THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1990 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 265, No. 5, Issue of February 15, pp. 2782-2788, 1990
Printed in U.S.A.

Genomic Structure of an Integrin α Subunit, the Leukocyte p150,95 Molecule*

(Received for publication, October 30, 1989)

Angel L. Corbi‡§, Julio Garcia-Aguilar‡§, and Timothy A. Springer¶

From the Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

The genomic structure of integrins is important to our understanding of the evolution of this complex family. The α subunit of the leukocyte integrin p150,95 (CD11c) is a transmembrane polypeptide of 1144 residues whose long extracellular region contains three putative divalent cation binding repeats and a 200-amino acid inserted or "I" domain. The p150,95 α subunit gene extends over 25 kilobases and is comprised of at least 31 exons grouped in five clusters. The I domain, which is only present in some integrins and is homologous to domains in von Willebrand factor, cartilage matrix protein, complement factor B and the α1 and α2 chains of collagen type VI, is distributed in four exons. Each one of the three divalent cation binding repeats is encoded by a separate exon. Surprisingly, a sequence homologous to the first two putative divalent cation binding repeats is present in an inverted orientation in the intron following the last exon of the I domain. Both the signal peptide and the transmembrane domain are split in two exons. Putative proteolytic cleavage sequences in other integrin α subunits align as inserts within the p150,95 α subunit gene falling at exon boundaries. The organization of the p150,95 α subunit gene provides further insights into the structure and evolution of the integrins.

The integrin gene family is composed of a group of structurally and functionally homologous cell surface glycoproteins involved in cell-cell and cell-extracellular matrix interactions (1). Structurally, all the integrins are noncovalently associated α/β heterodimers. There are three integrin subfamilies, each one sharing a unique β subunit (1-4). The β2 subfamily is comprised of three molecules, LFA-1,[1] Mac-1, and p150,95, which are leukocyte-specific and hence are termed the leukocyte integrins. They mediate cell-cell interactions in antigen-specific immune responses as well as antigen-independent inflammatory reactions (3). Leukocyte integrins are required for binding of circulating neutrophils and monocytes to endothelial cells and subsequent transmigration at sites of infection and inflammation, as dramatically illustrated by the lack of pus formation in patients who are genetically deficient in these molecules due to mutations in the common β subunit. Such patients have impaired wound healing, are susceptible to soft tissue infections, and often die in infancy (5). The β1 subfamily contains at least six members which include receptors for some extracellular matrix components such as collagen (6), laminin (7), and fibronectin (4, 8), and receptors which mediate cell-cell adhesion (9, 10). The β3 subfamily is also involved in cell-extracellular matrix interactions and includes the platelet glycoprotein IIb/IIIa and the vitronectin receptor.

p150,95 is primarily expressed on cells of the myeloid lineage, although it is also expressed on the surface of certain activated T and B lymphocytes (11, 12) and is a marker for hairy leukemia cells (13). The involvement of p150,95 in leukocyte adhesive functions has been demonstrated by inhibition with p150,95 α subunit-specific monoclonal antibodies. p150,95 plays a role in monocyte adhesion to endothelial cells (14), neutrophil adhesion to serum-coated surfaces (15), phagocytosis of latex particles, and chemotaxis by peripheral blood monocytes (14, 15). Furthermore, monoclonal antibodies against the p150,95 α subunit can inhibit CTL-target conjugate formation, demonstrating an involvement of p150,95 in CTL-mediated killing (12).

The α subunit of p150,95 is a transmembrane polypeptide which contains an extracellular domain of 1088 residues, a transmembrane domain of 26 amino acids, and a cytoplasmic tail of 30 amino acids (16), shows 63 and 37% identity to the α subunits of Mac-1 (17-19) and LFA-1 (20), respectively, and 20-25% identity to the α subunits of the integrins of the β1 and β3 subfamilies (21-25). The NH_2-terminal half of the extracellular domain includes three EF-hand-like divalent cation binding sequences (26, 27) and the 200-residue "inserted" (I) domain. The I domain is found in all three β2 subfamily α subunits (16-20) but in only one of the three sequenced β1 (21-23) and in none of the β3 subfamily α subunits (24, 25). Integrin I domains have significant homology to the A domains of von Willebrand factor (28), to the repeats of chicken cartilage matrix protein (29), to the repeats of the globular domain of the α1 and α2 polypeptides of collagen type VI (30), and to a region from the C3 binding complement components C2 (31) and factor B (32). Although cDNAs for several integrin α subunits have been cloned, the organization of the genes encoding these proteins is not presently known. Here we report the isolation of several genomic clones for the coding region of the α subunit of p150,95. The gene spans over 25 kb and is comprised of at least 31 exons. The signal peptide, like the transmembrane domain, is split in two exons. The I domain is distributed in four exons. Each of the putative cation binding domains is encoded in a different exon. Comparison of the genomic organization of the p150,95 α subunit with other proteins

* This work was supported by National Institutes of Health Grants CA 31799 and AI 20316, an American Cancer Society faculty award (to T. A. S.), and a Fogarty International fellowship (to A. L. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Servicios de Inmunologia (A. L. C.) and Cirugia (J. G.-A), Hospital de la Princesa, Madrid, Spain.

