OPEN ACCESS Freely available online

PLOS ONE

# Structural Characterization of Minor Ampullate Spidroin Domains and Their Distinct Roles in Fibroin Solubility and Fiber Formation

**Zhenwei Gao, Zhi Lin, Weidong Huang¤, Chong Cheong Lai, Jing-song Fan, Daiwen Yang***

Department of Biological Sciences, National University of Singapore, Singapore

## Abstract

Spider silk is protein fibers with extraordinary mechanical properties. Up to now, it is still poorly understood how silk proteins are kept in a soluble form before spinning into fibers and how the protein molecules are aligned orderly to form fibers. Minor ampullate spidroin is one of the seven types of silk proteins, which consists of four types of domains: N-terminal domain, C-terminal domain (CTD), repetitive domain (RP) and linker domain (LK). Here we report the tertiary structure of CTD and secondary structures of RP and LK in aqueous solution, and their roles in protein stability, solubility and fiber formation. The stability and solubility of individual domains are dramatically different and can be explained by their distinct structures. For the tri-domain miniature fibroin, RP-LK-$CTD_{Mi}$, the three domains have no or weak interactions with one another at low protein concentrations (<1 mg/ml). The CTD in RP-LK-$CTD_{Mi}$ is very stable and soluble, but it cannot stabilize the entire protein against chemical and thermal denaturation while it can keep the entire tri-domain in a highly water-soluble state. In the presence of shear force, protein aggregation is greatly accelerated and the aggregation rate is determined by the stability of folded domains and solubility of the disordered domains. Only the tri-domain RP-LK-$CTD_{Mi}$ could form silk-like fibers, indicating that all three domains play distinct roles in fiber formation: LK as a nucleation site for assembly of protein molecules, RP for assistance of the assembly and CTD for regulating alignment of the assembled molecules.

**Citation:** Gao Z, Lin Z, Huang W, Lai CC, Fan J-s, et al. (2013) Structural Characterization of Minor Ampullate Spidroin Domains and Their Distinct Roles in Fibroin Solubility and Fiber Formation. PLoS ONE 8(2): e56142. doi:10.1371/journal.pone.0056142

**Editor:** Markus J. Buehler, Massachusetts Institute of Technology, United States of America

**Received** July 1, 2012; **Accepted** January 7, 2013; **Published** February 13, 2013



**Funding:** This work was supported by a grant from Ministry of Education, Singapore (R154000453112). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: dbsydw@nus.edu.sg

¤ Current address: Ningxia Medical University, Yinchuan, Ningxia, People's Republic of China

## Introduction

Spider silk is an ideal super material due to its extraordinary mechanical properties compared to other available materials [1–3]. However, mass production of spider silk from nature is still impossible because the farming of spiders is limited by their cannibalistic and territorial behavior [4]. Thus, producing spider silk by recombinant biotechnology is one of the most promising alternatives [3,5–7]. Before achieving this goal, one needs to understand the molecular structures, self-assembly mechanism and fiber formation of spider silk proteins, which are affected by solvent environment and shear and elongational forces [3]. Female orb-weaving spiders can produce up to seven types of silk with different mechanical properties by using different types of silk proteins [2]. Until now, complete gene sequences have been known for only dragline silk (or major ampullate silk) from black widow spider (*Latrodectus hesperus*) [8], and only partial gene sequences have been determined for other spider silks. According to the known spider silk protein sequences, silk proteins share several common features in primary structures: (i) large number of amino acids, (ii) many highly repetitive units flanked by non-repetitive N-terminal domain (NTD) and C-terminal domain (CTD), and (iii) high abundance in Ala, Gly or/and Ser. Each repetitive unit of major ampullate spidroin (MaSp) contains multiple repeats of short simple motifs such as $(A)_n$ (n = 4–10) and GGX (X = A, Q or Y) [2,8]. Similarly, the repetitive unit of minor ampullate spidroin (MiSp) also has repeats of such short motifs [9]. Differently, each MiSp repetitive unit contains an additional relatively large domain that lacks repeats of short motifs [9,10]. Different from MaSp and MiSp, the repetitive units of aciniform and tubuliform silk proteins are complex and lack of the short motifs [11–15]. At present it is clear that the composition of the repetitive units varies from one type of silk protein to another, which is suggested to determine the mechanical properties of a given type of silk.

Unlike the repetitive units, CTDs are relatively conserved among all spider silk proteins except the web glue proteins and are assumed to perform the same function [16]. This also applies to NTDs [17]. Recent studies on two spider species have shown that the NTD of MaSp can regulate the self-assembly of silk proteins in a pH dependent manner [18] and the CTD of MaSp can prevent premature aggregation by stabilizing the solution state of silk proteins and direct the alignment of repetitive units to form well defined fibers [19–21]. In spite of the high sequence identity (58%) between the CTDs of MaSp ($CTD_{Ma}$) from *A. diadematus* and *E. australis*, these two domains have distinct biophysical properties.

For example, the intermolecular disulfide bridge in the $CTD_{Ma}$ from *A. diadematus* was considered to be important to the domain stability [21], but the disulfide bridge was found to have only slight contribution to the thermal stability of the domain from *E. australis* and to be not critical for fiber formation [19]. The CTD sequence identity among different types of silk proteins is lower than that among the same type of proteins from different species. For instance, the CTD of tubuliform spidroin (TuSp) and $CTD_{Ma}$ from the same *N. antipodiana* share 29% sequence identity. Due to the low sequence identity, the $CTD_{Ma}$ existed as a dimer [21] but the isolated CTD of TuSp ($CTD_{Tu}$) existed as oligomers in aqueous solution [22]. Therefore, the same type of CTDs from different species and different types of CTDs from the same species may display diverse biophysical properties. In order to demonstrate whether different CTDs perform the functions in the same or different molecular mechanisms, it is necessary to characterize the structures and biophysical properties of individual domains of different silk proteins and their functional roles in silk formation and protein storage.

We previously reported a MiSp clone, clone 145, from the total silk gland cDNA library of *N. antipodiana* [12]. The deduced amino acid (aa) sequence comprises one repetitive domain ($RP_{Mi}$, 128aa, previously named as spacer), one non-repetitive C-terminal domain ($CTD_{Mi}$, 107aa), and one linker domain ($LK_{Mi}$, 89aa, previously named as repetitive sequence) that links $RP_{Mi}$ and $CTD_{Mi}$ or in general links two structured domains (Figure 1). Until now, the N-terminal domain sequence has not been determined yet for any MiSps. $RP_{Mi}$ and $CTD_{Mi}$ are conserved among different spider species (Figure S1), but $LK_{Mi}$s vary significantly in number of amino acids among different repetitive units in the same MiSp [9,10]. Although the $CTD_{Mi}$ from *N. antipodiana* and $CTD_{Ma}$ from *A. diadematus* share 44% sequence identity, the $CTD_{Mi}$ contains no cysteine residues but the $CTD_{Ma}$ has one disulfide linkage between two molecules which can enhance the stability of $CTD_{Ma}$ [19,21]. Moreover, $RP_{Mi}$ is unique to MiSps and its functional roles in protein storage and fiber formation are unknown. Besides the difference in amino acid sequences, MaSp silk is elastic when stretched and MiSp displays irreversible deforming [9]. Thus, MiSp may adopt a different self-assembly and fiber formation mechanism than the well characterized MaSp. In this work, we report the three-dimensional (3D) structure of $CTD_{Mi}$ from *N. antipodiana*, the secondary structures of $RP_{Mi}$ and $LK_{Mi}$ and their roles in conferring protein stability, solubility and fiber formation.

## Materials and Methods

### Cloning of RP and LK Domains of MiSp from *N. antipodiana*

Forward (5′-gcaaatgctatgaacagtttacttggt-3′) and reverse (5′-attgcctaatgttgatacatatccacta-3′) primers were designed on the basis of the known sequence of $RP_{Mi}$ [12]. MiSp fragments each containing one LK domain flanked by partial $RP_{Mi}$ sequences were obtained by polymerase chain reaction (PCR) from genomic DNA.

### Protein Sample Preparation

The DNA sequence of our previously identified MiSp fragment (clone 145) [12] was confirmed here by PCR from our spider genomic DNA. The target genes encoding different MiSp regions ($CTD_{Mi}$, $RP_{Mi}$, RP-$LK_{Mi}$, LK-$CTD_{Mi}$, RP-LK-$CTD_{Mi}$) were amplified from clone 145 using specific primers and subcloned into a pET32-derived expression vector. The recombinant plasmids were transformed into *E. coli* BL21 strain (DE3). Cells were grown in LB or M9 medium at 37°C to an $OD_{600}$ of 0.6. Right after induction by 0.2 mM IPTG (isopropyl β-D-thiogalactoside), cells were shifted to 20°C and further cultured for 16 hrs. For $^{13}C$,$^{15}N$-labeled ($^{15}N$-labeled) samples, the cells were cultured in M9 medium which contained only $^{15}N$-labeled $NH_4Cl$ and $^{13}C$-labeled (non-labeled) D-glucose as the sole nitrogen and carbon source. After over-expressed, the proteins were purified by immobilized metal affinity chromatography, gel filtration and then ion exchange columns. All the proteins used here contained a 6xHis-tag and a thrombin cleavage sequence at the N-terminus.

