

Spider silk-like proteins derived from transgenic *Nicotiana tabacum*

Congyue Annie Peng · Julia Russo · Charlene Gravgaard · Heather McCartney · William Gaines · William R. Marcotte Jr.

Received: 5 June 2015 / Accepted: 3 March 2016 / Published online: 30 March 2016
© Springer International Publishing Switzerland 2016

Abstract The high tensile strength and biocompatibility of spider dragline silk makes it a desirable material in many engineering and tissue regeneration applications. Here, we present the feasibility to produce recombinant proteins in transgenic tobacco *Nicotiana tabacum* with sequences representing spider silk protein building blocks. Recombinant mini-spidroins contain native N- and C-terminal domains of major ampullate spidroin 1 (rMaSp1) or rMaSp2 flanking an abbreviated number (8, 16 or 32) of consensus repeat domains. Two different expression

plasmid vectors were tested and a downstream chitin binding domain and self-cleavable intein were included to facilitate protein purification. We confirmed gene insertion and RNA transcription by PCR and reverse-transcriptase PCR, respectively. Mini-spidroin production was detected by N-terminus specific antibodies. Purification of mini-spidroins was performed through chitin affinity chromatography and subsequent intein activation with reducing reagent. Mini-spidroins, when dialyzed and freeze-dried, formed viscous gelatin-like fluids.

Charlene Gravgaard and Heather McCartney have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:[10.1007/s11248-016-9949-1](https://doi.org/10.1007/s11248-016-9949-1)) contains supplementary material, which is available to authorized users.

C. A. Peng · J. Russo · C. Gravgaard · H. McCartney · W. Gaines · W. R. Marcotte Jr. (✉)
Department of Genetics and Biochemistry, Clemson University, 130 McGinty Court, 153 Robert F. Poole Agricultural Center, Clemson, SC 29634, USA
e-mail: marcotw@clemson.edu

Present Address:

C. Gravgaard
College of Pharmacy, Medical University of South Carolina, Charleston, SC 29425, USA

H. McCartney
Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA

Keywords Transgenic tobacco · Spider silk · Spidroin · Intein · Affinity chromatography

Introduction

Spiders can make up to six different proteinaceous silks that display a variety of remarkable physicochemical properties (Altman et al. 2003; Gosline et al. 1986). Spider silks, particularly dragline silk, have attracted much attention for their potential exploitation in the development of new materials (Allmeling et al. 2008; Hedhammar et al. 2010; Schacht and Scheibel 2011). The combination of strength and elasticity in dragline fibers suggests that materials based on these fibers will outperform currently available man-made materials such as steel, nylon and Kevlar (Gosline et al. 1999). Spider silks are also

biocompatible and biodegradable which makes them ideal materials for a variety of biomedical applications (Hedhammar et al. 2010). However, to take advantage of these remarkable qualities, sufficient amounts of the constituent proteins will need to be readily available. This cannot be accomplished by harvesting the spinning dope from spiders as a single spider produces <1 mg per day (Seidel et al. 1998). In addition, farming of spiders is impractical due to territorial and cannibalistic behavior.

The spider silk with the greatest mechanical toughness is major ampullate or dragline silk. It is made of two large, repetitive proteins, the major ampullate spidroins 1 and 2, MaSp1 and MaSp2 (Hinman and Lewis 1992; Xu and Lewis 1990). The full length sequences that encode these proteins have been unraveled (Ayoub et al. 2007). The general architecture of these proteins (Fig. 1) is similar in that both have large, central glycine- and alanine-rich repeat domains that consist of approximately 100 copies of shorter imperfect repeat blocks (Bini et al. 2004; Ayoub et al. 2007). MaSp1 and MaSp2 repeat blocks differ in the length (~30 and ~40 amino acids, respectively) and the nature of the repeat motifs. While each repeat block in MaSp1 and MaSp2 contains a poly-alanine tract, the amino acid triplet GGX is common in the repeat block of MaSp1 whereas MaSp2 contains proline rich motifs (GPGGY and GPGQQ) in the repeat domain (Bini et al. 2004). The repeat domain of MaSp1 is virtually devoid of proline (Hinman and Lewis 1992). The repeat domains of other silks (e.g. tubuliform and flagelliform silks), while conserved within and across species, differ dramatically from the MaSps (Huang et al. 2006). The

differences in repeat domains have been shown to be directly associated with the differences in tensile properties of various spider silks (Xia et al. 2010).

