

# Development and Applications of a Calmodulin-Based Fusion Protein System for the Expression and Purification of WW and Zinc Finger Modules

Christopher G. Toomey<sup>1</sup>, David Weiss<sup>1</sup>, Alan Chant<sup>2</sup>, Megan Ackerman<sup>1</sup>, Bethany A. Ahlers<sup>3,4</sup>, Ying-Wai Lam<sup>3,4</sup>, Christopher Ricciardi<sup>1</sup>, Dana Bourne<sup>5</sup>, Christina M. Kraemer-Chant<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, Saint Michael's College, Colchester, VT, USA

<sup>2</sup>Department of Chemistry, University of Vermont, Burlington, VT, USA

<sup>3</sup>Department of Biology, University of Vermont, Burlington, VT, USA

<sup>4</sup>Vermont Genetics Network Proteomics Facility, University of Vermont, Burlington, VT, USA

<sup>5</sup>Department of Biochemistry, University of Vermont, Burlington, VT, USA

Email: \*cchant@smcvt.edu

**How to cite this paper:** Toomey, C.G., Weiss, D., Chant, A., Ackerman, M., Ahlers, B.A., Lam, Y.-W., Ricciardi, C., Bourne, D. and Kraemer-Chant, C.M. (2017) Development and Applications of a Calmodulin-Based Fusion Protein System for the Expression and Purification of WW and Zinc Finger Modules. *Advances in Biological Chemistry*, 7, 89-106.

<https://doi.org/10.4236/abc.2017.72006>

**Received:** January 20, 2017

**Accepted:** April 22, 2017

**Published:** April 25, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Calmodulin from *Homo sapiens* is an  $\alpha$ -helical calcium-binding protein that expresses to high levels in *E. coli*. When the N-terminus of a calmodulin variant is bound to  $\text{Ca}^{2+}$ , it undergoes a conformational change, exposing hydrophobic pockets. This property can be utilized for purification purposes, as these pockets bind to phenyl sepharose resin with high affinity. Washing with EDTA chelates the  $\text{Ca}^{2+}$  ions from the protein, inducing a conformational change back to the more folded state and eluting the protein from the column. We describe herein the use of a protein expression and purification technique using the calmodulin variant and a short linker for proteolytic cleavage by the mutant NIa-Pro tobacco etch virus protease. We have shown this approach to be useful in obtaining purified quantities of various small proteins that could not be expressed using other methods, including high enough concentrations of a designed WW domain protein for NMR structural analysis. We have also obtained promising results on the usefulness of this procedure to express and purify zinc finger proteins without the addition of zinc ions or other cofactors.

## Keywords

Biomaterials, Protein Purification, Expression Systems, Protein Expression

---

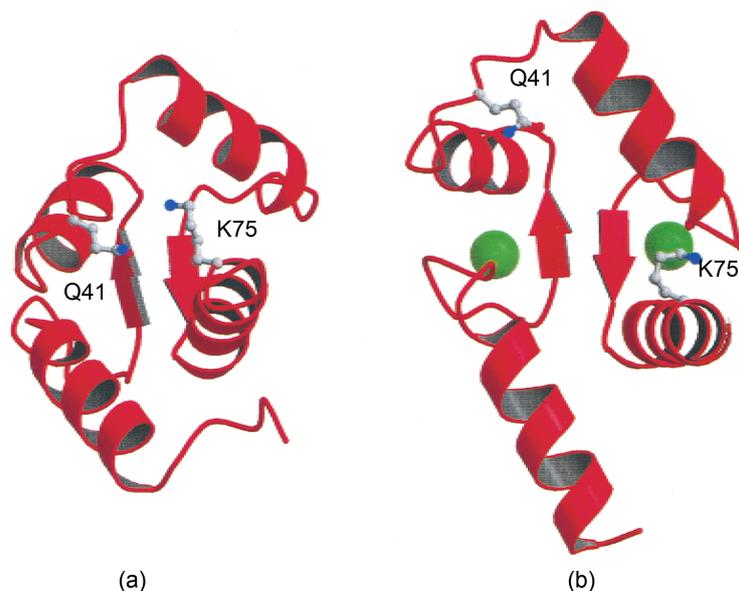
## 1. Introduction

Analysis of native and mutant protein structure and function is dependent on

first obtaining purified protein. One of the more common approaches involves the use of bacterial systems for expression of protein, followed by affinity or size-exclusion resins for protein purification. One of the most useful affinity tags is the (His)<sub>6</sub> tag. This method involves the addition of a six-histidine stretch at the N- or C-terminus of the target protein. Usually, the His tag will not alter the structure of the protein; however, we have shown in a previous study that the His tag can actually disrupt the function and/or folding of certain proteins, which can be problematic [1].

The work described herein uses a fusion protein system to increase expression and ease of purification of selected small proteins. It involves the N-terminus of a mutant calmodulin protein from *H. sapiens* (N-Cam. Y [2]) fused genetically to the target protein, with a site for proteolysis by the mutant N1a-Pro tobacco etch mosaic virus protease (TEV) located between the C-terminus of N-Cam. Y and the N-terminus of the protein of interest. N-Cam. Y is a calcium-binding protein that exposes hydrophobic pockets upon the binding of Ca<sup>2+</sup> ions. This characteristic can be utilized for purification purposes, as these hydrophobic pockets bind to the hydrophobic phenyl sepharose resin with high affinity when Ca<sup>2+</sup> is present. Washing the protein-bound column with EDTA chelates the Ca<sup>2+</sup> ions from the protein, which undergoes a conformational change back to its free state (Figure 1).

TEV cleaves a specific sequence of amino acids (sequence ENLYFQ/GS for the original version of the protease, and ENLYFQ/G for the newer version, where represents the site of proteolytic cleavage by TEV). According to the manufacturer (Promega, in this work), the most current iteration of the TEV protease



**Figure 1.** Ca<sup>2+</sup>-induced conformational changes in the N-terminal domain of calmodulin. (a) Ribbon diagrams show the differences in the conformations of Ca<sup>2+</sup>-free N-Cam (Kuboniwa *et al.* 1995) and (B) Ca<sup>2+</sup>-saturated N-Cam (Chattopadhyaya *et al.* 1992). Green spheres represent the Ca<sup>2+</sup> ions. Figure taken from Ababou and Desjarlais [10].

cleaves effectively to ~95% completion in most non-SDS buffers overnight. The result is a target protein with two additional residues (Gly-Ser) at the N-terminus with the older TEV protease, and one Gly residue at the N-terminus with the newer protease.