### NMR Spectroscopy

All NMR experiments were performed on a Bruker 800 MHz NMR spectrometer at 25°C. Non-labeled $CTD_{Mi}$ (~0.6 mM), $RP_{Mi}$ (~0.6 mM) and RP-LK-$CTD_{Mi}$ (~0.6 mM and ~3 mM) in 10 mM phosphate buffer (pH 6.8) were used to acquire one-dimensional (1D) $^1H$ NMR spectra. The 1D NMR spectra were recorded using the water gate W5 pulse scheme with 64 scans and an interscan delay of 2 s. $^{15}N$-labeled $CTD_{Mi}$ (~0.5 mM) and $RP_{Mi}$ (~0.5 mM) in 10 mM phosphate buffer were employed to record 2D $^1H$-$^{15}N$ HSQC spectra. The samples used for structure determination of $CTD_{Mi}$ contained 1 mM $^{13}C$,$^{15}N$-labeled protein, 10 mM phosphate buffer (pH 6.8), 5 mM EDTA, 50 mM NaCl and 0.01% sodium azide. To obtain sequence-specific assignments and nuclear Overhauser effects (NOEs), the following spectra of $CTD_{Mi}$ were recorded: 2D $^1H$-$^{15}N$ HSQC, 2D $^1H$-$^{13}C$ HSQC, 3D HNCA, 3D HN(CO)CA, 3D MQ-CCH-TOCSY[23], 4D time-shared $^{13}C$, $^{15}N$-edited NOESY [24]. Inter-molecular NOEs were identified from a $^{13}C$,$^{15}N$-filtered 3D experiment on a sample containing 50% $^{13}C$,$^{15}N$-labeled and 50% unlabeled proteins [25]. This sample was prepared by mixing equal amount of labeled and unlabeled proteins in 8 M urea for 2 hrs and then removing the urea by dialysis against 10mM phosphate buffer. All the spectra were processed by NMRpipe and analyzed by following the strategy described previously [26] and using NMRspy and XYZ4D (http://yangdw.science.nus.edu.sg/Software&Scripts/XYZ4D/index.htm). CYANA [27] was employed for structure calculation. 10 dimer structures with the

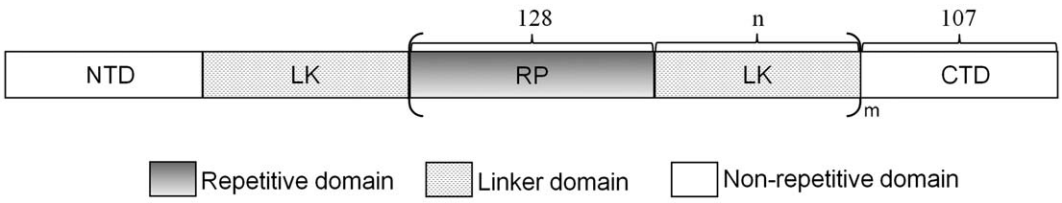


**Figure 1. Molecular organization of MiSp.** The amino acid (aa) number of each domain is indicated above the corresponding bar, if it is available. The number of aa in LK domain (n) varies from 83–174. The total number of RP and LK in the MiSp sequence is still unknown and is denoted as "m". The sequences of the NTD and its adjacent LK domain are unknown.
doi:10.1371/journal.pone.0056142.g001

lowest final target function values were chosen from 100 calculated ones. PROCHECK-NMR [28] was used to evaluate the quality of the structures. Protein structure validation software (PSVS) [29] (http://psvs-1_4-dev.nesg.org/) was utilized to assess the all-atom steric clashes [30].

## Circular Dichroism and Protein Unfolding

All circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a thermal controller. A 0.1 cm path length cuvette was used for all CD experiments. The far-UV spectrum of $RP_{Mi}$ was recorded using a 20 μM protein in 10 mM sodium phosphate at pH 6.8. Both urea- and thermal-induced unfolding processes were monitored at 222 nm using samples with 10 μM protein, 10 mM sodium phosphate at pH 6.8. Except for RP-LK-$CTD_{Mi}$, urea denaturation curves for other MiSp constructs were analyzed with the following equation derived from a two-state unfolding model [31].

$$I_{obs} = \frac{\alpha + \beta C}{1 + \exp[-m(C_m - C)]} \quad (1)$$

where $I_{obs}$ is the experimental signal intensity in the presence of $C$ molar urea, $\alpha$ and $\beta$ the intercept and slope of the pre-transition zone respectively, $C_m$ is the urea concentration at the transition midpoint, and $m$ is the slope at the transition midpoint. For urea-denaturation of RP-LK-$CTD_{Mi}$, the experimental data were fitted using a linear combination of two two-state unfolding equations. Eq. 1 was also used to obtain $T_m$ by replacing $C$ and $C_m$ by $T$ and $T_m$, respectively, where $T$ is temperature, and $T_m$ is the temperature at the transition midpoint.

## Size Exclusion Chromatography

A Superdex$^{TM}$ 75 PG (GE Healthcare) column with a total volume of 120 ml was used to run all the protein samples. The running buffer for $RP_{Mi}$ contained 10 mM sodium phosphate (pH 6.8) with or without 100 mM NaCl. For other samples, only 10 mM sodium phosphate at pH 6.8 was used. The flow rate used was 1 ml/min, and fractions were collected every 2 ml. The fractions were analyzed by SDS-PAGE to confirm which peak in the UV absorbing profile corresponded to the target protein. A molecular mass standard set consisting of Ribomuclease A (13.7 kDa), Chymotrypsinogen A (25 kDa), Ovalbumin (43 kDa), and BSA (67 kDa) was chromatographed to estimate the apparent molecular weights of target proteins.

## Protein Solubility

The purified protein samples in respective 10 mM sodium phosphate and 10 mM Tris buffers (pH 7.0) were concentrated using centrifugal filter units with 3 kDa cutoff membrane at centrifugal force of 3000×g. When the protein concentration was >5 mg/ml, 2 μl samples were regularly taken out from the solution until precipitate or gel was observed. Otherwise, larger volumes of samples were taken for concentration measurements. To determine protein concentrations, the samples taken were diluted in the same buffers as those used for the protein samples. The concentrations were measured using the absorbance at 280 nm and also estimated using SDS PAGE.

## Shear Force-Induced Aggregation

To study protein aggregation induced by shear force that plays a critical role in the natural silk spinning process, samples of 2 ml with 0.05 mg/ml proteins and 10 mM phosphate buffer (pH 6.8) were placed into a UV/Vis cuvette with a small magnetic star bar stirring at 500 rpm, 25°C. The turbidity of the samples was monitored by measuring $OD_{350}$ on a BIO-RAD Smart Spec$^{TM}$ Plus Spectrophotometer at a series of time intervals.

To determine the effect of sodium chloride and sodium phosphate on the aggregation of RP-LK-$CTD_{Mi}$, shear force-induced aggregation experiments were performed under two salt concentrations: 0 and 200 mM. The samples (1 mg/ml) placed in a 2 ml eppendorf tube were shaken at 150 rpm, 25°C in an incubation shaker. At different time points, the samples were taken out. After removing the precipitate by centrifuge, the concentration of the soluble portion was measured and then the total amount of precipitated protein was calculated.

## Scanning Electron Microscopy

1 ml purified protein sample containing 5 mg/ml RP-LK-$CTD_{Mi}$ in 10 mM sodium phosphate buffer (pH 6.8) was placed into a 2 ml eppendorf tube and the sample was shaken at 200 rpm, 25°C for 5 minutes in an incubation shaker. Then, silk-like fibers formed in the tube were picked out by a needle. SEM micrographs of the fibers were observed on a JEOL JSM-6510 and photographed at a voltage of 15 kV and room temperature (24–26°C).

## Prediction of Disorder, Hydrophobicity and Aggregation Propensity

The disordered residues in $LK_{Mi}$ were predicted using PONDR-FIT (http://www.disprot.org/pondr-fit.php). If the disordered score of a residue is >0.5, this residue is considered as disordered [32]. The aggregation-prone regions in $LK_{Mi}$ were predicted using Zyggregator (http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php). When a region of several consecutive residues each have aggregation scores larger than 1, this region is considered to be prone to aggregate [33]. The hydrophobicity plot of $LK_{Mi}$ was obtained using Protscale (http://web.expasy.org/cgi-bin/protscale/protscale.pl) with the scale option of Hphob./Roseman [34].

