Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by the association of LIM and SH3 domains

Catherine Tomasetto, Christel Moog-Lutz, Catherine H. Régnier, Valérie Schreiber, Paul Basset, Marie-Christine Rio*

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/U184 INSERM/ULP, BP 163, 67404 Illkirch Cédex, CU de Strasbourg, France

Received 23 July 1995

Abstract MLN 50 was previously identified in a cDNA library of breast cancer metastasis. In this study, we show that MLN 50, which is expressed at a basal level in normal tissues, is overexpressed in 8% of human breast carcinomas most often together with *c-erb*B-2. MLN 50 cDNA encodes a putative protein of 261 residues, named Lasp-1 (LIM and SH3 protein) since it contains a LIM motif and a domain of *Src* homology region 3 (SH3) at the amino- and the C-terminal parts of the protein, respectively. Thus, Lasp-1 defines a new LIM protein subfamily.

Key words: Lasp-1; LIM; SH3; Breast cancer; c-*erb*B-2

1. Introduction

Breast cancer is one of the most common types of cancer affecting women. Despite improvement in the early detection of small tumors and in treatment, associated metastases remain the leading cause of mortality for breast cancer patients [1]. MLN 50 (Lasp-1) cDNA was identified from a breast cancer-derived metastatic lymph node cDNA library by differential hybridization using malignant (metastatic lymph node) vs. non-malignant (breast fibroadenoma and normal lymph node) tissues. Chromosomal mapping showed that the Lasp-1 gene is localized on the q12–q21 region of the long arm of chromosome 17. This region is known to be altered in 20–30% of breast cancers, the most common modification being the amplification of the proto-oncogene *c-erb*B-2 [2–4]. In breast cancer cell lines we found that overexpression of Lasp-1 RNA was correlated with amplification of the gene and to *c-erbB*-2 amplification/overexpression, suggesting that Lasp-1 and *c-erbB*-2 belong to the same amplicon [5]. The aim of this study was to determine the frequency of Lasp-1 overexpression in human breast cancer and to characterize the structure of the putative protein encoded by this gene.

2. Materials and methods

2.1. Tissues

Surgical specimens obtained from the Hôpitaux Universitaires de Strasbourg were frozen in liquid nitrogen for RNA extraction.

2.2. RNA preparation and analysis

Surgical specimens were homogenized in guanidinium isothiocyanate lysis buffer and RNAs purified by centrifugation through cesium chloride cushions [6]. RNAs were fractionated by electrophoresis on 1% agarose, 2.2 M formaldehyde gels [7], transferred to nylon membrane (Hybond N; Amersham, Arlington Heights, IL) and immobilized by baking for 2 h at 80°C.

2.3. Probe preparation and hybridization

The Lasp-1 probe corresponded to a 1.0-kb *Bam*HI fragment released from MLN 50 cDNA (Accession X82456) [5] subcloned into pBluescript (Stratagene, La Jolla, CA). The c-erbB-2-specific probe was a *Eco*RI internal fragment of MLN 19 cDNA [5]. The RNA loading control probe 36B4 was an internal 0.7-kb *Pst*I fragment [8].

Northern blots were hybridized at 42°C in 50% formamide, 5 × SSC, 0.4% ficoll, 0.4% polyvinylpyrrolidone, 20 mM sodium phosphate pH 6.5, 0.5% SDS, 10% dextran sulfate and 100 μg/ml denatured salmon sperm DNA, for 36–48 h with ^{32}P-labeled probes [9] diluted to 0.5–1.10^6 cpm/ml. Stringent washings were performed at 60°C in 0.1 × SSC and 0.1% SDS. Blots were subjected to autoradiography at −80°C for 24 h.

2.4. Sequence analysis

Sequence analysis was performed using the GCG sequence analysis package (Wisconsin package version 8.0, Genetics Computer Group, Madison, WI). The Lasp-1 cDNA sequence and its deduced putative protein sequence were used to search the complete combined GenBank/EMBL database (Release 84/42) and the complete SwissProt database (Release 31) with BLAST [10] and FastA [11] programs, respectively. The LIM motif and other consensus sequences of Lasp-1 were further identified by the Motif program in the PROSITE dictionary (Release 12). The sequence alignments were obtained automatically by using the program PileUp [12].

3. Results and discussion

3.1. Lasp-1 is overexpressed in breast carcinomas

To study Lasp-1 mRNA distribution, we carried out Northern blot analysis using the cDNA probe (see Section 2). A single 4.0-kb mRNA band was detected at a basal level in all normal human tissues examined, including lymph node, skin, lung, stomach, colon and liver (data not shown), in non-malignant breast fibroadenomas (Fig. 1A, lanes 13–17) and in breast hyperplasia (Fig. 1A, lane 18). High levels of Lasp-1 mRNA were found in 8% (5/61) of primary breast cancers (Fig. 1A, lane 8 and data not shown) and in 40% (2/5) of metastatic lymph nodes derived from breast cancer (Fig. 1A, lanes 1 and 2, and data not shown). *c-erb*B-2 hybridization of the same blots showed that Lasp-1 and *c-erb*B-2 were simultaneously overexpressed in most human breast cancer tissues (Fig. 1A, lanes 1, 2 and 8). Nevertheless, some malignant samples displayed c-erbB-2 overexpression independantly of Lasp-1 overexpression (Fig. 1A, lane 12). Thus, these results showed that Lasp-1 which is widely expressed at a basal level in normal tissues is overexpressed in some breast cancers and their derivative metastatic lymph nodes. It has been shown previously that, in breast cancer cell lines, overexpression of Lasp-1 is correlated

*Corresponding author. Fax: (33) (88) 65 32 01.

SSDI 0014-5793(95)01040-8

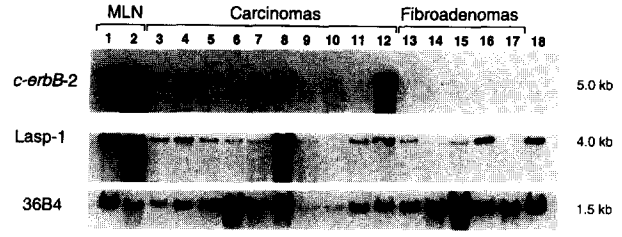


Fig. 1. Northern blot analysis of Lasp-1 mRNA expression in human breast tissues. Total RNA (10 μg) were loaded, transferred and hybridized with ^{32}P-labeled probes specific for c-erbB-2, Lasp-1 and to the RNA loading control 36B4. Approximative transcript sizes are indicated (right). RNA extracted from breast metastatic lymph nodes (MLN; lanes 1 and 2), carcinomas (lanes 3–12), fibroadenomas (lanes 13–17) and hyperplasia (lane 18).

to Lasp-1 gene amplification [5]. Moreover, Lasp-1 has been mapped to the human chromosome 17 close to the *c-erb*B-2 oncogene [2,5]. Altogether, these results suggest that amplification of the q11–q21 region of chromosome 17, one of the most common events occurring in human breast cancers and presumed to target the *c-erb*B-2 gene [3,4], could lead to Lasp-1 overexpression.

3.2. Lasp-1 cDNA sequence

The complete Lasp-1 cDNA sequence has been established from four independent cDNA clones. Both sense and antisense strands have been sequenced. The longest cDNA clone contained 3848 bp (Fig. 2A), a size consistent with that of the Lasp-1 transcript, suggesting that this clone should correspond to the full length cDNA. The first ATG codon (nucleotide position 76–78, Fig. 2A) had a favorable context for initiation of translation [13] and a classical AATAAA poly(A) addition signal sequence [14] was located 13 bp upstream of the poly(A) stretch (Fig. 2A). At the nucleotide level, sequence homologies were found with 22 expressed sequence tag (EST) [15]. Some of these sequences were redundant and they were mostly located at the 3′ untranslated end of the molecule (Fig. 2B). These EST were identified from different human cDNA libraries established from normal tissues (prostate gland, liver and fetal spleen and brain) and cells (white blood and pancreatic islet cells). The presence of Lasp-1 transcripts in all these samples is in good agreement with our finding of an ubiquitous expression of Lasp-1 mRNA (data not shown).

3.3. Lasp-1 putative protein sequence

The deduced Lasp-1 open reading frame encoded a 261 amino acid protein (Fig. 2A), with a molecular weight of 30 kDa and a pHi of 6.5. The putative protein showed several consensus sequences [16]: an amidation site (GGKR, residues 203–206), several phosporylation sites characteristic of cAMP- and cGMP-dependent protein kinases (RRDS, 143–146), casein kinase II (SGGE, 134–137; SAAD, 214–217; SFQD, 222–225), proteine kinase C (TEK, 15–17; TCK, 34–36; SYR, 151–153) and tyrosine kinases (KGYEKKPY, 45–52; RDSQDGSSY, 144–152). Moreover, a LIM cysteine-rich domain [17] was identified at the amino terminal end (residues 1–51) and a *Src* homology region 3 (SH3) domain [18] at the C-terminal end (residues 197–261) of the protein. In addition, the primary sequence of Lasp-1 that we deduced contained two probable tyrosine phosphorylation sites (Y_{52} and Y_{152}). It has previously been reported that *Src* homology region 2 (SH2) domains [19,20] bind to phospho-tyrosine [21], and that short tripeptides, following the phospho-tyrosine residue, are responsible for the specificity of this interaction [22]. Interestingly, the tripeptides following the Lasp-1 putative phospho-tyrosine residues (Y_{52} and Y_{152}) are compatible with those already described for SH2 binding motifs [22], suggesting that Lasp-1 could interact with SH2 containing proteins.