The proteins that initiated development of this expression and purification scheme were a wild-type WW domain from human peptidyl-prolyl *cis-trans* isomerase (hPin1) and computationally-designed proteins based on the WW family. WW domains are protein-protein interaction modules that are named for the two highly conserved Trp residues. WW domains are typically between 34 - 40 amino acids long, which makes them one of the smallest naturally occurring monomeric protein domains. The WW domain from hPin1 is a characteristic WW domain comprised of 34 amino acids forming three antiparallel  $\beta$ -sheets. Due to its size, it can be difficult to express and purify successfully; the Oschkinat group [3] found that the WW domain from human Yes-associated protein 65 was only expressible as a fusion protein, and our work on the WW domain from hPin1 required use of the fusion protein described herein, as expression using other methods yielded poor results [4].

We have also used this system to express and purify a specific zinc finger module (ZF23 as labeled by the Zinc Finger Consortium). Zinc finger motifs, or domains, are independently folded domains found within zinc finger proteins and contain a two-stranded antiparallel  $\beta$ -sheet as a  $\beta$ -hairpin, an  $\alpha$ -helix, and a zinc (II) ion. They are usually quite small (roughly 30 amino acids) and are usually expressed as His-tagged constructs for ease of purification.

There are few published studies of the usefulness of calmodulin for protein expression and purification; other groups have applied the method for the expression of plant proteins [5], in protein expression using insect cells [6], or with a selective factor Xa cleavage site present for removal of the calmodulin tag [7], for example. Our system is novel in that it is used for expression in *E. coli*, and has at the N-terminus of the target protein the N-Cam. Y tag, and includes a proteolytic site for removal of the tag. We have found that calmodulin typically overexpresses to high levels in *E. coli*, enabling large amounts of the target protein to be obtained; having N-Cam. Y N-terminal to our target protein exploits this characteristic. This method has proven to be essential for the overexpression and purification of several wild-type and novel proteins, one to a high enough concentration and purity for structural characterization by NMR [4].

## 2. Materials and Methods

### 2.1. Obtaining the Gene for the WW Domain Fusion Protein

Unless otherwise mentioned, all DNA oligomers for the WW portion of the project were purchased from Invitrogen Life Technologies at a 50 nmol scale with standard desalting purification, all enzymes and enzymatic buffers were purchased from New England Biolabs, all competent cells were purchased as a glycerol stock from Novagen, except for DH5 $\alpha$  competent cells, which were originally purchased from Invitrogen Life Technologies as a glycerol stock, and

all chemicals were purchased from J. T. Baker, except for chloramphenicol,  $\text{MgCl}_2$ , and DNase I, which were purchased from USB. For the zinc finger portion of the project, tryptone, Amp, and EDTA were purchased from Acros; APS was purchased from National Diagnostics, and NaCl was purchased from Sigma Aldrich. All other chemicals were purchased from Fisher Scientific unless otherwise mentioned.

The genes coding for the N-Cam. Y/TEV/WW constructs were designed using the typical codon preference in *E. coli* BL21 to select for the frequency of codon usage. The final results were expressible genes that coded for the proteins of interest.

## 2.2. Obtaining the Gene for the Zinc Finger Fusion Protein

The gene coding for the N-Cam/TEV/ZF23 was designed using the typical codon preference in *E. coli* BL21 to select for the frequency of codon usage. An NdeI site was placed at the 5' end of the gene and two stop codons combined with a BamHI site were placed at the 3' end. The gene, subcloned into the NdeI and BamHI sites of a pET21-a(+) vector, was ordered from Genscript. These restriction enzyme sites were selected as they would result in the gene of interest occurring some distance from the His tag found in the pET-21 systems and with the T7 tag (which occurs between NdeI and BamHI in this pET construct) removed. The sample was provided as a 4  $\mu\text{g}$  lyophilized sample, which was resuspended in 50  $\mu\text{L}$  nuclease-free water before use.

## 2.3. WW Expression in BL21 or HMS174 Bacterial Cells

1  $\mu\text{L}$  of dilutions of each plasmid (1, 1/10, 1/100, 1/1000) was transformed into 50  $\mu\text{L}$  of BL21 or HMS174 competent cells and plated. After ~22 h, one colony was inoculated into 10 mL LB containing 10  $\mu\text{L}$  of 25 mg/mL chloramphenicol and 40  $\mu\text{L}$  of 25 mg/mL Amp. This was placed in an orbital shaker at 37°C and allowed to shake overnight at 150 rpm.

The next morning, 5 mL of the overnight growth was inoculated into 1 L of LB containing 0.5 mL of 25 mg/mL chloramphenicol and 4 mL of 25 mg/mL Amp. This was allowed to shake at 250 rpm until the  $\text{OD}_{600}$  was 0.5 - 0.8; IPTG was then added (5 mL of 100 mM IPTG per liter of cells). The cells were harvested by centrifugation (5000 $\times$  g for 10 minutes) after at least 3 - 4 h post-induction and the cell pellets were resuspended in 20 mL TE per liter of LB. A crystal of DNase I was added and the cell suspensions were frozen overnight for increased cell lysis. After ~16 h, the cell paste was thawed, stirred for 30 minutes, and homogenized using a standard cell homogenizer. 0.351 g NaCl and 0.123 g  $\text{MgCl}_2$  were added to each sample and the suspensions were stirred for 20 minutes. The cell suspension was centrifuged at 12,000 $\times$  g for 20 minutes, and the supernatant was saved for column purification.

## 2.4. Purification of WW Fusion Proteins

The main method of purification of the WW fusion proteins involved column

chromatography using phenyl sepharose resin. Two other methods, utilized only when this primary method did not work, were also employed and are described in the Supporting Information.

After removing the supernatant from the cell solution as described above, the appropriate volume (~20 mL) was made 5 mM Ca<sup>2+</sup> and 500 mM Na<sup>+</sup>, doubled in volume with high-salt Ca<sup>2+</sup> wash buffer (50 mM Tris base, 500 mM NaCl, 1 mM CaCl<sub>2</sub>, pH 7.5), placed on ice, and applied to a newly-poured and washed phenyl sepharose column (~20 mL pre-swelled resin volume, Pharmacia, Peapack, NJ, CL-4B) at low pressure using a BioRad LP system. The system was washed with the high-salt Ca<sup>2+</sup> wash buffer until the UV absorbance returned to the pre-loading value. The protein was eluted with 5 mM EDTA buffer (50 mM Tris base, 500 mM NaCl, 5 mM EDTA, pH 7.5). The samples containing the desired protein as determined by SDS-PAGE were pooled and reserved for either HPLC or direct proteolytic cleavage.

### 2.5. Cleavage of WW Fusion Proteins with TEV Protease

Once the fusion protein was purified from the background proteins, it was cleaved by the TEV protease method. The older version of the TEV protease, which cleaves ENLYFQ/GS, was used for the WW fusion proteins. Fusion protein post-purification (about 50 μM) was incubated at room temperature with protease in a 200:1 protein: protease ratio for a maximum of 16 h. After 3 h, cleavage was roughly 75% complete as determined by SDS-PAGE; near-100% cleavage was reached overnight. Approximately 5 mL of purified fusion protein (~100 μM) was fully cleaved with 5 μL protease after 2 - 3 days at room temperature.

