Nature (London) 297, 655–659.
Lewin, B. (1980) in *Gene Expression*, Vol. 2, pp 824–826, Wiley-Interscience, New York.
Lodish, H. F. (1971) *J. Biol. Chem. 246*, 7131–7138.
Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.
Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–702.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, A Laboratory Manual*, pp 65–66, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Nevins, J. R., & Darnell, J. E. (1978) *Cell (Cambridge, Mass.) 15*, 1477–1493.
Nevins, J. R., Blanchard, J. M., & Darnell, J. E. (1980) *J. Mol. Biol. 144*, 377–386.
Nickerson, J. M., & Fuller, G. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 303–307.
Palmiter, R. D. (1974) *Biochemistry 13*, 3606–3615.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Radloff, R., Bauer, W., & Vinograd, J. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1514–1521.
Rixon, M. W., Chan, W.-Y., Davie, E. W., & Chung, D. W. (1983) *Biochemistry* (first paper of three in this issue).
Setzer, D. R., McGrogan, M., Nunberg, J. H., & Schimke, R. T. (1980) *Cell (Cambridge, Mass.) 22*, 361–370.
Setzer, D. R., McGrogan, M., & Schimke, R. T. (1982) *J. Biol. Chem. 257*, 5143–5147.
Smith, G. E., & Summers, M. D. (1980) *Anal. Biochem. 109*, 123–129.
Southern, E. (1975) *J. Mol. Biol. 98*, 503–517.
Töpfer-Petersen, E., Lottspeich, F., & Henschen, A. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 1509–1513.
Tosi, M., Young, R. A., Hagenbüchle, O., & Schibler, U. (1981) *Nucleic Acids Res. 9*, 2313–2323.
Uzan, G., Ardaillou, N., Larrieu, M. J., Kahn, A., & Marguerie, G. (1981) *Thromb. Haemostasis 46*, 180.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.
Watt, K. W. K., Takagi, T., & Doolittle, R. F. (1979) *Biochemistry 18*, 68–76.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 381–395.
Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., & Treiber, G. (1970) *Virology 40*, 734–744.

Characterization of a Complementary Deoxyribonucleic Acid Coding for the γ Chain of Human Fibrinogen†

Dominic W. Chung, Wai-Yee Chan,‡ and Earl W. Davie*

ABSTRACT: A number of cDNAs coding for the γ chain of human fibrinogen have been isolated from a liver cDNA library by employing a synthetic nucleotide mixture as a probe. One of the positive clones was then employed to screen the entire cDNA library of 18 000 recombinants, yielding 320 positive clones for the γ chain. The largest cDNA was 1638 base pairs in length and contained 10 base pairs of poly(G) at the 5′-end followed by 71 base pairs of noncoding nucleotides. The next 78 base pairs coded for a leader sequence that was 26 amino acids in length and included a methionine start signal and a typical hydrophobic core. The following 1233 base pairs coded for 411 amino acids that are present in the mature protein followed by a stop codon of TAA, 207 base pairs of noncoding nucleotides, a poly(A) track of 15 base pairs, and 22 base pairs of poly(C). Specific regions of the cDNA of the γ chain were then compared with the cDNAs for the α and β chains of human fibrinogen.

The γ chain of human fibrinogen (M_r 46 500) consists of 411 amino acids (Lottspeich & Henschen, 1977; Henschen & Lottspeich, 1977) and a single carbohydrate chain linked to Asn-52 (Iwanaga et al., 1968). Two γ chains, together with two α and two β chains, are cross-linked by intrachain disulfide bonds and form a distended polydomainal structure containing three nodules. An internal portion of the γ chain (amino acid residues 24–134) is bounded by unique cystine residues and is located in the interdomainal connecting structure linking the three nodules (Doolittle et al., 1978). The carboxyl-terminal region of the γ chain participates in fibrin polymerization (Olexa & Budzynski, 1981) and in intermolecular covalent cross-linking catalyzed by factor $XIII_a$ (Chen & Doolittle, 1970).

The complete amino acid sequence of the γ chain of human fibrinogen has been determined by amino acid sequencing techniques (Lottspeich & Henschen, 1977). This chain contains the unique sequence of Trp-Trp-Met-Asn-Lys starting with Trp-334. From this amino acid sequence, it can be predicted that the mRNA for the γ chain should contain the tetradecamer sequence of 5′-UGG-UGG-AUG-AA$^{U}_{C}$-AA-3′. By using the approach of Wallace et al. (1981) and Suggs et al. (1981), a DNA mixture complementary to these two tetradecamers has been employed as a hybridization probe for the identification of cDNAs coding for the γ chain of human fibrinogen. The largest cDNA that was identified was then

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received February 14, 1983.* This work was supported in part by Grants HL 16919 and HL 28598 from the National Institutes of Health. D.W.C. is an Established Investigator of the American Heart Association.

‡ Present address: Department of Pediatrics, University of Oklahoma, Oklahoma City, OK 73190.

0006-2960/83/0422-3250$01.50/0

sequenced, and its properties were compared with human cDNAs coding for the α and β chains of human fibrinogen (Rixon et al., 1983; Chung et al., 1983).