# Results and Discussion

## Sequences of RP and LK Domains

Our PCR results from the genomic DNA show that all the repetitive domains in the MiSp from *N. antipodiana* are identical. At present, the exact number of repeats has not been determined yet because the repetitive feature of the $RP_{Mi}$ in DNA. We identified 5 types of linker domains with different size ranging from 83 to 174 aa in genomic DNA (Figure S2). Glycine (45–48%) and alanine (33–39%) are dominant in linker domains, which are consistent with previous reports [9,10]. $RP_{Mi}$ is highly conserved among different species (Figure S1B). Interestingly, the linker domain between the CTD and RP domains of *N. antipodiana* ($LK_{Mi}$) obtained here is much shorter than that of *N. clavipes* [9].

## Solution Structures of $CTD_{Mi}$, $RP_{Mi}$ and $LK_{Mi}$

In aqueous solution, $CTD_{Mi}$ formed a stable homodimer as evidenced by size exclusion chromatography (SEC, Figure S3). The structure of $CTD_{Mi}$ was determined using distance and dihedral restraints derived from multidimensional NMR spectroscopy (Figure 2A and Table 1). Overall, the structure of $CTD_{Mi}$ adopts a globular fold of two twisted five-helix bundles (α1 [$Gly^{18}$-$Leu^{28}$], α2 [$Ala^{31}$-$Val^{45}$], α3 [$Leu^{55}$-$Ser^{69}$], α4 [$Asp^{75}$-$Ser^{97}$], α5 [$Val^{107}$-$Met^{122}$] which pack in parallel to form a homodimer. α5 is swapped to stabilize the dimeric structure. The major dimer interface involves helices α1/α5', α4/α4' and α5/α1'. Many hydrophobic residues are located in the interface and are in close

contact, suggesting that hydrophobic interactions are the dominant factor for holding the two monomers together. Similarly, hydrophobic interactions among different helices in each monomer (involving 26 hydrophobic residues) are critical for the stability of the monomer. In addition, α4 is connected with α1 and α2 through two salt bridges R27-E77 and R36-E85 in each monomer. The formation of the R36-E85 salt bridge is evident from the extremely large chemical shift of $^{1}H_{\varepsilon}$ of R36 (11.7 ppm) and the observation of $^{1}H_{\eta 1}$ (10.5 ppm) and $^{2}H_{\eta 2}$ (5.8 ppm) of R36. Although $^{1}H_{\eta 1}$ and $^{2}H_{\eta 2}$ of R27 were not detected, the side-chains of R27 and E77 are in close proximity to be able to form a salt-bridge. Mutation of R27 into A27 reduced the transition temperature of thermal denaturation ($T_m$) by ~20°C (Figure S4), confirming the presence of the R27-E77 salt bridge.

The overall structure of $CTD_{Mi}$ is very similar to the previously reported structure of $CTD_{Ma}$ [21] with a Dali Z-score of 15. In addition, both $CTD_{Mi}$ and $CTD_{Ma}$ contain two intra-molecular salt bridges and have many hydrophilic residues located on the surface. Nevertheless, there are several key differences in local structures. 1) For $CTD_{Mi}$ dimer, eight negatively charged carboxyl groups (four in each monomeric unit: E32, D61, D75, D103) are exposed on the protein surface (Figure 2B), but no net charges on the surface of $CTD_{Ma}$ (Figure S5). Note that each $CTD_{Ma}$ monomer contains only two negatively charged carboxyl groups and two positively guanidinium groups which form two salt bridges [21] and are buried. 2) $CTD_{Mi}$ contains no cysteine residues and there is no intermolecular disulfide bridge, but one intermolecular disulfide bond exists in the $CTD_{Ma}$ dimer [21]. 3) There are more hydrophobic residues located in between α5 and α3 and between α5 and α1' in $CTD_{Mi}$ than in $CTD_{Ma}$ (Figure S6).

$LK_{Mi}$ (89 aa) contains 46.1% Gly and 32.3% Ala (Figure S7A). It was predicted to be intrinsically disordered (Figure 3A). Except the region of G54-Y70, most residues have hydrophobic scores larger than zero (Figure S7B), implying that $LK_{Mi}$ has low water solubility. To determine experimentally the secondary structure of $LK_{Mi}$, we tried to produce it in *E. coli*, but the production was not successful because it was degraded rapidly during the purification process. Thus we used the bi-domain fragment, LK-$CTD_{Mi}$. A comparison of the $^{1}H$-$^{15}N$ HSQC spectra of the LK-$CTD_{Mi}$ and $CTD_{Mi}$ reveals that the backbone $^{1}H$-$^{15}N$ correlation peaks for the residues from the LK domain are located in the range of 7.7–8.5 ppm in the $^{1}H$ dimension and most Gly and Ala $^{1}H$-$^{15}N$ correlations are clustered together (Figure 4A). This result shows that the LK domain is indeed intrinsically disordered. Except the correlation peaks from the N-terminal region of the isolated $CTD_{Mi}$ (e.g., V17, G18 and T20), other peaks from the isolated

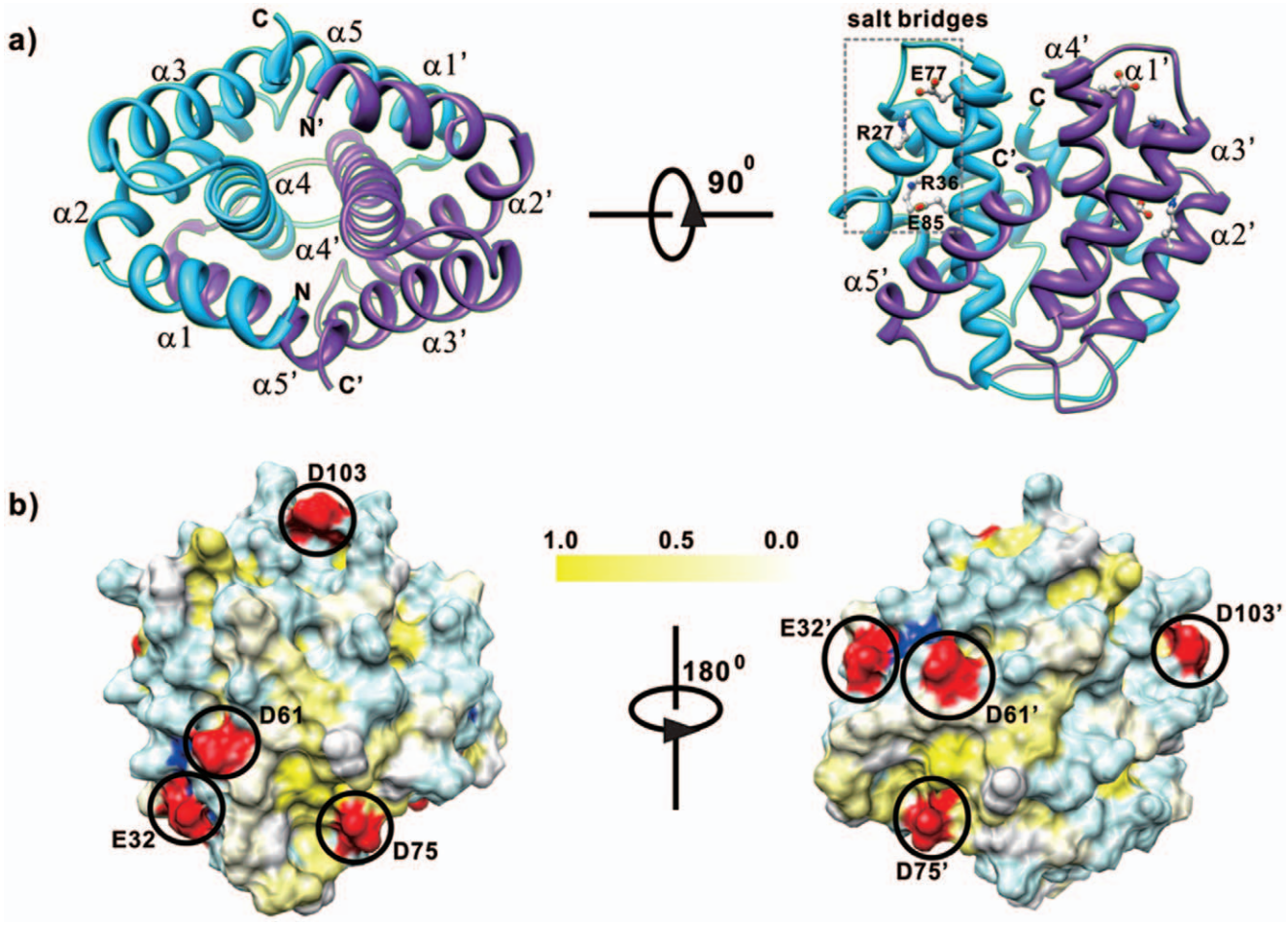


**Figure 2. Structure and surface plots of CTD.** a) Cartoon drawing of the lowest energy conformer of CTD. The two monomers are indicated by two different colors. b). Hydrophobic and charged surface of CTD. Hydrophobic residues are colored by a scale based on normalized hydrophobicity values [34]: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues are colored by blue, red and light blue.
doi:10.1371/journal.pone.0056142.g002