Following cleavage, another phenyl sepharose column was run to remove any of the proteins containing the N-Cam. Y portion. The protein sample was placed in a Falcon tube with approximately 9.5 mL of cleaned phenyl sepharose. A total of 1.25 mL of 100 mM CaCl<sub>2</sub> was added to induce conformational change in N-Cam. Y and concomitant binding to the phenyl sepharose resin. This was rocked on a platform for 1 h at room temperature and was then poured into a column. Approximately 5 mL of the sample (containing the WW domains) was collected under atmospheric pressure; this was dialyzed into 100 mM NaCl, 1 mM CaCl<sub>2</sub> and 10 mM MOPS (Sigma), pH 7.2 with a Spectra/Por membrane (Spectrum, Rancho Dominguez, CA; MW cutoff of 3500 Da).

Alternatively, the cleaved sample was HPLC-purified. This involved filtering the sample with a 0.45 μm Acrodisc filter (Pall Corporation, East Hills, NY) and loading 5 mL aliquots onto a reverse-phase HPLC system washed with 100% HPLC-grade water/0.1% TFA (buffer A). A gradient of 90% acetonitrile (Burdick and Jackson)/10% HPLC-grade water/0.1% TFA (buffer B) was initialized; as the run progressed, increasing concentration of buffer B was mixed with buffer A and added to the column. The peaks were collected and lyophilized. The resulting protein powder was resuspended in distilled water and analyzed by SDS-PAGE.

## 2.6. Analysis and Characterization of the Native and Designed WW Proteins

The identity and integrity of all purified WW proteins were qualitatively analyzed by SDS-PAGE and confirmed by MALDI mass spectrometry at the Pennsylvania State University mass spectrometry facility. Concentrations of the proteins were calculated using the absorbance at 280 nm in non-denaturing conditions and the absorbance coefficient as calculated from the primary structure. Certain WW proteins were characterized using standard biophysical techniques as described previously [4].

## 2.7. Expression and Purification of ZF23 Fusion Protein in BL21

The ZF23 fusion protein was transformed into BL21 competent *E. coli* cells and plated following standard procedures. After overnight incubation at 37°C, one colony was selected and inoculated into a Falcon tube containing 5 mL LB and 20 µL of 25 mg/mL Amp, and the tube was shaken for ~16 h at 37°C at 180 rpm in an orbital shaker. The next morning, all of the volume from the tube was placed into 125 mL of LB in a glass 250-mL Erlenmeyer flask and shaken for ~6 h. The sample was visually inspected for sufficient OD. The 125-mL sample was then induced with 125 µL of 100 mM IPTG. After ~16 h, the sample was centrifuged at 15,000× g for 15 minutes. Cell pellets were subjected to 4 repetitions of 30-minute freeze-thaw cycles and sonicated three times on high for 2 minutes. Pellets were then resuspended in a total volume of 25 mL of high-salt Ca<sup>2+</sup> wash buffer and centrifuged at 15,000× g for 15 minutes. The supernatant was removed and loaded onto a phenyl sepharose column (volume ~1 mL resin) that had been prewashed with the high-salt Ca<sup>2+</sup> wash buffer. The column was rinsed thoroughly with buffer and the protein of interest was eluted with 5 mM EDTA buffer. In total, 18,500-µL fractions were collected; while no UV monitoring of the column was done, SDS-PAGE analysis of each fraction was performed. Two fractions that contained the highest concentration of protein as determined by SDS-PAGE were concentrated using a Centricon device with an Ultracel-3 membrane (Millipore, YM-3) and digested with the TEV protease (purchased from Promega as ProTEV Plus, 5 u/µL) following the manufacturer-supplied protocol. Samples at 0, 4, and 22 h were taken for SDS-PAGE analysis and bands of interest were excised for LC-MS/MS analysis.

## 2.8. ZF23 Fusion Protein Identification by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Tryptic digestion was performed as described previously [8] and mass spectrometry was performed as described in the Supplemental Information section.

# 3. Results

## 3.1. Expression and Purification of WW Domains

Overall, the use of the system led to the successful expression and purification of

7 target proteins. The only wild-type protein studied was the wild-type WW domain from human Pin1 (hPin1 WW), which was selected only as basis for the computational design of the other WW-like proteins. The others were all computationally-designed proteins based on the WW family. (The computational design methods are published elsewhere [4]) The results are summarized in **Tables 1(a)-(c)**; deviations from standard procedures and results are described

**Table 1.** (a) Names, sequences, and molecular weights of each WW protein expressed and purified using the approach described herein. (b) Results of expression and purification of each WW protein. (c) Summary of final results from expression and purification of WW proteins.

(a)					
Name of protein	Amino acid sequence		Calculated molecular weight (MW), Da		
hPin1 WW	KLPPGWEKRMSSGRVYFNFHITNASQWERPSG		4168		
SPA-WW1	DLPDGWKVLVVAGSAVIYFFNEASGIKQKTRPTS		3839		
SPANS-WW1	NLPSGWTTPRTKSGASENYYYNKETNEVTNTRPTD		4036		
SPANS-WW2	SLPSGWTQLTKASDDTTYYNKTTDVVTNTRPTD		3985		
SPANS-WW3	ELPSGWTTPRKKDGASEFFYYNENTSEVQGTTRPTD		N/A		
K8P wt mutant	KLPPGWEPRMSSGRVYFNFHITNASQWERPSG		4137		
N17F					
SPANS-WW1 mutant	NLPSGWTTPRTKSGASEFYYNKKETNEVTNTRPTD		N/A		
W29A wt mutant	KLPPGWEKRMSSGRVYFNFHITNASQAERPSG		4052		

(b)					
Name	Expressed 3 hr post-IPTG induction? <sup>a</sup>	Protein soluble? <sup>b</sup>	Protein ~90% pure after phenyl sepharose purification? <sup>b</sup>	Concentration of final fusion protein after lyophilization ( $\mu$ M) <sup>c</sup>	Successful cleavage (~90%) by TEV? <sup>d</sup>
hPin1 WW	Yes	Yes	Yes	50	Yes
SPA-WW1	Yes	Yes	Could not purify	N/A	N/A
SPANS-WW1	Yes (4 hr)	Yes	Yes <sup>e</sup>	70 - 100	Yes (3 mL)
SPANS-WW2	Yes	Yes	Yes (plus basal cleavage in TEV)	N/A	Yes
SPANS-WW3	Yes	Yes	Yes	N/A	Yes (3 days)
K8P	Yes	Yes	Yes	N/A	Yes (10 mL)
N17F	Yes	Yes	Yes	N/A	Yes
W29A	Yes	Yes	Yes	N/A	Yes

<sup>a</sup>As determined qualitatively by SDS-PAGE. <sup>b</sup>As determined by SDS-PAGE *via* analysis of the supernatant after centrifugation of the cell pellet. <sup>c</sup>Calculated by UV spectroscopy at 280 nm as described in text. <sup>d</sup>As determined by SDS-PAGE after 2 days, unless otherwise noted. <sup>e</sup>SPANS-WW1 eluted from the phenyl sepharose column at a lower volume of elution buffer than the wild-type WW, which could have indicated a decreased affinity for phenyl sepharose. This may have pointed to a molten or unfolded protein that interacted with the N-Cam. Y fusion protein.