The repeat domains are flanked on both the N- and C-terminal sides by short non-repetitive sequences of approximately 150 amino acids each (N- and C-terminal domains; NTD and CTD). The NTDs and CTDs are conserved not only between MaSp1 and MaSp2 but also among other silks in the same species and with other species as well (Beckwitt and Arcidiacono 1994; Garb et al. 2010; Motriuk-Smith et al. 2005). This suggests that these non-repetitive domains have a common function in silk biology and the self-assembly process.

Recent studies on the functional role(s) of the NTD and CTD strongly suggest they are involved in regulating the self-assembly process. The CTD forms homodimers through one disulfide bond and two salt bridges (Ittah et al. 2007). Changes at the dimer interface and salt bridge region are critical in protein transition from soluble protein to solid fiber. NTD also forms homodimer through salt bridges (Hagn et al. 2010). The strength of the salt bridges is relevant to the ionic strength of the sodium chloride solution as demonstrated by a computer simulation study (Gronau et al. 2013). Reducing pH induces a local structure change that helps to stabilize NTD dimerization (Gaines et al. 2010) and initiate protein secondary structure transition from coil to β -sheet (Dicko et al. 2004). Inclusion of this domain in the recombinant protein construct may provide a valuable control point of the assembly process that is lacking in repeat domain-only fiber production. In addition, one or both terminal domains may also function as solubility

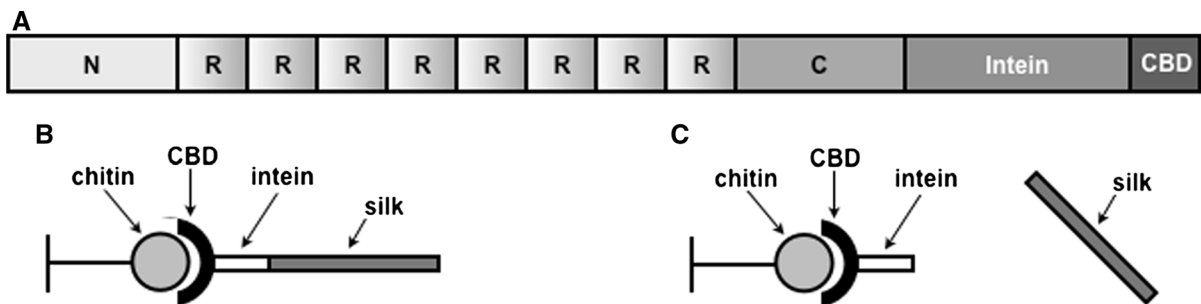


Fig. 1 Diagrammatic representation of recombinant mini-spidroin protein domain architecture and purification strategy. **a** Mini-spidroin structure. Each construct contains NTD and CTD flanking repeat domain multimers (R8 shown) followed by

the intein-CBD domain used for protein purification. **b**, **c** Purification strategy. rMaSp-intein-CBD protein pre-intein activation (**b**) and post-intein activation (**c**)

enhancers thereby allowing the high protein concentrations found in the lumen of the spider silk gland (Huemmerich et al. 2004a, b). While it has been demonstrated that fibers can be formed from repeat-only domains, inclusion of the NTD and CTD may help to stabilize proteins expressed in heterologous systems and may also allow the development of novel spinning technologies (Ittah et al. 2006).

Due to the limitations inherent in harvesting silks directly from spiders, there has been widespread use of genetic engineering for the production of recombinant spidroin-like proteins. The expression systems employed thus far include prokaryotes (*E. coli* and *Salmonella*: Arcidiacono et al. 1998; Lewis et al. 1996; Widmaier et al. 2009; Xia et al. 2010), lower eukaryotes (yeasts: Fahnestock and Bedzyk 1997; Gaines and Marcotte 2011), transgenic plants (*Arabidopsis*, potato, and tobacco: Scheller et al. 2001; Yang et al. 2005; Hauptmann et al. 2013a, b, 2015; Weichert et al. 2014; Menassa et al. 2004; Patel et al. 2007), mammalian cell lines (Lazaris et al. 2002), insects (silkworm: Miao et al. 2006; Teulé et al. 2011) and animals (mice and goats: Service 2002; Xu et al. 2007). Each of these approaches has met with some success although none has yet surmounted the largest problem, scalable expression and purification of recombinant spidroin-like proteins for materials development.