**Table 1.** Experimental restraints and structural statistics for ten lowest-energy NMR structures out of 100 calculated structures.

| **NMR distance constraints** | |
|---|---|
| Total intra-molecular NOE | 2697 |
| Intra-residue | 899 |
| Sequential | 736 |
| Medium-range (1<\|i-j\|<5) | 760 |
| Long-range (\|i-j\|≥5) | 302 |
| Inter-molecular NOE | 171 |
| Dihedral restraints* | 116 |
| Structural statistics | |
| Violation | mean±SD |
| Distance violation (Å) | 0.42±0.06 |
| Dihedral angle violation (°) | 3.46±0.58 |
| Max. dihedral angle violation (°) | 4.64 |
| Max. distance constraint violation (Å) | 0.49 |
| Ramachandran plot region (all residues) [%] | |
| most favored | 90.1 |
| additionally allowed | 9.3 |
| generously allowed | 0.3 |
| disallowed | 0.3 |
| Mean RMS deviation from the average coordinates (Å)† | |
| Backbone atoms ($C^{\alpha}$, C′, N, O) | 0.50±0.14 |
| All heavy atoms | 0.82±0.11 |
| MolProbity clash score | −6.48 |

*Dihedral angle constraints were generated by TALOS based on $C_{\alpha}$ and $C_{\beta}$ chemical shifts.

†Average rms deviation in the structural region (residue 17–123) was calculated among 10 refined structures.

doi:10.1371/journal.pone.0056142.t001

$CTD_{Mi}$ have the same $^{1}H$ and $^{15}N$ chemical shifts as those from the CTD in the bi-domain LK-$CTD_{Mi}$. Note that V17 is the N-terminal end residue of the CTD domain and is the connection site of the LK and CTD in the LK-$CTD_{Mi}$ construct. The signal of G48 in the bi-domain was weak and had the same chemical shifts as the G48 in the isolated CTD although it is not visible in Figure 4A. The results indicate that there are no or only weak interactions between the two domains at a protein concentration equal to or less than 0.5 mM. To evaluate the aggregation propensity of the $LK_{Mi}$, we predicted the aggregation-prone regions. The prediction shows that $LK_{Mi}$ contains three aggregation-prone regions with aggregation propensity scores >1: Y4-A12, A27-A37 and G67-A74 (Figure 3B).

$RP_{Mi}$ was folded into mainly α-helical structure in water as shown by its circular dichroism (CD) spectrum (Figure 5). It existed mainly in a monomeric form at low protein concentration (<1 mg/ml) in the presence of 100 mM NaCl, but mainly in a dimer together with small oligomers in the absence of salt as indicated by SEC (Figure S3). Very likely, the monomer, dimer and larger oligomers are in dynamic equilibrium since a single SEC peak was observed. The 1D $^{1}H$ NMR spectrum of $RP_{Mi}$ shows that this domain adopts a folded 3D structure since its methyl proton signals display good dispersion with one methyl at −0.13 ppm (Figure 4B). The line width of the peak at −0.13 ppm was 40 Hz, which was significantly larger than that for the methyl signal of $CTD_{Mi}$ at 0.03 ppm (26 Hz) (Figure 4B). Because the NMR line width is proportional to the molecular size, the apparent size of $RP_{Mi}$ must be significantly larger than the size of $CTD_{Mi}$ dimer under the condition of 0.6 mM protein, 10 mM sodium phosphate and pH 6.8. As mentioned earlier, $RP_{Mi}$ (128 aa) consists of only ~20% more aa than $CTD_{Mi}$ (107 aa). Therefore, the NMR result further shows that $RP_{Mi}$ should exist in equilibrium between dimers and small oligomers at low salt concentration (10 mM sodium phosphate) and relatively high protein concentration (0.6 mM). Due to the oligomerization-prone feature, we have not solved $RP_{Mi}$'s 3D structure yet.

At low protein concentrations (<1 mg/ml), RP-LK-$CTD_{Mi}$ existed as a dimer on the basis of the SEC data (Figure S3). The dimer should be mediated through the CTD domain since the dimerization of $CTD_{Mi}$ is independent of protein concentration. A comparison of the 1D $^{1}H$ NMR spectra of $RP_{Mi}$, $CTD_{Mi}$ and RP-LK-$CTD_{Mi}$ (Figure 4B) reveals that no strong interactions exist among the three different domains in RP-LK-$CTD_{Mi}$ because the isolated methyl signals from the tri-domain protein have nearly the same chemical shifts as those in the isolated individual domains. The line width of the methyl signals at 0.03 ppm and −0.13 ppm in the isolated $CTD_{Mi}$ and $RP_{Mi}$ were 26 Hz and 40 Hz respectively, very similar to those of the tri-domain fragment (27 Hz and 42 Hz) at a concentration of 0.6 mM, although the tri-domain is almost three times larger than the individual $CTD_{Mi}$ and $RP_{Mi}$ in molecular weight. This can be explained by the independent motions of the $CTD_{Mi}$ dimeric unit and $RP_{Mi}$ domain due to the high flexibility of the disordered linker domain. When the RP-LK-$CTD_{Mi}$ concentration was increased from ~0.6 mM to ~3 mM, the line width of the signal at 0.03 ppm (from the CTD) increased slightly from 27 Hz to 30 Hz, while the line width of the signal at −0.13 ppm (from the RP domain) increased dramatically from 42 Hz to 90 Hz. The result indicates that the tri-domain molecules assemble together to form oligomers through the weak association of RP and LK domains from different molecules.

## Stability of $CTD_{Mi}$, $RP_{Mi}$, LK-$CTD_{Mi}$ and RP-LK-$CTD_{Mi}$

Full length silk proteins are extremely water soluble and stable when stored in the silk glands [35]. To understand how silk proteins are stored stably at high concentration, we investigated the stability and solubility of individual protein domains and their dependences on salt and protein concentrations that change significantly when the proteins pass through the spinning duct [3]. Although $CTD_{Ma}$ and $CTD_{Mi}$ have similar overall structures, their chemical and thermal stabilities are significantly different. The transition midpoints in urea ($C_m$) and temperature ($T_m$) denaturation of $CTD_{Mi}$ were ~4.8 M urea and ~71°C, respectively (Figure 6A and Figure S4, blue line), which are significantly larger than those of $CTD_{Ma}$ (~2 M urea at 10 mM phosphate and ~2.8 M urea at 500 mM NaCl, 64°C at 10 mM phosphate) [21]. The result indicates $CTD_{Mi}$ is much more stable than $CTD_{Ma}$. Interestingly, NaCl had nearly no effect on the chemical stability of $CTD_{Mi}$ (Figure 6A), while NaCl could stabilize $CTD_{Ma}$ [21]. The stability of $CTD_{Mi}$ was independent of protein concentration when the concentration was below 0.2 mM, but $CTD_{Ma}$ was much more stable against urea denaturation at a protein concentration of 5 μM than 0.2 mM [21].

To examine the importance of the solvent-exposed charges to the stability of $CTD_{Mi}$, we prepared four conserved single-point mutants (E32Q, D61N, D75N and D103N) and one double-point mutant (E32Q/D75N). E32Q, D75N and D103N mutants showed significantly lower $C_m$ values than the wild type $CTD_{Mi}$ although the mutation of D61N had only a slight effect on the