Experimental Procedures

A mixture of synthetic tetradecamers with the specific sequences of 5′-TT$^{G}_{A}$TTCATCCACCA-3′ was purchased from P-L Biochemicals. T4 polynucleotide kinase polymerase I, the Klenow fragment of polymerase I, and restriction endonucleases were purchased from Bethesda Research Laboratories, New England Biolabs, or Amersham, and ^{32}P-labeled nucleotides were purchased from New England Nuclear.

Labeling of Oligonucleotides. Oligonucleotides were labeled at the 5′-end to high specific radioactivity for use as hybridization probes by transfer of [^{32}P]phosphate from [γ-^{32}P]ATP in the presence of T4 polynucleotide kinase. The radiolabeled oligonucleotides were removed from unreacted [γ-^{32}P]ATP by gel filtration on Sephadex G-50 superfine (Wallace et al., 1981). Specific activities of ~5 × 10^8 cpm/μg of synthetic nucleotide were routinely attained.

Screening of cDNA Clones. A collection of transformed *Escherichia coli*, carrying recombinant plasmids of human liver cDNA inserts in the *Pst*I site of plasmid pBR322, was kindly provided by Drs. S. L. C. Woo and T. Chandra (Chandra et al., 1983). The collection of transformants was plated, transferred to Whatman 541 filter paper, amplified, and prepared for hybridization by the method of Gergen and coworkers (Gergen et al., 1979). Hybridization to the labeled synthetic oligonucleotide mixture was performed at 34 °C in buffer containing 90 mM Tris-HCl,[1] pH 7.5, 1 mM EDTA, and 0.9 M NaCl with 0.5% detergent NP40 as described by Wallace and co-workers (Wallace et al., 1981). After hybridization, the filters were washed at room temperature with four changes of the same buffer. The filters were then dried and exposed to X-ray films for about 5 h with intensifying screens. Optimum hybridization conditions were obtained by using about 1.5 ng of labeled probe per filter and a hybridization temperature of 34 °C.

Subsequent screening of the cDNA collection with radiolabeled cDNA for the γ chain was performed by the same procedure except that the hybridization temperature was raised to 68 °C. DNA sequence determinations were performed by the method of Maxam & Gilbert (1980) as described (Rixon et al., 1983).

Containment. Experiments were performed in compliance with N1H Guidelines for Recombinant DNA Research.

Results

A synthetic mixture of two oligonucleotides (5′-TT$^{G}_{A}$TTCATCCACCA-3′) was radiolabeled and used as a probe to screen approximately 1000 recombinants of a human liver cDNA library. The probe was 14 nucleotides in length and was employed under stringent conditions of hybridization. A total of 18 positive colonies were identified, and small amounts of plasmid DNA from each colony were isolated. Preliminary studies by restriction mapping indicated that most of the cDNA inserts shared a common restriction map. The plasmid containing the longest insert was designated pHIγ1[2] and was further examined by restriction mapping (Figure 1)

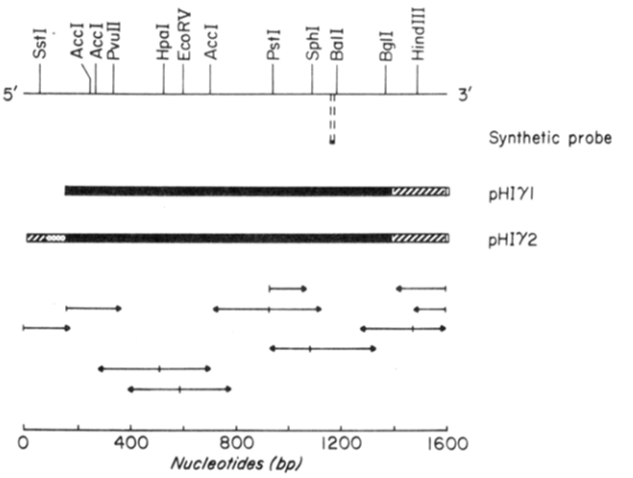

FIGURE 1: Restriction map and sequencing strategy of two cDNAs (pHIγ1 and pHIγ2) for the γ chain of human fibrinogen. The two cDNAs are represented by slashed bars, the signal peptide is represented by a dotted bar, the coding regions are represented by solid bars, and the poly(A) extensions are represented by open bars. Sites of labeling are indicated by vertical lines, and the arrows indicate the direction and extent of sequence determined. The dotted line below the restriction map shows the site of hybridization for the synthetic probe.

and sequence analyses. The cDNA insert from this plasmid contained a single internal *Pst*I site near the middle. Accordingly, cleavage of the plasmid by *Pst*I gave rise to two fragments. The fragment from the 5′-end was 770 base pairs in length, and the fragment from the 3′-end was 700 base pairs in length. A sequence analysis of the 5′-fragment indicated that the insert coded for the γ chain of human fibrinogen and extended to Val residue 2 in the amino-terminal end of the γ chain. Sequence analyses on the 3′-fragment showed that the 3′ noncoding region was 207 base pairs in length and was followed by a poly(A) tail.

