Correspondence

Are β-thymosins WH2 domains?

John Edwards

Available online 28 July 2004

In their review "WH2 domains: a small, versatile adapter for actin monomers", Paunola et al. [1] drew attention to sequence similarity between members of two families of actin-binding sequences, WH2 domains and β-thymosins. The authors represented the two as a single family, by extending the original definition of WH2 domains from 18 to ~35 residues. From sequence similarity, they inferred common ancestry. The enlarged WH2 domain is gaining currency [2–4], yet evidence for such a relationship is lacking.

β-Thymosins are ~43 residue peptides with a role in actin dynamics as monomer-sequestering agents [5]. Recently, a number of proteins containing internal tandem repeats of sequences very similar to monomeric β-thymosins have been identified, first in *Drosophila* and *Caenorhabditis* [6], more recently in *Ciona*, *Dermacentor* and *Hermissenda* [7]. From studies of the *Drosophila* protein ciboulot [8], the role of these proteins in regulation of actin filaments is likely to be different from monomeric thymosins. Homologues of β-thymosins have not been identified in single-celled organisms. WH2 domains [9] (from Wiskott–Aldrich homology 2) are 18-residue sequences that also confer actin binding, are known only as modular parts of larger proteins, but are widely distributed in phylogeny, being found in bacteria (hypothetical proteins in *Rickettsia montana* and *Vibrio parahaemolyticus*), certain viruses of arthropods, single-celled eukaryotes and metazoans, although not plants.

Co-alignment of monomeric thymosins, thymosin repeats and WH2 domains [1] usefully highlighted similarity between the conserved C-terminus of WH2 domains and the hexapeptide motif LKKTET of thymosins, the latter residues long implicated in actin binding [10,11]. However, aligning thymosin repeats and WH2 domains separately (Fig. 1) shows that their patterns of conservation are not co-extensive: there is strong conservation of residues in thymosin repeats C-terminal of this motif, whereas conservation of WH2 domains is N-terminal from it. C-terminal sequences flanking the 18-mer WH2 domains are very heterogeneous, some reach C-terminus of the protein within the conserved span of thymosin repeats, and in others, adjacent WH2 domains overlap into this span.

The similarity between these two families of actin binding modules is too low for their relatedness to be detected by similarity searches. For example, a Hidden Markov Model [15] constructed from the alignment of all 28 known tandem thymosin repeats, used to search non-redundant proteins, finds monomeric thymosins, but not WH2 domains. Conversely, an HMM based on the WH2 PFAM seed [16], with or without augmentation by 18 downstream residues, does not detect thymosins.

The LKKTET motif, or variants, has been found in actin-binding proteins unrelated to thymosins or WH2 domains, such as vertebrate protein kinases C-ϵ (LKKQET) [17] and so may have evolved independently in response to a shared mode of binding to actin. A suggested variant FKHVXPN in the link region of gelsolin is of particular interest, since an X-ray crystallographic study [2] has revealed how and where this binds to actin.

The possibility that the WH2 resemblance to thymosin LKKT might reflect commonality in mechanism of binding to actin is strengthened by the existence of a subset of WH2 domains in which the similarity to the thymosin motif extends to a decapeptide, KLKKAETNDR cf thymosin KLKKTET-

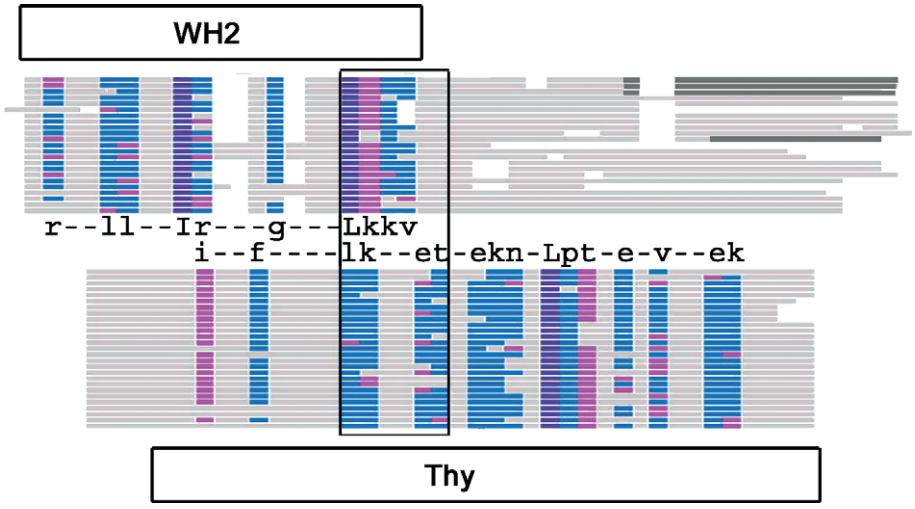

Fig. 1. Distinct patterns of conservation of WH2 and β-thymosin domains. Twenty-four 18-residue WH2 domains were selected at random from 185 currently identified by SMART [12] and each extended C-terminally to 36 residues, or to the protein C-terminus if shorter. All 27 β-thymosin domains from thymosin repeat proteins (i.e., excluding highly conserved monomeric thymosins) identified by SMART were extended three residues N-terminally. Each family was separately aligned with ClustalW [13] and the alignment coloured as a Texshade fingerprint [14]. Purple: all identical, blue: conserved (threshold 50%), cyan, similar. Dark shading indicates position of second tandem WH2 domains. Extent of SMART WH2 and Thy domains, and position of thymosin "hexamotif" boxed.

