MOLECULAR AND CELLULAR BIOLOGY, Apr. 1990, p. 1818–1821
0270-7306/90/041818-04$02.00/0

Vol. 10, No. 4

Correction of the N-Terminal Sequences of the Human Plastin Isoforms by Using Anchored Polymerase Chain Reaction: Identification of a Potential Calcium-Binding Domain

CHING-SHWUN LIN,[1] RUEDI H. AEBERSOLD,[2] AND JOHN LEAVITT[1]*

California Institute for Medical Research, 2260 Clove Drive, San Jose, California 95128,[1] and Biomedical Research Centre, University of British Columbia, Vancouver, British Columbia V5T 1W5, Canada[2]

Received 31 October 1989/Accepted 21 December 1989

Plastins are a family of at least three cytoplasmic protein isoforms that are expressed differentially between cells of the hematopoietic lineages and cells of solid tissues. Expression of the L-plastin isoform appears to be restricted to replicating blood cells, and the two T-plastin isoforms appear to be restricted to replicating cells of solid tissues. However, L-plastin is induced in many human solid tumor-derived cells. We used the anchored polymerase chain reaction technique to amplify and clone the missing 5′ ends of plastin mRNAs. We found that both plastin isoforms contain a potential calcium binding site near the N terminus.

Plastin was first noted as a polypeptide that appeared to be induced in human fibrosarcomas and other human tumor-derived cell lines (2, 6–8). Initially, expression of a single plastin species was thought to be restricted in normal cells to hematopoietic cell lineages and activated in other cell types as a consequence of neoplastic transformation (2, 8). When plastin cDNAs were isolated from the human fibrosarcoma cell line HuT-14, we discovered that at least two distinct isoforms were coexpressed in this neoplastic cell line: L-plastin, which was normally expressed in blood leukocytes, and T-plastin, which was found in normal cells derived from solid tissues such as fibroblasts, endothelial and epithelial cells, keratinocytes, and melanocytes (8; J. Leavitt and C.-S. Lin, unpublished results). The high cytoplasmic abundance of L-plastin and its normal restriction to expression in circulatory cells led to speculation that its induction in tumor cells may govern some phenotypic characteristics of neoplastic cells, such as reduced cytoplasmic spreading on substrates and reduced anchorage dependence, both of which are normal properties of blood cells (2, 4). Initially, no clues as to the cellular function of plastin isoforms were apparent when we examined the amino acid sequences of the two isoforms. Other than the appearance of L-plastin in human cancer cells and its cytoplasmic location, our knowledge of plastins was limited to the observation that L-plastin was a substrate for phosphorylation at serine residues (1, 2, 5, 12).

A number of subsequent observations led us to question whether we had correctly identified the N-terminal amino acid sequence and the 5′ ends of these cDNAs for both plastin isoforms. Initially, we identified the N-terminal sequence (8) on the basis of the finding that the open reading frames for both isoform cDNAs ended at a methionine residue (amino acid 58 in Fig. 1). We were unable to identify the N terminus of L-plastin by microsequencing because of N blocking (8). Second, the L-plastin clone contained an insert cDNA of 3.7 kilobases, the size of the L-plastin mRNA. Because the L-plastin cDNA was the size of the mRNA, we assumed that at least part of the upstream noncoding sequence of this cDNA represented the 5′ untranslated region of the L-plastin mRNA. We then attempted to use this putative 5′ sequence to map the transcription start site in genomic clones in order to locate the L-plastin gene promoter. Even though the L-plastin genomic clones used in this work contained at least 22 kilobases of DNA upstream from the putative initiation codon (characterized by sequence analysis), we could not identify a sequence in this DNA clone that corresponded to the 5′ (untranslated) sequence found in the L-plastin cDNA. Third, we discovered that the 5′ end of the T-plastin cDNA was actually a second cDNA encoding another protein. This second cDNA was linked to the T-plastin cDNA at the two overlapping nucleotides (TC) adjacent to the putative initiation codon for T-plastin (now known to be codon 58; Fig. 1). Fourth, the molecular weight of each of the two plastin isoforms determined by amino acid sequences predicted by the cDNAs was 64,000, whereas the apparent molecular weight determined by migration rates in sodium dodecyl sulfate-polyacrylamide gels was 68,000 (2, 8). We initially accepted this discrepancy, since migration rates in gels are not strictly determined by molecular size; for example, tropomyosins 1 and 3 (9) and mutant and wild-type β-actin (10, 16) are two sets of related proteins that have the same amino acid length but different migration rates in sodium dodecyl sulfate-polyacrylamide gels. The concerns listed above, considered together, raised the possibility that both the L- and the T-plastin cDNA clones were the products of spurious fusions between two unrelated cDNAs ligated together during library preparation. This artifact was also found in a cDNA clone from the same library that encoded the full length of tropomyosin 3 (9).