The use of a large biomass plant like tobacco provides advantages over other approaches for recombinant spidroin-like protein production. With the demonstration that these proteins can be purified from crude plant extracts (see below), optimization of the processes involved will make the approach essentially unlimited in scalability.

Here, we present the assembly of mini-spidroin genes that encode native MaSp1 and MaSp2 NTD and CTD sequences flanking an abbreviated number of the respective consensus repeat domains. Downstream of the mini-spidroin CTD is a self-cleavable modified minimal intein from the bacterium *Mycobacterium xenopi* (*Mxe*, Telenti et al. 1997) fused to a chitin binding domain (CBD; Evans et al. 1999). The *Mxe* intein variant used here (N198A) is disabled for cleavage at the C-terminus of the intein but retains the ability to cleave at the intein N-terminus when activated by a reducing agent. The CBD is derived from the C-terminal domain of chitinase A1 from *Bacillus circulans* WL-12 (Watanabe et al. 1994). We demonstrate that upon introduction into *Nicotiana*

tabacum, the genes are expressed and mini-spidroin proteins can be purified by chitin affinity chromatography and intein activation. Mini-spidroins were dialyzed and concentrated through freeze-drying, resulting in gelatin-like liquids.

Materials and methods

Mini-spidroin gene construction and generation of transgenic plants

A synthetic multi-cloning site (MCS), generated by annealing and extension of overlapping oligonucleotides, was used to replace the MCS in pT7T3 α -19. The synthetic MCS was used to assemble the mini-spidroin genes and contained the following restriction enzyme sites: HindIII-AscI-NcoI-XbaI-XmaI/SmaI-AgeI-NgoMIV-BglII-BamHI-NotI-EcoRI. Isolation of native NTD sequences for MaSp1A and MaSp2 has been described previously (Gaines and Marcotte 2008). Shorter fragments containing the native NTD sequences lacking the predicted signal peptide were generated using primers (Table 1) such that each NTD contained a SmaI/XmaI site at one end and an AgeI site at the other end. Similarly, shorter fragments containing the native CTD sequences were generated using primers (Table 1) such that each CTD contained an NgoMIV site at one end and an BglII site at the other end.

Alignment of published MaSp1 and MaSp2 repeat domains provided a consensus amino acid sequence for each:

NH₂-GGAGQGGYGGGLGGQGAGRGGQGA-GAAAAAA-COOH for MaSp1

NH₂-GPGQQGPGGYGPGQQGPG-GYGPQQGPGSPGSAAAAAAA-COOH for MaSp2

Reverse translation provided a nucleotide sequence optimized for tobacco codon usage. Consensus nucleotide repeats were created by annealing and extension of overlapping oligonucleotides (Table 1) such that the repeat domains contained an AgeI site at one end and an NgoMIV site at the other.

The *Mxe* *gyrA* intein and adjacent CBD were amplified from plasmid pTWIN1 (New England Biolabs, Cat. No. N6951S) using primers (Table 1)