doi:10.1016/j.febslet.2004.07.038

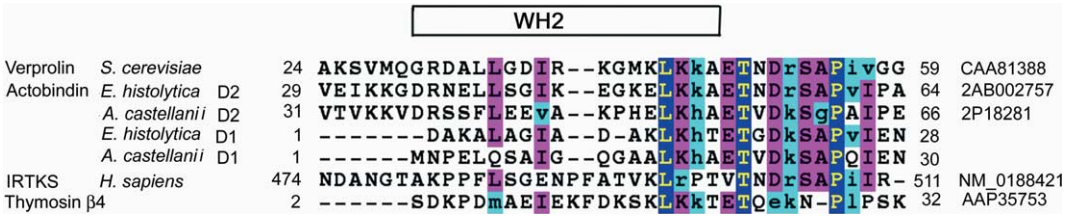

Fig. 2. Alignment of the thymosin-like WH2 domains of verprolin and actobindin, with thymosin β4 and the C-terminus of human insulin receptor tyrosine kinase substrate (IRTKS).

QEK (Fig. 2). However, even for this closest-to-thymosin subset of WH2 domains, searches fail to establish significant *overall* similarity between the notional 35-mer WH2 domain and thymosins.

The *Acanthamoeba* protein actobindin, discovered when WH2 domains were unknown [18], is consistently regarded as a dimeric thymosin, the first of the thymosin repeat proteins [19]. However, the two repeats in actobindin and also those of its putative *Entamoeba* homologue closely match the verprolin subset of WH2 domains in a 14 residue sequence consisting of four residues of the WH2 domain plus a further 10 flanking residues (Fig. 2). The more C-terminal *Entamoeba* repeat is a canonical WH2 domain and its *Acanthamoeba* homologue is closely related. The more N-terminal, shorter actobindin repeats, although less recognisable as WH2 domains, share the 14-residue sequence. (A similar 14-residue sequence occurs at the C-terminus of members of the IRSp53 protein family [20]. In MIM, ABBA1 and ABB2 it C-terminally flanks WH2 domains, whilst in IRTKS and FLJ22582 the rest of the WH2 consensus is absent [20].) The hypothesis that actobindin repeats are related in evolution to WH2 domains rather than β–thymosins, despite their functional similarity to thymosin repeat proteins [19], would make good sense of the phylogeny, since the latter are otherwise found only in metazoans.

It is axiomatic that "motif identity in the absence of overall sequence similarity is not a reliable indicator of homology." [21] Given their limited sequence similarity, which is as readily explained by convergence, common ancestry of WH2 domains and β-thymosins is not supported by available data.

References

[1] Paunola, E., Mattila, P.K. and Lappalainen, P. (2002) FEBS Lett. 513, 92–97.
[2] Irobi, E., Burtnick, L.D., Urosev, D., Naryan, K. and Robinson, R.C. (2003) FEBS Lett. 552, 86–90.
[3] Ksiazek, D., Brandstetter, H., Israel, L., Bourenkov, G.P., Katchalova, G., Janssen, K.P., Bartunik, H.D., Noegel, A.A., Schleicher, M. and Holak, T.A. (2003) Structure 11, 1171–1178.
[4] Mattila, P.K, Salminen, M., Yamashiro, T. and Lappailainen, P. (2003) J. Biol. Chem. 278, 8452–8459.
[5] Huff, T., Muller, C.S.G., Otto, A.M., Netzker, R. and Hannappel, E. (2001) Int. J. Biochem. Cell Biol. 33, 205–220.
[6] Manuel, M., Kruse, M., Müller, W.E.G. and Le Parco, Y. (2000) J. Mol. Evol. 51, 378–381.
[7] Crow, T., Redell, J.B., Tian, L.M., Bian, J.X. and Dash, P.K. (2003) J. Neurosci. 23, 3415–3422.
[8] Boquet, I., Boujemas, R., Carlier, M.-F. and Preat, T. (2000) Cell 102, 797–808.
[9] Machesky, L.M., Insall, R.H. and Volkman, L.E. (2001) Trends Cell Biol. 11, 286–287.
[10] Vancompernolle, K., Goethals, M., Huet, C., Louvard, D. and Vandekerckhove, J. (1992) EMBO 11, 4739–4746.
[11] Van Troys, M., Dewitte, D., Goethals, M., Carlier, M.-F., Vandekerckhove, J. and Ampe, C. (1996) EMBO J. 15.
[12] Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) Proc. Natl. Acad. Sci. USA 95, 5857–5864.
[13] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[14] Beitz, E. (2000) Bioinformatics 16, 135–139.
[15] Eddy, S.R. (2001). Available from: http://hmmer.wustl.edu/.
[16] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L.L. (2002) Nucleic Acids Res. 30, 276–280.
[17] Prekeris, R., Hernandez, R.M., Mayhew, M.W., White, M.K. and Terrian, D.M. (1998) J. Biol. Chem. 273, 26790–26798.
[18] Lambooy, P.K. and Korn, E.D. (1986) J. Biol. Chem. 261, 17150–17155.
[19] Hertzog, M.H., Yarmolas, E.G., Didry, D., Bubb, R. and Carlier, M.-F. (2002) J. Biol. Chem. 277, 14786–14792.
[20] Yamagishi, A., Masuda, M., Ohki, T., Onishi, H. and Mochizuki, N. (2004) J. Biol. Chem. 279, 14929–14936.
[21] Pearson, W.R. (2000) Methods Mol. Biol. 132, 185–219.

Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK