US00RE38981E

(19) **United States**

(12) **Reissued Patent**

Foster et al.

(10) **Patent Number: US RE38,981 E**

(45) **Date of Reissued Patent: *Feb. 14, 2006**

(54) **DNA SEQUENCE CODING FOR PROTEIN C**

(75) Inventors: **Donald C. Foster**, Seattle, WA (US); **Earl W. Davie**, Bellevue, WA (US)

(73) Assignee: **Board of Regents of the University of Washington**, Seattle, WA (US)

(*) Notice: This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/217,105**

(22) Filed: **Aug. 13, 2002**

Related U.S. Patent Documents

Reissue of:

(64) Patent No.: **4,968,626**
Issued: **Nov. 6, 1990**
Appl. No.: **06/766,109**
Filed: **Aug. 15, 1985**

(63) Continuation of application No. 09/882,150, filed on Jun. 15, 2001, now Pat. No. Re. 37,958.

(51) **Int. Cl.**

C12N 15/00	(2006.01)
C12N 9/64	(2006.01)
C07H 15/12	(2006.01)

(52) **U.S. Cl.** **435/320.1**; 435/226; 435/69.1; 435/440; 435/252.33; 536/23.2

(58) **Field of Classification Search** 435/69.1, 435/226, 320.1, 440, 252.33; 536/23.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,775,624 A * 10/1988 Bang et al.
4,784,950 A * 11/1988 Hagen et al.

FOREIGN PATENT DOCUMENTS

EP 138222 * 4/1985
WO WO85/00521 * 2/1985

OTHER PUBLICATIONS

Griffin et al. (1981) J. Clin. Invest., 68:1370–1373.*
Stenflo et al. (1982), Journal of Biochemistry, 257: 12180–12190.
Ferlund et al. (1982), Journal of Biochemistry, 257: 12170–12179.
Beckman et al. (1985), Nucleic Acids Research, 13: 6233–6247.
Foster et al. (1984), PNAS USA, 81 : 4766–4770.
Foster et al. (1985), PNAS USA, 82: 4673–4677.
Degan et al. (1983), Biochemistry, 22: 2087–2092.
Long, G. et al. (1984), PNAS, 81: 5653–5656.
Kaufman et al.(1982), Molecular and Cell Biology, 2:1304–1319.
Kaufman (1985), PNAS USA, 82: 689–693.
Hermonat et al. (1984), PNAS USA, 81 : 6466–6740.
Esmon et al. (1981), PNAS USA, 78: 2249–2252.
Ginsburg et al. (1985), Science, 228: 1401–1406.
Katayama et al. (1979), PNAS USA, 76: 4990–4994.
Kisiel et al. (1977), Biochemistry, 16: 5824–5831.
Walker et al. (1979), Biochim. et Biophys. Acta, 571: 333–342.
McMullen et al. (1983), Biochim. et Biophys. Res. Comm., 115: 8–14.
Beckmann et al. (1985), Fed. Proc., 44: 1069.
Kisiel et al. (1981), Methods of Enzymology, 80:320–332.
Kisiel et al. (1979), Journal of Clinical Investigation, 64: 761–769.
Van Hinsbergh et al. (1985), Blood, 65: 444–451.
Kisiel et al. (1983), Behring Inst. Mitt., 73: 29–42.
Gardiner et al. (1983), Progress in Hematology, 265–278.
Comp et al. (1981), Journal of Clinical Investigation, 68: 1221–1228.
Sakata et al. (1985), PNAS USA, 82: 1121–1125.
Broekmans et al. (1983), New England Journal of Medicine, 309: 340–344.
Seligsohn et al. (1984), New England Journal of Medicine, 310: 559–562.
Marlar et al. (1982), Blood, 59: 1067–1072.

* cited by examiner

Primary Examiner—James Ketter
(74) *Attorney, Agent, or Firm*—Christensen O'Connor Johnson Kindness PLLC

(57) **ABSTRACT**

Genomic and cDNA sequences coding for a protein having substantially the same biological activity as human protein C are disclosed. Recombinant plasmids and bacteriophage transfer vectors incorporating these sequences are also disclosed.

9 Claims, 9 Drawing Sheets

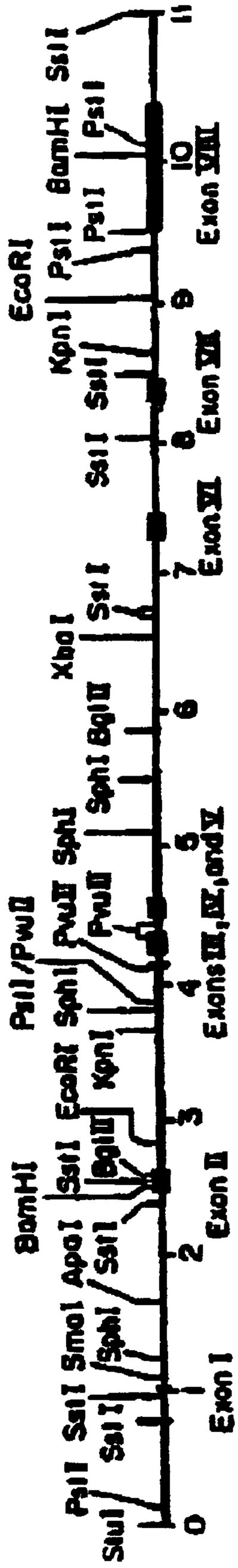
0
1
2
3
4
5
6
7
8
9
10
11
Exon I
Exons II
Exons III,IV,V
Exon VI
Exon VII
Exon VIII
BamHI
PstI/PwuI
SalI
SmaI
ApaI
SphI
EcoRI
KpnI
BglII
XbaI
StuI
PstI
PwuI

