Characterization of a cDNA Coding for Human Factor XII (Hageman Factor)†

Benito G. Que and Earl W. Davie*

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received September 3, 1985

ABSTRACT: Affinity-purified antibody against human factor XII (Hageman factor) has been radiolabeled with ^{125}I and employed as a probe to screen a human liver cDNA expression library prepared in λgt11. Approximately 3.5×10^6 recombinant phages were screened for factor XII, and two positive clones were identified and plaque purified. The largest cDNA coding for factor XII was 1571 base pairs in length and coded for amino acid residues 127–596 in the mature protein, a termination codon of TGA, a 3′ noncoding sequence of 147 nucleotides, and a poly(A) tail of 11 nucleotides. The second clone contained an insert of 1334 base pairs and coded for amino acid residues 200–596. The amino acid sequence predicted by the cDNAs was in excellent agreement with that previously determined by amino acid sequence analysis. The amino acid and DNA sequences in human factor XII showed considerable homology with the corresponding domains in other serine proteases, including prothrombin, plasminogen, tissue plasminogen activator, and urokinase.

Factor XII (Hageman factor) (Ratnoff & Colopy, 1955) is a single-chain glycoprotein (M_r 76 000) that is present in plasma as a zymogen form of a serine protease (Fujikawa & Davie, 1981). When the contact phase of coagulation or fibrinolysis is initiated, factor XII is converted to factor XII_a by the cleavage of a single internal Arg–Val peptide bond. This reaction occurs in the presence of kallikrein, an ionic surface, and high molecular weight kininogen (Griffin & Cochrane, 1979; Heimark et al., 1980). Human factor XII_a is a serine protease containing a heavy chain (350 residues) and a light chain (243 residues), and these two chains are held together by a disulfide bond (Fujikawa & McMullen, 1983; McMullen & Fujikawa, 1985). The heavy chain contains several potential domains (McMullen & Fujikawa, 1985), including type I and type II finger structures as reported in fibronectin (Petersen et al., 1983; Skorstengaard et al., 1984), two potential growth factor domains (Gregory & Preston, 1977), and a kringle structure (Magnusson et al., 1975). The heavy chain is responsible for the binding of factor XII (or factor XII_a) to an anionic surface, such as kaolin (Revak et al., 1974). The light chain of factor XII_a contains the serine protease portion of the molecule and is responsible for the catalytic function of the enzyme (Dunn et al., 1982; Fujikawa & McMullen, 1983). Factor XII_a activates several plasma proteins, such as prekallikrein, factor XI, and factor VII, by limited proteolysis. In these reactions, specific internal arginine-containing peptide bonds are cleaved.

In order to gain further understanding of the biosynthesis, regulation, and structure of the gene for factor XII, a cDNA coding for the human protein has been isolated and characterized. This clone was isolated from a cDNA expression library prepared in λgt11 from human liver poly(A) RNA.

Materials and Methods

Restriction endonucleases, M13mp11 RF-DNA, T4 DNA ligase, calf intestinal phosphatase, the Klenow fragment of DNA polymerase I, and Bal-31 nuclease were purchased from either Bethesda Research Laboratory or New England Biolab.

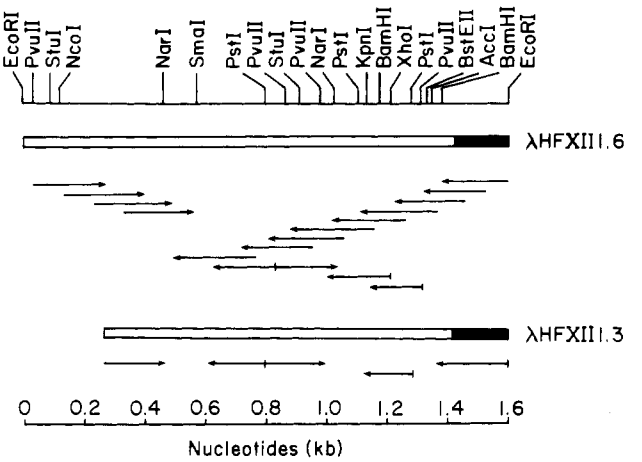

FIGURE 1: Restriction map and sequencing strategy for the cDNA inserts in λHFXII1.6 and λHFXII1.3 that code for human factor XII. The extent of sequencing is shown by the length of each arrow, and the direction of the arrow indicates the strand that was sequenced.