Table 1 Oligonucleotide primers

Primer	Sequence
Sp syn MCS F	5'-GCAAGCTTATAGGGCGCGCCGCCCATGGGCTCTAGA ACTCCCGGGAGCACCGGTGCGG-3'
Sp syn MCS R	5'-CGAATTCGCGGCCGCTTAGGATCCACCAGATCTTGTGCC GGCCGCACCGGTGCTCCCGGG-3'
MaSp1A NTD F	5'-GCCCCCGGGCAAAACACCCCATGGTCA-3'
MaSp1A NTD R	5'-GCCACCGGTACCACCGTAAGATACTTCATT-3'
MaSp2 NTD F	5'-GGCCCCGGGCAAGCAAGGTCTCCATGG-3'
MaSp2 NTD R	5'-GCCACCGGTACCGCCGTATGATACTTCATT-3'
MaSp1 repeat F	5'-GTGACCGGTGGAGCTGGACAAGGAGGATATGGAGGAC TTGGAGGTCAGGGTGCAGGTAGAG-3'
MaSp1 repeat R	5'-CTCGCCGGCAGCAGCAGCAGCAGCTCCAGCTCCTTGT CCTCCTCTACCTGCACCCTGAC-3'
MaSp2 repeat F	5'-GGCACCGGTCCAGGTCAACAAGGACCTGGTGGATATG GTCCAGGTCAGCAAGGACCTGGAGGTTACGGACCAG-3'
MaSp2 repeat R	5'-GCCGCCGGCAGCAGCAGCAGCTGCTGCTGCTGATCCT GGTCCAGAAGGACCTTGCTGACCTGGTCCGTAACCTC-3'
MaSp1 CTD F	5'-CGGCCGGCGCCTCTGCAGCTGCTT-3'
MaSp1 CTD R	5'-GCCAGATCTATGATGATGATGATGATGCTCGAGACCAAG AGCTTGATAAATTGATTGGC-3'
MaSp2 CTD F	5'-GAGCCGGCTCCGCTGCAGCTTCTC-3'
MaSp2 CTD R	5'-GCCAGATCTATGATGATGATGATGATGCTCGAGAAATGC ACTCAAAACAGATTGGC-3'
Intein-CBD F	5'-GGCAGATCTTTTCAATATTGCATCACGGGAGATGCA-3'
Intein-CBD R	5'-GCCGGATCCCCTCCTG-3'

Terminal restriction sites are underlined. Internal overlap of MCS and repeat domain primers are in bold

to produce a fragment that contained a BglII site on one end and a BamHI site on the other end. After amplification and cloning, the single NgoMIV and AgeI sites contained in the region encoding the intein-CBD were mutagenized to facilitate later cloning. These changes result in a single amino acid change in the region containing the intein and CBD.

The nucleotide fragments described above were assembled into the synthetic MCS. The translational start and stop codons were provided by the NcoI site in the MCS and the pTWIN1-derived intein-CBD fragment, respectively. To generate multimers of the repeat domains, monomeric copies of the repeat sequence (containing AgeI and NgoMIV ends) were gel-purified, self-ligated and subsequently re-digested with AgeI and NgoMIV. Tetramers of the repeat domain were then gel-purified and ligated between the NTD and CTD coding regions. For larger order multimers (8 copies and 16 tandem copies of the repeat domain), shorter multimers were pyramided on top of one another.

Plant expression plasmids pKM12 (Dey and Maiti 1999) and pKLP36 (Maiti and Shepherd 1998) were obtained from Dr. Indu Maiti (Kentucky Tobacco Research and Development Center). Both contain caulimovirus promoter regions from (mirabilis mosaic and peanut chlorotic streak, respectively), an MCS and *rbcs* terminator as well as a kanamycin cassette for selection of transgenic plants. The expression plasmid MCS was replaced with the synthetic MCS described above to generate pKM12' and pKLP36'. Assembled mini-spidroin constructs were inserted into pKM12' and pKLP36' as AscI/NotI fragments. Constructs were named for the mini-spidroin and repeat number (e.g. pKM12-Sp1R8 contains the MaSp1 NTD, eight copies of the Sp1 repeat domain and the MaSp1 CTD followed by the intein CBD domain). Plasmid pKM12' and pKLP36' containing either MaSp1 (with eight or 16 copies of the repeat domain) or MaSp2 (with eight, 16, or 32 copies of the repeat domain) were generated and introduced into *Agrobacterium*. *Agrobacterium*-mediated transformation into *Nicotiana tabacum* and

selection were essentially as described (Fisher and Gultinan 1995). Primary transformants (T0) were selected on kanamycin-containing medium, transferred to soil and grown to maturity. Seeds from T0 plants were germinated on kanamycin-containing medium and the seedlings (T1) were transferred to soil. At least 12 T1 plants for each construct were evaluated for the presence of the transgenes using PCR for both NTD and intein-CBD regions and all were positive. At least two T1 plants per construct (up to six) were tested for protein expression by immunoblot (see “Materials and methods” section and below for details). For each construct, at least two protein expression-positive lines were confirmed except for MaSp2R8, for which only a single line was identified. The protein-positive T1 plants were grown to maturity and T2 plants were used for all subsequent analyses.