FIG. 1

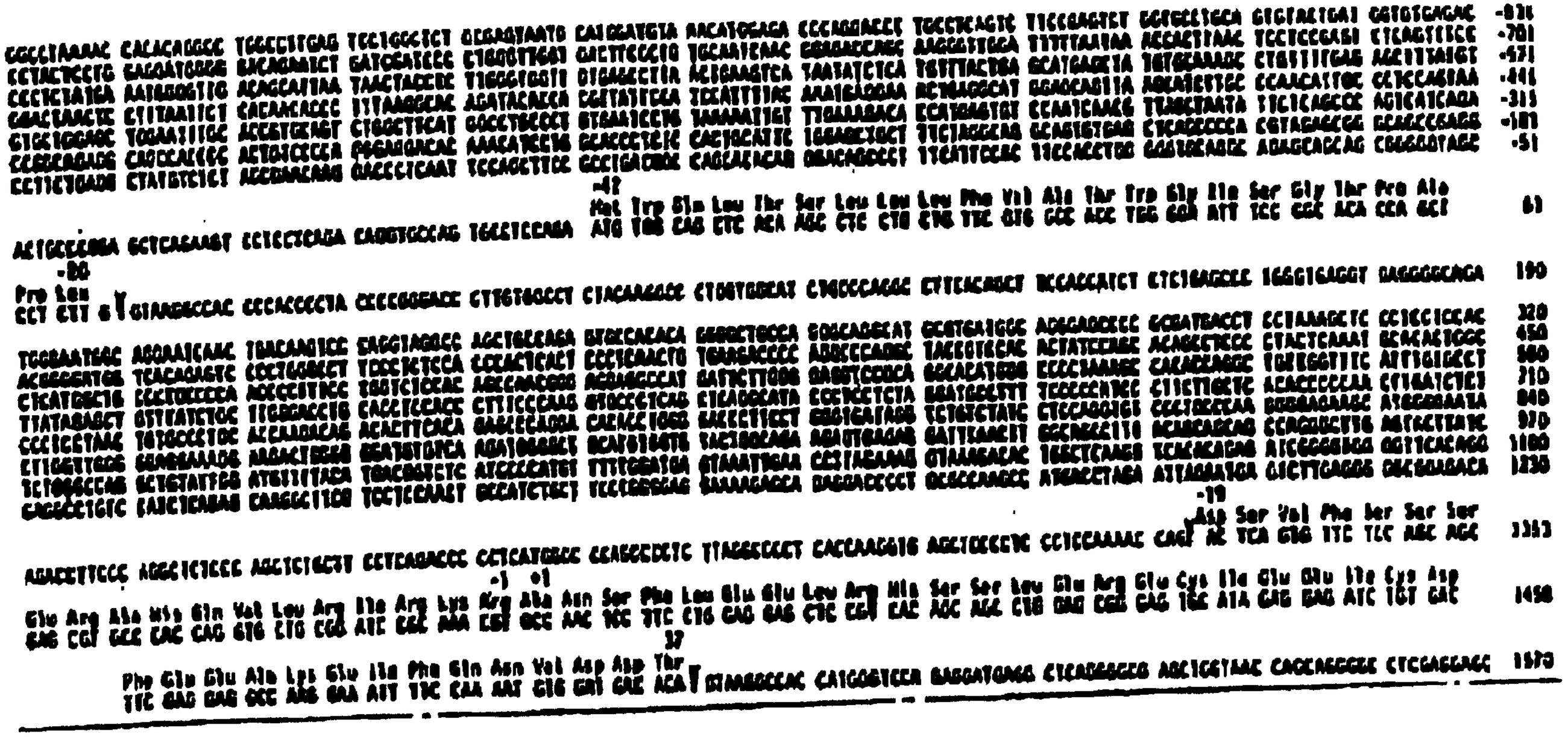

FIG. 2A

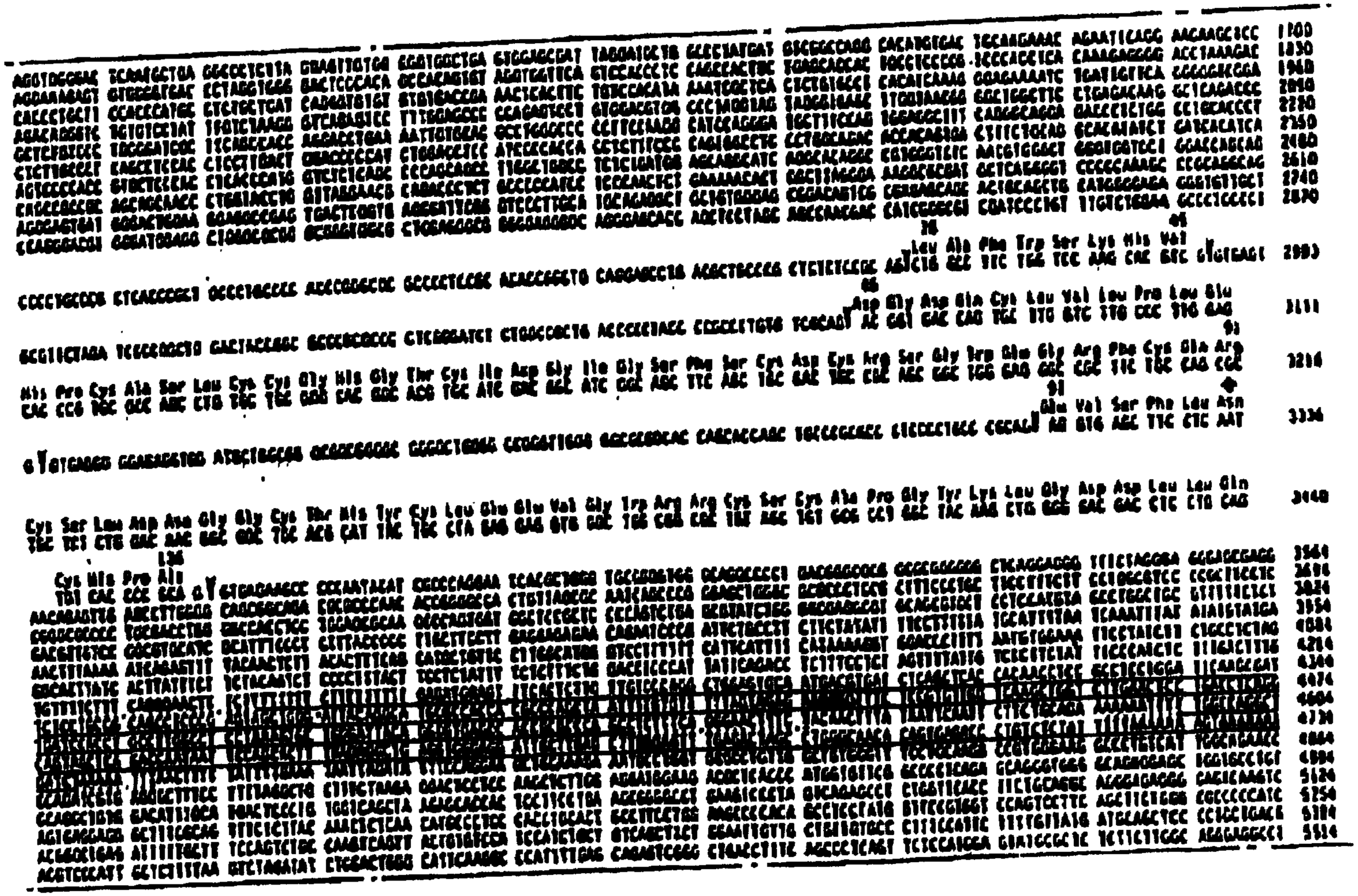

FIG. 2B

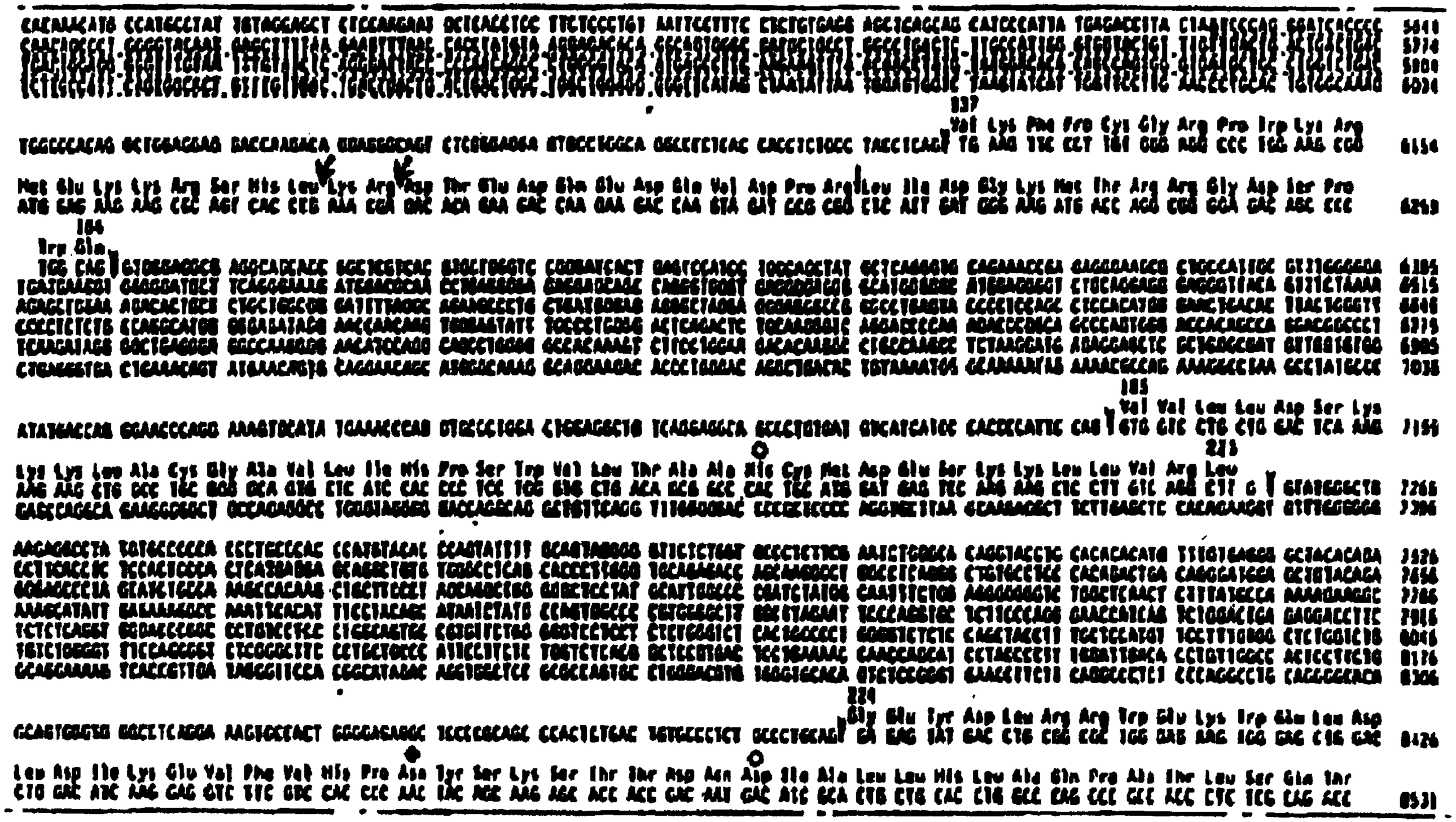

FIG. 2C

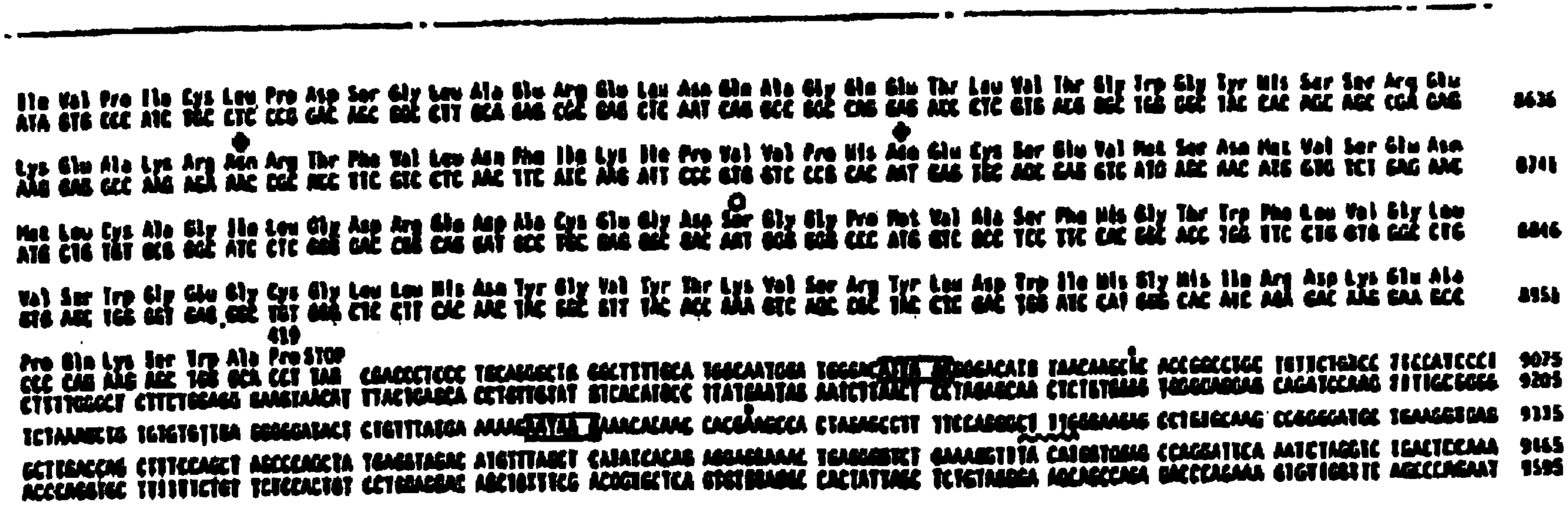

FIG. 2D

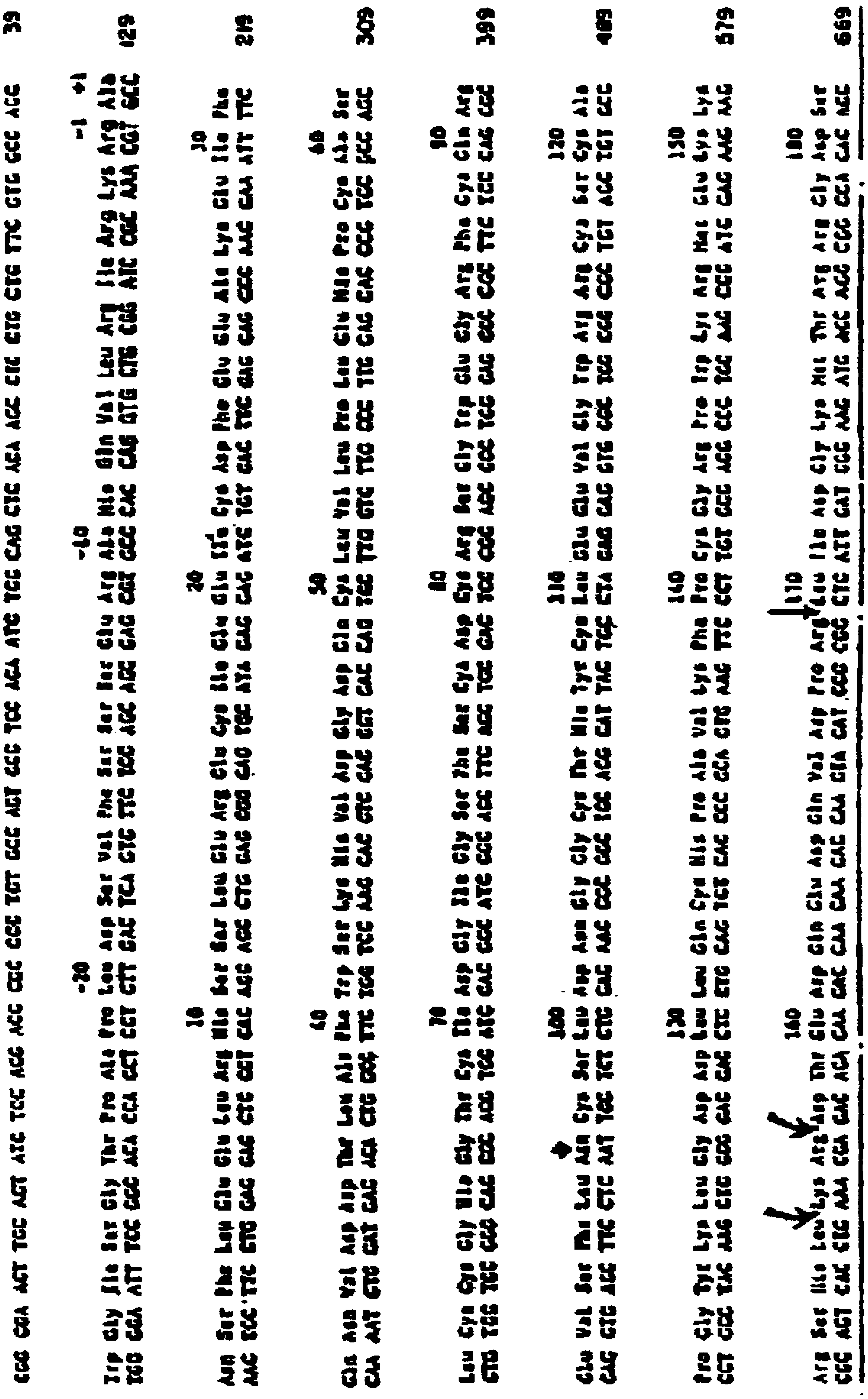

FIG. 3A

```
                                                    190                                                 200                                                 210
Pro Trp Gln Val Val Leu Leu Asp Ser Lys Lys Lys Leu Ala Cys Gly Ala Val Leu Ile His Pro Ser Trp Val Leu Thr Ala Ala His
CCC TGG CAG GTG GTC CTG CTG GAC TCA AAG AAG AAG CTG GCC TGC GGG GCA GTG CTC ATC CAC CCC TCC TGG GTG CTG ACA GCG GCC CAC      759

                                                    220                                                 230                                                 240
Cys Met Asp Glu Ser Lys Lys Leu Leu Val Arg Leu Gly Glu Tyr Asp Leu Arg Arg Trp Glu Lys Trp Glu Leu Asp Leu Asp Ile Lys
TGC ATG GAC GAG TCC AAG AAG CTG CTT GTC AGG CTT GGA GAG TAT GAC CTG CGG CGC TGG GAG AAG TGG GAG CTG GAC CTC GAC ATC AAG      849

