

# Compensatory Evolution of a WW Domain Variant Lacking the Strictly Conserved Trp Residue

Hayato Yanagida · Tomoaki Matsuura ·  
Tetsuya Yomo

Received: 5 September 2007 / Accepted: 19 November 2007 / Published online: 18 December 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Replacement of conserved amino acid residues during evolution of proteins can lead to divergence and the formation of new families with novel functions, but is often deleterious to both protein structure and function. Using the WW domain, we experimentally examined whether and to what degree second-site mutations can compensate for the reduction of function and loss of structure that accompany substitution of a strictly conserved amino acid residue. The W17F mutant of the WW domain, with substitution of the most strictly conserved Trp residue, is known to lack a specific three-dimensional structure and shows reduced binding affinity in comparison to the wild type. To obtain second-site revertants, we performed a selection experiment based on the proline-rich peptide (PY ligand) binding affinity using the W17F mutant as the initial sequence. After selection by ribosome display, we were able to select revertants that exhibited a maximum ninefold higher affinity to the PY ligand than the W17F mutant and showed an even better affinity than the wild type. In addition, we

found that the functional restoration resulted in increased binding specificity in selected revertants, and the structures were more compact, with increased amounts of secondary structure, in comparison to the W17F mutant. Our results suggest that the defective structure and function of the proteins caused by mutations in highly conserved residues occurring through divergent evolution not only can be restored but can be further improved by compensatory mutations.

**Keywords** WW domain · Conserved residue · Divergent evolution · Directed evolution · Compensatory mutation · Second-site revertant · Unfolded protein · Ribosome display

## Introduction

Protein families consist of proteins with similar biochemical function and three-dimensional structure. The primary sequences of proteins in a family contain conserved amino acid residues, which usually serve important roles in determining the protein function and structure. However, during divergent evolution, conserved residues are not invariable, and may be replaced to create a new function. As a result, when subfamilies within a protein family are compared, some amino acid residues are found to be replaced in moving from one subfamily to another, while remaining invariant within each subfamily. Such amino acid residues specifically conserved within a subfamily are known to be involved in determining the functional specificity of the subfamily (Innis et al. 2000; Pritchard and Dufton 1999; Sowa et al. 2001), such as substrate specificity. Thus, replacement of conserved residues may lead to diversification and the production of a family with a new

---

H. Yanagida · T. Yomo  
Graduate School of Frontier Biosciences, Osaka University,  
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

T. Matsuura · T. Yomo (✉)  
Department of Bioinformatics Engineering, Graduate School of  
Information Science and Technology, Osaka University, Osaka,  
Japan  
e-mail: yomo@ist.osaka-u.ac.jp

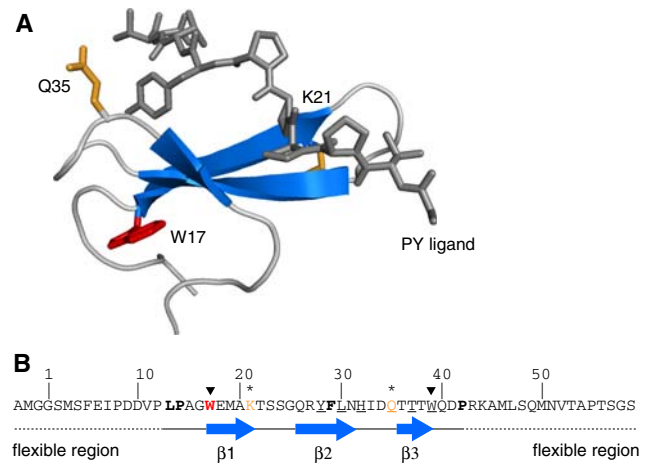
T. Matsuura  
Precursory Research for Embryonic Science and Technology  
(PRESTO), Japan Science and Technology Agency (JST),  
Tokyo, Japan

T. Yomo  
Exploratory Research for Advanced Technology (ERATO),  
Japan Science and Technology Agency (JST), Tokyo, Japan

function. On the other hand, mutations of conserved residues generally have deleterious effects on the protein's function and/or stability (Guo et al. 2004). These observations indicate that mutations in the conserved residues may alter the functional specificity or eliminate the mutated proteins from the evolutionary tree.

During molecular evolution, compensatory mutations are known to suppress deleterious mutational effects in functional and/or structural properties of proteins (Kimura 1991; Kondrashov et al. 2002). This compensatory mechanism has been investigated empirically. For example, the complete loss of lysozyme activity in a defective T4 lysozyme mutant with substitution of the Trp residue in the hydrophobic core could be restored through other compensatory mutations (Jucovic and Poteete 1998), and similar observations were made with structural destabilization of p53 protein mutants (Nikolova et al. 2000). These results indicate that there may be compensatory mutations that can neutralize the deleterious effects caused by replacement of highly conserved residues. To experimentally investigate the presence of such compensatory mutations, we focused on the most strictly conserved residue within the WW domain family, and examined whether and to what degree defective function and structure caused by replacement of this residue can be compensated by second-site mutations. This is important to understand the possible path taken in the evolution of proteins when a conserved residue is replaced.