Protein extraction and purification

Leaf tissues (routinely multiples of 50 g) were ground either manually or using a blender in Tris-buffered saline (1X TBS; 20 mM Tris-HCl pH 7.4, 140 mM NaCl) supplemented with 0.1 % Tween 20 (1X TBST) and the proteinase inhibitors PMSF (1 mM) and TLCK (0.1 mM). The ratio of tissue (T, in g) to extraction buffer volume (E, in mls) was 1:10 (T:E) unless noted otherwise. For small scale extractions and screening purposes, four 0.5 cm leaf punches were manually ground in 250 μ l of extraction buffer. Tissue debris was removed from the crude extracts by either centrifugation (for leaf punches) or filtration through Miracloth (EMD Millipore) followed by 3MM filter paper (for larger scale extraction). Chitin resins (New England BioLabs Inc. #S6651L) were washed with 10 resin volumes of nano-pure water followed by five resin volumes of 1X TBS, pH 7.4 prior to addition to the crude extracts (1–2 ml resin/100 ml crude extract) and incubation at 4 °C overnight with mixing. The following day, the chitin beads were collected and washed with \geq 20 volumes of 1X TBST until visually most or all of the chlorophyll was removed. Mini-spidroins were released from the fusion proteins by incubating the chitin resins with three resin volumes of intein activation buffer (20 mM Tris-HCl, pH 8 supplemented with 500 mM NaCl, 30 mM DTT and 0.1 % Tween 20) at 16 °C for 40 h with mixing. Eluate was collected and the chitin beads washed twice with one resin volume each of 1X TBST. These

were pooled with the eluates after which the mini-spidroins were concentrated ten-fold using Pellicon XL 50 Ultrafiltration Cassettes (EMD Millipore, Cat No. PXC010C50). The concentrated mini-spidroins were dialyzed against 5 mM ammonium bicarbonate at 4 °C overnight. Dialysis tubing was moved into fresh buffer and dialysis was continued for an additional 8 h with fresh buffer every 4 h. The mini-spidroins were recovered from the tubing and pre-frozen at –80 °C. Freeze drying was performed using a Labconco lyophilizer (FreeZone 2.5) with the temperature and vacuum settings at –50 °C and 0.08 mBar, respectively, until all visible ice crystals were gone. The resulting material was a viscous gelatin-like fluid.

Immunodetection

Samples were separated on 8 % SDS–polyacrylamide gels and electro-transferred to PVDF membranes. Recombinant MaSp1A-NTD was used to generate a custom polyclonal antibody (Rockland Immunochemicals, Inc.). The crude antibody was affinity-purified using Affi-10-immobilized MaSp1A-NTD and then cross-adsorbed against total protein isolated from non-transgenic tobacco (Affi-10 and Affi-15 immobilized). The resulting antibody recognizes both MaSp1 and MaSp2 NTDs and was used as the primary antibody. The secondary antibody was goat anti-rabbit conjugated with alkaline phosphatase. The detection reagent for alkaline phosphatase was Lumi-Phos WB (Pierce, Cat. No. #34150) and membranes were imaged using a Fujifilm LAS-1000plus imager.

Results

Mini-spidroin coding regions were assembled as described in “Materials and methods” section. The encoded mini-spidroins consisted of MaSp1 or MaSp2 NTD and CTD sequences flanking either 8, 16 or 32 copies of the respective repeat domain and followed by the intein-CBD. Attempts to generate larger repeat domains have been unsuccessful. Because the coding regions differ only in the number and nature of the repeat domains, these are hereafter referred to as rMaSp1R8, rMaSp1R16, rMaSp2R8, rMaSp2R16 and rMaSp2R32. All cloning junctions were confirmed by sequencing. Each assembled mini-spidroin gene was

transferred into the plant expression plasmids pKM12' and pKLP36' and introduced into *Nicotiana tabacum*.

