Vol 437|22 September 2005|doi:10.1038/nature03990

Natural-like function in artificial WW domains

William P. Russ[1], Drew M. Lowery[2], Prashant Mishra[1], Michael B. Yaffe[2] & Rama Ranganathan[1]

Protein sequences evolve through random mutagenesis with selection for optimal fitness[1]. Cooperative folding into a stable tertiary structure is one aspect of fitness, but evolutionary selection ultimately operates on function, not on structure. In the accompanying paper[2], we proposed a model for the evolutionary constraint on a small protein interaction module (the WW domain) through application of the SCA, a statistical analysis of multiple sequence alignments[3,4]. Construction of artificial protein sequences directed only by the SCA showed that the information extracted by this analysis is sufficient to engineer the WW fold at atomic resolution. Here, we demonstrate that these artificial WW sequences function like their natural counterparts, showing class-specific recognition of proline-containing target peptides[5–8]. Consistent with SCA predictions, a distributed network of residues mediates functional specificity in WW domains. The ability to recapitulate natural-like function in designed sequences shows that a relatively small quantity of sequence information is sufficient to specify the global energetics of amino acid interactions.

The basic premise of the SCA is that in accord with the cooperative nature of amino acid interactions in determining protein stability and function, the evolutionary constraint on proteins should be a distributed (rather than intrinsic) property of amino acid positions. That is, the conservation of amino acids at one site should be the result of constraints on that site taken independently, and of the constraints arising through energetic interactions with other positions. For example, consider the SCA for an alignment of 292 members of the WW domain family (Fig. 1). This small protein interaction module adopts a curved three-stranded β-sheet structure with a binding site for proline-containing peptides formed on the concave surface of the sheet (Fig. 1a). The binding surface includes an X-Pro binding site (positions 19 and 30, shown in blue as CPK representation; Fig. 1a) that recognizes the canonical proline in target peptides, and a specificity site formed by residues in β2 and the β2–β3 loop (positions 21, 23 and 26, shown in yellow as CPK representation; Fig. 1a)[9]. WW domains are classified into four groups based on target peptide sequence motifs: group I, PPxY[6]; group II, PPLP[7]; group III, PPR[5]; and group IV, pS/pT-P[8], where x stands for any amino acid. The SCA output for the WW family is a matrix of coupling values organized such that columns represent positions in the WW alignment, and rows represent sites where perturbations are