                                                    250                                                 260                                                 270
Glu Val Phe Val His Pro Asn Tyr Ser Lys Ser Thr Thr Asp Asn Asp Ile Ala Leu Leu His Leu Ala Gln Pro Ala Thr Leu Ser Gln
GAG GTC TTC GTG CAC CCC AAC TAC AGC AAG AGC ACC ACC GAC AAT GAC ATC GCA CTG CTG CAC CTG GCC CAG CCC GCC ACC CTC TCG CAG      939

                                                    280                                                 290                                                 300
Thr Ile Val Pro Ile Cys Leu Pro Asp Ser Gly Leu Ala Glu Arg Glu Leu Asn Gln Ala Gly Gln Glu Thr Leu Val Thr Gly Trp Gly
ACC ATA GTG CCC ATC TGC CTG CCG GAC AGC GGC CTT GCA GAG CGC GAG CTC AAT CAG GCC GGC CAG GAG ACC CTC GTC ACG GGC TGG GGC     1029

                                                    310                                                 320                                                 330
Tyr His Ser Ser Arg Glu Lys Glu Ala Lys Arg Asn Arg Thr Phe Val Leu Asn Phe Ile Lys Ile Pro Val Val Pro His Asn Glu Cys
TAC CAC AGC AGC CGA GAG AAG GAG GCC AAG AGA AAC CGC ACC TTC GTG CTC AAC TTC ATC AAG ATT CCC GTG GTG CCG CAC AAT GAG TGC     1119

                                                    340                                                 350                                                 360
Ser Glu Val Met Ser Asn Met Val Ser Glu Asn Met Leu Cys Ala Gly Ile Leu Gly Asp Arg Gln Asp Ala Cys Glu Gly Asp Ser Gly
AGC GAG GTC ATG AGC AAC ATG GTG TCT GAG AAC ATG CTG TGT GCG GGC ATC CTC GGC GAC CGG CAG GAT GCC TGC GAG GGC GAC AGT GGC     1209

                                                    370                                                 380                                                 390
Gly Pro Met Val Ala Ser Phe His Gly Thr Trp Phe Leu Val Gly Leu Val Ser Trp Gly Glu Gly Cys Gly Leu Leu His Asn Tyr Gly
GGG CCC ATG GTC GCC TCC TTC CAC GGC ACC TGG TTC CTG GTG GGC CTG GTC AGC TGG GGT GAG GGC TGT GGA CTC CTT CAC AAC TAC GGC     1299