Purified rMaSp1A-NTD protein was used to generate and affinity purify an anti-NTD antibody. The purified antibody detects rGST-MaSp1A NTD (Fig. A1, Lane 1) but does not cross react with any protein from wild-type tobacco leaf crude extract (Fig. A1, Lane 2). Figure 2 and Fig. A1 (Lanes 3 and 4) show the antibody readily detects both rMaSp1 and rMaSp2 proteins containing 8 and 16 copies of the repeat domains in crude extracts from transgenic plants. The expected MWs for the full-length rMaSp1- and rMaSp2-intein-CBD fusions are indicated in the legend and the corresponding bands are marked on the image with asterisks. Interestingly, the intein domain in the rMaSp2-R8 fusion proteins appears to spontaneously activate at some level. This is evidenced by the presence of a polypeptide containing the MaSp2-NTD at a molecular weight consistent with removal of the 28 kDa intein-CBD domain (Fig. 2, Lane 2; Fig. 4b). A lower molecular weight band is also seen in Fig. 2, Lane 4 (rMaSp2R16) but the size is not consistent with simple removal of the intein-CBD

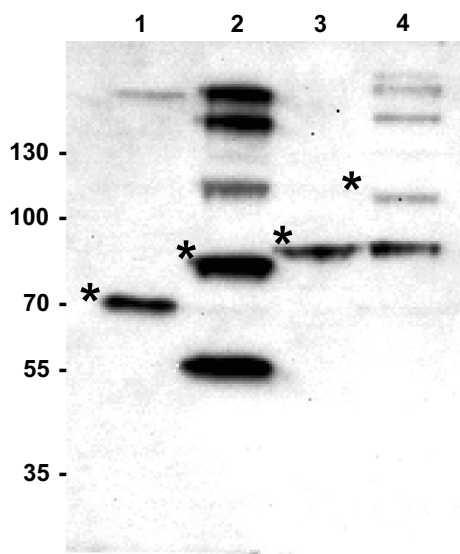


Fig. 2 Immunodetection of recombinant mini-spidroins in crude leaf extracts. Leaf punches were hand-ground in microfuge tubes in the presence of 250 μ l of extraction buffer. Each lane represents 25 μ l of crude extract (\sim 20 μ g TSP). Full length recombinant proteins are marked by asterisks. Lane 1, rMaSp1R8 (73 kDa); Lane 2, rMaSp2R8 (81 kDa); Lane 3, rMaSp1R16 (92 kDa); Lane 4, rMaSp2R16 (108 kDa). Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left

domain. The identity of this band is not clear but may indicate anomalous migration of rMaSp2 proteins with larger numbers of repeat domains. While the rMaSp1-R8 and -R16 samples in Fig. 2 do not display this same activity, it has also been observed in some rMaSp1 samples (Fig. A2). It is not clear if this activity is present in planta but, if that were true, one might expect all (or at least most) of the recombinant protein to be cleaved. However, there is clearly abundant full-length protein present in all crude extracts. This suggests that intein activation may be occurring at some point in the extraction/purification procedure or there may be a unique structural arrangement that can form under some circumstances in the rMaSp fusions that leads to intein activation/cleavage. It is also possible to see some bands at higher molecular weights, particularly in the rMaSp2 samples, that likely correspond to multimers of either full-length fusion proteins, mini-spidroins lacking the intein-CBD or combinations thereof.

Figure 3 shows immuno-detection of rMaSp2 proteins containing 32 copies of the repeat domain in crude extracts from both pKM12 and pKLP36 transgenic lines. The expected MW for the full-length rMaSp2R32-intein-CBD fusion is indicated in the legend. It seems that spontaneous activation of intein-CBD in the case of rMaSp2R32 protein is much less than seen for the rMaSp2R8 and rMaSp2R16 proteins. The signal strength of the rMaSp2R32 bands of the two individual pKLP36 lines (Fig. 3, Lanes 2, 3) indicates the protein amount varies in each transgenic line even when under control of the same promoter.

As part of the development of our purification protocol, we compared hand and blender grinding of leaf tissue at two different tissue:extract ratios (T:E = gram:ml). Based on immunodetection (Fig. A3), hand grinding at a T:E of 1:5 is slightly better than the other conditions tested and there was little difference in the laddering observed. The lower leaf tissue to buffer ratio does reduce the amount of extraction buffer which would be an advantage in scalable productions. Other than multimerization and intein spontaneous cleavage, protein degradation is minimized in our current process.