(c)

Name	Final concentration post-cleavage and final lyophilization ( $\mu\text{M}$ ) <sup>a</sup>	Molecular weight as determined by MALDI mass spectrometry	Further characterization performed?
hPin1 WW	410	4168.18 and 4185.02	CD [4]
SPANS-WW1	180	4040.62 <sup>b</sup> and 4316.42 (~50/50 mix)	CD
SPANS-WW2	260 - 1850	3986.32 and 4261.67 (~50/50 mix)	CD, 1-D and 2-D NMR (COSY, TOCSY, NOESY) [4]
SPANS-WW3	1450 - 1610	N/A	CD
K8P	460	4139.35 and 4414.99 (~50/50 mix)	CD
N17F	1750	N/A	CD
W29A	1970	N/A	CD

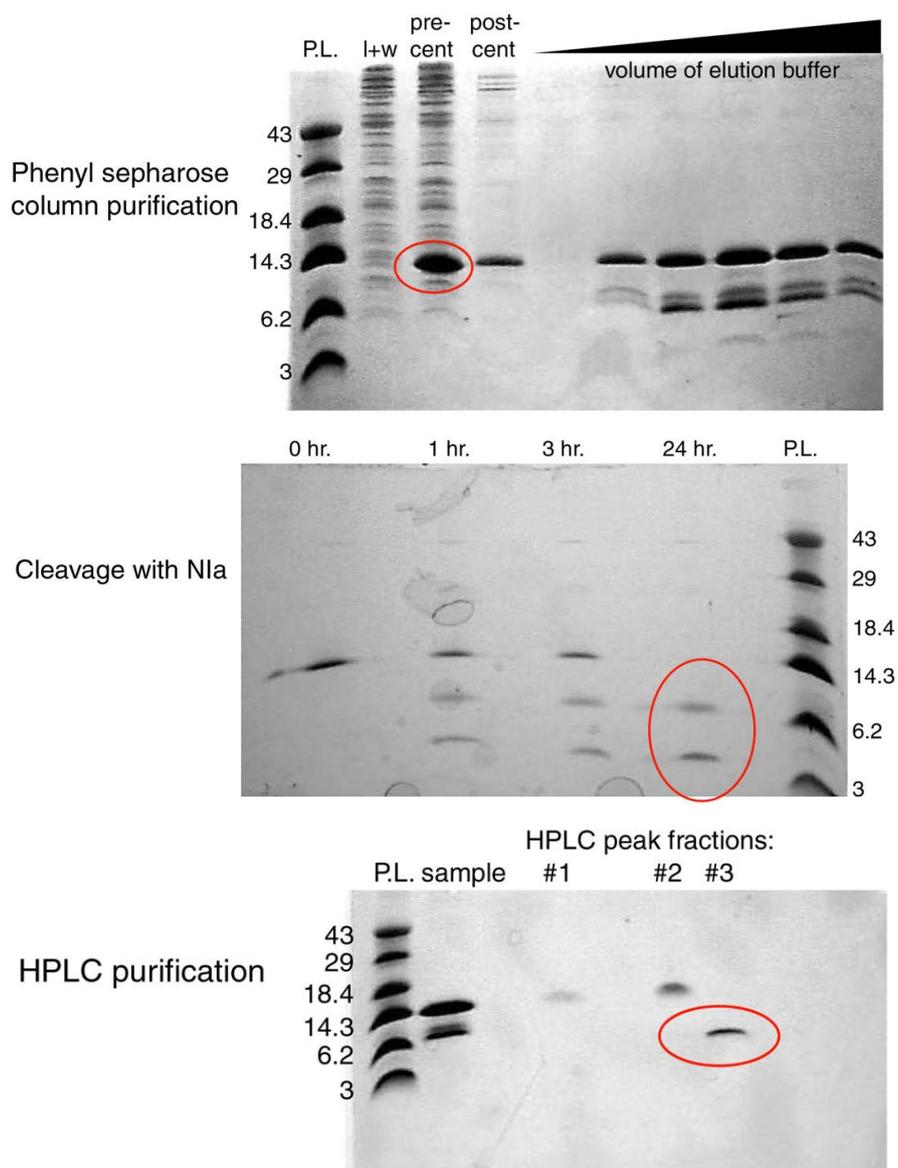
<sup>a</sup>As determined by UV spectroscopy at 280 nm as described in the text. <sup>b</sup>The difference from the calculated molecular weight was thought to be caused by inadequate calibration of the instrument.

below. Quantification of protein expression at each step was not performed; however, qualitative analysis using SDS-PAGE was performed at various steps in the expression and purification process. In a number of cases as indicated in **Table 1(c)**, CD was done on the purified proteins (unpublished results, [9]). Full characterization of one of the designed proteins (SPANS-WW2) by 1-D and 2-D NMR was performed and is published elsewhere [4]. A representative family of gels for each expression and purification step is shown in **Figure 2**.

### 3.2. Unsuccessful Expression and Purification of SPA-WW1, a Designed WW Domain

Several protocols for purification of SPA-WW1, a computationally-designed protein based on the hPin1 WW backbone, were attempted without success. Following our standard protocol, a native phenyl sepharose column was run; unfortunately, all of the target protein was present in the load and wash, indicating that the N-Cam. Y/TEV/SPA-WW1 fusion protein had no discernable affinity to phenyl sepharose under the conditions chosen. An anionic exchange column was attempted, as N-Cam. Y has a high affinity for anionic resin. However, neither a “typical” anionic exchange column, using 50 mM Tris and 50 mM NaCl, pH 7.0 as the low-salt wash, nor a low-salt anionic exchange column (10 mM Tris, 10 mM NaCl, pH 7.0 as the wash), led to adequate purification of the designed protein. An additional anionic exchange column, this time including 6 M urea for denaturation of the proteins, was unsuccessful as well (Supplemental Information, **Figure S1**).

