Converging on proline: the mechanism of WW domain peptide recognition

Ali Zarrinpar and Wendell A. Lim

Two new structures reveal the general rules of proline-rich motif recognition by WW domains and show that they are strikingly similar to those used by SH3 domains.

Specific protein–protein interactions are essential for the precise transmission of information in eukaryotic signal transduction pathways. Surprisingly, however, only a small number of modular protein recognition domains mediate the majority of protein–protein connections[1–3]. Some of the most common classes of modular recognition domains are those that recognize proline-rich motifs[4]. These domain families include, among others, SH3, EVH1 and WW domains. The importance of this class of interactions is underscored by the recent finding that in both the *Drosophila melanogaster* and the *Caenorhabditis elegans* genomes, the most frequently occurring motifs are proline-rich domains[5].

These observations raise several fundamental questions. How are proline motifs recognized? Why are they so commonly used in signaling pathways? And given the abundance of diverse proline motifs, how is sequence specificity achieved both across and within domain families? Two new high resolution structures of WW domains in complex with their proline-rich ligands, presented on pages 634 and 639 of this issue of *Nature Structural Biology* shed new light onto these issues[6,7].

WW domains

WW domains, named after a pair of conserved tryptophans, are highly compact (35–45 residues) modular domains that adopt an antiparallel three-stranded fold[8,9]. Although it has long been clear that WW

