Volume 311, number 3, 295–298 FEBS 11683 October 1992

cDNA encoding the chicken ortholog of the mouse *dilute* gene product

Sequence comparison reveals a myosin I subfamily with conserved C-terminal domains

Gabriele Sanders, Beate Lichte*, Helmut E. Meyer and Manfred W. Kilimann

Institut für Physiologische Chemie, Ruhr-Universität Bochum, Postfach 102148, W-4630 Bochum 1, Germany

Received 20 July 1992; revised version received 11 September 1992

We report the cDNA-deduced primary structure of the chicken counterpart of the murine *dilute* gene product, a member of the myosin I family. Comparison of the chicken and mouse sequences reveals a distinct pattern of domains of high and low sequence conservation. An internal deletion of 25 amino acids probably reflects differential mRNA processing. Compared with other myosin heavy chain molecules, sequence similarity is highest with the *MYO2* gene product of *Saccharomyces cerevisiae*. The *MYO2* protein, implicated in vectorial vesicle transport, is homologous to the *dilute* protein over practically its entire length. In addition, the C-terminal domain of the *dilute* protein is highly similar to a putative glutamic acid decarboxylase sequence cloned from mouse brain. Alternatively, this closely related clone might represent an isoform of the *dilute* protein derived from a second gene, potentially involved in genetic conditions related to dilute.

Dilute; Myosin heavy chain; Glutamic acid decarboxylase; Cytoskeleton; Vesicle transport; Synapse

1. INTRODUCTION

The formation and functioning of synaptic contacts between neurons involve a variety of motility events. For example, the migration and navigation of growth cones and the establishment of synapses, or the tightly controlled exocytosis and recycling of neurotransmitter vesicles demand efficient and highly organized mechanisms of intracellular transport and membrane trafficking. Identification of the underlying molecular machinery is therefore essential for an understanding of the ontogeny and functioning of the nervous system.

To identify new protein components of the neuronal synapse, we have screened brain cDNA libraries with antisera raised against synaptic plasma membranes [1]. One of the clones isolated in these experiments has turned out to be the chicken counterpart of the recently characterized protein that is defective in the mouse *dilute* mutations [2]. The *dilute* protein is a member of the myosin I family, a group of proteins presumed to drive the movement of membranes along actin filaments [3]. The *dilute* mutation impairs the formation of dendritic processes of melanocytes, and most mutated alleles (*dilute-lethal*) also produce a neurological defect, suggesting a role for this protein in the formation or functioning of cellular processes [2].

Comparison between the chicken and mouse sequences defines a pattern of domains with high or low sequence conservation. Sequence comparison of the *dilute* protein with other proteins identifies two other gene products with which it shares sequence similarity in its C-terminal domain, with implications regarding its possible function and the molecular nature of related mutations.

Present address: Institut für Biologie III, Universität Freiburg, Schänzlestr. 1, W-7800 Freiburg, Germany.

Correspondence address: M.W. Kilimann, Institut für Physiologische Chemie, Ruhr-Universität Bochum, Postfach 102148, W-4630 Bochum 1, Germany. Fax: (49) (234) 7094-193.