Deoxynucleoside and dideoxynucleoside triphosphates were supplied by P-L Biochemicals, and $Na^{125}I$ was purchased from New England Nuclear. The λgt11 cDNA library prepared from human liver was kindly provided by Dr. Savio L. C. Woo (Kwok et al., 1985) and was screened with ^{125}I-labeled antibody by the method of Young and Davis (1983a,b).

Affinity-purified rabbit antibody against human factor XII was generously provided by Dr. Kazuo Fujikawa and Lee Hendrickson. It was raised by injection of 0.5 mg of the purified protein into rabbits, and blood was withdrawn 5 weeks after the first injection. The IgG fraction was prepared by the method of Harboe and Ingild (1973) and dialyzed against 0.05 M Tris-HCl [tris(hydroxymethyl)aminomethane hydrochloride] buffer, pH 8, containing 0.15 M NaCl (TBS). The rabbit antibody was then passed through a column containing 2 mL of human IgG–Sepharose, and the unadsorbed fraction was applied to a 4-mL column containing human factor XII–Sepharose previously equilibrated with TBS. After the column was washed extensively wth 1 L of TBS, the adsorbed antibody was eluted with 0.2 M glycine–HCl buffer, pH 2.5, and the eluate mixed immediately with 1 M Tris-HCl buffer,

† This work was supported in part by Research Grant HL 16919 from the National Institutes of Health.