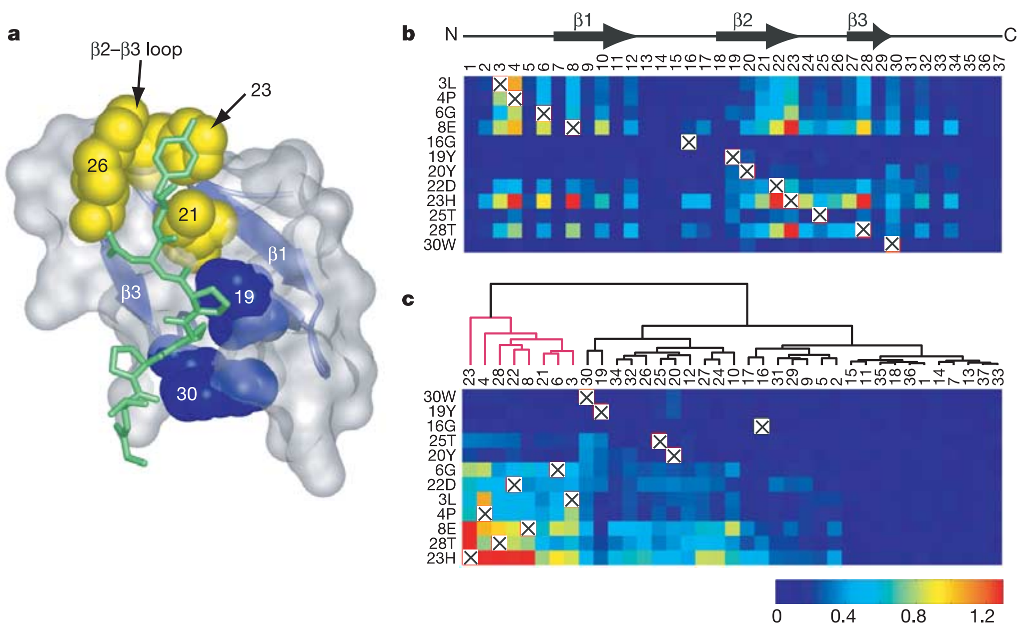

Figure 1 | Statistical coupling analysis for the WW domain, a proline-binding protein interaction module. a, The peptide-binding surface of the WW domain contains two structurally defined pockets: the X-Pro binding site (in blue) and a specificity site (in yellow). Shown is the Nedd4.3 WW domain (Protein Data Bank 1I5H) bound to a group I peptide (in green)[30]. **b**, A matrix of coevolution scores between all WW positions (columns) and 12 moderately conserved positions (rows) in an alignment of 292 WW domains. The colour scale ranges from blue (no coevolution) to red (maximal coevolution). Checked boxes mark the trivial self-correlation of sites in the SCA method. **c**, Hierarchical clustering shows that a single group of eight positions (marked in red) share a pattern of strong mutual coevolution.

[1]Howard Hughes Medical Institute and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050, USA. [2]Center for Cancer Research, Department of Biology and Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

introduced to interrogate evolutionary coupling[4] (Fig. 1b). Thus, each pixel shows the coevolution score for one pair of WW sites. Hierarchical clustering of this matrix reveals a remarkably simple global organization of conserved evolutionary interactions between amino acids. WW positions fall into two main clusters, one that shows minimal coupling to other sites, and one that comprises eight positions (marked in red) related by strong mutual coevolution (Fig. 1c).

Is the amino acid composition at sites plus the information in the SCA matrix all the sequence information required to specify the WW domain? The accompanying paper provides the first phase of this experimental test by showing that artificial sequences designed using only these parameters adopt a stable WW-like tertiary structure[2]. However, a key further test is the sufficiency of this information for specifying function. If the SCA matrix captures the fitness constraints on the WW family, then the artificial sequences should show class-specific recognition of proline-containing sequences and binding affinities like those of natural WW domains.

We developed an oriented peptide library binding assay for measuring WW domain specificity, and studied a set of natural and artificial sequences. Four biotinylated degenerate peptide libraries were constructed, each oriented around one group-specific WW recognition motif, and binding was detected using an enzyme-linked immunosorbent assay (ELISA) (Fig. 2a, b; see also Supplementary Fig. 1a). For example, the group-I-oriented peptide library was biotin-Z-GMAxxxPPxYxxxAKKK, where Z is 6-amino-hexanoic acid and x stands for any amino acid except cysteine (theoretical degeneracy of 8.9×10^8 sequences). A fifth proline-oriented library was also made as a control for nonspecific binding. For a group I domain (Nedd4.3, or N39 by the numbering system in the accompanying paper) the peptide library assay confirms specific interaction with the PPxY-containing sequences[10] (Fig. 2a). Interestingly, CC43, an artificial WW domain created through SCA-based protein design[2], also specifically interacts with the PPxY peptide library (Fig. 2b), suggesting that CC43 functions as a group I WW domain. To confirm these results independently, we carried out phage-display and fluorescence-based quantitative binding assays for N39 and CC43. Consistent with the peptide library screens, both N39 and CC43 selectively isolate PPxY-containing sequences from a random 12-mer phage display library (Fig. 2c, d). In addition, both of these domains show similar binding affinities to a model group I peptide (N39, dissociation constant (K_d) = 11.2 ± 1.2 μM; CC43, K_d = 1.7 ± 0.1 μM; mean ±s.d.; Fig. 2e). Taken together, these data validate the oriented peptide library assay for classification of WW domain specificity, and demonstrate that one artificial WW domain, CC43, displays group I-specific binding.

Figure 3 summarizes the results of peptide library screening for 27 randomly chosen natural and ten natively folded artificial WW domains[2]. The Pin1 WW domain was added to the list of natural domains tested because it is the best-characterized member of the group IV specificity class[11]. The data are shown in clustered matrix format, with each row showing the binding for one domain to all five oriented peptide libraries normalized between the minimum (white) and the maximum (black) observed signal to allow comparison of specificity profiles. Group I WW domains fell into two distinct subclusters (I and Im (marginal group I)) that significantly differed in the degree of specificity for the PPxY motif library (Supplementary Fig. 1). Group III domains bound to the PPR-oriented peptide library as expected but also showed binding to the PPxY library and weak binding to PPLP and proline-alone libraries. The binding to the PPxY library may simply reflect the fact that this library contains a non-trivial fraction of PPRY sequences. Nevertheless, the overall profiles show that group III domains display far more relaxed specificity than group I domains (Supplementary Fig. 1). A similar ambiguity in the binding specificity of group III domains has been reported previously[12]. Finally, a minor fraction of the total natural sequences tested (2 out of 28) bound to pSP- and PPLP-oriented libraries, and were scored as group IV and group II, respectively.

Artificial (CC) WW domains are functionally indistinguishable from natural WW sequences (Fig. 3). Of the ten artificial domains studied, six displayed group I specificity (one is Im) and four showed group III specificity, numbers that approximately reflect the distribution of these sequences in our MSA and in prior samplings of WW specificity[13,14]. Quantitative binding isotherms measured for all group I artificial sequences show a range of binding affinities typical for natural WW domains (Fig. 3, rightmost column). To probe further the similarity between natural and designed WW sequences, we carried out a saturation mutagenesis study of the group I and group III peptide ligands. In this experiment, every position of the target peptide is replaced individually to each of the 20 amino acids, and the effect on binding is measured by detecting protein hybridization to the array of peptide variants[14]. Supplementary Fig. 2 shows these data for two natural WW domains (N39, group I (ref. 10), and N31, group III (refs 5, 14)), and two artificial domains (CC43, group I, and CC20, group III), demonstrating that the designed sequences show a pattern of sensitivity to ligand mutagenesis similar to that for natural sequences. Thus, the amino acid composition at sites plus the information contained in the SCA matrix encodes the sequence rules for quantitatively specifying function in the WW domain.

No artificial sequences with group II or group IV specificity were identified. However, the paucity of group II and group IV sequences

