

# Structure of Human Clathrin Light Chains

## CONSERVATION OF LIGHT CHAIN POLYMORPHISM IN THREE MAMMALIAN SPECIES\*

(Received for publication, July 22, 1988)

Antony P. Jackson‡ and Peter Parham

From the Department of Cell Biology, Stanford University, Stanford, California 94305

**Complementary DNAs (cDNA) encoding the brain and non-brain forms of the human clathrin light chains LC<sub>a</sub> and LC<sub>b</sub> have been isolated, sequenced, and compared with their homologues in cow and rat. The significant differences that distinguish LC<sub>a</sub> from LC<sub>b</sub> and the brain from non-brain forms show remarkable preservation in all three species. These features include the position and sequence of the brain-specific inserts, a totally conserved region of 22 residues near the amino terminus, the LC<sub>b</sub>-specific phosphorylation site, the heavy chain binding site, and a distinctive pattern of cysteine residues near the carboxyl terminus. Unorthodox sequences for translation initiation and polyadenylation are found for LC<sub>b</sub> contrasting with LC<sub>a</sub> which exhibits orthodox regulatory sequences.**

Small insertions in human LC<sub>a</sub> revealed a duplicated sequence of 13 residues that flank the 22-residue conserved region. Only the carboxyl-terminal copy of this sequence is present in LC<sub>b</sub>. All sequences are consistent with the heavy chain binding site comprising an  $\alpha$ -helical central region of the light chains. The hydrophobic face of this helix, which is presumed to interact with the heavy chain, is highly conserved between LC<sub>a</sub> and LC<sub>b</sub>, whereas the hydrophilic face shows considerable divergence. To help define the carboxyl-terminal limit of the heavy chain binding region, the epitope recognized by the CVC.6 monoclonal antibody was localized to residues 192–208 of LC<sub>a</sub> with glutamic acid 198 being of most importance.

The faithful preservation of clathrin light chain polymorphism in three mammalian species provides evidence supporting a functional diversification of the brain and non-brain forms of LC<sub>a</sub> and LC<sub>b</sub>.

In cells of higher vertebrates clathrin-coated pits and vesicles participate in receptor-mediated pathways of endocytosis and intracellular transport of macromolecules. Polymerization of clathrin triskelions on the cytoplasmic face of membranes occurs concomitantly with the clustering of receptors and formation of a coated pit. Invagination then permits formation of a closed vesicle coated with a characteristic clathrin lattice of hexagons and pentagons. On entry of the

vesicle into the cytoplasm the clathrin coat is depolymerized and recycled (1, 2).

Each clathrin triskelion consists of three heavy chains and three light chains. In the cow two genes code for distinct light chains, LC<sub>a</sub> and LC<sub>b</sub>,<sup>1</sup> that are expressed in all tissues. These genes undergo tissue-specific mRNA splicing in the brain to yield larger forms of LC<sub>a</sub> and LC<sub>b</sub>, containing additional insertion sequences of 30 and 18 amino acids, respectively (3, 4).

The precise role of the clathrin light chains is uncertain. *In vitro* evidence that they bind calmodulin and are essential for the activity of an uncoating ATPase points to them being regulatory elements in clathrin function. Neither is the purpose of the light chain polymorphism established. One clue is that LC<sub>b</sub> is specifically phosphorylated both *in vitro* (5) and *in vivo* (6) by a casein-kinase II-like activity.

Clathrin has also been studied in yeast, a single-celled eukaryotic organism (7, 8). The role of clathrin in endocytosis, secretion, and intracellular transport has still to be defined but, given the nature of this organism, is unlikely to involve the targeted uptake of a large number of specific proteins. It is therefore of interest that yeast clathrin contains only a single light chain.<sup>2</sup> This raises the possibility that the evolution of light chain polymorphism is associated with the development in complex multicellular organisms of receptor-mediated systems for the uptake of multiple macromolecules. The validity of this idea will hinge on the extent to which polymorphic features of the clathrin light chains are preserved in different species. To examine this question we have determined the primary structure of four forms of human clathrin light chains and compared them with light chains from the cow and the rat.

### MATERIALS AND METHODS

**cDNA Libraries**—Human brain-type light chains were isolated from a human retinal cDNA library kindly supplied by Dr. J. Nathans (9). As a source of non-brain light chains, a cDNA library was constructed from the human lymphocyte cell line T 7527 by the method of Gubler and Hoffman (10). Size-selected double-stranded cDNA (0.6 to 4 kb) was ligated into the *EcoRI* site of the vector  $\lambda$ gt10 and propagated in *Escherichia coli* strain BNN 102 (11). Both libraries were screened with full-length bovine light chain clones, labeled by the primer extension method (12) to a specific activity of  $5 \times 10^8$  cpm/ $\mu$ g.

**Sequencing Strategy**—Positive clones were isolated and inserts subcloned in both orientations into M13mp18. Sequences were determined by the dideoxy chain termination method (13). Complete sequences on both strands were acquired by sequencing restriction enzyme subfragments from within the clones and by the use of internally priming oligonucleotides. In many cases oligonucleotide primers originally designed for sequencing the bovine light chains (3) worked satisfactorily for the human clones. Sequences were compiled

\* This work was supported by Grant CD 234 from the American Cancer Society. 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04174.

‡ Supported by a postdoctoral fellowship from the Muscular Dystrophy Association.

<sup>1</sup> The abbreviations used are: LC<sub>a</sub> and LC<sub>b</sub>, clathrin light chains a and b, respectively.

<sup>2</sup> L. Silveira and R. Schekman, personal communication.

on a VAX computer and analyzed using the programs of the University of Wisconsin Genetics Computer Group (14).

**Synthetic Peptides and Assay**—Peptides were made using F-moc chemistry on a Biosearch 9600 automatic synthesizer. Solid phase radioimmune assays were as described (15) with the following modifications. To reduce background 10% normal rabbit serum plus 4 mg/ml bovine serum albumin were included in the wash buffers and the solutions containing the first step monoclonal antibodies and the second step  $^{125}\text{I}$ -F(ab')<sub>2</sub> fragments of rabbit anti-mouse immunoglobulin. 0.02% sodium dodecyl sulfate and 0.1% Triton X-100 were also included in the wash buffers.

