Structure of the Gene for Human Coagulation Factor V[†,‡]

Larry D. Cripe,[§] Karen D. Moore, and William H. Kane*,[∥]

Division of Hematology–Oncology, Departments of Medicine and Pathology, Box 3656 Duke University Medical Center, Durham, North Carolina 27710

Received December 13, 1991; Revised Manuscript Received February 5, 1992

ABSTRACT: Activated factor V (Va) serves as an essential protein cofactor for the conversion of prothrombin to thrombin by factor Xa. Analysis of the factor V cDNA indicates that the protein contains several types of internal repeats with the following domain structure: A1–A2–B–A3–C1–C2. In this report we describe the isolation and characterization of genomic DNA coding for human factor V. The factor V gene contains 25 exons which range in size from 72 to 2820 bp. The structure of the gene for factor V is similar to the previously characterized gene for factor VIII. Based on the aligned amino acid sequences of the two proteins, 21 of the 24 intron–exon boundaries in the factor V gene occur at the same location as in the factor VIII gene. In both genes, the junctions of the A1–A2 and A2–A3 domains are each encoded by a single exon. In contrast, the boundaries between domains A3–C1 and C1–C2 occur at intron–exon boundaries, which is consistent with evolution through domain duplication and exon shuffling. The connecting region or B domain of factor V is encoded by a single large exon of 2820 bp. The corresponding exon of the factor VIII gene contains 3106 bp. The 5′ and 3′ ends of both of these exons encode sequences homologous to the carboxyl-terminal end of domain A2 and the amino-terminal end of domain A3 in ceruloplasmin. There is otherwise no homology between the B domain exons. These data provide further insight into the evolutionary relationships within this family of related plasma proteins and provide a basis from which to begin the investigation of the cellular regulation of factor V biosynthesis and characterization of molecular defects in congenital factor V deficiency.

The generation of thrombin by the prothrombinase complex is a critical event in hemostasis and thrombosis. This complex consists of factor Xa, factor Va, prothrombin, calcium, and a phospholipid or cellular surface. Factor Va participates as an essential protein cofactor in the activation of prothrombin by factor Xa (Kane & Davie, 1988). Most of the factor V in whole blood circulates in the plasma; however, 25% of the factor V in blood is stored in platelet α-granules (Tracy et al., 1982). Platelet factor V is released during platelet activation and appears to be important for platelet surface prothrombin activation and normal hemostasis (Miletich et al., 1978). The major sites of factor V biosynthesis appear to include both the liver (Wilson et al., 1984) and megakaryocytes (Gewirtz et al., 1986). We (Kane & Davie, 1986; Kane et al., 1987) and others (Jenny et al., 1987) have previously determined the predicted amino acid sequence of human factor V by cDNA cloning. The mature protein consists of 2196 amino acids. Analysis of the factor V cDNA indicates that the protein contains several types of internal repeats organized with the following domain structure: A1–A2–B–A3–C1–C2. The A, B, and C domains each contain approximately 350, 836, and 150 amino acids, respectively. The primary sequence of human factor V is 40% identical to human coagulation factor VIII except in the B domain where there is no homology (Toole et al., 1984; Vehar et al., 1984). The role of factor VIIIa in the activation of factor X is analogous to the role of factor Va in prothrombin activation. Thus, the factor X activation complex consists of factor IXa, factor VIIIa, factor X, calcium, and a phospholipid or cellular surface. The A domains present in factor V and factor VIII are approximately 30% identical to the triplicated A domains present in the plasma copper-binding protein ceruloplasmin (Ortel et al., 1984). The C domains in factor V and factor VIII are also approximately 40% identical to the duplicated C domains present in a recently characterized murine breast epithelial cell protein (Stubbs et al., 1990).

The amino acid sequence homology among this family of proteins suggests that the genes for these proteins have evolved through a process of gene duplication and exon shuffling (Gilbert, 1985). The gene for factor VIII has been characterized (Gitschier et al., 1984), and it spans >180 kb on human chromosome Xq28. One striking feature of the factor VIII gene is that the entire B domain or connecting region is encoded by a particularly large exon of 3106 bp. The gene for human factor V has been localized to human chromosome 1q21-25 (Wang et al., 1988). Recent gene mapping studies (Watson et al., 1990) indicate that the gene for human factor V is located within a 300-kb region that also includes the genes for the selectin family of leukocyte adhesion molecules. We now report the cloning and characterization of the gene for human factor V. Our data provide evidence that gene duplication and exon shuffling occurred during the evolution of this family of related plasma proteins and provide a basis from which to begin the investigation of the cellular regulation of factor V biosynthesis and characterization of molecular defects in congenital factor V deficiency (parahemophilia).

Materials and Methods

Materials. All restriction enzymes and DNA-modifying enzymes were purchased from Bethesda Research Laboratories or United States Biochemicals. ^{32}P-Labeled nucleotides and

[†] This work was supported by Grants HL43106 and DK01965 from the National Institutes of Health, Searle Scholar Award 89-G-130, and Established Investigator Award 91-181 from the American Heart Association and Smith-Kline Beecham Inc., with funds contributed in part by the North Carolina Affiliate of the American Heart Association.

[‡] The genetic sequence for factor V given in this paper has been submitted to GenBank under Accession Number J05368.

[§] Supported by NIH Program Physician Scientist Award.

[∥] AHA Established Investigator; 1989 Searle Scholar.

0006-2960/92/0431-3777$03.00/0 © 1992 American Chemical Society

[α-^{35}S]dATPαS were obtained from Amersham. The Sequenase Version 2.0 DNA sequencing kit was obtained from United States Biochemicals. DNA amplification kits were obtained from Perkin-Elmer Cetus.

Isolation of Genomic Clones for Factor V. A human lymphocyte genomic library in λEMBL 3 (Clontech) and a human lung fibroblast genomic library in λFIX (Stratagene) were screened by the in situ plaque hybridization method of Benton and Davis (1977). A full-length factor V cDNA probe (Kane et al., 1990) was used to screen approximately 1×10^6 recombinant phage from the λFIX library. Approximately 8×10^5 phage from the λEMBL3 library were screened in a similar fashion using a restriction fragment (*Eco*RI–*Acc*I) corresponding to nucleotides 1–269 of the factor V cDNA. The cDNA probes were labeled with [^{32}P]dCTP to a specific activity of $>10^8$ cpm/μg in the presence of random hexanucleotide primers and DNA polymerase I (Feinberg & Vogelstein, 1983). Positive genomic clones were plaque purified using standard methods (Sambrook et al., 1989).

Southern Blot Analysis of Genomic Clones and Human Genomic DNA. Phage DNA from the genomic clones and high molecular weight human genomic DNA from human peripheral blood leukocytes were isolated using standard methods (Sambrook et al., 1989). The DNA was digested with the restriction enzymes *Sst*I and/or *Xba*I, separated on 0.7% agarose gels, and transferred to nitrocellulose membranes (Southern, 1975). The membranes were then hybridized with ^{32}P-labeled factor V cDNA probes or oligonucleotides in order to determine the location of exons and to verify that no rearrangements occurred during the construction of the phage library. The filters were washed for 30 min at moderate (0.2× SSC,[1] 65 °C) or high (0.2× SSC, 70 °C) stringency, and autoradiographs were exposed using XAR-5 film with intensifying screens (Du Pont). Restriction fragments isolated from *Sst*I or *Xba*I digests of the genomics clones were subcloned into pBluescript (Stratagene). These plasmid subclones were then mapped by restriction enzyme digestion and Southern blotting. These data were used to construct a genomic restriction map for the enzymes *Sst*I and *Xba*I. The gene map was confirmed by characterization of at least two independent clones for the gene and by Southern blot analysis of human genomic DNA.

