Published in final edited form as:

Protein Expr Purif. 2012 October ; 85(2): 187–199. doi:10.1016/j.pep.2012.08.003.

Engineering of an elastic scaffolding polyprotein based on an SH3-binding intrinsically disordered titin PEVK module

Wanxia Li Tsai^a, Jeffrey G. Forbes^b, and Kuan Wang^c

Muscle Proteomics and Nanotechnology Section, Laboratory of Muscle Biology, NIAMS, NIH, DHHS, USA 20892-8024

Abstract

Titin is a large elastic protein found in muscle that maintains the elasticity and structural integrity of the sarcomere. The PEVK region of titin is intrinsically disordered, highly elastic and serves as a hub to bind signaling proteins. Systematic investigation of the structure and affinity profile of the PEVK region will provide important information about the functions of titin. Since PEVK is highly heterogeneous due to extensive differential splicing from more than one hundred exons, we engineered and expressed polyproteins that consist of a defined number of identical single exon modules. These customized polyproteins reduce heterogeneity, amplify interactions of less dominant modules, and most importantly, provide tags for atomic force microscopy and allow more readily interpretable data from single-molecule techniques.

Expression and purification of recombinant polyprotein with repeat regions presented many technical challenges: recombination events in tandem repeats of identical DNA sequences exacerbated by high GC content, toxicity of polymer plasmid and expressed protein to the bacteria; early truncation of proteins expressed with different numbers of modules; and extreme sensitivity to proteolysis. We have investigated a number of *in vitro* and *in vivo* bacterial and yeast expression systems, as well as baculoviral systems as potential solutions to these problems. We successfully expressed and purified in gram quantities a polyprotein derived from human titin exon 172 using *Pichia pastoris* yeast. This study provides valuable insights into the technical challenges regarding the engineering and purification of a tandem repeat sequence of an intrinsically disordered biopolymer.

Keywords

tandem repeat biopolymer; functionalized elastomer

Introduction

Intrinsically disordered and repetitive regions are abundant in the genomes of many organisms [1–3]. These regions have important functions in the regulation of cell signal pathways [4–7]. They have been implicated in many illnesses such as cancer, diabetes and

© 2012 Elsevier Inc. All rights reserved.

^cCorresponding address: Institute of Chemistry, Academia Sinica, 128 Academia Rd. Sec.2, Nankang, Taipei 115, Taiwan. Wangk@gate.sinica.edu.tw. Phone No. 011-886-2-2789-8500. Fax No. 011-886-2-2651-1488.

^aPresent address: Building 10, 6D46, Bethesda, MD 20892-8024, USA. liwan@mail.nih.gov

^bPresent address: TWIC, Inc, 4423 Lehigh Ave, #172, College Park, MD 20740, USA. jgf@twicinc.com

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

neurodegenerative diseases [8–13]. Therefore, it becomes increasingly important to study the functions of these repetitive and intrinsically disordered regions.

Titin is a large elastic muscle protein (3–4 MDa) in skeletal and cardiac muscles. It consists mainly of numerous immunoglobulin-like domains, fibronectin type III domains, and the PEVK region [14]. Titin plays a critical role in maintaining the stability of sarcomeres and assembling nascent sarcomeres during the development of muscle cells [15, 16]. Recent studies have revealed the role of the titin kinase and the PEVK region in the signaling pathways that regulate muscle function [17, 18]. The PEVK region of titin is expressed by differential splicing from a large number of exons, e.g. exons 109–224 in humans, and consists of tandem repeats of numerous 28-amino acid motifs enriched in proline, glutamic acid, valine, and lysine. The PEVK region of titin has the unique function of regulating sarcomere tension upon stretching [19, 20].

We have studied the physical and biochemical characteristics of synthetic modules and cloned fragments of the PEVK region of titin extensively [21, 22]. Our studies indicated that the PEVK region of titin is highly charged with gel-like, intrinsically disordered structures [23, 24]. The PEVK region of titin behaves like a spring where part of its elasticity may be due to charge-charge interactions between side chains of the molecules, since the persistence length of the region increases with an elevation in ionic strength [21, 25, 26]. We also discovered that the left handed PPII helices in many PEVK modules of titin bind to Src homolog domains (SH3) of nebulin and other signaling proteins [27, 28]. Such PPII sequences thus may mediate interactions between nebulin and titin [29–31].

To further understand the possible role of PPII in the PEVK region in the signaling functions and mechanical properties of titin, we engineered a recombinant polypeptide with 15 identical repeat modules to facilitate purification and immobilization. This single module consisted of 28 amino acids (PEPPKEVVPEKKAPVAPPKKPEVPPVKV) derived from the human titin exon 172. This engineered protein, designated as [PPE172₁₅] contains 15 tandem PEVK repeats and is tagged at both ends with his₆ tag, S-protein tag, and biotin protein ligase (BirA) [32] on the N-terminus and a fluorescein arsenical hairpin binder (FlAsH) tag [33, 34] on the C-terminus of the polymer. The defined number of identical modules in [PPE172₁₅] reduces heterogeneity of its behavior and allows for more readily interpretable binding interaction measurements. The multiple affinity tags on [PPE172₁₅] can facilitate the investigation of force-related protein interactions and provide more readily interpretable data from single-molecule techniques such as atomic force microscopy (AFM).

The cloning and expression of a tandem repeat sequence in large quantities was technically challenging [35–40]. The tandem repeats in the DNA sequence coding for [PPE172₁₅] caused many problems in engineering the plasmids. The repetitive sequences with high GC content were difficult to digest by restriction enzymes and amplify by PCR reactions. The newly cloned plasmids were not stable; random recombination events and deletions of nucleotides readily occurred at all the stages of cloning. The polymer plasmids and expressed protein were also toxic to the bacteria. [PPE172₁₅] was expressed in truncated forms or with different numbers of modules. Fig. 1 illustrates the major steps and difficulties encountered in the process of cloning and expression of [PPE172₁₅]. We compared different methods of protein expression and overcame many technical challenges during the process of engineering the tandem repeat sequence. We further studied the binding and biochemical characteristics of the newly engineered polypeptide. Therefore, this study offers practical details and guidance on the purification and expression of tandem repeats and intrinsically disordered recombinant proteins.

Materials and Methods

Materials

The *Pichia* expression system, *BL21(DE3) pLysS* and *Stbl2* competent cells, and Gateway cloning technology were obtained from Invitrogen. His-tag monoclonal antibody, pET-30a vector, and S-protein agarose were purchased from Novagen. Rapid translation system (RTS 500) *E. coli* HY kit, and complete mini EDTA-free protease inhibitor cocktail tablet were purchased from Roche Molecular Biochemicals. Exo mung bean deletion kit and quickchange XL site-directed mutagenesis kit were purchased from Stratagene. HRP-rabbit anti-mouse IgG was obtained from Zymed Laboratories Inc. Wetware FIAsh protein labeling kit was purchased from Aurora Biosciences Corporation, San Diego. The nickel sepharose 6 fast flow column resin was purchased from GE Amersham Biosciences. The Affi-gel 10 was purchased from Bio-Rad.

Methods

In vivo expression and purification of the [PPE172₁₅]

***E. coli* expression system:** The polyprotein vector containing the 15 monomers of titin exon 172 [PPE172₁₅] with FIAsh and BirA tags in pENTR11 entry vector was purified and digested with *NotI* and *NcoI* restriction endonucleases, and ligated into the pET-30a(+) protein expression vector containing the kanamycin-resistance marker. This expression vector, designated as pET30a-[PPE172₁₅], was transformed into *E. coli BL21(DE3) pLysS* cells for expression. Single colonies were inoculated into four 50 ml tubes of LB medium with 25 µg/ml chloramphenicol, 40 µg/ml kanamycin and 1% glucose, and then incubated at 30 °C overnight. The 50 ml cultures were transferred into 8 L of the same medium and incubated at 30 °C with a shaker at 200 rpm until the OD reached A₅₅₀ = 0.6–0.8. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 1 mM was used to induce protein expression for 4 hours. The bacteria were collected by centrifugation at 6500 × *g* for 10 min and lysed in a French press cell (3 × 1500 psi) in lysis buffer (20 mM sodium phosphate, 0.3 M NaCl, 5 mM imidazole, 10 µg/ml leupeptin, and 1 complete EDTA-free protease inhibitor cocktail tablet per 50 ml buffer pH 8.0). The cell lysate was centrifuged at 45,000 × *g* for 15 min at 4 °C and the supernatant was loaded to a nickel-chelating sepharose column (Amersham Pharmacia Biotech) equilibrated with 20 mM sodium phosphate, 0.3 M NaCl pH 8.0. The column was washed with 10 volumes of column buffer and the [PPE172₁₅] was eluted with 0.25 M imidazole, 20 mM sodium phosphate, 0.3 M NaCl pH 8.0. The eluent was then applied to an S-protein agarose (Novagen) column in 20 mM tris-HCl, 150 mM NaCl pH 7.5 and the S-protein column was washed with 25 volumes of column buffer and eluted with 7 M guanidine chloride in the same column buffer. Due to the multiple cysteine amino acids on the FIAsh tag, the eluted [PPE172₁₅] was reduced with 40 times molar excess with tris (2-carboxyethyl) phosphine hydrochloride (TCEP) under nitrogen gas at 4 °C for 60 min. The reduced [PPE 172₁₅] was loaded onto an Affi-gel 501 organomercurial agarose column in 20 mM tris-HCl, 1 mM EDTA, 0.5 M NaCl pH 8.0. The column was washed with 25 column-volume buffer and eluted with 100 mM 1,4-dithiothreitol in the same buffer with pH 8.0. The eluted [PPE172₁₅] was quickly frozen in liquid nitrogen and stored at –80 °C. The total yield of purified protein was about 2 mg/liter of culture medium and was measured spectrophotometrically (extinction coefficient at 280 nm is 0.21 for 1 mg/ml).