                                                    400                                                 410                                                 419
Val Tyr Thr Lys Val Ser Arg Tyr Leu Asp Trp Ile His Gly His Ile Arg Asp Lys Glu Ala Pro Gln Lys Ser Trp Ala Pro STOP
GTT TAC ACC AAA GTC AGC CGC TAC CTG GAC TGG ATC CAT GGG CAC ATC AGA GAC AAG GAA GCC CCC CAG AAG AGC TGG GCA CCT TAG CGA     1389
```

FIG. 3B

CCC TGG CTG CAG GCC TCG GCT TTT GCA TCC CAA TGG ATG GCA CAT TAA AGG GAC ATG TAA CAA GCA CAC CGG GCT GCT GTT CTG TCC TTC 1479

CAT CCG TCT TTT CGC CTC TTC TGG AGC GAA GTA ACA TTT ACT GAC CAC CTG TTG TAT GTC ACA TGC CTT ATC AAT AGA ATC TTA ACT CCT 1569

AGA GCA ACT CTG TCG GGT CGG GAG GAG CAG ATC GAA GTT TTC CGG GGT CTA AAG CTG TGT GTC TTG ACG GGG ATA CTC TGT TTA TGA AAA 1659

AGA ATA AAA AAC AGA ACC ACG AAA AAA AAA $_{3'}$ 1689

FIG. 3C

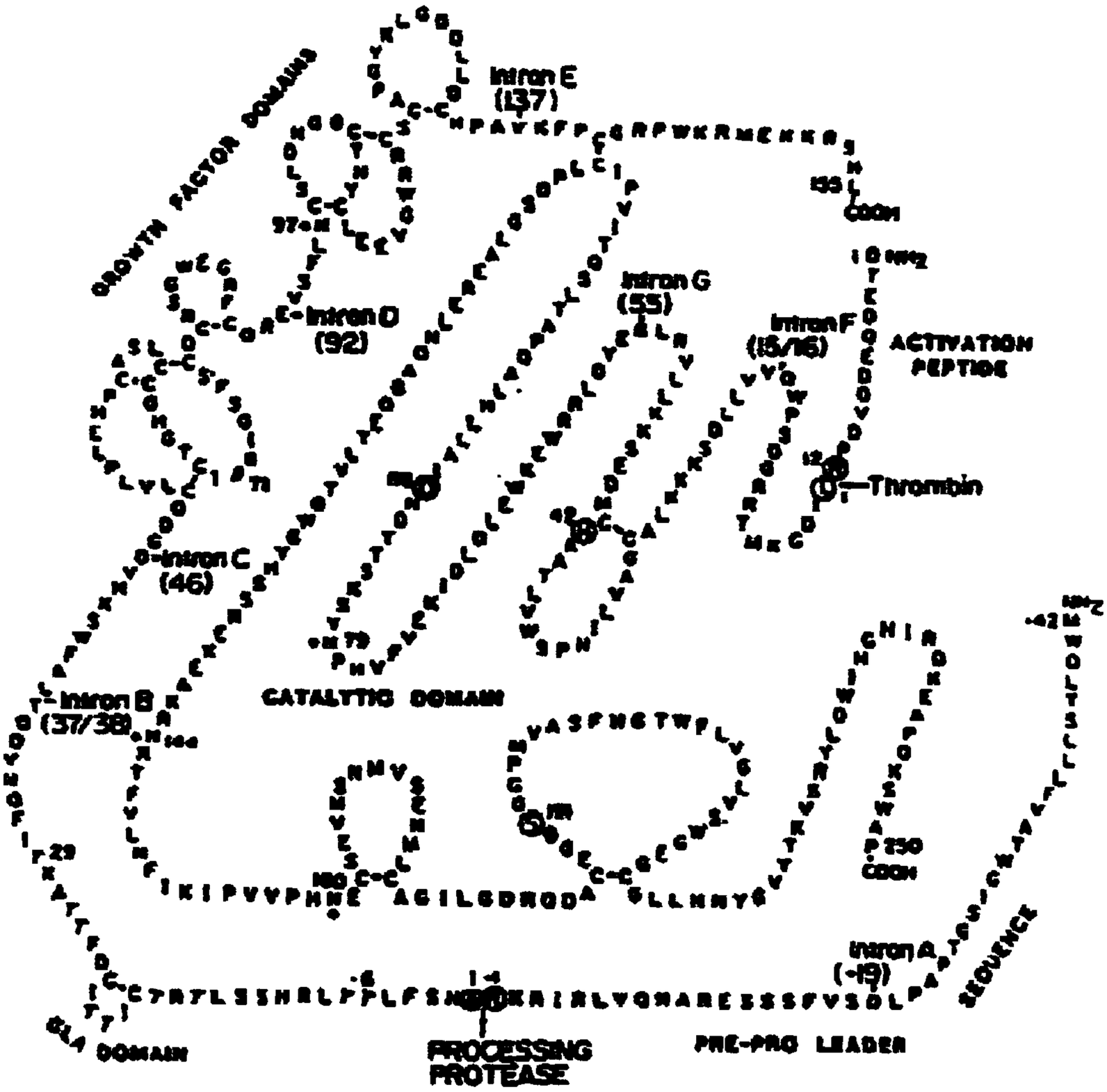
GROWTH FACTOR DOMAINS
Intron E
(137)
Intron D
(92)
Intron C
(46)
Intron B
(37/38)
Intron G
(55)
Intron F
(15/16)
ACTIVATION PEPTIDE
Thrombin
CATALYTIC DOMAIN
Intron A
(-19)
PRE-PRO LEADER
PROCESSING PROTEASE
COOH