The WW domain is one of the smallest protein modules (30–40 residues), consisting of a three-stranded antiparallel  $\beta$ -sheet structure, which mediates protein–protein interaction by recognizing proline-rich peptide sequences (Bork and Sudol 1994; H.I. Chen and Sudol 1995; Macias et al. 1996). The WW domain carries two highly conserved Trp residues separated by a stretch of 20–22 amino acids, for which it is named. Especially, the N-terminal Trp is almost perfectly conserved among more than 2500 WW domains determined to date, as revealed by analysis using the SMART server (Schultz et al. 1998). A previous study on the human Yes-associated protein (hYAP) WW domain demonstrated the functional and structural importance of the two highly conserved Trp residues (Koepl et al. 1999a; Macias et al. 1996) and indicated that the N-terminal Trp17 is involved in the hydrophobic patch and is essential for maintaining the overall  $\beta$ -sheet structure (Fig. 1). Indeed, replacement of Trp17 with Phe resulted in a significant loss of structural integrity and reduced binding affinity ( $K_d = 15.1 \mu\text{M}$ ) to the PY ligand (EYPPYPPPP-YPSG) compared to the wild type ( $K_d = 5.9 \mu\text{M}$ ) (Koepl et al. 1999a). Thus, the W17F mutant of the hYAP WW domain lacks the strictly conserved residue and the overall structural integrity, and exhibits reduced binding function (Koepl et al. 1999a).



**Fig. 1** The primary and three-dimensional structures of the hYAP WW domain. (A) Three-dimensional structure of the wild-type hYAP WW domain (V11–R43) bound to the gPY ligand (GTPPPYTVG) (Macias et al. 1996). Side chains of Trp17, Lys21, and Gln35 are depicted. Trp17 is the primary mutated residue highly conserved in WW domains, and Lys21 and Gln35 are the residues substituted most frequently in the enriched clones. The figure was prepared using PyMOL (<http://www.pymol.org>). (B) The sequence of the wild-type WW domain used in this study. Residues that form a hydrophobic patch, important for stabilizing the overall structure of the domain, are shown in boldface, and residues that make direct contact with the PY ligand are underlined. The two highly conserved residues, Trp17 and Trp39, are indicated by filled triangles above the sequence. Lys21 and Gln35, the residues substituted most frequently in the enriched clones, are indicated by stars above the sequence. The regions forming  $\beta$ -sheets ( $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ ) and the flexible regions are indicated below the sequence by arrows and dashed lines, respectively

In this study, we investigated whether and to what extent decreased binding function and structural disorder in the W17F mutant of the hYAP WW domain can be restored with second-site mutations introduced through the evolutionary process. We performed a selection experiment using the W17F mutant as a starting point and selected for variants (second-site revertants) that were able to bind the PY ligand with improved affinity, but preserving the W17F mutation, using ribosome display (Hanes and Plückthun 1997; Matsuura et al. 2007). After four rounds of selection, we obtained a second-site revertant, whose affinity to the PY ligand was 9-fold higher than that of the original W17F mutant and even 3.6-fold higher than that of the wild type. We also evaluated the binding specificity and structural properties of the selected revertants and found that the binding was highly specific for the recognition motif of the PY ligand. Moreover, the selected revertants were found to have a decreased apparent molecular weight and increased secondary structure content of the WW fold. Our results suggest that deleterious effects caused by mutations in highly conserved residues occurring through divergent evolution not only can be restored but can be improved even further by compensatory mutations. In addition, the

functional significance of unstructured proteins, such as the W17F mutant, in protein evolution is discussed.

## Materials and Methods

### Synthesis of PY Ligands

Two types of PY ligand were used for selection and binding experiments: ePY (EYPPYPPPYPSG) (Sudol 1998) and gPY (GTPPPYTVG) (H. I. Chen and Sudol 1995). All peptides were synthesized by Gene Design Inc. (Osaka, Japan) with standard Fmoc protocols, purified by reverse-phase HPLC, and confirmed by MALDI-TOF mass spectrometry. Biotinylation of ePY ligand was performed by labeling the N-terminal amine of the ligand with EZ-link NHS-SS-biotin (Pierce, Rockford, IL, USA) in accordance with the manufacturer's instructions.

### DNA Constructs

Plasmid pRD-WW is a plasmid encoding human Yes-associated protein WW domain (Koepp et al. 1999a) and a spacer sequence for ribosome display format (Matsuura and Plückthun 2003). Plasmid pRD-W17F encoding the W17F mutant of the WW domain was generated by site-directed mutagenesis using the primers W17F+ (5'-TCTGCCAGCAGGTTTCGAAATGGCAA-3') and W17F- (5'-TTTGCCATTTTCGAAACCTGCTGGCAGA-3') with a QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) in accordance with the manufacturer's instructions. To prepare the randomly mutated library of the W17F mutant, the DNA encoding the W17F gene was amplified by error-prone PCR using  $\Delta$ Tth DNA polymerase (Toyobo, Osaka, Japan) and the primers WW-forward (5'-ACGACAAAGCCATGGGTGGT-3') and WW-reverse (5'-ATATAGGATCCACTGGTCGGGGCTGTGACG-3') as described previously (Arakawa et al. 1996). The PCR product was then digested and ligated into the vector pRD-n1n2\_2 (Matsuura and Plückthun 2003) using the *NcoI*/*Bam*HI restriction sites. The ligation products were directly PCR amplified using the primers SDA-pqe (5'-AGACCAC AACGGTTTCCTCTAGAAATAATTTTGTAACTT TAAGAAAGAGGAGAAATTA ACTATGAGA-3') and pqe1- (5'-GATCTATCAACAGGAGTCCAAGCTCA-3'). The products were subsequently amplified again by PCR with the primers T7B (5'-ATACGAAATTAATACGACTACTATAGGGAGACCACAACGG-3') and T3te\_pD- (5'-CGGCCACCCGTGAAGGTGAGCC-3') to introduce the 5' and 3' stem loops and to make the final construct for *in vitro* transcription (Matsuura and Plückthun 2003). PCR was performed using Ex Taq polymerase (TaKaRa, Otsu,

Japan). The PCR products were used directly for *in vitro* transcription using T7 RNA polymerase and purified as described previously (Hanes and Plückthun 1997).

