Cell, Vol. 80, 693–694, March 10, 1995, Copyright © 1995 by Cell Press

Letter to the Editor

A Phosphotyrosine Interaction Domain

A major step in the understanding of signal transduction came with the realization that Src homology 2 (SH2) domains bind peptide sequences containing phosphotyrosine (Pawson and Schlessinger, 1993). Shc is an SH2 domain protein that becomes tyrosine phosphorylated after activation of cells by various cytokines and growth factors. This phosphorylation enables Shc to bind to Grb2, coupling Shc to the Ras activation pathway (Rozakis-Adcock et al., 1992). Surprisingly, Kavanaugh and Williams (1994) as well as our group (Blaikie et al., 1994) have recently identified a region in Shc distinct from the SH2 domain that can also bind tyrosine-phosphorylated proteins. This region appears structurally unrelated to SH2 domains and may impart on Shc its unique ability to bind to the Asn-Pro-X-Tyr(P) motif found in many tyrosine-phosphorylated proteins, including growth factor receptors (Obermeier et al., 1993; Stephens et al., 1994; Campbell et al., 1994).

The minimum binding domain of the amino terminus of Shc that can mediate phosphotyrosine-dependent interactions has been defined as amino acids 46–209 of the 52 kDa form of Shc (Blaikie et al., 1994). We have subjected this region of Shc and a related protein, Sck (Kavanaugh and Williams, 1994), to various data base search methods (Koonin et al., 1994) in order to identify other proteins that might contain homologous domains. Indeed, we were able to retrieve nine distinct proteins (Figures 1 and 2) that contained a similar domain, several of which have been sequenced in different organisms. We propose the name PID (for phosphotyrosine interaction domain) for this protein–protein interaction motif. The average length of the PID is about 160 amino acids, but it can vary considerably. Structural predictions suggest that it is a globular domain, and the arrangement of the helices and β strands indicates the presence of an antiparallel β sheet (Figure 1). Although all the sequences share common features such as hydrophobicity patterns and conserved motifs, only two positions appear to be invariant (Gly-58 and Ser-151 of p52 Shc).

The biochemical data obtained with the PID of Shc indicate that it may have a general role in tyrosine kinase signal transduction. This is supported by the identification of a PID in Disabled, a tyrosine-phosphorylated protein found in Drosophila (Gertler et al., 1993). A mouse gene encoding a mitogen-responsive phosphoprotein with a PID very similar to Disabled was also found in the GenBank data base (accession number U18869; p96 in Figures 1 and 2), as was a partial sequence of a putative human ortholog (Mok et al., 1994; DOC-2 in Figure 1). Another member of the PID family, numb, has not been suspected to be part of a tyrosine kinase signaling pathway but has an important role in determining cell fates in the Drosophila nervous system (Rhyu et al., 1994).

The *X11* gene was originally isolated as a candidate gene for Friedreich ataxia (Duclos et al., 1993). Its PID has the closest sequence similarity to those of Sck and Shc (29% sequence identity). The X11 proteins contain two successive disc homologous region (DHR) repeats of about 100 amino acids downstream of the PID (Figure 2).