3.4. Lasp-1 LIM domain

The LIM domain is an arrangment of eight cysteine and histidine residues (C-X_2-C-$X_{16/23}$-H-X_2-C-X_2-C -X_2-C-$X_{16/21}$-C-$X_{2/3}$-C/D/H) and is found in a number of invertebrate and vertebrate proteins [17]. The LIM domain defines a zinc binding structure that is properly folded in presence of Zn. The LIM generic name was drawn from the names of the first three proteins identified containing such a domain (lin-11, Isl-1 and mec-3). The family of LIM containing proteins is continously increasing and can be subdivided into distinct groups [17]. One group designated LIM-HD, includes proteins having two LIM domains associated with a homeodomain (lin-11, Isl-1, mec-3). Another group designated LIM-only, includes proteins exhibiting one (CRIP), two (CRP, TSF3, RBTN1, RBTN2, RBTN3) or three (zyxin) LIM domains. Recently, a new group designated LIM-k has been described; it includes proteins having two LIM domains associated with a kinase domain [23].

Sequence alignment of the Lasp-1 LIM domain showed a high homology (66% identity and 80% similarity) with that of the putative *C. elegans* YLZ4 protein (Accession P34417), although the overall protein homology was low (36% identity and 55% similarity). The LIM domain of YLZ4 does not perfectly fit the LIM consensus, since the first two cysteines are spaced by four instead of two residues, leading to a gap in the alignment (Fig. 3A). Among other LIM containing proteins, besides the conserved LIM residues, additional homologies were found in the human cysteine-rich intestinal protein CRIP [24], the rat cysteine-rich protein CRP [25] and the sun flower transcription factor SF3 (TSF3) [26]. The physiological function of these proteins is not yet known, although a role for CRIP in intestinal zinc absorption has been suggested. Moreover, CRP was identified as a binding partner for another LIM protein, the zyxin, presumed to have regulatory or signalling functions in focal adhesion plaques [27–29]. The interaction between these two proteins is mediated through specific LIM domain interactions [30]. Thus, the LIM domain can be considered as a protein/protein binding modular interface, similar to the SH2 and SH3 domains [30].

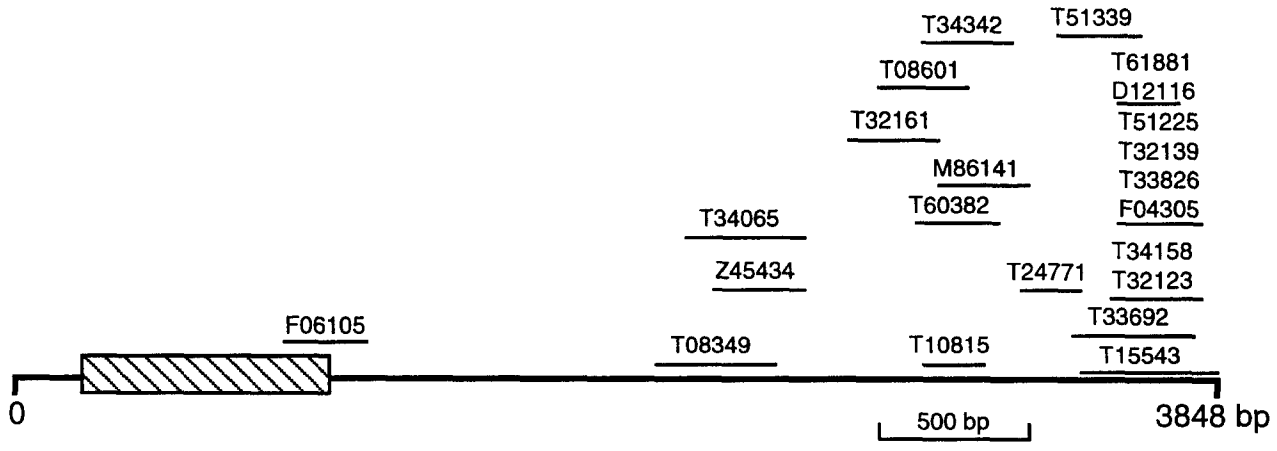
→

Fig. 2. Nucleotide and deduced amino acid sequences of human Lasp-1. (A) Nucleotides and amino acids are numbered on the left and right, respectively. The residues involved in the LIM domain are underlined and printed in bold characters and those of the SH3 domain in bold characters (●) indicates tyrosine residues that are possibly phosphorylated. An asterisk denotes the termination codon. The signal for polyadenylation is underlined. (B) The Lasp-1 cDNA is schematically represented (0–3848 bp), the shaded box indicates the protein-coding region. The relative positions of the different EST, homologous to Lasp-1, are indicated together with their corresponding length and accession numbers.