FIG. 4

DNA SEQUENCE CODING FOR PROTEIN C

Matter enclosed in heavy brackets [ ] appears in the original patent but forms no part of this reissue specification; matter printed in italics indicates the additions made by reissue.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of application Ser. No. 09/882,150, filed Jun. 15, 2001, now U.S. Pat. No. RE 37,958, issued Jan. 7, 2003, which is a reissue of U.S. Pat. No. 4,968,626, issued on Nov. 6, 1990 from application Ser. No. 06/766,109, filed Nov. 6, 1990, and is related to U.S. Pat. No. 5,073,609 (a division of U.S. Pat. No. 4,968,626) and to U.S. Pat. No. 5,302,529 (which is a continuation of U.S. Pat. No. 4,968,626).

GOVERNMENT SUPPORT

This invention was made with government support under National Institutes of Health grant number HL 16919. The government has certain rights in the invention.

TECHNICAL FIELD

The present invention relates to sequences coding for plasma proteins in general and, more specifically, to a DNA sequence which codes for a protein having substantially the same structure and/or activity of human protein C.

BACKGROUND ART

Protein C is a zymogen, or precursor, of a serine protease which plays an important role in the regulation of blood coagulation and generation of fibrinolytic activity in vivo. It is synthesized in the liver as a single-chain polypeptide which undergoes considerable processing to give rise to a two-chain molecule comprising heavy (Mr=40,000) and light (Mr=21,000) chains held together by disulphide bonds. The circulating two-chain intermediate is converted to the biologically active form of the molecule, known as "activated protein C" (APC), by the thrombin-mediated cleavage of a 12-residue peptide from the amino-terminus of the heavy chain. The cleavage reaction is augmented in vivo by thrombomodulin, an endothelial cell cofactor (Esmon and Owen, Proc. Natl. Acad. Sci. USA 78: 2249–2252, 1981).

Protein C is a vitamin K-dependent glycoprotein which contains approximately eleven residues of gammacarboxyglutamic acid (gla) and one equivalent of betahydroxyaspartic acid which are formed by post-translational modifications of glutamic acid and aspartic acid residues, respectively. The post-translational formation of specific gamma-carboxyglutamic acid residues in protein C requires vitamin K. These unusual amino acid residues bind to calcium ions and are believed to be responsible for the interaction of the protein with phospholipid, which is required for the anticoagulant activity of protein C.

In contrast to the coagulation-promoting action of other vitamin K-dependent plasma proteins, such as factor VII, factor IX, and factor X, activated protein C acts as regulator of the coagulation process through the inactivation of factor Va and factor VIIIa by limited proteolysis. The inactivation of factors Va and VIIIa by protein C is dependent upon the presence of acidic phospholipids and calcium ions. Protein S has been reported to regulate this activity by accelerating the APC-catalyzed proteolysis of factor Va (Walker, J. Biol. Chem. 255: 5521–5524, 1980).

Protein C has also been implicated in the action of plasminogen activator (Kisiel and Fujikawa, Behring Inst. Mitt. 73: 29–42, 1983). Infusion of bovine APC into dogs results in increased plasminogen activator activity (Comp and Esmon, J. Clin. Invest. 68: 1221–1228, 1981). Recent studies (Sakata et al., Proc. Natl. Acad. Sci. USA 82: 1121–1125, 1985) have shown that addition of APC to cultured endothelial cells leads to a rapid, dose-dependent increase in fibrinolytic activity in the conditioned media, reflecting increases in the activity of both urokinase-related and tissue-type plasminogen activators by the cells. APC treatment also results in a dose-dependent decrease in anti-activator activity.

Inherited protein C deficiency is associated with recurrent thrombotic disease (Broekmans et al., New Eng. J. Med. 309: 340–344, 1983; and Seligsohn et al., New Eng. J. Med. 310: 559–562, 1984) and may result from genetic disorder or from trauma, such as liver disease or surgery. This condition is generally treated with oral anti-coagulants. Beneficial effects have also been obtained through the infusion of protein C-containing normal plasma (see Gardiner and Griffin in Prog. in Hematology, ed. Brown, Grune & Stratton, NY, 13: 265–278). In addition, some investigators have discovered that the anti-coagulant activity of protein C is useful in treating thrombotic disorders, such as venous thrombosis (WO 85/00521). In some parts of the world, it is estimated that approximately 1 in 16,000 individuals exhibit protein C deficiency. Further, a total deficiency in protein C is fatal in newborns.

While natural protein C may be purified from clotting factor concentrates (Marlar et al., Blood 59: 1067–1072) or from plasma (Kisiel, ibid), it is a complex and expensive process, in part due to the limited availability of the starting material and low concentration of protein C in plasma. Furthermore, the therapeutic use of products derived from human blood carries the risk of disease transmission by, for example, hepatitis virus, cytomegalovirus, or the causative agent of acquired immune deficiency syndrome (AIDS). In view of protein C's clinical applicability in the treatment of thrombotic disorders, the production of useful quantities of protein C and activated protein C is clearly invaluable.

DISCLOSURE OF INVENTION

Briefly stated, the present invention discloses a DNA sequence which codes for a protein having substantially the same biological activity as human protein C.

In addition, the present invention discloses a recombinant plasmid or bacteriophage transfer vector comprising a cDNA sequence comprising the protein C gene cDNA sequence. The amino acid and DNA sequences of this cDNA coding for human protein C are also disclosed.

Other aspects of the invention will become evident upon reference to the detailed description and attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a restriction enzyme map of the genomic DNA coding for human protein C.

FIG. **2** illustrates the complete genomic sequence, including exons and introns for human protein C. Arrowheads indicate intron-exon splice junctions. The polyadenylation or processing sequences of A-T-T-A-A-A and A-A-T-A-A-A at the 3' end are boxed. ♦, potential carbohydrate binding sites; ↙, apparent cleavage sites for processing of the connecting dipeptide; ↓, site of cleavage in the heavy chain when protein C is converted to activated protein C; ●, sites of polyadenylation.