### Ribosome Display

*In vitro* translation was carried out using an *E. coli* S-30 extract system (H. Z. Chen and Zubay 1983) or PURE system: an *E. coli*-based reconstituted *in vitro* translation system (Post Genome Institute, Tokyo) (Shimizu et al. 2001), essentially as described previously (Hanes and Plückthun 1997). Following a 7-min translation reaction at 37°C, the reaction was stopped by fourfold dilution with ice-cold wash buffer (WBKT; 50 mM Tris-acetate, pH 7.5, 150 mM NaCl, 50 mM magnesium acetate, 0.5 M KCl, 0.1% Tween 20) supplemented with 2.5 mg/ml heparin and 0.5% BSA. After centrifugation at 14,000g for 5 min, the supernatant containing ternary complexes of mRNA, encoded protein, and ribosomes was placed on ice and used for subsequent experiments.

Microtiter plates (Maxisorb; Nunc, Roskilde, Denmark) were coated with 100  $\mu$ L of 2  $\mu$ g/ml neutravidin (Pierce) dissolved in phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) at 4°C overnight, and then blocked by incubation with 5% BSA for 2 h at room temperature. After three washes with 300  $\mu$ L of PBST (PBS supplemented with 0.1% Tween 20), the plates were incubated with 100  $\mu$ L of 0.5  $\mu$ M biotinylated ePY ligand (biotin-EYPPYPPPYPSG) for 0.5 h at room temperature. After five washes with 300  $\mu$ L of ice-cold WBKT, aliquots of 100  $\mu$ L of the supernatant containing the ternary complexes were applied to each well of the microtiter plates, which were then shaken at 4°C for 1 h. After five washes with 300  $\mu$ L of WBKT, the retained ternary complexes were eluted with 200  $\mu$ L of ice-cold elution buffer (50 mM DTT, 50  $\mu$ g/ml *Saccharomyces cerevisiae* RNA in WBKT), which cleaves the S-S bond between the biotin moiety and the PY ligand, for 0.5 h at 4°C, and the released mRNA was purified using the SV Total RNA Isolation System (Promega, Madison, WI, USA).

The recovered mRNA was reverse transcribed and amplified by PCR using Taq polymerase (Promega) as described previously (Matsuura and Plückthun 2003). The PCR products were analyzed by agarose gel electrophoresis and used directly, or purified from the gel and reamplified if necessary. The resulting product was digested with *NcoI* and *Bam*HI and ligated again into the vector pRD-n1n2\_2 using the *NcoI*/*Bam*HI restriction sites. The mRNA was prepared from the ligation product as described above and then used for subsequent panning rounds. It should be noted that errors can be introduced not only in the first

round but also during the subsequent rounds, but with lower error rates. Although the ribosome display experiments were carried out using two different *in vitro* translation systems, S30 extract and the PURE system, we found no differences between the sequences of the selected clones, but the enrichment factor judged by the band intensity after RT-PCR was markedly better with the PURE system (data not shown).

#### Expression and Purification of the Selected Clones

The DNA pool after selection rounds was cloned into the *NcoI/BamHI* sites of pQE-pDNB (Matsuura and Plückthun 2003) to add the His-tag at the N-terminus used as an affinity tag for purification and an enterokinase cleavage site to remove the His-tag if necessary. *E. coli* XL1-blue (Stratagene) harboring the plasmid encoding the recombinant protein was grown in LB medium containing ampicillin (50 µg/ml) at 37°C to an OD<sub>600</sub> of 0.5–1.0. Expression was induced by addition of 1 mM IPTG followed by an additional 3-h incubation at 37°C. The proteins were purified by immobilized metal ion affinity chromatography (IMAC) using Ni-NTA superflow (Qia-gen, Hilden, Germany) under denaturing conditions, in accordance with the manufacturer's instructions. Although all WW variants existed in the soluble fraction, they were purified under denaturing conditions because the variants may be sensitive to protease digestion. Purified proteins were then successfully refolded by dialyzing against HBST (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20). When necessary, the His-tag was cleaved from the target protein by enterokinase (EKMax; Invitrogen, San Diego, CA, USA). Finally, the cleaved WW domains were purified to homogeneity by gel filtration chromatography on a HiLoad 16/60 Superdex 30 column (Amersham Bioscience, Piscataway, NJ, USA) using HBST as a running buffer for the subsequent BIAcore experiment. The purity was confirmed by SDS-PAGE and MALDI-TOF mass spectrometry (Voyager-DE STR; Applied Biosystems, Foster City, CA, USA). Note that Tween20 was removed from the sample prior to mass spectrometry by dialysis.

#### Surface Plasmon Resonance (SPR)

SPR experiments were carried out on a BIAcore 3000 instrument (BIAcore, Uppsala, Sweden). All experiments were performed at 25°C in HBST at a flow rate of 20 µl/min. The biotinylated ePY ligand was immobilized on an SA chip (BIAcore) in accordance with the manufacturer's instructions. The amount of PY ligand immobilized on the flow cells was between 50 and 250 resonance units (RU).

An uncoated flow cell was used as a reference cell. It should be noted that some variants showed stronger responses in the reference cell than those in the PY ligand immobilized cell, resulting in negative responses. This may have been because of enrichment of molecules such as streptavidin binders. Dissociation constants ( $K_d$ ) were determined with protein concentrations ranging from 0.1 to 20 µM (seven different concentrations were used for each experiment) and measurement of the RU at equilibrium. Assuming 1:1 binding of the analyte and the ligand, RU at equilibrium ( $R_{eq}$ ) obtained with different concentrations of analyte ( $C$ ) was fitted to the following equation:

$$R_{eq} = \frac{R_{max} C}{K_d + C} \quad (1)$$

where  $R_{max}$  is a constant and  $K_d$  (M) is the dissociation constant. Data are presented as the averages of measurements in three independent flow cells (Table 1).