## RESULTS AND DISCUSSION

**Nucleotide Sequences of cDNA Encoding Human Clathrin Light Chains**—Human retinal (9) and lymphocyte cDNA libraries were screened with full-length cDNA probes encoding bovine brain and non-brain forms of the clathrin light chains: LC<sub>a</sub> and LC<sub>b</sub>. Preliminary analysis by Western blotting having shown that retina predominantly expressed the brain forms of the clathrin light chains and that Epstein-Barr virus transformed B lymphoblastoid cell lines expressed the non-brain forms. Positive clones were obtained from both libraries at a frequency of about 1 in  $5 \times 10^4$  for each probe. This permitted the isolation of full-length cDNA clones encoding the brain (retina) and non-brain (lymphocyte) forms of human LC<sub>a</sub> and LC<sub>b</sub>. The complete nucleotide sequences of these clones were determined and were compared with the sequences of cDNA encoding the corresponding light chains from bovine and rat (Fig. 1, *a* and *b*). As no previous comparison of the bovine (3) and rat (4) sequences has been made, we shall make a general assessment of the similarities between the light chains in all three species.

For both LC<sub>a</sub> and LC<sub>b</sub> there is 90–94% identity in the coding region when nucleotide sequences from different species are compared. The human sequences are slightly more similar to those of the cow than the rat (Table I). The relationships between the brain and non-brain forms of LC<sub>a</sub> and LC<sub>b</sub> are preserved in the three species. Inserted sequences of 90 nucleotides for LC<sub>a</sub> and 54 nucleotides for LC<sub>b</sub> are found in precisely the same positions in the brain forms (Fig. 1, *a* and *b*). Outside this region, the non-brain forms show no nucleotide substitutions as would be expected if the tissue-specific differences are due to differential mRNA splicing of single genes encoding LC<sub>a</sub> and LC<sub>b</sub> (3, 4).

It is striking that in the case of LC<sub>a</sub> there are no nucleotide substitutions in the brain-specific insert, whereas substitutions are found in the flanking regions. For LC<sub>b</sub> there are no differences in the brain-specific inserts of cow and human. However, one coding and three silent substitutions are found with rat LC<sub>b</sub> (Fig. 2*b*). In the case of LC<sub>a</sub> it is clear that there has been selection against both coding and non-coding substitutions in the brain-specific inserts. Some synonymous changes may not be neutral because the concentrations of synonymous tRNA molecules vary widely, and this may affect the rate of protein synthesis or even the way a protein folds. Alternatively, the sequence of the mRNA itself may have to fold into a critical secondary structure for correct tissue-specific splicing. Partial analysis of the bovine LC<sub>a</sub> gene has revealed that the brain-specific insert is composed of two exons; the first exon encodes residues 158–175 and the second encodes residues 176–187.<sup>3</sup> Thus, structural constraints on the LC<sub>a</sub> gene may be unusually severe in this part of the sequence because of several different splicing patterns that could potentially occur. This is consistent with the observation of Kirchhausen *et al.* (4) who have described a rat brain

LC<sub>a</sub> cDNA clone containing only the first 18 amino acids of the brain-specific insert.

In the 5'-untranslated regions differences were found between the brain and non-brain forms of human LC<sub>a</sub> and LC<sub>b</sub> (Fig. 1, *a* and *b*). However, we believe these reflect a cloning artifact rather than a difference in the genomic sequences encoding the brain and non-brain forms. The first 9 bases of the cDNA for human brain LC<sub>a</sub> and the first 46 bases of the cDNA for human brain LC<sub>b</sub> are inverted with respect to their non-brain counterparts. The second strand of the retinal cDNA library, which was used to isolate the brain forms, was constructed by the S1 nuclease digestion method, which is known to generate inverted sequences at the 5' end (16). In contrast, the lymphocyte cDNA library was constructed by the RNase H1 technique (10), which is not prone to this particular artifact. We therefore believe that the correct sequence is represented by the lymphocyte cDNA.

**cDNA Encoding LC<sub>a</sub> and LC<sub>b</sub> Have Distinct Regulatory Sequences in the Untranslated Regions**—The sequences of the 5'- and 3'-untranslated regions for LC<sub>a</sub> and LC<sub>b</sub> show little homology. This pattern is typical of that found in members of a multigene family (17). In addition, the untranslated regions have undergone a more rapid divergence between the three species than has the coding region (Table I). In particular, it was necessary to introduce gaps into each sequence in order to maximize the homology (Fig. 1, *a* and *b*). This shows that a number of insertions and deletions have occurred during the evolution of the light chain genes and suggests that these regions for the most part are under less selective pressure.

Exceptions are the sequences surrounding the initiation codons, which are well conserved between the three species. From comparison of sequences derived from various eukaryotic genes, Kozak (18) has obtained a consensus for this element. In this sequence of CC A/G CC ATGG the positioning of a purine at three bases upstream from the ATG start codon (underlined) is particularly important. For LC<sub>a</sub> there is an exact match to the consensus sequence. In contrast, the corresponding region of LC<sub>b</sub> deviates at 4 of the 6 positions, although the critical purine at position -3 is conserved. Although not unique to LC<sub>b</sub>, such extreme divergence from the consensus has only been encountered in about one-tenth of the genes examined (18).

A second and more dramatic example of an unusual regulatory sequence in LC<sub>b</sub> is found in the polyadenylation signal in the 3'-untranslated region. Again one finds that the LC<sub>a</sub> gene has an orthodox polyadenylation sequence of AATAAA (19) which starts 19–23 base pairs upstream from the poly(A) tail (Fig. 1*A*, and Ref. 23). For LC<sub>b</sub> a conventional polyadenylation signal does not exist. In human and rat LC<sub>b</sub> cDNA a related sequence TTAAA occurs 18–30 base pairs upstream of the poly(A) sequence (Fig. 1*b*). A homologous sequence is not found in the bovine LC<sub>b</sub> cDNA which does, however, contain the sequence AAAATT about 40 bases from the poly(A) tail. It is possible that these two distinct sequences act as polyadenylation signals, although they do not correspond to any of the known variants (20).

Extensive deviation from the consensus polyadenylation sequence can prevent correct RNA processing and decrease the levels of mature message (20). Thus the combination of a weak polyadenylation signal together with an unorthodox initiation signal may make expression of the LC<sub>b</sub> gene inefficient relative to LC<sub>a</sub>. This could account for the difficulty in detecting the biosynthesis of LC<sub>b</sub> but not LC<sub>a</sub> by incorporation of [<sup>35</sup>S]methionine (an amino acid more commonly found in LC<sub>b</sub> than LC<sub>a</sub>) in tissue culture cells (21).

<sup>3</sup> H. F. Seow, A. P. Jackson, and P. Parham, unpublished data.