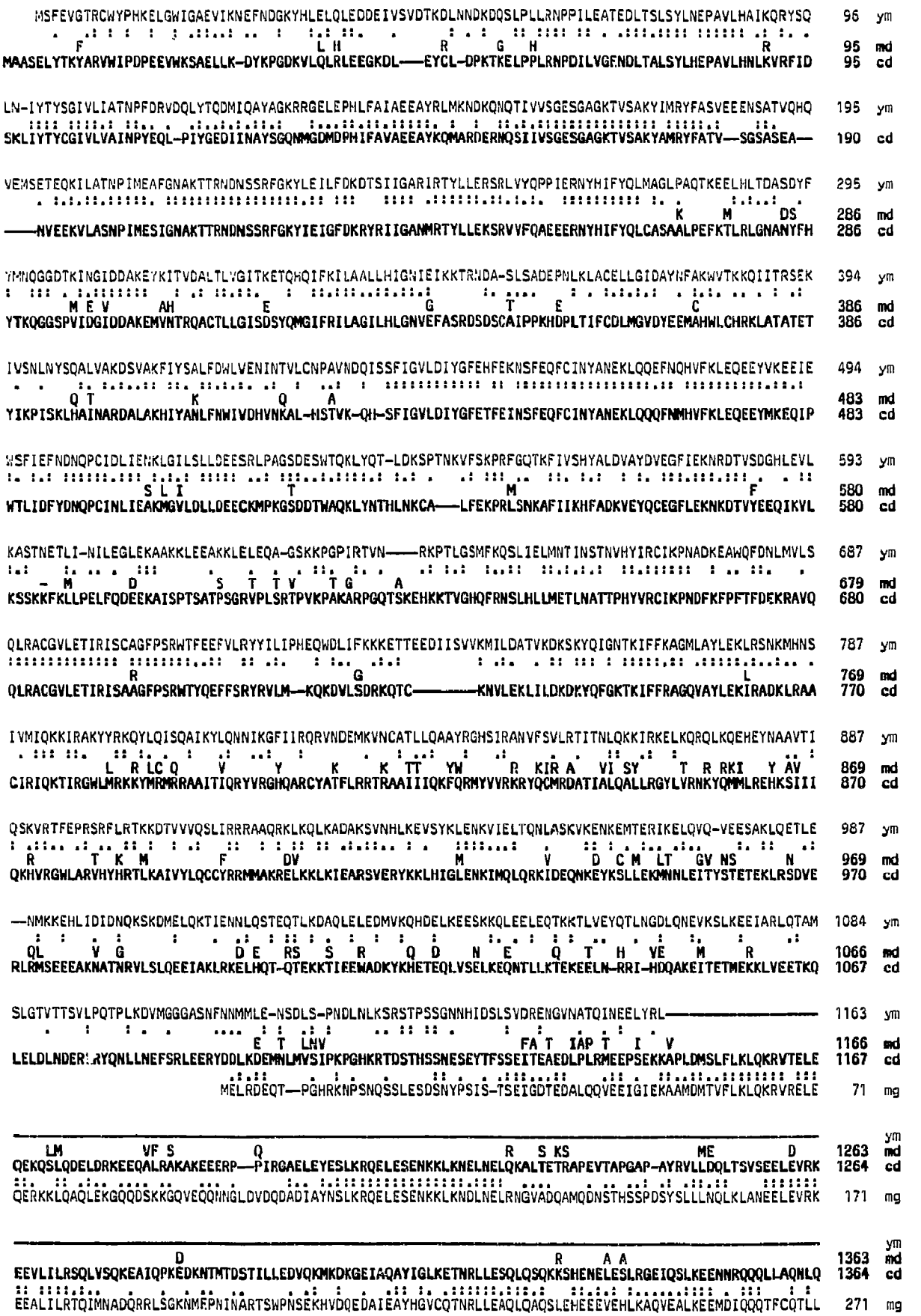
2. EXPERIMENTAL

A partial chicken *dilute* cDNA, dilute-8.6, was isolated from a brain cDNA library by immunoscreening as described [1]. Its sequence was extended by two rounds of rescreening with 5′-terminal cDNA fragments as hybridization probes. Sequencing was carried out in both directions by the chain termination method. Northern blotting was performed as described [1].

3. RESULTS AND DISCUSSION

High amino acid sequence similarity (90% overall identity) and practically identical length identify the protein encoded by our cDNA with high likelyhood as the chicken ortholog of the mouse *dilute* gene product (Fig. 1), which is a member of the myosin heavy chain (type I) family. Northern blot analysis detects an mRNA of ~7.5 kb at high abundance in chicken forebrain, but not in liver and muscle (Fig. 2).