Solution affinities of WW variants for five gPY ligands were measured by competition BIAcore analysis. Competition experiments were carried out using the same SA chip coated with ePY ligand as described above. Samples of 5 µM WW variant were coincubated with gPY ligand at a series of concentrations (0.01–100 µM) for at least 1 h at 4°C before injection. Plateau responses ( $R_{eq}$ ) of sensorgrams were plotted against the corresponding gPY ligand concentration ( $L_i$ ), and the dissociation constant in solution ( $K_i$ ) was calculated using Eq. 2:

$$R_{eq} = \frac{R_{max}}{1 + \frac{2K_d}{\sqrt{(L_i - C + K_i)^2 + 4K_i C} - (L_i - C + K_i)}} \quad (2)$$

where  $C$  (M) is the concentration of the WW variant (here 5 µM),  $K_d$  (M) is the dissociation constant between WW variant and the ePY ligand estimated using Eq. 1,  $R_{max}$  is a constant, and  $K_i$  is the dissociation constant in solution. Errors for all measurements ranged between 5% and 20% (Table 1).

#### Biophysical Characterization of the Selected Variants

Circular dichroism (CD) spectroscopy was performed with a Jasco J-720WI spectropolarimeter. All measurements were performed at 25°C. The data were normalized to molar ellipticity with a path length of 0.1 cm. The His-tagged WW domain variants were dissolved in TBS (25 mM Tris-HCl, 137 mM NaCl, 2.7 mM KCl, pH 7.4) to a final concentration of 50 µM.

Gel filtration experiments were performed with a SMART system using a Superdex 75 column (Amersham Biosciences) at room temperature. The absorbance data were recorded at 280 nm. A total volume of 25 µl of 10 µM sample was injected and run at a flow rate of 80 µl/

**Table 1** Dissociation constants ( $\mu\text{M}$ ) of WW domain variants<sup>a</sup> to PY ligands

PY ligand variants	sequence <sup>b</sup>	Wild-type	W17F	sk1	se3	pe1	pe2
ePY	EYPPY <b>PP</b> YPSG	13.8 ± 3.2	34.9 ± 7.0	6.9 ± 1.4	14.7 ± 2.7	5.1 ± 0.4	3.9 ± 1.1
gPY	GTP <b>PP</b> YTVG	123	272	94	292	93	94
gPY-P6Y	GTP <b>PP</b> YTVG	281	355	159	435	138	151
gPY-P3I	GTI <b>PP</b> YTVG	155	618	316	918	318	298
gPY-P4S	GTP <b>SP</b> YTVG	1052	823	661	1256	625	675
gPY-P4W	GTP <b>WP</b> YTVG	2444	1047	2025	6967	5274	2665
$K_d^{\text{P4S}} / K_d^{\text{gPY}}$		9	3	7	4	7	7
$K_d^{\text{P4W}} / K_d^{\text{gPY}}$		20	4	22	24	57	28

<sup>a</sup> Dissociation constants ( $K_d$ ) for ePY were measured by direct binding to immobilized ligand, whereas  $K_d$  for gPY and its variants were measured by competition experiments (see Materials and Methods). Errors in measurements of gPY ligands were between 5% and 20%

<sup>b</sup> Bold letters indicate the position of the recognition sequence, PPxY

min using HBST as a running buffer. Standards used for molecular weight calibration were bovine albumin (66 kD), carbonic anhydrase (29 kD), cytochrome *c* (12.4 kD), and aprotinin (6.5 kD) (all from Sigma, St. Louis, MO, USA).

## Results

### Directed Evolution of the W17F Mutant of the WW Domain

To prepare the W17F mutant gene library, we introduced random mutations into the full length of the W17F gene using error-prone PCR. We found an average of 3.2 amino acid changes per clone in the initial library generated by error-prone PCR. This library was then subjected to selection experiments using ribosome display based on binding to the PY ligand, ePY (EYPPYPPPPYPSG). In ribosome display, rare molecules in the initial library (i.e., those with improved binding) are enriched, and thus the proportion of such molecules is expected to increase through selection rounds. After four rounds of selection and amplification, we observed a clear increase in the amount of recovered mRNA, indicating enrichment of molecules with improved binding. It should be noted that errors could be introduced during the amplification step between the rounds of selection, as we used Taq polymerase for amplification, which does not exhibit a high fidelity (see Materials and Methods for details).

After the fourth round, the enriched pool was cloned and isolated clones were subjected to sequence analysis. None of 51 clones sequenced contained stop codons, deletions, or insertions. No mutations were observed at position F17, indicating that all of the selected clones still lacked the Trp

residue in the N-terminus. It should be noted that ribosome display experiments were performed using two different *in vitro* translation systems: *E. coli* S30 extract and the PURE system (see Materials and Methods). Nevertheless, we obtained similar clones containing the same prevalent mutations in these two independent experiments, indicating that these prevalent mutations are not neutral but are likely to exhibit compensatory effects. Five of 51 clones were found to have identical mutations, which we named sk1 (K21R/Q35R/M50V) (Fig. 2B), while most of the other clones were different from each other. Several prevalent mutations were identified among the mutations in the above 51 clones. The most frequent mutations were K21R and Q35R, which were found in 29 (56.9%) and 14 (27.4%) of the 51 clones, respectively (Fig. 1B), and were also present in the enriched clone sk1. Both mutated residues are located in the  $\beta$ -sheet structure region, and the Q35 residue makes direct contact with the PY ligand (Macias et al. 1996), while the K21 residue does not (Fig. 1A). We also found prevalent M50V and N51D mutations in the flexible region of the C-terminus of the WW domain (Fig. 1B), which were found in 14 (27.4%) and 6 (11.8%) of the 51 clones examined, respectively.

### PY Ligand Binding Properties of Selected Variants

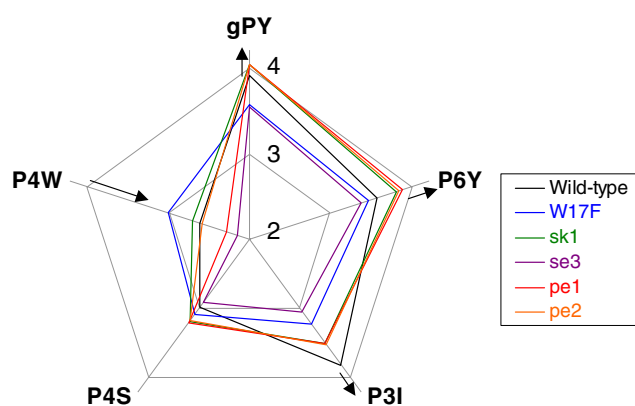
To evaluate the restoration of binding function, we chose 16 representative clones, including sk1, which have different combinations of second-site mutations, and performed preliminary binding analysis by surface plasmon resonance (SPR) (Fig. 2). All 16 clones were purified via the His-tag and then used for the binding assay. Binding assays were carried out by injecting 10  $\mu\text{M}$  samples onto an ePY ligand immobilized sensor chip and the responses



hYAP WW domain have been reported previously (H. I. Chen et al. 1997), and we therefore had sufficient information to design the gPY variants used in this study. The mutations in gPY-P3I and -P6W are outside the recognition motif PPxY, whereas those in gPY-P4S and -P4W are within the recognition motif (Table 1).

Binding affinities of WW variants for each PY ligand are plotted on each axis of a radar plot in Fig. 3, with the binding properties shown in a pentagon shape. The wild type showed a distorted shape from an equilateral pentagon with bias toward gPY, P6Y, and P3I, indicating that its binding is specific for the PPxY sequence, as reported previously (H. I. Chen and Sudol 1995). Unlike the wild type, the W17F mutant showed a nearly equilateral pentagonal plot, indicating that it has broad specificity compared to the wild type. The selected variants excluding se3 showed similar distorted plots, indicating that they have improved affinity to gPY, P6Y, and P3I, similar affinity to P4S, and reduced affinity to P4W compared with the W17F mutant. These results indicate that the selected variants acquired a high degree of specificity for the PPxY sequence.

To quantitatively define the binding specificity for the PPxY motif, we next calculated two ratios of  $K_d$  between the gPY and the gPY-P4S or -P4W, both of which contain the mutation within the PPxY motif (Table 1). Binding specificity can be defined as the ratio of  $K_d$  between the sequences with the highest and lowest affinity, as described previously for DNA binding specificity (Bulyk et al. 2001). For example, the specificity of the wild-type WW domain against the PPxY motif can be defined as  $9 = (1052/123) = (K_d \text{ for gPY-P4S}/K_d \text{ for gPY})$  or  $20 = (2444/123) = (K_d \text{ for gPY-P4W}/K_d \text{ for gPY})$ , while that of the



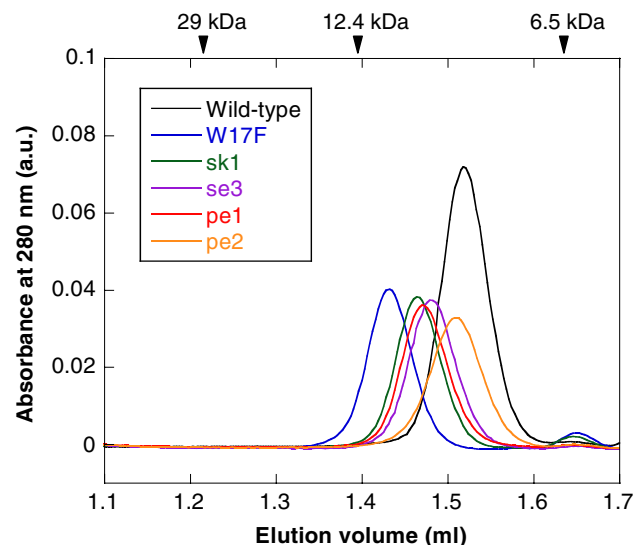
**Fig. 3** Radar plot of binding specificities of the selected variants (sk1, se3, pe1, and pe2), wild type, and W17F mutant. The values of  $-\log K_d$  for each PY ligand are plotted on each axis. The  $K_d$  values are listed in Table 1. Arrows show the directions of changes in binding affinity of the selected variants from the W17F mutant. The binding specificity can be evaluated from the shape of the pentagon. See text for details

W17F mutant is  $3 = (823/272)$  or  $4 = (1047/272)$ , respectively (Table 1). Those with higher specificities showed larger numbers, and thus binding of the W17F mutant to the PY ligand is less specific than that of the wild type. All of the selected variants exhibited increased specificity relative to the W17F mutant (Table 1). These results indicate that by replacing the highly conserved Trp with Phe in the hYAP WW domain, both the binding affinity and the specificity for the PY ligand were decreased, and the introduction of compensatory mutations through evolutionary experiments recovered not only the affinity but also the binding specificity.

### Biophysical Properties of Selected Variants

The W17F mutant is largely unfolded, as suggested by the lack of downfield or upfield resonances in the  $^1\text{H}$  nuclear magnetic resonance (NMR) spectrum typically observed with native protein (Koepp et al. 1999a). We carried out a gel filtration experiment and also performed CD spectroscopy to investigate how the structural features of the functionally restored proteins have been altered in comparison with the parent W17F mutant.