Baculovirus expression system: The baculovirus expression was conducted by Kemp Biotechnologies Inc. (Frederick, MD). Briefly, the [PPE172₁₅] insert was excised from plasmid at the *EcoRI* and *NdeI* sites and ligated into the pVL1393 transfer vector with *EcoRI* and *NotI* sites to construct pVL1393-[PPE172₁₅] transfer vector. The pVL1393-[PPE172₁₅] plasmids were amplified in *E. coli* and further purified. The pVL1393-[PPE172₁₅] vectors were co-transfected with linearized DNA of *Autographa californica*

nuclear polyhedrosis virus (AcNPV) in the Sf9 insect cells (*Spodoptera frugiperda*). The homologous recombinant virus was then amplified by infecting fresh Sf9 insect cells at 0.3% (v/v) and 3% (v/v) multiplicity of infection (MOI) for 0, 24, 48 and 72 hrs. The cell lysates were analyzed with SDS-PAGE gels to check the [PPE172₁₅] expression.

Cell-free expression and purification of polyprotein [PPE172₁₅]: Roche rapid translation system *RTS-500* was used to express pET30a-[PPE172₁₅] plasmid. The feeding and reaction solutions were provided in the expression kit (Roche Applied Science) and prepared according to the product manual. Briefly, the reaction solution included *E. coli* lysate, amino acid mix, and reaction buffer with 15 µg pET30a-[PPE172₁₅] plasmid DNA. The feeding solution had feeding mix, amino acid mix, and methionine. One complete mini EDTA free protease inhibitor cocktail tablet (Roche Applied Science) was added to 11 ml feeding solution. The reaction and feeding solutions were loaded into reaction and feeding compartments, respectively. The reaction was carried out in the RTS Proteomaster instrument at 30 °C with 150 rpm for 17–24 hrs. The reaction solution was taken out of the compartment and centrifuged at 10,000 × *g* for 5 min at 4 °C. The supernatant was mixed with 0.5 ml nickel-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen) in 50 mM sodium phosphate, 0.3 M NaCl pH 8.0. The sample mixture was rotated at 4 °C for 60 min and packed into an econo-column (Bio-Rad). The resin was washed with 20 columns volume of 50 mM sodium phosphate, 5 mM imidazole, 5 mM β-mercaptoethanol, 1.5 M NaCl pH 8.0, and 50 mM sodium phosphate, 0.3 M NaCl, 20 mM imidazole pH 8.0, respectively. The polyprotein was eluted with 0.25 M imidazole, 50 mM sodium phosphate, and 0.3 M NaCl pH 8.0. Pooled fractions were treated at eight times molar excess of protein concentration with TCEP under nitrogen gas at 4 °C for 60 min. The pool was applied to an organomercurial agarose Affi-gel 501 column (1.0 × 5 cm; mercury content is 2.51 µmol/ml) equilibrated with 20 mM tris-HCl, 0.5 M NaCl, 1 mM EDTA pH 8.0. The column was washed with 20 volumes of column buffer, and the polyprotein was eluted with 100 mM DTT in the same buffer pH 8.0. The fractions were quickly frozen in liquid nitrogen and stored in –80 °C. The total yield of *RTS-500* per reaction was about 1 mg.

Yeast expression system and large scale expression: *Pichia pastoris* strains *X-33* and *GS115* were used as hosts for expression. The pPICZαA and pPICZ, which contains the zeocin resistance genes and alcohol oxidase (AOX1) promoters with and without α-factor secretion signal, were used as the expression vectors (Invitrogen). The pPICZαA and pPICZ vectors were digested with restriction endonuclease *EcoRI* and pET30a-[PPE172₁₅] was digested with *NdeI*. The 3′-overhang DNA created by enzyme digestion was blunt-ended with 1 U of Klenow in NEBuffer2. The blunt ends of the vectors were ligated and transformed into *E. coli Top 10* for propagation. The linearized pPICZαA-[PPE172₁₅] (10 µg) and pPICZ were electroporated into *Pichia pastoris X-33* and *GS115* with a 0.2 cm cuvette containing 80 µl pre-washed yeast cells at a voltage of 2.0 kV using Bio-rad Micropulser pre-programmed “pic” setting. 1 ml of 1 M sorbitol was immediately added to cuvettes and incubated at 30 °C for 60 min. The mixture was spread onto YPDS plates 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, 1 M sorbitol, and 1.5% (w/v) bacteriological agar with 100 µg/ml zeocin and incubated at 30 °C for 3 days. The methanol utilization plus (Mut⁺) phenotype of the *Pichia pastoris X-33* and *GS115* were determined using minimal dextrose histidine (MDH) and minimal methanol histidine (MMH) plates. The Mut⁺ and Mut^s transformants were selected by picking single colonies from YPDS plates containing the zeo^R and patched onto MDH and MMH plates since Mut^s transformants do not produce alcohol oxidase and cannot metabolize methanol as a carbon source. After 2 days incubation at 30 °C, *Pichia* yeast cells grown normally on both YPDS and MMH plates were selected and used for the expression host cells.

The *Pichia* yeast cells *X-33* with pPICZαA-[PPE172₁₅] were scaled up and expressed with the fermentation process in the 10 L bench-top fermentor vessel (New Brunswick BioFlo 3000). The components of the fermentation basal salts medium with PTM₁ trace salts were as described [41]. A single colony was picked from YPDS plates (100 μg/ml zeocin) and inoculated into 325 ml MGY medium with shaking at 250 rpm and at 30 °C for 24 hrs. The log phase medium was injected into 7 L basal salts medium with PTM₁. The OD 550_{nm} and the dissolved oxygen were monitored. 50% glycerol with 12 ml/L PTM₁ was added at 1 ml/min for 72 hours until the cell mass reached 252 g/L wet weight. The protein expression was induced by adding methanol at 0.8 ml/min for 24 hrs. The pH was adjusted by adding 5 M potassium hydroxide or ammonium hydroxide during the fermentation process.

The [PPE172₁₅] was secreted into growth medium and separated from the yeast cells by centrifugation at 6500 × *g* for 10 min and then filtered through a 0.2 μm filter. The [PPE172₁₅] was precipitated with 50% ammonium sulfate at 4 °C for 30 min and centrifuged at 18,000 × *g* for 10 min. The pellet was suspended into 10 mM potassium phosphate pH 7.8 and dialyzed into the same buffer at 4 °C overnight. The [PPE172₁₅] was loaded onto a hydroxylapatite column equilibrated in 10 mM potassium phosphate pH 7.8. The column was washed with 10 volumes of the same buffer and the [PPE172₁₅] was eluted with 1 M potassium phosphate pH 7.8. The yield was over 1 gram/L. For the N¹⁵ labeled [PPE172₁₅], FM22 media [41] with 10 g/l (15NH₄)₂SO₇ (Cambridge Isotope Laboratories, Inc) was used and processed as described above.

The [PPE172₁₅] was digested with *PNGase F* (BioLabs) to check the glycosylation of the yeast expressed proteins according to the manufacturer's instructions. Briefly, different amounts of the [PPE172₁₅] (10, 18, and 36 μg) were digested with 10 units *PNGase F* for 1 hour at 37 °C in G7 reaction buffer, respectively. The digestion mixtures were visualized with 4–12% Bis-Tris NuPAGE gel.

Affinity purification of the [PPE172₁₅] with organomercurial agarose gel—The Affi-gel 501 organomercurial agarose was prepared according to manufacturer's protocol (Bio-Rad). Briefly, the Affi-gel 10 gel resin was washed with 3 volumes of anhydrous isopropyl alcohol and then reacted with 1.15% of p-aminophenylmercuric acetate in dimethylformamide for 4 hours at the ambient temperature. The unreacted succinimide groups were blocked with 0.8% (v/v) of ethanolamine HCl, pH 8.0 for 1 hour at the same temperature. The agarose gel was washed with 2 volumes of dimethylformamide and 6 volumes of anhydrous isopropyl alcohol. The organomercurial agarose gel resins were experimentally determined to have 8 mg/ml of bovine hemoglobin binding capacity with 2.51 μmoles Hg/ml.

Generating short segments of the [PPE172₁₅] DNA sequence for the plasmid sequencing—The [PPE172₁₅] plasmid was sequenced from both ends of the polymer insert. The pET30a-[PPE172₁₅] vector was first digested with *KpnI* and *NcoI* restriction endonuclease, then using the *Exonuclease III* to digest the 3' hydroxyl terminal of double strand DNA for 1 to 5 min. The reaction was stopped by deactivating the *Exonuclease III* at 68 °C for 15 min. The remaining undigested single strand DNA extensions were digested with *Mung bean nuclease* at 30 °C for 30 min. The deleted vectors at different time points were purified by 0.2% SDS, 25 mM tris-HCl, 0.8 M lithium chloride pH 8.8, and 2.5 parts of 1:1 phenol-chloroform. The mixture was spun at 16,000 × *g* for 5 min with an Eppendorf 5415C centrifuge, a final concentration of 0.2 M sodium acetate and 60% cold ethanol added to the supernatant, and the mixture placed on dry ice for 10 min. The mixture was then spun at 16,000 × *g* for 20 min. 80% ethanol was added to the pellet, spun again, and the pellet air dried. The purified DNA was suspended into 10 mM tris-HCl, 1 mM EDTA pH 8.0 and run on a 1% low melting point agarose gel. The different lengths of deleted pET30a-

[PPE172₁₅] inserts were gel purified and ligated with T₄ ligase. The recirculated pET30a-[PPE172₁₅] vectors were transformed into *E. coli* *Stbl2* competent cells and grown overnight on Luria-Bertani (LB) medium agar plates with kanamycin. The single colonies were picked and plasmids were purified by the miniprep procedure (Qiagen). The purified DNA was sequenced by GENEWIZ (South Plainfield, NJ).

Western blot and FIAsh labeling for the tag detection—Polyacrylamide gel electrophoresis was performed in 4–12% NuPAGE Bis-Tris gel with MES SDS running buffer (Invitrogen). The [PPE172₁₅] was transferred electrophoretically onto a PVDF membrane (Invitrogen) using a TE22 Mighty Small Transphor tank transfer unit (Amersham Biosciences) at 50 V constant voltage for 3 hours at 4 °C. NuPAGE transfer buffer with 10% methanol was used.