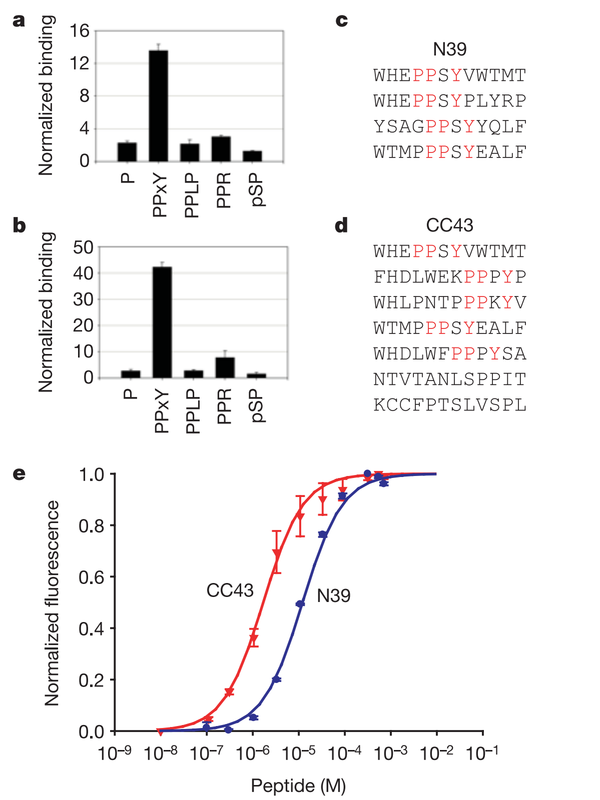

Figure 2 | Assays for binding affinity and specificity in WW domains. **a**, **b**, Oriented peptide library screening (see Methods) of binding specificity for N39, a natural domain with group I (PPxY) specificity (**a**), or CC43, an artificial WW domain showing 37% average and 68% top-hit identity to natural WW domains in the MSA (**b**). Binding is reported as fold above background in the absence of target peptides. **c**, **d**, Phage display analysis of binding specificity for N39 and CC43 WW domains; both domains select PPxY-containing sequences from a random 12-mer peptide library. **e**, Binding isotherms for N39 (K_d = 11.2 ± 1.2 μM) and CC43 (K_d = 1.7 ± 0.1 μM), assayed by Trp fluorescence quenching using a model group I peptide (EYPPYPPPPYPSG). Error bars represent standard deviation from at least four independent assays.

in our alignment of natural WW domains and the relatively small number of artificial domains tested so far probably account for the lack of these specificity classes. These classes may emerge from larger-scale design of artificial WW sequences.

Given these results, can we infer which positions constitute the sequence determinants for WW binding specificity? The eight positions comprising the primary cluster of coevolving residues in the WW family (Fig. 1c) are highlighted in yellow in Fig. 3. Sequences experimentally found to display group I binding specificity, whether natural or designed, strictly conserve a specific sequence motif in the coevolving cluster (^{3}L^{4}P^{6}G^{8}E^{21}I/V^{22}D/N^{23}H^{28}T). In contrast, the average sequence identity in non-cluster positions is barely different for group I sequences in comparison with all WW sequences: 40.9 ± 2.9% within group I sequences and 35 ± 3.9% overall (mean ± s.d.). These results strongly suggest that this network of mutually evolving residues is the major determinant of group I binding specificity. In accord with this conclusion, marginal group I domains or the weakly specific group III domains display considerable variation within this cluster. A larger study of designed sequences may help to distil the minimal sequence profiles of these and other WW groups.

The spatial organization of the network residues provides an unexpectedly distributed picture of binding specificity in the WW domain. Rather than being restricted to the ligand-binding surface, the eight network residues are organized into a physically contiguous network linking the primary specificity determining pocket (positions 21, 23 and 28) with residues on the opposite side (3 and 4) through a few intervening residues (6, 8 and 22) (Fig. 4a). The coevolution of these positions predicts that some residues act at long range in mediating peptide binding, and the network amino acids act cooperatively in determining the binding free energy. Previous mutagenesis studies already suggest that network residues at or close to the peptide binding surface (8, 21, 23 and 28) mediate binding specificity[15–18], and structural work in the dystrophin WW domain provides a mechanistic understanding for the contribution of these residues in group I domains[19]. To test more of the network for contribution to peptide binding, and to test the prediction of cooperative action, we carried out thermodynamic double mutant cycle analysis[20,21] to measure the energetic coupling between mutations at binding-site position 28 and positions 3, 8 and 23 in the Nedd4.3 WW domain. In the mutant cycle method, the effect of one mutation on the equilibrium dissociation constant for peptide binding is measured in two conditions: (1) the wild-type background ($X1 = K_d^{M1}/K_d^{WT}$), and (2) in the background of a second mutation ($X2 = K_d^{M1,M2}/K_d^{M2}$) (Fig. 4b). The ratio of these effects gives the coupling parameter Ω, a measure of the degree of interaction between the two mutations. If the two mutations are thermodynamically independent, the effect of the first mutation is the same in conditions 1 and 2, and $\Omega = 1$. If $\Omega \neq 1$, then the two mutations act cooperatively.