A

```
   1 GCCTCCCGCCAGCTCGCCTCGGGGAACAGGACGCGCGTGAGCTCAGGCGTCCCCGCCCCAGCTTTTCTCGGAACCATGAACCCCAACTGCGCCCGGTGCG
                                                                  M  N  P  N  C  A  R  C  G    9

 101 GCAAGATCGTGTATCCCACGGAGAAGGTGAACTGTCTGGATAAGTTCTGGCATAAAGCATGCTTCCATTGCGAGACCTGCAAGATGACACTGAACATGAA
       K  I  V  Y  P  T  E  K  V  N  C  L  D  K  F  W  H  K  A  C  F  H  C  E  T  C  K  M  T  L  N  M  K   42

 201 GAACTACAAGGGCTACGAGAAGAAGCCCTACTGCAACGCACACTACCCCAAGCAGTCCTTCACCATGGTGGCGGACACCCCGGAAAACCTTCGCCTCAAG
       N  Y  K  G  Y  E  K  K  P  Y  C  N  A  H  Y  P  K  Q  S  F  T  M  V  A  D  T  P  E  N  L  R  L  K   75
                                    •
 301 CAACAGAGTGAGCTCCAGAGTCAGGTGCGCTACAAGGAGGAGTTTGAGAAGAACAAGGGCAAAGGTTTCAGCGTAGTGGCAGACACGCCCGAGCTCCAGA
       Q  Q  S  E  L  Q  S  Q  V  R  Y  K  E  E  F  E  K  N  K  G  K  G  F  S  V  V  A  D  T  P  E  L  Q  R 109

 401 GAATCAAGAAGACCCAGGACCAGATCAGTAATATAAAATACCATGAGGAGTTTGAGAAGAGCCGCATGGGCCCTAGCGGGGGCGAGGGCATGGAGCCAGA
       I  K  K  T  Q  D  Q  I  S  N  I  K  Y  H  E  E  F  E  K  S  R  M  G  P  S  G  G  E  G  M  E  P  E  142

 501 GCGTCGGGATTCACAGGACGGCAGCAGCTACCGGCGGCCCCTGGAGCAGCAGCAGCCTCACCACATCCCGACCAGTGCCCCGGTTTACCAGCAGCCCCAG
       R  R  D  S  Q  D  G  S  S  Y  R  R  P  L  E  Q  Q  Q  P  H  H  I  P  T  S  A  P  V  Y  Q  Q  P  Q   175
                                    •
 601 CAGCAGCCGGTGGCCCAGTCCTATGGTGGCTACAAGGAGCCTGCAGCCCCAGTCTCCATACAGCGCAGCGCCCCAGGTGGTGGCGGGAAGCGGTACCGCG
       Q  Q  P  V  A  Q  S  Y  G  G  Y  K  E  P  A  A  P  V  S  I  Q  R  S  A  P  G  G  G  G  K  R  Y  R  A 209

 701 CGGTGTATGACTACAGCGCCGCCGACGAGGACGAGGTCTCCTTCCAGGACGGGGACACCATCGTCAACGTACAGCAGATCGACGACGGCTGGATGTACGG
       V  Y  D  Y  S  A  A  D  E  D  E  V  S  F  Q  D  G  D  T  I  V  N  V  Q  Q  I  D  D  G  W  M  Y  G  242

 801 GACGGTGGAGCGCACCGGCGACACGGGGATGCTGCCGGACAACTACGTGGAGGACATCTGAACCCGGAGCGCCCCCATCTGTCTTCAGCACATTCCACGG
       T  V  E  R  T  G  D  T  G  M  L  P  D  N  Y  V  E  D  I  *                                       261

 901 CATCGCATCCGTCCTGGGCGTGAGCCGTCCATTCTTCAGTGTCTCTGTTTTTTAAAACCTGCGACAGCTTGTGATTCCTACCCCTCTTCCAGCTTCTTTT
1001 GCCAACTGAAGCCTTCTTCTGCCACTTCTGCGGGCTCCCTCCTCTGGCAGGCTTCCCCCGTGATCGACTTCTTGGTTTTCTCTCTGGATGGAACGGGTAT
1101 GGGCCTCTCTGGGGGAGGCAGGGCTGGAATGGGAGACCTGTTGGCCTGTGGGCCTCACCTGCCCCTCTGTTCTCTCCCCTCACATCCTCCTGCCCAGCTC
1201 CTCACATACCCACACATTCCAGGGCTGGGGTGAGCCTGACTGCCAGGACCCCAGGTCAGGGGCTCCCTACATTCCCCAGAGTGGGATCCACTTCTTGGTT
1301 CCTGGGATGGCGATGGGGACTCTGCCGCTGTGTAGGGACCAAGTGGGATGGGCTCTACCTCTCTTTCTCAAAGAGGGGGCTCTGCCCACCTGGGGTCTCTC