§ The contribution to this paper by A. L. C. and J.G.-A. is equal, and the order of authorship is arbitrary.

¶ To whom correspondence should be addressed.

[1] The abbreviations used are: LFA-1, lymphocyte function-associated antigen 1; VLA, very late antigen; SDS, sodium dodecyl sulfate; kb, kilobase(s); bp, base pair(s).

provides useful information about the structure and evolution of this important family of proteins.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Genomic Clones—10^6 colonies of a human genomic library constructed in the cosmid vector pCOS2EMBL (obtained from Dr. Hans Lehrach, EMBL, Heidelberg, Germany) (33) were grown on nitrocellulose filters overlaying LB/kanamycin plates. Replica filters were obtained as described (34) and washed in 0.1 × SSC, 0.5% SDS at 65 °C with shaking for 16 h and with frequent changes of the solution. Filters were prehybridized overnight in 6 × SSC (0.6 M sodium chloride, 0.06 M sodium citrate), 1% SDS, 1 × Denhardt's, and 100 μg/ml of herring sperm DNA at 65 °C. Hybridization was carried out in the same solution including 10^6 cpm/ml of the nick-translated 1.2-kb *Eco*RI-*Pst*I fragment from the p150,95 α subunit cDNA clone λX47 as probe (16). The filters were washed twice in 2 × SSC, 0.5% SDS at room temperature for 30 min followed by 0.1 × SSC, 0.5% SDS at 65 °C for 30 min. Nitrocellulose filters were then exposed to preflashed XAR-5 film for 48 h with intensifying screen. Additional genomic clones were isolated by screening 5 × 10^5 plaque-forming units from a human genomic library in λEMBL3 (donated by Dr. S. Orkin, Children's Hospital, Boston, MA) (35) and 5 × 10^5 colony-forming units from a pWE15 cosmid library (a gift from Dr. G. Evans, Salk Institute, San Diego, CA) (36). Nitrocellulose filter plaque lifts or replica filters, respectively, were screened as described above except that a 0.6-kb *Eco*RI fragment and the whole coding region from λX47 were used as probes.

Positive clones were isolated by repeated subcloning and their inserts were analyzed by restriction enzyme mapping using single, double, and partial digestions (37). Further mapping and location of exons was done by probing Southern blots of those digestions with different fragments from the p150,95 α subunit cDNA clone λX47 using standard techniques (37).

Sequencing of the Genomic Clones—Initial sequencing of the clone pCOS2EMBL 8.1 was done by the Sanger dideoxynucleotide method (38) after subcloning different fragments into M13mp18 and -19 using standard techniques (37). Small genomic fragments of pCOS2EMBL 8.1, pWE15 30.1, and λEMBL 22.1.2.1 containing several exons were cloned into pUC13, -18, -19, pGEM3, or pGEM7Z for preparation of nested deletions using the Erase-a-base system (Promega Biotec) (39). The missing areas between deletions and selected exon/intron boundaries were sequenced using specific oligonucleotide primers, based on the genomic or cDNA sequence, on double-stranded plasmid DNA.

RESULTS

Isolation, Mapping, and Sequencing of the p150,95 α Subunit Gene—We have previously described the isolation and sequencing of the cDNA clone λX47, which contains the entire coding information for the α subunit of the leukocyte integrin p150,95 (16). Southern blot analysis of human genomic DNA with different fragments from λX47 as probes showed that the p150,95 α subunit is encoded by a single copy gene (16, 40). A ^{32}P-labeled 1.2-kb *Eco*RI-*Pst*I fragment from λX47 was used to screen 10^6 cosmid clones from a human placenta genomic library in pCOS2EMBL, and a single cosmid clone (pCOS2EMBL 8.1) was obtained. Restriction mapping, Southern hybridization with different fragments of λX47, and DNA sequencing demonstrated that pCOS2EMBL 8.1 contained the 3′-untranslated region and the COOH-terminal 3/4 of the p150,95 α subunit coding region (Fig. 1*A*). To isolate the missing portion of the p150,95 α subunit gene, a λEMBL3 genomic library and a pWE15 cosmid library were screened with the most 5′ fragment of the p150,95 α subunit cDNA (a 0.6-kb *Eco*RI fragment) and with the whole coding region, respectively. A single λEMBL3 clone (λEMBL3 22.1.2.1) was isolated, and sequence analysis demonstrated that it contains the 5′-untranslated region of λX47 and a portion of the gene encoding the first 124 residues of the mature p150,95 α subunit (Fig. 1*A*). Restriction map analysis of three pWE15 cosmid clones indicated that one of them (pWE15 30.1) overlaps with both pCOS2EMBL 8.1 and λEMBL3 22.1.2.1 and contained the entire coding sequence for the p150,95 α