DNA Sequence Analysis of the Isolated Genomic Clones. DNA sequences were determined by the dideoxy chain termination method (Sanger et al., 1977) using either universal or synthetic oligonucleotide primers. Single-stranded templates were prepared using M13mp18/19 or pBluescript (Sambrook et al., 1989). Plasmids employed for double-stranded sequencing were purified by cesium chloride banding (Sambrook et al., 1989). Approximately 95% of the DNA sequence for exons and exon–intron boundaries were sequenced on both strands. Sequences that deviated from previously published cDNA sequences were verified on independent templates on both strands. DNA sequences were stored and analyzed using a Macintosh cx computer and MacVector version 3.5 software (IBI). In most cases, exons were aligned to the genomic *Xba*I and *Sst*I map using restriction sites present in the exon or exon–intron boundary sequence. The locations of exons 15–18 were determined by PCR (Saiki et al., 1988) using the following pairs of oligonucleotide primers: 5′-ATACCTACGTATGGCATG-3′ (exon 15) and 5′-GAGTCTCCAAGAACTTCG-3′ (exon 16); 5′-ATTATCAGAAGAGCAAGG-3′ (intron 16) and 5′-CAGATTGCCTTTTCCCTG-3′ (intron 17); 5′-GGGATGATCTACAGCTTG-3′ (exon 17) and 5′-TAGTAATAGGCACTCTCC-3′ (intron 18). Typically, reactions contained 100 ng of a bacteriophage clone template, 1 μM each oligonucleotide, and 2.5 units of *Taq*I polymerase. The DNA was amplified for 40 cycles of denaturation (1 min at 94 °C), annealing (2 min at 45–60 °C), and extension (2 min at 72 °C) using a Coy thermal cycler. The magnesium chloride concentration was varied between 1 and 6 mM in order to optimize results. Reaction products were analyzed on 1% agarose gels with ethidium bromide staining.

Results and Discussion

Isolation and Characterization of Factor V Genomic Clones. Forty-five genomic phage clones were isolated from the λFIX library using a ^{32}P-labeled full-length factor V cDNA probe. These clones contained inserts ranging from 12 to 30 kb. Preliminary analysis indicated that 24 of these isolates represented unique overlapping clones. Southern blotting and partial DNA sequence analysis indicated that the sequences 5′ to nucleotide 327 in the factor V cDNA were not present in these clones. Subsequently, five clones were isolated from the λEMBL3 library by hybridization with a probe containing nucleotides 1–269 of the factor V cDNA. Restriction and Southern analysis of these clones demonstrated that three were unique overlapping clones with inserts ranging from 12 to 16 kb. Partial DNA sequence analysis of these clones indicated that they contained the sequences corresponding to nucleotides 1–326 of the factor V cDNA. Hybridization experiments demonstrated no overlap between the two sets of overlapping clones (see below).

Structure of the Human Factor V Gene. Twenty of the independent genomic clones were chosen for detailed characterization. Restriction fragments from the eight clones shown in Figure 1 were also subcloned into pBluescript and characterized as described under Materials and Methods. Characterization of *Sst*I and or *Xba*I digests of these clones or plasmid subclones by Southern blotting provided the data used to construct a restriction map of the factor V gene for these enzymes. It is possible that some small genomic fragments (<250 bp) containing intron sequences may not have been detected in our analysis. Restriction mapping of the genomic clones indicated that the gene was comprised of at least 17 *Sst*I fragments ranging in size from 0.7 to 12 kb (Figure 1). Fourteen of these fragments contained exons as judged by Southern blotting experiments. No genomic clones completely spanned the second intron. The sizes of the cloned genomic *Sst*I fragments were consistent with the size of the fragments observed when human genomic DNA was probed with factor V cDNA probes. This indicates that no major rearrangements occurred during the construction of the genomic libraries. When genomic DNA was probed with factor V cDNA probes encoding the heavy chain (nt 1–2074) or the B domain (nt 2268–4785), only the predicted hybridizing fragments were observed consistent with a single copy of the factor V gene. However when genomic DNA was probed with probes encoding the light-chain region (nt 5001–6903, 6000–6590, or 6591–6903), several additional hybridizing bands were observed after high stringency washes.[2] These additional bands may represent genomic DNA that is sequence related to the factor V gene.

The structure of the human factor V gene is shown in Figure 1. The human factor V gene spans >80 kb of DNA. The

[1] Abbreviations: SSC, 0.01 M sodium citrate buffer, pH 7.4, 0.15 M sodium chloride; PCR, polymerase chain reaction.

[2] L. D. Cripe and W. H. Kane, unpublished observations.

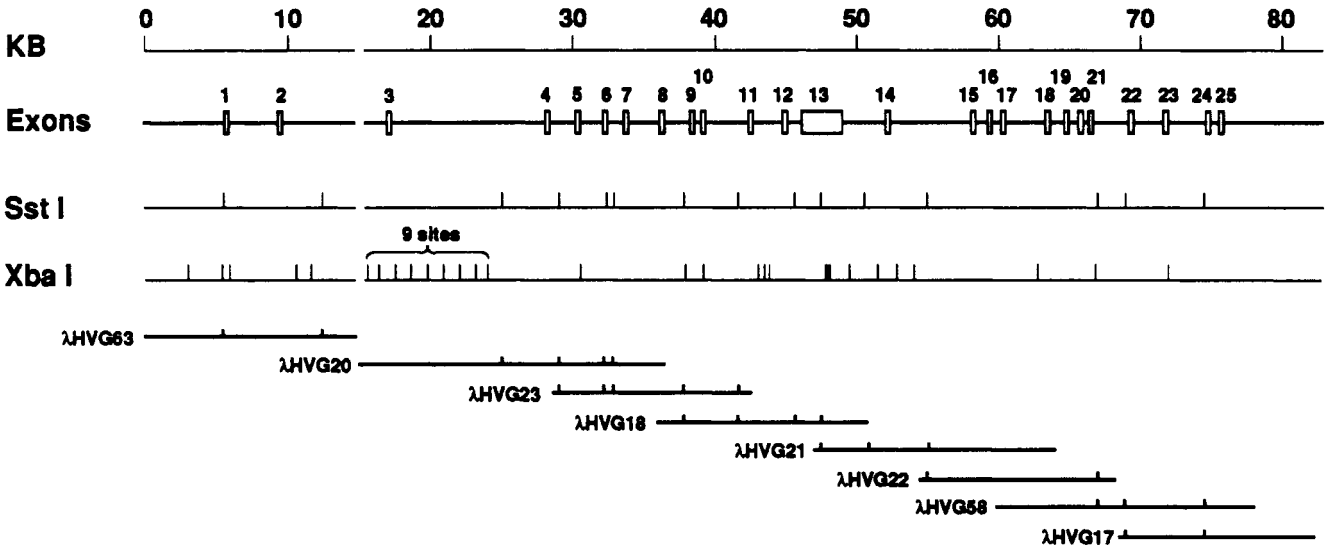

FIGURE 1: Organization of the human coagulation factor V gene. The size scale in kilobases (kb) is shown at the top of the figure. The second line is schematic representation of the human factor V gene drawn to scale with exons indicated by the rectangles connected by lines representing the introns. The number for each exon is indicated. Genomic restriction sites for *Sst*I and *Xba*I are depicted below. The order of the eight *Xba*I fragments that are bracketed was not determined. The location for eight of the unique overlapping human genomic clones is indicated by the solid lines at the bottom of the figure. The identity of each clone is shown at the right of the bar, and the vertical lines indicate the location of genomic *Sst*I sites contained in each clone. Clone λHVG63 was obtained from the λEMBL 3 library; the remainder of the clones were isolated from the λFIX library. There is no overlap between clones λHVG63 and λHVG20. This gap occurs in intron 2 of the factor V gene and it is indicated in the map by the interrupted lines.