Western blotting was conducted by using monoclonal antibody against rabbit skeletal muscle titin (RT11) and his-tag monoclonal antibodies, and was detected by horseradish peroxidase-conjugated secondary antibodies (rabbit anti-mouse IgG, IgM and IgA (H + L)). Titin antibody RT11 reacted positively with titin exon 172 PEVK module in the [PPE172₁₅]. S-protein HRP conjugate was used for detection of S-tag on the expressed proteins. Affinity purified rabbit polyclonal antibody (IgG) against human nebulin SH3 C-terminus peptide (VQRTGRTGMLPANYVEC) [42] was used to detect nSH3.

The tetracysteine Tag (CCXXCC) was detected with the FIAsh reagent. The [PPE172₁₅] was treated with Laemmli sample buffer (30 mM tris-HCl, 1% SDS, 12.5% glycerol pH 6.8 plus 50 mM tris (2-carboxyethyl) phosphine hydrochloride (TCEP) at 99 °C for 3 min and cooled down before adding FIAsh dye (final concentration 1.5 μM). The reaction was overnight at 4 °C in the dark. The [PPE172₁₅] samples were separated in 4–12% Bis-Tris gel with MES SDS running buffer (Invitrogen), and scanned with Molecular Dynamics Storm 860 (excited at 450 nm with emission at 535 nm).

Circular dichroism spectroscopy—The circular dichroism (CD) spectra were recorded on a JASCO-J-600 spectropolarimeter (Easton, MD), and the detailed procedure has been described [28]. Briefly, yeast expressed [PPE172₁₅] at a concentration of 0.289 mg/ml (5.35 μM) in 20 mM potassium phosphate pH 7.0, was measured at 4, 25, 50, 75 °C, respectively in a cuvette with a 0.1 cm path length. Thermal titration of the [PPE172₁₅] was monitored by following the change in CD spectra at 201 nm between 4 °C to 75 °C.

Gel filtration chromatography—The elution behavior of *E. coli* expressed [PPE172₁₅] in several types of gel filtration chromatography was investigated. Gel filtration resins of Sephadex G-25, G-50, G75, Hitrap desalting column (GE Healthcare), and Bio-gel P4 polyacrylamide gel (Bio-Rad) were equilibrated in 20 mM tris-HCl, 0.5 M NaCl pH 7.5. The [PPE172₁₅] was dialyzed into the same buffer plus 2 mM TCEP and 1 mM EDTA. The [PPE172₁₅] was centrifuged at 10,000 × *g* for 20 min at 4 °C prior to loading.

The K_{av} of the [PPE172₁₅] was calculated according to $K_{av} = (V_e - V_o) / (V_t - V_o)$ where V_o is the void volume, V_t is the total volume of the column, and V_e is the elution volume of [PPE172₁₅]. The high salt in the column buffer was used to minimize nonspecific ionic interactions between the [PPE172₁₅] and the column resins.

Limited digestion by Enterokinase—The yeast expressed [PPE172₁₅] was digested with recombinant *Enterokinase* (BioLabs) in 10 mM potassium phosphate, 150 mM KCl, 10 mM dithiothreitol (DTT) pH 7.0 with 0.004% (*w/w*) and 0.04% (*w/w*) protein to *Enterokinase* ratio. The reaction was carried out at room temperature at 1, 3, 6, and 10 hours, and was terminated with 25 mM phenylmethanesulfonylfluoride (PMSF). The digested

[PPE172₁₅] was treated with Laemmli sample buffer with DTT and boiled in a water bath for 1 min. The protein samples were loaded onto NuPAGE 4–12% Bis-Tris gel and run in the NuPAGE MES SDS running buffer (Invitrogen) at 100V constant voltage for 1 hour.

Enzyme linked immunosorbent assay—The yeast expressed [PPE172₁₅] was adsorbed to microtiter plates (50 μ l at 10 μ g/ml in 20 mM sodium phosphate, 1 mM EDTA, 150 mM NaCl pH 7.0) overnight at 4 °C. The [PPE172₁₅] coated plates were washed three times with TBS-T and blocked with 0.2% bovine serum albumin in TBS-T for 1 hour at 4 °C and 25 °C, respectively. 50 μ l nebulin SH3 (nSH3) from 0 to 80 μ M in 20 mM potassium phosphate, 50 mM NaCl pH 7 was then added to the [PPE172₁₅] coated plates. The plates were then incubated for 1 hour at either 4 °C or 25 °C. After washing the plates three times with TBS-T, rabbit polyclonal anti-peptide antibodies to nebulin SH3 and mouse monoclonal anti-titin antibody (RT11) were incubated for 1 hour at 4 °C or 25 °C in 0.2% BSA in TBS-T. Plates were washed three times with TBS-T and then incubated with peroxidase-conjugated rabbit anti-mouse antibody or goat anti-rabbit antibody (Zymed, S. San Francisco, CA) in the same blocking buffer for 1 hour at the same temperature. The binding was detected by adding ABTS-H₂O₂ substrate in 100 mM citrate buffer pH 4.2 and absorbance measured at 405 nm (Molecular Dynamics Biolumin 960).

Surface plasmon resonance spectroscopy—The binding of nebulin SH3 (nSH3) to yeast expressed [PPE172₁₅] was studied by surface plasmon resonance with a Biacore 1000 (GE Healthcare) instrument. The [PPE172₁₅] in 10 mM potassium phosphate buffer pH 7.0 was immobilized onto CM5 chips (GE Healthcare). The surfaces were activated with 1:1 molar ratio of 0.4 M 1-ethyl-3-(3-dimethylpropyl)-carbodiimide to 0.1 M N-hydroxysuccinimide at flow rate 10 μ l/min for 10 min. The [PPE172₁₅] was then immobilized onto the surface until there were no changes in the RU reading. The excess reactive sites on the surface were deactivated by injecting 1 M ethanolamine-HCl, pH 8.0 at 10 μ l/min for 7 min. Then nSH3 in running buffer (10 mM potassium phosphate, 150 mM KCl pH 7.0) was injected at 5 μ l/min onto the surface of the chip at room temperature. The concentration was increased stepwise from 0.0025 to 3 mM; while the increase in resonance unit (RU) was monitored. The binding constant was calculated from the titration curve by subtracting the baseline RU from the RU at the different concentrations of nSH3.

The binding constants of the ELISA and SPR were calculated according to a 3 parameter single rectangular hyperbola regression using Sigmaplot 9.0. The equation of $y = y_0 + (ax / b + x)$ was used, where b is the K_d .

Results

Engineering the polyprotein [PPE 172₁₅] vector

We engineered a recombinant polyprotein construct, the [PPE172₁₅], which consists of a defined number and order of identical modules. An 84-base pair DNA oligonucleotide from titin PEVK region at exon172 (5'-ccagagcctcccaagaagtagttcctgaaaagaagcgccagtggctcctcctaaaaagcctgaagtccacctgttaaagtg) was synthesized, and used as a single monomer unit for [PPE172₁₅] vector construction. Synthetic DNA oligonucleotides were filled in with polymerase and cloned into pbluescript vector. The monomer sequence with *EarI* site was annealed specifically and seamlessly into the polymer sequence [43–45], which was further ligated into pbluescript via *EarI* sites. The ligation product was cloned into *DH5a* and colonies selected. The longest clone selected had a total of 15 monomers.

The FLAsH binding tetracysteine sequence (CCXXCC) [33] was cloned into pENTR11 (Invitrogen) entry vector between the end of 15 monomers and C-terminus stop codon. The

biotin holoenzyme synthetase (BirA) sequence (AGLNDIFEAQKIEWHEG) [32] was cloned into pENTR11 entry vector at the N-terminus of the beginning of the 15 monomers. The *EarI* sites were added to pENTR11 via PCR using *Pfu* polymerase and methyl-dCTP. The polymer sequence and pENTR11 entry vector were ligated and transformed into *E. coli* *Stbl2* cells. The constructed peptide sequence was as follows:

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAMAGLNDIFEAQKIEWHEG[PEPPKEVVPEKKAPVAPPKKPEVPPVKV]₁₅
PSWEAAAREACCRECCARA.

The newly engineered [PPE172₁₅] vector had fifteen 28-amino acid repeats, multiple tags, his₆ tag, S-tag, and BirA [32] biotinylation sequence, on the N-terminus, and a FAsH tag [33, 34] on the C-terminus of the polymer as shown in the Fig. 2. The biotinylation BirA tag of the [PPE172₁₅] makes it possible to study protein interactions for a variety of commercially available reagents. The FAsH tag is a short tetracysteine sequence (CCXXCC), and is capable of binding to a chemically modified fluorescent dye that fluoresces when bound to the peptide sequence. It is stable and versatile for both affinity purifications and detection studies. Therefore, these affinity tags on [PPE172₁₅] not only facilitate the purification of the protein from both ends to ensure a full-length protein is isolated, but also makes it possible to attach [PPE172₁₅] to tips and surfaces of AFM for single molecule techniques studies. More readily interpretable data can then be generated from these techniques.

The [PPE172₁₅] recombinant protein expression

The challenges in the expression of a repetitive DNA sequence include toxicity of expressed proteins to expression host cells, proteolysis, random recombination events, and a very low yield of expressed protein. We expressed and purified both *in vitro* and *in vivo* bacteria systems, and viral expression systems as potential solutions to these problems. We successfully expressed [PPE172₁₅] in large quantities in a yeast expression system. We will address each expression system respectively; and combine the purification of the expressed [PPE172₁₅] for each host system in one section.