**a**

```

HUMAN BRAIN TYPE LCa 1 GCCACCGCTGTGGTGTCCGGTGGGTCGGTGGTTTTTGTCTCACCGTTG 48
HUMAN LYMPHOCYTE LCa agoggtggatgcccggc-----
COW BRAIN TYPE LCa t tgcctctgatagcaggcgcaagcacagcgccgctgctggagt-----tt-ct-----g--c--g-*****
RAT BRAIN TYPE LCa tctccgctcgtcctccggtagcagccgccagcagcagtgctccgctgctgggc-g-----c-a-----c---*g-t--gg-t-c-

HUMAN BRAIN TYPE LCa GTGTCGGTCCGTTTCAGTTG*CCCGCCATGGCTGAGCTGGATCCGTTGGCGCCCTCGCGCCCTCGCGGTTCCCGCGCTGGGGAAC 138
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa aa--g-----cg-ac-t*-----t-a-----t-*****-g-----c-----
RAT LCa -----gc--gt--g--gg--t-----t-----a-t--t--c--t--t--g--c--t--c-----c-----

HUMAN BRAIN TYPE LCa GGAGTGGCCGGCCCGGGGAAGAAGACCCGGCTGCGCCCTTCTGGCGCAGCAAGAGAGCCAGATTGCGGGCATCGAGAACGACGAGGCC 228
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----*****-----g--t--t--c-----g-----a-----
RAT BRAIN TYPE LCa -----g--g-----g-----c-----g-----t-----a-----

HUMAN BRAIN TYPE LCa TTCGCCATCCTGGACGCGCGGCCCGCCCGGGCCCGAGCCGACGCGGAGCCCGCGGGGGTCCGGATGCTGTTGATGGAGTAATGAATGGT 318
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----at-----t-----at--a-----g-c-----a
RAT BRAIN TYPE LCa -----a-----g--g-a-----t-----g-----g-----c-----

HUMAN BRAIN TYPE LCa GAATACTACCAGGAAAGTAATGGTCCAACAGACAGTTATGCAGCTATTTACAAGTGGATCGATTGCAGTCAGAGCCTGAAAGTATCCGT 408
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----c--g-----g--g-----
RAT BRAIN TYPE LCa -----g--c-----c-----c-----g-----g-----

HUMAN BRAIN TYPE LCa AAATGGAGAGAAGAACAATAATGGAACGCTTGAAGCCCTTGATGCCAATTTCTCGGAAGCAAGAAGCAGAGTGGAAAGAAAAGGCAATAAAG 498
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----g--c-----g-----g-----a-----
RAT BRAIN TYPE LCa -----g--g--ca--g--c-----c-----g-----g-----g--g-----

HUMAN BRAIN TYPE LCa GAGCTAGAAGAATGGTATGCAAGACAGGACGAGCAGCTACAGAAAACAAAAGCAAACACAGGCTGGCAGATGAAGCTTTCTACAACAA 588
HUMAN LYMPHOCYTE LCa -----*****-----
COW BRAIN TYPE LCa -----g--c--g-----g--g--t-----g-----c-----
RAT BRAIN TYPE LCa -----g--g--c--g--t-----g-----c--g-----

HUMAN BRAIN TYPE LCa CCCTTCGCTGACGTGATTGGTTATGTCAACAACATAAACCATCCTTGCTACAGCCTAGAACAGGCAGCAGAAGAAGCCCTTGTAAATGAC 678
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----
RAT BRAIN TYPE LCa -----c-----

HUMAN BRAIN TYPE LCa ATTGACGAGTCGTCGCCAGGCACTGAGTGGGAACGGGTGGCCCGGCTGTGTGACTTTAACCCCAAGCTAGCAAGCAGGCCAAGATGTC 768
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----g-----c-----c-----a-----
RAT BRAIN TYPE LCa -----a-----g-----c-----c-----a-----

HUMAN BRAIN TYPE LCa TCCCGCATCGGCTCAGTCCTCATCTCCCTCAAGCAGGCCCGCTGGTGCAGTGAAGGCCACCCCTGGGAAACACTACATCTGCAATATC 858
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----t--t-----c-----g-----
RAT BRAIN TYPE LCa -----t-----g--g--c-----t-----g--c--g--c--*

HUMAN BRAIN TYPE LCa TTAATCCTACTCAGTGAAGCTCTTCACAGTCATTGGATTAATATGTTGAGTTCCTTTGGACCAACCTTTTGTCTTTAGAGTTGTCA 948
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----g--a-----*-----a-----
RAT BRAIN TYPE LCa *-----g--gca*-----cc-----*-----a-----

HUMAN BRAIN TYPE LCa TTGTTTGATGATGATGTTTCCTTCCTCAACTGTGTTCTCCCTGGCATTGAGAGAGAGGG*AGAGGA*GGAAGAGGAAGGGGAGGGAA 1038
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----****-----a-----g--gt--gc--c-----*-----g
RAT BRAIN TYPE LCa -----****-----a--a--*-----*-----*-----g

HUMAN BRAIN TYPE LCa GCTTCCCAAGAGTAGCCTCAACCTGTGCTTCTGTGCAATTATCTGAGATAATAATTCTGTTTCAAACTGT (A)n 1106
HUMAN LYMPHOCYTE LCa -----
COW BRAIN TYPE LCa -----c-----cg-c-----a-----t-----*--c--g(A)n
RAT BRAIN TYPE LCa -----*-----c-----t-----(A)n

```

**FIG. 1. Nucleotide sequences of human brain and lymphocyte LC<sub>a</sub> (a) and human brain and lymphocyte LC<sub>b</sub> (b) compared with the corresponding sequences from the cow (3) and rat (4). Dash, positions identical to human brain type LC<sub>a</sub> and LC<sub>b</sub>; asterisk, deletions relative to these sequences. The start codon is indicated by an arrow and the stop codon is overlined. Possible polyadenylation signals are hatched; the positions of the brain-specific inserts are boxed; and the positions of the inverted sequences at the extreme 5' ends of the brain LC<sub>a</sub> and LC<sub>b</sub> clones are highlighted in bold. a, the direct repeat sequences encoding the extra amino acids found in the human amino terminus of LC<sub>a</sub> compared with the cow are boxed. b, the inverted repeats present in the 3'-untranslated region are boxed.**

**b**

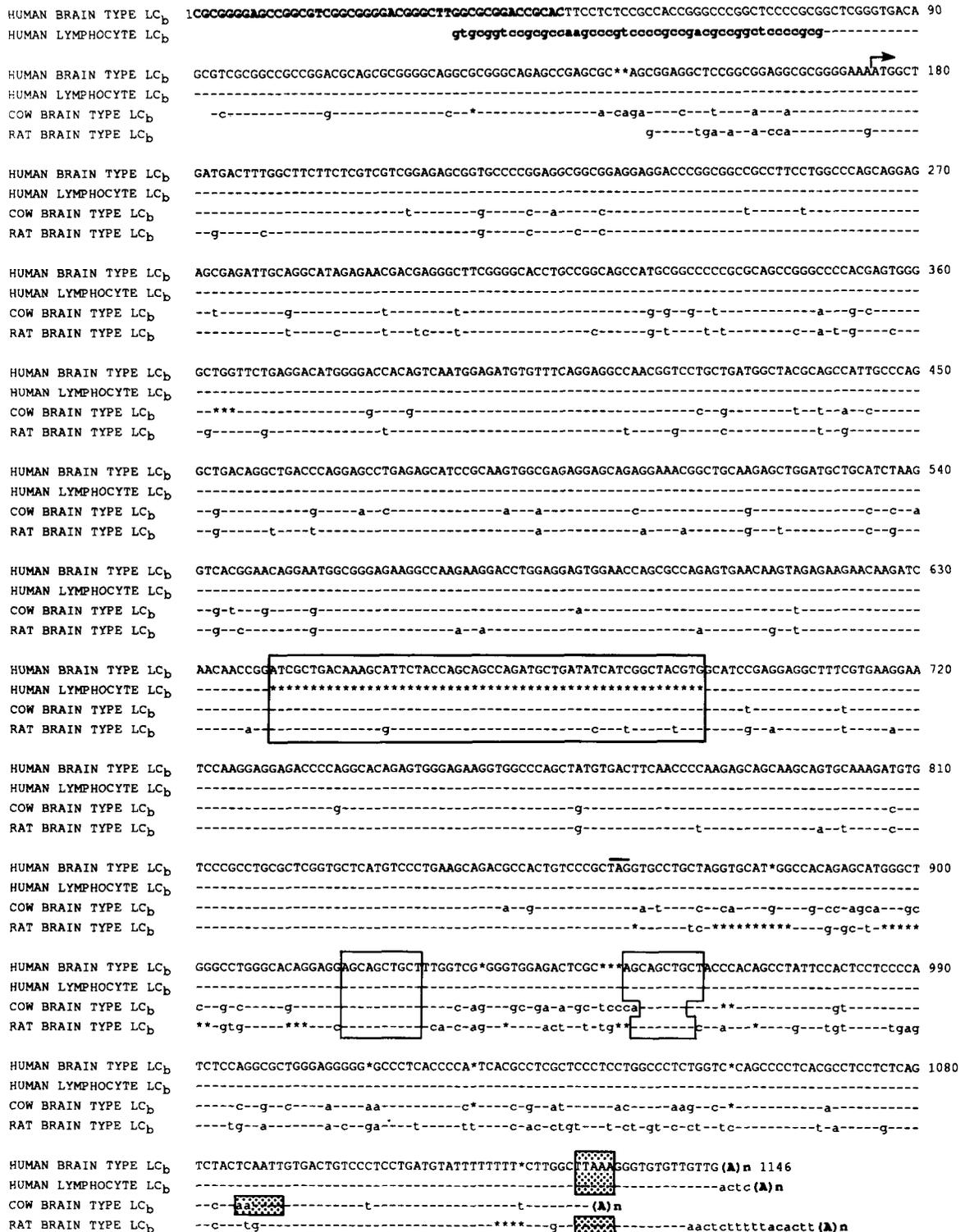


FIG. 1—continued

An interestingly conserved feature in the 3'-noncoding region of LC<sub>b</sub> is the occurrence of poly(T) sequences adjacent to a G T-rich area (Fig. 1b). Several eukaryotic genes contain downstream elements containing these structures, where they function together with the polyadenylation signal to define the mRNA cleavage site. However, these elements do not normally occur in mature mRNA as is found for LC<sub>b</sub> (19, 20, 22, 23). Another curious feature of the 3'-untranslated region

of LC<sub>b</sub> is the set of two inverted repeats that may be capable of forming a hairpin loop structure (Fig. 1b). Since similar sequences are found in all three species, it is possible that they may play a role in the transcriptional control of the LC<sub>b</sub> gene.

*Comparison of Structure of Clathrin Light Chains*—Similarities between the protein sequences of clathrin light chains from different species are high, giving a range of 95–98%

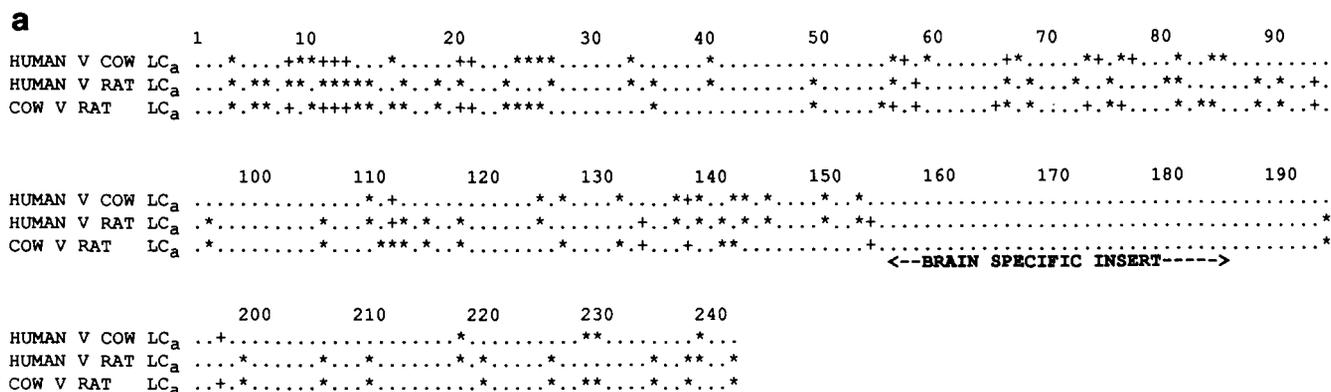
TABLE I  
Percent nucleotide homologies between light chains