One additional finding was in conflict with our initial conclusion relating to the sizes of the plastin polypeptides. We initially detected a ninth oligopeptide sequence by microsequencing that corresponded to a calcium-binding sequence found in troponin C and calmodulin (3), yet we did not find this sequence encoded by either the T- or L-plastin cDNA. We dismissed this sequence as a potential artifact resulting from the small amount of plastin used for microsequencing. Nevertheless, our inability to find the calcium-binding sequence encoded in plastin cDNA sequences, and difficulties in tracking the transcription start site in plastin genomic clones, led us to question whether we had correctly identified the N-terminal sequence of both plastin isoforms.

PCR amplification of the 5′ ends of plastin mRNAs. Our

* Corresponding author.

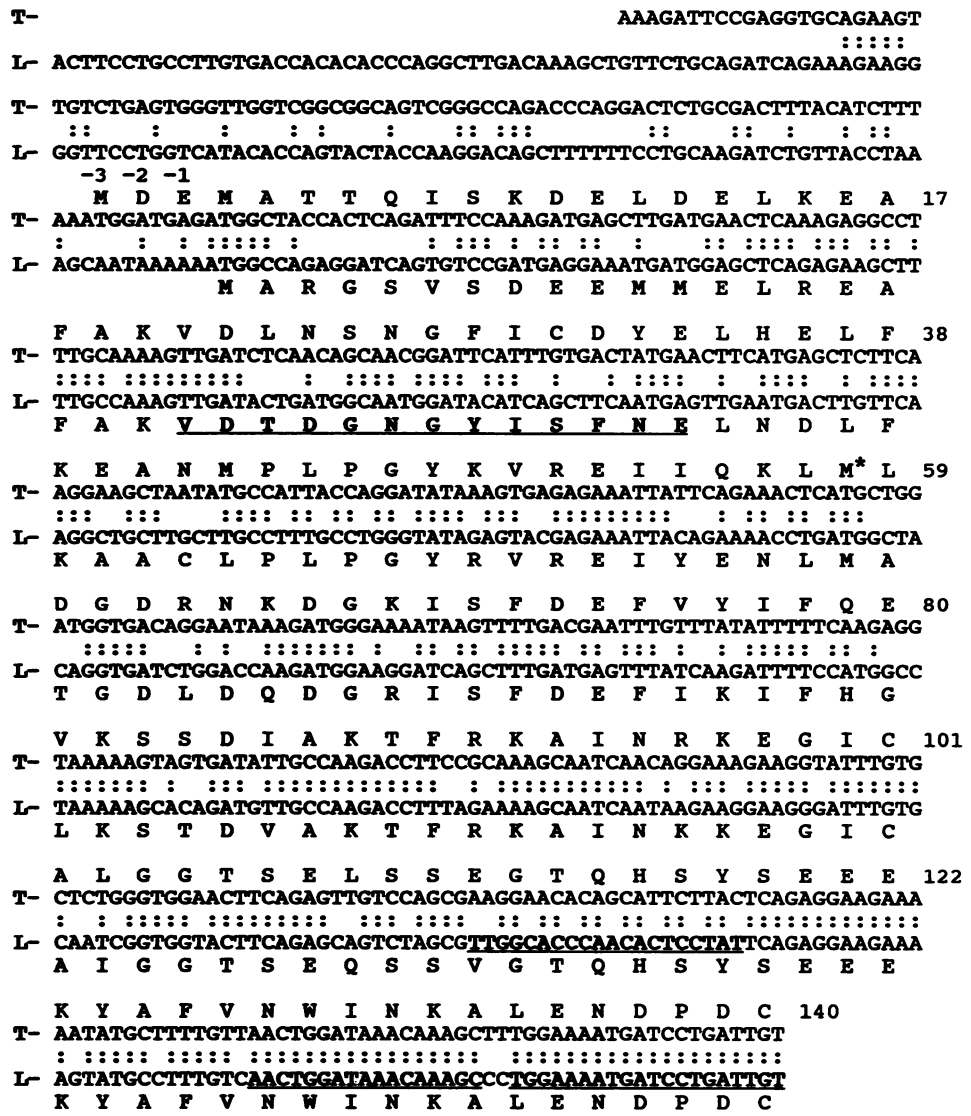

FIG. 1. Nucleotide and deduced amino acid sequences of the 5′ ends of plastin mRNAs derived by PCR amplification and cloning of the 5′ cDNA sequences synthesized from plastin mRNAs. Identical nucleotides between the two isoforms are indicated by a colon. The calcium-binding domain is underlined under the L-plastin peptide sequence. The three oligonucleotide primers used in the PCR amplification experiments are also underlined (see text for explanation). The asterisk on amino acid residue 58 indicates the first codon of the previously published sequences (8).

desire to identify the correct 5′ mRNA sequence for both plastin isoforms led us to use the anchored polymerase chain reaction (PCR) technique (11) to amplify the 5′ ends of both plastin mRNAs (Fig. 1). To synthesize the first cDNA strand, the primer 5′-ACAATCAGGATCATTTTCCA-3′ (1 μg) was annealed to HuT-14 poly(A) RNA (8 μg), and cDNA was synthesized by reverse transcription (10-μl reaction volume containing 250 mM KCl and 10 mM Tris hydrochloride [pH 8.3]; the reaction mixture was incubated first at 80°C for 3 min and then at 49°C for 45 min). A 20-μl amount of reverse transcription buffer (24 mM Tris hydrochloride [pH 8.3], 16 mM $MgCl_2$, 8 mM dithiothreitol, 0.4 mM each of the four deoxyribonucleotides), 1 μl of [α-^{32}P]dCTP (3,000 Ci/mmol), and 0.5 μl (100 U) of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc.) were then added to the 10-μl primer-template mixture. After 1 