Prokaryotic hosts expression—Elastomeric proteins such as elastin, abductin, resilin, spider silk, and the streptococcal B1 immunoglobulin-binding domain of protein G (GB1) have been cloned and expressed in bacterial hosts [46–53]. Since the elastomeric protein contains repetitive sequences and intrinsically disordered regions, the problems of premature terminations and limited yields of the expressed proteins were encountered in these studies as well [54]. In order to find more suitable hosts for the repetitive sequence expression, we cloned the [PPE172₁₅] into pET-30a(+) expression vector and tested different bacterial hosts for [PPE172₁₅] expression. The pET-30a-[PPE172₁₅] plasmid was transformed into *BL21-CodonPlus (DE3)-RP*, *BL21-SI*, *BLR(DE3) pLysS*, *RosettaBlue(DE3) pLysS*, and *BL21(DE3) pLysS*. The [PPE172₁₅] was expressed in LB medium in these hosts. Although these *BL21* derivatives contain extra tRNA genes such as argU and proL [55] to enhance expressions of toxic proteins with repetitive sequences, we could not detect expression of the [PPE172₁₅] in these hosts with western blots. In addition, the growth rates of bacterial cells with the [PPE172₁₅] plasmid were significantly slower than normal rates of cell growth.

Studies have shown that addition of glucose in the LB medium can decrease levels of basal expression of a target gene. It can also help to maintain plasmid stability, since it reduces the cAMP stimulation of T7 RNA polymerase [56, 57]. Therefore, we modified the expression environments by adding 1% glucose. The expression of the [PPE172₁₅] in the LB culture with 1% glucose was significantly increased upon 1 mM IPTG induction as shown in the

Fig. 3. In addition, adding 1% glucose in LB agar plates before inoculation also increased the level of the [PPE172₁₅] expression in comparison with the LB agar plates without glucose.

In addition, we found that the toxicity of the [PPE172₁₅] to bacterial hosts was further reduced by decreasing the temperature during expression. When the temperature of the expression medium was reduced from 37 °C to 30 °C during cell growth, the bacteria grew better and expressed more [PPE172₁₅]. At each stage of expression, we controlled the growth of the culture cells within the logarithmic phase. We found that mid-log phase induction of expression increased the quantity of the expressed [PPE172₁₅].

We successfully expressed the full-length [PPE172₁₅] in *E. coli* BL21(DE3) *pLysS* hosts (Fig. 4). We further purified the [PPE172₁₅] with affinity NiNTA and organomercurial agarose columns sequentially. The total yield of purified [PPE172₁₅] was approximately 2 mg/liter of culture medium.

Baculoviral expression—Insect cell expression systems have many advantages over other expression systems. They can produce large quantities of recombinant proteins in a short period of time. Since expressed recombinant proteins in insect cells can be posttranslationally modified, they can have increased biological activity and antigenicity [58]. Insect cells also have chaperonins that are required for the proper folding of some proteins and are absent in prokaryotic cells. Recombinant proteins with repetitive sequences such as human type III collagens were expressed successfully in an insect system [59, 60]. Collagens resemble the characteristics of the [PPE172₁₅] sequence, since they consist of many repeating triplet sequences enriched in proline amino acids. Therefore, the [PPE172₁₅] was further investigated in the viral expression system. The [PPE172₁₅] DNA sequence was cloned into a pV1393 expression vector. The pV1393-[PPE172₁₅] vectors were then recombined with the *Autographa californica nuclear polyhedrosis* virus (AcNPV) genome. *Spodoptera frugiperda* Sf9 insect cells were then infected, and the [PPE172₁₅] was expressed and monitored by the gel electrophoresis (Fig. 5A).

In the baculoviral expression, there was one major [PPE172₁₅] band detected at 3% multiplicity of infection (MOI) at 24 hours from western blots. The number of [PPE172₁₅] related bands increased with the expression time, which was revealed by western blots with specific monoclonal antibody (RT11) to titin PEVK region. At 48 and 72 hours, numerous forms of [PPE172₁₅] were expressed. In the baculovirus system, [PPE172₁₅] showed an array of distinguishable bands (Fig. 5A). The mass difference between each band of the [PPE172₁₅] varied from 3000 to 7000 Da. Many clearly defined lower molecular weight [PPE172₁₅] bands were detected on the western blots. Most interestingly, we detected multiple bands of polyprotein higher than the molecular weight of [PPE172₁₅]. The amount of expressed [PPE172₁₅] with higher molecular weights increased after 24 hours of expression in the viral system. The creation of new and larger forms of [PPE172₁₅] in the viral expression system may be caused by recombination events during the process of protein expression or DNA transcription. The recombination events of the polyprotein seemed to follow the modular sizes of the [PPE172₁₅]. The molecular weight of each band of the polyprotein was changed roughly in multiples of the single module of [PPE172₁₅].

Repetitive nucleotide sequences in the eukaryotic genome are a very common feature [61, 62]. During DNA replication processes, the single-stranded DNA regions with repetitive sequences can spontaneously insert or delete some tandem repeats. Therefore, it causes mismatches of the tandem repeats, and results in recombinant of DNA sequences [63]. These mismatches and recombination of DNA structures cause a newly synthesized DNA

with a different length and various numbers of repeats [64, 65]. The aberrant forms of [PPE172₁₅] in viral expression may be the result of recombination events during expression.

Cell-free expression system—Directly coupled transcription and translation systems provide an easy and fast way to produce many recombinant proteins in a cell-free system. Therefore, the pET-30a-[PPE172₁₅] plasmid was expressed *in vitro* by an *E. coli* coupled transcription and translation system according to the product manual (Roche rapid translation system). Briefly, 15 µg pET30a-[PPE172₁₅] plasmid DNA was added to the reaction solution, one complete mini EDTA free protease inhibitor cocktail tablet was added to the feeding solution. Both solutions were loaded to the designated compartment respectively. The reaction was carried out in the RTS Proteomaster instrument at 30 °C with 150 rpm for 17–24 hrs. The expressed [PPE172₁₅] was then purified with NiNTA and organomercurial agarose columns. As shown in Fig. 5B, there were significant problems with premature termination and/or proteolysis of the [PPE172₁₅]. [PPE172₁₅] was expressed in a ladder pattern. Increasing the concentration of protease inhibitors in the reaction mixture and varying the reaction time did not significantly prevent the formation of truncated [PPE172₁₅] during expression.

The truncation of [PPE172₁₅] in the cell-free protein expression could be caused by translational error [66], since repetitive DNA sequences can cause ribosome frameshifts during the translation process [67]. The frequency of frameshifts increases significantly if there is a limited amount of transfer ribonucleic acids. Since 29.3% of the amino acid composition in the [PPE172₁₅] sequence was proline, an insufficient supply of proline aminoacyl-transferase may have increased the frequency of ribosome frameshifts. Since the proline codon is used less frequently by bacterial hosts, the difficulty in [PPE172₁₅] expression may have been exacerbated [68, 69]. In addition, truncated forms of [PPE172₁₅] can be caused by proteolysis during the expression process, due to the inherent sensitivity of intrinsically disordered proteins to proteolysis.

***Pichia pastoris* yeast expression**—The yeast system, as a eukaryotic host, has an advantage over bacterial systems in terms of protein expression and folding. Since there were 15 identical modules in [PPE172₁₅], the expression host cells were placed under extreme stress by the demand of the repetitive sequence during transcription and translation. This exacerbated the potential for translational errors [70]. We compared the demands of codon usage in the [PPE172₁₅] sequence to the usage in *E. coli* and yeasts. The percentage of the codon usage for yeast cells was relatively close to the requirement of the four major amino acids (PEVK) in the sequence of [PPE172₁₅]. Therefore, the yeast system was a more suitable host for the expression of [PPE172₁₅] [55, 71]. In addition, *Pichia pastoris* transformed with pPICZα expression vectors can transport expressed proteins with the secretion signal into the expression medium via secretory pathways. This mechanism could further reduce the stress imposed by the tandem repeat sequences on the expression hosts. Although recombinant proteins with repetitive sequences have been expressed in the yeast *Pichia* [72, 73], the yields of expressed protein vary for the secretory pathway in the yeast system [74].

Therefore, we investigated the expression of [PPE172₁₅] in the yeast *Pichia* system. We cloned the [PPE172₁₅] into pPICZα and pPICZ expression vectors, and electroporated into *Pichia x-33* and *GS115* cells. Numerous transformants for each expression vector and cell types were screened and selected for the expression of the [PPE172₁₅]. As indicated in Fig. 5D, the transformants in *x-33* cells with pPICZα vector were the most successful clones in expression of the [PPE172₁₅]. We expressed the [PPE172₁₅] in the *Pichia X-33* cells with pPICZαA-[PPE172₁₅] in YPD medium at 30 °C. The expressed [PPE172₁₅] was further sequentially purified by NiNTA and organomercurial agarose columns. Fig. 5C indicated the

successful purification of the [PPE172₁₅] in the *Pichia* system. The yeast expression system produced large quantities of the expressed polypeptide via secretion pathway. It overcame the difficulties associated with the expression of repetitive regions in recombinant polypeptide. Production of [PPE172₁₅] was increased significantly. The yield was over 1 gram/L. We expanded the expression of [PPE172₁₅] in a bench top fermentor with FM22 medium, which made large quantities of [PPE172₁₅] available for biophysical studies. We further expressed [PPE172₁₅] with the stable isotope nitrogen-15 in minimal media for NMR studies. In addition, we digested the different concentrations of the [PPE172₁₅] with endoglycosidases, *PNGase F*, which removes all attached N-linked carbohydrate such as mannose from the proteins. The results, Fig. 5E indicated that there was no significant glycosylation of the expressed [PPE172₁₅].

Purification of the expressed [PPE172₁₅]

The [PPE172₁₅] was engineered to have 15 identical modules with multiple tags at both ends of the polymer. With these engineered tags, studies of [PPE172₁₅] with single molecule technique such as AFM and Laser Tweezers can generate more interpretable data to understand the mechanical function of the PEVK region of the titin.

In order to attach both C- and N-termini of [PPE172₁₅] to surfaces and AFM tips for nanomechanical measurements, tags at both ends must be present. The full-length nature of [PPE172₁₅] expressed by each host system was confirmed and purified by affinity purification. Due to the high frequency of recombination events during cloning and expression, we purified both *E. coli*, cell-free, and yeast expressed [PPE172₁₅] first with a Ni-NTA column to catch the N-terminus His affinity tags, and then with an organomercurial agarose column to catch the C-terminus tetracysteine sequence (CCXXCC) FLAsH tags. The purified proteins were thus highly enriched in full-length [PPE172₁₅]. The quality of the purified [PPE172₁₅] was routinely checked by labeling with FLAsH dye prior to SDS gel analysis to confirm the presence of the tetracysteine sequence at the C-terminus. In addition, [PPE172₁₅] was further checked by western blotting with anti-His tag and RT11 antibodies. The RT11 antibody is monoclonal and targeted directly to the PEVK module in the full-length [PPE172₁₅] [22]. Results from the purification and labeling confirmed the full-length expression of the [PPE172₁₅] in Fig. 4 for *E. coli* expression and in Fig. 5B for cell-free express and Fig. 5C for yeast expression.