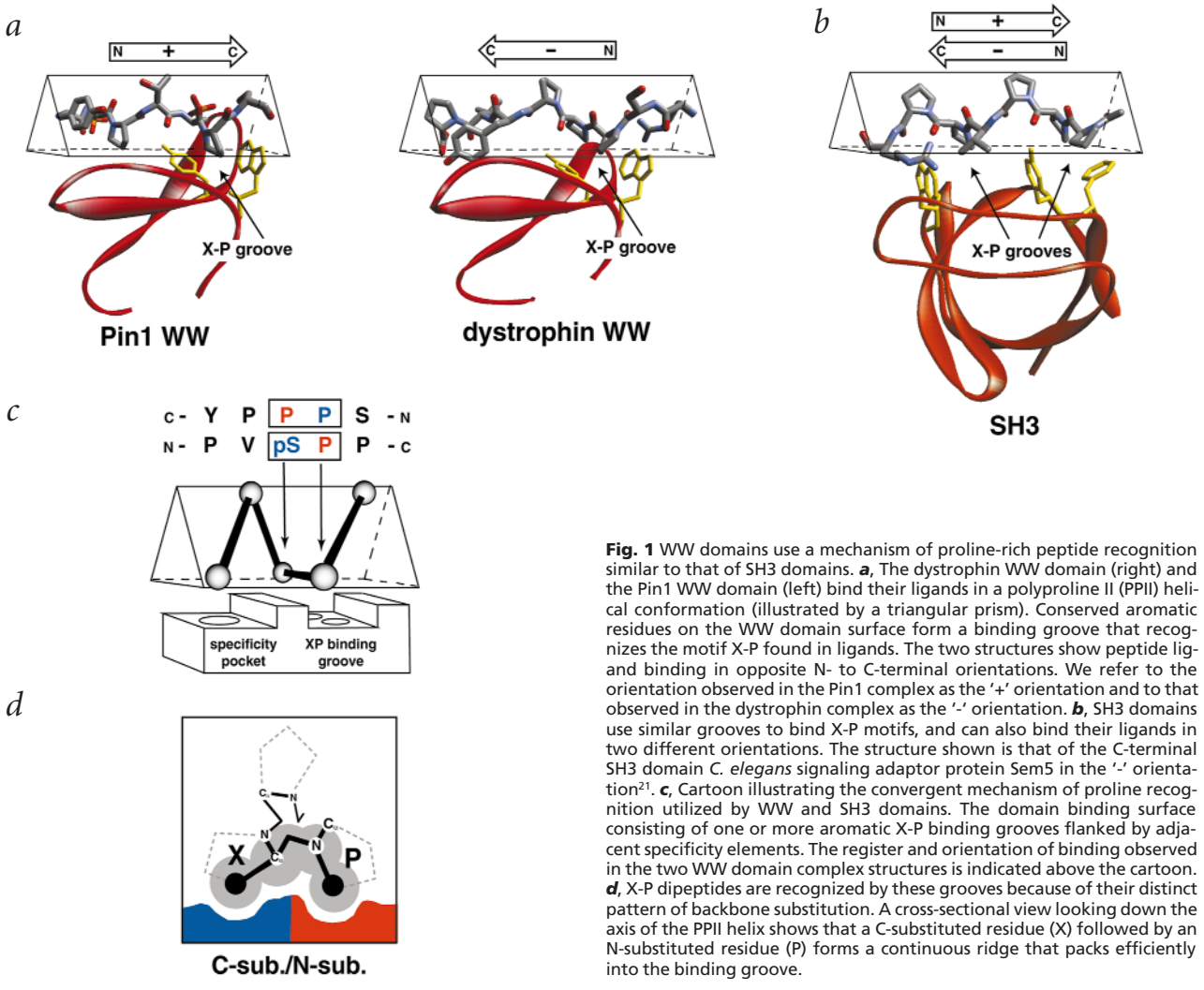

Fig. 1 WW domains use a mechanism of proline-rich peptide recognition similar to that of SH3 domains. ***a***, The dystrophin WW domain (right) and the Pin1 WW domain (left) bind their ligands in a polyproline II (PPII) helical conformation (illustrated by a triangular prism). Conserved aromatic residues on the WW domain surface form a binding groove that recognizes the motif X-P found in ligands. The two structures show peptide ligand binding in opposite N- to C-terminal orientations. We refer to the orientation observed in the Pin1 complex as the '+' orientation and to that observed in the dystrophin complex as the '-' orientation. ***b***, SH3 domains use similar grooves to bind X-P motifs, and can also bind their ligands in two different orientations. The structure shown is that of the C-terminal SH3 domain *C. elegans* signaling adaptor protein Sem5 in the '-' orientation[21]. ***c***, Cartoon illustrating the convergent mechanism of proline recognition utilized by WW and SH3 domains. The domain binding surface consisting of one or more aromatic X-P binding grooves flanked by adjacent specificity elements. The register and orientation of binding observed in the two WW domain complex structures is indicated above the cartoon. ***d***, X-P dipeptides are recognized by these grooves because of their distinct pattern of backbone substitution. A cross-sectional view looking down the axis of the PPII helix shows that a C-substituted residue (X) followed by an N-substituted residue (P) forms a continuous ridge that packs efficiently into the binding groove.