As N-Cam. Y is stable at high temperatures (and therefore a fusion protein containing N-Cam. Y may also be thermostable), a heat gradient experiment involving heating the fusion proteins to various temperatures followed by cen-



**Figure 2. Representative SDS-PAGE protein gels of the scheme utilized to purify WW designs.** (Top) Purification of protein using phenyl sepharose resin. “1 + w” (load and wash of column); “pre-cent” (precentrifugation) and “post-cent” (postcentrifugation) of WW fusion protein (specifically SPANS-WW2 in these images). Black triangle indicates increasing volume of EDTA-containing elution buffer. The red ellipse indicates potential presence of the N-Cam. Y/TEV/WW fusion protein. (Middle) SDS-PAGE of the fusion protein before and after proteolysis by the TEV protease. The red ellipse indicates the presence of the N-Cam. Y/TEV fusion protein (top band) and a WW protein (bottom band). (Bottom) HPLC purification after proteolysis. “sample”, sample loaded onto HPLC column. #1-#3 indicate fractions collected of HPLC peaks. In all images, the resolving gel was 14% acrylamide/0.4% bis-acrylamide; the stacking gel was 0.4% acrylamide/0.1% bis-acrylamide. P. L., protein ladder; values given are approximate molecular weights of bands in kDa.

trifugation was performed. However, there was a significant amount of background proteins in each sample that precluded the useful application of this method. After multiple purification efforts, this design was abandoned. Full

Methods details for these purification schemes are included in the Supplemental Information.

### 3.3. Alternative Cleavage Patterns in Fusion Proteins

As shown in **Table 1(c)**, almost all of the proteins were present as two populations by mass spectrometry of the cleaved purified target proteins. This mass may have been due to proteolysis by a cellular protease at a site located two residues N-terminal to the TEV protease cleavage site; instead of the GS tag, a tag of FQGS N-terminal to the target protein was most likely present. The addition of the FQ residues would have led to a ~270 Da molecular weight increase, which is the difference between the molecular weights as determined by this procedure.

The presence of SPANS-WW2 fusion protein in the fractions obtained from phenyl sepharose purification was confirmed by SDS-PAGE analysis, which also showed that a band with a lower molecular weight, within error of the molecular weight of SPANS-WW2 by itself, was present in the load and wash solutions (**Figure 2 top**). This pointed toward a small amount of basal proteolytic cleavage of the fusion protein prior to addition of the TEV protease. This result indicates basal proteolysis of the fusion protein prior to phenyl sepharose and TEV protease cleavage. The TEV linker was designed to be flexible for protease accessibility; perhaps there was a background cellular protease that was capable of cleaving this linker over time. Whether or not the number of protein molecules containing the extra two residues depended on how promptly the cells were processed was not determined.

### 3.4. Zinc Finger/Fusion Protein Expression and Purification

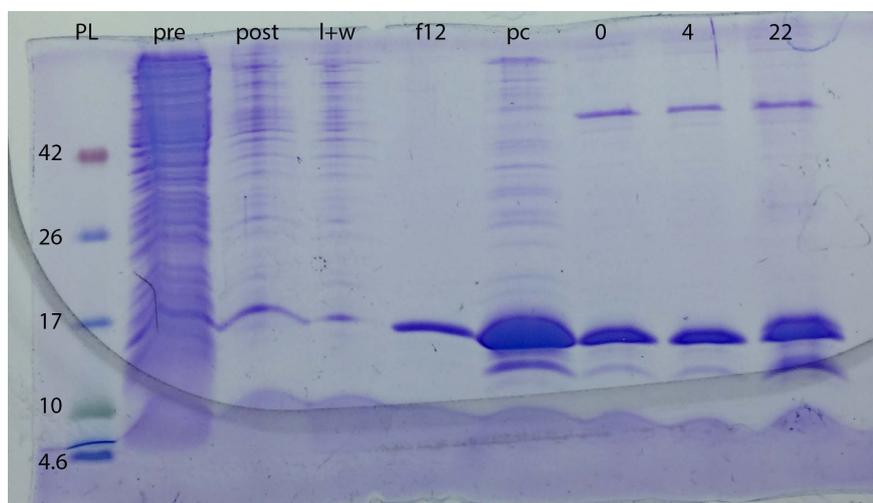
ZF23 has the following sequence:

PGERPFMCTWSYCGKRFRSDDLQRHKRTH

Overnight (~18 h) induction with IPTG resulted in overexpressed protein located at approximately 14 kDa by SDS-PAGE (the molecular weight of the N-Cam. Y/ZF23 protein is calculated to be 13.4 kDa) (**Figure 3**). The band of interest was excised and analyzed by mass spectrometry.

Two peptides from the ZF23 fusion protein were identified by MS based proteomic approach, and the corresponding annotated MS/MS spectra of the two identified peptides are shown in **Figure 4**. The fact that the peptides were identified with good XCorr scores and the tandem mass spectra exhibit a continuous stretch of b- or y-ion series indicate a confident identification of the ZF fusion protein (**Figure 4(a)**). As the identification was made from a partially purified fraction, the ZF fusion protein was identified along with other *E. coli* proteins with a similar molecular weight (15 - 17 kDa) (data included in Supporting Information). Nonetheless, a good sequence coverage of ZF-domain containing protein was obtained (**Figure 4(b)**, 26.3%).