The purified full-length bacterial and yeast expressed [PPE172₁₅] was further confirmed by LC-MALDI-TOF mass spectrometry with the molecular mass of 54,180 Da. The mass value in the mass spectrometry was slightly higher than the calculated [PPE172₁₅] molecular weight according to the amino acids sequence. The difference could be caused by attachments of residual metal ions in the buffer or matrix to the highly charged [PPE172₁₅] [75–77]. The extinction coefficient was calculated (http://expasy.org/proteomics/protein_characterisation_and_function) at 280 nm as 0.21 for 1 mg/ml, due entirely to the end tags and the absence of aromatic amino acids in the PEVK module.

For conformational studies, the [PPE172₁₅] was subject to gel filtration chromatography. We observed that the *E. coli* and yeast expressed [PPE172₁₅] appeared to interact significantly and variably with the gel filtration matrix under the experimental conditions. As shown in Table 1, the K_{av} varied with the gel filtration media even under a strongly reducing condition to keep the tetracysteines from cross linking [PPE172₁₅] into aggregates. In the Sephadex gels, the K_{av} was about 2.5 times larger than the K_{av} of bovine serum albumins (BSA, globular protein at 68 kDa). Fortunately, its K_{av} on a desalting column (HiTrap, GE Health Sciences) was as small as 0.077. [PPE172₁₅] was eluted near the void volume of the column as expected in the desalting column.

In general, the affinity of [PPE172₁₅] toward the gel filtration media still allowed for effective purification, but rendered an uncertain estimation of hydrodynamic property (Stokes radius). Since [PPE172₁₅] is intrinsically disordered with unpredictable shape and flexibility, it was not useful to estimate its molecular weight or the oligomeric state on gel filtration columns. Specifically, there were no suitable calibration markers, with similar shapes and flexibility, that correlate Stokes radius to the molecular weight of intrinsically disordered proteins. The interactions of the [PPE172₁₅] with gel filtration media further compounded the problem. Similar problems were also evident on gel electrophoresis. Since the [PPE172₁₅] migrated much slower in the gels, and thus a higher molecular weight than its calculated molecular weight (54,000 Da based on protein sequence) was observed in gel electrophoresis system. This abnormal mobility might be caused by unusual SDS binding, the abnormal conformation of SDS/protein complexes at any given SDS-protein ratio and/or potential interaction of the complex with the gel material.

Plasmid sequencing

We further confirmed the sequence of the [PPE172₁₅] at the DNA level. Since the total length of the [PPE172₁₅] DNA sequence was approximately 1500 base pairs (bp), and standard DNA sequence reactions give a reliable reading for only 400–500 base pairs, it was difficult to confirm the accuracy of the [PPE172₁₅] sequence. Therefore, we generated a set of nested deletion inserts from the pET-30a-[PPE172₁₅] plasmid. The [PPE172₁₅] plasmid was sequentially digested with *KpnI*, *NcoI*, and *Exonuclease III* sequentially. The remaining undigested single strand DNA extensions were further digested with *Mung bean nuclease*. Three DNA segments with 200 bp, 850 bp, and 1200 bp were created from the [PPE172₁₅] plasmids. These new DNA segments were gel-purified and ligated with T₄ ligase. The recirculated [PPE172₁₅] plasmids were transformed into *StbI2* cells. The plasmids were purified and sequenced. The results of the sequences for the set of deletion inserts were aligned with each other using the website <http://www.ncbi.nlm.nih.gov/blast>. A 1.5 kb [PPE172₁₅] sequence was confirmed with correct reading frames and stop codons for expression.

Disordered in the [PPE172₁₅] conformation

The intrinsically disordered structures of [PPE172₁₅] were similar to that of the PEVK modules of titin, as measured by near UV circular dichroism spectroscopy (CD) and its thermal responses. As shown in Fig. 6, the strong negative ellipticity minimum at 201 nm and a shoulder near 220 nm in the CD spectra indicated a polyproline II (PPII) left-handed helix structure of the [PPE172₁₅]. The CD spectra of the [PPE172₁₅] measured over the range of 2 °C to 75 °C and showed a well-defined isodichroic point between 210 nm and 212 nm, which was completely reversible when it was cooled. Thermal melting curves were monitored by continuous measurements of the CD spectral value at 201 nm from 4 °C to 75 °C. The melting curve consisted of two linear lines that intercepted at 58 °C, with no sigmoidal melting characteristic of globular proteins (data not shown). The CD studies of the [PPE172₁₅] were indicative of an ensemble of malleable conformations of the PEVK modules, including PPII helices that interconvert to other secondary structures gradually with the rising temperature [28, 31].

Limited digestion of [PPE172₁₅]

Intrinsically disordered proteins are generally prone to proteolytic digestion due to their rather open structure that exposes proteolytic sites. To investigate the polymeric structure, we subjected yeast expressed [PPE172₁₅] to limited digestion conditions with *Enterokinase* that cuts mainly between glutamic acid (E) and lysine (K) residues. The [PPE172₁₅] was digested with 0.004% (w/w) and 0.04% (w/w) protein to *Enterokinase* ratio. The time course of digestion was followed by SDS gels. As shown in Fig. 7, the digested [PPE172₁₅] at

different enzyme concentrations displayed a ladder of bands with increasing mobility on the gel electrophoresis. The spacing between these bands was estimated to be 3 kDa, corresponding to the size of one PEVK module. This was confirmed by LC-MALDI-TOF/TOF mass spectrometry. The ladder pattern of 15 bands of progressively digested PEVK module confirmed the 15-mer structure of [PPE172₁₅]. The pattern of the digested [PPE172₁₅] was reminiscent of those observed during the expression of the [PPE172₁₅] in the baculovirus system.

[PPE172₁₅] binds multiple nebulin SH3 with modest affinity

Our previous studies indicated that many titin PEVK modules contain a single SH3 binding motif [17]. SH3 is a 60-amino acid recognition domain, and it serves important regulatory functions by binding to proline rich motifs of signaling proteins [78]. Since [PPE172₁₅] was expected to contain 15 tandem SH3 binding sites, we investigated this interaction by both ELISA and surface plasmon resonance (SPR) methods. In ELISA, the surface adsorbed yeast expressed [PPE172₁₅] bound to the soluble nSH3 with an isotherm. The binding curves were fitted with a dissociation constant (K_d) of 7 μ M at 4 °C (Fig. 8A) and 23 μ M at 25 °C (Fig. 8B). The K_d was consistent with the affinity constants of the binding of nSH3 to a single PEVK module in the solution as studied by the fluorescence enhancement of nSH3 [31]. The increase in affinity at the lower temperature may result from an increase in PPII content that formed at the nSH3 binding site. Anti-titin monoclonal antibody, RT11, showed a strong binding to [PPE172₁₅] with a K_d of 0.47 μ M at 4 °C and 0.3 μ M at 25 °C. Interestingly, titin RT11 inhibited the binding of nSH3 to [PPE172₁₅] in the ELISA. The K_d decreased to 9.69 μ M at 4 °C and 471 μ M at 25 °C in the presence of 50 μ g/ml of RT11. These data suggested that the RT11 epitope is at the nSH3 binding motif of the PEVK module and would be a useful probe of the interplay of PEVK and SH3 signaling pathways.

The interactions between the immobilized yeast expressed [PPE172₁₅] and soluble nSH3 were studied by SPR. The [PPE172₁₅] was directly immobilized onto sensor chips and the binding of nSH3 to the [PPE172₁₅] chip was studied by continuously injecting nSH3 into the [PPE172₁₅] immobilized chip surfaces at different concentrations without regeneration. The K_d of the [PPE172₁₅] was calculated as 340 μ M at room temperature by assuming 15 independent sites (Fig. 8C). The discrepancies in the binding constants between the SPR (340 μ M) and the ELISA (23 μ M) may be caused by the differences in the data treatment and/or the negatively charged surface of the CM5 chip that was lower in affinity. Furthermore, the binding of nSH3 in a dynamic flow system may be affected by on/off kinetics of the gel layer on the chip surfaces; it could further affect the binding capability of nSH3.

Discussion

Successful cloning and expression of repetitive and intrinsically disordered sequences of recombinant proteins has become increasingly important for two reasons. First, recent advances in tissue engineering and regenerative medicine have created a pressing need for effective biopolymers to generate supporting scaffolds. The future of regenerative medicine requires these scaffolds for cells to adhere and grow into functional units. The increasing demand for a suitable biopolymer will promote interest in engineering tandem repeat recombinant polyproteins [79–86]. Second, the availability of genomic sequences in various organisms makes many new open reading frames interesting for further study of their functions [87, 88]. Many of these sequences are with repetitive and intrinsically disordered regions [89–91]. Therefore, successful cloning and expression of these regions are the first steps for examining their biological roles [38, 92]. The knowledge and ability to express and purify a repetitive and intrinsically disordered recombinant protein such as [PPE172₁₅] will be useful for future research.