We also assessed our ability to affinity purify the recombinant mini-spidroins rMaSp1R8 and rMaSp2R8 and subsequently remove the CBD tag by intein activation by monitoring various points in the purification process (Fig. 4). It is clear that the crude extract

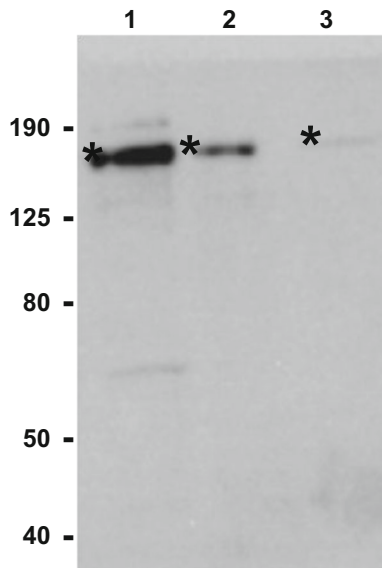


Fig. 3 Immunodetection of recombinant mini-spidroin rMaSp2R32 (162 and 134 kDa, with and without the intein-CBD, respectively) in crude leaf extracts. Leaf punches were hand-ground in microfuge tubes in the presence of 250 μ l of extraction buffer. Each lane represents 25 μ l of crude extract (\sim 20 μ g TSP). Full length recombinant proteins are marked by asterisks. Lane 1, pKM12'-Sp2R32 (Line 1.2.1); Lane 2, pKLP36'-Sp2R32 (Line 2.3); Lane 3, pKLP36'-Sp2R32 (Line 3.5). Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A-NTD. Molecular weight marker sizes (in kDa) are shown to left

was essentially depleted of full-length fusion protein post-chitin bead binding (Fig. 4a, b, Lanes 2 and 3). It is also possible to demonstrate that the full-length fusion proteins are efficiently bound to the chitin beads (Fig. 4a, b, Lanes 4, 15 μ l of a 50 % slurry of the beads). We routinely note the presence of cleaved fusion protein in the chitin bead fraction for the rMaSp2 proteins suggesting that additional spontaneous intein activation is occurring after binding but prior to incubation with intein activation buffer. After activation of the intein and concentration of the pooled eluates/wash fractions, the majority of the protein is present as monomer rMaSp protein (Fig. 4a, b Lanes 5). While there was some residual protein that remained bound to the beads post-cleavage (compare Lanes 4 and 6, 15 μ l of a 50 % slurry of the beads each), it was minimal. It is common to see variable amounts of laddering in different extractions of the rMaSp proteins. This is particularly true, and much more easily seen, in the rMaSp2 samples. It is not clear at this time if this is

