FEBS Letters 519 (2002) 178–180 FEBS 26107

An EVH1/WH1 domain as a key actor in TGFβ signalling

Isabelle Callebaut*

Systèmes Moléculaires and Biologie Structurale, LMCP, CNRS UMR 7590, Universités Paris 6 and Paris 7, Case 115, 4 place Jussieu, 75252 Paris cedex 05, France

Received 28 March 2002; accepted 4 April 2002

First published online 29 April 2002

Edited by Veli-Pekka Lehto

Abstract EVH1 (enabled VASP (vasodilator-stimulated protein) homology 1)/WH1 (WASP (Wiskott–Aldrich syndrome protein) homology 1) domains, present in Ena VASP and WASP, are protein interaction modules specialised in binding proline-rich ligands. An EVH1/WH1 domain is here identified in the recently cloned SMIF protein, a key protein in transforming growth factor-β (TGFβ) signalling which was not yet related to defined domains. The SMIF EVH1/WH1 domain interacts with the proline-rich Smad4 activation domain, leading to translocation of so-formed complex to the nucleus where SMIF possesses strong intrinsic TGFβ-inducible transcriptional activity. This finding highlights the pivotal role that the EVH1/WH1 family of domains play in multiple eukaryotic signal transduction pathways.

Key words: SMIF; Smad4; Sequence analysis; PSI-BLAST; Hydrophobic cluster analysis

1. Introduction

The transforming growth factor-β (TGFβ) cytokine superfamily regulates the proliferation of many cell types and includes signalling molecules such as the TGFβs, bone morphogenetic proteins (BMPs), activins, inhibins and Müllerian-inhibiting substance [1–3]. The activated receptors of this superfamily phosphorylate their associated SMAD proteins which propagate the signal through homo- and hetero-oligomeric interactions. The tumour suppressor Smad4 plays a central role in the signalling mediated by the TGFβ superfamily as it is the shared hetero-oligomerisation partner of the other SMAD proteins. When assembled with its partners, Smad4 translocates to the nucleus where transcriptional complexes are formed with specific co-factors.

Recently, a new protein called SMIF was reported as forming a TGFβ/BMP4-inducible complex with Smad4, that translocates to the nucleus and possesses transcriptional activity [4]. Deletion analyses located the Smad4-interacting domain of SMIF in the N-terminal 100 amino acids, whereas the SMIF-interacting region of Smad4 was observed in the Smad4 linker region C-terminus (amino acids 275–308), a region which is proline-rich and which is included in the Smad4 activation domain (SAD; amino acids 275–322). As noticed by the authors, sequence analysis using the SMIF sequence as query failed to identify significant similarities with other proteins or defined domains and thus let this protein orphan from family.

Here, I show that the N-terminal region of SMIF (amino acids 1–131), including the Smad4-interacting domain, belongs to the EVH1 (enabled VASP (vasodilator-stimulated protein) homology 1)/WH1 (WASP (Wiskott–Aldrich Syndrome protein) homology 1) family of domains [5,6], a now well-described family within the growing number of modules that bind proline-rich ligands [7,8]. EVH1/WH1 domains are present in various proteins involved in the maintenance of cytoskeletal integrity such as Ena, VASP and WASP, mutated in a X-linked recessive disorder. The EVH1/WH1 family also includes, among others, the neuronal Homer protein that binds to glutamate and inositol triphosphate receptors, the Spred protein, a suppressor of Ras signalling [9] and the Drosophila still life protein type 1 (SIF1) protein, that interacts with Rho-like GTPases and participates in the organisation of actin cytoskeleton [10]. EVH1/WH1 domains of the mammalian Ena (Mena) and Ena/VASP-like proteins, whose structures have been solved in complex with peptides, specifically recognise proline-rich ligands that adopt a polyproline II conformation (PPII, left-handed helix with three residues per turn) [11,12]. Similarly to SH3 domains, the peptide-binding sites contain critical aromatic amino acids (Y16, W23, F77 in the Mena sequence) that interact with prolines through a V-shaped concave groove (Fig. 1). EVH1/WH1 can however adapt a variety of proline-rich ligands, as exemplified by those specific of the Homer/Vesl family, adopting only partly a PPII conformation [13]. Although the binding specificities of the WH1 domain of WASP and N (neural)-WASP are not defined at the peptide level, this one binds a C-terminal, proline-rich region of WASP-interacting protein (WIP) and several mutations within the WASP WH1 domain that cause Wiskott–Aldrich syndrome impair this interaction [14–16].

The here-reported identification of an EVH1/WH1 domain in the SMIF N-terminal sequence is consistent with its observed ability to interact with a proline-rich region, and provides thus new insights in the characterisation of this interaction at the molecular level.

2. Materials and methods

The non-redundant database (NR) at NCBI was searched using PSI-BLAST [17]. The SMART domain database [18] was used to

*Fax: (33)-1 44 27 45 87.
E-mail address: isabelle.callebaut@lmcp.jussieu.fr (I. Callebaut).