The results of the gel filtration experiments indicated that all six proteins were monomeric, and thus the increase in affinity was not due to oligomerization (Fig. 4). The elution volume of the wild-type protein was somewhat lower than expected for natural globular proteins of this



**Fig. 4** Gel filtration analysis of the selected variants (sk1, se3, pe1, pe2), wild type, and W17F mutant. The absorbance data were recorded at 280 nm. Protein concentration was  $10 \mu\text{M}$  in all cases. Elution volumes of the molecular weight standards are shown at the top: carbonic anhydrase (29 kD), cytochrome *c* (12.4 kD), and aprotinin (6.5 kD)

size. From the elution volume of the wild type and those of the molecular weight standards, the apparent molecular weight of the wild type (actual MW: 6.6 kDa) was estimated to be 30% larger (8.7 kDa) than that of native globular proteins of the same size, most likely because of the flexible region in both termini of the protein as seen in the NMR structure (Macias et al. 1996). In agreement with the previous report that the W17F mutant is largely unstructured, the elution volume of the W17F mutant was larger than that of the wild type, and the apparent molecular weight of the W17F mutant was 90% (12.5 kDa) greater than that of a globular protein of the same size. The peaks of the selected variants (sk1, se3, pe1, and pe2) were distributed between those of the wild type and the W17F mutant, and thus the selected variants had lower apparent molecular weights than the unstructured W17F mutant and were therefore more compact than the parent molecule, despite the lack of the Trp17 residue constituting the hydrophobic core in the wild type. Among the selected variants, pe2 showed the largest elution volume, which was almost the same as that of the wild type.

We used far-UV CD spectroscopy to compare the secondary structures of the selected variants, the wild type, and the W17F mutant in the presence and absence of the ePY ligand (Fig. 5). The spectrum of the wild-type WW domain exhibited the maximum peak at 230 nm as reported previously (Koeppf et al. 1999a), resulting from the aromatic contribution to the far-UV CD spectrum (Krittana and Johnson 1997; Woody 1994). This distinctive feature has therefore been used as a qualitative signature of the three-stranded  $\beta$ -sheet of the WW fold (Dalby et al. 2000; Jiang et al. 2001; Koeppf et al. 1999b). The spectra for the W17F mutant and the selected variants showed blue-shifted peaks at 225 nm with peak intensities that were different from each other (Fig. 5A). Shifted peaks at 225–230 nm are often seen in other WW domains (Kraemer-Pecore et al. 2003; Przedziak et al. 2006). The differences in peak wavelength between the wild type and the selected variants, including the W17F mutant, are likely due to the

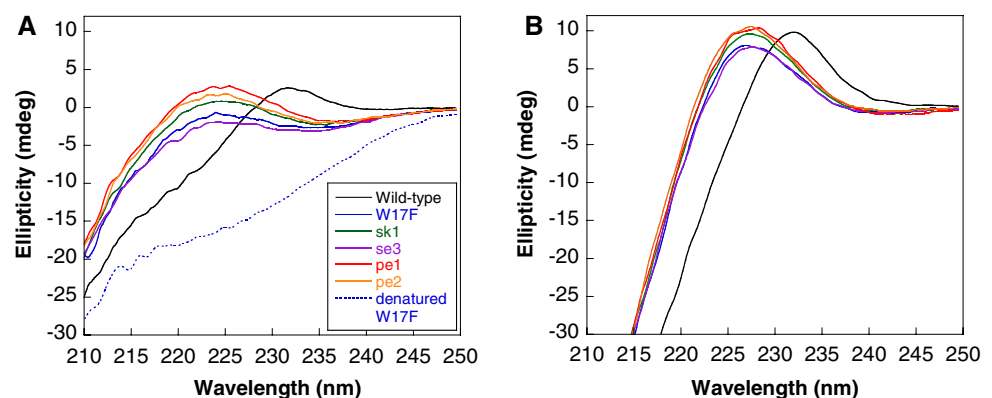
changes in the environment of the aromatic residues resulting from the lack of N-terminal Trp in these proteins. An NMR study showed that the W17F mutant folds into a well-defined WW domain structure upon ligand binding (Koeppf et al. 1999a), and in the CD spectra this resulted in an increase in the peak at 225 nm (see below). Therefore, the shifted peaks at 225 nm can also be used as a signature of the folded WW domain structure.

The W17F mutant yielded a significant decrease in the peak intensity compared to the wild type, as reported previously (Koeppf et al. 1999a). Upon thermal denaturation at 80°C, this peak disappeared completely, resembling a typical random-coil spectrum (Fig. 5A). Thus, the W17F mutant had decreased  $\beta$ -sheet structure of the WW domain, but not a typical random-coil structure, as determined by CD spectroscopy. The spectra of the selected variants (sk1, pe1, and pe2) showed higher intensities of the peaks around 225 nm compared with the W17F mutant, suggesting an increase in their  $\beta$ -sheet structure in comparison to the W17F mutant. The variant se3 had a smaller peak at 225 nm than the W17F mutant, possibly because the se3 sequence contained a mutation at Ile7, which covers the hydrophobic patch and is therefore essential for folding (Macias et al. 1996). Addition of the ePY ligand to the selected variants resulted in an increase in positive ellipticity in the 225- to 230-nm range characteristic of a folded WW domain (Fig. 5B), indicating conformational changes when bound to the PY ligand, as seen in both the wild type and the W17F mutant.