	5'-Untranslated region	Coding region	3'-Untranslated region
Human LC <sub>a</sub> vs. bovine LC <sub>a</sub>	72.4	94.2	93.4
Human LC <sub>a</sub> vs. rat LC <sub>a</sub>	71.3	91.7	94.0
Bovine LC <sub>a</sub> vs. rat LC <sub>a</sub>	62.2	91.5	91.9
Human LC <sub>b</sub> vs. bovine LC <sub>b</sub>	87.3	92.2	76.0
Human LC <sub>b</sub> vs. rat LC <sub>b</sub>	64.3	90.9	74.5
Bovine LC <sub>b</sub> vs. rat LC <sub>b</sub>	67.9	90.2	72.1

identity between any pair of LC<sub>a</sub> or LC<sub>b</sub> sequences (Table II). The substitutions are generally conservative in nature. This contrasts with the range of 59–60% identity which is obtained when LC<sub>a</sub> and LC<sub>b</sub> sequences are compared. Thus, the differences between LC<sub>a</sub> and LC<sub>b</sub>, as well as their occurrence, have been conserved during the diversification of three orders of mammals. It is therefore likely that these differences reflect

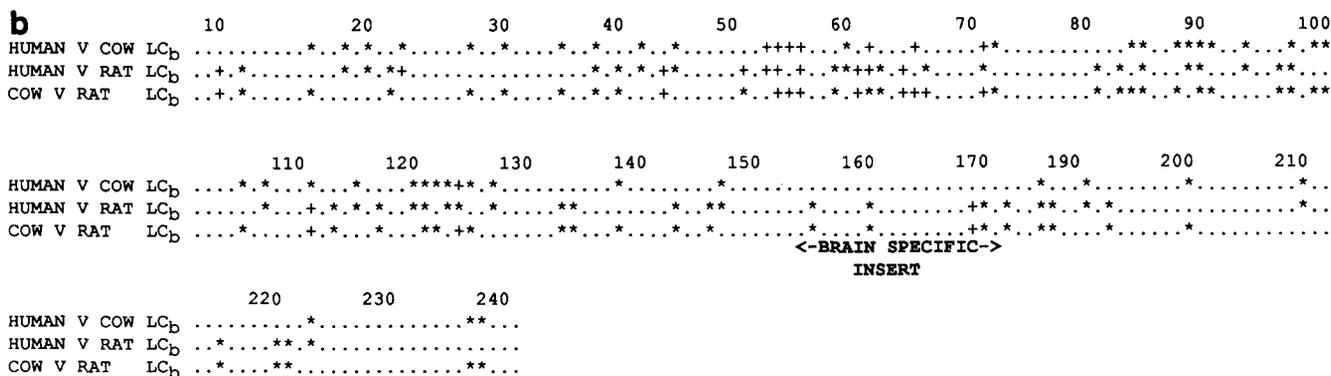
a functional diversification of the LC<sub>a</sub> and LC<sub>b</sub> light chains. A striking pattern of differences between LC<sub>a</sub> and LC<sub>b</sub> is seen in the first 24 residues of the amino-terminus (Fig. 3). This region of LC<sub>b</sub> is rich in serines and is the site for LC<sub>b</sub>-specific phosphorylation, a functional property that distinguishes the two classes of clathrin light chain (24). This region of LC<sub>a</sub> contains no serines and its length differs between species, an extra five amino acids being found in the human and rat compared with the cow. The extra bases found in the human and rat represent simple repeats of adjacent sequences, a pattern of species variation that has been noted in other genes (25). Efstratiadis *et al.* (17) have proposed a mechanism in which tandem repeats can be lost during DNA replication. The LC<sub>a</sub> sequences can be readily interpreted in terms of this model, which would imply that the two extra blocks of sequences were present in the common ancestor of rat, cow, and human but were subsequently lost along the lineage leading to the cow.

NUCLEOTIDE COMPARISON FOR LC<sub>a</sub>



\* silent codon change.  
+ coding change.

NUCLEOTIDE COMPARISON FOR LC<sub>b</sub>



\* silent codon change.  
+ coding change.

FIG. 2. Pairwise comparison of the nucleotide sequences for the human, cow, and rat LC<sub>a</sub> (a) and LC<sub>b</sub> (b) showing the distribution of silent (\*) and non-silent (+) codon changes.

TABLE II  
Percent homologies between light chain proteins

	Human LC <sub>a</sub>	Bovine LC <sub>a</sub>	Rat LC <sub>a</sub>	Human LC <sub>b</sub>	Bovine LC <sub>b</sub>	Rat LC <sub>b</sub>
Human LC <sub>a</sub>	100	96.7	98.0	60.1	60.1	59.2
Bovine LC <sub>a</sub>		100	95.5	59.6	60.0	59.2
Rat LC <sub>a</sub>			100	59.2	59.3	59.2
Human LC <sub>b</sub>				100	96.9	95.6
Bovine LC <sub>b</sub>					100	95.1
Rat LC <sub>b</sub>						100

The presence of these extra residues combined with the substitution of alanine at position 9 in human LC<sub>a</sub> permitted identification of a similarity between residues 7–16 of LC<sub>a</sub> and 47–59 of LC<sub>b</sub>. Dot matrix analysis confirmed the significance of this homology in the nucleotide sequences (Fig. 4a). Thus, 29 of 39 base pairs are shared by nucleotides 94–132 of LC<sub>a</sub> and nucleotides 306–344 of LC<sub>b</sub>. This is greater than the 20 of 39 base pairs that are shared by positions 306–344 of LC<sub>b</sub> and positions 220–258 of LC<sub>a</sub>, which are the homologous regions when alignment is continued from the more conserved 3' end of the coding region (Fig. 4b). This repeat motif in the LC<sub>a</sub> sequence is most likely the result of a duplication event during the evolution of the LC<sub>a</sub> gene. Ongoing analysis of the exon-intron organization of the LC<sub>a</sub> gene may produce evidence in support of this hypothesis.

At present the functional consequences of the duplicated

sequences can only be guessed. However, two features may be of significance; first, the duplicated sequences of LC<sub>a</sub> flank the most conserved region (residues 23–44) of the clathrin light chains. Second is the distinction that the single LC<sub>b</sub> copy of this sequence is on the carboxyl-terminal side of the conserved region, whereas the more closely related LC<sub>a</sub> sequence is on the amino-terminal side (Fig. 3). This shuffling of sequences in the primary structure may change the juxtaposition of functionally important sites in LC<sub>a</sub> and LC<sub>b</sub>.

