Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2829–2832, May 1987
Genetics

Hereditary thrombophilia: Identification of nonsense and missense mutations in the protein C gene

GIOVANNI ROMEO*†, H. JANE HASSAN‡, SUSANNE STAEMPFLI*, LAURA RONCUZZI*†, LUCIANO CIANETTI‡, ANTONELLA LEONARDI‡, VICENTE VICENTE§, PIER MANNUCCIO MANNUCCI¶, ROGIER BERTINA||, CESARE PESCHLE‡, AND RICCARDO CORTESE*

*European Molecular Biology Laboratory, 6900 Heidelberg, Federal Republic of Germany; †Laboratorio di Genetica Molecolare, Istituto G. Gaslini, Genova, Italy; ‡Department of Hematology, Istituto Superiore di Sanita, Rome, Italy; §Catedra de Hematologia, Hopital Clinico, Salamanca, Spain; ¶Hemophilia and Thrombosis Center A. Bianchi Bonomi, University of Milan, Milan, Italy; ||University of Leiden, Haemostasis and Thrombosis Research Unit, 2300 RC Leiden, The Netherlands

Communicated by L. L. Cavalli Sforza, December 22, 1986

ABSTRACT **The structure of the gene for protein C, an anticoagulant serine protease, was analyzed in 29 unrelated patients with hereditary thrombophilia and protein C deficiency. Gene deletion(s) or gross rearrangement(s) was not demonstrable by Southern blot hybridization to cDNA probes. However, two unrelated patients showed a variant restriction pattern after *Pvu* II or *Bam*HI digestion, due to mutations in the last exon: analysis of their pedigrees, including three or seven heterozygotes, respectively, with ≈50% reduction of both enzymatic and antigen level, showed the abnormal restriction pattern in all heterozygous individuals, but not in normal relatives. Cloning of protein C gene and sequencing of the last exon allowed us to identify a nonsense and a missense mutation, respectively. In the first case, codon 306 (CGA, arginine) is mutated to an inframe stop codon, thus generating a new *Pvu* II recognition site. In the second case, a missense mutation in the *Bam*HI palindrome (GGATCC → GCATCC) leads to substitution of a key amino acid (a tryptophan to cysteine substitution at position 402), invariantly conserved in eukaryotic serine proteases. These point mutations may explain the protein C-deficiency phenotype of heterozygotes in the two pedigrees.**

Hereditary thrombophilia, an autosomal dominant disease with a prevalence of 1 in 7500, is characterized by recurrent venous thrombosis in heterozygotes and neonatal purpura fulminans in homozygotes, due to a deficiency of any of three antithrombotic proteins, i.e., antithrombin III (1), protein S (2), or protein C (3). Protein C is a precursor of a plasma serine protease (4–6), exerting its anticoagulant activity through inactivation of factors Va and VIIIa, as well as stimulation of fibrinolysis (3). Its cDNA (7–9) and gene (10, 11) have been cloned and sequenced, and the gene has been mapped on chromosome 2 (12). Two major groups of hereditary protein C deficiency have been described (13): type I shows an equivalent reduction of both enzymatic activity and antigen concentration, and type II shows a decrease of the former parameter only. Our studies focused on 50 normal subjects and 29 protein C-deficient subjects who had a variety of phenotypic abnormalities. In two families with protein C deficiency of type I, all of the heterozygotes had (*i*) a nonsense mutation, causing premature termination of protein C synthesis, or (*ii*) a missense mutation, leading to the substitution of a key tryptophan with cysteine at residue 402. Normal subjects expressed neither of these mutations.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Blood Samples. Peripheral blood samples, from 50 normal subjects and 29 unrelated protein C deficient patients, were collected in a 10% volume of 3.2% (wt/vol) trisodium citrate. Blood cells, washed twice with PBS (137 mM NaCl/2.7 mM KCl/8 mM Na_2HPO_4/1.5 mM KH_2PO_4, pH 7.4), were stored at −20°C until utilized.

DNA Analysis. High molecular weight DNA was extracted from leukocytes by standard techniques (14), digested with restriction endonucleases (Boehringer Mannheim; New England Biolabs), electrophoresed on 0.8% agarose gels, transferred to nitrocellulose filters (BA-85, Schleicher & Schuell, West Germany) (15), and finally hybridized in 50% (vol/vol) formamide at 42°C overnight to 2×10^7 cpm of protein C cDNA probe (7, 8). The probe was ^{32}P-labeled by nick-translation to a specific activity of $3–7 \times 10^8$ dpm/μg. Filters were washed under stringent conditions (final wash: 15 mM NaCl/1.5 mM trisodium citrate/0.1% $NaDodSO_4$, pH 7 at 65°C) and exposed at −70°C to Kodak SO-282 x-ray films in X-omatic intensifying screen cassette.