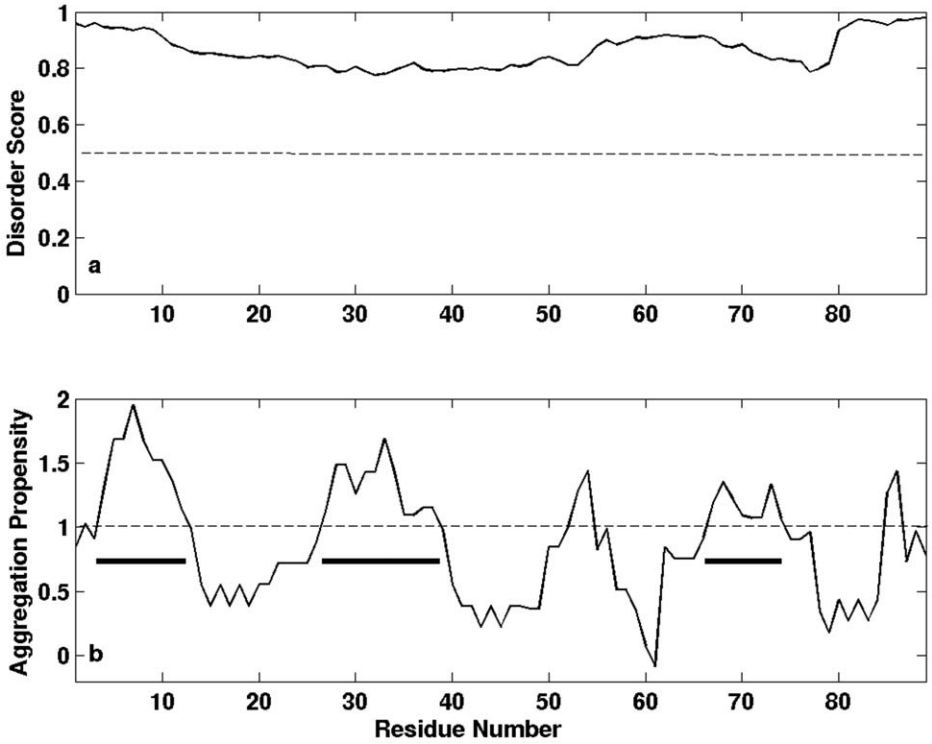


**Figure 3. Predicted disordered residues (a) and aggregation regions (b) in $LK_{Mi}$.** Disordered residues have scores >0.5. The aggregation-prone regions with aggregation propensity scores >1 are indicated by bars in panel b.
doi:10.1371/journal.pone.0056142.g003

stability. Moreover, double mutation reduced the $C_m$ from ~4.8 M to ~3.2 M urea (Figure 7, filled square and unfilled triangle). The results indicate that the solvent-exposed negatively charged residues are critical to the stability of $CTD_{Mi}$. Interestingly, these negatively charged residues are conserved or partially conserved in all MiSp, but absent in the $CTD_{Ma}$ of *A. diadematus* (Figure S1A). Besides the solvent-exposed negative charges, other factors such as hydrophobic interaction and hydrogen bonding which are slightly different in the two CTDs may also contribute to their significant difference in stability.

The chemical stability of LK-$CTD_{Mi}$ and $CTD_{Mi}$ was nearly identical (Figure 6A and 6B). This result shows that $LK_{Mi}$ has no obvious effects on $CTD_{Mi}$'s stability, implying that $LK_{Mi}$ does not interact with $CTD_{Mi}$ and confirming the conclusion drawn from the comparison of 2D HSQC spectra. The $T_m$ of LK-$CTD_{Mi}$ was about 4°C lower than that of $CTD_{Mi}$. This should not have resulted from the interaction of LK and CTD, but could be caused by the gradual slight aggregation of LK-$CTD_{Mi}$ during the temperature ramping process. It is noteworthy that a small of amount of precipitate was observed only for LK-$CTD_{Mi}$ and RP-LK-$CTD_{Mi}$ in the thermal denaturation process. Similar to $CTD_{Mi}$, LK-$CTD_{Mi}$ was not influenced in stability by salt when NaCl concentration was below 500 mM.

$RP_{Mi}$ also displayed a typical two-state unfolding profile, but showed a much lower stability than $CTD_{Mi}$ (Figure 6C). Its $C_m$ went up from 1.4 M to 2.3 M urea when NaCl concentration was increased from 0 to 500 mM. This salt-dependent stability is similar to $CTD_{Ma}$ but different from $CTD_{Mi}$. Thermal denaturation (Figure S8, black and blue lines) also shows $RP_{Mi}$ is much less stable than $CTD_{Mi}$ ($T_m$: 53 vs 71°C).

Unlike individual domains, the tri-domain fragment, RP-LK-$CTD_{Mi}$, unfolded in two steps with the increase of urea concentrations. The denaturation curves were fitted (Figure 6D), and the extracted first and second $C_m$ values were the same as the $C_m$ values of the isolated $RP_{Mi}$ and $CTD_{Mi}$, respectively. This suggests $RP_{Mi}$ tends to unfold first, followed by the unfolding of $CTD_{Mi}$. The result also indicates that the three domains of RP-LK-$CTD_{Mi}$ have no or very weak interactions in solution at low protein concentrations (<10 μM). This conclusion is consistent with that drawn from the NMR data analysis, and is further supported by the fact that the first-step unfolding of RP-LK-$CTD_{Mi}$ is obviously dependent on salt ($C_m$: 1.4–2.3 M) and the second-step unfolding is nearly independent of salt ($C_m$: 4.7–4.9 M) (Figure 6D). The thermal denaturation data (Figure S8, cyan line) also indicates RP-LK-$CTD_{Mi}$ unfolds in two steps in a non-cooperative way and the CTD in the tri-domain protein cannot stabilize the RP. Taken together, in spite of the high stability of $CTD_{Mi}$, the tri-domain protein can easily undergo conformational changes due to the low stability of $RP_{Mi}$ when the protein is in a dimeric structure. To achieve high stability, the tri-domain protein and full length MiSp may assemble to form high order structures like oligomers. Due to the absence of the RP domain in MaSp, MaSp and MiSp may use different mechanisms to achieve high stability.

## Solubility of $CTD_{Mi}$, $RP_{Mi}$, LK-$CTD_{Mi}$ and RP-LK-$CTD_{Mi}$

In 10 mM Tris buffer (pH 7.0), $CTD_{Mi}$, LK-$CTD_{Mi}$, RP-LK-$CTD_{Mi}$, $RP_{Mi}$ and RP-$LK_{Mi}$ could be concentrated to about 300, 200, 150, 60 and 5 mg/ml before the observation of precipitate or gel. In 10 mM sodium phosphate (pH 7.0), the solubility of each protein was nearly the same as that in 10 mM Tris, indicating that the solubility is not affected by buffer. RP-$LK_{Mi}$ had the lowest solubility and was prone to precipitate. Other domains or fragments did not precipitate during the concentration process,

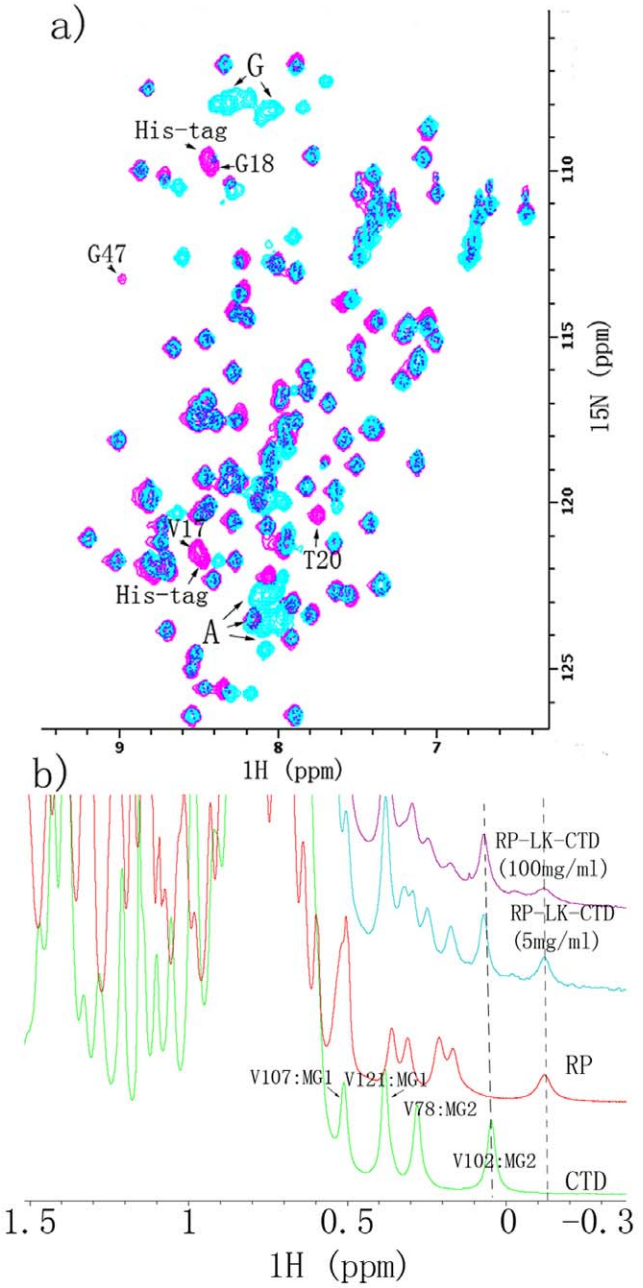


**Figure 4. NMR spectra of LK-CTD$_{Mi}$, CTD$_{Mi}$, RP$_{Mi}$ and RP-LK-CTD$_{Mi}$.** a). Overlay of 2D $^{1}$H-$^{15}$N HSQC spectra of isolated CTD$_{Mi}$ (red) and di-domain LK-CTD$_{Mi}$ (cyan). The signals from the His-tag in the isolated CTD$_{Mi}$ were labeled. Correlation peaks from Ala and Gly of the LK domain in LK-CTD$_{Mi}$ are clustered in small regions and are indicated by arrows. b). stacked plot of 1D $^{1}$H spectra (−0.5 ppm –1.6 ppm) of CTD$_{Mi}$ (green, 0.6 mM), RP$_{Mi}$ (red, 0.6 mM), RP-LK-CTD$_{Mi}$ (cyan, 0.6 mM) and RP-LK-CTD$_{Mi}$ (pink, 3 mM).
doi:10.1371/journal.pone.0056142.g004

but they formed gel when their concentrations were above their corresponding maxima. As shown in Figure 2B, CTD$_{Mi}$ is purely negatively charged and very polar on its surface. The electrostatic repulsion among negatively charged dimeric CTD$_{Mi}$ can prevent self-assembly for the formation of random aggregates. Therefore, the high hydrophilicity and unique charge of CTD$_{Mi}$ surface explains its extremely high solubility. On the basis of the high solubility and dimerization feature, CTD$_{Mi}$ has been used to