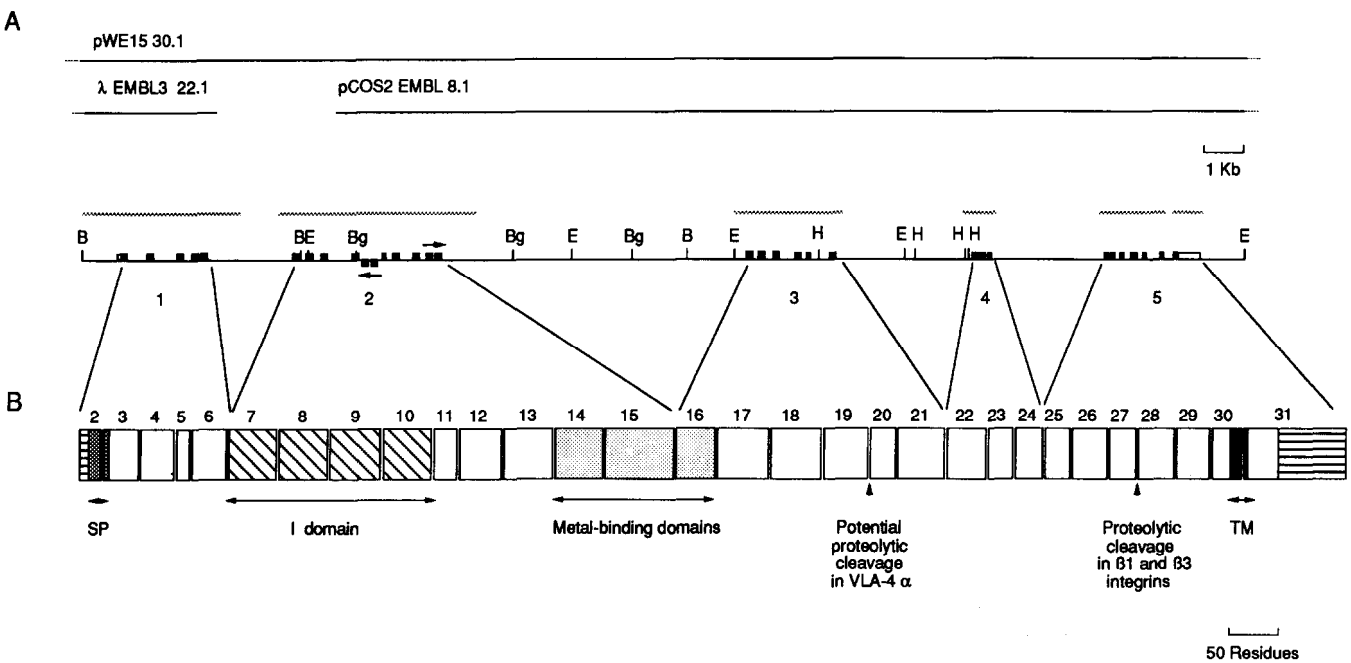

FIG. 1. **Genomic organization of the p150,95 α subunit gene.** *A*, physical map of the p150,95 α subunit gene. The combined restriction map of the three overlapping genomic clones for four restriction endonucleases is depicted with an accuracy of ±250 bp. The restriction enzymes indicated are *Bam*HI (*B*), *Bgl*II (*Bg*), *Eco*RI (*E*), and *Hin*dIII (*H*). *Filled boxes* represent coding exons, and *empty boxes* represent the 5′- and 3′-untranslated region. The *dashed lines* above the physical map indicate the regions that have been sequenced. Exon clusters are numbered. *B*, schematic representation of the exons encoding the p150,95 α subunit gene. Exon length is drawn to scale. The clustering of the exons is indicated. The different domains are based on the cDNA-derived amino acid sequence. *SP*, signal peptide; *TM*, transmembrane domain.

FIG. 2. **Nucleotide sequence of the p150,95 α subunit gene exons and their borders, and the amino acid sequence.** Exonic sequences are in *upper case*. Intron sequences are *lower case*, and only that portion significant for the consensus rules for splicing is shown. The protein sequence is shown in *one-letter code* below the first base of each codon. The *numbers* on the *left side* represent the residue number of the mature polypeptide; every 10th amino acid is marked with a *dot*. Exons are numbered as described in the text.

subunit as well as the 3′-untranslated region and all but the six most 5′ nucleotides of the cDNA 5′-untranslated region (Fig. 1*A*). The restriction maps of λEMBL3 22.1.2.1, pWE15 8.1, and pCOS2λEMBL 8.1 were obtained independently, and overlapping regions were concordant in every case.

Organization of the p150,95 α Subunit Gene—The organization of the p150,95 α subunit gene was deduced from the restriction map, Southern hybridization, and composite nucleotide sequence of the λEMBL3 22.1.2.1, pWE15 30.1, and pCOS2EMBL 8.1 genomic clones. The gene spans over 25 kb and is comprised of at least 31 exons groups in five clusters (Fig. 1, *A* and *B*). The sequences of these clusters, comprising 14 kb, have been deposited in GenBank.[2] The exonic sequences are identical to the cDNA sequence with the exception of four nucleotides (cDNA positions 2329–2332). This results in a leucine to aspartic acid change at amino acid position 737 (Fig. 2). All the intron/exon boundaries are in agreement with the consensus rule for splicing (41). The introns separating the different clusters are between 2.1 and 6.9 kb long, with an average of 3.7 kb (Table I). Inside each cluster the length of the introns varies between 84 and 619 bp, with an average of 275 bp. The exons range in size from 71 to 210 bp, with an average size of 118 bp excluding exon 31 for which the 3′ extent of the 3′-untranslated region has not been determined.