Phylogenetic sequence conservation is distributed



very unevenly, defining three main regions that correlate remarkably well with the domain pattern previously deduced from comparison to other myosin heavy chains [2]. The N-terminal ~780 amino acids display 94% sequence identity; this region coincides with the actin-binding head domain common to all myosin heavy chains. In the central region of ~270 amino acids, sequence similarity is much lower (only 76% identity). This region harbours the five tandem repeats that were proposed to be potential calmodulin binding sites (amino acids 767–888 of the chicken sequence), followed by a stretch of heptad repeats proposed to mediate dimerization through formation of a coiled coil [2]. At the beginning of the C-terminal domain, sequence conservation gradually increases again, and, downstream from amino acid 1,260, reaches an exceptionally high 97.5% identity (with 12 of the 14 replacements among 567 amino acids being conservative). This indicates particularly stringent sequence requirements for the (as yet unknown) function of this domain.

Compared to the mouse sequence, the cDNA from chicken has an internal deletion of 75 nucleotides, resulting in the omission of 25 amino acids downstream from amino acid 1,387 of the predicted protein. This deletion probably reflects differential RNA processing, and, apart from it, the chicken and mouse molecules differ in length by only one amino acid. The sequence gap is also found in the homologous clone from mouse that is discussed below.

Comparison with other sequences in the database identified two molecules with which the *dilute* protein is most closely related. Among all myosin I and II sequences available, the *dilute* protein is most closely related to the *MYO2* protein of yeast, which is involved in intracellular vectorial membrane vesicle transport required for yeast budding [4]. Already in the head and

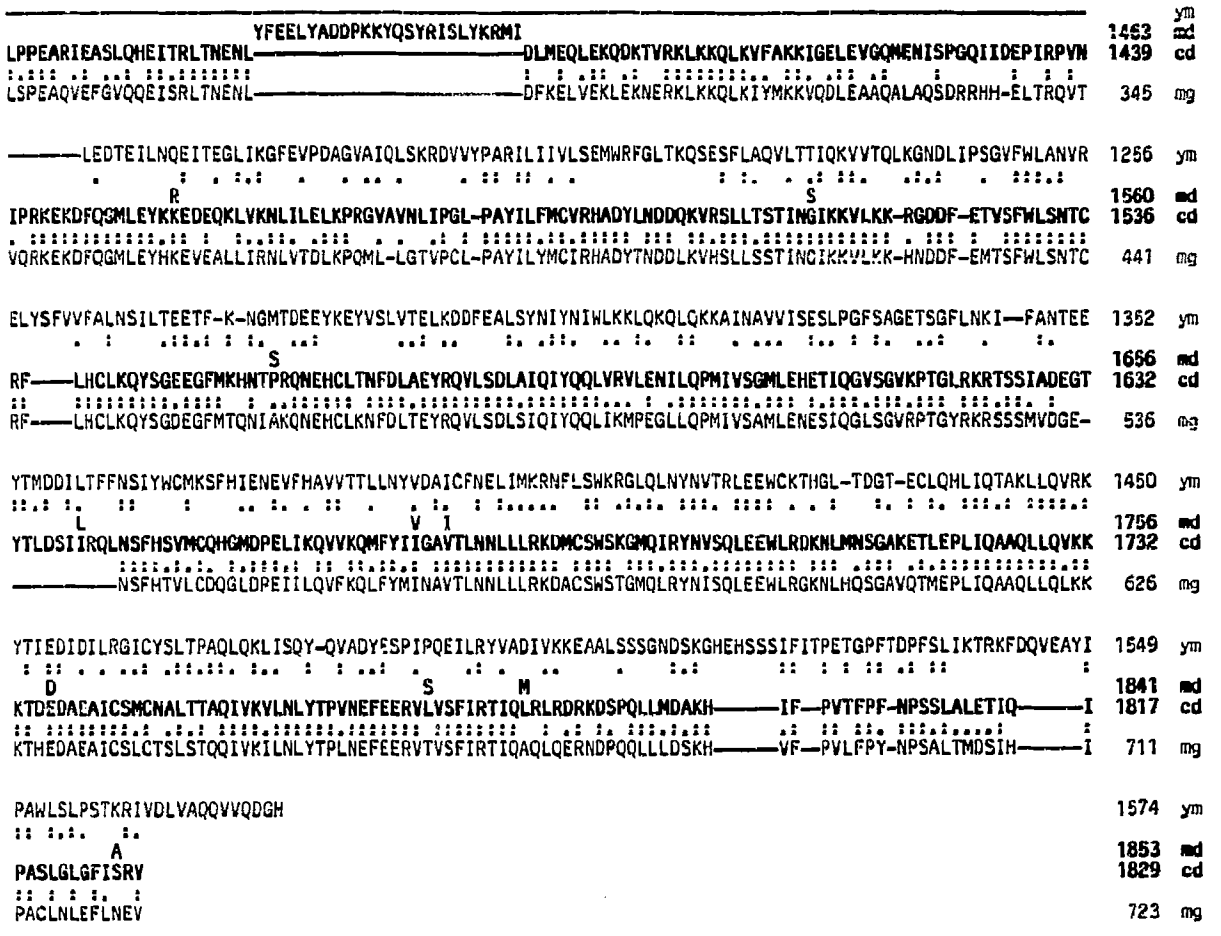


Fig. 1. Alignment of the deduced amino acid sequences of the chicken (cd) and mouse (md) *dilute* proteins, the yeast *MYO2* protein (ym) and the putative GAD from mouse brain (mg). Colons indicate amino acid identities and dots indicate conservative replacements, in comparison with the chicken *dilute* sequence. Gaps are indicated by dashes. The *dilute* sequences are given in bold print, and amino acids of the mouse *dilute* sequence are only indicated where they differ from the chicken *dilute* sequence. The chicken *dilute* cDNA sequence was compiled from three clones, dilute-8.6 (3′ end to codon 1,032), dilute-8.6–74 (upstream to codon 177) and dilute-8.6–3 (upstream to the 5′ end). The whole nucleotide sequence comprises 99 nucleotides (nt) of 5′ untranslated sequence, a reading frame of 5,487 nt (encoding a polypeptide of M_r 212,381), and a 3′ untranslated sequence of 1,339 nt, lacking a poly(A) tail. It has been deposited in the EMBL database under accession number X67251.

←

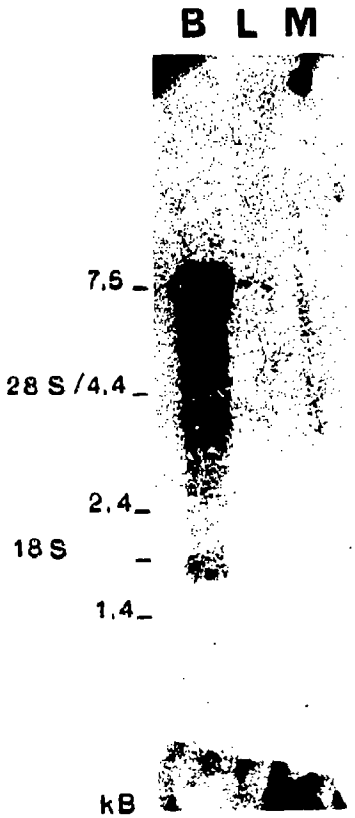


Fig. 2. RNA blot analysis. 10 μg of poly(A)$^+$ RNA from chicken forebrain (B), liver (L) and muscle (M) were resolved on a 1% agarose gel, transferred to a nylon membrane and hybridized with the cDNA of clone dilute-8.6 (which represents the C-terminal coding sequence and the 3′ untranslated sequence; cf. legend to Fig. 1) labelled by nick-translation. The positions of DNA (in kb) and rRNA molecular size standards are given at the left margin.

central domains (amino acids 1–1,040, including the putative calmodulin-binding and coiled-coil domains that both molecules also seem to have in common), similarity to the *MYO2* protein (42% identity) is substantially higher than to any other myosin heavy chains (not shown). Moreover, there is sequence conservation in the C-terminal domains. A stretch between amino acids 1,686 and 1,767 of the *dilute* sequence has 44% identity with the *MYO2* sequence, whereas the rest of the *MYO2* sequence can be aligned with lower similarity (21% identity; Fig. 1).