## Discussion

In the process of molecular evolution, compensatory mutations are thought to act as a salvage mechanism against deleterious changes in functional properties of proteins (DePristo et al. 2005; Jucovic and Poteete 1998; Kimura 1991). In the present study, we examined whether and to what degree the defective W17F mutant of the WW

**Fig. 5** Far-UV CD spectra of the selected variants (sk1, se3, pe1, pe2), wild type, and W17F mutant at 25°C in the (A) absence or (B) presence (100  $\mu$ M) of the PY ligand. In (A), the spectrum of the W17F mutant measured at 80°C is shown with a dashed line. Protein concentration was 50  $\mu$ M in all cases





domain, containing substitution in a strictly conserved residue, could be restored through second-site mutations, providing additional evidence that a defective protein variant could be restored through second-site mutations. The results indicated that only a few second-site mutations, such as K21R/Q35R, were sufficient to restore the reduced binding function of the W17F mutant and even exceeded that of the wild type. In addition, improvements in binding specificity and structural properties accompanying functional restoration were also seen, as observed previously (Eaton et al. 1995; Ruan et al. 1998; Starovasnik et al. 1997). Thus, we have shown that second-site mutations not only can compensate for lowered function and structural loss due to replacement of conserved residues, but also can create proteins with improved functionality over the original protein. These observations suggest the possible path taken in protein evolution when a conserved residue is replaced; divergence after deleterious mutation in a conserved residue of a protein may not necessarily be eliminated but may be maintained through compensatory mutations in protein evolution.

Compensatory mutations are usually defined as a pair of substitutions at different sites in a protein that have a harmful effect independently but result in recovery of the original function in combination. This type of effect has been reported to take place between amino acid residues in close proximity in the folded state of a protein (Kimura 1991). On the other hand, none of the selected variants in the present study had direct or proximal mutations to the Phe17 residue. In a previous study using T4 lysozyme, most of the functional revertants carried primary-site mutations (Jucovic and Poteete 1998). However, we did not find any primary-site mutations, including simple F17W reversion, probably because F17W reversion requires two specific transversions in the DNA sequence, the occurrence of which is rare. In the wild-type structure, second-site mutations that were preserved among selected variants, such as K21R, Q35R, and the C-terminal mutations, were all located on the surface of the WW domain, and relatively remote from the site of the primary W17F mutation in the hydrophobic patch (Fig. 1A). These mutations are therefore likely to have compensatory effects independent of the site of primary mutation, i.e., additivity effects (Wells 1990). It is also possible that such mutations are “global suppressor” mutations, which are known to suppress more than one primary mutation at different sites (Poteete et al. 1997; Shortle and Lin 1985). Among the preserved mutations, only the Q35 residue comes into direct contact with the PY ligand in the three-dimensional structure of the wild-type protein (Macias et al. 1996). Therefore, although the physicochemical details have yet to be determined, the Q35R mutation appeared to contribute greatly to the functional restoration in the selected variants. In contrast,

the K21 and C-terminal residues are not located in positions where they can make contact with the hydrophobic core or the ligand in the wild-type three-dimensional structure. Interestingly, improvements in structural properties, such as amount of secondary structural elements and apparent molecular radius, occurred without mutation to residues related to structural stability, such as the hydrophobic core. Further investigations are required to determine the individual contributions of these mutations and their effects.

In this study, the defective W17F mutant due to replacement of a highly conserved residue was found to have relaxed binding specificity compared with the wild-type protein (Fig. 3). As a result, the W17F mutant showed a higher binding affinity than the wild type toward gPY-P4W, which does not have the consensus sequence of the PY ligand, PPxY. These observations suggest that mutation of the Trp17 residue, a conserved residue in the WW domain, disrupts the structure but generates functional promiscuity due to relaxation of binding specificity. Recent studies also showed that proteins with structural plasticity displayed functional promiscuity. For example, an antibody with conformational diversity showed multispecificity, and natively unfolded proteins, a number of which have recently been found in eukaryotic cells, have been shown to be multifunctional (James et al. 2003; Tompa et al. 2005). Functional promiscuity generated by such structural plasticity may be an important property of a protein as a source of evolution generating proteins with new functions (James and Tawfik 2003).

The results of the present study have shown that even when a deleterious mutation occurs in a highly conserved residue of a protein during divergent evolution, it is possible that subsequent divergence may be maintained and can be improved even further through compensatory mutations. In addition, proteins with disrupted structure through mutation in conserved residues may become sources of divergent evolution, leading to the generation of families with new functions due to functional promiscuity. The above observations indicated that new functional proteins with novel binding specificities can be generated from the W17F mutant. For example, in comparison with the wild type, the W17F mutant is more likely to evolve toward proteins with novel binding specificity for P4W in evolutionary experiments. This may be true when selecting for ligands with structures that are markedly different from the original PY ligand, as the wild type requires global rearrangement of its structure to fit the novel ligand. On the other hand, there may be cases in which the wild type is preferred when selecting for ligands with similar conformation to the original PY ligand, as the wild type already has a rigid structure and requires only a few amino acid changes in the ligand-binding surface for improved affinity.

Further investigations will help to understand the quantitative effects of structural plasticity generated by the unfolded molecule on the ability to evolve new functions.