The first identified exon encodes 52 bp of the 5′-untranslated region of λX47 as well as the NH_2-terminal 12 residues of the signal peptide. The presence of an acceptor splicing sequence 5′ of this exon (Fig. 2) and the absence from it of the first six nucleotides from the λX47 cDNA clone suggest the presence of at least one additional exon. To signify this we have designated this exon as number 2 (Fig. 1*B* and Table I). Primer extension analysis using an oligonucleotide complementary to the first 27 bp coding the signal peptide has confirmed that the 5′-untranslated region of the p150,95 α subunit mRNA is longer than the 57 bp contained in λX47 (data not shown). The signal peptide is split, with the remaining 7 residues encoded in exon 3 together with the initial 28 amino acids of the mature polypeptide. One exon has been sequenced for the murine Mac-1 gene (42); it has precisely

[2] The GenBank accession numbers for clusters 1, 2, 3, and 4 are M29482, M29483, M29484, and M29485, respectively. The 5′- and 3′-sequenced regions of cluster 5 are M29486 and M29487, respectively.

TABLE I

Exons and introns of the p150,95 α subunit

Exon number	Domain[a]	Exon length	Intron length	Intron phase
		bp	*bp*	
2	5′-UT + SP (12 aa)	89	602	1
3	SP (7 aa)	106	619	2
4		104	300	1
5		71	154	0
6		112	2100 ± 200	1
7	I domain	131	86	0
8	I domain	146	243	2
9	I domain	154	604	0
10	I domain	151	615	1
11		74	163	0
12		130	412	1
13		143	181	0
14	Metal-binding domain	141	88	0
15	Metal-binding domain	210	6880 ± 200	0
16	Metal-binding domain	131	124	2
17		163	132	1
18		155	577	0
19		132	87	0
20		74	582	2
21		142	3700 ± 200	0
22		117	84	0
23		80	98	2
24		84	2300 ± 200	2
25		79	115	0
26		108	117	0
27		84	199	0
28		114	142	0
29		102	271	0
30	TM (22 aa)	111	250 ± 50	0
31	TM (4 aa) + CD + 3′-UT	>506		

[a] Abbreviations used are: UT, untranslated region; SP, signal peptide; aa, amino acids; TM, transmembrane domain; CD, cytoplasmic domain.

```
Hu. p150    ctctttcccag CC TTA GCA ACT TCT CTA GGT TTC AAC TTG GAC
-6                        L   A   T   S   L   G   F   N   L   D
                          *   *                   *   *   *   *   *
Mu. Mac-1   tatccttccag CC CTG GCC TTG TGT CAT GGC TTC AAT CTG GAC
-6                        L   A   L   C   H   G   F   N   L   D

Hu. p150    ACA GAG GAG CTG ACA GCC TTC CGT GTG GAC AGC GCT GGG TTT
5           T   E   E   L   T   A   F   R   V   D   S   A   G   F
            *   *                   *                       *   *
Mu. Mac-1   ACT GAA CAT CCC ATG ACC TTC CAG GAG AAT GCA AAA GGC TTT
5           T   E   H   P   M   T   F   Q   E   N   A   K   G   F

Hu. p150    GGA GAC AGC GTG GTC CAG TAT GCC AAC TCC TG gtgaggcccag
19          G   D   S   V   V   Q   Y   A   N   S
            *           *   *   *
Mu. Mac-1   GGA CAG AAT GTG GTC CAG CTT GGC GGA ACC AG gtaaagtacct
19          G   Q   N   V   V   Q   L   G   G   T
```

FIG. 3. **Alignment of exon 3 of the human p150,95 α subunit gene with the homologous region of the murine Mac-1 α subunit gene.** Exon and intron sequences are in *upper* and *lower case*, respectively. Amino acid residues appear below the nucleotide sequence and are numbered as in their respective mature proteins. Conserved amino acid residues are marked by an *asterisk*.

the same boundaries as p150,95 exon 3 (Fig. 3). The remainder of the first exonic cluster, including three more exons, continues until the beginning of the I domain.

Cluster two begins with the exons encoding the I domain, concludes with the exons encoding the first two cation binding repeats, and includes three exons in between these regions. By using the intron/exon organization as a criterion to define its limits, the I domain of the p150,95 α subunit extends from residues 125 to 318 and is encoded in four exons of 43, 48, 50, and 51 residues, respectively (Figs. 1 and 2). Comparison of the intron/exon organization of the I domain of the p150,95 α subunit gene with the organization of the homologous regions of the von Willebrand factor (43), factor B (44), and chicken cartilage matrix protein (45) genes shows a disparity in the number, length, and even phase class (46) of the introns within the I domain (Fig. 4). However, the introns flanking both the 5′ and 3′ ends of the I domain are conserved as phase class 1 in all cases in which this has been determined, consistent with the notion that this domain can be inserted as an evolutionary unit (Fig. 4).