A cDNA fragment from the 3′-end was then isolated by digestion with *Pst*I and *Hin*dIII. This fragment, which was 554 base pairs in length, was free of poly(G/C) and poly(A) sequences and was used as a hybridization probe to screen the entire human cDNA collection of 18 000 recombinant plasmids. A total of 320 positive clones were identified in this screening. Also, 17 of the original 18 positive colonies that were identified by the synthetic probe were confirmed to be cDNA clones for the γ chain of human fibrinogen. Thus, the conditions of hybridization with the synthetic probes were such that hybrids with mispaired bases were unstable.

Isolates of cDNA for the γ chain were arbitrarily grouped into subsets according to the intensity of their hybridization signal. Representative plasmids from each group were then examined, and plasmid pHIγ2, which extended farthest into the 5′-end of the insert for the γ chain, was isolated and characterized. The restriction map for this plasmid showed that it overlapped with plasmid pHIγ1 but extended approximately 160 base pairs upstream from this plasmid (Figure 1). The complete nucleotide sequence and the predicted amino acid sequence derived from plasmid pHIγ2 were then determined (Figure 2). The insert was 1638 base pairs in length, including poly(G) and poly(C) linkers of 10 and 22 base pairs at the 5′- and 3′-ends, respectively. The cDNA included 71 base pairs of noncoding nucleotides at the 5′-end followed by 78 nucleotides coding for a signal peptide of 26 amino acids. Two stop condons were identified upstream from the signal peptide, starting at nucleotides 33 and 42. This indicates that the codon for the methionine at position −26 is the correct initiation codon. The leader sequence is a typcial signal peptide

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane.

[2] The nomenclature of pHIγ1 is as follows: p, plasmid; H, human, I, fibrinogen; γ, γ chain; 1, first plasmid identified.

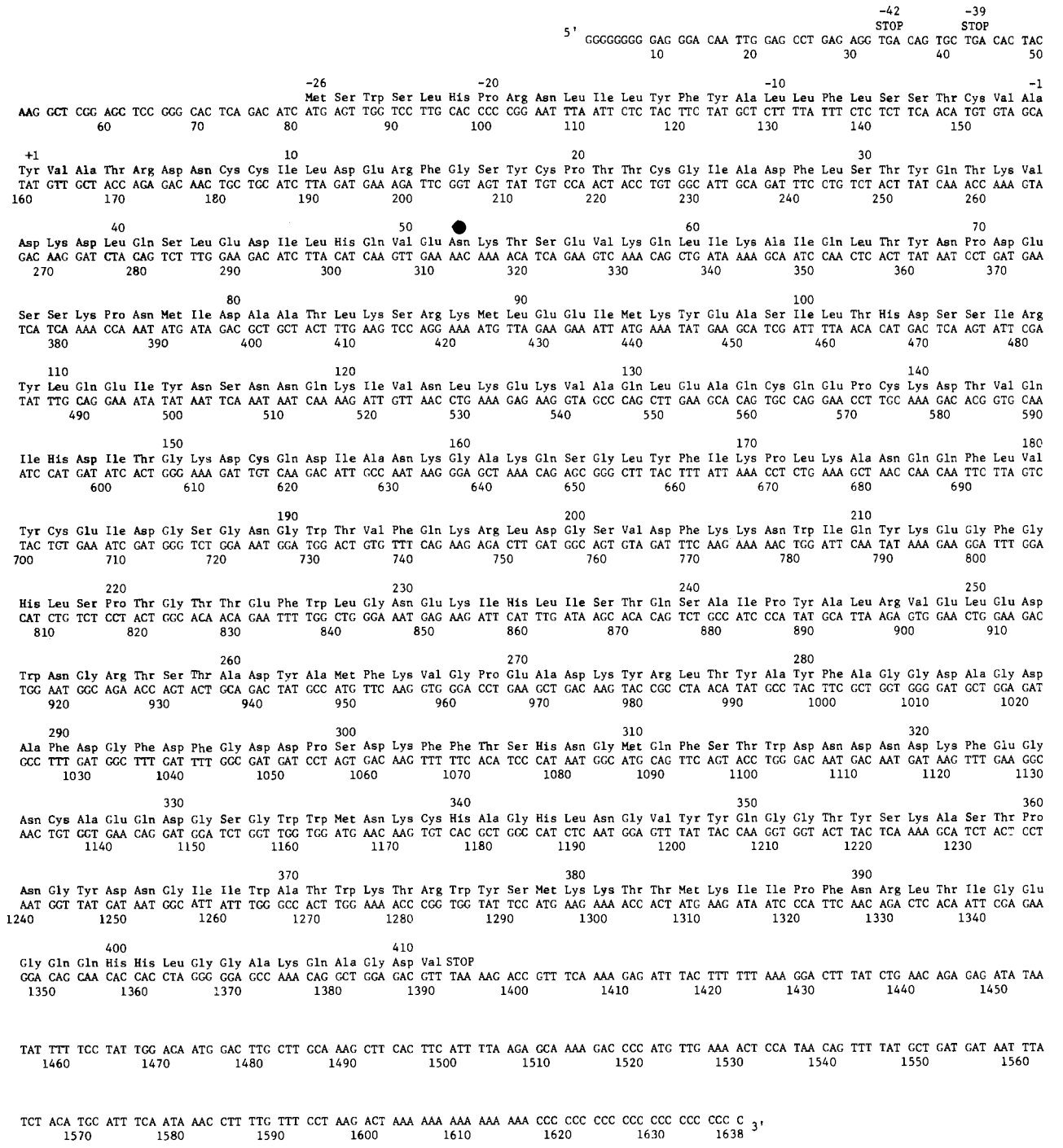

FIGURE 2: Complete nucleotide sequence of insert of plasmid pHIγ2 for the γ chain of human fibrinogen. The nucleotide sequence of the coding strand along with the predicted amino acid sequence are shown. Amino acid residues −26 to −1 represent the signal peptide. The attachment for carbohydrate on Asn-52 is indicated by (♦).