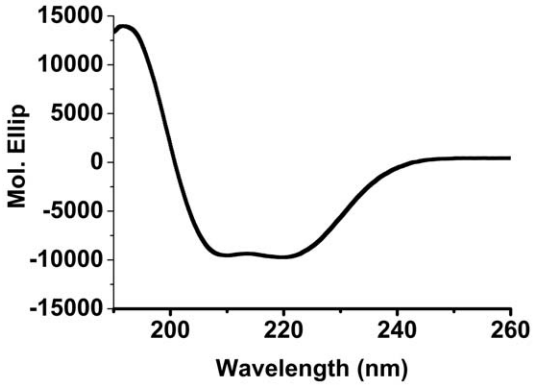


**Figure 5. Circular Dichroism (CD) spectrum of RP$_{Mi}$.**
doi:10.1371/journal.pone.0056142.g005

generate large sized silk-like proteins for strong silk fiber production [36].

Although RP$_{Mi}$ was easy to form small oligomers, its water solubility was still quite high. This may be achieved by burying some solvent-exposed hydrophobic patches through formation of oligomeric structures. LK-RP$_{Mi}$ was much less soluble than RP$_{Mi}$, demonstrating that the solubility of LK$_{Mi}$ should be significantly lower than 5 mg/ml. LK$_{Mi}$'s low solubility agrees with its high hydrophobicity (Figure S7B). Interestingly, LK-CTD$_{Mi}$ and RP-LK-CTD$_{Mi}$ were very soluble. This may be explained by the presence of the highly soluble CTD$_{Mi}$ through mutual compensation in solubility. Most likely, however, the high solubility of the LK-CTD$_{Mi}$ and RP-LK-CTD$_{Mi}$ is achieved through an alternative mechanism by forming oligomers. With this mechanism, the poorly soluble domains or fragments assemble to form oligomers through the aggregation-prone regions in LK$_{Mi}$ or/and RP$_{Mi}$, leading to partial burial of solvent-exposed hydrophobic regions and then resulting in high solubility of the entire protein. The presence of such oligomers in the sample of 3 mM RP-LK-CTD$_{Mi}$ is evidenced by the observation of the significant increase of the line width of methyl proton NMR signals from the RP domain rather than from the CTD domain (Figure 4B). Similar to RP-LK-CTD$_{Mi}$, the full length MiSp (which comprises about 15 repeats of RP-LK$_{Mi}$) may also exist in oligomers in the silk gland where the protein concentrations can reach up to ~50% w/w [35].

Our results also suggest that CTD$_{Mi}$ and RP-LK$_{Mi}$ play distinct roles in maintaining MiSp proteins in a highly water soluble form, i.e., RP-LK$_{Mi}$ initiates the oligomerization through weak hydrophobic interactions among LK$_{Mi}$ and RP$_{Mi}$ domains and forms the core region of the oligomers, while CTD$_{Mi}$ prevents MiSp from forming precipitate by staying outside the oligomer core. The structure of the oligomers formed by MiSp fragments seems quite different from that by TuSp fragments which resembles a micelle-like structure [22]. The different structures may result from the significant differences in the LK and CTD domains: LK$_{Mi}$ (89aa) is much larger and more hydrophobic than the linker region between the RP and CTD of TuSp1 (48aa); the isolated CTD of TuSp exists as oligomers but CTD$_{Mi}$ as dimers in aqueous solution [22]. MaSp was also proposed to form a micelle-like structure in which the repetitive domains are inside the micelle and CTD domains are outside [21]. Because of the significant difference in amino acid sequences, different types of silk proteins may use different ways to form high order structures for stable storage. In all the cases, however, CTDs are located outside the assembled structures to enhance the solubility of the assembled form. To

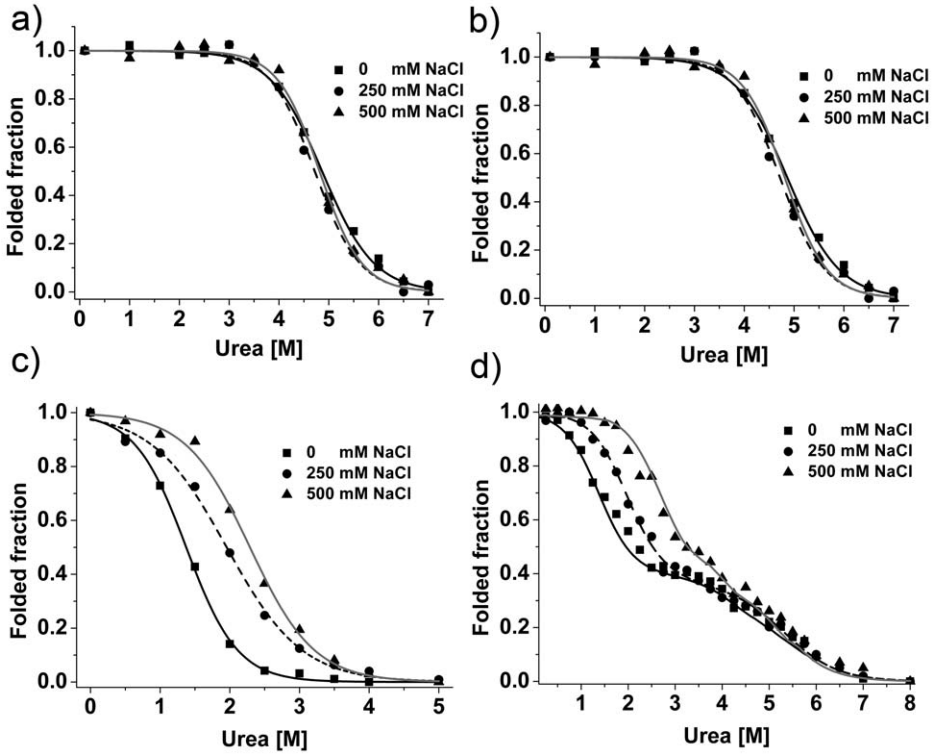


**Figure 6. Urea-induced denaturation of MiSp fragments under different NaCl concentrations monitored by CD.** a) $CTD_{Mi}$; b) LK-$CTD_{Mi}$; c) $RP_{Mi}$; d) RP-LK-$CTD_{Mi}$. The lines represent the fitting curves. All the data were normalized.
doi:10.1371/journal.pone.0056142.g006

fully understand why silk fibroins are highly soluble in silk glands, studies on the full length fibroins or large fragments including N- and C-terminal domains and several repetitive units are necessary.

## Stability against Shear Force

In the natural silk spinning process, silk proteins pass through the spinneret in the silk gland and then become silk fibers. During this spinning, the proteins undergo conformational changes after encountering shear and elongational forces [3].

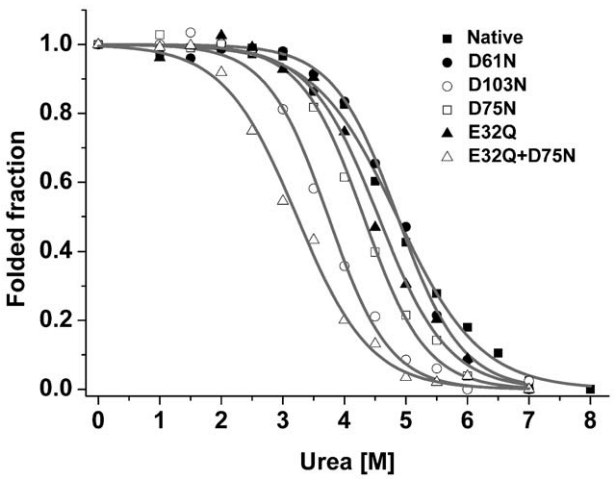


**Figure 7. Urea-induced denaturation of $CTD_{Mi}$ mutants. The solid lines represent the fitting curves.**
doi:10.1371/journal.pone.0056142.g007

Here, we studied the effect of shear force on protein stability and aggregation by stirring protein solutions. In the absence of stirring (mechanical shear force), $RP_{Mi}$, RP-$LK_{Mi}$, RP-LK-$CTD_{Mi}$, LK-$CTD_{Mi}$, $CTD_{Mi}$ and maltose binding protein (MBP) could maintain a soluble state under the condition of 0.05 mg/ml protein, 10 mM phosphate, pH 6.8 and 25°C without detectable precipitate within two days. In the presence of stirring, however, all of them tended to aggregate to form visible precipitate that is detectable at 350 nm. The changes in the amount of aggregated proteins with time are shown in Figure 8A. The $C_m$ and $T_m$ values of MBP in the absence of its ligand are 3.3 M urea and 63°C respectively [37], indicating MBP is more stable than $RP_{Mi}$ but less stable than $CTD_{Mi}$. Figure 8A shows that the aggregation rates of $RP_{Mi}$, $CTD_{Mi}$ and MBP are inversely proportional to their thermal or chemical stability. The aggregation should occur through partial protein unfolding and then assembly of the partially unfolded molecules. Note that the partially unfolded proteins have more solvent-exposed hydrophobic residues than the folded ones. Therefore the more stable a protein is, the slower the protein unfolding is, and the slower the shear-force-induced aggregation is.