h of incubation at 37°C, the mixture was extracted with phenol-chloroform, and nucleic acids were precipitated twice with ammonium acetate and ethanol. For tailing of cDNAs, the product of first-strand cDNA synthesis was suspended in 40 μl of water and mixed with 5 μl of 10 mM dGTP, 5 μl of tailing buffer (1.4 M potassium cacodylate [pH 6.9], 0.3 M Tris base, 10 mM $CoCl_2$, 2 mM 2-mercaptoethanol), and 1 μl (15 U) of terminal deoxynucleotide transferase (International Biotechnologies, Inc.). For PCR amplification, a total of three primers were used in two PCR experiments. One primer, 5′-CTCGAGCTCCCCCCCCCC CCCC-3′, was used as the upstream primer in both attempts. It contained a *Sac*I site (GAGCTC) for convenient cloning and a dC tail for annealing with the dG tail of the cDNAs. The other two primers were downstream primers. One of them, 5′-GGGGGCCCGCTTTGTTTATCCAGTT-3′, was used in the first attempt and contained an *Apa*I site (GGG CCC) for convenient cloning and 17 bases that are complementary to both plastin isoform mRNAs starting at codon 128 (Fig. 1). The other downstream primer, 5′-AAAAGGGC

CCATAGGAGTGTTGGGTGCCAA-3', which was used in the second attempt, also contained an *Apa*I site, but the 3'-end sequence of 20 bases was complementary only to L-plastin isoform mRNAs starting at codon 112 (Fig. 1). The PCR reactions were set up as follows. Half of the dG-tailed cDNA product (in 87.5 μl of water) was mixed with 10 μl of PCR buffer (100 mM Tris [pH 8.3], 500 mM KCl, 15 mM $MgCl_2$, 0.1% [wt/vol] gelatin), 10 μl of a mixture of 2 mM each of the four deoxyribonucleotides, 1 μg (in 1 μl of water) of each of the upstream and downstream primers, and 0.5 μl (2.5 U) of *Taq* polymerase (Perkin Elmer-Cetus). This mixture was overlayed with 100 μl of mineral oil and incubated in a Temperature Cycler (Ericomp) for a total of 30 cycles. The cycles consisted of 94°C, 1 min, 45°C (first five cycles) or 55°C (last 25 cycles), 1.5 min, and 72°C, 2.5 min. The reaction was finished with an extra 10-min incubation at 72°C. For cloning and sequencing of amplification products, the PCR products were either digested with *Sac*I and *Apa*I or attached to *Eco*RI linkers after a kinase reaction. Both modified PCR products were then cloned into Bluescript plasmid (Strategene) and sequenced with Sequenase (U.S. Biochemicals).