**Acknowledgments** The authors thank Drs. I. Urabe and K. Tsukada for helpful discussions, and the Post Genome Institute (Tokyo) for providing the PURE system. This research was partially conducted at Open Laboratories for Advanced Bioscience and Biotechnology (OLABB), Osaka University. This research was supported in part by the Special Coordination Funds for Promoting Science and Technology: Yuragi Project and the Global COE (Centers of Excellence) Program of the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

## References

- Arakawa T, Jongsareejit B, Tatsumi Y, Tanaka K, Ikeda K, Komatsubara H, Inoue H, Kawakami B, Oka M, Emi S, Yomo T, Shima Y, Negoro S, Urabe I (1996) Application of N-terminally truncated DNA polymerase from *Thermus thermophilus* (delta Tth polymerase) to DNA sequencing and polymerase chain reactions: comparative study of delta Tth and wild-type Tth polymerases. *DNA Res* 3:87–92
- Bork P, Sudol M (1994) The WW domain: a signalling site in dystrophin? *Trends Biochem Sci* 19:531–533
- Bulyk ML, Huang X, Choo Y, Church GM (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* 98:7158–7163
- Chen HI, Sudol M (1995) 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
- Chen HI, Einbond A, Kwak SJ, Linn H, Koepf E, Peterson S, Kelly JW, Sudol M (1997) Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J Biol Chem* 272:17070–17077
- Chen HZ, Zubay G (1983) Prokaryotic coupled transcription-translation. *Methods Enzymol* 101:674–690
- Dalby PA, Hoess RH, DeGrado WF (2000) Evolution of binding affinity in a WW domain probed by phage display. *Protein Sci* 9:2366–2376
- DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6:678–687
- Eaton BE, Gold L, Zichi DA (1995) Let's get specific: the relationship between specificity and affinity. *Chem Biol* 2:633–638
- Guo HH, Choe J, Loeb LA (2004) Protein tolerance to random amino acid change. *Proc Natl Acad Sci USA* 101:9205–9210
- Hanes J, Plückthun A (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 94:4937–4942
- Innis CA, Shi J, Blundell TL (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. *Protein Eng* 13:839–847
- James LC, Tawfik DS (2003) Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28:361–368
- James LC, Roversi P, Tawfik DS (2003) Antibody multispecificity mediated by conformational diversity. *Science* 299:1362–1367
- Jiang X, Kowalski J, Kelly JW (2001) Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. *Protein Sci* 10:1454–1465
- Jucovic M, Poteete AR (1998) Delineation of an evolutionary salvage pathway by compensatory mutations of a defective lysozyme. *Protein Sci* 7:2200–2209
- Kimura M (1991) Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc Natl Acad Sci USA* 88:5969–5973
- Koepf EK, Pettrassi HM, Ratnaswamy G, Huff ME, Sudol M, Kelly JW (1999a) Characterization of the structure and function of W → F WW domain variants: identification of a natively unfolded protein that folds upon ligand binding. *Biochemistry* 38:14338–14351
- Koepf EK, Pettrassi HM, Sudol M, Kelly JW (1999b) WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci* 8:841–853
- Kondrashov AS, Sunyaev S, Kondrashov FA (2002) Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci USA* 99:14878–14883
- Kraemer-Pecore CM, Lecomte JT, Desjarlais JR (2003) A *de novo* redesign of the WW domain. *Protein Sci* 12:2194–2205
- Krittanaï C, Johnson WC (1997) Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Anal Biochem* 253:57–64
- Macias MJ, Hyvonen M, Baraldi E, Schultz J, Sudol M, Saraste M, Oshkinat H (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* 382:646–649
- Matsuura T, Plückthun A (2003) Selection based on the folding properties of proteins with ribosome display. *FEBS Lett* 539:24–28
- Matsuura T, Yanagida H, Ushioda J, Urabe I, Yomo T (2007) Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. *Biochem Biophys Res Commun* 352:372–377
- Nieba L, Krebber A, Plückthun A (1996) Competition BIAcore for measuring true affinities: large differences from values determined from binding kinetics. *Anal Biochem* 234:155–165
- Nikolova PV, Wong KB, DeDecker B, Henckel J, Fersht AR (2000) Mechanism of rescue of common p53 cancer mutations by second-site suppressor mutations. *EMBO J* 19:370–378
- Poteete AR, Rennell D, Bouvier SE, Hardy LW (1997) Alteration of T4 lysozyme structure by second-site reversion of deleterious mutations. *Protein Sci* 6:2418–2425
- Pritchard L, Dufton MJ (1999) Evolutionary trace analysis of the Kunitz/BPTI family of proteins: functional divergence may have been based on conformational adjustment. *J Mol Biol* 285:1589–1607
- Przedziak J, Tremmel S, Kretzschmar I, Beyermann M, Bienert M, Volkmer-Engert R (2006) Probing the ligand-binding specificity and analyzing the folding state of SPOT-synthesized FBP28 WW domain variants. *Chembiochem* 7:780–788
- Ruan B, Hoskins J, Wang L, Bryan PN (1998) Stabilizing the subtilisin BPN' pro-domain by phage display selection: How restrictive is the amino acid code for maximum protein stability? *Protein Sci* 7:2345–2353
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 95:5857–58
- Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol* 19:751–755
- Shortle D, Lin B (1985) Genetic analysis of staphylococcal nuclease: identification of three intragenic “global” suppressors of nuclease-minus mutations. *Genetics* 110:539–555
- Sowa ME, He W, Slep KC, Kercher MA, Lichtarge O, Wensel TG (2001) Prediction and confirmation of a site critical for

- effector regulation of RGS domain activity. *Nat Struct Biol* 8:234–237
- Starovasnik MA, Braisted AC, Wells JA (1997) Structural mimicry of a native protein by a minimized binding domain. *Proc Natl Acad Sci USA* 94:10080–10085
- Sudol M (1998) From Src Homology domains to other signaling modules: proposal of the 'protein recognition code.' *Oncogene* 17:1469–1474
- Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 30:484–489
- Wells JA (1990) Additivity of mutational effects in proteins. *Biochemistry* 29:8509–8517
- Woody RW (1994) Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur Biophys J* 23:253–262