Consistent with earlier studies[17,19], the E8A, H23A and T28A mutations all affected binding of a PPxY-containing peptide (Fig. 4c). However, L3A also had a significant effect (5.15 ± 0.99 fold; mean ± s.d.), although located on the opposite surface from the peptide-binding site. In addition, mutant cycle analyses for the T28A mutation with each of the three other mutations show Ω values

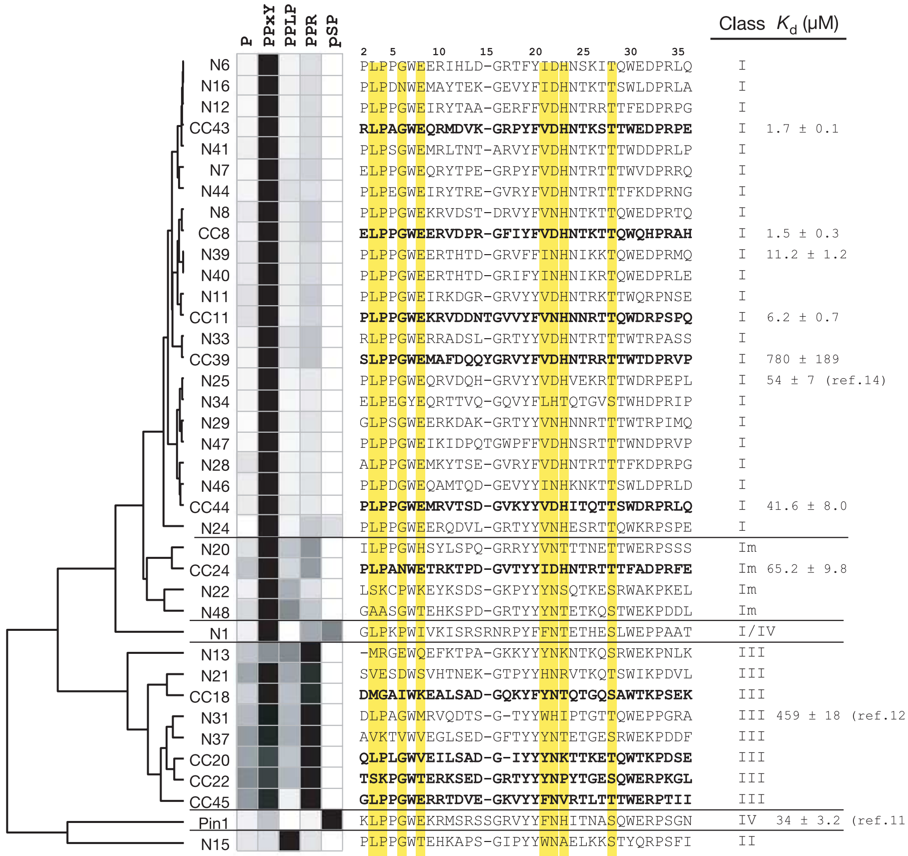

Figure 3 | A summary of functional measurements for all tested natural and artificial WW domains. The matrix shows the results of oriented peptide library assays, clustered to reveal the similarities in binding specificity for different domains. The colour scale ranges from white (minimum signal) to black (maximum signal) for each domain. Class assignments for domains are shown at the right. Dissociation constants for selected group I domains were either measured by fluorescence binding assay (N39 and all artificial sequences) or were derived from literature references. Sequences are aligned per the MSA, and positions corresponding to the cluster of eight coevolving residues (Fig. 1c) are highlighted in yellow. SCA-based designed sequences are shown in bold. Measurements are mean ± s.d.

that significantly differ from unity (Fig. 4c; see also Supplementary Fig. S3). Specifically, the effects of mutations at 3, 8 and 23 are either diminished (L3A and H23A) or abrogated (E8A) in the background of T28A (Supplementary Fig. 3). Thus T28A is thermodynamically coupled to mutations at 3, 8 and 23. These results support the model that a distributed and cooperative network of residues predicted by the SCA contributes to peptide recognition in the WW domain.

It is perhaps surprising that all folded artificial WW domains could be classified into a known functional group. The SCA emphasizes the deeply conserved couplings between sequence positions in a protein family while down-weighting or even ignoring less conserved couplings. These weak couplings typically arise from small clades of more recently diverged sequences in the MSA, which are expected to contain many positional correlations yet to be relaxed through variation. If these recent branchings of the phylogenetic tree also hold important information about the physical chemistry of specific binding in extant proteins, we might have expected many folded but functionally undifferentiated WW domains in SCA-based design. The data here suggest that at least as defined by *in vitro* assays, information specifying binding specificity is captured in positional interactions that are in the deep evolutionary record. Future studies of functional complementation *in vivo* will help further to address the completeness of the SCA-based sequence information in specifying natural-like proteins.