1401 TCCCTACCTCCCTCCTCAGGGGCAACAACAGGAGAATGGGGTTCCTGCTGTGGGGCGAATTCATCCCCTCCCCGCGCGTTCCTTCGCACACTGTGATTTT
1501 GCCCTCCTGCCCACGCAGACCTGCAGCGGGCAAAGAGCTCCCGAGGAAGCACAGCTTGGGTCAGGTTCTTGCCTTTCTTAATTTTAGGGACAGCTACCGG
1601 AAGGAGGGGAACAAGGAGTTCTCTTCCGCAGCCCCTTTCCCCACGCCCACCCCCAGTCTCCAGGGACCCTTGCCTGCCTCCTAGGCTGGAAGCCATGGTC
1701 CCGAAGTGTAGGGCAAGGGTGCCTCAGGACCTTTTGGTCTTCAGCCTCCCTCAGCCCCCAGGATCTGGGTTAGGTGGCCGCTCCTCCCTGCTCCTCATGG
1801 GAAGATGTCTCAGAGCCTTCCATGACCTCCCCTCCCCAGCCCAATGCCAAGTGGACTTGGAGCTGCACAAAGTCAGCAGGGACCACTAAATCTCCAAGAC
1901 CTGGTGTGCGGAGGCAGGAGCATGTATGTCTGCAGGTGTCTGACACGCAAGTGTGTGAGTGTGAGTGTGAGAGATGGGGCGGGGGTGTGTCTGTAGGTGT
2001 CTCTGGGCCTGTGTGTGGGTGGGGTTATGTGAGGGTATGAAGAGCTGTCTTCCCCTGAGAGTTTCCTCAGAACCCACAGTGAGAGGGGAGGGCTCCTGGG
2101 GCAGAGAAGTTCCTTAGGTTTTCTTTGGAATGAAATTCCTCCTTCCCCCCATCTCTGAGTGGAGGAAGCCCACCAATCTGCCCTTTGCAGTGTGTCAGGG
2201 TGGAAGGTAAGAGGTTGGTGTGGAGTTGGGGCTGCCATAGGGTCTGCAGCCTGCTGGGGCTAAGCGGTGGAGGAAGGCTCTGTCACTCCAGGCATATGTT
2301 TCCCCATCTCTGTCTGGGGCTACAGAATAGGGTGGCAGAAGTGTCACCCTGTGGGTGTCTCCCTCGGGGGCTCTTCCCCTAGACCTCCCCCTCACTTACA
2401 TAAAGCTCCCTTGAAGCAAGAAAGAGGGTCCCAGGGCTGCAAAACTGGAAGCACAGCCTCGGGGATGGGGAGGGAAAGACGGTGCTATATCCAGTTCCTG
2501 CTCTCTGCTCATGGGTGGCTGTGACAACCCTGGCCTCACTTGATTCATCTCTGGTTTTCTTGCCACCCTCTGGGAGTCCCCATCCCATTTTCATCCTGAG
2601 CCCAACCAGGCCCTGCCATTGGCCTCTTGTCCCTTGGCACACTTGTACCCACAGGTGAGGGGCAGGACCTGAAGGTATTGGCCTGTTCAACAATCAGTCA
2701 TCATGGGTGTTTTTGTCAACTGCTTGTTAATTGATTTGGGGATGTTTGCCCCGAATGAGAGGTTGAGGAAAAGACTGTGGGTGGGGAGGCCCTGCCTGAC
2801 CCATCCCTTTTCCTTTCTGGCCCCAGCCTAGGTGGAGGCAAGTGGAATATCTTATATTGGGGGCTCGGGGAGGCAGAGAATCTCTTGGGAG
2901 TCTTGGGTGGCGCTGGTGCATTCTGTTTCCTCTTGATCTCAAAGCACAATGTGGATTTGGGGACCAAAGGTCAGGGACACATCCCCTTAGAGGACCTGAG
3001 TTTGGGAGAGTGGTGAGTGGAAGGGAGGAGCAGCAAGAAGCAGCCTGTTTTCACTCAGCTTAATTCTCCTTCCCAGATAAGGCAAGCCAGTCATGGAATC
3101 TTGCTGCAGGCCCTCCCTCTACTCTTCCTGTCCTAAAAATAGGGGCCGTTTTCTTACACACCCCCAGAGAGAGGAGGGACTGTCACACTGGTGCTGAGTG
3201 ACCGGGGGCTGCTGGGCGTCTGTTCTTTACCAAAACCATCCATCCCTAGAAGAGCACAGAGCCCTGAGGGGCTGGGCTGGGCTGGGCTGAGCCCCTGGTC
3301 TTCTCTACAGTTCACAGAGGTCTTTCAGCTCATTTAATCCCAGGAAAGAGGCATCAAAGCTAGAATGTGAATATAACTTTTGTGGGCCAATACTAAGAAT
3401 AACAAGAAGCCCAGTGGTGAGGAAAGTGCGTTCTCCCAGCACTGCCTCCTGTTTTCTCCCTCTCATGTCCCTCCAGGGAAAATGACTTTATTGCTTAATT
3501 TCTGCCTTTCCCCCCTCACACATGCACTTTTGGGCCTTTTTTTATAGCTGGAAAAAACAAAATACCACCCTACAAACCTGTATTTAAAAAGAAACAGAAA
3601 TGACCACGTGAAATTTGCCTCTGTCCAAACATTTCATCCGTGTGTATGTGTATGTGTGTGAGTGTGTGAAGCCGCCAGTTCATCTTTTTATATGGGGTTG
3701 TTGTCTCATTTTGGTCTGTTTTGGTCCCCTCCCTCGTGGGCTTGTGCTCGGGATCAAACCTTTCTGGCCTGTTATGATTCTGAACATTTGACTTGAACCA
3801 CAAGTGAATCTTTCTCCTGGTGACTCAAATAAAAGTATAATTTTTA
```