Due to the repetitive sequences in the [PPE172₁₅] DNA, random recombination events and deletions of nucleotides readily occurred during the cloning process. We tested several strains of *E. coli* hosts and found that *max efficiency Stbl2* with *recA1* was the best host for the unstable DNA plasmid. The newly constructed DNA plasmids containing the repetitive sequences were very unstable. The DNA plasmids in the transformed *E. coli* cells had to be instantly stored at −20 °C. At each stage of the plasmid propagation, it was critical to avoid reaching the saturation stage in LB growth medium and to store the plasmid in *E. coli* collected during the logarithmic phase of the growth curve. These careful storage steps could reduce the instability of DNA plasmid throughout the process. Additionally, secondary structures of DNA and RNA interfered with each step of DNA manipulation due to the repetitive sequences. For example, it was difficult to cut the [PPE172₁₅] DNA near the *EarI* restriction enzyme sites. We used *Eam110II* enzyme as an isoschizomer to *EarI*, since they have the same recognition sequence. We prioritized the order of enzyme digestions by linearizing the [PPE172₁₅] plasmids with other enzymes before the *EarI* digestion. Fresh *EarI* restriction enzymes were also added to the reaction mixture multiple times during the digestions process. These steps helped to overcome the difficulties in engineering the [PPE172₁₅] DNA polymer.

In this study, use of the *Pichia* yeast expression host via the secretion pathway was most suitable for protein expression of the tandem repeat sequence [70, 93–96]. Codon usage bias in *Pichia* and *E. coli* could contribute to the success of the expression of the [PPE172₁₅] in the yeast system. The relative synonymous codon usage (RSCU) value is a useful tool to assess the adaptation of expression hosts to the demand of the expression sequence. For instance, the RSCU value for three proline codons used in [PPE172₁₅] was 3.975 for yeast, but only 0.711 for bacteria [71, 97, 98]. This indicates that yeast is a better host system for protein expression in this study. The effective number of codon (ENC) is a good indicator for the usage of the codon for the viral expression system. For example, proline has an average ENC of 25, which is considered as a strong codon bias [99]. Therefore, it may not be a good host for expression of [PPE172₁₅].

Nutrient composition in the medium is an important factor in the process of recombinant protein expression. Studies have shown that the formulation of growth medium can alter how bacterial cells utilize nutrients via different fermentation processes during the growth [100]. In this study, we changed the expression environment for *E. coli* expression hosts by adding glucose to attenuate the bacterial expression process, and by lowering medium temperature to increase the [PPE172₁₅] expression. These changes can decelerate the translation assembling process; decrease the speed of biosynthesis by the expression hosts, and therefore increased expression of the recombinant polyprotein.

This study represents the successful expression of a recombinant polyprotein with tandem repeats and intrinsically disordered regions. The technical challenges that we overcame in this work can aid future research by providing information about pitfalls during the design and engineering of tandem repeat DNA sequences.

Acknowledgments

We thank Mr. Gustavo Gutierrez-Cruz for technical support and suggestions for the work. We are grateful to our NIAMS colleagues and the anonymous reviewers for helpful comments.

References

1. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic acids research*. 2012; 40:D571–579. [PubMed: 22135293]

2. Ge Y, Wagner MJ, Siciliano M, Wells DE. Sequence, higher order repeat structure, and long-range organization of alpha satellite DNA specific to human chromosome 8. *Genomics*. 1992; 13:585–593. [PubMed: 1639387]
3. Warburton PE, Waye JS, Willard HF. Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol Cell Biol*. 1993; 13:6520–6529. [PubMed: 8413251]
4. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002; 323:573–584. [PubMed: 12381310]
5. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*. 2008; 9:S1.
6. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry*. 2002; 41:6573–6582. [PubMed: 12022860]
7. Hartwell LH, Kastan MB. Cell cycle control and cancer. *Science*. 1994; 266:1821–1828. [PubMed: 7997877]
8. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YYJ, Romero P, Cortese MS, Uversky VN, Dunker AK. Rational drug design via intrinsically disordered protein. *Trends Biotechnol*. 2006; 24:435–442. [PubMed: 16876893]
9. Prahlad V, Morimoto RI. Neuronal circuitry regulates the response of *Caenorhabditis elegans* to misfolded proteins. *PNAS*. 2011; 108:14204–14209. [PubMed: 21844355]
10. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, Bates GP, Davies SW, Lehrach H, Wanker EE. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell*. 1997; 90:549–558. [PubMed: 9267034]
11. Lee JE, Cooper TA. Pathogenic mechanisms of myotonic dystrophy. *Biochem Soc Trans*. 2009; 37(Pt 6):1281–1286. [PubMed: 19909263]
12. Muragaki Y, Mundlos S, Upton J, Olsen BR. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science*. 1996; 272:548–551. [PubMed: 8614804]
13. Krawczak M, Cooper DN. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environments. *Human Genetics*. 1991; 86:425–441. [PubMed: 2016084]
14. Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H, Labeit S. The complete gene sequence of titin, expression of an unusual ~700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circulation Research*. 2001; 89:1065–1072. [PubMed: 11717165]
15. Granzier HL, Labeit S. Titin and its associated proteins: The third myofilament system of the sarcomere. *Adv Protein Chem*. 2005; 71:89–119. [PubMed: 16230110]
16. Tskhovrebova L, Trinick J. Titin: properties and family relationships. *Nature reviews Molecular cell biology*. 2003; 4:679–689.
17. Ma K, Forbes JG, Gutierrez-Cruz G, Wang K. Titin as a giant scaffold for integrating stress and Src homology domain 3-mediated signaling pathways - The clustering of novel overlap ligand motifs in the elastic PEVK segment. *J Biol Chem*. 2006; 281:27539–27556. [PubMed: 16766517]
18. Lange S, Xiang F, Yakovenko A, Vihola A, Hackman P, Rostkova E, Kristensen J, Brandmeier B, Franzen G, Hedberg B, Gunnarsson LG, Hughes SM, Marchand S, Sejersen T, Richard I, Edstrom L, Ehler E, Udd B, Gautel M. The kinase domain of titin controls muscle gene expression and protein turnover. *Science*. 2005; 308:1599–1603. [PubMed: 15802564]
19. Trombitás K, Wu Y, McNabb M, Greaser M, Kellermayer MS, Labeit S, Granzier H. Molecular basis of passive stress relaxation in human soleus fibers: assessment of the role of immunoglobulin-like domain unfolding. *Biophys J*. 2003; 85:3142–3153. [PubMed: 14581214]
20. Kellermayer MS, Smith SB, Granzier HL, Bustamante C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science*. 1997; 276:1112–1116. [PubMed: 9148805]

21. Forbes JG, Flaherty DB, Ma K, Qadota H, Benian GM, Wang K. Extensive, modular intrinsically disordered segments in *C elegans* TTN-1, implications in filament binding, elasticity and oblique striation. *J Mol Biol.* 2010; 398:672–689. [PubMed: 20346955]
22. Gutierrez-Cruz G, Van Heerden AH, Wang K. Modular motif, structural folds and affinity profiles of the PEVK segment of human fetal skeletal muscle titin. *J Biol Chem.* 2001; 276:7442–7449. [PubMed: 11084039]
23. Jin AJ, Forbes JG, Wang K. Modulation of titin PEVK elasticity as studied via atomic force spectroscopy. *Biophys J.* 2002; 82:399a–399a. [PubMed: 11751326]
24. Jin AJ, Forbes JG, Wang K. Atomic force microscopy studies of a titin PEVK segment. *Biophys J.* 2000; 78:381a–381a.
25. Forbes JG, Jin AJ, Ma K, Wang K. Elasticity of the PEVK segment of titin is dominated by sequence-specific charge pairing interactions. *Biophys J.* 2003; 84:140a–140a.
26. Forbes JG, Jin AJ, Ma K, Gutierrez-Cruz G, Tsai WL, Wang K. Titin PEVK segment: charge-driven elasticity of the open and flexible polyampholyte. *J Muscle Res Cell Motil.* 2005; 26:291–301. [PubMed: 16465472]
27. Ma K, Wang K. Interaction of nebulin SH3 domain with titin PEVK and myopalladin: implications for the signaling and assembly role of titin and nebulin. *Febs Lett.* 2002; 532:273–278. [PubMed: 12482578]
28. Ma K, Kan L, Wang K. Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. *Biochemistry.* 2001; 40:3427–3438. [PubMed: 11297408]
29. Ma K, Wang K. Binding of copper(II) ions to the polyproline II helices of PEVK modules of the giant elastic protein titin as revealed by ESI-MS, CD, and NMR. *Biopolymers.* 2003; 70:297–309. [PubMed: 14579303]
30. Ma K, Wang K. Nebulin SH3 domain binds to multiple proline-rich peptides of titin PEVK and myopalladin: Implications for the targeting pathway of nebulin to the IZI assembly. *Biophys J.* 2002; 82:400a–400a.
31. Ma K, Wang K. Malleable conformation of the elastic PEVK segment of titin: non-co-operative interconversion of polyproline II helix, beta-turn and unordered structures. *Biochem J.* 2003; 374:687–695. [PubMed: 12816538]
32. Barker DF, Campbell AM. Genetic and biochemical characterization of the *birA* gene and its product: Evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in *Escherichia coli*. *J Mol Biol.* 1981; 146:469–492. [PubMed: 6456358]
33. Hoffmann C, Gaietta G, Zürn A, Adams SR, Terrillon S, Ellisman MH, Tsien RY, Lohse MJ. Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nature Protocols.* 2010; 5:1666–1677.
34. Adams SR, Tsien RY. Preparation of the membrane-permeant biarsenicals FIAsh-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nature Protocols.* 2008; 3:1527–1534.
35. Borgi I, Gargouri A. A spontaneous direct repeat deletion in the pGEX fusion vector decreases the expression level of recombinant proteins in *Escherichia coli*. *Protein Expression and Purification.* 2008; 60:15–19. [PubMed: 18434194]
36. Chow DC, Dreher MR, Trabbic-Carlson K, Chilkoti A. Ultra-high expression of a thermally responsive recombinant fusion protein in *E. coli*. *Biotechnol Prog.* 2006; 22:638–646. [PubMed: 16739944]
37. Welch M, Villalobos A, Gustafsson C, Minshull J. You're one in a googol: optimizing genes for protein expression. *Journal of the Royal Society Interface the Royal Society.* 2009; 6:S467–S476.
38. Godiska R, Mead D, Dhodda V, Wu CC, Hochstein R, Karsi A, Usdin K, Entezam A, Ravin N. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic acids research.* 2010; 38:e88. [PubMed: 20040575]
39. Cèbe R, Geiser M. Rapid and easy thermodynamic optimization of the 5'-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*. *Protein Expr Purif.* 2006; 45:374–380. [PubMed: 16125408]