DNA Cloning and Sequencing. Genomic DNA from patient S.V. was partially digested with *Mbo* I restriction endonuclease and inserted into the *Bam*HI site of the λEMBL3 vector (16), according to standard procedures (17). Plaques were screened with protein C cDNA probe (7), and positive recombinant clones were characterized by restriction endonuclease mapping. An ≈1-kilobase (kb) *Sac* I fragment containing the last part of exon 9 was subcloned in M13mp19 (18) and sequenced by dideoxy chain-termination (19).

Genomic DNA from patient M.L. was digested to completion with *Bam*HI. The 9.5-kb DNA region was gel-isolated and inserted into the *Bam*HI site of λEMBL3 (16). Plaques (5×10^4 plaques) were screened with the cDNA probe (7), and several clones containing the DNA segment encoding protein C were isolated. A 411-base-pair *Pst* I–*Sac* I segment from exon 9 was subcloned into M13mp18 (18) and sequenced (19).

RESULTS

Southern Blot Analysis. Our studies include 50 normal subjects and 29 protein C-deficient pedigrees with a variety of phenotypic abnormalities. Twenty-eight families had type I-deficient patients, of which 26 had only heterozygotes with 50% reduction of both enzymatic activity and antigen level, and 2 had also homozygous patients. A single type II pedigree included only heterozygotes.

DNA samples from the 29 unrelated patients and 50 normal individuals were digested with eight different restriction enzymes (*Bam*HI, *Eco*RI, *Pst* I, *Pvu* II, *Bgl* II, *Hin*dIII, *Taq* I, and *Xba* I): no deletion(s) or gross rearrangement(s) of the

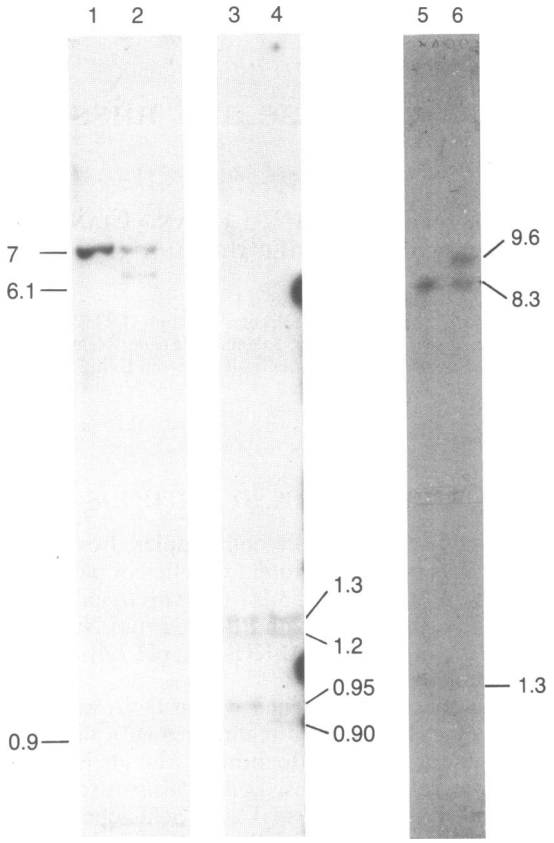

FIG. 1. Southern blots of genomic DNA from protein C-deficient patients and control subjects. *Pvu* II digestion of DNA from a normal individual (lane 1) and patient S.V. (lane 2). Mapping of the new *Pvu* II site: double digestion with *Sac* I and *Pvu* II in a normal subject (lane 3) and S.V. (lane 4). *Bam*HI digestion of DNA from a normal individual (lane 5) and patient M.L. (lane 6). The 9.6-kb fragment is generated by disappearance of the last exon *Bam*HI site (10). The 8.3- and 1.3-kb fragments identify the normal allele.

protein C gene was demonstrated by Southern blot hybridization to cDNA probes (results not shown).

Two patients with protein C deficiency of type I, S.V. (*cf.* family 2 in ref. 20) and M.L. (see ref. 21), showed a variant restriction pattern after *Pvu* II and *Bam*HI digestion, respectively. The normal *Pvu* II pattern (Fig. 1*A*, lane 1) consists of a single 7-kb fragment (*cf.* ref. 10 and Fig. 2). In patient S.V. (Fig. 1*A*, lane 2) two additional fragments of 6.1 and 0.9 kb were observed, indicating presence of a new *Pvu* II restriction site in the last part of exon 9 (Figs. 1*B* and 2), mapped by double digestion with *Sac* I and *Pvu* II on the 3′ side of the *Sac* I site.