B



A

				Identity	Similarity
Lasp-1	H. sapiens	(1-51)	..MNPNCAR. .CGKIVYPTE KVNCLDKFWH KACFHCETCK MTLNMKNYKG YEKKPYCNAH Y.....PKQ	%	%
YLZ4	C. elegans	(1-52)	..-SKKC--E DC--T---V- ELK----V-H -QC-KCTVCG ---------- -D-R--CDPH -.....--T	66	80
hCRIP	H. sapiens	(1-55)	...M-KCPK. .CN-E--FA- R-TS-G-D-H RPCLKC-KCG K--TSGGHAE H-G---C-HP CYVAMFG--G	46	55
rCRP2	R. norvegicus	(1-56)	..-ASKCPK. .CD-T--FA- --SS-G-D-H -FCLKC-RCN K--TPGGHAE HDG--FC-KP CYATLFG--G	46	55
-		(119-180)	TGEPNMCP-. .CN-R--FA- --TS-G-D-H RPCLRC-RCS K--TPGGHAE HDGQ--CHKP CYG.....	44	56
TSF3	H. annuus	(5-64)	TGTTQKCT.. VCE-T--LVD -LVANQRVYH --C-RCHHCN S--KLS-FNS FDGVV-CRHH FDQLFKRTGS	35	52
-		(104-162)	EGTRDKCN.. ACA-----I- R-KVDGTAYH R-C-KCCHGG C-ISPS--IA H-GRL-CKHH HIQLFKKKGN	40	54

B

				Identity	Similarity
Lasp-1	H. sapiens	(196-261)	RSAP GGGGKRYRAV YDYSAADEDE VSFQDGDTIV NVQQIDDGWM YGTVERTGDT GMLPDNYVED I*	%	%
YLZ3	C. elegans	(134-200)	I-PT -KA-FAVK-I ---A---K-- I--LE--I-- -CEK------ T---Q--LQW ----A---QP HK	57	74
EMS1	H. sapiens	(486-550)	DEYE NDL-YTAV-L ---Q--GD-- I--DPDDI-T -IEM-----W R-VCK..-RY -LF-A----L RQ*	44	65
ABP1	S. cerevisiae	(526-592)	PEKK PKENPWAT-E ---D--EDN- LT-VEN-K-I -IEFV--D-W L-EL-KD-SK ----S---SL GN*	33	60
h/fyn	H. sapiens	(76-141)	-TRG -T-VTLFV-L ---E-RT--D L--HK-EKFQ ILNSSEGD-W EARSLT--E- -YI-S---AP VD	35	56
h/src	H. sapiens	(78-144)	AGPL A--VTTFV-L ---ESRT-TD L--KK-ERLQ I-NNTEGD-W LAHSLS--Q- -YI-S---AP SD	33	55
h/frg	H. sapiens	(71-135)	-GVS -I-VTLFI-L ---E-RT--D LT-TK-EKFH ILNNTEGD-W EARSLSS-K- -CI-S---AP VD	32	52
h/yes	H. sapiens	(85-152)	PAGL T--VTIFV-L ---E-RTTED L--KK-ERFQ IINNTEGD-W EARSIA--KN -YI-S---AP AD	30	50