The aggregation rates of $CTD_{Mi}$ and $RP_{Mi}$ were substantially accelerated by covalently linking $LK_{Mi}$ domain to them respectively (Figure 8A). This result can be explained by the high aggregation propensity and low water solubility of the $LK_{Mi}$ domain (Figure 3B and Figure S7B). Although $LK_{Mi}$ is intrinsically disordered, stirring still greatly enhanced the aggregation rate of LK-$CTD_{Mi}$ and RP-$LK_{Mi}$, implying that the aggregation-prone regions are partially protected in the bi-domain protein fragments and shear force can reduce the protection. The protection may be achieved by partial local folding of the aggregation regions of

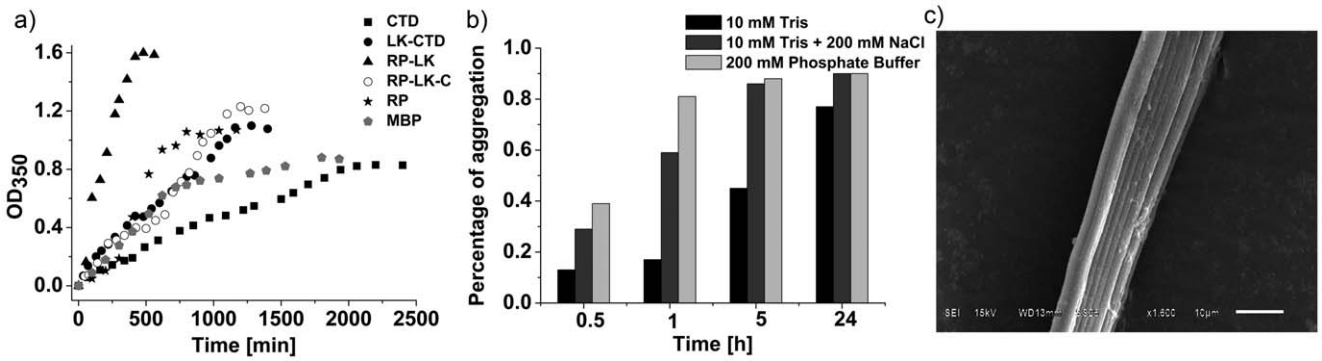


**Figure 8. Shear-force-induced aggregation of MiSp fragments and MBP monitored by $OD_{350}$ (a), dependence of aggregation of RP-LK-$CTD_{Mi}$ on salts (b), and SEM of silk fibers formed by RP-LK-$CTD_{Mi}$ (c).** The scale bar is 10 μm.
doi:10.1371/journal.pone.0056142.g008

$LK_{Mi}$ or by the weak interaction between the disordered domain and folded domain. In any cases, $LK_{Mi}$ plays a predominant role in the aggregation process of the protein fragments with an $LK_{Mi}$ domain. Although RP-$LK_{Mi}$, LK-$CTD_{Mi}$ and RP-LK-$CTD_{Mi}$ all contain an $LK_{Mi}$ domain, RP-$LK_{Mi}$ that is lacking a $CTD_{Mi}$ displayed the highest aggregation rate. This result shows that $CTD_{Mi}$ can slow down the aggregation rate and may play a role in regulating the assembly of silk protein molecules to form ordered structures.

Both NaCl and $Na_3PO_4$ were able to enhance the aggregation of RP-LK-$CTD_{Mi}$ in the presence of shear force in similar rates (Figure 8B). In the case of MaSp, the effect of NaCl on the aggregation of CTD and RP-CTD was much less pronounced than that of $Na_3PO_4$ [21]. This result suggests that the fibroin storage and/or assembly conditions in MiSp and MaSp spider glands may be different.

## Fiber Formation

All the single and bi-domain constructs underwent nonspecific aggregation or precipitation in aqueous solution upon gentle shaking. Under the same condition, however, RP-LK-$CTD_{Mi}$ could form small fibers with well-aligned structure and smooth surface even at a low protein concentration of ~0.3 mg/ml (Figure 8C). The diameters of the formed fibers ranged from ~2–10 μm, similar to that of the native MiSp silk [38]. The result reveals that all the three domains should participate in the fine-tuned process of fiber formation. Previous studies on MaSp and TuSp have shown that the minimum sequence requirement for a silk protein fragment to form silk fibers is that the fragment should contain a RP region and a terminal domain [20–22]. A recent study has revealed that the RP domain of aciniform spidroin alone could form silk fibers [39]. Therefore, MiSp fragments and full length MiSp may adopt a different fiber formation mechanism from other spider silk proteins.

Based on its low solubility (<<5 mg/ml), high hydrophobicity and aggregation propensity (Figure 3B, Figure S7B) and high aggregation rate in the presence of shear force (Figure 8A), $LK_{Mi}$ may act as a nucleation site to initiate the assembly of RP-LK-$CTD_{Mi}$ molecules through hydrophobic interactions among LK domains. Since $RP_{Mi}$ is prone to form oligomers and is unstable against shear force and chemical and thermal denaturation, it may assist the LK domain to assemble silk protein molecules together and play a dominant role in conformational changes upon shear force. In the absence of $CTD_{Mi}$, MiSp fragments such as RP-$LK_{Mi}$ and $RP_{Mi}$ formed only precipitate, indicating $CTD_{Mi}$ is essential to silk fiber formation. The folded $CTD_{Mi}$ may regulate the alignment of the assembled molecules by controlling the assembling rate since it can slow down the aggregation rate of RP-$LK_{Mi}$ (Figure 8A), which leads to controlled formation of well-defined fibers rather than non-specific aggregation.

## Conclusions

$CTD_{Mi}$, $RP_{Mi}$ and $LK_{Mi}$ have very distinct stability and solubility, which can be explained by their different structures, and each play specific roles in conferring the stable storage of MiSp fragments *in vitro* or full length MiSp in the silk gland. Due to the oligomerization-prone feature of $RP_{Mi}$ and $LK_{Mi}$, they are able to initiate the oligomerization through weak hydrophobic interactions among $LK_{Mi}$ and $RP_{Mi}$ domains and form the core region of the oligomers. On the other hand, because of the high solubility, $CTD_{Mi}$ may prevent the MiSp fragments or full length MiSp from forming precipitate by staying outside the oligomer core.

Shear force greatly accelerates protein aggregation through protein partial unfolding. In the presence of shear force, the aggregation rate of a folded protein is inversely proportional to its thermal or chemical stability; while the aggregation rate of a multi-domain protein containing both folded and disordered domains is determined mainly by the property of the disordered domain and the solubility of the entire protein. Although all MiSp domains investigated here could self-assemble in the presence of shear force, only the tri-domain RP-LK-$CTD_{Mi}$ formed well defined silk fibers, indicating that all three domains play distinct roles in fiber formation. According to our experimental data, we propose that the LK domain serves as a nucleation site to assemble different molecules together and CTD domains enable the arrangement of the assembled molecules in a highly ordered manner in the presence of shear force. Although MiSp, MaSp and TuSp fragments assemble in different ways, the relatively conserved CTD domains seem to play the same function, i.e., maintaining the assembled form in a highly soluble state before fiber formation and regulating the alignment of assembled molecules to form silk fibers. Due to the significant differences in biophysical properties among different types of CTDs and in primary structures and properties among different types of RPs, the molecular mechanisms of self-assembly and fiber formation for different types of silk proteins can be different. To reveal the detailed mechanisms, further studies on the structures of the assembled forms are required.