5'-Terminal sequences of plastin mRNAs. In the first PCR amplification, we used an oligonucleotide primer that was homologous to both T- and L-plastin-coding sequences (codons 128 to 133; Fig. 1). The product was a single band of DNA in an agarose gel of about 500 base pairs. After *Sac*I-*Apa*I double digestion, this band was no longer visible. We therefore opted to clone this undigested PCR product with *Eco*RI linkers. Three clones were isolated that contained identical sequences which matched 229 bases with the 5' end of the T-plastin cDNA that we previously published (Fig. 1; 8). This PCR-cloned sequence extended for another 265 bases upstream and contained an additional 60 codons of the T-plastin open reading frame heading upstream from the 5' end of the previously published coding sequence for T-plastin. This PCR sequence contained a *Sac*I site at base 206 from its 5' end that could explain the disappearance of the DNA band after a *Sac*I digestion.

In the second PCR amplification attempt using an oligonucleotide primer specific for L-plastin only (complementary to codons 112 to 118; Fig. 1), we obtained a single DNA fragment approximately 500 base pairs in length. The amplified DNA was cloned after *Sac*I-*Apa*I digestion or attachment of *Eco*RI linkers. Eleven clones were sequenced, and all exhibited identical sequences except for one that was shorter at its 5' end. The longer sequence overlapped with 183 base pairs of the 5'-end sequence previously reported for L-plastin. The sequence extended for another 308 base pairs upstream, which contained an additional 57 codons of the L-plastin open reading frame heading upstream from the 5' end of the previously published coding sequence for L-plastin (8). Surprisingly, the L-plastin 5' sequence also contained a *Sac*I site at position 278 (Fig. 1); therefore, the larger PCR clones that were obtained after *Sac*I-*Apa*I double digestion were in fact cloned from partially digested PCR fragments. We have successfully used the 5' end of the new cDNA to identify the gene promoter and transcription start site through analysis of the structure of a genomic clone of the L-plastin gene (Lin and Leavitt, unpublished results). Therefore, we have concluded that this additional cDNA sequence encodes the true N-terminal amino acid sequence for L-plastin (Fig. 1).

Like the previously published plastin cDNA sequences, the PCR-amplified L- and T-plastin cDNA sequences have a high degree of identity with each other except for the 5'-end noncoding regions. However, the T-plastin sequence has three extra potential codons upstream from the methionine codon (Met-1), the codon analogous to the first methionine codon of L-plastin (Fig. 1). If both methionines (Met-1 and Met-minus 3) are used as initiation codons, the three upstream codons (encoding Met-Glu-Asp) would generate a second T-plastin polypeptide of 630 amino acids differing from the other T isoform (627 amino acid) by approximately two charge units. Both T isoforms were translated in vitro in equal amounts from mRNA that hybridized to the single T-plastin cDNA, and the T-plastin cDNA used as a probe also detected only one mRNA species in Northern (RNA) analysis (8). Therefore, the presence of two possible initiation codons (codons 1 and −3) in the T-plastin cDNA raises the possibility that both T-plastin isoforms may be generated from a single mRNA species. At present, however, we cannot rule out the alternative possibility that there are two distinct T-plastin mRNAs that have the same length. To explore the origin of the two T isoforms further, we constructed two T-plastin cDNAs that start at the 5' end with either codon 1 or codon −3. Using these cDNAs, we expressed in *Escherichia coli* two T-plastin polypeptides with the same isoelectric points and apparent molecular weights as the two T-plastins in human cells. In *E. coli*, the longer cDNA supports the synthesis of only the more acidic (pI 5.4) T isoform and the shorter cDNA supports the synthesis of the more basic (pI 5.5) T isoform (8; Lin and Leavitt, unpublished results).