nucleotide sequences of exons and exon–intron boundaries were sequenced by the dideoxy chain termination method. The gene consists of 25 exons which encode the factor V mRNA as determined by comparison with the cloned cDNA (Kane & Davie, 1986; Kane et al., 1987; Jenny et al., 1987). The exons range in size from 72 to 2820 bp (Table I). The genomic clones include ~5.4 kb of DNA 5′ to exon 1 and ~7 kb 3′ to exon 25. The introns in the factor V gene range in size from 0.4 to ≥11 kb (Table II). There is no cross-hybridization between clones λHVG63 and λHVG20 indicating that these clones are missing a portion of intron 2. Splice donor and acceptor sequences agree with the GT-AG rule and conform to the consensus sequence of Mount (1982) except for intron 6. Intron 6 agrees with the consensus sequence except that the donor–acceptor sequence is GC-AT. This variant donor splice site sequence accounts for the majority of nonconforming donor splice sites in nonimmunoglobulin genes (Shapiro & Senepathy, 1987).

Comparison of Factor V Genomic and cDNA Sequences. The sequence of the factor V genomic clones is in excellent agreement with the two previously published cDNAs (Kane & Davie, 1986; Kane et al., 1987; Jenny et al., 1987). There are seven positions with differences between the genomic sequence and the two published cDNAs (Table III). In two cases the differences occur in domain A1 at nucleotides 481 and 628. These differences are silent in that they do not affect the predicted amino acid sequence; however, our previously reported cDNA (Kane et al., 1987) contains a *Pst*I site at position 477 which is not present in the genomic clones or the cDNA sequence reported by Jenny (Jenny et al., 1987). The five remaining differences occur within the B domain of factor V. Four of these differences result in amino acid substitutions.

The human factor V cDNAs isolated by Kane (Kane et al., 1987) and Jenny (Jenny et al., 1987) include 5′ untranslated regions of 76 and 90 nucleotides, respectively. The first 20 nucleotides of these sequences are not identical. Comparison of the cDNA sequences to the genomic sequence helps to explain these differences. The sequence reported by Jenny (Jenny et al., 1987) includes the sequence GAATTCCG which is not present in the gene (Figure 2). This sequence corresponds to the *Eco*RI linker used to construct the cDNA library. The sequence of the Jenny cDNA following the linker agrees completely with the genomic sequence. The first 11 nucleotides of the cDNA sequence that we reported previously do not agree with the genomic sequence reported here. This discrepancy appears to be an artifact that occurred near adjacent *Alu*I sites during the construction of the cDNA library. Thus, nucleotides 1–16 of our cDNA correspond to the comple-

Table I: Exon Organization of the Human Factor V Gene[a]

exon	cDNA	length	amino acids	domain
1	14–234	221	53	5′ untranslated sequence, signal sequence, A1
2	235–326	90	30	A1
3	327–449	124	41	A1
4	450–662	212	71	A1
5	663–806	144	48	A1
6	807–1028	221	74	A1
7	1029–1194	166	56	A1-A2
8	1195–1372	178	59	A2
9	1373–1472	99	33	A2
10	1473–1687	215	72	A2
11	1688–1838	151	50	A2
12	1839–2051	213	71	A2
13	2052–4872	2820	940	A2-A3, connecting region or B domain, 18 and 9 amino acid tandem repeats, thrombin cleavage sites
14	4873–5047	175	59	A3
15	5048–5284	237	79	A3
16	5285–5495	211	70	A3
17	5496–5675	180	60	A3
18	5676–5792	117	39	A3
19	5793–5864	72	24	C1
20	5865–5968	104	35	C1
21	5969–6124	156	52	C1
22	6125–6269	145	48	C1
23	6270–6421	152	51	C2
24	6422–6604	182	61	C2
25	6605–6890	286	48	C2, 3′ untranslated sequence

[a] Numbering of nucleotide residues and protein domains corresponds to the previously published sequences (Kane & Davie, 1988). Amino acids with interrupted codons were assigned to the exon containing two of the three codon nucleotides.

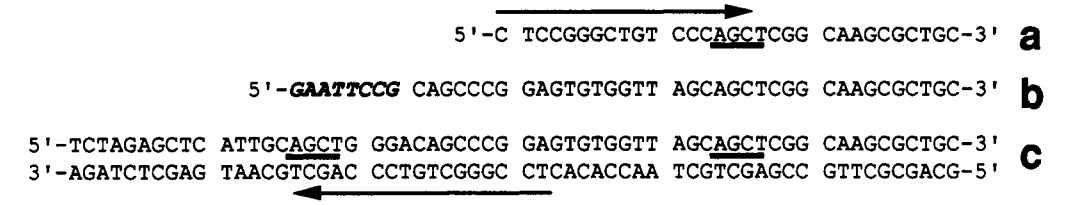

FIGURE 2: Comparison of the factor V genomic sequence with 5′ untranslated regions of published cDNA sequences. The sequences corresponding to the cDNA reported by (a) Kane (Kane et al., 1987), (b) Jenny (Jenny et al., 1987), and (c) the genomic sequence reported here are listed. Sequences corresponding to the *Eco*RI linker are shown in bold print. *Alu*I restriction sites are underlined. The first 18 nucleotides of the cDNA reported by Kane (Kane et al., 1987) differ from the sequence reported by Jenny (Jenny et al., 1987) but are identical to complementary strand genomic sequence (arrows).

Table II: Intron–Exon Splice Junctions of the Human Factor V Gene[a]

Intron	Exon	Intron	Exon	Intron size (kb)	Intron Type
1	TCAAG	**GT**AACTCA.....TTATTTCC**AG**	TTTGAATC	3.8	II
2	TTCAG	**GT**AAGAAC.....GTTTTCAC**AG**	GACTTCTT	>10	I
3	AGAAG	**GT**AAGATA.....ACTCCTCC**AG**	GTGCTTCT	11	I
4	AAAAG	**GT**AAGAAC.....CCTTTCAT**AG**	GGACCCTA	2.2	I
5	GCCAG	**GT**AACACG.....CTTTTCTC**AG**	ATATAACA	1.9	I
6	GCAAG	**G**CAAGAAA.....CCTAACTC**AG**	CTGGGATG	1.4	I
7	GACAA	**GT**GAGTTG.....TCACTTTA**AG**	AAAATACA	2.4	II
8	TCAAA	**GT**AGTAAC.....ATTACTTG**AG**	ATCGTGTT	1.9	0
9	CTCAG	**GT**TTGAAT.....TTTCTTTC**AG**	GCAGGAAC	0.6	I
10	TACAG	**GT**ACTTTT.....TGTCCTCC**AG**	AGGGCAGC	3.1	0
11	GAGCA	**GT**AAGTCA.....TCACTTTC**AG**	CTATCAAT	2.2	I
12	TGTTG	**GT**GAGTAA.....TCTTCTGT**AG**	GAACTTGG	1.1	I
13	CAAAG	**GT**TTGGCC.....ATTATTTC**AG**	GGAAACAG	2.9	II
14	TCCAA	**GT**AAGTCT.....CTTTTCCT**AG**	GTTCGTTT	5.8	0
15	ACCCA	**GT**ACTTCT.....TGTTTTCC**AG**	GAAAAAGA	1.2	0
16	TCACG	**GT**ATTTTT.....TTGTCTTC**AG**	CCATTAAT	1.0	I
17	GCCTG	**GT**AAAGAT.....TCTATTTT**AG**	GTTCATTT	3.0	I
18	CAGAG	**GT**ATCACA.....TTTTTCAT**AG**	ACTGTAGG	1.3	I
19	TCTGG	**GT**AAGTTG.....CTTGAGTT**AG**	GTTACTGG	1.1	I
20	TCCAG	**GT**TTGTCT.....ATATTGGC**AG**	GTGGACAT	0.4	0
21	TGATG	**GT**TTGTGT.....TTTCTTCT**AG**	TATTTTAA	2.7	0
22	AAATG	**GT**AAGGTA.....CTTTTGAT**AG**	GATGTTCC	2.3	I
23	CCAAG	**GT**CAAGTA.....ATTTTCAT**AG**	GCAAACAA	2.8	0
24	ACAAG	**GT**AGAGTG.....TTTTGGTC**AG**	ATTTTTGA	1.0	0
Consensus					
	C	A	TTTTTT T		
	AG	GT AGT.....	N AG		
	A	G	CCCCCC C		