containing a hydrophobic core rich in leucine, phenylalanine, isoleucine, and tyrosine residues and an uncharged amino acid (Ala) at the cleavage site for signal peptidase (Blobel et al., 1979). It showed very little homology or similarity to the signal peptide for the α and β chains (Rixon et al., 1983; Chung et al., 1983) in terms of size or specific amino acid sequences (Figure 3).

The coding sequence for the 411 amino acid residues present in the mature γ chain was 1233 nucleotides in length and was followed by a stop codon of TAA, at 3′ noncoding region of 207 bases, and a poly(A) extension of 15 base pairs. The amino acid composition of the mature γ chain as predicted from the cDNA sequence was determined as follows: ASP_{32}, Asn_{23}, Thr_{28}, Ser_{25}, Glu_{22}, Pro_{11}, Gly_{35}, Ala_{26}, Val_{14}, Met_8, Ile_{25}, Leu_{26}, Tyr_{20}, Phe_{18}, Lys_{34}, His_{10}, Arg_{10}, $^1/_2$-Cys_{10}, and Trp_{10}. The molecular weight for the protein free of carbohydrate was calculated as 46 406. This is equivalent to a molecular weight

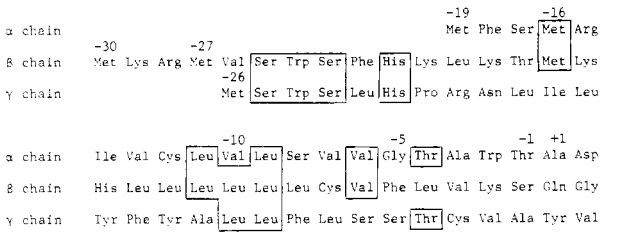

FIGURE 3: Comparison of amino acid sequences of signal peptides of the α, β, and γ chains of human fibrinogen.