40. Candelas G, Candelas T, Ortiz A, Rodríguez O. Translational pauses during a spider fibroin synthesis. *Biochemical and Biophysical Research Communications*. 1983; 116:1033–1038. [PubMed: 6651838]
41. Plantz BA, Sinha J, Villarete L, Nickerson KW, Schlegel VL. *Pichia pastoris* fermentation optimization: energy state and testing a growth-associated model. *Applied Microbiology and Biotechnology*. 2006; 72:297–305. [PubMed: 16493554]
42. Wang K, Knipfer M, Huang QQ, Heerden AV, Hsu LC, Gutierrez G, Quian XL, Stedman H. Human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture. *The Journal of Biological Chemistry*. 1996; 271:4304–4314. [PubMed: 8626778]
43. Won JI, Barron AE. A New Cloning Method for the Preparation of Long Repetitive Polypeptides without a Sequence Requirement. *Macromolecules*. 2002; 35:8281–8287.
44. McMillan RA, Lee TAT, Conticello VP. Rapid Assembly of Synthetic Genes Encoding Protein Polymers. *Macromolecules*. 1999; 32:3643–3648.
45. Lu Q. Seamless cloning and gene fusion. *Trends Biotechnol*. 2005; 23:199–207. [PubMed: 15780712]
46. Prince JT, McGrath KP, DiGirolamo CM, Kaplana DL. Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry*. 1995; 34:10879–10885. [PubMed: 7662669]
47. Arcidiacono S, Mello C, Kaplan D, Cheley S, Bayley H. Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. *Appl Microbiol Biotechnol*. 1998; 49:31–38. [PubMed: 9487707]
48. Fahnestock SR, Irwin SL. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl Microbiol Biotechnol*. 1997; 47:23–32. [PubMed: 9035407]
49. Szela S, Avtges P, Valluzzi R, Winkler S, Wilson D, Kirschner D, Kaplan DL. Reduction-oxidation control of beta-sheet assembly in genetically engineered silk. *Biomacromolecules*. 2000; 1:534–542. [PubMed: 11710178]
50. Lewis RV, Hinman M, Kothakota S, Fournier MJ. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. *Protein Expression and Purification*. 1996; 7:400–406. [PubMed: 8776759]
51. Lyons RE, Lesieur E, Kim M, Wong DCC, Huson MG, Nairn KM, Brownlee AG, Pearson RD, Elvin CM. Design and facile production of recombinant resilin-like polypeptides: gene construction and a rapid protein purification method. *Protein Engineering Design & Selection*. 2007; 20:25–32.
52. Lv S, Dudek DM, Cao Y, Balamurali MM, Gosline J, Li H. Designed biomaterials to mimic the mechanical properties of muscles. *Nature*. 2010; 465:69–73. [PubMed: 20445626]
53. Cao Y, Li H. Engineered elastomeric proteins with dual elasticity can be controlled by a molecular regulator. *Nature Nanotechnology*. 2008; 3:512–516.
54. Wong PF, Kaplan DL. Genetic engineering of fibrous proteins: spider dragline silk and collagen. *Advanced Drug Delivery Reviews*. 2002; 54:1131–1143. [PubMed: 12384311]
55. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol*. 1995; 6:494–500. [PubMed: 7579660]
56. Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene*. 1998; 209:95–103. [PubMed: 9524234]
57. Zhang Y, Taiming L, Liu J. Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in *Escherichia coli*. *Protein Expression and Purification*. 2003; 29:132–139. [PubMed: 12729734]
58. Ikonomou L, Schneider YJ, Agathos SN. Insect cell culture for industrial production of recombinant proteins. *Appl Microbiol Biotechnol*. 2003; 62:1–20. [PubMed: 12733003]
59. Dagger JM, Dagger AL, Notbohm H, Fietzek PP, Dagger TP, Dagger KK. Expression of wild-type and modified proalpha chains of human type I procollagen in insect cells leads to the formation of stable $[\alpha 1(I)]_2\alpha 2(I)$ collagen heterotrimers and $[\alpha 1(I)]_3$ homotrimers but not $[\alpha 2(I)]_3$ homotrimers. *The Journal of Biological Chemistry*. 1997; 272:21824–21830. [PubMed: 9268313]

60. Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko KI. Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. *J Biol Chem.* 1996; 271:11988–11995. [PubMed: 8662631]
61. Lovett ST, Feschenko VV. Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. *Proc Natl Acad Sci USA.* 1996; 93:7120–7124. [PubMed: 8692955]
62. Warren ST. The expanding world of trinucleotide repeats. *Science.* 1996; 271:1374–1375. [PubMed: 8596908]
63. Oliveira PH, Prather KJ, Prazeres DMF, Monteiro GA. Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. *Applied Microbiology and Biotechnology.* 2010; 87:2157–2167. [PubMed: 20496146]
64. Levinson G, Gutman GA. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution.* 1987; 4:203–221. [PubMed: 3328815]
65. Kubis S, Schmidt T, Heslop-Harrison JS. Repetitive DNA Elements as a Major Component of Plant Genomes. *Annals of Botany.* 1998; 82(supplement A):45–55.
66. Ramachandiran V, Kramer G, Horowitz PM, Hardesty B. Single synonymous codon substitution eliminates pausing during chloramphenicol acetyl transferase synthesis on *Escherichia coli* ribosomes in vitro. *Febs Lett.* 2002; 512:209–212. [PubMed: 11852081]
67. Lindsley D, Gallant J. On the Directional Specificity of Ribosome Frameshifting at a Hungry Codon. *P Natl Acad Sci USA.* 1993; 90:5469–5473.
68. Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev.* 1996; 60:512–538. [PubMed: 8840785]
69. Flick K, Ahuja S, Chene A, Bejarano MT, Chen QJ. Optimized expression of *Plasmodium falciparum* erythrocyte membrane protein I domains in *Escherichia coli*. *Malaria J.* 2004; 3
70. Sallach RE, Conticello VP, Chaikof EL. Expression of a recombinant elastin-like protein in *pichia pastoris*. *Biotechnol Prog.* 2009; 25:1810–1818. [PubMed: 19827084]
71. Wadal K, Wada Y, Ishibashi F, Gojobori T, Ikemura T. Codon usage tabulated from the GenBank Genetic Sequence Data. *Nucleic acids research.* 1992; 20(suppl):2111–2118. [PubMed: 1598238]
72. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI. Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *The EMBO Journal.* 1997; 16:6702–6712. [PubMed: 9362485]
73. Fahnestock SR, Bedzyk LA. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Applied Microbiology and Biotechnology.* 1997; 47:33–39. [PubMed: 9035408]
74. Fahnestock SR, Yao Z, Bedzyk LA. Microbial production of spider silk proteins Reviews in Molecular. *Biotechnology.* 2000; 74:105–119.
75. Spickermann J, Martin K, Räder HJ, Müllen K. Quantitative analysis of broad molecular weight distributions obtained by matrix-assisted laser desorption ionisation-time-of-flight mass spectrometry. *European Journal of Mass Spectrometry.* 1996; 2:161–165.
76. Klee JE. Mass spectrometry of step-growth polymers. *European Journal of Mass Spectrometry.* 2005; 11:591–610. [PubMed: 16322666]
77. Zenobi R, Knochenmuss R. Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews.* 1998; 17:337–366.
78. Pawson T. Protein modules and signalling networks. *Nature Protocols.* 1995; 373:573–580.
79. Velema J, Kaplan D. Biopolymer-based biomaterials as scaffolds for tissue engineering. *Adv Biochem Eng Biotechnol.* 2006; 102:187–238. [PubMed: 17089791]
80. Jaya S, Durance TD. Tailor-made biopolymers porous scaffold fabrication for tissue engineering: application of radiant energy in the form of microwave under vacuum. *Biomed Mater Eng.* 2008; 18:357–366. [PubMed: 19197112]
81. Langer R, Tirrell DA. Designing materials for biology and medicine. *Nature.* 2001; 428:487–492. [PubMed: 15057821]

82. Daamena WF, Veerkampa JH, van Hestb JCM, van Kuppevelt TH. Elastin as a biomaterial for tissue engineering. *Biomaterials*. 2007; 28:4378–4398. [PubMed: 17631957]
83. Maskarinec SA, Tirrell DA. Protein engineering approaches to biomaterials design. *Current Opinion in Biotechnology*. 2005; 16:422–426. [PubMed: 16006115]
84. Qiu W, Huang Y, Teng W, Cohn CM, Cappello J, Wu X. Complete recombinant silk-elastinlike protein-based tissue scaffold. *Biomacromolecules*. 2010; 11:3219–3227. [PubMed: 21058633]
85. Megeed Z, Cappello J, Ghandehari H. Genetically engineered silk-elastinlike protein polymers for controlled drug delivery. *Adv Drug Deliv Rev*. 2002; 54:1075–1091. [PubMed: 12384308]
86. Ghandehari H, Cappello J. Genetic engineering of protein-based polymers: potential in controlled drug delivery. *Pharmaceutical Research*. 1998; 15:813–815. [PubMed: 9647343]
87. Kahvejian A, Quackenbush J, Thompson JF. What would you do if you could sequence everything? *Nat Biotechnol*. 2008; 26:1125–1133. [PubMed: 18846086]
88. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008; 26:1135–1145. [PubMed: 18846087]
89. Kouprina M, Leem S, Solomon G, Ly A, Koriabine M, Otstot J, Pak E, Dutra A, Zhao S, Barrett JC, Larionov V. Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. *EMBO Rep*. 2003; 4:257–262. [PubMed: 12634842]
90. Ashley CT, Warren ST. Trinucleotide repeat expansion and human disease. *Annual Review of Genetics*. 1995; 29:703–728.
91. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*. 1998; 27:573–580. [PubMed: 9862982]
92. Morag AS, Saveson CJ, Lovett ST. Expansion of DNA repeats in escherichia coli: effects of recombination and replication functions. *J Mol Biol*. 1999; 289
93. Sinclair G, Choy FY. Synonymous codon usage bias, the expression of human glucocerebrosidase in the methylotrophic yeast, *Pichia pastoris*. *Protein Expr Purif*. 2002; 26:96–105. [PubMed: 12356476]
94. Outchkourov NS, Stiekema WJ, Jongsma MA. Optimization of the expression of equistatin in *Pichia pastoris*. *Protein Expression and Purification*. 2002; 24:18–24. [PubMed: 11812218]
95. Romanos MA, Scorer CA, Clare JJ. Foreign gene expression in yeast: a review. *Yeast*. 1992; 8:423–488. [PubMed: 1502852]
96. Cregg JM, Vedvick TS, Raschke WC. Recent advances in the expression of foreign genes in *Pichia pastoris*. *Nature Biotechnology*. 1993; 11:905–910.
97. Sharp PM, Li WH. The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*. 1987; 15:1281–1295. [PubMed: 3547335]
98. Hu S, Li L, Qiao J, Guo Y, Cheng L, Liu J. Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. *Protein Expression and Purification*. 2006; 47:249–257. [PubMed: 16403645]
99. Jiang Y, Deng F, Wang H, Hu Z. An extensive analysis on the global codon usage pattern of baculoviruses. *Archives of Virology*. 2008; 153:2273–2282. [PubMed: 19030954]
100. Tripathi NK, Shrivastva A, Biswal KC, Rao PVL. Optimization of culture medium for production of recombinant dengue protein in *Escherichia coli*. *Industrial Biotechnology*. 2009; 5:179–183.
101. Johnson EM, Berk DA, Jain RK, Deen WM. Diffusion and partitioning of proteins in charged agarose gels. *Biophysical Journal*. 1995; 68:1561–1568. [PubMed: 7787041]
102. Root DD, Wang K. Calmodulin-Sensitive Interaction of Human Nebulin Fragments with Actin and Myosin. *Biochemistry*. 1994; 33:12581–12591. [PubMed: 7918483]