Phenyl sepharose purification of the expressed protein post-centrifugation

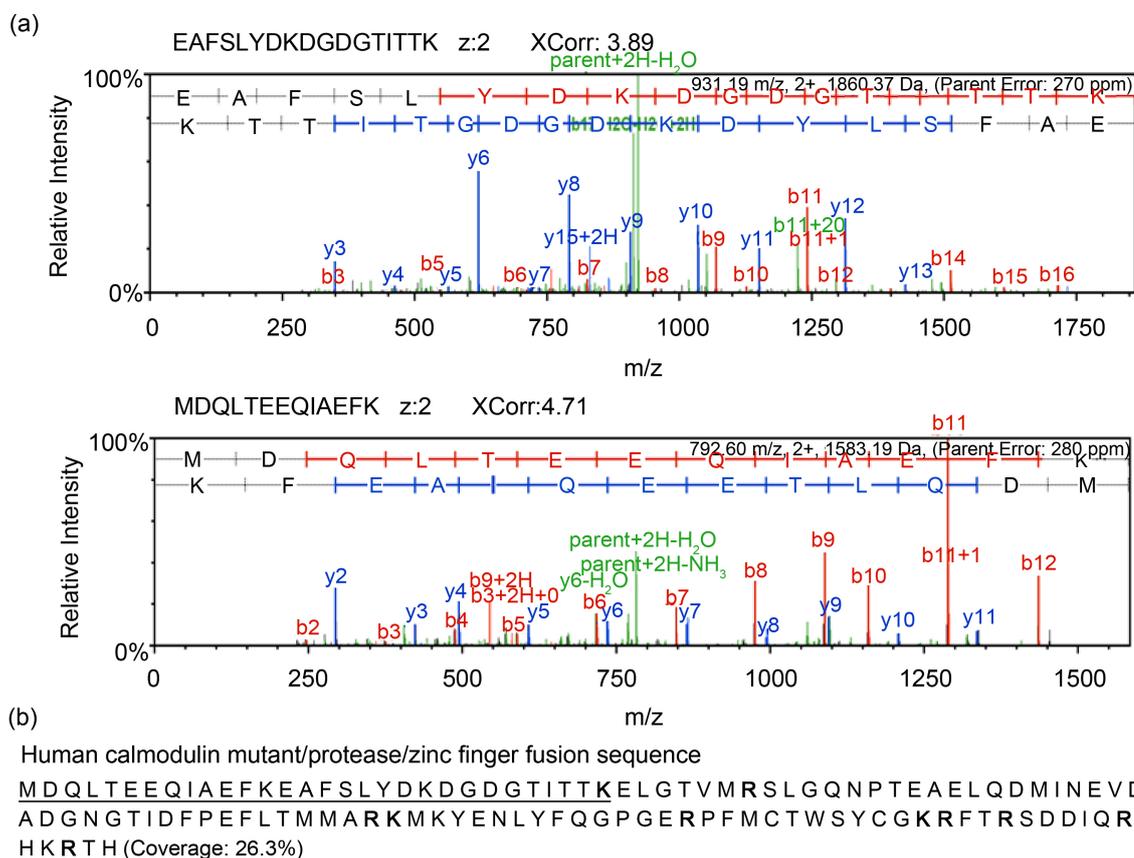


**Figure 3.** Gel of expression, purification, and digestion of N-Cam. Y/TEV/ZF23 construct. PL = protein ladder (MW values are labeled to left of gel); pre is pre-induction; post is post-IPTG induction (after ~22 h); l + w is phenyl sepharose column flow-through and load/wash; f12 is the twelfth fraction collected after EDTA wash started; pc is post-concentration; 0, 4, and 22 are 0-, 4-, and 22-h post-TEV protease addition. The band running at ~15 kDa is the fusion protein; the band running at ~12 kDa is the calmodulin/TEV without ZF23. Note presence of this band in the concentrated sample; background cleavage is therefore present before TEV addition as with the WW results. The large MW (~50 kDa) band in the last three lanes is due to the TEV protease.

resulted in pure protein (better than ~90% as seen by SDS-PAGE analysis). As was seen in the WW results, so much protein was expressed that a small amount was also present in the column flow-through and load/wash because the resin volume was not large enough to bind to all of the protein. The protein eluted in 18 fractions of 500  $\mu$ L each. Digestion with the TEV protease resulted in partial cleavage that increased with time. Complete cleavage was not achieved within 22 h, due most likely to the high concentration of fusion protein. The ZF23 band was not visible on the protein gel due to its size, but the calmodulin band is easily seen as increasing in intensity as the reaction proceeds. A gel of the combined results is seen in **Figure 3**.

#### Potential Drawbacks to the Applicability of This Method

The usefulness of this method, beyond its failure to yield pure SPA-WW1, is quite impressive. It was the only method that provided a significant amount of pure SPANS-WW2 protein for NMR analysis, for example. However, it does have its drawbacks. As shown in **Figure 3** (top), a not-insignificant amount of expressed protein is lost in the insoluble fraction after centrifugation. This is evident when the “pre-cent” and “post-cent” lanes are compared to each other. Although this is a qualitative method, it is obvious that some protein is lost in centrifugation. In addition, because of the spontaneous proteolysis described above, if HPLC is used for purification prior to addition of the protease, some of the target protein may be lost if care is not taken to collect all relevant HPLC peaks. In addition, the TEV activity in the ZF fusion protein appears somewhat



**Figure 4.** Identification of the N-Cam. Y/ZF23 fusion protein. (a) MS/MS spectra of the identified peptides. The product ion mass spectra were searched against the *E. coli* and *Homo sapiens* protein databases using the SEQUEST search engine. The annotated MS/MS spectra of the peptides identified with the highest XCorr score of the ZF23 fusion protein are shown. Z: charge state; XCorr, cross correlation. (b) Sequence coverage of the N-Cam. Y/ZF23 fusion protein. The target protein was identified with a coverage of 26.3%. Arginines and lysines in the sequence are in bold to indicate possible trypsin cleavage sites.

less efficient than in the WW fusion proteins. However, this may be due to concentration of the fusion protein and will be analyzed further in future studies, as use of qualitative methods such as SDS-PAGE can prove problematic in drawing quantitative comparisons.

#### 4. Discussion

The construct described herein, which is based on the N-terminus of a calmodulin mutant, allows for the straightforward expression and purification of a large amount of protein in *E. coli* cells. It has proven successful for the expression and purification of a number of small proteins, including WW domains and a zinc-finger domain. We were unable to obtain several WW proteins—both wild-type and a few designed variants—through any other method, including solid-phase synthesis, and after this system proved useful for the wild-type and initial designed WW proteins, we stopped expression and purification attempts using other systems to focus solely on methods using this fusion system. Use of this system allowed us to obtain enough pure WW for study, including NMR analy-

sis of SPANS-WW2. Indeed, one large-scale (~4 L) expression of SPANS-WW2 provided enough protein for 2-D NMR characterization.

We have also shown that this system appears useful for the expression and purification of a zinc-finger protein, without the addition of zinc ions or other cofactors. Initial protein gel and mass spectrometry results indicate that we have been able to overexpress and purify a large amount of the fusion protein. Future goals include purifying the digested fusion protein using the phenyl sepharose procedure described above to obtain pure ZF23 protein, in parallel with expression and purification of the His-tagged version of the same zinc finger module. The functionality of this protein will be tested by following change in fluorescence with the addition of the DNA binding partner. Because zinc finger modules bind to three-base regions of DNA with high specificity and affinity, we can determine whether the binding region of ZF23 expressed with this fusion protein has been altered by following the fluorescence change upon addition of the target DNA to verify that this is a valid assumption.

This system has also proven useful for the expression of a computationally-designed protein based on the cold shock protein from *Bacillus subtilis* (CspB; unpublished results REF THESIS). An inclusion body preparation was performed using urea and Triton X-100 as per standard protocol. The protein remained soluble when dialyzed into dilute (0.6 and 0.3 M) guanidine solution in TE, but the resultant protein could not be purified by HPLC. A gene with N-Cam. Y-Cys-mutant identity was constructed for 2-nitro-5-thiocyanobenzoic acid cleavage prior to the Cys residue in the presence of chaotropic agent. This system yielded high amounts of soluble protein, which were applied to a phenyl sepharose column in the presence of calcium as discussed above. The protein showed high affinity to the phenyl sepharose resin, but the samples were not very pure; the project was stopped so that the WW studies could be performed, and further computational design studies using the CspB backbone were never attempted in Reference 9.