[a] Intron numbers are indicated in the first column. The nucleotide sequences of the exon–intron boundaries are shown for each intron. The splice donor and acceptor dinucleotides are in boldface type and match the consensus GT...AG except for intron 6. The sizes of introns in the human factor V gene were determined by restriction enzyme digestion or PCR. Type 0 indicates a splice between codons, type I and type II indicate a splice after the first or second nucleotide of a codon, respectively. The consensus sequence proposed by Mount (1982) is shown at the bottom of the table.

Table III: Differences in Nucleotide and Predicted Amino Acid Sequences among Factor V cDNA and Genomic Clones[a]

nucleotide no.	genomic codon	Jenny et al. cDNA codon	Kane et al. cDNA codon	amino acid
481	CGG	GCG	GCA	Ala
628	TCG	TCG	TCT	Ser
2377	TCA	TCG	TCG	Ser
2649	AAA	AGA	AGA	Lys to Arg
2670	CAT	CGT	CGT	His to Arg
2849	AAG	GAG	GAG	Lys to Glu
3930	CTT	CTT	ATT	Leu to Ile

[a] The first column indicates the nucleotide number corresponding to the cDNA sequences reported by Kane (Kane & Davie, 1986; Kane et al., 1987). The second column indicates the sequence of the codon determined from the factor V genomic clones in the present work. The nucleotide that differs from previously reported nucleotides is shown in bold print. The second and third columns show the codons reported by Jenny (Jenny et al., 1987) and Kane (Kane & Davie, 1986; Kane et al., 1987). The fifth column indicates the amino acid encoded by the genomic sequence followed by the amino acid encoded by a cDNA sequence if different.

mentary strand of the genomic sequence terminating in an upstream *Alu*I site (Figure 2). RNAse protection assays with a probe that extends 147 bases upstream from the ATG initiation codon demonstrate two protected fragments of 97 and 103 nucleotides.[3] Primer extension assays will be used to verify these as authentic transcription start sites. Thus, the first exon of the factor V gene encodes a 5′ untranslated region of 97 or 103 nucleotides, a 28 amino acid leader peptide, and the first 24 amino acids of the mature protein.

Exon–Intron Organization of the Human Factor V Gene. Comparison with the Factor VIII Gene and Correlation with Protein Domains. The exon–intron organization of the gene for human factor V is almost identical to the organization of the gene for human factor VIII (Gitschier et al., 1984). The genes for factors V and VIII are composed of 25 and 26 exons, respectively. The larger mRNA of factor VIII relative to factor V is primarily due to the 1.8 kb of 3′ untranslated

[3] L. D. Cripe and W. H. Kane, unpublished observations.

sequence in factor VIII. The gene for factor V spans ≥80 kb whereas the gene for factor VIII is considerably larger at ~180 kb. This difference is largely due to the fact that six of the introns in the factor V gene are much smaller than the corresponding introns in the factor VIII gene. We have previously used the computer programs of Dayhoff to align the amino acid sequences for factor V and factor VIII (Kane & Davie, 1988). Figure 3 shows the aligned amino acid sequences near the exon–intron boundaries. Inspection of the corresponding genomic DNA sequences for factor V and factor VIII reveals that in 21 of 24 cases the exon–intron boundaries occur at precisely the same location in the aligned sequence. The 3′ boundary of exon 5 of the factor V gene corresponds to the 3′ boundary of exon 6 in the gene for factor VIII. This can be explained either by the gain or the loss of an intron subsequent to the divergence of these two genes. The 3′ exon–intron boundaries for exon 9 and 13 and the 5′ exon–intron boundary for exon 14 are not precisely conserved as judged by the protein alignment. These discrepancies may either be due to inaccuracies in the protein alignment or alternatively to the phenomenon of intron drift (Craik et al., 1983).

Although the complete structure of the gene for ceruloplasm has not been reported, a pseudogene for ceruloplasmin on chromosome 8 has been characterized as well as several clones for the functional ceruloplasmin gene on chromosome 3 (Koschinsky et al., 1987; Royle et al., 1987). Two exon–intron boundaries for the ceruloplasmin gene were reported. One boundary, corresponding to the 3′ region of exon 9 in the factor V gene, does not occur at the same position as the boundary in either the factor V or factor VIII gene on the basis of the protein alignments. The second boundary corresponds to the 5′ end of exon 13 in the factor V gene and exon 14 in the factor VIII gene. This boundary occurs at precisely the same location in all three genes on the basis of the protein alignments.

Figure 4 shows the location of introns and protein domains for factor V and factor VIII. Within the individual A domains of both proteins, the locations of some introns are conserved suggesting that they existed in the primordial gene for this domain whereas the others were either gained or lost subsequent to the duplication events. The sequences encoding the A1–A2 and A2–A3 domain boundaries are each contained within a single exon in both genes. There is amino acid homology between factor V, factor VIII, and ceruloplasmin in the sequences encoded by the 5′ and 3′ ends of these exons; however, there are major differences in the amino acid sequences encoded by the central portion of each exon (Ortel et al., 1984; Jenny et al., 1987; Kane et al., 1987). Thus, exon 8 of the factor VIII gene contains a 33 amino acid insert that includes 15 acidic residues which is not present in exon 7 of the factor V gene. Futhermore, the exons encoding the domain A2–A3 boundaries in the genes for factor V and factor VIII are both extremely large and encode the entire B domain for the protein. Exon 13 in the factor V gene spans 2820 bp whereas the exon 14 in the factor VIII gene is 3106 bp in length. The exons encoding the B domain in factor V and factor VIII are much larger in size than the reported average size for exons in vertebrate genes of 133 bp (Smith, 1988). Recently, however, large exons ranging from 1.5 to 8.7 kb have been described in a number of genes or gene families including apolipoprotein B (Ludwig et al., 1987), the large subunit of RNA polymerase II (Ahearn et al., 1987), nuclear hormone receptor genes (Faber et al., 1989), collagen genes (Muragaki et al., 1991), and mucin genes (Ligtenberg et al., 1990; Toribara et al., 1991).