The normal *Bam*HI pattern (Fig. 1*C*, lane 1) includes two fragments (8.3 and 1.3 kb; *cf.* ref. 11 and Fig. 2). Patient M.L. shows an additional 9.6-kb band (Fig. 1*C*, lane 2), thus indicating loss of the *Bam*HI restriction site in the last exon (10).

Further analysis was carried out on the pedigrees of the first and second patient. All three and seven heterozygotes in pedigrees S.V. (20) and M.L. (21), respectively, showed the variant fragment(s) described above. Conversely, the restriction pattern of all unaffected members in these two families was equivalent to that in the other 27 pedigrees, as well as in 50 normal individuals (data not shown).

Sequence Analysis. To assess whether the variant patterns may reflect mutations resulting in protein C deficiency, we have isolated and sequenced exon 9 of the protein C gene encompassing these mutations. The DNA sequence of patient S.V. shows that codon 306 (CGA, arginine) is mutated into an inframe stop codon (TGA), thus generating a new *Pvu* II recognition site (Fig. 3*A*). The DNA sequence from patient M.L. shows a single transversion in the *Bam*HI palindrome (GGATCC → GCATCC), thereby causing tryptophan to cysteine substitution at position 402 (Fig. 3*B*).

DISCUSSION

Protein C deficiency represents a major cause of thrombotic disease (3, 20–24) through the malfunction of an antithrombotic regulatory system including protein C, protein S, and an endothelial cell cofactor, thrombomodulin (24). However, the molecular basis of protein C deficiency is still almost completely unknown. Our study focused on 29 unrelated patients with deficiency of this factor, leading to a variety of phenotypic abnormalities: two patients with protein C deficiency of type I had a variant restriction pattern after *Pvu* II or *Bam*HI digestion. Sequencing of the last exon of the protein C gene from these subjects allowed us to identify two different point mutations: in the first case codon 306 (CGA, arginine) is mutated into a stop codon (TGA), while in the second case a missense mutation causes substitution of tryptophan (TGG) with a cysteine (TGC) codon at amino acid position 402 (Fig. 3*B*).

Since we have not sequenced the whole protein C gene from these patients, it cannot be excluded that other mutations might be present in other parts of the gene. However, the point mutations reported here seem sufficient to explain the clinical phenotype. Thus, the two variant restriction patterns are 100% correlated with protein C deficiency in both pedigrees, but are absent in all other normal or affected subjects. Furthermore, the 50% reduction of protein C level documented in these heterozygotes may be easily attributed to the nonsense mutation in the pedigree of patient S.V. but also presumably to the missense mutation in the pedigree of

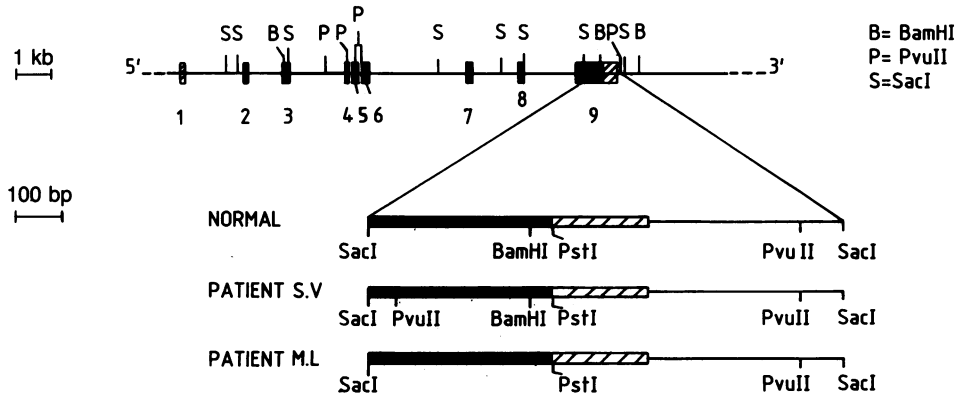

FIG. 2. Exon-intron structure of human protein C gene. Relevant restriction sites are indicated. bp, Base pairs.