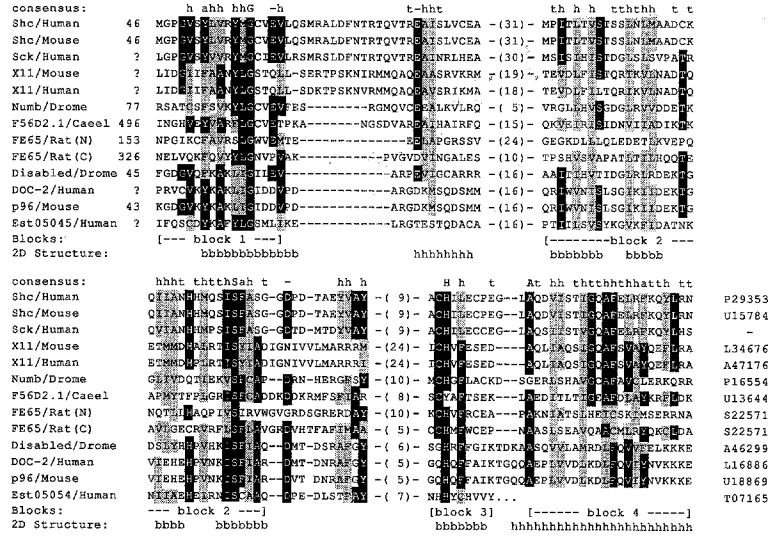

Figure 1. Sequence Alignments of the PID

Initial data base searches with the putative binding domains of Shc and Sck were performed using the BLAST series of programs (Altschul et al., 1994, and references therein). F56D2.1, X11, and numb had scores indicative of a homology (probability of matching by chance, $p < 0.03$). Their consistent alignments (i.e., they all matched the same regions of the query protein) prompted a detailed analysis. Additional BLAST runs with the candidate sequences using PROFILE (Gribskov et al., 1987), as well as iterative motif searches (Tatusov et al., 1994) with the four most conserved blocks (marked in the figure), confirmed the similarity and identified the remaining proteins of the set. For example, F56D2.1 and Disabled or p96 have significant BLASTP p values of $<3.3 \times 10^{-6}$. Furthermore, all sequences match each of the blocks independently with high significance as computed using MOST (Tatusov et al., 1994); e.g., all sequences match block 2, with probabilities of occurring by chance of $p < 1 \times 10^{-8}$. The consensus is indicated on the top line: capital letters, highly conserved residues; h, hydrophobic; a, aromatic; t, polar/turnlike; minus, acidic residues. White letters on black background denote residues conserved in more than 50% of the proteins. Stippled residues underline the hydrophobic or aromatic nature of a position in agreement with the consensus. Predicted secondary (two-dimensional) structures (Rost and Sander, 1994) are indicated by h (helices) and b (β strands). Dashes and numbers in brackets indicate gaps in the alignment. The accession numbers for the proteins are listed at the far right. The sequence for Sck has not been deposited and is derived from the literature (Kavanaugh and Williams, 1994). FE65 has an amino-terminal (N) and carboxy-terminal (C) PID.

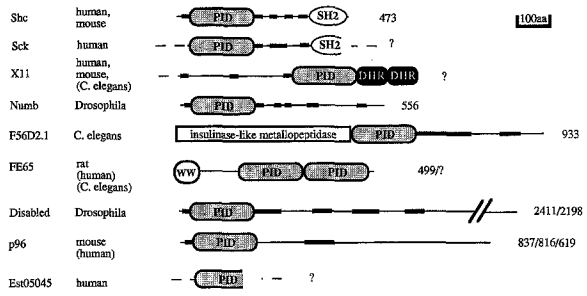

Figure 2. Schematic Representation of Proteins Containing a PID

The domains represented are these: DHR (Ponting and Philip, 1995), WW (Bork and Sudol, 1994), SH2 (Pawson and Schlessinger, 1993), and insulinase-like metallopeptidase (Rawlings and Barrett, 1991). Compositionally biased and probably nonglobular segments such as proline-rich regions, as indicated by the program SEG (Wootton, 1994), are shown by thick lines. Species names in parentheses indicate the presence of partial sequences in the data base (mostly ESTs) that appear to represent orthologs of these proteins. The numbers to the far right indicate the number of amino acids in the proteins; Disabled, p96, and FE65 appear to have alternatively spliced forms. The dashed lines indicate that only fragments have been sequenced.

The DHR domain was first found in several proteins that localize at postsynaptic, septate, and tight junctions, but its presence in various other intracellular signaling proteins is now established (Ponting and Philip, 1995). FE65 was originally proposed as a transcriptional activator (Duilio et al., 1991). However, the presence of a WW domain (Figure 2; Bork and Sudol, 1994) as well as two PIDs suggests that it is likely to be involved in signal transduction. *F56D2.1* is a putative gene identified in the Caenorhabditis elegans genome sequencing project (Wilson et al., 1994) that contains an amino-terminal metalloprotease domain. Various expressed sequence tags (ESTs) cover parts of the PID, but, with the exception of human *Est05045*, they seem to encode the previously identified genes with PIDs from other species. Nevertheless, these ESTs demonstrate a broad species range and a considerable expression level, and their degree of sequence similarity indicates conservation of the domain during evolution.