However, the sequence most closely related to the *dilute* protein that was found in the database aligns to almost the complete C-terminal domain with 57% identity. This is a cDNA from mouse brain that was described as an isoform of glutamic acid decarboxylase (GAD), based on immunological evidence and GAD activity of bacterial extracts expressing its β-galactosidase fusion protein [5]. Thus, the *dilute* protein might be a chimeric molecule with an enzyme moiety fused to a myosin head domain, analogous to the *Drosophila ninaC* gene product [6]. In this case, it might be functionally linked to the biosynthesis of the neurotransmitter, GABA, which plays roles as a signal molecule and a trophic factor inside and outside of the nervous system [7]. However, this putative GAD sequence is completely unrelated to all the other GADs that have been cloned, and also unrelated to other pyridoxal phosphate-dependent decarboxyases (dopa decarboxylase, histidine decarboxylase, tryptophan decarboxylase) which share sequence similarity among each other and with GAD [8–10]. In particular, the clone of Huang et al. [5], and likewise the C-terminal *dilute* domain, lack the tetrapeptide consensus sequence, NPHK, of the pyridoxal phosphate binding site.

Alternatively, whatever the functional properties of the C-terminal domain may be, the clone isolated by Huang et al. [5] could be an incomplete cDNA of an isoform of the *dilute* protein derived from a second gene. It will be of great interest to see whether its sequence can be extented further upstream to encode a full-length *dilute* homolog. The existence of multiple genes encoding proteins similar to *dilute* would offer explanations for other genetic conditions related to *dilute* which have a similar phenotype *(ashen, leaden)* or suppress it (*dilute suppressor*) [2,11,12].

Acknowledgements: This work was supported by the Fonds der Chemischen Industrie. B.L. was supported through a predoctoral fellowship from the Evangelisches Studienwerk Villigst. M.W.K. is a Heisenberg Fellow of the Deutsche Forschungsgemeinschaft.

REFERENCES

[1] Lichte, B., Veh, R.W., Meyer, H.E. and Kilimann, M.W. (1992) EMBO J. 11, 2521–2530.
[2] Mercer, J.A., Seperack, P.K., Strobel, M.C., Copeland, N.G. and Jenkins, N.A. (1991) Nature 349, 709–713.
[3] Pollard, T.D., Doberstein, S.K. and Zot, H.G. (1991) Annu. Rev. Physiol. 53, 653–681.
[4] Johnston, G.C., Prendergast, J.A. and Singer, R.A. (1991) J. Cell Biol. 113, 539–551.
[5] Huang, W.-M., Reed-Fourquet, L., Wu, E. and Wu, J.-Y. (1990) Proc. Natl. Acad. Sci. USA 87, 8491–8495.
[6] Montell, C. and Rubin, G. (1988) Cell 52, 757–772.
[7] Erdö, L.F. and Wolff, J.R. (1990) J. Neurochem. 54, 363–372.
[8] Bu, D.F., Erlander, M.G., Hitz, B.C., Tillakaratne, N.J.K., Kaufman, D.L., Wagner-McPherson, C.B., Evans, G.A. and Tobin, A.J. (1992) Proc. Natl. Acad. Sci. USA 89, 2115–2119.
[9] De Luca, V., Marineau, C. and Brisson, N. (1989) Proc. Natl. Acad. Sci. USA 86, 2582–2586.
[10] Vaaler, G.L., Brasch, M.A. and Snell, E.E. (1986) J. Biol. Chem. 261, 11010–11014.
[11] Moore, K.J., Swing, D.A., Rinchik, E.M., Mucenski, M.L., Buchberg, A.M., Copeland, N.G. and Jenkins, N.A. (1988) Genetics 119, 933–941.
[12] Moore, K.J., Seperack, P.K., Strobel, M.C., Swing, D.A., Copeland, N.G. and Jenkins, N.A. (1988) Proc. Natl. Acad. Sci. USA 85, 8131–8135.