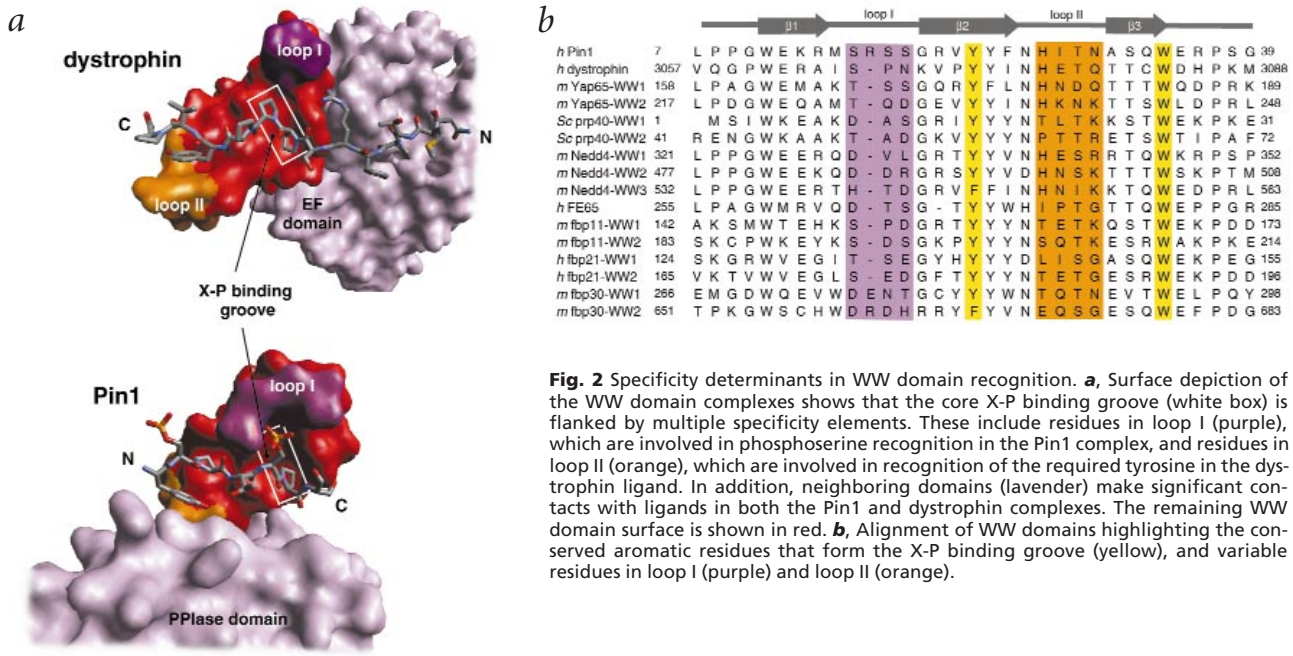

Fig. 2 Specificity determinants in WW domain recognition. ***a***, Surface depiction of the WW domain complexes shows that the core X-P binding groove (white box) is flanked by multiple specificity elements. These include residues in loop I (purple), which are involved in phosphoserine recognition in the Pin1 complex, and residues in loop II (orange), which are involved in recognition of the required tyrosine in the dystrophin ligand. In addition, neighboring domains (lavender) make significant contacts with ligands in both the Pin1 and dystrophin complexes. The remaining WW domain surface is shown in red. ***b***, Alignment of WW domains highlighting the conserved aromatic residues that form the X-P binding groove (yellow), and variable residues in loop I (purple) and loop II (orange).

domains recognize proline-containing sequences, the mechanism of recognition has been difficult to elucidate for several reasons. First, WW domains have highly diverse sequence preferences. For example, WW domains of both the Yes-associated protein YAP65 and dystrophin prefer the motif Pro-Pro-X-Tyr (P-P-X-Y)[10,11], while the Pin1 WW domain binds phosphoSer-Pro (pS-P) motifs[7]. Therefore, it has been a challenge to clearly delineate any common pattern of recognition across the family, aside from the general preference for proline[12]. Second, a lack of high resolution structural data has hampered a clear understanding of the mechanism of WW domain recognition.

Two groups have now determined high resolution structures of WW domain–ligand complexes. Huang *et al.*[6] have determined the structure of a WW domain from dystrophin in complex with a P-P-X-Y ligand from β-dystroglycan. This interaction is essential for the formation of the dystrophin complex[11], and its disruption can result in Duchenne or Becker muscular dystrophy[13]. Verdecia *et al.*[7] have determined the structure of the WW domain-containing protein Pin1, in complex with a phosphoserine peptide from the C-terminal domain of RNA polymerase II. Pin1 is a mitotic peptidyl-prolyl isomerase (PPIase) that binds phosphoSer (pS) or phosphoThr (pT) followed by a proline (pS-P or pT-P)[14]. Such motifs are generated by cyclin-dependent kinases, and their recognition by Pin1 is proposed to regulate mitotic progression[15,16].

A convergent mechanism of proline recognition

Both WW domain complexes show striking similarity to SH3 and other proline recognition complexes, suggesting that these evolutionarily unrelated domains have converged upon a similar solution for proline recognition (Fig. 1). First, the peptide ligands are recognized in a polyproline II (PPII) helix, a three-residue per turn left-handed helix that is preferentially formed by proline-rich sequences[17]. Second, the domain surface that recognizes key prolines is composed of a series of nearly parallel aromatic residues. These aromatic residues form a series of ridges and grooves on the domain surface, against which the PPII helix packs. The WW domains have a single groove formed by a conserved Tyr and Trp (Fig. 1*a*) while SH3 domains have a pair of grooves formed by conserved Trp, Tyr, and Phe residues (Fig. 1*b*)[18,19]. Third, proline residues are recognized by these grooves in a nearly identical manner. Each groove actually recognizes a pair of residues of the sequence X-P, where X is a variable amino acid (Fig. 1*c*). The WW domain interface has at its core one X-P binding groove, while the SH3 domain interface has two successive X-P grooves. Profilin, an actin monomer binding protein, also uses this mechanism of proline recognition, and has two X-P grooves on its binding surface.