0006-2960/86/0425-1525$01.50/0

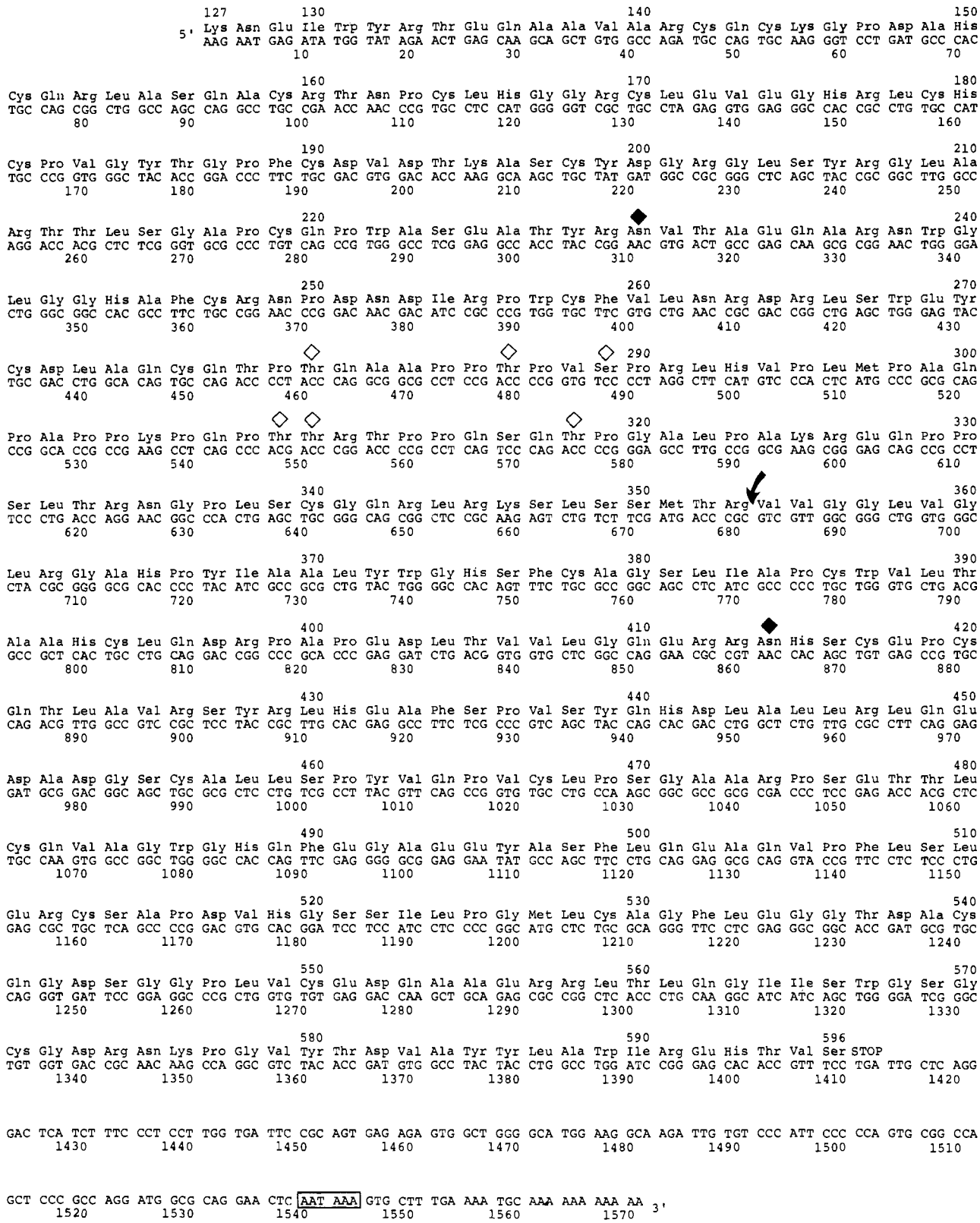

FIGURE 2: Complete nucleotide sequence for the cDNA insert in λHFXII1.6 coding for human factor XII. The predicted amino acid sequence is also shown starting with amino acid residue 127 in the single polypeptide chain. The solid diamonds indicate the N-glycosylation sites, and the open diamonds represent the proposed O-glycosylation sites (McMullen & Fujikawa, 1985). The solid arrows indicate the site of cleavage in the molecule when factor XII is converted to factor XII_a by plasma kallikrein. The processing or polyadenylation sequence of AATAAA is shown in boxes.

pH 8.0. Fractions that contained protein were pooled, dialyzed, and concentrated against TBS by employing a MicroProDiCon concentrator. The human IgG–Sepharose and human factor XII–Sepharose were prepared by coupling 0.5 g of activated CH-Sepharose (Pharmacia) and 2 mL of Sepharose 4B to 4 mg of each purified protein, as described in the instructions provided by Pharmacia. The affinity-purified antibody to human factor XII was radiolabeled with ^{125}I, as described by Fraker and Speck (1978). Positive clones were isolated and plaque purified, and phage DNA was prepared by the plate–

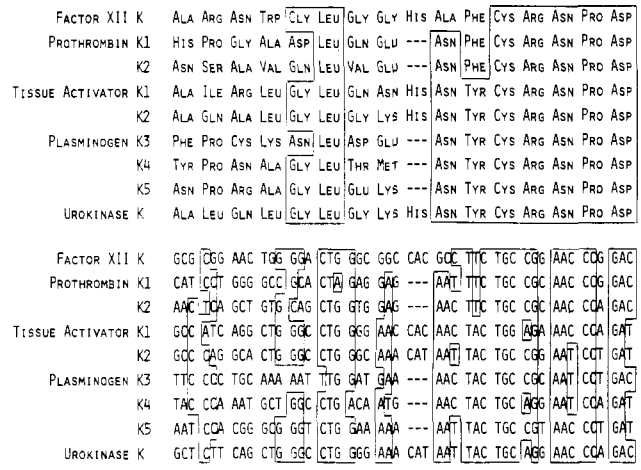

FIGURE 3: Comparison of a portion of the amino acid and DNA sequence of the kringle in factor XII with the analogous sequences in prothrombin, tissue plasminogen activator, plasminogen, and urokinase. Five or more amino acids or nucleotides that are identical in each of the domains are shown in boxes.

lysis method (Maniatis et al., 1982). Restriction enzyme mapping, Bal-31 deletion experiments, and subcloning of cDNA fragments into M13mp11 were carried out as described in the Amersham cloning and sequencing manual. Dideoxy sequencing reactions were performed with [α-^{35}S]-thio-dATP purchased from Amersham and run on buffered gradient gels as described (Biggin et al., 1983). More than 90% of each strand of the cDNA insert was sequenced twice. RNA was prepared by the procedure of Chirgwin et al. (1979). RNA and DNA gel electrophoresis and Southern blotting were carried out according to published methods (Southern, 1975). Peripheral blood DNA was kindly provided by Dr. Barbara Schach.

Results and Discussion

Two positive clones were identified from a human liver λgt11 cDNA expression library of approximately 3.5×10^6 recombinant phages when it was screened with an affinity-purified antibody against human factor XII by the method of Young and Davis (1983a,b). The two positive phages (λHFXII1.6 and λHFXII1.3) were isolated and plaque purified, and their DNA was isolated. The cDNA inserts from λHFXII1.6 and λHFXII1.3 were then subcloned into the replicative form of M13mp11, and a partial restriction map was prepared (Figure 1). The nucleotide sequences of the inserts were then determined by the dideoxy method using the strategy shown in the bottom of Figure 1. The cDNA insert from λHFXII1.6 was found to contain 1571 nucleotides and coded for amino acid residues 127–596 of the mature protein present in plasma (Figure 2). The carboxyl-terminal serine coded by nucleotides 1408–1410 was followed by a termination codon of TGA, 147 nucleotides of 3′ noncoding sequence, and a poly(A) tail of 11 nucleotides. The polyadenylation or processing signal AATAAA (Proudfoot & Brownlee, 1976) was present 15 nucleotides upstream from the poly(A) tail. The cDNA insert in λHFXII1.3 contained 1344 nucleotides and coded for amino acids 200–596 of the mature protein. It differed slightly from the cDNA insert in λHFXII1.6 at the 3′ end of the noncoding region, in that it contained two extra nucleotides (TG) following nucleotide 1560 (C) just prior to the poly(A) tail.

The amino acid sequence of factor XII deduced from the cDNA sequences in λHFXII1.6 and λHFXII1.3 was in complete agreement with that found by amino acid sequence analysis (Fujikawa & McMullen, 1983; McMullen & Fujikawa, 1985) except for glycine-360. This residue was shown to be an alanine by protein sequence analysis. It seems unlikely that this minor difference is due to a cloning artifact, since glycine-360 was predicted by both λHFXII1.6 and λHFXII1.3. Accordingly, it appears probable that a true polymorphism exists in the gene for human factor XII at this site.

Human factor XII is a multifunctional protein. Starting from the N-terminal end, it contains a potential type II finger domain, a growth factor domain, a type I finger domain, a second growth factor domain, a kringle domain, and a serine protease or catalytic domain (McMullen & Fujikawa, 1985). The finger domains were originally identified in fibronectin (Petersen et al., 1983). The largest cDNA clone isolated in this investigation is a partial clone starting with amino acid 127 in the type I finger domain. This cDNA spans about 80% of the total amino acid sequence of factor XII. Accordingly, it lacks the leader sequence that is typical of secreted proteins (Blobel et al., 1979), as well as the coding region for the type II finger domain and the first growth factor domain. When preparations of mRNA isolated from human liver, baboon kidney, and pancreatic carcinoma cell lines were subjected to gel electrophoresis, transferred, blotted onto nitrocellulose, and hybridized with a nick-translated cDNA insert from λHFXII1.6, a single distinct band equivalent to about 2100 nucleotides was detected in human liver (data not shown). Accordingly, the largest clone isolated for factor XII lacks about 500 nucleotides on its 5′ end.

The domain structures and catalytic region of factor XII show considerable amino acid sequence homology with the analogous structures present in prothrombin (Magnusson et al., 1975; Degen et al., 1983), tissue plasminogen activator (Pennica et al., 1983; Pohl et al., 1984), urokinase (Gunzler et al., 1982; Steffens et al., 1982; Verde et al., 1984), and plasminogen (Magnusson et al., 1976) (Figure 3). Consequently, considerable DNA sequence identity was also noted in the domains for these proteins, as shown in Figure 3 where the same portions of the kringle structures are compared.

The gene for factor XII is at least 20 kilobases in size, as shown by Southern blotting. In these studies, restriction enzyme digestion of human DNA was carried out with *Eco*RI, *Bam*HI, or *Kpn*I, and the cDNA insert in λHFXII1.6 was employed as a probe. This probe, however, may not detect all of the 5′ end of the gene since it starts with amino acid 127 in human factor XII. Accordingly, the gene for factor XII may be considerably larger than 20 kilobases. Whether introns are located between the coding regions for the various potential domains in factor XII remains to be determined. In the vitamin K dependent proteins that participate in blood coagulation, the introns separate some of the protein sequences into well-defined domains. For instance, the coding regions for the Gla domains and the kringle or the growth factor domains in the vitamin K dependent plasma proteins are separated by introns (Degen et al., 1983, 1985; Anson et al., 1984; Yoshitake et al., 1985; Foster et al., 1985). Also, introns separate the first and second growth factor domains in factor IX and protein C. In prothrombin, an intron separates kringle 1 and kringle 2. Kringle 1, however, contains an additional intron within its coding structure. A similar situation has been reported for the genes for other proteins, such as the LDL receptor (Sudhof et al., 1985). The gene for this protein contains 18 exons, and 13 of the exons code for protein sequences that are homologous to sequences in other proteins, including the growth factor domains. Accordingly, it seems probable that a similar situation may exist for the gene for human factor XII.