Highlights

Successfully engineered titin PEVK polyproteins for nanomechanical measurements.

Investigated systematically expression hosts for proteins with tandem repeats.

Expressed and purified gram quantities of secreted polyproteins in a yeast system.

Facilitated studies of the role of force in titin's function.

Figure 1.
The process and challenges of engineering the [PPE172₁₅] protein

Figure 2.
Schematic diagram of the engineered polyprotein [PPE172₁₅]

Figure 3. Effect of glucose on expression level of [PPE172₁₅] in *E. coli*

Addition of 1% glucose in LB medium increased the level of *E. coli* expressed [PPE172₁₅]. The [PPE172₁₅] was separated on a 4–12% Bis-Tris gel with MES SDS running buffer stained with Coomassie blue. Lane 1: Myofibril standard. Lanes 2: SeeBlue prestained protein standard. Lanes 3 and 5: *E. coli* cell lysates before 1 mM IPTG induction. Lanes 4 and 6: *E. coli* cell lysates after 1 mM IPTG induction. Lanes 3 and 4: LB medium with 1% glucose. Lanes 5 and 6: LB medium without 1% glucose. *The [PPE172₁₅].

Figure 4. Expression and purification of the [PPE172₁₅] from *E. coli* hosts

Analysis of [PPE172₁₅] on a 4–12% NuPAGE Bis-tris gel. (A) Gel stained with Coomassie blue. (B) The C-terminal tetracysteine sequence (CCXXCC) of [PPE172₁₅] was labeled with FLAsH reagent. Lane 1: Invitrogen SeeBlue plus2 prestained standard. Lane 2: Nebulin SH3. Lane 3: NA4 (nebulin recombinant fragments) [102]. Lane 4: [PPE172₁₅]. *The bright band of lane 1 in B is myoglobin-red in Invitrogen SeeBlue plus 2 standards, which can be detected at 535 nm by the instrument.

Figure 5. Expression and purification of the [PPE172]₁₅ in different hosts

A. Western blots of the baculovirus expressed [PPE172]₁₅ using anti-his antibodies. Lane 1: His Std: Qiagen 6× His protein ladder. Lane 2 & 3: The [PPE172]₁₅ expressed in *BL21(DE3)pLysS* cells. Lanes 4, 5, 6, 7: The baculovirus expressed [PPE172]₁₅ at multiplicity of infection (MOI) of 0.3% v/v at 0, 24, 48, 72 hours, respectively. Lanes 8, 9, 10, 11: The MOI of 3% expressed at 0, 24, 48, 72 hours, respectively.

B. The prokaryotic cell-free translation system expressed [PPE172]₁₅ was separated on a 4–20% Tris-glycine gel with silver staining. Lane 1: Invitrogen SeeBlue plus2 pre-stained standard. Lane 2: The [PPE172]₁₅ purified from organomercurial agarose columns. Lane 3: [PPE172]₁₅ purified from NiNTA columns. Lane 4: Western blots for [PPE172]₁₅ with anti-titin antibodies. Lane 5: Western blots for [PPE172]₁₅ using anti-His antibodies.

C. Analysis of the yeast expressed [PPE172]₁₅ with 4–12% NuPAGE Bis-Tris gel stained with Coomassie blue stain. Lane 1: TP1 (recombinant titin PEVK fragment) [22]. Lane 2:

Invitrogen SeeBlue plus2 prestained marker. Lanes 3 & 4: [PPE172₁₅] at different amounts of loading.

D. Screening of [PPE172₁₅] expression in the *Pichia pastoris* yeast system. The pPICZ α -[PPE172₁₅] and pPICZ-[PPE172₁₅] expression vectors were electroporated into yeast *X-33* and *GS115* cells. The transformants were screened for the expression of the [PPE172₁₅]. The yeast cell pellets and supernatants were analyzed with 4–12% NuPAGE Bis-Tris gel stained with Coomassie blue stain. Lanes 1 & 2: The transformants of *X-33* with pPICZ α -[PPE172₁₅] vectors. Lane 3: Invitrogen SeeBlue plus2 prestained standard. Lane 4: TP1. Lane 5: *E. coli* expressed [PPE172₁₅]. Lane 6 & 7: The transformants of *GS115* cells with pPICZ α -[PPE172₁₅]. Lane 8 & 9: The transformants of *GS115* cells with pPICZ-[PPE172₁₅] vectors.

E. Digestion of the yeast expressed [PPE172₁₅] with *PNGase F* to check for glycosylation of the proteins. Lane 1: Invitrogen SeeBlue plus2 prestained marker. Lanes 2, 4, 6: The digested [PPE172₁₅] at 10, 18, 36 μ g of *PNGase*. Lanes 3, 5, 7: Nondigested [PPE172₁₅]. Lanes 8 & 9: Digested and nondigested rabbit IgG at 20 μ g, respectively. Lane 10: CandyCane glycoprotein molecular weight standard (Molecular Probes).

Figure 6. Circular dichroism of the intrinsically disordered [PPE172]₁₅

The circular dichroism (CD) spectra of the yeast expressed [PPE172]₁₅. The CD spectra were obtained at 4 °C, 25 °C, 50 °C, and 75 °C. Each spectrum was scanned eight times at 20 nm/min with a time constant of 1 s. The [PPE172]₁₅ was in 20 mM potassium phosphate pH 7.0.

Figure 7. Proteolysis of tandem repeats of the [PPE172₁₅]

Yeast expressed [PPE172₁₅] was digested with *Enterokinase* in 10 mM potassium phosphate, 150 mM KCl, 10 mM DTT pH 7.0 with 0.004% (w/w) and 0.04% (w/w) protein to *Enterokinase* ratio. The digested [PPE172₁₅] was loaded onto NuPAGE 4–12% Bis-Tris gel and run in the NuPAGE MES SDS running buffer (Invitrogen). Lane 1: SeeBlue Plus2 molecular weight marker (Invitrogen). Lane 6: Undigested [PPE172₁₅]. Lanes 2 & 7: 1 hour digestion. Lanes 3 & 8: 3 hours digestion. Lanes 4 & 9: 6 hours digestion. Lanes 5 & 10: 10 hours digestion.

Figure 8. Interactions of the [PPE172]₁₅ with nebulin SH3

A& B: ELISA study. Yeast expressed [PPE172]₁₅ was adsorbed to a microtiter plate (10 μ g/ml in 20 mM sodium phosphate, 1 mM EDTA, 150 mM NaCl pH 7.0) and incubated with nebulin SH3 (20 mM potassium phosphate, 50 mM NaCl pH 7.0) at 4 °C and 25 °C.

Binding was detected with rabbit polyclonal anti-peptide antibodies to nebulin SH3. The dissociated binding constants were 7 μ M at 4 °C and 23 μ M at 25 °C.

C: Surface plasmon resonance (SPR) study. Yeast expressed [PPE172]₁₅ in 10 mM potassium phosphate buffer pH 7.0 was immobilized onto CM5 chips with 0.4 M of 1-ethyl-3-(3-dimethylpropyl)-carbodiimide and 0.1 M N-hydroxysuccinimide. 1 M ethanolamine-HCl was used to deactivate excess active groups on the surface. Nebulin SH3 (nSH3) in running buffer (10 mM potassium phosphate buffer, 150 mM KCl pH 7.0) was injected from 0.0025 to 3 mM into the surface of the chip at room temperature.

Table 1

The partition coefficients (K_{av}) of *E. coli* expressed [PPE172₁₅] were determined by gel filtration chromatography with different types of column resins

Gel Filtration Column Resins	The [PPE172 ₁₅]	TCEP ^a
HiTrap Desalting	0.077	0.615
Sephadex G-25	1.3	2.1
Sephadex G-50	1.34	2.36
Sephadex G-75	1.22	2.5
Bio-Gel P-4	2.09	2.63

^aTris (2-carboxyethyl) phosphine hydrochloride (TCEP) is a reducing agent with molecular weight of 250.2. We used the small TCEP as a control for column volumes in the gel filtration columns. The average equilibrium K_{av} is 0.54 in SP-sepharose for BSA and ovalbumin [101]. [PPE172₁₅] has a much higher K_{av} than globular proteins.