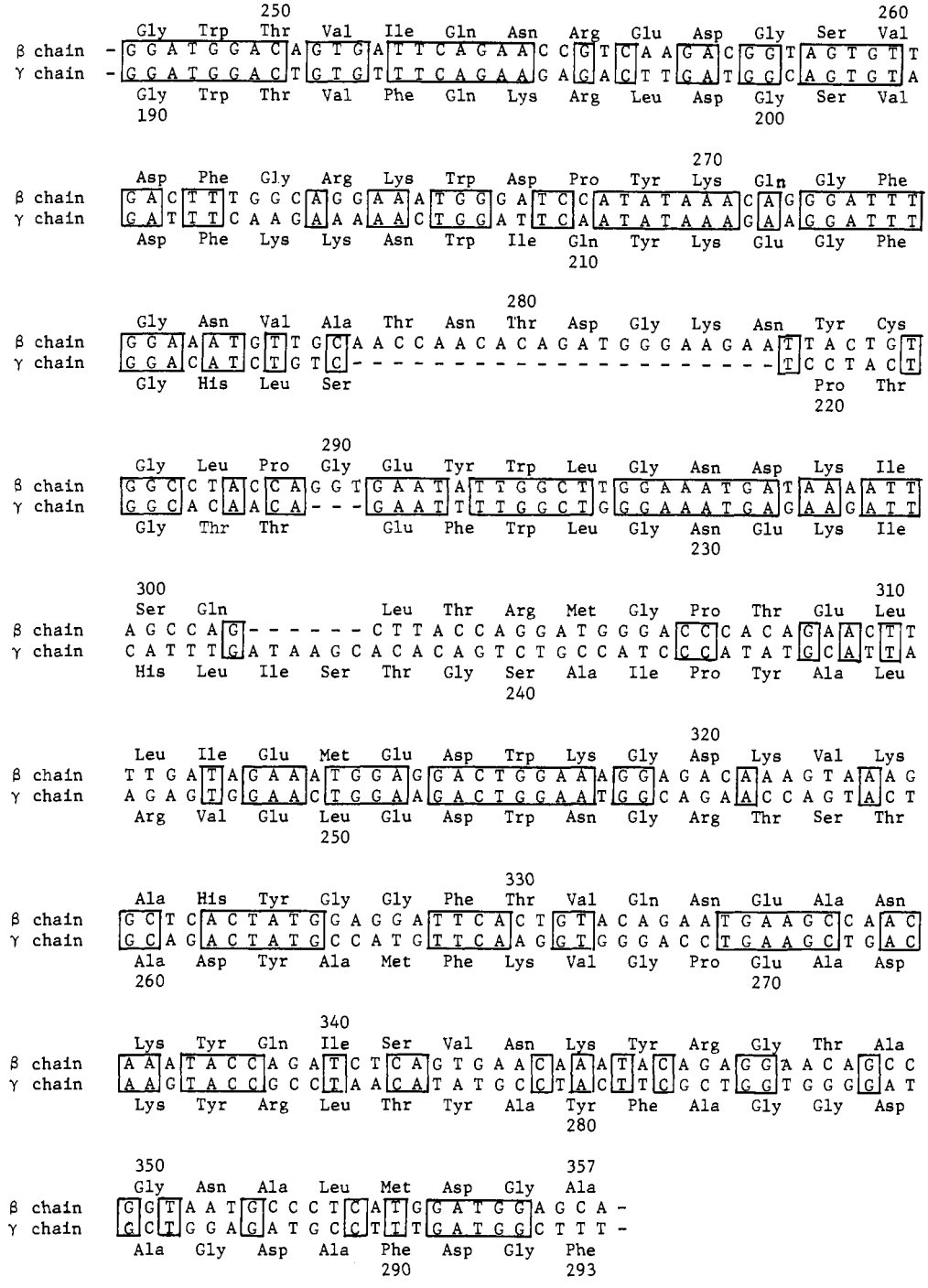

FIGURE 4: Comparison of a portion of nucleotide sequence of cDNAs for the β and γ chains of human fibrinogen. The cDNA sequence for the β chain includes nucleotides 822–1151. For the γ chain, it includes nucleotides 726–1037. The data for the β chain were from Chung et al. (1983).

of 48 906 upon the addition of 2500 g of carbohydrate. Only one potential amino acid sequence for the attachment of a carbohydrate chain is present in the γ chain, starting with residue 52 (Iwanaga et al., 1968).

Analyses of the codon usage for the γ chain as well as the α and β chains (Rixon et al., 1983; Chung et al., 1983) showed two interesting aspects (Table I). First, the pattern is biased and is similar to those of other human proteins (Chen et al., 1982). Codons containing the dinucleotide CG (for example, UCG for serine, ACG for threonine, GCG for alanine, and CCG for proline, as well as CGN for arginine) are least utilized. This is consistent with the observation that a low frequency of occurrence of the dinucleotide combination CG is universal among eucaryotic genes (Nussinov, 1981). Second, the coding portions of the human γ chain have a significantly higher A+T content than that of the rat (59 vs. 51%) (Table II). However, the amino acid sequences are highly conserved. In order to minimize the amount of change reflected in the amino acid sequence, this shift in A+T content is accommodated mainly in the third position of codons. As shown in Table II, A+T residues account for 61% of all the base pairs used in the third position of codons in the human γ chain, while in the rat it is only 43%.

The amino acid sequence of the mature γ chain as predicted from the cDNA was in complete agreement with that reported by Lottspeich & Henschen (1977). The 3′ noncoding region is rich in A and T residues (68%), and the putative polyadenylation signal (AATAAA) is located 19 base pairs from

Table I: Codon Usage for the α, β, and γ Chains of Human Fibrinogen

		α	β	γ
Ala	GCG	1	0	0
	GCA	6	5	8
	GCT	7	9	12
	GCC	10	9	8
Arg	AGG	12	9	1
	AGA	14	10	6
	CGG	0	1	2
	CGA	6	1	1
	CGT	7	1	0
	CGC	3	5	1
Asn	AAT	13	21	16
	AAC	16	11	8
Asp	GAT	23	12	18
	GAC	12	16	13
Cys	TGT	2	10	7
	TGC	7	2	4
Gln	CAG	11	14	13
	CAA	7	12	11
Glu	GAG	17	9	2
	GAA	27	21	20
Gly	GGG	15	3	5
	GGA	30	18	14
	GGT	14	12	10
	GGC	13	9	6
His	CAT	8	4	7
	CAC	8	5	4
Ile	ATA	6	3	5
	ATT	8	6	13
	ATC	5	7	8
Leu	TTG	7	7	6
	TTA	7	5	8
	CTG	9	7	7
	CTA	4	7	3
	CTT	4	8	4
	CTC	2	2	5
Lys	AAG	16	14	14
	AAA	24	21	19
Met	ATG	12	16	9
Phe	TTT	11	4	10
	TTC	9	8	10
Pro	CCG	1	2	0
	CCA	10	10	4
	CCT	20	5	7
	CCC	7	6	1
Ser	AGT	19	7	7
	AGC	19	5	3
	TCG	1	0	1
	TCA	12	7	7
	TCT	31	8	8
	TCC	9	5	4
Thr	ACG	3	1	1
	ACA	13	9	9
	ACT	19	5	12
	ACC	15	8	7
Trp	TGG	11	13	11
Tyr	TAT	4	11	15
	TAC	5	10	7
Val	GTG	9	8	4
	GTA	5	8	4
	GTT	9	6	5
	GTC	9	5	2
stop	TGA	0	0	0
	TAG	1	1	0
	TAA	0	0	1

the poly(A) extension (Proudfoot & Brownlee, 1976).