The B domain of factor V is not homologous to any known protein; however, it shares several features with the family of mucin glycoproteins (Ligtenberg et al., 1990; Toribara et al., 1991; Gum et al., 1989; Porchet et al., 1991). These similarities may be an example of convergent evolution. Thus, the genes for mucins and factor V contain large exons which encode amino acid sequences that are rich in hydroxyl amino acids and proline. These proteins contain tandem repeats of short amino acid sequences and appear to be heavily glycosylated with both O- and N-linked oligosaccharide chains. In each of these proteins the amino acid sequences and tandem repeat structures appear to be unique. The genes for mucins have been shown to be polymorphic, containing a variable number of tandem repeats (Ligtenberg et al., 1990; Toribara et al., 1991; Gum et al., 1989; Porchet et al., 1991). Thus far, this kind of polymorphism has not been observed in the gene for human factor V.[4] Factor V contains 31 repeats of a nine amino acid sequence. One of the cDNA clones that was isolated from a HepG2 cDNA library contained an in-frame deletion of precisely 11 of the tandem nine amino acid repeats (Kane & Davie, 1986; Kane et al., 1987). It will be of interest to determine whether this was an artifact of the cDNA library construction or a polymorphism in the human factor V gene. On the other hand, the predicted amino acid sequence for bovine factor V contains eight fewer nine amino acid tandem repeats and only a single copy of the 17 amino acid sequence that is duplicated in the human protein (Guinto et al., 1989). The B domain of factor VIII is also rich in carbohydrate and hydroxyl amino acids; however, it does not contain any tandem repeat structures (Toole et al., 1984; Vehar et al., 1984).

In contrast to the sequences encoding the A1–A2 and A2–A3 domain boundaries, the sequences encoding the A3–C1 and C1–C2 domain boundaries are not each contained within a single exon. In these cases the domain boundaries in the protein correlate precisely with type I exon–intron boundaries in the gene. Recently, Stubbs et al. (1990) cloned the cDNA for a murine breast epithelial cell protein of unknown function that consists of a single epidermal growth factor like domain followed by the two C-type domains. The C domains in the epithelial cell protein are approximately 40% identical to the C domains in factor V and factor VIII. The structure of the gene encoding this protein is not known. On the basis of the amino acid sequence homology, we predict that the exon–intron organization of this gene will be very similar to the genes for factor V and factor VIII.

Evolutionary and Functional Considerations. The function of the large connecting region or B domains in factor V and factor VIII remains poorly understood. Both proteins circulate in plasma as relatively inactive cofactors. Activation of both molecules by thrombin yields an active species comprising a calcium-dependent heterodimer composed of heavy and light chains. The heavy chains contain domains A1–A2, whereas the light chains contain domains A3–C1–C2. The B domains contain all three of the thrombin cleavage sites in factor V and two of the three thrombin cleavage sites in factor VIII. Following thrombin activation, the B domains of both proteins are released as very large activation peptides comprising >40% of the mass of the original molecule. Recent ultrastructural data obtained using electron microscopy suggests that the B domains of factor V and factor VIII form a rodlike tail extending from a globular domain containing the heavy chain and light chain (Mosesson et al., 1990a,b). The B domains of factor V and factor VIII are not required for generation