FIG. **3** depicts the amino acid and DNA sequences for a cDNA coding for human protein C.

FIG. **4** illustrates a proposed model for the structure of human protein C.

BEST MODE FOR CARRYING OUT THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to set forth definitions of certain terms to be used hereinafter.

Biological Activity: A function or set of functions performed by a molecule in a biological context (i.e., in an organism or an in vitro facsimile). Biological activities of proteins may be divided into catalytic and effector activities. Catalytic activities of the vitamin K-dependent plasma proteins generally involve the specific proteolytic cleavage of other plasma proteins, resulting in activation or deactivation of the substrate. Effector activities include specific binding of the biologically active molecule to calcium or other small molecules, to macromolecules, such as proteins, or to cells. Effector activity frequently augments, or is essential to, catalytic activity under physiological conditions.

For protein C, biological activity is characterized by its anticoagulant and fibrinolytic properties. Protein C, when activated, inactivates factor Va and factor VIIIa in the presence of phospholipid and calcium. Protein S appears to be involved in the regulation of this function (Walker, ibid). Activated protein C also enhances fibrinolysis, an effect believed to be mediated by the lowering of levels of plasminogen activator inhibitors (van Hinsbergh et al., Blood 65: 444–451, 1985). As more fully described below, Exons VII and VIII are primarily responsible for the catalytic activity of protein C.

Transfer Vector: A DNA molecule which contains, inter alia, genetic information which ensures its own replication when transferred to a host microorganism strain. Examples of transfer vectors commonly used for recombinant DNA are plasmids and certain bacteriophages. Transfer vectors normally include an origin of replication and sequences necessary for efficient transcription and translation of DNA.

As noted above, protein C is synthesized as a single-chain polypeptide which undergoes considerable processing to give rise to a two-chain molecule; a heavy chain (M_r 41,000) and a light chain (M_r 21,000), held together by a disulfide bond.

Within the present invention, a λgtll cDNA library was prepared from human liver mRNA. This library was then screened with 125_I labeled antibody to human protein C. Antibody-reactive clones were further analyzed for the synthesis of a fusion protein of B-galactosidase and protein C in the λgtll vector.

One of the clones gave a strong signal with the antibody probe and was found to contain an insert of approximately 1400 bp. DNA sequence analysis of the DNA insert revealed a predicted amino acid sequence which shows a high degree of homology to major portions of the bovine protein C, as determined by Fernlund and Stenflo (J. Biol. Chem. 257: 12170–12179; J. Biol. Chem. 257: 12180–12190). Chem. 257: 12170

The DNA insert contained the majority of the coding region for protein C beginning with amino acid 65 of the light chain, including the entire heavy chain coding region, and proceeding to the termination codon. Further, following the stop codon of the heavy chain, there are 294 base pairs of 3' noncoding sequence and a poly (A) tail of 9 base pairs.

The processing or polyadenylation signal A-A-T-A-A-A was present 13 base pairs upstream from the poly (A) tail in this cDNA insert. This sequence is one of two potential polyadenylation sites.

The cDNA sequence also contains the dipeptide Lys-Arg at position 156–157, which separates the light chain from the heavy chain and is removed during processing by proteolytic cleavage. Upon activation by thrombin, the heavy chain of human protein C is cleaved between arginine-12 and leucine-13, releasing the activation peptide.

In order to obtain the remainder of the light chain coding sequence (amino acids 1–64), a human genomic library in λ Charon 4A phage was screened for genomic clones of human protein C using the cDNA described above as a hybridization probe. Three different λ Charon 4A phage were isolated that contained overlapping inserts for the gene coding for protein C.

The position of exons on the three phage clones were determined by Southern blot hybridization of digests of these clones with probes made from the 1400 bp cDNA described above. The genomic DNA inserts in these clones were mapped by single and double restriction enzyme digestion followed by agarose gel electrophoresis, Southern blotting, and hybridization to radiolabeled 5' and 3' probes derived from the cDNA for human protein C, as shown in FIG. **1**.

DNA sequencing studies were performed using the dideoxy chain-termination method. As shown in FIG. **2**, the nucleotide sequence for the gene for human protein C spans approximately 11 kb of DNA. These studies further revealed a potential pre-pro leader sequence of 42 amino acids. Based on homology with the leader sequence of bovine protein C in the region −1 to −20, it is likely that the pre-pro leader sequence is cleaved by a signal peptidase following the Ala residue at position −10. Processing to the mature protein involves additional proteolytic cleavage following residue −1 to remove the amino-terminal propeptide, and at residues 155 and 157 to remove the Lys-Arg dipeptide which connects the light and heavy chains. This final processing yields a light chain of 155 amino acids and a heavy chain of 262 amino acids.

As noted above, the protein C gene is composed of eight exons ranging in size from 25 to 885 nucleotides, and seven introns ranging in size from 92 to 2668 nucleotides. Exon I and a portion of Exon II code for the 42 amino acid pre-pro peptide. The remaining portion of Exon II, Exon III, Exon IV, Exon V, and a portion of Exon VI code for the light chain of protein C. The remaining portion of Exon VI, Exon VII, and Exon VIII code for the heavy chain of protein C. The amino acid and DNA sequences for a cDNA coding for human protein C are shown in FIG. **3**.

The location of the introns in the gene for protein C are primarily between various functional domains. Exon II spans the highly conserved region of the leader sequence and the gamma-carboxyglutamic acid (gla) domain. Exon III includes a stretch of eight amino acids which connect the Gla and growth factor domains. Exons IV and V each represent a potential growth factor domain, while Exon VI covers a connecting region which includes the activation peptide. Exons VII and VIII cover the catalytic domain typical of all serine proteases.

The amino acid sequence and tentative structure for human pre-pro protein C are shown in FIG. **4**. Protein C is shown without the Lys-Arg dipeptide, which connects the light and heavy chains. The location of the seven introns (A through G) is indicated by solid bars. Amino acids flanking

known proteolytic cleavage sites are circled. ♦designates potential carbohydrate binding sites. The first amino acid in the light chain, activation peptide, and heavy chain start with number 1, and differ from that shown in FIGS. **2** and **3**.