Identification of a potential calcium-binding domain in the plastin sequence. A ninth oligopeptide derived from microsequencing of L-plastin, which contained the consensus calcium-binding site, was not encoded in the original cDNAs (8). The sequence of this oligopeptide closely matched those of the three calcium-binding sequences of human calmodulin and the single calcium-binding sequence of human skeletal muscle troponin C (Fig. 2; 3, 15). This sequence was encoded in the PCR-amplified N-terminal domain of both plastin isoforms, starting at amino acid 21 and ending at amino acid 33 (Fig. 1). The oligopeptide sequence obtained by microsequencing of L-plastin isolated from a T-lymphoblastoid cell line matched the L-plastin sequence derived from human fibrosarcoma mRNA amplified by PCR at 9 of 11 residues and exhibited conservative differences at the remaining 2 residues. The conserved amino acids that potentially interact with calcium ion or are important in EF-hand helix formation (X, Y, Z, G, I, −X, and −Z [3, 15]) are indicated above these sequences in Fig. 2. Protein microsequencing detected a conservative split sequence for L-plastin at amino acid 24 (Asp and Glu), which indicated the presence of an allelic polymorphism at this locus in the lymphoblastoid cell line used as the source of L-plastin (8). In the analogous sequence of T-plastin, which diverged from L-plastin at 6 of 13 amino acids, the oxygen-containing side chain at position Y was donated by an asparagine residue (Fig. 2) instead of an aspartic acid, conserving the calcium-chelating potential of this sequence (3, 15). The L-plastin sequence matched the three calmodulin sequences collectively at 11 of 13 positions and the troponin C sequence at 8 of 13 positions.

The discovery of a potential calcium-binding site in L-plastin is consistent with the identity of this isoform with a protein previously identified in human macrophages by Pacaud (13) and Pacaud and Molla (14) that exhibited the same molecular weight (approximately 70,000) and isoelectric point (pI 5.3) as L-plastin in two-dimensional gels. We have previously shown that L-plastin is one of the most

		X		Y		Z	G		I	-X			-Z
L-plastin microsequencing	val	asp	ile	asp/glu	gly	asn	gly	tyr		ala	phe		glu
	**	**		**	**	**	**	**			**		**
L-plastin PCR sequence	val	asp	thr	asp	gly	asn	gly	tyr	ile	ser	phe	asn	glu
	**	**				**	**		**				**
T-plastin PCR sequence	val	asp	leu	asn	ser	asn	gly	phe	ile	cys	asp	tyr	glu
	**	**		**	**	**	**		**		**		**
human calmodulin-1	val	asp	ala	asp	gly	asn	gly	thr	ile	asp	phe	pro	glu
		**		**	**	**	**	**	**	**			**
human calmodulin-2	phe	asp	lys	asp	gly	asn	gly	tyr	ile	ser	ala	ala	glu
		**		**	**		**						**
human calmodulin-3	ala	asp	ile	asp	gly	asp	gly	gln	val	asn	tyr	glu	glu
	**	**		**	**		**		**		**		**
human troponin C	val	asp	glu	asp	gly	ser	gly	thr	ile	asp	phe	glu	glu

FIG. 2. Comparison of the predicted calcium-binding domains of plastin isoforms with related domains of calmodulin and troponin C. Matched residues are indicated in bold letters with double asterisks. Matched residues important for calcium ion interaction (X, Y, Z, −X, and −Z) or for EF-hand helix structure (G and I) are indicated above the sequences (3, 15).

abundant proteins of human macrophages, especially of this size and charge (2). The macrophage protein was shown to bind to polymerized actin in a calcium-dependent manner (13, 14). Also, Matsushima and co-workers (12) have purified L-plastin from human macrophages (identified as L-plastin by sequence [8]) and have demonstrated that it is phosphorylated at serine residues.

We thank Russell Doolittle (University of California at San Diego) for helping us identify the calcium-binding sequence through searching his protein sequence data base. We also thank Paul Matsudaira (Whitehead Institute, Massachusetts Institute of Technology) for helpful comments on the manuscript. We acknowledge the assistance of Ross Eccleshall and Maureen Wang in our initial PCR amplification attempt.