[4] L. D. Cripe and W. H. Kane, unpublished observations.

```
                                        .....EXON 1.....GAG CCC ACA AAC TCA AG GTAACTCA    V
                                                        Glu Pro Thr Asn Ser Se
                                                        Leu Pro Val Asp Ala Ar
                                        .....EXON 1.....CTG CCT GTG GAC GCA AG GTAAAGGC    VIII

          TTATTTCCAG  T TTG AAT CTT TCT GTA.....EXON 2.....CAA TCT ACC ATT TCA G  GTAAGAAC    V
                      r Leu Asn Leu Ser Val               Gln Ser Thr Ile Ser G
                      g Phe Pro Pro Arg Val               Arg Pro Pro Trp Met G
          CTTCACGCAG  A TTT CCT CCT AGA GTG.....EXON 2.....AGG CCA CCC TGG ATG G  GTAATGAA    VIII

          GTTTTCACAG GA CTT CTT GGG CCT ACT.....EXON 3.....AGT AAA TTA TCA GAA G  GTAAGATA    V
                     ly Leu Leu Gly Pro Thr               Ser Lys Leu Ser Glu G
                     ly Leu Leu Gly Pro Thr               Trp Lys Ala Ser Glu G
          ATTCTTACAG GT CTG CTA GGT CCT ACC.....EXON 3.....TGG AAA GCT TCT GAG G  GTGAGTAA    VIII

          ACTCCTCCAG GT GCT TCT TAC CTT GAC.....EXON 4.....CTT ATC TGT AAA AAA G  GTAAGAAC    V
                     ly Ala Ser Tyr Leu Asp               Leu Ile Cys Lys Lys G
                     ly Ala Glu Tyr Asp Asp               Leu Val Cys Arg Glu G
          CCTGCTATAG GA GCT GAA TAT GAT GAT.....EXON 4.....CTA GTA TGT AGA GAA G  GTAAGTGT    VIII

          CCTTTCATAG GG ACC CTA ACT GAG GGT.....EXON 5.....AAT GGG ACA ATG CCA G  GTAACACG    V
                     ly Thr Leu Thr Glu Gly               Asn Gly Thr Met Pro A
                     ly Ser Leu Ala Lys Glu               Asn Arg Ser Leu Pro G
          TCTTTTTCAG GG AGT CTG GCC AAG GAA....EXON 5-6....AAC AGG TCT CTG CCA G  GTATGTAC    VIII

          CTTTTCTCAG AT ATA ACA GTT TGT GCC.....EXON 6.....CCA AAA CAT TTG CAA G  GCAAGAAA    V
                     sp Ile Thr Val Cys Ala               Pro Lys His Leu Gln A
                     ly Leu Ile Gly Cys His               Ser Ser His Gln His A
          TACTTTACAG GT CTG ATT GGA TGC CAC.....EXON 7.....TCT TCC CAC CAA CAT G  GTAATATC    VIII

          CCTAACTCAG CT GGG ATG CAG GCT TAC.....EXON 7.....CCA GCG AAT ATG GAC AA GTGAGTTG    V
                     la Gly Met Gln Ala Tyr               Pro Ala Asn Met Asp Ly
                     sp Gly Met Glu Ala Tyr               Leu Ala Pro Asp Asp Ar
          CTGACTCCAG AT GGC ATG GAA GCT TAT.....EXON 8.....CTC GCC CCC GAT GAC AG GTAAGCAC    VIII

          TCACTTTAAG  A AAA TAC AGG TCT CAG.....EXON 8.....AGA GAC ACA CTC AAA    GTAGTAAC    V
                      s Lys Tyr Arg Ser Gln               Arg Asp Thr Leu Lys
                      g Ser Tyr Lys Ser Gln               Gly Asp Thr Leu Leu
          TCTTATACAG  A AGT TAT AAA AGT CAA.....EXON 9.....GGA GAC ACA CTG TTG    GTAAGTTG    VIII

          ATTACTTGAG    ATC GTG TTC AAA AAT.....EXON 9.....TCT TCT TTC ACC TCA G  GTTTGAAT    V
                        Ile Val Phe Lys Asn               Ser Ser Phe Thr Ser G
                        Ile Ile Phe Lys Asn               Leu Pro Lys G
          ATACTTTCAG    ATT ATA TTT AAG AAT.....EXON 10....TTA CCA AAA G          GTAAATAT    VIII

          TTTCTTTCAG GC AGG AAC AAC ACC ATG.....EXON 10....AGG CGA GGA ATA CAG    GTACTTTT    V
                     ly Arg Asn Asn Thr Met               Arg Arg Gly Ile Gln
                     ly Val Lys His Leu Lys               Gln Arg Gly Asn Gln
          TTGTGGGTAG GT GTA AAA CAT TTG AAG.....EXON 11....CAA AGA GGA AAC CAG    GTGAGTTC    VIII

          TGTCCTCCAG    AGG GCA GCA GAC ATC.....EXON 11....TCA AAC ATC ATG AGC A  GTAAGTCA    V
                        Arg Ala Ala Asp Ile               Ser Asn Ile Met Ser T
                        Ile Met Ser Asp Lys               Ser Asn Ile Met His S
          TAATTAACAG    ATA ATG TCA GAC AAG.....EXON 12....TCC AAC ATC ATG CAC A  GTGAGTAA    VIII

          TCACTTTCAG CT ATC AAT GGC TAT GTG.....EXON 12....ACA ATG GAT AAT GTT G  GTGAGTAA    V
                     hr Ile Asn Gly Tyr Val               Thr Met Asp Asn Val G
                     er Ile Asn Gly Tyr Val               Ser Met Glu Asn Pro G
          GTTTTTGCAG GC ATC AAT GGC TAT GTT.....EXON 13....TCG ATG GAA AAC CCA G  GTTAGTTA    VIII

          ATTATTTCAG GA ACT TGG ATG TTA ACT.....EXON 13....CAA AG                 GTTTGGCC    V
                     ly Thr Trp Met Leu Thr               Gln Ar
                     ly Leu Trp Ile Leu Gly               Ser Ser Pro His Val Leu Arg Asn Ar
          TCATCTCCAG GT CTA TGG ATT CTG GGG.....EXON 14....AGC TCC CCA CAT GTT CTA AGA AAC AG GTATGAAT   VIII

TTAATTTCAG  G GAA ACA GAT ATT GAA GAC TCT GAT.....EXON 14....GAT GAT GTT ATC CAA    GTAAGTCT    V
            g Glu Thr Asp Ile Glu Asp Ser Asp               Asp Asp Val Ile Gln
                                     g Ala                  Glu Asp Asn Ile Met
            TTTTCTCCAG                 G GCT.....EXON 15....GAA GAT AAT ATC ATG    GTGAGTTA    VIII


          CTTTTCCTAG    GTT CGT TTT AAA AAT.....EXON 15....TCA GCT GTG AAC CCA    GTACTTCT    V
                        Val Arg Phe Lys Asn               Ser Ala Val Asn Pro
                        Val Thr Phe Arg Asn               Ser Asp Val Asp Leu
          TGTTCTACAG    GTA ACT TTC AGA AAT.....EXON 16....TCT GAT GTT GAC CTG    GTGAGTTA    VIII

          TGTTTTCCAG    GAA AAA GAT ATT CAC.....EXON 16....TCC CAT GAG TTT CAC G  GTATTTTT    V
                        Glu Lys Asp Ile His               Ser His Glu Phe His A
                        Glu Lys Asp Val His               Asn Tyr Arg Phe His A
          CCCTCCCTAG    GAA AAA GAT GTG CAC.....EXON 17....AAT TAT CGC TTC CAT G  GTAATATA    VIII

          TTGTCTTCAG CC ATT AAT GGG ATG ATC.....EXON 17....TGG CCC CTT CTG CCT G  GTAAAGAT    V
                     la Ile Asn Gly Met Ile               Trp Pro Leu Leu Pro G
                     la Ile Asn Gly Tyr Ile               Tyr Asn Leu Tyr Pro G
          CCTTCTCCAG CA ATC AAT GGC TAC ATA.....EXON 18....TAC AAT CTC TAT CCA G  GTATGAGC    VIII
```

```
TCTATTTTAG GT TCA TTT AAA ACT CTT.....EXON 18....CTT ATC ATG GAC AGA G  GTATCACA    V
           ly Ser Phe Lys Thr Leu                Leu Ile Met Asp Arg A
           ly Val Phe Glu Thr Val                Leu Val Tyr Ser Asn L
TTTTTATAAG GT GTT TTT GAG ACA GTG.....EXON 19....CTG GTG TAC AGC AAT A  GTGAGTAG    VIII

TTTTTCATAG AC TGT AGG ATG CCA ATG.....EXON 19....GCT TCA GAG TTT CTG G  GTAAGTTG    V
           sp Cys Arg Met Pro Met                Ala Ser Glu Phe Leu G
           ys Cys Gln Thr Pro Leu                Ala Ser Gly Gln Tyr G
TTCATTTCAG AG TGT CAG ACT CCC CTG.....EXON 20....GCT TCA GGA CAA TAT G  GTAAATAC    VIII

CTTGAGTTAG GT TAC TGG GAG CCC AGA.....EXON 20....AAA CCT TGG ATC CAG    GTTTGTCT    V
           ly Tyr Trp Glu Pro Arg                Lys Pro Trp Ile Gln
           ly Gln Trp Ala Pro Lys                Phe Ser Trp Ile Lys
TTGGGCAAAG GA CAG TGG GCC CCA AAG.....EXON 21....TTT TCT TGG ATC AAG    GTTAGAAA    VIII

ATATTGGCAG    GTG GAC ATG CAA AAG.....EXON 21....ACA AGG AAT GTG ATG    GTTTGTGT    V
              Val Asp Met Gln Lys                Thr Arg Asn Val Met
              Val Asp Leu Leu Ala                Thr Gly Thr Leu Met
TAATTGGTAG    GTG GAT CTG TTG GCA.....EXON 22....ACT GGA ACC TTA ATG    GTATGTAA    VIII

TTTCTTCTAG    TAT TTT AAT GGC AAT.....EXON 22....GGT TGT GAG GTA AAT G  GTAAGGTA    V
              Tyr Phe Asn Gly Asn                Gly Cys Glu Val Asn G
              Val Phe Phe Gly Asn                Gly Cys Asp Leu Asn S
CTCCATACAG    GTC TTC TTT GGC AAT.....EXON 23....GGC TGT GAT TTA AAT A  GTAAGTGC    VIII

CTTTTGATAG GA TGT TCC ACA CCC CTG.....EXON 23....GCC TGG CAA GCC AAG    GTCAAGTA    V
           ly Cys Ser Thr Pro Leu                Ala Trp Gln Ala Lys
           er Cys Ser Met Pro Leu                Ala Trp Arg Pro Gln
TTTCTTTGAG GT TGC AGC ATG CCA TTG.....EXON 24....GCC TGG AGA CCT CAG    GTAAGAGG    VIII

ATTTTCATAG    GCA AAC AAC AAT AAG.....EXON 24....TCC ATG GTG GAC AAG    GTAGAGTG    V
              Ala Asn Asn Asn Lys                Ser Met Val Asp Lys
              Val Asn Asn Pro Lys                Asn Gly Lys Val Lys
TTGCCCTCAG    GTG AAT AAT CCA AAA.....EXON 25....AAT GGC AAA GTA AAG    GTAAGCTG    VIII

TTTTGGTCAG    ATT TTT GAA GGA AAT.....EXON 25....
              Ile Phe Glu Gly Asn
              Val Phe Gln Gly Asn
CCTCTTTCAG    GTT TTT CAG GGA AAT.....EXON 26....
```

FIGURE 3: Comparison of exon–intron junction sequences of the genes for factors V and VIII. The amino acid sequences of human factor V and factor VIII were aligned by the method of Dayhoff as previously reported (Kane & Davie, 1988). The first line of each comparison is the genomic sequence for factor V determined in the present work. The second line is the corresponding translated amino acid sequence for factor V. The third line is the translated amino acid sequence for factor VIII which has been aligned to the factor V amino acid sequence. The fourth line in each series is the corresponding sequence for the factor VIII gene determined by Gitschier (Gitschier et al., 1984). Genomic exon sequences are shown as triplets, and the translated amino acid sequence for split codons is indicated by the split in the three-letter amino acid code. The number for each exon in the factor V and the factor VIII genes is indicated. Exon 5 in the factor V gene contains the sequences corresponding to exons 5 and 6 in the factor VIII gene.