The presence of the PID in several otherwise unrelated regulatory proteins suggests a general role for this domain in protein–protein interactions and signal transduction. The exact binding preference of the different PIDs needs to be determined experimentally, but it is tempting to speculate that proteins containing PIDs are involved in tyrosine kinase signaling pathways.

Peer Bork* and Benjamin Margolis†
*Max Delbrück Center for Molecular Medicine
13122 Berlin-Buch
Federal Republic of Germany
and European Molecular Biology Laboratory
69012 Heidelberg
Federal Republic of Germany
†Department of Pharmacology
and Kaplan Cancer Center
New York University Medical Center
New York, New York 10016

References

Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994). Nature Genet. *6*, 119–129.

Błaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B. (1994). J. Biol. Chem. *269*, 32031–32034.

Bork, P., and Sudol, M. (1994). Trends Biochem. Sci. *19*, 531–533.

Campbell, K. S., Ogris, E., Burke, B., Su, W., Auger, K. R., Druker, B. J., Schaffhausen, B. S., Roberts, T. M., and Pallas, D. C. (1994). Proc. Natl. Acad. Sci. USA *91*, 6344–6348.

Duclos, F., Boschert, U., Sirugo, G., Mandel, J. L., Hen, R., and Koenig, M. (1993). Proc. Natl. Acad. Sci. USA *90*, 109–113.

Duilio, A., Zambrano, N., Mogavero, A. R., Ammendola, R., Cimino, F., and Russo, T. (1991). Nucl. Acids Res. *19*, 5269–5274.

Gertler, F. B., Hill, K. K., Clark, M. J., and Hoffmann, F. M. (1993). Genes Dev. *7*, 441–453.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Proc. Natl. Acad. Sci. USA *84*, 4355–4358.

Kavanaugh, W. M., and Williams, L. T. (1994). Science *266*, 1862–1865.

Koonin, E. V., Bork, P., and Sander, C. (1994). EMBO J. *13*, 493–504.

Mok, S. C., Wong, K. K., Chan, R. K., Lau, C. C., Tsao, S. W., Knapp, R. C., and Berkowitz, R. S. (1994). Gynecol. Oncol. *52*, 247–252.

Obermeier, A., Lammers, R., Wiesmuller, K. H., Jung, G., Schlessinger, J., and Ullrich, A. (1993). J. Biol. Chem. *268*, 22963–22966.

Pawson, T., and Schlessinger, J. (1993). Curr. Biol. *3*, 434–442.

Ponting, P., and Philip, C. (1995). Trends Biochem. Sci., in press.

Rawlings, N. D., and Barrett, A. J. (1991). Biochem J. *275*, 389–391.

Rhyu, M. S., Jan, L. Y., and Jan, Y. N. (1994). Cell *76*, 477–491.

Rost, B., and Sander, C. (1994). Proteins *19*, 55–72.

Rozakis-Adcock, M., McGlade, J., Mbamalu, G., Pelicci, G., Daly, R., Li, W., Batzer, A., Thomas, S., Brugge, J., Pelicci, P. G., Schlessinger, J., and Pawson, T. (1992). Nature *360*, 689–692.

Stephens, R. M., Loeb, D. M., Copeland, T. D., Pawson, T., Greene, L. A., and Kaplan, D. R. (1994). Neuron *12*, 691–705.

Tatusov, R. D., Altschul, S. F., and Koonin, E. V. (1994). Proc. Natl. Acad. Sci. USA *91*, 12091–12095.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. (1994). Nature *368*, 32–38.

Wootton, J. C. (1994). Comput. Chem. *18*, 269–285.