This work was supported by Public Health Service grants CA34763 and CA49423 (to J.L.) from the National Institutes of Health.

LITERATURE CITED

1. **Anderson, N. L., M. A. Gemmell, P. M. Coussens, S.-I. Murau, and E. Huberman.** 1985. Specific protein phosphorylation in human promyelocytic HL-60 leukemia cells susceptible or resistant to induction of cell differentiation by phorbol-12-myristate-13-acetate. Cancer Res. **45**:4995–5003.
2. **Goldstein, D., J. Djeu, J. Latter, S. Burbeck, and J. Leavitt.** 1985. Abundant synthesis of the transformation-induced protein of neoplastic human fibroblasts, plastin, in normal lymphocytes. Cancer Res. **45**:5643–5647.
3. **Kretsinger, R. H.** 1980. Structure and evolution of calcium-modulated proteins. Crit. Rev. Biochem. **8**:119–174.
4. **Leavitt, J.** 1989. Changes in gene expression associated with neoplastic transformation of human fibroblasts: the abundant cellular proteins, p. 239–266. *In* G. Milo (ed.), Human fibroblast transformation: molecular and genetic mechanisms. CRC Press, Inc., Boca Raton, Fla.
5. **Leavitt, J., G. Bushar, T. Kakunaga, H. Hamada, T. Hirakawa, D. Goldman, and C. Merril.** 1982. Variations in expression of mutant b-actin accompanying incremental increases in human fibroblast tumorigenicity. Cell **28**:259–268.
6. **Leavitt, J., D. Goldman, C. Merril, and T. Kakunaga.** 1982. Changes in gene expression accompanying chemically induced malignant transformation of human fibroblasts. Carcinogenesis **3**:61–70.
7. **Leavitt, J., and T. Kakunaga.** 1980. Expression of a variant form of actin and additional polypeptide changes following chemical-induced in vitro neoplastic transformation of human fibroblasts. J. Biol. Chem. **255**:1650–1661.
8. **Lin, C.-S., R. H. Aebersold, S. B. Kent, M. Varma, and J. Leavitt.** 1988. Molecular cloning and characterization of plastin: a human leukocyte protein expressed in transformed human fibroblasts. Mol. Cell. Biol. **8**:4659–4668.
9. **Lin, C.-S., and J. Leavitt.** 1988. Cloning and characterization of a cDNA encoding the transformation-sensitive tropomyosin isofrom 3 from tumorigenic human fibroblasts. Mol. Cell. Biol. **8**:160–168.
10. **Lin, C.-S., S.-Y. Ng, P. Gunning, L. Kedes, and J. Leavitt.** 1985. Identification and order of sequential mutations in β-actin genes isolated from increasingly tumorigenic human fibroblast strains. Proc. Natl. Acad. Sci. USA **82**:6995–6999.
11. **Loh, E. Y., J. F. Elliott, S. Cwirla, L. L. Lanier, and M. Davis.** 1989. Polymerase chain reaction with a single-sided specificity: analysis of T cell receptor d chain. Science **243**:217–220.
12. **Matsushima, K., M. Shiroo, H.-F. Kung, and T. D. Copeland.** 1988. Purification and characterization of a cytosolic 65-kilodalton phosphoprotein in human leukocytes whose phosphorylation is augmented by stimulation with interleukin-1. Biochemistry **27**:3765–3770.
13. **Pacaud, M.** 1986. Separation and identification of the major constituents of cytoplasmic gels from macrophages. Eur. J. Biochem. **156**:521–530.
14. **Pacaud, M., and A. Molla.** 1987. Cytoplasmic gels from macrophages: evidence for the involvement of four proteins in the calcium-dependent solation of actin networks. Eur. J. Biochem. **165**:139–145.
15. **Strynadka, N. C. J., and M. N. G. James.** 1989. Crystal structures of the helix-loop-helix calcium binding proteins. Annu. Rev. Biochem. **58**:951–998.
16. **Vandekerckhove, J., J. Leavitt, T. Kakunaga, and K. Weber.** 1980. Coexpression of a mutant β-actin and two normal β- and γ-cytoplasmic actins in a stably transformed human cell line. Cell **22**:893–899.