Discussion

The cDNAs for the γ chain of fibrinogen constitute approximately 1.7% of the cDNA collection employed in these studies. Together with those identified for the α chain (1.3%) (Rixon et al., 1983) and the β chain (2.5%) (Chung et al., 1983), fibrinogen cDNAs comprise approximately 5.6% of the human liver cDNA collection. If one assumes that the frequency of occurrence of cDNAs in a large collection reflects the relative abundance of the corresponding mRNAs, the percentage of fibrinogen clones correlates well with the estimated total mRNA level of 5.8% determined by cell-free translation of total liver mRNA and specific immunoprecipitation of the translation products (D. W. Chung, unpublished results).

Table II: Distribution of Bases in Three Positions of Codons in Coding Regions of cDNAs for the γ Chains of Human and Rat Fibrinogen

	human				rat[a]			
position:	first	second	third	total	first	second	third	total
G	131	78	76	285	125	87	108	320
A	138	168	120	426	141	163	80	384
T	99	103	147	349	93	96	109	298
C	70	89	95	254	79	92	141	312

	human	rat
A+T in first position	54%	53%
A+T in second position	62%	59%
A+T in third position	61%	43%
A+T overall	59%	52%

[a] Data from Crabtree & Kant (1982).

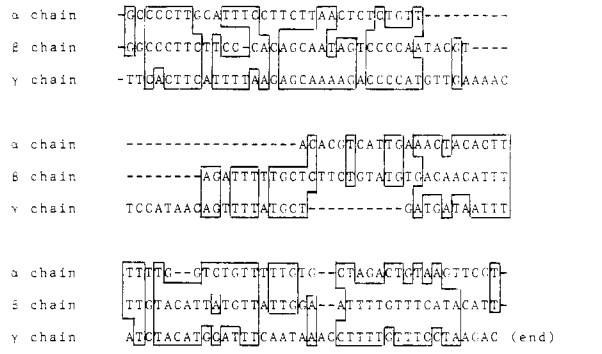

FIGURE 5: Comparison of nucleotide sequences in the 3′ noncoding regions of cDNAs for the α, β, and γ chains of human fibrinogen. The cDNA sequences for the α, β, and γ chains include nucleotides 2010–2097, 1444–1547, and 1492–1601, respectively.

Comparisons of the amino acid sequences of the α, β and γ chains showed that the three chains are related and probably have evolved from a common ancestor (Doolittle et al., 1979; Doolittle, 1980; Henschen et al., 1980). The β and the γ chains, however, share a significantly higher degree of homology with each other than with the α chain. It has been estimated that the divergence of the β and γ chains occurred about 500 million years ago, while that of the α chain occurred about 1500 million years ago (Henschen et al., 1980; Doolittle, 1980). The characterization of a cDNA for each of the three chains of human fibrinogen makes it possible to compare the signal sequences as well as the nucleotide sequences for the three chains.

The amino-terminal region containing the signal peptide region, the disulfide knot region, and the region forming the α helix are the most variable portions of the three chains in amino acid sequence and nucleotide sequence. For instance, only a small degree of homology in the nucleotide sequence ($\sim$30%) exists among the corresponding regions of the three chains that participate in interdomainal superhelix formation. Apparently, the evolutionary constraint imposed on the fibrinogen structure in this region is fairly relaxed, so that amino acid replacements that do not disrupt α helix formation and substitutions that do not affect potential sites for plasmin hydrolysis are permitted.

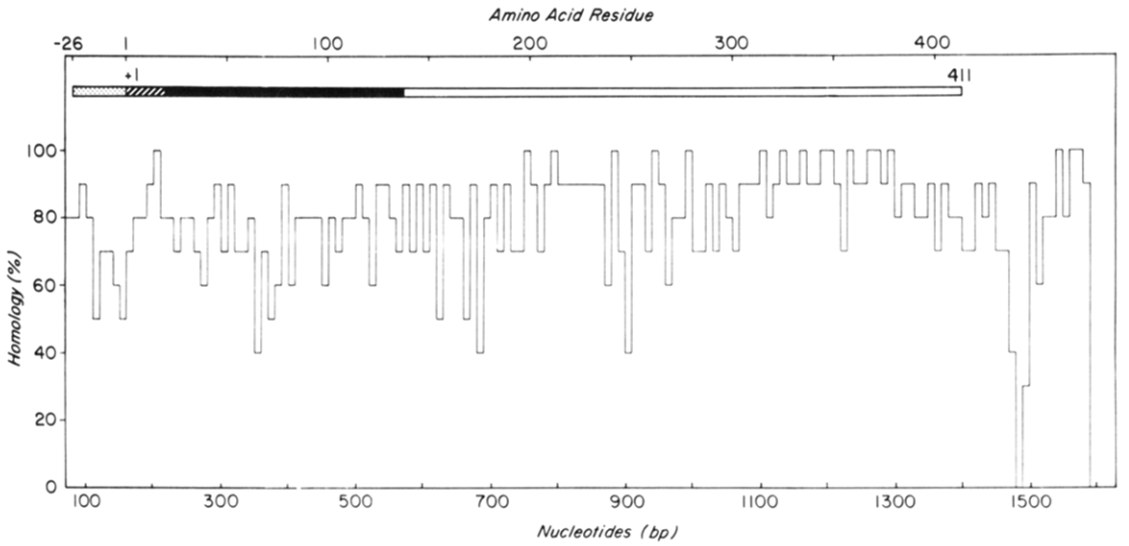

FIGURE 6: Comparison of nucleotide sequences of cDNAs for the γ chains of human and rat fibrinogen. Consecutive segments of 10 nucleotides from both sequences were compared, and the percent identity was recorded. The bar above the histogram represents the γ chain. The dotted bar represents the signal peptide (residues −26 to −1), the slashed bar represents the amino-terminal portion of the molecule (residues 1–18), the solid bar represents the "coiled-coil" domain, including the flanking disulfide rings (residues 19–139) (Doolittle et al., 1978), and the open bar represents the segment that constitutes a portion of the D domain (residues 140–411) (Marder et al., 1969). The sequence for the rat γ chain was taken from Crabtree & Kant (1982).