As we have shown in previous work, the addition of a His tag can be detrimental to protein functionality. Addition of a bulky His tag to zinc-finger protein, which is only 30 amino acids long, could significantly alter or affect its folding and function. This system has the benefit of not only potentially increasing expression of the protein in *E. coli* but also making it easy to purify the protein; our scheme adds only a Gly residue N-terminal to the protein of interest, which may affect folding and/or function to a lesser extent than the His tag. The addition of GS onto the SPANS-WW2 protein did not appear to affect folding or stability. It is hypothesized that this G remaining after the TEV digestion will not alter the overall conformation of the target protein as well, unlike the His tag has been shown to do.

This construct has been developed in such a way as to allow for simple sub-cloning of the gene coding for the protein of interest into the expression system using the flanking unique restriction enzymes. Future goals include testing the usefulness of this method in the expression and purification of other hard-to-

express and various poorly folded or insoluble proteins.

## Acknowledgements

We wish to acknowledge Dr. John R. Desjarlais, whose suggestion to develop this method further led to the successful expression and purification of the WW domains, and who graciously allowed us to take over the project for the zinc-finger studies. We are grateful to Dr. Song Tan (Pennsylvania State University) for a sample of the original TEV cleavage protease used in some of the WW work, to Dr. Jennifer Doudna (University of California, Berkeley) for the plasmid coding the original (His)<sub>6</sub>-tagged TEV cleavage protease mutant used in some of the WW work, and Marissa Smith (SUNY Upstate Medical University) for her assistance with a portion of the ZF project. C. T., D. W., M. A., C. R., and D. B. contributed to the ZF project and were supported by summer research grants from the Vice President of Academic Affairs at St. Michael's College. Work done by C.K-C. on the ZF protein was supported by a summer Faculty Development Committee Expense Defrayment Grant from St. Michael's College. All WW work was supported in part by a National Science Foundation Career Award CH 9876234 awarded to C.K-C.'s thesis advisor (John R. Desjarlais). All protein preparation for the zinc finger mass spectrometry was performed in the VGN Proteomics Facility. The Vermont Genetics Network Proteomics Facility is supported through NIH grant P20GM103449 from the INBRE Program of the National Institute of General Medical Sciences. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

## Conflict of Interest

The authors declare no commercial or financial conflict of interest.

## References

- [1] Chant, A., Kraemer-Pecore, C.M., Watkin, R. and Kneale, G.G. (2005) Attachment of a Histidine Tag to the Minimal Zinc Finger Protein of the *Aspergillus nidulans* Gene Regulatory Protein Area Causes a Conformational Change at the DNA-Binding Site. *Protein Expression and Purification*, **39**, 152-159.
- [2] Ababou, A., Shenvi, R.A. and Desjarlais, J.R. (2001) Long-Range Effects on Calcium Binding and Conformational Change in the N-Domain of Calmodulin. *Biochemistry*, **40**, 12719-12726.
- [3] Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. and Oschkinat, H. (1996) Structure of the WW Domain of a Kinase-Associated Protein Complexed with a Proline-Rich Peptide. *Nature*, **382**, 646-649.  
<https://doi.org/10.1038/382646a0>
- [4] Kraemer-Pecore, C.M., Lecomte, J.T. and Desjarlais, J.R. (2003) A *de Novo* Redesign of the WW Domain. *Protein Science*, **12**, 2194-2205.  
<https://doi.org/10.1110/ps.03190903>
- [5] Desai, U.A., Sur, G., Daunert, S., Babbitt, R. and Li, Q. (2002) Expression and Affinity Purification of Recombinant Proteins from Plants. *Protein Expression and Purification*, **25**, 195-202.

- [6] Li, C.Q., Ye, P., Cao, Z., Wang, H., Lu, L., Nicastro, P., Wood, E., Robert, J.J., Ouweland, W.H., Hill, F., Lopez, J.A. and Wardell, M.R. (2001) Expression of the Amino-Terminal Domain of Platelet Glycoprotein Ib $\alpha$ : Exploitation of a Calmodulin Tag for Determination of Its Functional Activity. *Protein Expression and Purification*, **22**, 200-210.
- [7] Schauer-Vukasinovic, V. and Daunert, S. (1999) Purification of Recombinant Proteins Based on the Interaction between a Phenothiazine-Derivatized Column and a Calmodulin Fusion Tail. *Biotechnology Progress*, **15**, 513-516.  
<https://doi.org/10.1021/bp990058l>
- [8] Hristova, M., Veith, C., Habibovic, A., Lam, Y.W., Deng, B., Geiszt, M., Janssen-Heininger, Y.M. and van der Vliet, A. (2014) Identification of Duox1-Dependent Redox Signaling through Protein S-Glutathionylation in Airway Epithelial Cells. *Redox Biology*, **2**, 436-46. <https://doi.org/10.1016/j.redox.2013.12.030>
- [9] Pecore, C.M.K. (2002) Computational Design and Experimental Characterization of Protein Domains. The Pennsylvania State University, State College.
- [10] Ababou, A. and Desjarlais, J.R. (2001) Solvation Energetics and Conformational Change in EF-Hand Proteins. *Protein Science*, **10**, 301-312.  
<https://doi.org/10.1110/ps.33601>