Integrin α subunits contain three or four tandem repeats of a putative cation binding motif which is similar to the "EF-hand" motif of Ca^{2+}-binding proteins (26, 27). There are three such repeats in the p150,95 α subunit and other leukocyte integrins (16–20). The three tandem repeats are significantly homologous to one another, suggesting evolution by gene duplication. The three putative divalent cation-binding domains are encoded in separate exons, supporting this hypothesis. Four repeats which are more NH_2-terminal and do not contain divalent cation-binding sites have been conjectured based on similarities to sequences flanking the divalent cation-binding sites in leukocyte integrins (16–20) as well as some β1 integrins (21, 22). These repeats, which do not show statistically significant homology to one another, do not correlate with the p150,95 intron/exon boundaries, and therefore their evolutionary significance remains unclear.

An unusual inverted sequence was found within cluster 2. This 430-bp region is 83.7% homologous in nucleotide sequence to the first two divalent cation binding exons and the intervening 88-bp intron. This inversion is present within the intron following exon 10, the last exon of the I domain (*boxes* below the *line* in Fig. 1*A*). Alignment of these sequences (Fig. 5) shows a good correlation between the exons encoding the first and second cation-binding domains and their inverted homologues, as demonstrated by their preserved open reading frame, similar length, agreement with the consensus rule for splicing, as well as the phase class of their flanking and intervening introns. Remarkably, the intronic sequence in between but not outside both exons is conserved. The fact that this inverted sequence has been detected in cosmid clones from two different libraries rules out the possibility of a cloning artifact or a postcloning recombination event. No further inverted regions have been found within the sequenced regions of the gene.

The exon encoding the third divalent cation binding repeat is separated from the first and second by a 6.8-kb intron and begins cluster 3, which includes five more exons (Fig. 1). While the introns flanking the exons encoding the first and second cation binding repeats are all phase 0, the intron 3′ of the exon encoding the third cation-binding domain is phase 2 (Table I).

The VLA-4 α subunit, unlike any other integrin α subunit, shows a variable extent of processing to two cleaved fragments of 70 and 80 kDa, respectively (22). Relative to other integrin α subunits, a 5-residue insert is present in VLA-4, which contains a dibasic putative protease cleavage site (22). This small insert corresponds exactly with the splicing site between exons 19 and 20 in the p150,95 exon cluster 3 (Fig. 1*B*).

Cluster four is the smallest with only three exons, while cluster five includes six exons, 25–31 (Fig. 1*A*). The α subunits of the fibronectin receptor (24), the glycoprotein gpIIb/IIIa (25), and VLA-5 (23) are proteolytically cleaved, generating a small COOH-terminal fragment that in the mature protein is attached to the rest of the α subunit by a disulfide bond. Sequence analysis has shown that the dibasic protease cleavage sites are located in a 30-residue region that is present only in those α subunits that undergo proteolytic cleavage (23–25,

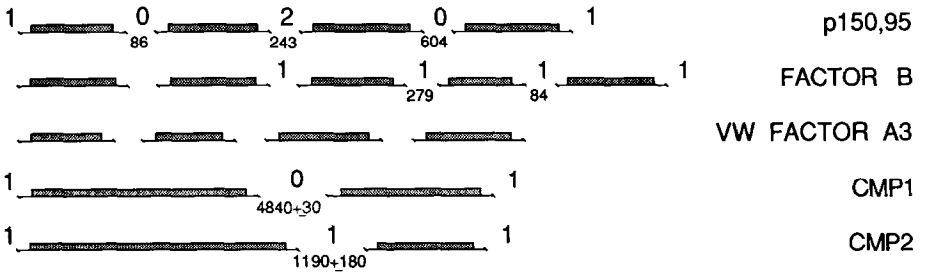

FIG. 4. **Intron-exon organization of I domains.** Structural features are from Refs. 43, 44, and 45. Exons are drawn to scale as *filled boxes*. When known, the length and phase of the introns, represented as *interrupted lines*, are indicated.

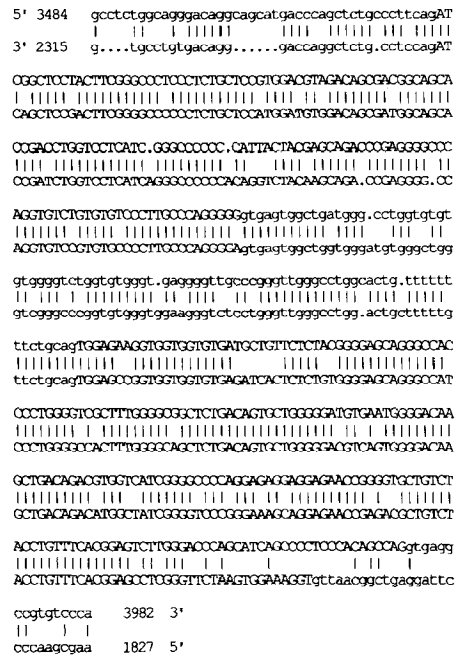

FIG. 5. **Sequence homology between the exons encoding the first and second divalent cation-binding domains and the inverted sequence.** p150,95 exon and intron sequences are indicated as *upper* and *lower case*, respectively. Putative exons and introns in the inverted sequences are shown similarly. The sequence alignment is extended beyond the area of homology.