An important consequence of this mechanism of recognition is that WW domains can bind peptides in either one of two possible orientations. Indeed, the ligand in the Pin1 WW domain complex ('+' orientation) binds in exactly the opposite orientation from that in the dystrophin WW domain complex ('-' orientation) (Fig. 1*a*). This unusual behavior, first observed in SH3 domains[20,21], is a direct consequence of the mechanism of proline recognition. The PPII helix and the grooves used to recognize it have an approximate two-fold rotational symmetry, both in shape and hydrogen bonding groups[21]. Thus, the ligand orientation can be reversed, and the same overall complementarity achieved. Like SH3 and WW domains, profilin also can bind proline-rich motifs in two possible orientations.

The discovery of reversible ligand binding to WW domains explains some of the previous difficulty in extracting a clear recognition consensus for WW domains from known ligands. Different binding orientations will require different ligand sequences. For example, flanking residues that are N-terminal to the core X-P motif in ligands that bind in one orientation would be C-terminal in ligands that bind in the opposite orientation. This orientational flexibility at first confounded attempts to decipher SH3 domain recognition motifs, as well. It is now firmly established, however, that for most SH3 domains there are two distinct ligand consensus motifs — K/R-X-X-P-X-X-P and X-P-X-X-P-X-K/R (where K or R is a required flanking residue) — each of which binds in a different orientation. The new structures now clearly establish this dual orientation framework for WW

domain binding. Armed with the knowledge that each WW domain is likely to have two classes of ligands, researchers will hopefully be able to re-evaluate and clarify much of the apparent complexity observed in peptide library screens.

Why have WW and other evolutionarily unrelated proline recognition domains converged upon this similar mechanism of proline recognition? Recognition of X-P pairs by aromatic grooves presents a highly specific way to recognize proline (Fig. 1*d*). Unlike all other amino acids, which, with the exception of glycine, have substituents at the Cα position, proline is alkylated at both the Cα and the amide-nitrogen. This means that the X-P dipeptide unit has the unique backbone substitution pattern of a C-substituted residue followed by an N-substituted residue. The C/N substituted pair is required because, in this arrangement, substituent groups are separated by only a single backbone carbon atom, forming a relatively continuous ridge that can pack efficiently into the aromatic grooves[18,19]. In this manner, the requirement for N-substitution very effectively excludes all amino acids other than proline, since proline is the only natural N-substituted residue. Because this mechanism focuses only on one unique property of proline — N-substitution — and not the entire proline ring, it allows recognition to be highly selective but of low affinity[22]. This is consistent with the micromolar affinities of isolated SH3 and WW domains for their natural ligands[4,22]. This type of binding mechanism may facilitate signaling interactions which must recognize ligands with high enough selectivity to maintain proper information flow, but with low enough affinity to allow sensitive and dynamic modulation in response to changing signals.

Sequence specificity

Since all WW domains share a core X-P binding groove, how do individual domains achieve specific recognition? Once again the new structures[6,7] reveal two general mechanisms of specific recognition: use of variable loops and neighboring domains (Fig. 2).

The WW domain fold has two variable loops that lie on the same surface as the aromatic X-P binding groove (Fig. 2*a*). We refer to these loops as loop I (between strands β1 and β2) and loop II (between strands β2 and β3) (Fig. 2*b*). Loop I directly abuts the side of the X-P binding groove. Indeed, in the '+' orientation Pin1 complex, it is the loop I residues Ser 16 and Arg 17 that directly contact the pSer side chain. In '-' orientation complexes loop I would abut the required Pro, and thus could not contribute to specificity. Thus it is predicted that only WW domains that have an arginine in the loop I sequence will be able to bind sequences with the core motif pS-P or pT-P, and these motifs will all bind in the '+' orientation. Loop II lies at the opposite end of the surface, and forms a predominantly hydrophobic pocket that is responsible for recognition of the Tyr residue within the P-P-X-Y motif bound by the dystrophin WW domain. This mechanism of specificity is also conceptually similar to that used by SH3 domains. SH3 domains have, flanking their core X-P binding grooves, two variable loops (the n-Src and RT loops) that play a key role in specificity[23].

