-1 cells transformed with human genomic DNA encoding erythropoietin. An experimental procedure was carried out on the cloned cDNA by sequencing its bases for purposes of determining the primary structure of the polypeptide which the gene encoded and this work led to the confirmation of the presence of an "additional" lysine residue at human erythropoietin position 116 vis-a-vis monkey erythropoietin as reported in the application.

12. That additional work carried out on a human cDNA clone (designated pHu13) involved its "restriction mapping" to ascertain its susceptibility to digestion with restriction endonuclease enzymes. This mapping work was performed in two ways for purposes of cross-checking the accuracy of the data obtained. In one procedure, the DNA sequence information was subjected to computer-assisted analysis for restriction endonuclease recognition sites. In the companion procedure, the cDNA itself was subjected to multiple restriction endonuclease enzyme digestions to generate digested samples which were then electrophoretically separated and "sized" by comparison to digestion fragments of a known (standard) DNA sequence treated with a specific restriction enzyme.

13. That the restriction endonuclease digestion study referred to above was carried out in the following manner:

A human erythropoietin clone, pHu13, was digested to completion with a variety of endonucleases including

-1- 361

PstI in 30 µl of buffer comprising 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl. The digested samples were then electrophoretically separated in a 1.5% agarose gel (130V, 2.5 hours), stained with ethidium bromide (1 µg/ml) for ten minutes and then destained in distilled water for ten minutes. A photograph of the gel was taken under ultraviolet light (300 nm) and a copy of the photograph is attached hereto as Exhibit A. The PstI digest of the pHu13 cDNA insert is present in lane 5 of the gel (marked with an arrow). Simultaneously with the digestion of the pHu13 insert, molecular "markers" consisting of a HindIII digest of lambda phage DNA and a HaeIII digest of ϕ X174 DNA were electrophoretically separated and the fragments are present in lane 1 of the gel. Sizes of the marker DNA fragments are noted in the left margin of the gel photograph.

14. That analysis of the fragments present in lane 5 of the gel reveals that the human erythropoietin cDNA clone includes four PstI recognition sites, generating a total of four DNA fragments. These fragments were approximately 175 base pairs, 210 base pairs, 600 base pairs and 3400 base pairs in length, with the last-mentioned fragment comprising vector plasmid DNA and portions of DNA from the cDNA insert up through the first PstI site at either end of the insert.

15. That attached hereto as Exhibit B is a copy of a computer-generated search for various six-base-pair restriction enzyme recognition sites, including PstI, in the cDNA encoding erythropoietin. Four PstI sites are noted at positions 218, 801, 976 and 1185 of the insert. Upon insertion into a plasmid having no PstI sites of its own, this cDNA should generate a composite plasmid which, upon digestion with PstI should generate three small fragments having sizes of 584, 176 and 210 base pairs together with one large fragment comprising plasmid DNA.

16. That the size of the PstI digestion fragments obtained from the PstI digestion and electrophoresis described in paragraphs 13

and 14 and represented in Figure I in Exhibit A are similar to the size of the PstI fragments calculated using the restriction map of the cDNA sequence of EPO in Exhibit B.

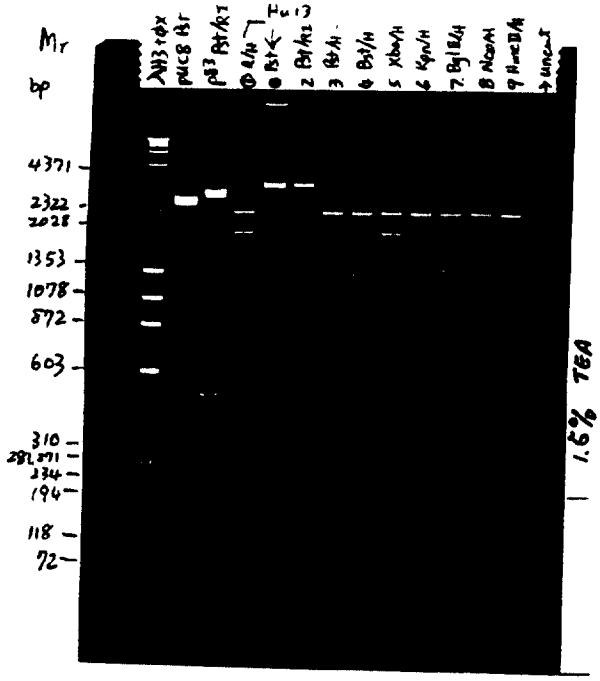
That all statements herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issuing thereon.

Date: 8/29/86

Fu-Kuen Lin

Lin(08.29.86)-

EXHIBIT A



384

FIG.6C

TABLE VI (cont'd.)

```
TGGTGGCCCCAAACCATACCTGAAACTAGGCAAGGAGCAAAGCCAGCAGATCCTACGCCTGTGGGCCAGGG

                                    27                  30
CCAGAGCCTTCAGGGACCCTTGACTCCCCGGGCTGTGTGCATTTCAG    Thr Gly Cys Ala Glu
                                                   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 GTGAGTCCTTTTTTTTTTTTTTTTCCTTTCTTTTGGAGAATCTCATT

TGCGAGCCCTGATTTTGGATGAAAGGGAGAATGATCGGGGGAAAGGTAAAATGGAGCCAGCAGAGATGAGGCT

GCCTGGGCGCAGAGGCTCACGTCTATAATCCCAGGCTGAGATGGCCGAGATGGGAGAATTGCTTGAGCCCT

GGAGTTTCAGACCAACCTAGGCAGCATAGTGAGATCCCCCATCTCTACAAACATTTAAAAAAATTAGTCAG

GTGAAGTGGTGCATGGTGGTAGTCCCAGATATTTGGAAGGCTGAGGCGGGAGGATCGCTTGAGCCCAGGAA

TTTGAGGCTGCAGTGAGCTGTGATCACACCACTGCACTCCAGCCTCAGTGACAGAGTGAGGCCCTGTCTCA
```

