Vol. 162, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 14, 1989 Pages 294-299

IDENTIFICATION OF PREVIOUSLY UNRECOGNIZED SEQUENCE MOTIFS AT THE EXTREME CARBOXYTERMINUS OF THE NEUROFILAMENT SUBUNIT NF-M

Gerry Shaw

Department of Neuroscience, JHM Health Center, Box J-244, Gainesville, FL 32610

Received May 26, 1989

Two previously unrecognized features of neurofilament architecture are revealed by careful analysis of published neurofilament sequences. 1. The extreme C-terminus of the NF-M tail contains two highly conserved homologous sequences each of 15 amino acids, with the consensus EEK-V-TKKVEK-TS, plus another very closely related 7 amino acid sequence. 2. The C-terminus of NF-M contains sequences of consensus K-SP or K--SP which in some species are multiply repeated, are probably phosphorylated, but are distinct from the more obvious KSP repeated sequences. Sequences related to both the K-SP and K--SP sequences are found in NF-H, microtubule associated proteins tau and MAP2, suggesting a further level of immunological and potential evolutionary relationship between neurofilaments and these microtubule associated proteins. The possible significance of these findings is discussed.

Neurofilaments are composed predominantly of three subunits, NF-L, NF-M and NF-H, usually referred to as the neurofilament triplet proteins (1,2). The most unusual features of these proteins are the extensive C-terminal "tail" sequences of NF-M and NF-H which are absent from other known 10nm filament subunits. Several electron microscopical studies have provided evidence that these tail sequences actually protrude as fine rodlets from the side of the neurofilament, the body of which contains the highly conserved α-helical "rod" domains of the triplet proteins (see 3 for example). Mammalian NF-M and NF-H tail sequences are built on the same general pattern, though appear to be only distantly related. At the N-terminal end of each is a short proline containing sequence called tail A (2), followed by a glutamic acid rich region, followed by a region containing repeated sequences based upon the tripeptide lysine-serine-proline (KSP), into which is inserted a further glutamic acid rich region in NF-M (4-11). At the extreme C-termini, after the last of the KSP repeats, are long regions rich in both lysine and glutamic acid. In mammalian NF-H the lysine and glutamic acid rich region is normally about 185 amino acids long, and,

0006-291X/89 $1.50

since it contains many proline residues, is here referred to as the KEP region. The analogous region of mammalian NF-M is about 220 amino acids long, has very few prolines, and is therefore here called the KE segment. The large KE and KEP segments, being apparently situated on the ends of extensions protruding from the side of the neurofilament seem to be ideally positioned to mediate interactions between neurofilaments and other cellular components. Understanding such interactions may give us insights into the still mysterious mechanisms of neurofilament transport and cross-linking, as well as the numerous types of pathological responses involving aberrant neurofilament expression. Accordingly, a careful analysis of newly published amino acid sequences of these two regions has been performed.

MATERIALS, METHODS AND RESULTS

Functionally important sequences are usually well conserved across species boundaries. Comparison of the known KE and KEP sequences shows that the KE segment is much more conserved (4-12). For instance, alignment of human and rat KEP segments can only be performed by introducing no fewer than 12 insertions (4-6). Out of 171 alignable amino acids, 34 (19.9%) are non-identical, and 11 (6.4%) of these differences are non-conservative. In contrast, comparison of the 4 known KE segments show a high degree of conservation of sequence, particularly at the extreme C-terminus (9-12, see figure 1). The mammalian sequences can be aligned perfectly and only 11 (9.1%) positions show amino acid heterogeneity. All of these variations are conservative. Alignment of the chicken sequence with the mammalian requires six breaks, but still reveals remarkable sequence conservation, particularly at the C-terminus. For example, the selected sequence between the two arrows in figure 1, consisting of 57 amino acids, gives perfect alignment for all species, and sequence divergence at only 6 positions (10.5%), only 1 (1.7%) varying between the mammalian sequences. The sequence variations are all highly or semi-conservative. This degree of conservation is much greater than that of any selected region of the KEP segment, and is particularly significant since currently known KEP sequences only cover mammalian species. For comparison, a segment of 68 amino acids corresponding to the C-terminal region of coil II in the α-helical rod, the most conserved 10nm filament domain, is known in all four species. These sequences align perfectly and only 5 (7.3%) of the amino acid positions show any variation. Only 1 (1.4%) varies within mammals. The degree of conservation of the selected region of the KE domain indicated in figure 1 therefore approaches that of the most conserved regions of the NF-M molecule.