Fig. 3. Comparison of Lasp-1 LIM and SH3 domains with those from various proteins. (A) The consensus LIM domain residues are in bold, identical residues are dashed, (.) indicates gap in the alignment. (B) SH3 conserved or semi-conserved residues in more than half of the aligned sequences are in bold. Identical residues are dashed. (.) indicates gaps in the alignment.

3.5. Lasp-1 SH3 domain

SH3 is a small protein domain of 60 amino acids, first identified as a conserved sequence in the non-catalytic amino terminal part of the *src* protein tyrosine kinase [19,20]. The function of SH3 domain remains unclear. A number of proteins involved in the tyrosine kinase signal transduction pathway contain SH3 domains [21]. This domain is also found in proteins of unrelated function, such as proteins associated with the cytoskeleton [18]. SH3 containing proteins are usually located close to the plasmic membrane, suggesting that the SH3 domain may be implicated in localizing them to this cell compartment. Accordingly, the involvement of the SH3 domain in the targetting of the adaptor molecule Grb2 has been demonstrated [31]. Hints concerning the function were obtained from the crystallographic resolution of different SH3 domains, showing that, in spite of limited sequence similarity, the overall secondary and tertiary structures are well conserved revealing a domain that is independantly folded in the protein [32]. The SH3 domain has been reported to be able to interact with proteins through specific proline-rich SH3 binding regions [33].

Sequence alignment revealed homology of the Lasp-1 C-terminal part with several SH3 domains (Fig. 3B), including that of EMS1 [34], a human homolog of the *src* tyrosine kinase substrate cortactin [35]. The strongest conservation of this SH3 domain (57% identity and 74% similarity) was found with the putative YLZ3 protein of *C. elegans* (Accession P34416), although the overall protein homology is low (23% identity and 40% similarity). Interestingly, on the F42H10.3 cosmid which contains part of the *C. elegans* chromosome III [36], the gene encoding YLZ3 lies next to the gene encoding YLZ4 which was previously shown to contain the LIM domain most homologous with that of Lasp-1 (Fig. 3A). This may reflect processes of evolution that can lead to the association in the same protein of two functional domains, originally encoded by two consecutive genes.

In conclusion, Lasp-1 carries a LIM and a SH3 domain, which represent modular structures present in various proteins, often in association with other functional domains. Thus, two LIM protein subfamilies have already been shown to contain a LIM domain associated with either a homeo or a kinase domain [17]. In addition, SH3 domains are often found in association with SH2, pleckstrin homology (PH) and kinase domains [37–39]. To date, Lasp-1, which is the first protein containing LIM and SH3 domains, defines a new LIM protein subfamily. These domains are known to be involved in protein/protein interactions important for different cellular processes, such as transcription, cell transformation and cell signalling [17,21,32]. The ubiquitous expression of Lasp-1 in human adult tissues suggests a basic cellular function for this protein. Lasp-1 is overexpressed in 8% of breast carcinomas. Thus, the Lasp-1 protein which may be involved as an adaptor molecule in a signalling pathway, may contribute to cell transformation and/or tumor progression.

Acknowledgements: We are indebted to P. Chambon for his support and interest in this work and we thank P. Simpson for critical reading of the manuscript. We also thank C. Wendling and S. Vicaire for technical assistance. This work was supported by funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Centre Hospitalier Universitaire Régional, the Mutuelle Générale de l'Éducation Nationale, the Groupement de Recherches et d'Études sur les Génomes (Grant 94/50), the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale Française, the Ligue Nationale contre le Cancer and a grant to P. Chambon from the Fondation Jeantet. V. Schreiber is a recipient of a post-doctoral fellowship from the Association pour la Recherche sur le Cancer.

References

[1] Helzlsouer, K.J. (1994) Curr. Opin. Oncol. 6, 541–548.
[2] Fukushige, S.I., Matsubara, K.I., Yoshida, M., Sasaki, M., Suzuki, T., Semba, K., Toyoshima, K. and Yamamoto, T. (1986) Mol. Cell. Biol. 6, 955–958.
[3] Slamon, D.J, Clark, G.M., Wong, S.G., Lewin, W.J., Ullrich, A. and McGuire, W.L. (1987) Science 235, 177–182.
[4] Slamon, D.J., Godolphin, W. Jones, L.A., Holt, J.A., Wong, S.G., Keith, D.E., Lewin, W.J., Stuart, S.G., Udove, J., Ullrich, A. and Press, M.F. (1989) Science 244, 707–712.
[5] Tomasetto, C., Régnier, C.H., Moog-Lutz, C., Mattei, M.G., Chenard, M.P., Lidereau, R., Basset, P. and Rio, M.C. (1995) Genomics, 28, 267–376.