Abbreviations: BMP, bone morphogenetic protein; VASP, vasodilator-stimulated protein; EVH1, enabled VASP homology 1; TGFβ, transforming growth factor-β; SIF1, still life protein type 1; WASP, Wiskott–Aldrich syndrome protein; WH1, WASP homology 1; HCA, hydrophobic cluster analysis

PII: S0014-5793(02)02751-5

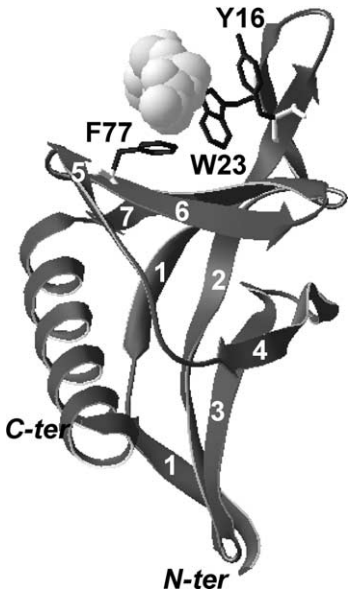

Fig. 1. Ribbon representation of the three-dimensional structure of the Mena EVH1 domain, in complex with the ActA FPPPP peptide (pdb 1evh). The three aromatic acids which are involved in ligand binding and are highly conserved in the EVH1/WH1 family are shown and labelled. Strands are labelled according to Fig. 2.

refine PSI-BLAST alignments (EVH1/WH1 domains: SMART accession number SM0461 (http://smart.embl-heidelberg.de)). The two-dimensional hydrophobic cluster analysis (HCA) [19,20] was also used, as it offers the possibility to add information about secondary structures to the lexical analysis of the considered sequences. Conservation of hydrophobic cluster features which participate in the protein core, together with sequence similarities, is associated with the maintenance of a similar structure and often allows the alignment procedure for highly divergent sequences (typically in the 10–20% sequence identity range, below the so-called twilight zone (25–30%)).

Visualisation of three-dimensional structures was performed using Swiss-PdbViewer [21].

3. Results and discussion

As stated by Bai et al. [4], no significant similarities could be detected to known mammalian proteins or defined domains using the SMIF sequence as query.

The belonging of the N-terminal domain of SMIF to the EVH1/WH1 family was significantly highlighted in a roundabout way, using the WH1 domain of human WASP as query (amino acids 1–170) in a PSI-BLAST [17] search (BLAST 2.2.2, non-redundant database at NCBI (887 672 sequences)). In this way, we significantly detected at convergence by iteration 15 all the known members of the EVH1/WH1 family, but also the N-terminal region of SMIF (amino acids 1–147), exactly matching the Smad4-interacting domain.

A multiple alignment against the EVH1/WH1 family was constructed on the basis of the PSI-BLAST results and refined using the EVH1/WH1 multiple alignment of the SMART database [18] and HCA [20] (Fig. 2). This alignment shows that despite low levels of sequence identity (below 20%), hydrophobic amino acids that participate in the maintenance of the fold (PH-fold superfamily) are conserved in the SMIF sequence (balls on Fig. 2), as well as amino acids which are involved in ligand binding (stars and arrows on Fig. 2). In particular, aromatic amino acids which are highly conserved in the EVH1/WH1 family and interacting with ligand are strictly conserved in the SMIF sequence (Fig. 2; Y36, W45 and F110 in the SMIF sequence, aligned with Y16, W23 and F77 of Mena, three amino acids which are also depicted on Fig. 1). A strict conservation of these critical positions of the alignment is also found in the putative Drosophila homologue of human SMIF (CG11183, GenBank identification (gi) number: 18485748).

It is worth noting that despite the conservation of residues that are critical to the fold and to the function, the EVH1/WH1 domain of SMIF was not detected when searching the domain databases (SMART [18] and Pfam [22] databases). This lack of detection can be due to sequence divergence