These findings suggest that the extreme C-terminal tail of NF-M has an important function, whilst the NF-H tail is less important to the neuron. A visual search for repeated amino acid sequences within the most conserved part of the KE segment revealed two closely related 15 amino acids sequences which are outlined in

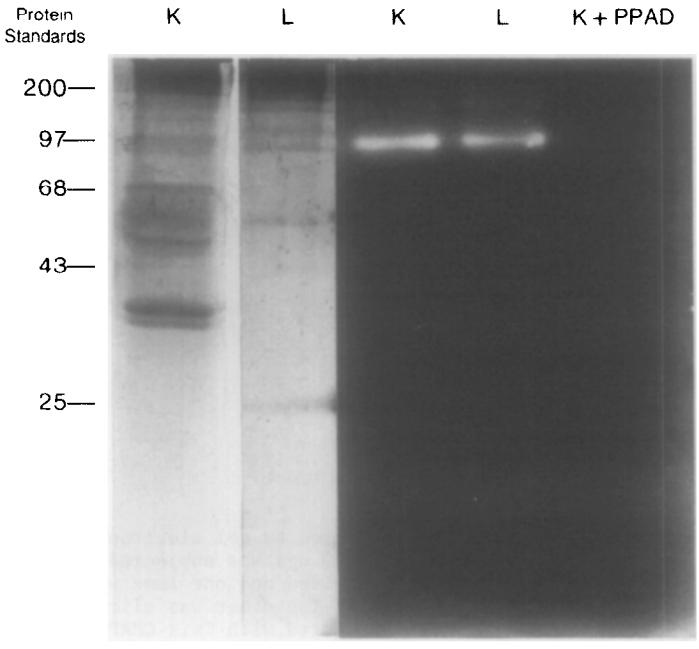

Figure 1. Identification of neutral endopeptidase by fluorescent product formation. Active fractions of neutral endopeptidase from either kidney (K) or ligament (L) membranes were pooled and concentated. Five µg of protein (containing approximately 100 ng of neutral endopeptidase) was applied to a stacking gel (8%) and current of 10 mA was applied to run the sample into the gel. Electrophoresis was at 20 mA at constant voltage. The gels were divided and incubated with substrate alone or in the presence of phophoramidon (PPAD). Mr indicated at the left was estimated from protein standards (Pharmacia) separated in another lane of the gel.

developed fluorescence was limited to a principal band with Mr of approximately 95,000 and a minor band at approximately 200,000 for either preparation. The complete lack of reaction in the gel that was impregnated with substrate in the presence of phosphoramidon (PPAD) suggests the specificity of the reaction for neutral endopeptidase.

In another experiment, partially-purified material from human kidney was separated on two gels run in parallel. One was sliced into 1 mm slices and the proteins were extracted from each slice with Tris-CHAPS buffer. The samples were assayed for neutral endopeptidase using the DAGNPG substrate according to Florentin et al (11). Figure 2 shows that both bands of the 95,000 material are active, but a major protein contaminant (Mr 50,00) had only slight activity. Thus, the substrate gel technique might be a useful preparative procedure when only small amounts of material are available.

Sequence	Position*	Species and Reference	
NF-M			
EKKKAESP	731-737	Human	(Myers et al. 1987)
KAESP	1-5	Pig	(Geisler et al. 1987)
EAGKVSP	9-15, 25-31	Pig	(Geisler et al. 1987)
EKAPSP	16-21, 32-37	Pig	(Geisler et al. 1987)
EVAKESP	637-643	Rat	(Napolitano et al. 1988)
DKKKAESP	661-667	Rat	(Napolitano et al. 1988)
EVTKESP	641-647	Mouse	(Levy et al. 1987)
DKKKAESP	665-670	Mouse	(Levy et al. 1987)
EKP-(S/T)P_{17} (consensus)	270-370	Chicken	(Zopf et al. 1988)
NF-H			
EKAPATP	845-851	Human	(Lees et al. 1988)
DKKKVTP	899-906	Human	(Lees et al. 1988)
EKTPATP	620-626	Rat	(Dautigny et al. 1988)
EKPKDSP	655-661	Rat	(Dautigny et al. 1988)
EKTLPTP	903-909	Mouse	(Julien et al. 1988)
EKKETP	932-937	Mouse	(Julien et al. 1988)
Tau			
KIATP	92-96	Human	(Goedert et al. 1988)
KIATP	81-85	Mouse	(Lee et al. 1988a)
MAP2			
EGKKETSP	660-664	Human	(Kosik et al. 1988)
ESKETP	714-717	Human	(Kosik et al. 1988)
ERRPSP	829-833	Human	(Kosik et al. 1988)
EPKDGSP	852-856	Human	(Kosik et al. 1988)
RISTP	945-949	Human	(Kosik et al. 1988)
EKEATP	934-939	Mouse	(Lewis et al. 1988)
KDTSP	1010-1014	Mouse	(Lewis et al. 1988)
EGKKETSP	1157-1162	Mouse	(Lewis et al. 1988)
EPKDGSP	1349-1353	Mouse	(Lewis et al. 1988)
RISTP	1442-1446	Mouse	(Lewis et al. 1988)