**Heavy Chain Binding Region**—All clathrin light chains share the property of binding to the clathrin heavy chain (26). Association assays *in vitro* shows LC<sub>a</sub> and LC<sub>b</sub> are equivalent competitors for the heavy chain, indicating that they have similar heavy chain binding regions (27). A variety of experimental and theoretical analyses point to the central region, residues 95 to 157 of the light chains, as being involved in heavy chain binding (Fig. 3). Monoclonal antibodies against this region are against cryptic epitopes of clathrin triskelions and inhibit *in vitro* association (15). This structure is predicted to be  $\alpha$ -helical, and sequence homology is found with the  $\alpha$ -helical domains of several intermediate filament proteins (3). In addition Kirchhausen *et al.* (4) have emphasized a set of heptad repeats found within this same region and suggest that the light chains form a coiled coil around the proximal arm of the heavy chain. In this region human LC<sub>a</sub> differs by 2 residues from the cow and 3 residues from the rat, whereas human LC<sub>b</sub> differs by only a single residue from the cow. These small numbers of conservative substitutions do

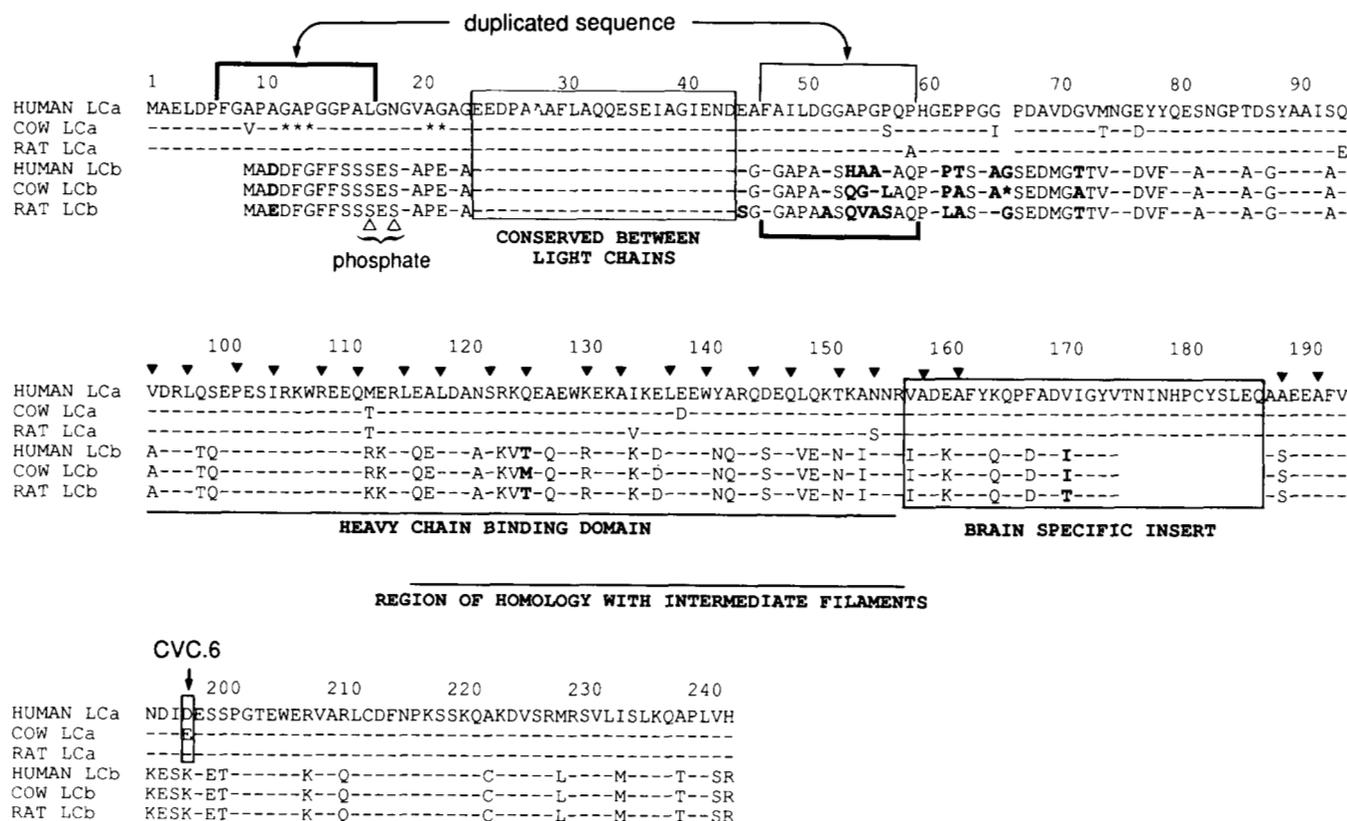


FIG. 3. Comparison of the amino acid sequences of brain LC<sub>a</sub> and LC<sub>b</sub> from human, cow, and rat. Numbers refer to those of the original bovine LC<sub>a</sub> sequence (3). Amino acid identities are shown by a dash; differences are as indicated; the differences in LC<sub>b</sub> between species are highlighted in bold. Asterisks, deletions. The positions of the heptad repeats (4) are denoted with filled triangles; the positions of the LC<sub>b</sub> phosphorylation sites (24) are shown with open triangles; and the location of the CVC-6 epitope is boxed. The duplicated sequence in LC<sub>a</sub> and its homologous region in LC<sub>b</sub> are indicated by square brackets. The region of the heavy chain binding site as mapped with monoclonal antibodies (15) and the region of maximum homology to the intermediate filaments (3) are also indicated.

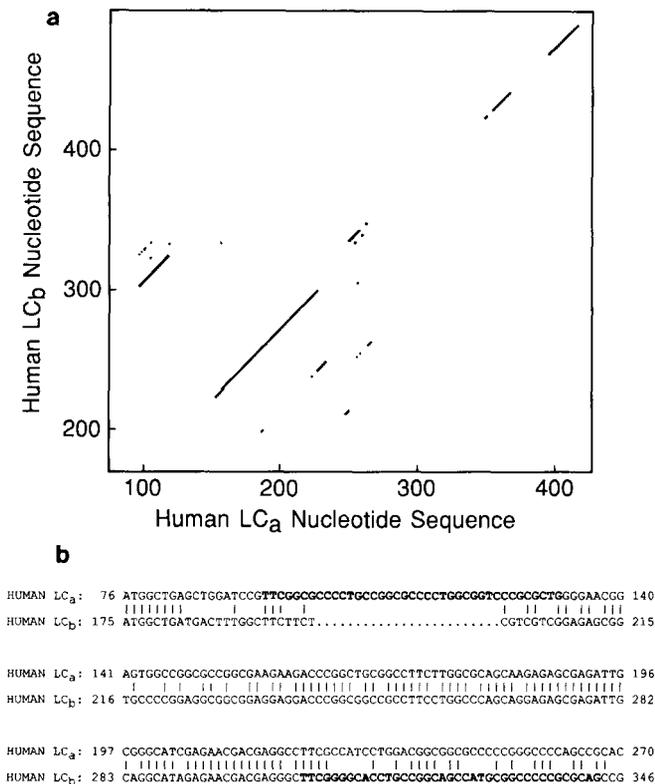


FIG. 4. *a*, dot matrix (14) comparison of human non-brain LC<sub>a</sub> and LC<sub>b</sub>. The sequences were compared processively over a window of 21 nucleotides, with matches of 14 or greater being recorded. *b*, nucleotide comparison between the amino-terminal region of LC<sub>a</sub> and LC<sub>b</sub>. The sequences, starting at the initiation methionine, are aligned for maximal homology with no gap penalties. The homology between residues 94–132 of LC<sub>a</sub> and residues 306–344 of LC<sub>b</sub> is shown in **bold**.