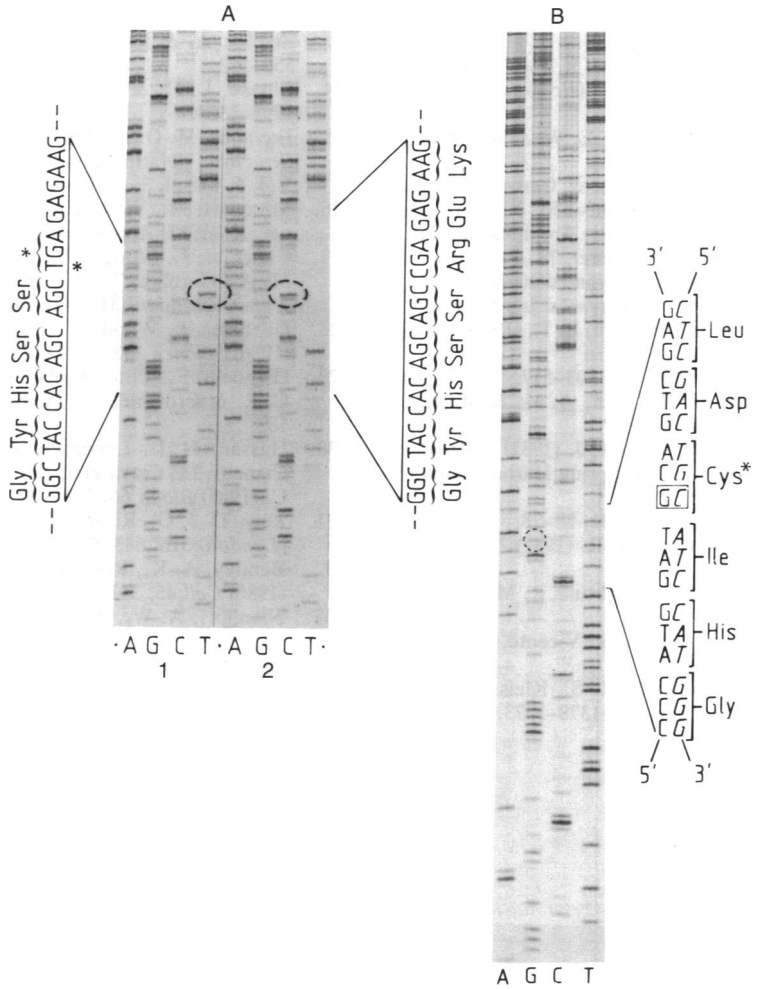

FIG. 3. (*A*) Nucleotide sequence of part of the *Sac* I–*Sac* I segment from exon 9 of the normal (*Left*) and the mutant (*Right*) allele of patient S.V. Nucleotide differences are circled. (*B*) The nucleotide sequence of part of protein C gene from patient M.L. The missense mutation is circled. The sequences in normal and italic letters indicate the sequence derived from the gel and the complementary strand, respectively. The asterisks indicate the nonsense codon generated by the cytidine to thymidine transition.

patient M.L. Tryptophan-402 is invariantly conserved in a biochemical domain present in all eukaryotic serine proteases (25), located in an α-helix region facing hydrophobic residues (26). Computer simulation of the three-dimensional structure of this domain in trypsin indicated that replacing the large tryptophan aromatic ring with the small cysteine hydrophilic side-chain engenders physicochemical constraints, leading to destabilization of the tertiary structure. Alternatively or additionally, improper disulfide bonds generated by the variant cysteine residue might interfere with the correct folding of the protein, thus leading to its destabilization or inactivation.

Growing evidence indicates that coagulation disorders may derive from a heterogeneous array of molecular lesions, as shown for hemophilia A (27, 28) or B (29–32). In particular, hemophilia B is caused by deficiency of factor IX, i.e., a serine protease strictly related to protein C in terms of sequence homology and intron-exon organization, although exerting an opposite biological activity. In B hemophiliacs the factor IX gene may bear either extensive deletions (29, 30) or splicing mutations (31) or missense mutation in the propeptide region (32). In protein C deficiency of type I the gene may be apparently silenced by yet different molecular abnormalities, i.e., either a nonsense mutation or substitution of the key tryptophan-402, invariantly conserved in eukaryotic serine proteases.

It is of interest to note that molecular studies on thalassemia syndromes (33, 34) have similarly shown that a large variety of abnormalities, affecting globin gene expression at the level of transcription, RNA processing, or translation, underlie the heterogenous array of thalassemic phenotypes.

We express our gratitude to J. Aznar, G. Mariani, I. Pabinger-Fashing, and N. Sala for supplying protein C deficient blood samples, E. W. Davie for kindly providing a protein C cDNA probe, M. Tripodi for advice, L. Philipson for critical reading of the manuscript, and H. Seifert and P. Alessandri for secretarial assistance. The work performed in Italy was in part supported by Consiglio Nazionale delle Ricerche, Rome, Italy, Progetti Finalizzati "Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie" (to G.R., P.M.M., C.P., R.C.) and "Oncologia" (to C.P. and R.C.).