Figure 2: K-SP, K--SP and related sequences in NF-M, NF-H and MAP proteins.

*The position is counted from the deduced N-terminus, except for human MAP2, chicken NF-M, pig NF-M and rat NF-H, for which the entire sequence is not known. In these cases the sequence is numbered from the first known amino acid.

During the analysis described above another feature of neurofilament architecture became clear. Chicken NF-M contains, following four tandem KSP repeats, an unusual repeated sequence found 17 times based on the consensus EKP-(S/T)P (12). Alignment of the chick NF-M molecule with mammalian sequences revealed the presence of homologous sequences in the same relative position at the N-terminal end of the KE segment in all species (see figure 2). Rat and mouse have two of these, but the first is missing in human. Interestingly, Geisler et al. (14) isolated tryptic phosphopeptide fragments of pig NF-M together forming a segment initiated by the sequence KAESP and containing at least two copies of each of the sequences KVSP and KAPSP. This segment presumably corresponds to a multiply repeated version of the KAESP sequence found in other mammals. Taken together these findings suggest the existence of a second type of sequence in NF-M, related to but distinct from the KSP repeats, which may be but is not necessarily multiply repeated and which is probably heavily phosphorylated. The consensus of these NF-M sequences in known mammals is K--SP and K-SP, preceded by glutamic or aspartic acid. KSP and related sequences are found in NF-H (4-8) and microtubule associated proteins MAP2 and tau (15-18). Interestingly, NF-H, MAP2 and tau also contain

several sequences fitting or closely related to the K--SP and K-SP motifs described here (see figure 2). NF-H also contains K---TP sequences, preceded with an acid, which could also belong to this family of sequences.

DISCUSSION

The most likely function for the -EK-V-TKKVVEK-TS sequences is to mediate an important protein-protein interaction, although other possibilities, such as solely structural or metal binding are conceivable. However, the apparent lack of any known non-neurofilament proteins containing similar motifs suggests that their function may be specialized, as would be expected for a neurofilament specific binding domain. Some interactions between neurofilaments and other cytoskeletal proteins, such as fodrin/spectrin, synapsin I, MAP2 and tau have been described (see 19-22). None of these binding interactions localize to the tails of either NF-H and NF-M, leaving the function of these regions something of a mystery. These NF-M repeats could interact with virtually any component of the cytoplasm, and may be involved in neurofilament transport, cross-linking or other important functions. Since these sequences are repeated it is perhaps reasonable to suppose that the putative ligand is also a repeated structure such as a polymer or a multisubunit complex.

The function of the unusual and interesting KSP repeated sequences in neurofilament tails is not known. Protein chemical and immunological studies suggest that phosphorylation on these repeats should have profound structural consequences, although no gross change in appearance of the neurofilament tail is apparent following enzymatic dephosphorylation (23). It has been suggested that phosphorylation may protect the neurofilament tails from proteolysis (24) or mediate low-affinity calcium binding (25). This report identifies another family of serine-proline (SP) containing sequences specific for a defined region of NF-M, which are probably phosphorylated, may be present as multiple copies, and which are related to sequences found in NF-H, tau and MAP2. Presumably the various SP and related sequences represent a family of motifs which have similar or at least related functions. The additional presence of K-SP, K--SP and related sequences in NF-M, NF-H, tau and MAP2 suggests further functional and evolutionary relationships as well as more extensive potential immunological similarities between these proteins than currently recognized. Such further immunological similarities are of significance since antibody studies have implicated all of these proteins in the formation of many pathological lesions, notably the neurofibrillary tangles of Alzheimer's disease. Perhaps some of the more questionable results may be due to the use of antibodies recognizing epitopes defined by the shared sequences described here.

Understanding the significance of these various sequence motifs, the identity of the components that may bind to them and the control and consequences of their phosphorylation should throw light on the function of the NF-M tail, of

neurofilaments in general, and also perhaps of microtubule associated proteins tau and MAP2.

ACKNOWLEDGMENTS

Supported by N.I.H. grants NS22695 and AG07470. I would like to thank Drs. Bob Cohen, Ken Kosik and Paul Chun for valuable discussion.

REFERENCES

1. Lasek, R. J., Phillips, L., Katz, M. J. and Autilio-Gambetti, L. (1985) Ann. New York Acad. Sci. 455:462-478.
2. Geisler, N., Fischer, S., Vandekerckove, J., Van Damme, J., Plessman, U. and Weber, K. (1985) EMBO J. 4:57-63.
3. Hisanaga, S-I. and Hirokawa, N. (1988) J. Mol. Biol. 202:297-305.
4. Lees, J. F., Shneidman, P. S., Skuntz, S. F., Carden, M. J. and Lazzarini, R. A. (1988) EMBO J. 7:1947-1955.
5. Dautigny, A., Pham-Dinh, D., Roussel, C., Felix, J. M., Nussbaum, J. L. and Jolles, P. (1988) Biochem. Biophys. Res. Comm. 154:1099-1106.
6. Breen, K. C., Robinson, P. A., Wion, D. and Anderton, B. H. (1988) FEBS letts. 241:129-136.
7. Julien, J-P., Cote, F., Beaudet, L., Sidky, M., Flavell, D., Grosveld, F. and Mushynski, W. (1988) Gene 68:307-314.
8. Mack, K., Currie, J. R. and Soifer, D. (1988) J. Neurosci. Res. 20:129-136.
9. Myers, M. W., Lazzarini, R. A., Lee V. M-Y., Schlaepfer, W. W. and Nelson, D. L. (1987) EMBO J. 6:1617-1626.
10. Napolitano, E. W., Chin, S. S. M., Colman, D. R. and Liem, R. K. H. (1987) J. Neurosci. 7:2590-2599.
11. Levy, E., Liem, R. K. H., D'Eustachio, P, Cowan, N. J. (1987) Eur. J. Biochem. 166:71-77.
12. Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E. D. and Betz, H. (1987) Genes and Development 1:699-708.
13. Edelman, A. M., Blumenthal, D. K. and Krebs, E. G. (1987) Ann. Rev. Biochem. 56:567-613.
14. Geisler, N., Vandekerckhove, J. and Weber, K. (1987) FEBS Letts. 221:403-407
15. Lee, G., Cowan, N. and Kirschner, M. (1988) Science 239:285-288.
16. Goedert, M., Wischik, C. M., Crowther, R. A., Walker, J. E. and Klug, A. (1988) Proc. Natl. Acad. Sci. USA 85:4051-4055.
17. Lewis, S. A., Wang, D. and Cowan, N. J. (1988) Science 242:936-939.
18. Kosik, K. S., Orecchio, L. D., Bakalis, S., Duffy, L. and Neve, R. L. (1988) J. Neurochem. 51:587-599.
19. Frappier, T., Rignouf, F. and Pradel, L. (1987) Eur. J. Biochem. 169,651-657.
20. Steiner, J. P., Ling, E. and Bennett, V. (1987) J. Biol. Chem. 262:905-914.
21. Heimann, R., Shelanski, M. J. and Liem, R. K. H. (1985) J. Biol. Chem. 260:12160-12166.
22. Miyata, Y., Hoshi, M., Nishida, E., Minami, Y. and Sakai, H. (1986) J. Biol. Chem. 261:13026-13030.
23. Hisanaga, S-H. and Hirokawa, N. (1989) J. Neurosci. 9:959-966.
24. Goldstein, M. E., Sternberger, N. H. and Sternberger, L. A. (1988) J. Neuroimmunol. 14:149-160.
25. Lefebvre, S. and Mushynski, W. E. (1988) Biochemistry 27:8503-8508.