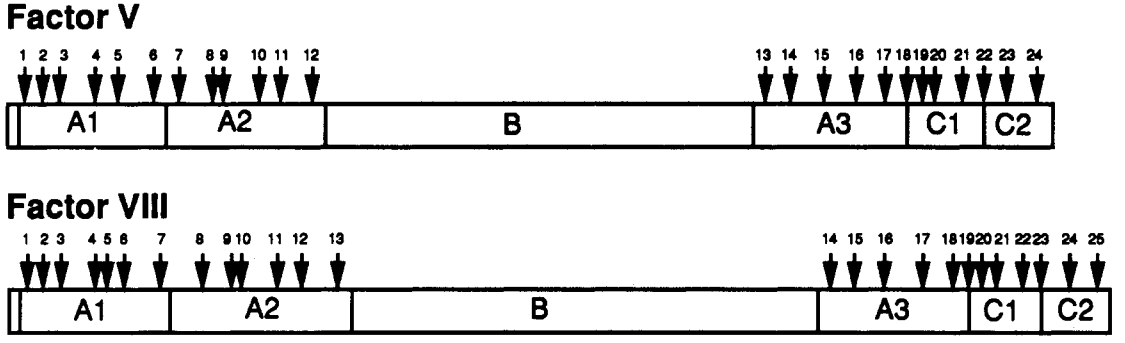

FIGURE 4: Comparison of the domain structures and exon–intron structures for factor V and factor VIII. The boxes represent the domain structures for factor V and factor VIII and are drawn to scale. The identities of the domains are indicated by the letters and correspond to the amino acid sequences described previously (Kane & Davie, 1988). The arrows indicate the location of introns which are removed from the mature mRNA. The number of each intron in the gene for factor V or factor VIII is indicated.

of procoagulant activity since deletion of a large portion of these regions in recombinant proteins does not reduce the procoagulant activity of either molecule (Toole et al., 1986; Kane et al., 1990). However, these observations do not exclude other functions for the B domains. Thus, the factor V deletion mutant appeared to have increased constitutive procoagulant activity suggesting that the entire B domain is necessary to inhibit cofactor activity (Kane et al., 1990). Complete deletion of the B domain results in a molecule with markedly diminished activity secondary to reduced rates of proteolytic activation by thrombin (Pittman et al., 1990). Furthermore, the B domain of factor VIII serves a unique function in that it contains the binding site for von Willebrand factor (Pittman & Kaufman, 1989). The von Willebrand factor–factor VIII interaction is required for the stabilization of factor VIII in vitro as well as in vivo (Kaufman et al., 1988).

Because of the large size of the exons encoding the B domains in factor V and factor VIII, we speculate that these exons arose through the insertion of a processed gene derived from reverse transcription of an mRNA. It is possible that the exons encoding the B domains of both factor V and factor VIII are derived from a single primordial cofactor gene. In this case, the lack of amino acid homology between the two cofactors would be due to divergent evolution. Analysis of the amino acid sequences for factor V and factor VIII from different species suggests that there has been relatively little

selective pressure to conserve amino acid sequences in the B domains. Thus, the predicted amino acid sequence for the B domain of bovine factor V is only 59% identical to the corresponding sequence of the human protein whereas there is 88% amino acid sequence identity in the heavy chain and light chain regions (Guinto et al., 1989). Similar divergence has been observed between human and porcine factor VIII sequences. Toole et al. (1986) reported ~50% amino acid identity in the B domain region whereas there was 80–85% amino acid identity in the heavy-chain and light-chain regions. Divergence within the B domain of human factor V is further suggested by the fact that the five discrepancies between reported genomic and cDNA sequences that result in missense mutations all occur within the B domain. While the available data are consistent with a single insertion event in the primordial cofactor gene followed by divergent evolution, it is also possible that two separate insertion events occurred subsequent to the divergence of the genes for factor V and factor VIII. An insertion event also appears to have occurred in the A1–A2 exon in the factor VIII gene which codes for an additional 33 amino acids that are not present in the corresponding region of factor V or ceruloplasmin. Characterization of related genes in primitive species may further clarify the evolution of this family of related genes.

Characterization of the gene for human factor V provides the groundwork for examining natural mutations that affect factor V function and result in the rare bleeding disorder congenital factor V deficiency. Functional characterization of these defects should provide insight into the regulation of the prothrombinase complex. The major sites of synthesis of factor V appear to be the hepatocyte and megakaryocyte (Wilson et al., 1984; Gewirtz et al., 1986). The acquired factor V deficiency associated with hepatic failure and the correlation of clinical severity with platelet factor V levels in parahemophilia indicate the importance of factor V synthesis at these sites (Miletich et al., 1978). Factor V has also been reported to be synthesized in other cell types including monocytes (Altieri & Edgington, 1989), endothelial cells (Shen & Edgington, 1991), T-cells (Shen et al., 1990), and vascular smooth muscle cells (Rodgers, 1988). The importance of cellular surfaces in regulating blood coagulation suggests a possible role for factor V associated with these cells. Characterization of the regulatory elements of the factor V gene may help explain the tissue-specific expression of this protein and the significance of extrahepatic synthesis in physiological and pathophysiological states.

Acknowledgments

We thank Mr. Thad Howard for his assistance with the genomic Southern blot experiments. We also thank the following individuals for helpful discussions and criticisms during the course of this work: Drs. G. R. Vandenbark, Michael F. Seldin, Rebecca Oakey, Thomas L. Ortel, and Charles S. Greenberg.

Registry No. V_a, 65522-14-7.

References

Ahearn, J. M., Barolomei, M. S., West, M. L., Cisek, L. J., & Corden, J. L. (1987) *J. Biol. Chem. 262*, 10695.

Altieri, D. C., & Edgington, T. S. (1989) *J. Biol. Chem. 264*, 2969.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180.

Craik, C. S., Rutter, W. J., & Fletterick, R. (1983) *Science 220*, 1125.

Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkman, A. O., & Trapman, J. (1989) *Mol. Cell. Endocrinol. 61*, 257.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6.

Gewirtz, A. M., Keefer, M., Doshi, K., Annamalai, A. E., Chiu, H. C., & Colman, R. W. (1986) *Blood 67*, 1639.

Gilbert, W. (1985) *Science 226*, 823.

Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Capon, D. J., & Lawn, R. M. (1984) *Nature 312*, 326.

Guinto, E. R., Odegaard, B., Mann, K. G., & MacGillivray, R. T. A. (1989) *Thromb. Haemostasis 62*, 80 (Abstract).

Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., & Kim, Y. S. (1989) *J. Biol. Chem. 264*, 6480.

Jenny, R. J., Pittman, D. D., Toole,J. J., Kriz, R. W., Aldape, R. A., Hewick, R. M., Kaufman, R. J., & Mann, K. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4846.

Kane, W. H., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6800.

Kane, W. H., & Davie, E. W. (1988) *Blood 71*, 539.

Kane, W. H., Ichinose, A., Hagen, F. S., & Davie, E. W. (1987) *Biochemistry 26*, 6508.

Kane, W. H., Devore-Carter, D., & Ortel, T. L. (1990) *Biochemistry 29*, 6762.