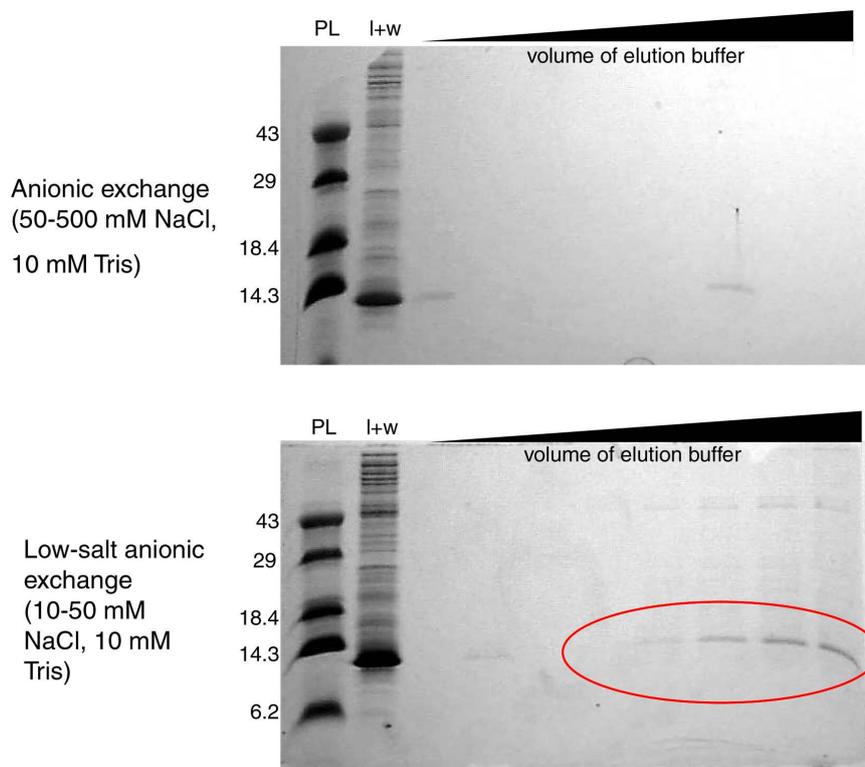
## Supporting Information

Another purification method, used only for SPA-WW1, involved use of an anionic exchange column with QAE Sephadex A-25 (Pharmacia) as the resin. After centrifugation of the post-induction sample was complete, the supernatant was again diluted two-fold, this time with a low-salt buffer (50 mM Tris base, pH 7.5; 50 mM NaCl). The anionic exchange column was loaded and washed with the low-salt buffer until the absorbance returned to the pre-load value. A 50% gradient of increasing high salt concentration was run using the low-salt buffer and a high-salt buffer (50 mM Tris base, pH 7.5; 1 M NaCl) to a maximum salt concentration of 25 mM Tris and 500 mM NaCl. When this yielded no useful results, a lower-salt concentration anionic exchange column was attempted to maximize the affinity of protein for the column resin; this involved using the previously described low-salt buffer in place of the high-salt buffer in the last procedure, and using a lower-salt buffer (10 mM Tris base, pH 7.5; 10 mM NaCl) as the loading and washing buffer. A denaturing anionic exchange column was also attempted, wherein 6 M urea was added to both the low- and high-salt buffers and the column was run as described.

Another method did not utilize column chromatography at all. Because of the high  $T_m$  and thermal stability of N-Cam. Y, a thermal stability experiment was performed to attempt to purify, at least partially, one of the designed WW fusion proteins. This involved centrifuging small volumes of the post-induction supernatant to remove any precipitates, heating the samples at the desired temperature (which ranged from 40°C to 90°C) for 10 minutes, centrifuging the samples for three minutes, removing the supernatant, discarding the pellets, and running an SDS-PAGE protein gel. The goal of this step was to eliminate, by unfolding and precipitation, the majority of the proteins less thermally stable than N-Cam. Y. It was hoped that upon centrifugation, the background proteins formed a pellet at the bottom of the tube, whereas the N-Cam. Y fusion protein would remain soluble. If the thermal stability purification proved successful, HPLC or other methods could have been performed to further purify the fusion protein away from the background proteins.

### ZF23 Fusion Protein Identification by Liquid Chromatography-Tandem mass Spectrometry (LC-MS/MS)

In brief, Coomassie Blue stained gel bands were destained with 50% ACN/50 mM  $\text{NH}_4\text{HCO}_3$  for 2 - 3 h. After reduction by 10 mM dithiothreitol and alkylation with 50 mM iodoacetamide, the gel pieces were washed with 50 mM  $\text{NH}_4\text{HCO}_3$ . They were then minced, dehydrated with ACN, dried in a SpeedVac, and subjected to digestion with trypsin (Promega, Madison, WI) for 18 h at 37°C. Peptides were extracted successively with 5% FA/50% ACN and 100% ACN, dried and kept in the freezer until LC-MS/MS analysis. The dried digests were re-suspended in 20  $\mu\text{l}$  of 2.5% ACN and 2.5% FA in water and 4  $\mu\text{l}$  of the digest were analyzed by LC-MS/MS on a LTQ mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto a 100  $\mu\text{m}$   $\times$  120 mm capillary fused



**Figure S1.** (Top) SDS-PAGE protein gel of the attempted purification of SPA-WW1 using anionic exchange. Note thick band in load + wash (I + w) was protein of interest, exhibiting low binding affinity for the resin. (Bottom) SDS-PAGE protein gel of the attempted purification of SPA-WW1 using anionic exchange with a reduced ionic strength. A gain, thick band in load + wash was protein of interest; circled bands indicate slightly increased binding affinity of protein to resin in lowered salt concentration.

silica column packed with MAGIC C18 (5  $\mu\text{m}$  particle size, 20 nm pore size, Michrom Bioresources, CA) at a flow rate of 500 nL/min. Peptides were separated by a gradient of 5% - 35% ACN/0.1% formic acid over 31 min, 35% - 100% ACN/0.1% FA in 1 min, and 100% ACN/0.1% FA for 11 min, followed by an immediate return to 2.5% ACN/0.1% FA and an isocratic hold at 2.5% ACN/0.1% FA until the next injection. Peptides were introduced into the linear ion trap via a nanospray ionization source and a laser pulled  $\sim 3$   $\mu\text{m}$  orifice with a spray voltage of 1.8 kV. Mass spectrometry data were acquired in a data-dependent "Top 10" acquisition mode, in which a survey scan from  $m/z$  360 - 2000 is followed by 10 CID tandem mass spectrometry MS/MS scans of the most abundant ions. MS/MS scans were acquired with the following parameters: isolation width: 2  $m/z$ , normalized collision energy: 35%, Activation Q: 0.250 and activation time = 30 ms. Dynamic exclusion was enabled (repeat count: 2; repeat duration: 30 sec; exclusion list size: 500; exclusion duration: 60 sec). The minimum threshold was 500.

Product ion spectra were searched using the SEQUEST HT engine implemented on the Proteome Discoverer 1.4 (Thermo Fisher Scientific, Waltham, MA, USA) against a curated Uniprot *Escherichia coli* (UP000019194) and *Homo sapiens* (UP000005640) database (the May 14, 2016 release was downloaded) with the

sequence of the N-Cam. Y/ZF23 fusion protein included. Search parameters were as follows: 1) Full trypsin enzymatic activity, 2) Two missed cleavages, 3) Peptides between the MW of 350 - 5000, 4) Mass tolerance at 2 Da for precursor ions and 0.8 Da for fragment ions, 5) Dynamic modifications on methionine (+15.9949 Da: oxidation), 6) 4 maximum dynamic modifications allowed per peptide; and 7) Static modification on cysteine (+57.0215 Da: carbamidomethylation). The result files were then further analyzed by Scaffold 4.3 (Proteome Software, Portland, OR, USA). The annotated MS/MS spectra of the peptides from the ZF23 fusion protein that were identified by SEQUEST HT were manually evaluated.



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [abc@scirp.org](mailto:abc@scirp.org)