47). This 30-residue region falls between exons 27 and 28 of the p150,95 α subunit gene (Fig. 1*B*).

The transmembrane domain, also included in cluster five, is encoded in two separate exons (Fig. 1*B*). Exon 30 encodes the last 15 residues of the extracellular domain together with the first 22 residues of the transmembrane domain, while exon 31 contains the last 4 residues of the transmembrane domain, the cytoplasmic domain, and at least 434 bp of the 3′-untranslated region, which includes the Alu sequence. Sequencing of the entire 1145-bp 3′-untranslated region has not been attempted, and therefore the possibility of additional introns cannot be excluded.

DISCUSSION

We have previously reported that the genes of the leukocyte integrin α subunits are present in the human genome as single copy genes on the short arm of chromosome 16 between bands p11 and p13.1 (40). However, no data have been available on the structure of these or any other integrin genes. In this paper we report the organization of an integrin α subunit, that of p150,95. The nucleotide sequence encoding the complete p150,95 α subunit polypeptide, as well as part of the 5′- and 3′-untranslated regions, is distributed in 31 exons that are grouped in five clusters and extend over 25 kb.

One exon from the integrin Mac-1 α subunit has been previously reported (42). Exon 3, which encodes the last 7 residues of the signal peptide and the initial 28 residues of the mature polypeptide, has exactly the same organization in the human p150,95 and mouse Mac-1 α subunit genes (42). The number of residues corresponding to the signal peptide and to the mature polypeptide as well as the phase of both flanking introns are preserved. This conserved intron/exon organization between different α subunits of different species suggests that the genomic organization of the other leukocyte integrin α subunits may be very similar. The almost identical distribution of the splicing sites in the region encoding the cation divalent binding domains of the p150,95 and LFA-1 α subunit[3] genes further supports this homology.

The I domain was initially described in the p150,95 α subunit as a region of 187 residues, from residue 149 to 335, that was absent from the α subunits of β1 and β2 integrins (16). Later its boundaries were defined by the presence of similar regions in the α subunits of Mac-1 (17–19), LFA-1 (20), and VLA-2 (21) and by its homology to nonintegrin protein domains (28–32). Using the initial criterion of absence from other integrin α subunits, it was difficult to define the exact beginning and end of the I domain, because there is little homology between these two classes of integrins just prior to and after the I domain. Since splicing sites often mark functional domains of proteins (48), we believe they are a more accurate criterion for defining the boundaries of the I domain, which we place at amino acids 125–318. This assignment is in good agreement with the definition of the I domain by homology to the A domains of von Willebrand factor (28), to sequences in complement components C2 (31) and factor B (32), to some regions of the α1 and α2 polypeptide of collagen type VI (30), and to the two repeating units of the chicken cartilage matrix glycoprotein (29).

Probably the best example of gene evolution by exon shuffling and intron insertion is the family of the serine proteases (49, 50). From the genomic organization of the different members of this family we have learned that only exons (or a group of exons) flanked at both ends by introns in the same phase class can undergo duplication into adjacent introns or insertion into introns of other genes by recombination; otherwise the reading frame would be disrupted (51). Comparison of the genomic organization of the p150,95 α subunit I domain with the organization of the homologus regions of the von Willebrand factor (43), factor B (44), and chicken cartilage matrix protein (45) genes has shown that while there is little consensus on the splice sites within the I domain, the splice sites at the beginning and end of the I domain are conserved as phase 1 introns. This supports the idea that the I domain has been inserted as a unit into a diverse range of proteins during evolution.

The different number, position, length, and phase of the introns within I domains can be explained by intron insertion

[3] R. Larson and T. A. Springer, unpublished data.

later in evolution. Intron insertion is considered a mechanism that can be exploited to accelerate protein evolution. Splice junctions are sites where sequence variation is more common, amino acids are almost always at the protein surface, and the length of turns can be varied by splice junction sliding (48). Different patterns of intron insertion may contribute to the development of the diverse binding specificities hypothesized for I domains. We therefore propose that the I domain was inserted as a primordial exon into ancestral genes for a number of different protein families. Subsequent evolution occurred by duplication and diversification of the entire gene to yield gene families (integrin α subunits, complement factor B and C2), duplication of the I domain within the gene (the repeats of von Willebrand factor and chicken cartilage matrix protein), and intron insertion and deletion within the I domain.

There is not an explanation at the moment for the presence of a sequence highly homologous to the exons encoding the first and second divalent cation binding repeats and their intervening intron in the antisense strand of the intron 3′ of the I domain. The exons in this inverted sequence do not contain stop codons, and thus there is no reason to consider them "pseudoexons." The fact that the homology also involves the intron between these exons suggests that a duplication with an inversion may have occurred. The high degree of sequence conservation suggests the inversion is quite recent. Alternatively, conservation may be through copying the regions against one another, as by gene conversion. The significance of a stem and loop structure which could form, with exons 11–15 looped out, is unclear. It is also possible that this inverted sequence is part of a functional gene contained on the antisense strand of the p150,95 α subunit gene locus. The existence of nested inverted genes encoding similar or unrelated proteins has recently been described both in humans (52) and Drosophila (53). Studies are under way to confirm this possibility.