Compared with the current field of protein design, the SCA-based protein design takes a completely different but complementary approach to understanding the design of natural proteins. Using atomistic energy functions that approximate the physical forces between atoms, several groups have now achieved spectacular successes in the re-design of natural folds[22–24], including WW domains[25], and even in the *de novo* construction of artificial folds[26,27]. These successes demonstrate the accuracy of the scoring functions used, but important unsolved issues remain. The basic design principle in atomistic protein design is optimization of a target potential function that produces sequences with deep thermodynamic minima in the native state and structures with high thermal stabilities[22,27]. However, natural proteins are thought to have native states with shallow energy minima, resulting in marginally stable but dynamic folds that are often capable of supporting more than one conformation. In a purely statistical and mechanism-free way, SCA-based protein design produces sequences that show the same marginal stability of natural proteins and function like natural proteins. It may be interesting to combine the mechanistic description of energetics from atomistic design with the sparse and distributed architecture of amino acid interactions in SCA to understand better the design of evolved proteins.

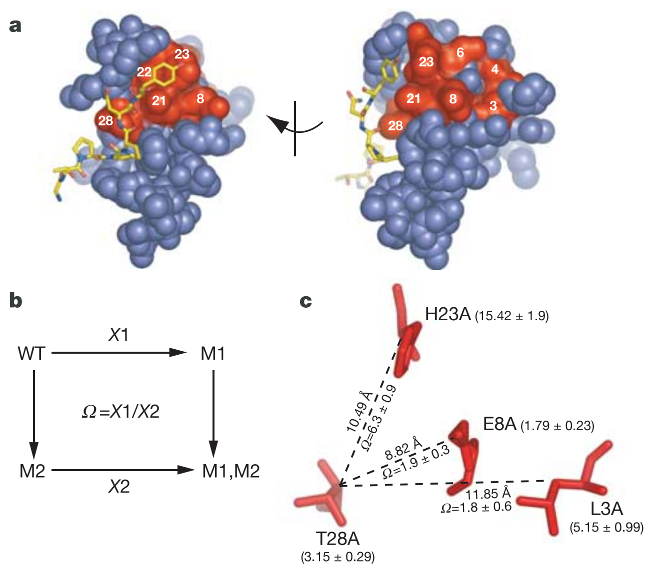

Figure 4 | A spatially distributed network underlying WW function. a, The cluster of eight coevolving residues (Fig. 1c) mapped on the Nedd4.3 WW domain structure (in red). The cluster forms a connected network that links binding site residues with the opposite side through a few intervening positions. **b**, The thermodynamic mutant cycle formalism. The fold effect of one mutation (M1) on an equilibrium constant is calculated in the wild-type background ($X1$) or in the background of the second mutation ($X2$). The coupling parameter Ω is the ratio of these fold effects ($X1/X2$); that is, the degree to which the effect of one mutation depends on the second. **c**, Mutant cycle analysis of selected coevolving positions. Residues are shown in the same orientation as in **a** (right), with distances between β-carbons of residues indicated. Single mutations at all sites affect peptide binding (fold effect relative to wild type in parentheses), and mutant cycle analysis demonstrates energetic coupling ($\Omega > 1$) between position 28 and positions 3, 8 and 23. Measurements are mean ± s.d.

METHODS

Statistical coupling analysis. SCA was conducted as previously described[3,4] on an alignment of 292 WW domain sequences, updated from the original alignment of 120 sequences in the accompanying paper using the March 2004 release of the non-redundant database. SCA results are essentially identical for the two alignments. Sequences were collected using PSI-BLAST (*e*-scores <0.001) and aligned using ClustalW[28] followed by manual adjustment. The alignment is available from our laboratory website (http://www.hhmi.swmed.edu/Labs/rr), and the code is available upon request.