As predicted from amino acid sequence comparison, the β and the γ chains show strong homology in certain portions of their chains that contribute to the D domain (Doolittle et al., 1979). Figure 4 shows an alignment of the two nucleotide sequences from which strong homology can be observed in a stretch of over 300 base pairs in both sequences. This homology in DNA sequence approaches 75–80% in certain short regions. This provides supportive evidence for some evolutionary relatedness of the two chains. Despite extensive changes that have occurred in other parts of the molecule, this region, which immediately follows the helical domain, is rather highly conserved.

The 3′ noncoding sequences of related genes tend to diverge faster than protein coding sequences and often have insertions and deletions in addition to base substitutions (Konkel et al., 1979). This is apparent in the comparison of the 3′ noncoding regions of the α, β and γ chains. An obvious difference is in the length of the 3′ noncoding regions and the use of polyadenylation signal. Because of the high content of A and T residues in this region, short spurious stretches of homology can be found among the three chains and can lead to an exaggerated estimation of homology. However, a significant extended region of homology occurs between the β and γ chains where an identity of more than 52% over a significant portion of the 3′ noncoding region occurs when the two are adjusted for optimal homology (Figure 5).

A comparison of the nucleotide sequences of the human and rat γ chains presents another way of examining the changes that have occurred in the evolution of the γ chain since its divergence from the β chain. Figure 6 shows the variation in the degree of homology along the two nucleotide sequences. It is evident from this comparison that regions of high homology are localized, particularly nucleotides 700–1200. Regions that changed most extensively from the β chain, namely, the signal peptide and the helical-coil region (nucleotides 80–550), are regions that diverged the most between the human and rat γ chains.

Variant forms of the γ chain differing in charge (Mosesson et al., 1972), in sialic acid content (Gati & Straub, 1978), and in size (Henschen & Edman, 1972; Mosher & Blout, 1973) have been reported. More recently, a functionally normal high molecular weight form (γ_B or γ′) from human fibrinogen has been chracterized (Francis et al., 1980; Wolfenstein-Todel & Mosesson, 1980, 1981). In the case of rat fibrinogen, alternative splicing of the mRNA for the γ chain has been shown to result in the formation of a different mRNA coding for the variant form (Crabtree & Kant, 1981, 1982). Occasional omission of the splicing of the seventh or last intervening sequence leads to the formation of a longer message. Continuous translation from the seventh exon into the seventh intervening sequence resulted in the substitution of the last four amino acids of the regular γ chain by a peptide of 12 amino acids in the γ′ form. It appears likely that a similar situation exists in human fibrinogen. Thus far, however, cDNA clones for a human γ′ chain have not been identified.

Acknowledgments

We thank Judy Morishima and Brad McMullen for excellent technical assistance and Drs. Kazuo Fujikawa, Benito Que, and Mark Rixon for helpful discussions.

Registry No. DNA (human liver clone pHIγ2 fibrinogen γ-chain-specifying mRNA complementary), 85882-39-9; fibrinogen (human liver clone pHIγ2 γ-chain precursor protein moiety reduced), 85882-42-4; fibrinogen (human liver clone pHIγ2 γ-chain protein moiety reduced), 85882-45-7.

References

Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H., & Lingappa, R. (1979) in *Secretory Mechanisms* (Hopkin, C. R., & Duncan, C. J., Eds.) Vol. 33, pp 9–36, Cambridge University Press, London.

Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1845–1848.

Chen, H. R., Dayhoff, M. O., Barker, W. C., Hunt, L. T., Yeh, L. S., George, D. G., & Orcutt, B. C. (1982) *DNA 1*, 365–374.

Chen, R., & Doolittle, R. F. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66*, 472–479.

Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., & Davie, E. W. (1983) *Biochemistry* (second paper of three in this issue).

Crabtree, G. R., & Kant, J. A. (1981) *J. Biol. Chem. 256*, 9718–9723.