Some integrin α subunits are proteolytically cleaved close to the COOH-terminal end of the molecule (23–25), while the VLA-4 α subunit is cleaved closer to the NH_2-terminal end of the molecule (22). Potential cleavage sites have been proposed in both cases in short sequences of approximately 30 and 5 residues, respectively, which are not present in the p150,95 α subunit (21–25, 47). In both cases the inserted sequence containing dibasic potential protease cleavage sites corresponds exactly with intron/exon boundaries in the p150,95 α subunit gene, supporting the idea that protease cleavage sites in some proteins correlate with splicing sites in their corresponding genes (48). Sequences around splice junctions usually show the highest degree of sequence variation between members of the same protein family (54). The sequences containing the dibasic protease cleavage sites may thus be the result of either intronic recombination by exon insertion or intron sliding by displacement of one of the intron/exon boundaries.

Recently, Brown *et al.* (55) published the genomic organization of the α subunit of the position-specific antigen 2 (PS2), a Drosophila integrin. It contains 12 exons. Alignment of p150,95 and PS2 α subunit sequences demonstrates significant correlation in their exon structures. PS2 exon 2 corresponds to a fusion of p150,95 exons 2 and 3, with the intervening sequence in the signal peptide eliminated. Exons 3, 4, and 5 of PS2 α correlate perfectly in intron position and phase with exons 4, 5, and 6 of p150,95 α. Exon 6 of PS2 α correlates perfectly with exon 11 of p150,95. The I domain exons 7–10 of p150,95 have no equivalent in PS2, and significantly PS2 exons 5 and 6 between which they would be inserted are separated by a phase I intron, which is the correct phase for insertion of I domains. Exon 7 in PS2 α corresponds to the beginning of p150,95 exon 12. The beginning and end of exon 8 of PS2, which is variably spliced and aligns poorly if at all with any p150,95 sequence, are the only exon boundaries in PS2 that do not correlate in position and phase with exon boundaries in p150,95. Exon 9 of PS2 is equivalent to a portion of p150,95 α exon 12 and a fusion of exons 13–20. Exon 10 of PS2 corresponds to exons 21 and 22 of p150,95. Exon 11 of PS2 aligns with exons 23–30 of p150,95 and contains additional serine-rich sequences not found in mammalian integrins at a position corresponding to the internal protease cleavage site in β1 and β3 integrins. The homology continues to the final exons, 12 of PS2 and 31 of p150,95, which begin after a conserved basic residue on the inner face of the membrane and continue into the 3′-untranslated region. These comparisons demonstrate significant conservation in exon organization of integrin α subunits, despite the evolutionary distance between insects and mammals and the likely disparity in function between PS2 and p150,95.

We have described the organization of the gene for an integrin α subunit. The intron/exon organization has revealed interesting features of importance for understanding the evolution of this gene. At least one exon is conserved for a different mouse leukocyte integrin α subunit, and considering the evolutionary distance between vertebrata and arthropoda the similarity to an insect integrin gene is remarkable. Further interesting comparisons await the characterization of other vertebrate integrin α subunit genes, particularly those of the α subunits of the β1 and β3 integrin subfamilies. Craik *et al.* (48) have shown evidence that the splicing sites in genes usually map to amino acid residues located at the protein surface. Information about the structure of a protein obtained from the organization of its gene may therefore be a guide in the design of experiments aimed to determine the structure-function relationship of the protein.

Acknowledgments—We gratefully acknowledge Drs. Hans Lehrach, Glen Evans, and Stuart Orkin for providing the genomic libraries, Richard Larson for communicating unpublished results, and Ken Barfield for help with figures. We thank Dr. Nick Brown for an alignment of the PS2 and p150,95 α subunit sequence.

REFERENCES

1. Hynes, R. O. (1987) *Cell* **48,** 549–554
2. Ruoslahti, E., and Pierschbacher, M. D. (1987) *Science* **238,** 491–497
3. Kishimoto, T. K., Larson, R. S., Corbi, A. L., Dustin, M. L., Staunton, D. E., and Springer, T. A. (1989) *Adv. Immunol.* **46,** 149–182
4. Hemler, M. E. (1988) *Immunol. Today* **9,** 109–113
5. Anderson, D. C., and Springer, T. A. (1987) *Annu. Rev. Med.* **38,** 175–194
6. Takada, Y., Wayner, E. A., Carter, W. G., and Hemler, M. E. (1988) *J. Cell. Biochem.* **37,** 385–393
7. Sonnenberg, A., Modderman, P. W., and Hogervorst, F. (1988) *Nature* **336,** 487–489
8. Wayner, E. A., and Carter, W. G. (1987) *J. Cell Biol.* **105,** 1873–1884
9. Takada, Y., Strominger, J. L., and Hemler, M. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 3239–3243
10. Holzmann, B., McIntyre, B. W., and Weissman, I. L. (1989) *Cell* **56,** 37–46
11. Miller, L. J., Schwarting, R., and Springer, T. A. (1986) *J. Immunol.* **137,** 2891–2900
12. Keizer, G. D., Borst, J., Visser, W., Schwarting, R., de Vries, J. E., and Figdor, C. G. (1987) *J. Immunol.* **138,** 3130–3136
13. Schwarting, R., Stein, H., and Wang, C. Y. (1985) *Blood* **65,** 974–983
14. te Velde, A. A., Keizer, G. D., and Figdor, C. G. (1987) *Immunology* **61,** 261–267