1. Thaler, E. & Lechner, K. (1981) *Clin. Hematol.* **10,** 369–390.
2. Bertina, R. M. (1985) *Haemostasis* **15,** 241–246.
3. Broeckmans, A. W. (1985) *Haemostasis* **15,** 233–240.
4. Stenflo, J. (1976) *J. Biol. Chem.* **251,** 355–363.
5. Fernlund, P. & Stenflo, J. (1982) *J. Biol. Chem.* **257,** 12170–12179.
6. Stenflo, J. & Fernlund, P. (1982) *J. Biol. Chem.* **257,** 12180–12190.
7. Foster, D. & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4766–4770.
8. Santamaria, R., Roncuzzi, L., Sbarra, D., Cortese, R. & Romeo, G. (1984) in *New Trends in Experimental Haematol-*

ogy, eds. Peschle C. & Rizzoli C. (Ares Serono Symposia, Rome, Italy), Review no. 7, pp. 296–298.
9. Beckmann, R. J., Schmidt, R. J., Santerre, R. F., Plutzky, J., Crabtree, G. R. & Long, G. L. (1985) *Nucleic Acids Res.* **13,** 5233–5246.
10. Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4673–4677.
11. Plutkzy, J., Hoskins, J. A., Long, G. L. & Crabtree, G. R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 546–550.
12. Rocchi, M., Roncuzzi, L., Santamaria, R., Archidiacono, N., Dente, L. & Romeo, G. (1986) *Hum. Genet.* **74,** 30–33.
13. Bertina, R. M., Broekmans, A. W., Krommeuhaevan Es, C., Van Wijngaarden, A. (1984) *Thromb. Haemostasis* **51,** 1–5.
14. Britten, R. J., Graham, D. E. & Nenfeled, B. R. (1974) *Methods Enzymol.* **29,** 363–383.
15. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
16. Frischauf, A. M., Lehrach, H., Poustka, A. M. & Murray, N. J. (1983) *Mol. Biol.* **170,** 827–842.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
18. Messing, J. (1983) *Methods Enzymol.* **101,** 28–78.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Broeckmans, A. W., Veltkamp, J. J., Bertina, R. M. (1983) *N. Engl. J. Med.* **309,** 340–344.
21. Gonzales, R., Alberca, I., Sala, N. & Vicente, V. (1985) *Tromb. Haemostasis* **53,** 320–322.
22. Griffin, J. H., Evatt, B., Zimmerman, T. S., Kleiss, A. G. & Wideman, C. (1981) *J. Clin. Invest.* **68,** 1370–1373.
23. Seligsohn, U., Berger, A., Aben, M., Rubin, L., Attias, D., Zivelin, A. & Rapaport, S. I. (1984) *N. Engl. J. Med.* **310,** 559–562.
24. Clouse, L. H. & Comp, P. C. (1986) *N. Engl. J. Med.* **314,** 1298–1304.
25. Dayhoff, M. O. (1978) in *Atlas of Proteins Sequence and Structure*, (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, p. 73.
26. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Hubner, R. (1983) *Acta Crystallogr.* **39,** 480–490.
27. Gitschier, J., Wood, W. I., Tuddenham, E. G. D., Schuman, M. A., Goralka, T. M., Chen, E. Y. & Lawn, R. M. (1985) *Nature (London)* **315,** 427–430.
28. Gitschier, J., Wood, W. T., Schuman, M. A. & Lawn, R. M. (1986) *Science* **232,** 1415–1416.
29. Giannelli, F., Choo, K. H., Rees, D. J. G., Boyd, Y., Rizza, C. R. & Brownlee, G. G. (1983) *Nature (London)* **303,** 181–182.
30. Hassan, H. J., Leonardi, A., Guerriero, R., Chelucci, C., Cianetti, L., Ciavarella, N., Ranieri, P., Pilolli, D. & Peschle, C. (1985) *Blood* **66,** 728–730.
31. Rees, D. J. G., Rizza, C. R. & Brownlee, G. G. (1985) *Nature (London)* **316,** 643–645.
32. Bentley, A. K., Rees, D. J. G., Rizza, C. & Brownlee, G. G. (1986) *Cell* **45,** 343–348.
33. Nienhuis, A. W., Anagnon, N. P. & Ley, T. J. (1984) *Blood* **63,** 738–758.
34. Kazazian, H., Orkin, S. H., Antonorakis, S. E., Sexton, P., Boehm, C. D., Goff, S. C. & Waber, P. G. (1984) *EMBO J.* **3,** 593–596.