ADDED IN PROOF

After the manuscript was submitted, Cool et al. (1985) reported the isolation of a cDNA for human factor XII.

ACKNOWLEDGMENTS

We thank Drs. Kazuo Fujikawa, Torben Petersen, Dominic Chung, Kotoku Kurachi, and Evan Sadler for helpful suggestions and assistance during the course of these investigations. We thank Dr. Savio L. C. Woo for the generous gift of the λgt11 cDNA library.

REFERENCES

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, J. A., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

Biggin, M. D., Gibson, T. T., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H., & Lingappa, R. (1979) *Symp. Soc. Exp. Biol. 33*, 9–36.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Cool, D. E., Edgell, C.-J. S., Louie, G. V., Zoller, M. J., Brayer, G. D., & MacGillivray, R. T. A. (1985) *J. Biol. Chem. 260*, 13666–13676.

Degen, S. J. F., MacGillivray, R. T. A., & Davie, E. W. (1983) *Biochemistry 22*, 2087–2097.

Degen, S. J. F., Rajput, B., Reich, E., & Davie, E. W. (1985) in *Protides of Biological Fluids* (Peeters, H., Ed.) Vol. 33, pp 47–50, Pergamon Press, New York.

Dunn, J. T., Silverberg, M., & Kaplan, A. P. (1982) *J. Biol. Chem. 257*, 1779–1784.

Foster, D. C., Yoshitake, S., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4673–4677.

Fraker, P. J., & Speck, J. C., Jr. (1978) *Biochem. Biophys. Res. Commun. 80*, 849–857.

Fujikawa, K., & Davie, E. W. (1981) *Methods Enzymol. 80*, 198–211.

Fujikawa, K., & McMullen, B. A. (1983) *J. Biol. Chem. 258*, 10924–10933.

Gregory, H., & Preston, B. M. (1977) *Int. J. Pept. Protein Res. 9*, 107–118.

Griffin, J. H., & Cochrane, C. G. (1979) *Semin. Thromb. Hemostasis 5*, 254–273.

Gunzler, W. A., Steffens, G. J., Otting, F., Kim, S.-M. A., Frankus, E., & Flohe, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 1155–1165.

Harboe, N., & Ingild, A. (1973) *Scand. J. Immunol., Suppl. 2*, 161–164.

Heimark, R. L., Kurachi, K., Fujikawa, K., & Davie, E. W. (1980) *Nature (London) 286*, 456–460.

Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K. J. H., & Woo, S. L. C. (1985) *Biochemistry 24*, 556–561.

Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) p 123, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G., & Claeys, H. (1976) in *Proteolysis and Physiological Regulation* (Robbins, D. W., & Brew, K., Eds.) pp 203–238, Academic Press, New York.

Malinowski, D. P., Sadler, J. E., & Davie, E. W. (1984) *Biochemistry 23*, 4243–4250.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, pp 371–372, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

McMullen, B. A., & Fujikawa, K. (1985) *J. Biol. Chem. 260*, 5328–5341.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.

Petersen, T. E., Thogersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L., & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 137–141.

Pohl, G., Kallstrom, M., Bergsdorf, N., Wallen, P., & Jornvall, H. (1984) *Biochemistry 23*, 3701–3707.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Ratnoff, O. D., & Colopy, J. E. (1955) *J. Clin. Invest. 34*, 602–613.

Revak, S. D., Cochrane, C. G., Johnson, A. R., & Hugli, T. E. (1974) *J. Clin. Invest. 54*, 619–627.

Skorstengaard, K., Thogersen, H. C., & Petersen, T. E. (1984) *Eur. J. Biochem. 140*, 235–243.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Steffens, G. J., Gunzler, W. A., Otting, F., Frankus, E., & Flohe, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 1043–1058.

Sudhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science (Washington, D.C.) 228*, 815–822.

Verde, P., Stoppelli, M. P., Galeffi, P., & Di Nocera, P. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4727–4731.

Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.

Young, R. A., & Davis, R. W. (1983a) *Science (Washington, D.C.) 222*, 778–782.

Young, R. A., & Davis, R. W. (1983b) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.