Oriented peptide library assay. Five biotinylated degenerate peptide libraries were synthesized using N-α-FMOC protected amino acids and standard BOP/HOBt coupling chemistry. The libraries were constructed to present either a proline-only control (biotin-Z-GMAxxxxPxxxxAKKK) or the four different characteristic WW domain binding motifs: group-I-oriented (biotin-Z-GMAxxxPPxYxxxAKKK-C), group-II-oriented (biotin-Z-GMAxxxPPLPxxx-AKKK), group-III-oriented (biotin-Z-GMAxxxPPRxxxAKKK) and group-IV-oriented (biotin-Z-GMAxxxxpSPxxxxAKKK), where Z is 6-aminohexanoic acid, pS is phosphoserine, and x denotes all amino acids except cysteine. Peptide libraries were immobilized onto pre-washed streptavidin-coated 96-well plates (10 μg per well) in phosphate-buffered saline plus 0.5% Tween-20 (PBST) at 4 °C for 1 h. Wells were washed in PBST and incubated with GST–WW domains (0.5 μg) for 2 h, washed, and detected by ELISA using a horseradish peroxidase-conjugated anti-GST antibody (Amersham) at 1:5,000 dilution for 1 h at 4 °C followed by reaction with 3,3′,5,5′-tetramethylbenzidine liquid substrate system (Sigma) for 5 min. Absorbance was monitored at 450 nm.

Phage display. Phage display was carried out using a commercial random 12-mer library (PhD Phage Display kit, NEB) per protocols supplied by the manufacturer. Phage isolates were sequenced after three rounds of amplification and nonspecific elution with glycine-HCl, pH 2.2.

Binding assays. Tryptophan fluorescence-based peptide binding assays were conducted on a PTI spectrofluorimeter, monitoring fluorescence emission at 340 nm (excitation at 295 nm). Group I peptide, EYPPYPPPPYPSG (Tufts Protein Chemistry Facility), was purified using reverse-phase high-performance liquid chromatography. Binding assays were conducted at 4 °C in buffer A (100 mM TrisHCL, pH 8.0, 100 mM NaCl) with 1 μM WW domain. WW domains were cloned into the pHIS8-3 vector, expressed and purified as described[2]. The assay follows the fraction of protein bound to peptide by the normalized fluorescence quenching of the Trp residue in the X-Pro binding pocket. Isothermal titration calorimetry measurements were made using a MicroCal VP-ITC calorimeter in buffer A at 4 °C, starting with 50–200 μM WW domain in the sample cell and titrating 0.5–3 mM of the group I peptide. Data were fit using MicroCal Origin software supplied by the manufacturer, using a one-site binding model. The marginal stability of WW domains required correction of apparent dissociation constants to account for the fraction of folded protein at the assay temperature. Assuming a two-state folding reaction, the correction applied was $K_{\mathrm{d}} = K_{\mathrm{d,app}}\left(\frac{K_{\mathrm{f}}}{1+K_{\mathrm{f}}}\right)$, where K_{f} is the equilibrium constant for the folding reaction. For each protein assayed, K_{f} was calculated using thermal denaturation studies carried out as in the accompanying paper[2], and assuming a previously reported value for the change in heat capacity for the unfolding reaction[29]. K_{f} values were: wild type, 38.8; L3A, 1.65; E8A, 15.7; H23A, 22.7; T28A, 32.0; L3A/T28A, 1.25; E8A/T28A, 8.54; H23A/T28A, 21.5.

Protein hybridization assay. Protein hybridization assays were performed as described[14]. Arrays of all single amino acid substitutions for a group I peptide (GTPPPPYTVG) and a group III peptide (PPGPPPRGPPP) were synthesized on membranes by the M.I.T. Biopolymers Laboratory. Membranes were washed in PBS plus 0.1% Tween-20 (PBST), blocked for 2 h at room temperature in PBST plus 5% non-fat dry milk, washed, and incubated at 4 °C overnight with 10–400 $\mu g\,ml^{-1}$ of purified GST–WW domains. Membranes were washed, treated with horseradish peroxidase-conjugated anti-GST antibodies (Sigma) in PBST plus 5% non-fat dry milk at 4 °C for 2 h, washed again, and bound WW domains were detected using the ECL kit (Amersham).