15. Anderson, D. C., Miller, L. J., Schmalstieg, F. C., Rothlein, R., and Springer, T. A. (1986) *J. Immunol.* **137,** 15–27
16. Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S., and Springer, T. A. (1987) *EMBO J.* **6,** 4023–4028
17. Corbi, A. L., Kishimoto, T. K., Miller, L. J., and Springer, T. A. (1988) *J. Biol. Chem.* **263,** 12403–12411
18. Arnaout, M. A., Gupta, S. K., Pierce, M. W., and Tenen, D. G. (1988) *J. Cell Biol.* **106,** 2153–2158
19. Pytela, R. (1988) *EMBO J.* **7,** 1371–1378
20. Larson, R. S.,Corbi, A. L., Berman, L., and Springer, T. A. (1989) *J. Cell Biol.* **108,** 703–712
21. Takada, Y., and Hemler, M. E. (1989) *J. Cell Biol.* **109,** 397–407
22. Takada, Y., Elices, M. J., Crouse, C., and Hemler, M. E. (1989) *EMBO J.* **8,** 1361–1368
23. Suzuki, S., Argraves, W. S., Arai, H., Languino, L. R., Pierschbacher, M. D., and Ruoslahti, E. (1987) *J. Biol. Chem.* **262,** 14080–14085
24. Argraves, W. S., Suzuki, S., Arai, H., Thompson, K., Pierschbacher, M. D., and Ruoslahti, E. (1987) *J. Cell Biol.* **105,** 1183–1190
25. Poncz, M., Eisman, R., Heidenreich, R., Silver, S. M., Vilaire, G., Surrey, S., Schwartz, E., and Bennett, J. S. (1987) *J. Biol. Chem.* **262,** 8476–8482
26. Kretsinger, R. H. (1980) *Crit. Rev. Biochem.* 119–174
27. Szebenyi, D. M. E., Obendorf, S. K., and Moffat, K. (1981) *Nature* **294,** 327–332
28. Shelton-Inloes, B. B., Titani, K., and Sadler, J. E. (1986) *Biochemistry* **25,** 3164–3171
29. Argraves, W. S., Deak, F., Sparks, K. J., Kiss, I., and Goetinck, P. F. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 464–468
30. Koller, E., Winterhalter, K. H., and Trueb, B. (1989) *EMBO J.* **8,** 1073–1077
31. Bentley, D. R. (1986) *Biochem. J.* **239,** 339–345
32. Mole, J. E., Anderson, J. K., Davison, E. A., and Woods, D. E. (1984) *J. Biol. Chem.* **259,** 3407–3412
33. Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B., and Lehrach, H. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 4129–4133
34. DiLella, A. G., and Woo, S. L. C. (1985) *Focus* **7,** 1–5
35. Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842
36. Evans, G. A., and Wahl, G. M. (1987) *Methods Enzymol.* **152,** 604–610
37. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
38. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
39. Henikoff, S. (1984) *Gene (Amst.)* **28,** 351–359
40. Corbi, A. L., Larson, R. S., Kishimoto, T. K., Springer, T. A., and Morton, C. C. (1988) *J. Exp. Med.* **167,** 1597–1607
41. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
42. Sastre, L., Roman, J. M., Teplow, D. B., Dreyer, W. J., Gee, C. E. Larson, R. S., Roberts, T. M., and Springer, T. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 5644–5648
43. Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., and Titani, K. (1986) *Cold Spring Harbor Sympo. Quant. Biol.* **51,** 515–523
44. Campbell, R. D., and Porter, R. R. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4464–4468
45. Kiss, I., Deák, F., Holloway, R. G., Jr., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., and Goetinck, P. F. (1989) *J. Biol. Chem.* **264,** 8126–8134
46. Sharp, P. A. (1981) *Cell* **23,** 643–646
47. Loftus, J. C., Plow, E. F., Jennings, L. K., and Ginsberg, M. H. (1988) *J. Biol. Chem.* **263,** 11025–11028
48. Craik, C. S., Sprang, S., Fletterick, R., and Rutter, W. J. (1982) *Nature* **299,** 180–182
49. Rogers, J. (1985) *Nature* **315,** 458–459
50. Patthy, L. (1985) *Cell* **41,** 657–663
51. Patthy, L. (1987) *FEBS Lett.* **214,** 1–7
52. Miyajima, N., Horiuchi, R., Shibuya, Y., Fukushige, S-I., Matsubara, K-I., Toyoshima, K., and Yamamoto, T. (1989) *Cell* **57,** 31–39
53. Henikoff, S., Keene, M. A., Fechtel, K., and Fristrom, J. W. (1986) *Cell* **44,** 33–42
54. Craik, C. S., Buchman, S. R., and Beychok, S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1384–1388
55. Brown, N. H., King, D. L., Wilcox, M., and Kafatos, F. C. (1989) *Cell* **59,** 185–195