Carbohydrate attachment sites are located at residue 97 in the light chain and residues 79, 144, and 160 in the heavy chain, according to the numbering scheme of FIG. **4**. The carbohydrate moiety is covalently linked to Asn, but Thr, Ser, or Gln may be substituted. In the majority of instances, the carbohydrate attachment environment can be represented by N-X-Ser or N-X-Thr, where N=Asn, Thr, Ser, or Gln, and X=any amino acid.

The catalytic domain of protein C, which is encoded by Exons VII and VIII, plays a regulatory role in the coagulation process. This domain possesses serine protease activity which specifically cleaves certain plasma proteins (i.e., factors Va and VIIIa), resulting in their acrivation or deactivation. As a result of this selective proteolysis, protein C displays anticoagulant and fibrinolytic activities.

The example which follows describes the cloning of DNA sequences encoding human protein C.

EXAMPLE

Restriction endonucleases and other DNA modification enzymes (e.g., T_4 polynucleotide kinase, bacterial alkaline phosphatase, Klenow DNA polymerase, T_4 polynucleotide ligase) may be obtained from Bethesda Research Laboratories (BRL) and New England Biolabs and are used as directed by the manufacturer, unless otherwise noted.

CLONING OF DNA SEQUENCES ENCODING HUMAN PROTEIN C

A cDNA coding for a portion of human was prepared as described by Foster and Davie (PNAS (USA) 81: 4766–4770, 1984, herein incorporated by reference). Briefly, a λgtll cDNA library was prepared from human liver mRNA by conventional methods. Clones were screened using 125_I-labeled affinity-purified antibody to human protein C, and phage were prepared from positive clones by the plate lysate method (Maniatis et al., ibid), followed by banding on a cesium chloride gradient. The cDNA inserts were removed using Eco RI and subcloned into plasmid pUC9 (Vieira and Messing, Gene 19: 259–268, 1982). Restriction fragments were subcloned in the phage vectors M13mp10 and m13mp11 (Messing, Meth. in Enzymology 101: 20–77, 1983) and sequenced by the dideoxy method (Sanger et al., Proc. Natl. Acad. Sci. USA 74: 5463–5467, 1977). A clone was selected which contained DNA corresponding to the known sequence of human protein C (Kisiel, ibid) and encoded protein C beginning at amino acid 65 of the light chain and extending through the heavy chain and into the 3' non-coding region. This clone was designated λHC1375.

The cDNA insert from λHC1375 was nick translated using α—^{32}P dNTP's and used to probe a human genomic library in phage λ Charon 4A (Maniatis et al., Cell 15: 687–702, 1978) using the plaque hybridization procedure of Benton and Davis (Science 196: 181–182, 1977) as modified by Woo (Meth. in Enzymology 68: 381–395, 1979). Positive clones were isolated and plaque-purified (by Foster et al., PNAS (USA) 82: 4673–4677, 1985, herein incorporated by reference).

Phage DNA was prepared from positive clones by the method of Silhavy et al. (Experiments with Gene Fusion, Cold Spring Harbor Laboratory, 1984). The purified phage DNA was digested with EcoRI and subcloned into pUC9 for further mapping and sequencing studies. Further analysis suggested that the gene for protein C was present in three EcoRI fragments. In order to generate overlapping protein C DNA sequences, purified phage DNA was digested with Bgl II and subcloned into pUC9.

The sequences of the EcoRI and Bgl II protein C fragments were determined by subcloning the fragments into M13 phage cloning vectors. Sequence analysis of the overlapping fragments established the DNA sequence of the entire protein C gene.

Alternatively, the complete DNA sequence has been determined using a second cDNA clone isolated from a λgtll cDNA library. This clone encodes a major portion of protein C, beginning at amino acid 24 and including the heavy chain coding region, termination codon, and 3' noncoding region. The insert from this λ phage clone was subcloned into pUC9 and the resultant plasmid designated pHC 6L.

This pHC 6L insert was nick translated and used to probe a human genomic library in phage λ Charon 4A. One genomic clone was identified which contained a 4.4 kb EcoRI fragment corresponding to the 5' end of the protein C gene. This phage clone was subcloned into pUC9 and the resultant plasmid designated pHCR 4.4. DNA sequence analysis revealed that the pHCR 4.4 insert comprised two exons, encoding amino acids –42 to –19, and amino acids –19 to 37. Thus, the DNA sequence of the entire protein C gene was established due to the overlapping sequences of pHC 6L (24 to 3' noncoding region) and pHCR 4.4 (–42 to 37).

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

We claim:

[1. An isolated human DNA sequence which codes for a protein having substantially the same biological activity as human protein C.**]**

[2. An isolated DNA sequence comprising the sequence of FIG. **2**, from bp 1 to bp 8972, which sequence codes for human protein C.**]**

[3. A bacterial plasmid for bacteriophage transfer vector comprising a cDNA sequence comprising the human protein C gene cDNA sequence.**]**

4. An isolated human DNA which codes for human protein C, wherein said DNA comprises a sequence which codes for amino acids 1 to 419 as shown in FIG. 3.

5. The isolated human DNA of claim 4, wherein said sequence codes for the amino acid sequence of FIG. 3, starting with methionine, number –42, and ending with proline, number 419.

6. The isolated human DNA of claim 4, wherein said DNA comprises nucleotides 127–1383 of FIG. 3.

7. The isolated human DNA of claim 6, wherein said DNA comprises nucleotides 1–1383 of FIG. 3.

8. The isolated human DNA of claim 4, wherein said DNA comprises nucleotides 1390–1500, 2963–2987, 3080–3217, 3320–3453, 6123–6265, 7139–7256, and 8386–8972 as shown in FIG. 2.

9. The isolated human DNA of claim 8, wherein said DNA comprises nucleotides 1–70, 1334–1500, 2963–2987, 3080–3217, 3320–3453, 6123–6265, 7139–7256, and 8386–8972 as shown in FIG. 2.

10. An isolated human DNA which codes for human protein C, wherein said human protein C comprises a light

chain as shown in FIG. 3 from amino acid number 1 to amino acid number 155, and a heavy chain as shown in FIG. 3 from amino acid number 158 to amino acid number 419.

11. An isolated human DNA which codes for human protein C, wherein said DNA consists of nucleotides 1–1383 of FIG. 3.

12. An isolated human DNA which codes for human protein C, wherein said DNA consists of nucleotides 1–70, 1334–1500, 2963–2987, 3080–3217, 3320–3453, 6123–6265, 7139–7256, and 8386–8972 as shown in FIG. 2.

* * * * *