The neighboring domains with which a WW domain is presented are also critical for specificity (Fig. 2*a*). The dystrophin WW domain alone cannot bind the dystroglycan ligand — the WW domain must be paired with the adjacent helical EF hand-like domain. The structure of the complex reveals that the two domains actually form a composite recognition surface; approximately half of the dystroglycan peptide ligand contacts only the EF-domain. Similarly, in the Pin1 complex, the peptide ligand makes significant contacts with the adjacent PPIase domain. In retrospect, the importance of higher order context in sequence specificity may also explain some of the difficulty in identifying a consensus WW binding motif, since much of the work done on specificity was performed using WW domains in the context of a whole or larger protein. Given their small size and the nature in which they are utilized in these two structures, it is possible that the primary function of WW domains is to act as an auxiliary recognition motif in tandem with other domains.

Rules of WW domain recognition

The two new studies[6,7] presented in this issue, in combination with earlier biochemical studies, allow the formulation of a basic set of rules unifying WW domain recognition. At the core of the WW domain is a highly conserved aromatic groove, which binds the motif X-P in a manner reminiscent of SH3 and other proline recognition domains. This dipeptide can bind the WW domain in two possible N- to C-terminal orientations. The orientation of binding and the specific sequence requirements flanking the X-P motif are determined by interactions with both variable loops and neighboring domains. These studies highlight the importance of solving multiple structures within a protein family and provide a unifying framework for identifying and characterizing new WW domain interactions.

Ali Zarrinpar and Wendell A. Lim are in the Department of Cellular and Molecular Pharmacology, Department of Biochemistry and Biophysics, and Program in Biological Sciences, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0450, USA. Correspondence should be addressed to W.A.L. email: wlim@itsa.ucsf.edu

1. Lee, C.H., Cowburn, D. & Kuriyan, J. *Methods Mol. Biol.* **84**, 3–31 (1998).
2. Sudol, M. *Oncogene* **17**, 1469–1474 (1998).
3. Pawson, T. & Nash, P. *Genes Dev.* **14**, 1027–1047 (2000).
4. Kay, B.K., Williamson, M.P. & Sudol, M. *FASEB J.* **14**, 231–241 (2000).
5. Rubin, G.M. *et al. Science* **287**, 2204–2215 (2000).
6. Huang, X. *et al. Nature Struct. Biol.* **7**, 634–638 (2000).
7. Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T. & Noel, J.P. *Nature Struct. Biol.* **7**, 639–643 (2000).
8. Macias, M.J. *et al. Nature* **382**, 646–649 (1996).
9. Sudol, M. *Prog. Biophys. Mol. Biol.* **65**, 113–132 (1996).
10. Chen, H.I. & Sudol, M. *Proc. Natl. Acad. Sci. USA* **92**, 7819–7823 (1995).
11. Rentschler, S. *et al. Biol. Chem.* **380**, 431–442 (1999).
12. Bedford, M.T., Sarbassova, D., Xu, J., Leder, P. & Yaffe, M.B. *J. Biol. Chem.* **275**, 10359–10369 (2000).
13. Roberts, R.G., Gardner, R.J. & Bobrow, M. *Hum. Mutat.* **4**, 1–11 (1994).
14. Lu, P.J., Zhou, X.Z., Shen, M. & Lu, K.P. *Science* **283**, 1325–1328 (1999).
15. Ho, C.K. & Shuman, S. *Mol. Cell* **3**, 405–411 (1999).
16. Steinmetz, E.J. *Cell* **89**, 491–494 (1997).
17. Creamer, T.P. *Proteins* **33**, 218–226 (1998).
18. Lim, W.A. & Richards, F.M. *Nature Struct. Biol.* **1**, 221–225 (1994).
19. Yu, H. *et al. Cell* **76**, 933–945 (1994).
20. Feng, S., Chen, J.K., Yu, H., Simon, J.A. & Schreiber, S.L. *Science* **266**, 1241–1247 (1994).
21. Lim, W.A., Richards, F.M. & Fox, R.O. *Nature* **372**, 375–379 (1994). [published erratum appears in *Nature* **374**, 94 (1995)].
22. Nguyen, J.T., Turck, C.W., Cohen, F.E., Zuckermann, R.N. & Lim, W.A. *Science* **282**, 2088–2092 (1998).
23. Musacchio, A., Wilmanns, M. & Saraste, M. *Prog. Biophys. Mol. Biol.* **61**, 283–297 (1994).