Kaufman, R. J., Wasley, L. C., & Dorner, A. J. (1988) *J. Biol. Chem. 263*, 6352.

Koschinsky, M. L., Chow, B. K., Schwartz, J., Hamerton, J. L., & MacGillivray, R. T. (1987) *Biochemistry 26*, 7760.

Ligtenberg, M. J. L., Vos, H. L., Gennissen, A. M. C., & Hilkens, J. (1990) *J. Biol. Chem. 265*, 5573.

Ludwig, E. W., Blackhart, B. D., Pierotti, V. I., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R. W., Levy-Wilson, B., & McCarthy, B. J. (1987) *DNA 6*, 363.

Miletich, J. P., Majerus, D. W., & Majerus, P. W. (1978) *J. Clin. Invest. 62*, 824.

Mosesson, M. W., Church, W. R., DiOrio, J. P., Krishnaswamy, S., Mann, K. G., Hainfeld, J. F., & Wall, J. S. (1990a) *J. Biol. Chem. 265*, 8863.

Mosesson, M. W., Fass, D. N., Lollar, P., DiOrio, J. P., Parker, C. G., Knutson, G. J., Hainfeld, J. F., & Wall, J. S. (1990b) *J. Clin. Invest. 85*, 1983.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459.

Muragaki, Y., Jacenko, O., Apte, S., Mattei, M., Ninomiya, Y., & Olsen, B. R. (1991) *J. Biol. Chem. 266*, 7721.

Ortel, T. L., Takahashi, N., & Putnamm, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4761.

Pittman, D. D., & Kaufman, R. J. (1989) *Thromb. Haemostasis 61*, 161.

Pittman, D. D., Thompkinson, K., Marquette, K., Jenny, R. J., Mann, K. B., & Kaufman, R. J. (1990) *Blood 76* (Suppl.) 433a (Abstract).

Porchet, N. N., Van Cong, J., Dufosse, J. P., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A., & Aubert, J. P. (1991) *Biochem. Biophys. Res. Commun. 175*, 414.

Rodgers, G. M. (1988) *Biochim. Biophys. Acta 968*, 17.

Royle, N. J., Irwin, D. M., Koschinsky, M. L., MacGillivray, R. T., & Hamerton, J. L. (1987) *Somatic Cell Mol. Genet. 13*, 285.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487.

Sambrook, J. E., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
Shapiro, M. B., & Senepathy, P. (1987) *Nucleic Acids Res. 15*, 7155.
Shen, N. L. L., & Edgington, T. S. (1991) *Arterioscler. Thromb. 11*, 1543a (Abstract).
Shen, N. L. L., Altieri, D. C., & Edgington, T. S. (1990) *FASEB J. 4*, A2275 (Abstract).
Smith, M. W. (1988) *J. Mol. Evol. 27*, 45.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503.
Stubbs, J. D., Lekutis, C., Singer, K. L., Bui, A., Yuzuki, D., Srinivasan, U., & Parry, G. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8417.
Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N., & Hewick, R. M. (1984) *Nature 312*, 342.
Toole, J. J., Pittman, D. D., Orr,E. C., Murtha, P., Wasley, L. C., & Kaufman, R. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5939.
Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M., & Kim, Y. S. (1991) *J. Clin. Invest. 88*, 1005.
Tracy, P. B., Eide, L. L., Bowie, E. J., & Mann, K. G. (1982) *Blood 60*, 59.
Vehar, G. A., Keyt, B., Eaton, D. L., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkins, R. N., Tuddenham, E. G. D., Lawn, R. M., & Capon, D. J. (1984) *Nature 312*, 337.
Wang, H., Riddell, D. C., Guinto, E. R., & MacGillivray, R. T. A. (1988) *Genomics 2*, 234.
Watson, M. L., Kingsmore, S. F., Johnston, G. I., Siegelman, M. H., Le Beau, M. M., Lemons, R. S., Bora, N. S., Howard, T. A., Weissman, I. L., McEver, R. P., & Seldin, M. F. (1990) *J. Exp. Med. 172*, 263.
Wilson, D. B., Salem, H. H., Mruk, J. S., Maruyama, I., & Majerus, P. W. (1984) *J. Clin. Invest. 73*, 654.

Identification of Amino Acids Modified by the Bifunctional Affinity Label 5′-(*p*-Fluorosulfonyl)benzoyl)-8-azidoadenosine in the Reduced Coenzyme Regulatory Site of Bovine Liver Glutamate Dehydrogenase†

Kenneth E. Dombrowski,‡ Yu-Chu Huang, and Roberta F. Colman*
Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716
Received December 24, 1991

ABSTRACT: Bovine liver glutamate dehydrogenase reacts with the bifunctional affinity label 5′-(*p*-(fluorosulfonyl)benzoyl)-8-azidoadenosine (5′-FSBAzA) in a two-step process: a dark reaction yielding about 0.5 mol of -SBAzA/mol of subunit by reaction through the fluorosulfonyl moiety, followed by photoactivation of the azido group whereby covalently bound -SBAzA becomes cross-linked to the enzyme [Dombrowski, K. E., & Colman, R. F. (1989) *Arch. Biochem. Biophys. 275*, 302–308]. We now report that the rate constant for the dark reaction is not reduced by ADP or GTP, but it is decreased 7-fold by 2 mM NADH and 40-fold by 2 mM NADH + 0.2 mM GTP, suggesting that 5′-FSBAzA reacts at the GTP-dependent NADH inhibitory site. The amino acid residues modified in each phase of the reaction have been identified. Modified enzyme was isolated after each reaction phase, carboxymethylated, and digested with trypsin, chymotrypsin, or thermolysin. The digests were fractionated by chromatography on a phenylboronate agarose column followed by HPLC. Gas-phase sequencing of the labeled peptides identified Tyr^{190} as the major amino acid which reacts with the fluorosulfonyl group; Lys^{143} was also modified but to a lesser extent. The predominant cross-link formed during photolysis is between modified Tyr^{190} and the peptide Leu^{475}-Asp^{476}-Leu^{477}-Arg^{478}, which is located near the C-terminus of the enzyme. Thus, 5′-FSBAzA is effective in identifying critical residues distant in the linear sequence, but close within the regulatory nucleotide site of glutamate dehydrogenase.

Bovine liver glutamate dehydrogenase [L-glutamate:NAD(P)$^+$ oxidoreductase (deaminating), EC 1.4.1.3] is an allosteric enzyme that is activated by ADP, but inhibited by GTP and high concentrations of NADH. The enzyme in its smallest, active form is a hexamer of six identical subunits (Goldin & Frieden, 1972; Julliard & Smith, 1979). For each subunit there are six binding sites for catalytic and regulatory purine nucleotides: one catalytic site that binds either NAD(H) or NAD(P)H (Goldin & Frieden, 1972), one additional coenzyme site which is regulatory (Krause et al., 1974), two ADP sites (Batra & Colman, 1986a), and two GTP binding sites as measured in the presence of NADH (one of high affinity and one of low affinity) (Pal & Colman, 1979).

Considerable information regarding the amino acid residues within the catalytic and regulatory sites of bovine liver glutatmate dehydrogenase has come from affinity labeling ex-

† This work was supported by USPHS Grant DK 37000 (to R.F.C.), NSF Grant DMB-9105116 (to R.F.C.), and NIH Postdoctoral Fellowship GM12936 (to K.E.D.).
‡ Present address: Department of Internal Medicine, Texas Tech University Health Sciences Center, 1400 Wallace Blvd., Amarillo, TX 79124.

0006-2960/92/0431-3785$03.00/0 © 1992 American Chemical Society