not affect the theoretical predictions. In contrast there are 22 differences between LC<sub>a</sub> and LC<sub>b</sub> in the 63 residues of this region, including 13 non-conservative substitutions (Fig. 3). Apart from three conservative changes in the heptad repeats (at positions 95, 126, and 131), most of the differences between light chains cluster on the opposite face of the helix from that predicted to interact with the clathrin light chain (Fig. 5). Thus, LC<sub>a</sub> and LC<sub>b</sub> may bind to the same site on the heavy chain via the conserved heptad repeats and yet present a markedly different structure along the side of the helix that is potentially exposed to the cytosol.

The precise boundaries of the heavy chain binding site are uncertain. However, at the amino-terminal side, residues 45–94 contain 7 out of 17 prolines in LC<sub>a</sub> and 5 out of the 13 prolines in LC<sub>b</sub> (Fig. 3). This region has a low probability of forming an  $\alpha$ -helix. Moreover, this same region contains 6 out of the 12 species differences for LC<sub>a</sub> and 11 of the 14 species differences for LC<sub>b</sub>. This implies that residues 45–94 are under less functional constraint than other areas of the molecule and are therefore less likely to contribute to heavy chain binding. To define the carboxyl-terminal limits of the heavy chain binding site, the identification of residues contributing to the epitope recognized by monoclonal antibody CVC.6 was undertaken.

**Glutamic Acid 198 Is Critical for the CVC.6 Epitope**—We have previously shown that CVC.6, an LC<sub>a</sub>-specific monoclonal antibody, binds to a tryptic peptide obtained from bovine LC<sub>a</sub> consisting of residues 188–208 (15). Within this sequence human, bovine, and rat LC<sub>a</sub> only differ at position 198, which is glutamic acid in the cow and aspartic acid in human and rat (Fig. 3). As CVC.6 is highly specific for bovine

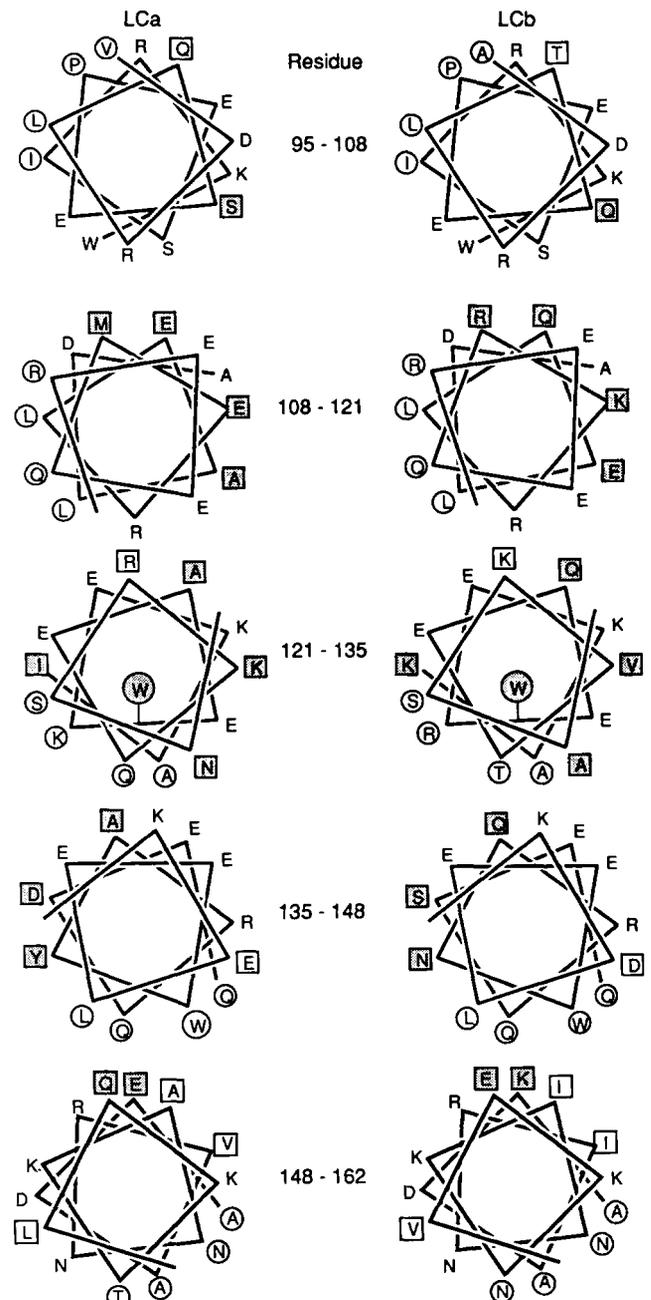


FIG. 5. Helix wheel projection of the heavy chain binding site of LC<sub>a</sub> and LC<sub>b</sub>. Amino acids of the heptad repeats are circled; amino acid differences in the rest of the sequence between LC<sub>a</sub> and LC<sub>b</sub> are within squares, with nonconservative changes shaded. To preserve the heptads, a "skip" residue at tryptophan 130 has been proposed (4). This is indicated by the shaded circle in projection.

LC<sub>a</sub>, showing no detectable reaction with human LC<sub>a</sub>, this suggests that a single methylene group in the residue at position 198 is absolutely critical for the CVC.6 epitope.