[6] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 52–94.
[7] Lehrach, H., Diamond, G., Wozney, J.M. and Boedtker, T. (1977) Biochemistry 16, 4743–4751.
[8] Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F. Krust, A. and Chambon, P. (1982) Nucleic Acids Res. 10, 7895–7903.
[9] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 112, 195–203.
[10] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[11] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448
[12] Feng, D.F. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.
[13] Kozak, M. (1987) Nucl. Acids Res. 15, 8125–8149.
[14] Wahle, E. and Keller, W. (1992) Annu. Rev. Biochem. 61, 419–440.
[15] Weinstock, K.G., Kirkness, E.F., Lee, N.H., Earle-Hughes, J.A. and Venter, J.C. (1994) Curr. Opin. Biotech. 5, 599–603.
[16] Kemp, B.E. and Pearson, R.B. (1990) Trends Biochem. Sci. 15, 342–346.
[17] Sanchez-Garcia, I. and Rabbitts, T.H. (1994) Trends Genet. 10, 315–320.
[18] Musacchio, A., Gibson, T., Lehto, V.P. and Saraste, M. (1992) FEBS Lett. 307, 55–61.
[19] Sadowski, I., Stone, J.C. and Pawson, T. (1986) Mol. Cell. Biol. 6, 4396–4408.
[20] Mayer, B.J., Hamaguchi, M. and Hanafusa, H. (1988) Nature (London) 332, 272–275.
[21] Schlessinger, J. (1994) Curr. Opin. Genet. Dev. 4, 25–30.
[22] Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King F., Roberts, T., Ratnofsky, S., Lechleider, R.J., Neel, B., Birge, R.B., Fajardo, J.E., Chou, M.M., Hanafusa, H., Schaffausen, B. and Cantley, L.C. (1993) Cell 72, 767–778.
[23] Mizuno, K., Okano, I., Ohashi, K., Nunoue, K., Kuma, K., Miyata, T. and Nakamura, T. (1994) Oncogene 9, 1605–1612.
[24] Tsui, S.K.W., Yam, N.Y.H., Yu Lee, C. and Waye, M.M.Y. (1994) Biochem. Biophys. Res. Commun. 205, 497–505.
[25] Liehaber, S.A., Every, J.G., Urbaneck, M., Wang, X. and Cooke, N.E. (1990) Nucl. Acids Res. 18, 3871–3879.
[26] Baltz, R., Evrard, J.C., Domon, C. and Steinmetz, A. (1992) Plant Cell 12, 1465–1466.
[27] Crawford, A.W., Michelsen, J.W. and Beckerle, M.C. (1992) J. Cell Biol. 116, 1381–1393.
[28] Crawford, A.W., Pino, J.D. and Beckerle, M.C. (1994) J. Cell Biol. 124, 117–127.
[29] Sadler, I., Crawford, A.W., Michelsen, J.W. and Beckerle, M.C. (1992) J. Cell Biol. 119, 1573–1587.
[30] Schmeichel, K.L. and Berckerle, M.C. (1994) Cell 79, 211–219.
[31] Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V. and Schlessinger, J. (1993) Cell, 74, 83–91.
[32] Morton, C.J. and Campbell, I.D. (1994) Curr. Biol. 4, 615–617.
[33] Alexandropoulos, K., Cheng, G. and Baltimore, D. (1995) Proc. Natl. Acad. Sci. USA 92, 3110–3114.
[34] Schuuring, E., Verhoeven, E., Wolter, J.M. and Michalides, J.A.M. (1992) Oncogene 7, 355–361.
[35] Wu, H., Reynolds, A.B., Kanner, S.B., Vines, R.R. and Parsons, T. (1991) Mol. Cell Biol. 11, 5113–5124.
[36] Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulton, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Wilkinson-Sproat, J. and Wohldman, P. (1994) Nature (London) 368, 32–38.
[37] Pawson, T. and Schlessinger, J. (1993) Curr. Biol. 3, 4345–4342.
[38] Mayer, B.J. and Baltimore, D. (1993) Trends Cell Biol. 3, 8–13.
[39] Mayer, B.J., Ren, R., Clark, K.L. and Baltimore, D. (1993) Cell 73, 629–630.