Crabtree, G. R., & Kant, J. A. (1982) *Cell (Cambridge, Mass.) 31*, 159–166.
Doolittle, R. F. (1980) in *Protides of Biological Fluids, Proceedings of the 28th Colloquium* (Peeters, H., Ed.) pp 41–46, Pergamon Press, New York.
Doolittle, R. F., Goldbaum, D. M., & Doolittle, L. R. (1978) *J. Mol. Biol. 120*, 311–325.
Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D., & Riley, M. (1979) *Nature (London) 280*, 464–468.
Francis, C. W., Marder, V. J., & Martin, S. E. (1980) *J. Biol. Chem. 255*, 5599–5604.
Gati, W. P., & Straub, P. W. (1978) *J. Biol. Chem. 253*, 1315–1321.
Gergen, J. P., Stern, R. H., & Wensink, P. C. (1979) *Nucleic Acids Res. 7*, 2115–2136.
Henschen, A., & Edman, P. (1972) *Biochim. Biophys. Acta 263*, 351–367.
Henschen, A., & Lottspeich, F. (1977) *Thromb. Res. 11*, 869–880.
Henschen, A., Lottspeich, F., Topfer-Petersen, E., Kehl, M., & Timpl, R. (1980) in *Protides of Biological Fluids, Proceedings of the 28th Colloquium* (Peeters, H., Ed.) pp 47–50, Pergamon Press, New York.
Iwanaga, S., Blömback, B., Gröndahl, N. J., Hessel, B., & Wallen, P. (1968) *Biochim. Biophys. Acta 160*, 280–283.
Konkel, D. A., Maizel, J. V., & Leder, P. (1979) *Cell (Cambridge, Mass.) 18*, 1273–1283.
Lottspeich, F., & Henschen, A. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 935–938.
Marder, V. J., Shulman, N. R., & Carroll, W. R. (1969) *J. Biol. Chem. 244*, 2111–2119.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Mosesson, M. W., Finlayson, J. S., & Umfleet, R. A. (1972) *J. Biol. Chem. 247*, 5223–5227.
Mosher, D. F., & Blout, E. R. (1973) *J. Biol. Chem. 248*, 6896–6903.
Nussinov, R. (1981) *J. Mol. Biol. 149*, 125–131.
Olexa, S. A., & Budzynski, A. Z. (1981) *Thromb. Haemostasis 46*, 161.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Rixon, M. W., Chan, W.-Y., Davie, E. W., & Chung, D. W. (1983) *Biochemistry* (first paper of three in this issue).
Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., & Itakura, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6613–6617.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.
Wolfenstein-Todel, C., & Mosesson, M. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5069–5073.
Wolfenstein-Todel, C., & Mosesson, M. W. (1981) *Biochemistry 20*, 6146–6149.

Novel Inactivators of Serine Proteases Based on 6-Chloro-2-pyrone†

Richard B. Westkaemper‡ and Robert H. Abeles*

ABSTRACT: The interaction of serine protease (esterases) with 6-chloro-2-pyrones was investigated. Time-dependent inactivation of chymotrypsin, α-lytic protease, pig liver elastase, and cholinesterase was found with 3- and 5-benzyl-6-chloro-2-pyrone, as well as 3- and 5-methyl-6-chloro-2-pyrone. No inactivation was observed with the unsubstituted 6-chloro-2-pyrone. The substituted pyrones did not inactivate papain or carboxypeptidase A, as well as a number of other nonproteolytic enzymes. The substituted chloropyrones, therefore, show considerable selectivity toward serine proteases. Analogues in which the 6-chloro substituent is replaced by H or OH do not inactivate. The presence of the halogen is, therefore, essential for inactivation. Chymotrypsin catalyzes the hydrolysis of 3-benzyl-6-chloro-2-pyrone. At pH 7.5, (*E*)-4-benzyl-2-pentenedioic acid is the major product, and 2-benzyl-2-pentenedioic anhydride is a minor product. The ratio of hydrolysis product found to the number of enzyme molecules inactivated varies from 14 to 40. The enzyme inactivated with the 3-benzyl compound does not show a spectrum characteristic of the pyrone ring. This suggests that inactivation by 3-benzyl-6-chloro-2-pyrone occurs in a mechanism-based fashion after enzymatic lactone hydrolysis. When the enzyme is inactivated with the 5-benzyl compound, absorbance due to the pyrone ring is observed. We suggest that inactivation occurs through an active site directed mechanism involving a 1,6-conjugate addition of an active site nucleophile to the pyrone ring.

The concept of suicide substrates or mechanism-based irreversible inactivators has led to the design of inhibitors for a number of enzyme classes (Abeles & Maycock, 1976; Rando, 1975; Seiler et al., 1978; Walsh, 1982). Inactivators of this type contain a latently reactive functional group that is unmasked only at the active site after enzymatic activation. Most mechanism-based irreversible inactivators described to date are for pyridoxal-dependent enzymes.

Serine proteases play a crucial role in many biochemical systems and disease states (Barrett, 1980). Human leucocyte elastase, for example, which causes the destruction of lung elastin, has been implicated in the development of emphysema (Mittman, 1972; Turino et al., 1974; Hance & Crystal, 1975; Boudier et al., 1981). For these reasons, we thought it desirable to explore the possibility of developing additional mechanism-based irreversible inactivators of serine proteases.

While this work was in progress, Chakravarty et al. (1982) reported the discovery of a new type of mechanism-based

† From the Graduate Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254. *Received October 12, 1982.* This is Publication No. 1447. Supported by a research grant to R.H.A. from the National Institutes of Health (GM 12633) and by a National Institutes of Health Postdoctoral Fellowship to R.B.W. (GM 07939).

‡ Present address: Department of Pharmaceutical Chemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23219.

0006-2960/83/0422-3256$01.50/0