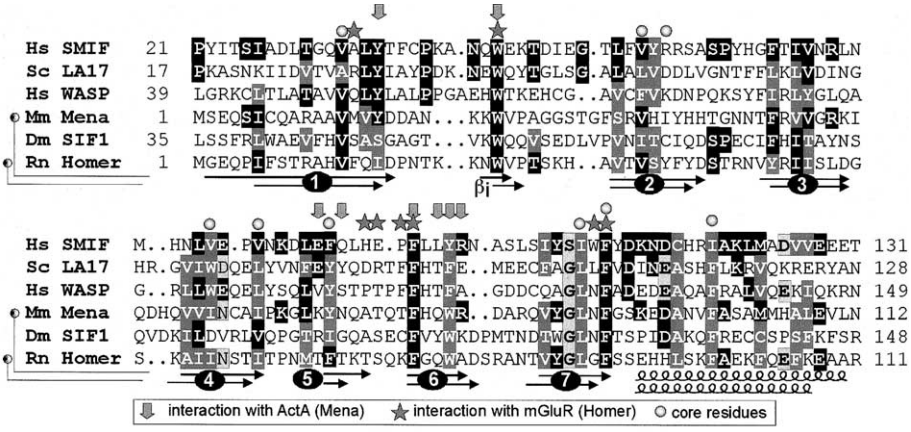

Fig. 2. Alignment of the SMIF N-terminal domain with several representative members of the EVH1/WH1 family. This alignment was deduced from the PSI-BLAST analysis and refined on the basis of the SMART alignments [18] and HCA [20]. Identical amino acids between SMIF and other EVH1/WH1 domains are on a black background, similar ones are shaded grey. Hydrophobic core residues are indicated (as defined for Mena (PDB 1evh)), as well as those positions that are involved in ligand binding. Positions of the regular secondary structures (seven β-strands and a small extra strand βi, one α-helix), as experimentally determined for Mena (PDB 1evh) and Homer (PDB 1ddv) are indicated below the sequences, according to Fig. 1. The highest similarities of the SMIF sequence with EVH1/WH1 domains are observed with LA17, the yeast homologue of WASP. Like many domains of this family, the EVH1/WH1 domain of SMIF is located N-terminal. GenBank identification (gi) numbers: Hs (*Homo sapiens*) SMIF, 8923767; Sc (*Saccharomyces cerevisiae*) LA17, 2498506; Hs WASP, 1722836; Mm (*Mus musculus*) Mena, 5107580; Dm (*Drosophila melanogaster*) SIF1, 6094287; Rn (*Rattus norvegicus*) Homer, 8569598.

around strands β4–β6 and to the associated difficulties that lexical procedures have to align correctly these regions.

Regarding the here-reported observations, the SMIF N-terminal domain can be thus predicted as a new class of functional EVH1/WH1 domain, consistently with its observed ability to interact with a proline-rich region of Smad4. A good candidate for the interacting peptide in Smad4 is the HPPMPP sequence (amino acids 291–296), whose structure (PDB 1dd1) can be well superimposed on the Mena EVH1 ligand sequence (PDB 1evh) (data not shown). Y301 and W302, both amino acids whose mutations have been shown to impair the interaction with SMIF [4], flank this Smad4 peptide. These could thus play a crucial role in the specificity towards the SMIF EVH1/WH1 domain, similarly to that played by residues flanking proline-rich sequences in well-described EVH1/WH1 domains.

In conclusion, the classification of SMIF as a new member of the EVH1/WH1 family brings new insights in the characterisation of the SMIF–Smad4 interaction at the molecular level but also highlights the critical role that the EVH1/WH1 plays in multiple signalling pathways.

Acknowledgements: I would like to thank Jean-Paul Mornon for critical reading of the manuscript and Pierre Dehoux for initiating our interest in EVH1/WH1 domains several years ago. This work was supported by the CNRS-INRA-INRIA-INSERM 'Action Bioinformatique'.

References

[1] Heldin, C., Miyazono, K. and ten Dijke, P. (1997) Nature 390, 465–471.
[2] Massague, J., Blain, S. and Lo, R. (2000) Cell 103, 295–309.
[3] Attisano, L. and Wrana, J. (2000) Curr. Opin. Cell Biol. 12, 235–243.
[4] Bai, R., Koester, C., Ouyang, T., Hahn, S., Hammerschmidt, M., Peschel, C. and Duyster, J. (2002) Nat. Cell Biol. 4, 181–190.
[5] Ponting, C. and Philips, C. (1997) J. Mol. Med. 75, 769–771.
[6] Callebaut, I., Cossart, P. and Dehoux, P. (1998) FEBS Lett. 441, 181–185.
[7] Kay, B., Williamson, M. and Sudol, M. (2000) FASEB J. 14, 231–240.
[8] Ball, L.J., Jarchau, T., Oschkinat, H. and Walter, U. (2002) FEBS Lett. 513, 45–52.
[9] Wakioka, T. et al. (2001) Nature 412, 647–651.
[10] Sone, M. et al. (1997) Science 275, 543–547.
[11] Fedorov, A., Fedorov, E., Gertler, F. and Almo, S. (1999) Nat. Struct. Biol. 6, 661–665.
[12] Prehoda, K., Lee, D. and Lim, W. (1999) Cell 97, 471–480.
[13] Beneken, J., Tu, J., Xia, B., Nuriya, M., Yuan, J., Worley, P. and Leahy, D. (2000) Neuron 26, 143–154.
[14] Martinez-Quiles, N. et al. (2001) Nat. Cell Biol. 3, 484–491.
[15] Ramesh, N., Anton, I., Hartwig, J. and Geha, R. (1997) Proc. Natl. Acad. Sci. USA 94, 14671–14676.
[16] Stewart, D., Tian, L. and Nelson, D. (1999) J. Immunol. 162, 5019–5024.
[17] Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Nucleic Acids Res. 25, 3389–3402.
[18] Letunic, I. et al. (2002) Nucleic Acids Res. 30, 242–244.
[19] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.-P. (1987) FEBS Lett. 224, 149–155.
[20] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J. (1997) Cell. Mol. Life Sci. 53, 621–645.
[21] Guex, N. and Peitsch, M.C. (1997) Electrophoresis 18, 2714–2723.
[22] Bateman, A. et al. (2002) Nucleic Acids Res. 30, 276–280.