Received 25 January; accepted 30 June 2005.

1. Voigt, C. A., Kauffman, S. & Wang, Z. G. Rational evolutionary design: the theory of *in vitro* protein evolution. *Adv. Protein Chem.* **55**, 79–160 (2000).
2. Socolich, M. *et al.* Evolutionary information for specifying a protein fold. *Nature* doi:10.1038/nature03991 (this issue).
3. Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
4. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69 (2003).
5. Bedford, M. T., Sarbassova, D., Xu, J., Leder, P. & Yaffe, M. B. A novel pro-Arg motif recognized by WW domains. *J. Biol. Chem.* **275**, 10359–10369 (2000).
6. Chen, H. I. & Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl Acad. Sci. USA* **92**, 7819–7823 (1995).
7. Ermekova, K. S. *et al.* The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of *Drosophila* enabled. *J. Biol. Chem.* **272**, 32869–32877 (1997).
8. Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* **283**, 1325–1328 (1999).
9. Zarrinpar, A. & Lim, W. A. Converging on proline: the mechanism of WW domain peptide recognition. *Nature Struct. Biol.* **7**, 611–613 (2000).
10. Kanelis, V., Rotin, D. & Forman-Kay, J. D. Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nature Struct. Biol.* **8**, 407–412 (2001).
11. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature Struct. Biol.* **7**, 639–643 (2000).
12. Kato, Y. *et al.* Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J. Biol. Chem.* **279**, 31833–31841 (2004).
13. Hu, H. *et al.* A map of WW domain family interactions. *Proteomics* **4**, 643–655 (2004).
14. Otte, L. *et al.* WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci.* **12**, 491–500 (2003).
15. Chen, H. I. *et al.* Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* **272**, 17070–17077 (1997).
16. Espanel, X. & Sudol, M. A single point mutation in a group I WW domain shifts its specificity to that of group II WW domains. *J. Biol. Chem.* **274**, 17284–17289 (1999).
17. Kasanov, J., Pirozzi, G., Uveges, A. J. & Kay, B. K. Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem. Biol.* **8**, 231–241 (2001).
18. Toepert, F., Pires, J. R., Landgraf, C., Oschkinat, H. & Schneider-Mergener, J. Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angew. Chem. Int. Edn Engl.* **40**, 897–900 (2001).
19. Huang, X. *et al.* Structure of a WW domain containing fragment of dystrophin in complex with β-dystroglycan. *Nature Struct. Biol.* **7**, 634–638 (2000).
20. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835–840 (1984).
21. Hidalgo, P. & MacKinnon, R. Revealing the architecture of a K^+ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**, 307–310 (1995).
22. Dahiyat, B. I. & Mayo, S. L. *De novo* protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
23. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304**, 1967–1971 (2004).
24. Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nature Struct. Mol. Biol.* **11**, 371–379 (2004).
25. Kraemer-Pecore, C. M., Lecomte, J. T. & Desjarlais, J. R. A *de novo* redesign of the WW domain. *Protein Sci.* **12**, 2194–2205 (2003).
26. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
27. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
28. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
29. Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H. & Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl Acad. Sci. USA* **98**, 13002–13007 (2001).
30. Delano, W. L. *The PyMOL Molecular Graphics System* ⟨http://www.pymol.org⟩ (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Ranganathan laboratory for advice and critical review of the manuscript, J. P. Noel for providing the pHIS8 expression vector, and K. Voegler for contributing to this project. This study was supported by the Robert A. Welch foundation (R.R.), the Mallinckrodt Foundation Scholar Award (R.R.), NIH grants (M.B.Y.), and a Burroughs-Wellcome Career Development Award (M.B.Y.). D.M.L. was supported by a Howard Hughes Medical Institute pre-doctoral award. W.P.R. is an associate and R.R. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.R. (rama.ranganathan@utsouthwestern.edu).