## Supporting Information

**Figure S1 Sequence alignments.** (a) C-terminal domains of MiSps from *Nephila antipodiana* (*N.a*), *Nephila clavipes* (*N.c*), *Latrodectus Hesperus* (*L.h*), *Lephilengys cruentata* (*L.c*) and *Uloborus diversus (U.d)* and MaSp from *Araneus diadematus* (*ADF-3*), (b) repetitive domains from *N.a, N.c, Nephilengys cruentata* (*N.c'*) and *Deinopis spinosa* (*D.s*).
(PDF)

**Figure S2 Sequence alignment of 5 types of linker domains from *Nephila antipodiana*.** Type 5 is the linker domain between the RP and CTD domains.
(PDF)

**Figure S3 Size exclusion chromatography profiles of $CTD_{Mi}$ (12.2 kDa), LK-$CTD_{Mi}$ (18.5 kDa), $RP_{Mi}$ (14.5 kDa), RP-$LK_{Mi}$ (20.8 kDa) and RP-LK-$CTD_{Mi}$ (33 kDa).** Molecular weight makers are indicated on the top. Except for one $RP_{Mi}$ (dashed curve) profile which was run in the presence of 100 mM NaCl, all other profiles were obtained under a buffer condition of 10 mM phosphate at pH 6.8.
(PDF)

**Figure S4 Temperature-induced denaturation of $CTD_{Mi}$ and its mutant.** The curves were fitted using Eq. 1. All samples contained 10 μM protein and 10 mM phosphate at pH6.8.
(PDF)

**Figure S5 Comparison of surface plots of $CTD_{Ma}$ (a) and $CTD_{Mi}$ (b).** Hydrophobic residues are colored by a scale based on normalized hydrophobicity values: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues (including all backbone and side-chain atoms) are colored by blue, red and light blue. Note that the exposed red and blue regions in the left panel are not from the charged carboxyl groups and guanidinium groups but from other parts of the charged residues.
(PDF)

**Figure S6 Comparison of hydrophobic interactions between α5 and α3 and between α5 and α1' for $CTD_{Ma}$ (a) and $CTD_{Mi}$ (b).** Yellow and green represent hydrophobic and non-hydrophobic residues, respectively. Here Thr is considered as hydrophobic.
(PDF)

**Figure S7 Amino acid sequence of $LK_{Mi}$ (a) and its hydrophobicity plot (b).**
(PDF)

**Figure S8 Temperature-induced unfolding of different MiSp fragments.** Except for RP-LK-CTD, the other curves were fitted using a two-state equation (Eq. 1). The curve for RP-LK-CTD was fitted using a linear combination of two two-state equations. All samples contained 10 μM protein and 10 mM phosphate buffer at pH 6.8.
(PDF)

## Author Contributions

Conceived and designed the experiments: DY. Performed the experiments: ZG ZL WH CCL JSF. Analyzed the data: ZG ZL. Wrote the paper: ZG ZL DY.

## References


1. Omenetto FG, Kaplan DL (2010) New opportunities for an ancient material. Science 329: 528–531.
2. Lewis RV (2006) Spider silk: Ancient ideas for new biomaterials. Chem Rev 106: 3762–3774.
3. Heim M, Keerl D, Scheibel T (2009) Spider silk: From soluble protein to extraordinary fiber. Angew Chem Int Ed 48: 3584–3596.
4. Greenwood MJ, McIntosh AR, Harding JS (2010) Behav Ecol 21: 1227–1235.
5. Spiess K, Lammel A, Scheibel T (2010) Recombinant spider silk proteins for applications in biomaterials. Macromol Biosc 10: 998–1007.
6. Rising A, Widhe M, Johansson J, Hedhammar M (2011) Spider silk proteins: Recent advances in recombinant production, structure-function relationships and biomedical applications. Cell Mol Life Sci 68: 169–184.
7. Teule F, Cooper AR, Furin WA, Bittencourt D, Rech EL, et al. (2009) A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning. Nature Protocols 4: 341–355.
8. Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY (2007) Blueprint for a high-performance biomaterial: Full-length spider dragline silk genes. Plos One 2.
9. Colgin MA, Lewis RV (1998) Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "Spacer regions". Protein Sci 7: 667–672.
10. Bittencourt D, Souto BM, Verza NC, Vineck F, Dittmar K, et al. (2007) Spidroins from the brazilian spider nephilengys cruentata (araneae: Nephilidae). Comp Biochem Physiol B-Biochem Mol Biol 147: 597–606.
11. Hayashi CY, Blackledge TA, Lewis RV (2004) Molecular and mechanical characterization of aciniform silk: Uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol Biol Evol 21: 1950–1959.
12. Huang W, Lin Z, Sin YM, Li D, Gong Z, et al. (2006) Characterization and expression of a cdna encoding a tubuliform silk protein of the golden web spider nephila antipodiana. Biochimie 88: 849–858.
13. Garb JE, Hayashi CY (2005) Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proc Nat Acad Sci USA 102: 11379–11384.
14. Tian MZ, Lewis RV (2005) Molecular characterization and evolutionary study of spider tubuliform (eggcase) silk protein. Biochemistry 44: 8006–8012.
15. Zhao AC, Zhao TF, SiMa YH, Zhang YS, Nakagaki K, et al. (2005) Unique molecular architecture of egg case silk protein in a spider, nephila clavata. J Biochem 138: 593–604.
16. Challis RJ, Goodacre SL, Hewitt GM (2006) Evolution of spider silks: Conservation and diversification of the c-terminus. Insect Mol Biol 15: 45–56.
17. Rising A, Hjalm G, Engstrom W, Johansson J (2006) N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. Biomacromolecules 7: 3120–3124.
18. Askarieh G, Hedhammar M, Nordling K, Saenz A, Casals C, et al. (2010) Self-assembly of spider silk proteins is controlled by a ph-sensitive relay. Nature 465: 236–U125.
19. Hedhammar M, Rising A, Grip S, Martinez AS, Nordling K, et al. (2008) Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from euprosthenops australis: Implications for fiber formation. Biochemistry 47: 3407–3417.
20. Stark M, Grip S, Rising A, Hedhammar M, Engstrom W, et al. (2007) Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. Biomacromolecules 8: 1695–1701.
21. Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, et al. (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. Nature 465: 239–U131.
22. Lin Z, Huang WD, Zhang JF, Fan JS, Yang DW (2009) Solution structure of eggcase silk protein and its implications for silk fiber formation. Proc Nat Acad Sci USA 106: 8906–8911.
23. Yang DW, Zheng Y, Liu DJ, Wyss DF (2004) Sequence-specific assignments of methyl groups in high-molecular weight proteins. J Am Chem Soc 126: 3710–3711.
24. Xu YQ, Long D, Yang DW (2007) Rapid data collection for protein structure determination by NMR spectroscopy. J Am Chem Soc 129: 7722–7723.
25. Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, et al. (1997) Methods for measurement of intermolecular noes by multinuclear NMR spectroscopy: Application to a bacteriophage lambda n-peptide/boxb rna complex. J Am Chem Soc 119: 6711–6721.
26. Xu YQ, Zheng Y, Fan JS, Yang DW (2006) A new strategy for structure determination of large proteins in solution without deuteration. Nature Methods 3: 931–937.
27. Herrmann T, Guntert P, Wuthrich K (2002) Protein nmr structure determination with automated noe assignment using the new software candid and the torsion angle dynamics algorithm dyana. J Mol Biol 319: 209–227.
28. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) Aqua and procheck-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8: 477–486.
29. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66: 778–95.
30. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35: W375–83.

31. Yang SA, Noble CG, Yang DW (2009) Characterization of dlc1-sam equilibrium unfolding at the amino acid residue level. Biochemistry 48: 4040–4049.
32. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) Pondr-fit: A meta-predictor of intrinsically disordered amino acids. Biochimi Biophy Acta 1804: 996–1010.
33. Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M, et al. (2005) Prediction of "Aggregation-prone" And "Aggregation-susceptible" Regions in proteins associated with neurodegenerative diseases. J Mol Biol 350: 379–392.
34. Roseman MA (1988) Hydrophilicity of polar amino-acid side-chains is markedly reduced by flanking peptide-bonds. J Mol Biol 200: 513–522.
35. Hijirida DH, Do KG, Michal C, Wong S, Zax D, et al. (1996) C-13 NMR of Nephila clavipes major ampullate silk gland. Biophy J 71: 3442–3447.
36. Lin Z, Deng Q, Liu XY, Yang D (2012) Engineered Large Spider Eggcase Silk Protein for Strong Artificial Fiber. Adv Mater. doi: 10.1002/adma.201204357. Epub ahead of print.
37. Beena K, Udgaonkar JB, Varadarajan R (2004) Effect of signal peptide on the stability and folding kinetics of maltose binding protein. Biochemistry 43: 3608–3619.
38. La Mattina C, Reza R, Hu X, Falick AM, Vasanthavada K, et al. (2008) Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver latrodectus hesperus. Biochemistry 47: 4692–4700.
39. Wang S, Huang W, Yang DW (2012) NMR structure note: repetitive domain of aciniform spidroin 1 from Nephila antipodiana. J Biomol NMR. 54: 415–20.