A synthetic peptide corresponding to residues 188–208 of bovine LC<sub>a</sub> was found to bind strongly to the CVC.6 antibody in a solid phase radioimmunoassay. This showed that the epitope was purely the result of primary sequence and not to a post-translational modification present in the previously studied "natural" peptide. The homologous human LC<sub>a</sub> peptide, differing by just the substitution of aspartic acid for glutamic acid at position 198, showed no binding over background levels as determined with an irrelevant peptide. To provide a more sensitive and quantitative comparison of the

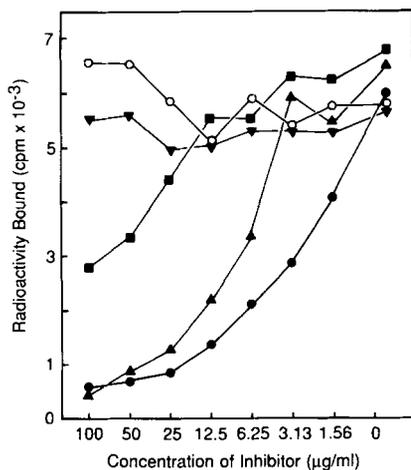


FIG. 6. Inhibition of CVC.6 binding by synthetic peptides. The binding of CVC.6 to a solid phase peptide corresponding to residues 188–208 of bovine LC<sub>a</sub> in the presence of fluid phase peptides was measured. Inhibitory peptides were bovine LC<sub>a</sub> residues 188–208 (●), bovine LC<sub>a</sub> residues 192–208 (▲), bovine LC<sub>a</sub> residues 193–208 (■), bovine LC<sub>a</sub> residues 195–208 (▼), and human LC<sub>a</sub> residues 188–208 (○).

peptide-antibody interaction an assay in which the capacity of soluble peptides to inhibit the binding of CVC.6 to a solid phase was measured. As shown in Fig. 6 a strong and titratable inhibition was seen with the bovine peptide. In contrast the human peptide gave no inhibition even at a concentration 50-fold greater than was required to see an effect with the bovine peptide.

To define further the CVC.6 epitope a nested set of peptides derived from the bovine LC<sub>a</sub> sequence was made. These had a common-carboxyl terminus at residue 208 and stepwise amino-termini from residues 188 to 199. Strong inhibition was seen with peptides having amino-terminal residues 188–192, significant but weak inhibition was seen with peptides having amino-terminal residues 193–194, and no inhibition was seen with the shorter peptides (Fig. 6).

These results prove that the glutamic acid residue at position 198 of bovine LC<sub>a</sub> is absolutely critical for the CVC.6 epitope which is formed by residues 192–208 of bovine LC<sub>a</sub>. Although flanking residues are clearly important in forming the epitope, they confer no significant interaction with CVC.6 in the absence of glutamic acid at position 198.

CVC.6 binds equally well to free light chains and to light chains that are associated with heavy chains in triskelions and coated vesicles (15). This indicates that interaction of this antibody with its epitope is not perturbed by heavy chain binding. As the linear CVC.6 epitope minimally consists of residues 192–198, it is highly unlikely that this sequence is involved in heavy chain binding. These results strongly suggest that the heavy chain binding site does not extend beyond residue 192, consistent with the organization of the heptad repeats which end at residue 193 (Fig. 3).

The region from the end of the heptad repeats to the

carboxyl terminus is almost perfectly conserved between species. However, there are 14 substitutions out of 51 residues between LC<sub>a</sub> and LC<sub>b</sub>, including distinct patterns of cysteine residues which *in vitro* can readily form disulfide bonds (28). This comparison suggests that this carboxyl-terminal region may, in addition to the amino-terminal region, provide a site of functional diversification for clathrin light chains LC<sub>a</sub> and LC<sub>b</sub>.

*Acknowledgments*—We thank C. Van Wandelen for technical assistance and B. Bruechert for preparation of the manuscript.

#### REFERENCES

- Brown, M. S., Anderson, R. G. W., and Goldstein, J. L. (1983) *Cell* **32**, 663–667
- Pearse, B. M. F., and Bretscher, M. S. (1981) *Annu. Rev. Biochem.* **50**, 85–101
- Jackson, A. P., Seow, H.-F., Holmes, N., Drickamer, K., and Parham, P. (1987) *Nature* **326**, 154–159
- Kirchhausen, T., Scarmato, P., Harrison, S. C., Monroe, J. J., Chow, E. P., Mattaliano, R. J., Ramachandran, K. L., Smart, J. E., Ahn, A. H., and Brosius, J. (1987) *Science* **236**, 320–324
- Bar-Zvi, D., and Branton, D. (1986) *J. Biol. Chem.* **261**, 9614–9621
- Cantournet, B., Creuzet, C., Komano, O., and Loeb, J. (1987) *FEBS Lett.* **220**, 143–148
- Payne, G. S., and Schekman, R. (1985) *Science* **230**, 1009–1014
- Lemmon, S. K., and Jones, E. W. (1987) *Science* **238**, 504–509
- Nathans, J., Thomas, D., and Hogness, D. S. (1986) *Science* **232**, 193–202
- Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25**, 263–269
- Watson, C. J., and Jackson, J. F. (1985) in *DNA Cloning: A Practical Approach* (Glover, D. M. ed) Vol. 1, pp. 79–88, IRL Press, Oxford
- Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
- Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3963–3965
- Devereux, J., Haerberli, P., and Smithies, O. (1984) *Nucleic Acid. Res.* **12**, 387–395
- Brodsky, F. M., Galloway, C. J., Blank, G. S., Jackson, A. P., Seow, H.-F., Drickamer, K., and Parham, P. (1987) *Nature* **326**, 203–205
- Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E., and Goodman, H. M. (1979) *Nucleic Acids. Res.* **7**, 1137–1146
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., Deriel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulters, C. C., and Proudfoot, N. J. (1980) *Cell* **21**, 653–668
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872
- Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263**, 211–214
- Birnsteil, M. L., Busslinger, M., and Strub, K. (1985) *Cell* **41**, 349–359
- Brodsky, F. M. (1985) *J. Cell Biol.* **101**, 2055–2062
- McLauchlan, J., Gaffney, D., Whitton, J. L., and Clements, J. B. (1985) *Nucleic Acids Res.* **13**, 1347–1368
- Platt, T. (1986) *Annu. Rev. Biochem.* **55**, 339–372
- Hill, B. L., Drickamer, K., Brodsky, F. M., and Parham, P. (1988) *J. Biol. Chem.* **263**, 5499–5501
- deJong, W. W., and Ryden, L. (1981) *Nature* **290**, 157–159
- Ungewickell, E. (1983) *EMBO J.* **2**, 1401–1408
- Winkler, F. K., and Stanley, K. K. (1983) *EMBO J.* **2**, 1393–1400
- Parham, P., Brodsky, F. M., and Drickamer, K. (1988) *Biochem. J.*, in press.