3.85

FIG.6D

TABLE VI (cont'd.)

```
AAAAAGAAAAGAAAAAAGAAAAATAATGAGGGCTGTATGGAATACATTCATTATTCATTCACTCACTCACT

CACTCATTCATTCATTCATTCATTCAACAAGTCTTATTGCATACCTTCTGTTTGCTCAGCTTGGTGCTTGG

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

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

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

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

GAAGGGTCTTGCTAAGGAGTACAGGAACTGTCCGTATTCCTTCCCTTTCTGTGGCACTGCAGCGACCTCCT

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

374

366

EXHIBIT NO. 9

Exhibit No. 9

FIG.6A

TABLE VI

```
AAGCTTCTGGGCTTCCAGACCCAGCTACTTTGCGGAACTCAGGCAACCCAGGCATCTCTGAGTCTCCGCCCA

AGACCGGGATGCCCCCCAGGGGAGGTGTCCGGGAGCCCCAGCCTTTCCCAGATAGCACGCTCCGCCAGTCCC

AAGGGTGCGCAACCGGCTGCACTCCCCTCCCGCGACCCAGGGCCCGGGAGCAGCCCCCATGACCCACACGC

ACGTCTGCAGCCAGCCCCGCTCACGCCCCGGCGAGCCTCAACCCAGGCGTCCTGCCCCTGCTCTGACCCCGG

GTGGCCCCTACCCCTGGCGACCCCTCACGCACACAGCCTCTCCCCCACCCCCACCCGCGCACGCACACATG

CAGATAACAGCCCCCGACCCCCGGCCAGAGCCGXAGAGTCCCTGGGCCACCCCGGCCCGCTCGCCTGCCGCTG

CGCCGCACCGCGCTGTCCTCCCGGAGCCGGACCGGGGCCACCGCGCGCCCXGCTCTGCTCCGACACCGCGCCC

CTTGGACAGCCGCCCTCTCCTCTAGGCCCGTGGGCTGGCCCTGCACCGCCGAGCTTCCCGGGATGAGGXX

                                -27          -24
CCCGGTGACCGGGCGCGCCCCAAGTCGCTGAGGGACCCCGGCCAAGCGCGGAG  ATG GGG GTG CAC G

GTGAGTACTCGCGGGCTGGGCGCTCCCGGCGCGGCCGGCTTCCTGTTTGAGCGGGGATTTAGCGCCCCGGCT
```

369

FIG.6B

TABLE VI (cont'd.)

```
ATTGGCCAAGAGGTGGCTGGGTTCAAGGACCGGCGACTTGTCAAGGACCCCCGGAAGGGGGGAGGGGGGGTGGG

GCAGCCTCCACGTGCCCGCGGGGACTTGGGGGAGTTCTTGGGGATGGCAAAAACCTGGCCTGTTGAGGGGCA

CAGTTTGGGGTTGGGGAGGAGGTTTGGGGTTCTGCTGTGCAGTTGTGTCGTTGTCAGTGTCTCG[I.S.]

TTGCACACGCACAGATCAATAAGCCAGAGGCAGCACCTGAGTGCTTGCATGGTTGGGACAGGAAGGACGAG

CTGGGGCAGAGACGTGGGGATGAAGGAAGCTGTCCTTCCACAGCCACCCTTCTCCCCCCCCGCCTGACTCT

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

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

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

  26
Thr
ACG GTGAGACCCCCTTCCCCAGCACATTCCACAGAACTCACGCTCAGGGCTTCAGGGAACTCCTCCCAGAT

CCAGGAACCTGGCACTTGGTTTGGGGTGGAGTTGGGAAGCTAGACACTGCCCCCCCTACATAAGAATAAGTC
```

378

Exhibit No. 10

FIG.6E

TABLE VI (cont'd.)

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

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

166
Asp Arg OP
GAC AGA TGA CCAGGTGTGTCCACCTGGGCATATCCACCACCTCCCTCACCAACATTGCTTGTGCCACA

CCCTCCCCCGCCACTCCTGAACCCCGTCGAGGGGCTCTCAGCTCAGCGCCAGCCTGTCCCATGGACACTCC

AGTGCCAGCAATGACATCTCAGGGGCCAGAGGAACTGTCCAGAGAGCAACTCTGAGATCTAAGGATGTCAC

AGGGCCAACTTGAAGGGCCCAGAGCAGGAAGCATTCAGAGAGCAGCTTTAAACTCAGGGACAGAGCCATGC

TGGGAAGACGCCTGAGCTCACTCGGCACCCTGCAAAATTTGATGCCAGCACGCTTTGGAGGCCATTTAC

CTGTTTTCGCACCTACCATCAGGGACAGGATGACCTGGAGAACTTAGGTGGCAAGCTGTGACTTCTCCAGG

TCTCACGGGCATGGGCACTCCCTTGGTGGCAAGAGCCCCCTTGACACCGGGGTGGTGGGAACCATGAAGAC

AXGATXGGGGCTGGCCTCTGGCTCTCATGGGGTCCAAGTTTTGTGTATTCTCAACCTATTGACAGACTGAA

ACACAATATGAC

392