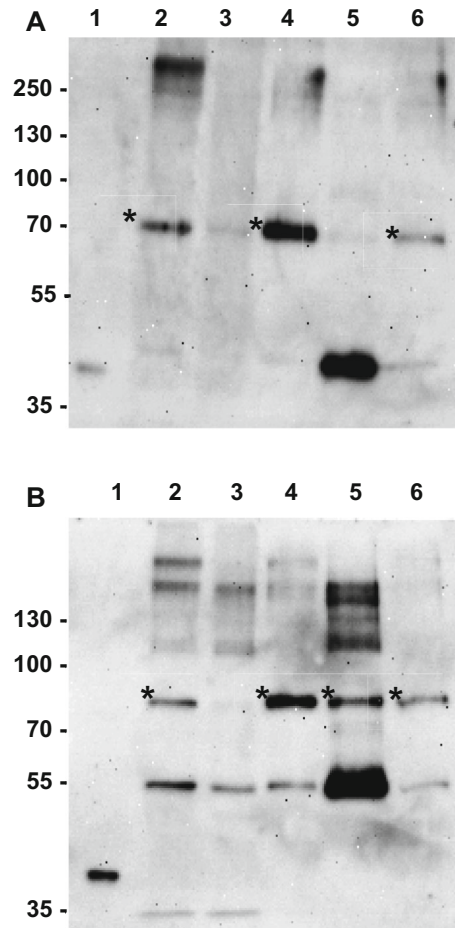


Fig. 4 Purification of recombinant mini-spidroins from crude leaf extracts. a rMaSp1R8; B, rMaAp2R8. Lanes 1, positive control rGST-MaSp1A NTD fusion protein (43 kDa), 12 ng; Lanes 2, 25 μ l of crude extracts (\sim 20 μ g TSP). Full length recombinant proteins are marked by asterisks; Lanes 3, crude extract post chitin binding (25 μ l, \sim 20 μ g TSP); Lanes 4, chitin beads post binding (15 μ l of 50 % slurry); Lanes 5, pooled intein cleavage buffer and wash post concentration (5 μ l, \sim 15 μ g protein); Lanes 6, chitin beads post intein activation and wash (15 μ l of 50 % slurry). Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left

due to differences in individual extractions from a given plant or to differences among siblings in a particular line.

Mini-spidroin yield in percentage of total soluble proteins were estimated based on the extraction and purification process shown in Fig. 4. Expression levels for untargeted (cytoplasmic) mini-spidroin yields in tobacco leaves are estimated to be 0.7 and 1.9 % TSP

for rMaSp1R8 and rMaSp2R8, respectively (Table 2). Purified mini-spidroins were dialyzed against 5 mM ammonium bicarbonate and lyophilized. After extensive freeze drying, the mini-spidroins formed viscous gelatin-like fluids.

Discussion

Protein production in transgenic plants

Plants are attractive hosts for the production of recombinant proteins for a variety of reasons including increased likelihood of appropriate post-translational modifications compared to prokaryotic expression systems and elimination of the risk of contaminating human/mammalian pathogens, a serious concern in many eukaryotic expression systems. Plants expressing significant levels of recombinant proteins also have the potential to provide a cost-effective way to scale up production.

However, target protein production levels can vary greatly across different transgenic lines (Fig. 3) and even among individuals within a single transgenic line. In some cases, recombinant protein is no longer detectable after a few generations even when the seedlings are germinated under selection. This may be due to gene silencing, a common obstacle for plant-derived proteins and can be especially true if the gene contains highly repetitive sequences (Stam et al. 1997). And, as has been noted recently, the cost and efficiency of downstream technical aspects of “molecular farming” can present significant challenges. In an elastin-like peptide (ELP)—MaSp1 repeat fusion protein study, transglutaminase was required to crosslink lysine and glutamine residue *in vitro* to produce a multimerized nearly native sized ELP-silk-like protein (Weichert et al. 2014).

Despite these concerns, plants continue to be a frequent choice for heterologous gene expression. Not surprisingly, targeting recombinant protein production to various locations/organelles in plants has been

shown to have an effect on expression levels. Expression of a spider silk-like protein (tandem MaSp1 repeat domains) in either the apoplast or ER lumen of Arabidopsis leaves resulted in 8.5 and 6.7 % total soluble protein (TSP), respectively (Yang et al. 2005). That same study found that expression levels in seeds harboring the ER targeting construct can be even higher (18 % TSP). Our yields in tobacco leaves are estimated to be 0.7 and 1.9 % TSP for rMaSp1R8 and rMaSp2R8, respectively. This is significantly lower than the levels cited above for Arabidopsis but our expression levels are comparable to the 2 % TSP yield of an ER-targeted MaSp1 repeat domain construct produced in tobacco leaves (Scheller et al. 2001).

We also see rMaSp-NTD positive bands at higher molecular weights in the crude leaf extracts and at some steps in the purification process that likely correspond to self-assembled multimers. This type of aggregation, that may also display laddering, has been reported by others (Menassa et al. 2004; Sponner et al. 2004, 2005; Guehrs et al. 2008) and these could represent associations of full-length fusion proteins, recombinant proteins lacking the intein-CBD or combinations thereof.

The rMaSp proteins expressed here are qualitatively different than those in many other studies in that we have incorporated native spider silk NTD and CTD domains (in addition to the intein-CBD self-cleavable affinity tag). Spider silk-like proteins with only the repeat domain frequently display solubility issues due to the high hydrophobicity of the repeat domains or premature aggregation (Ramensee et al. 2008; Zhang et al. 2008) whereas our mini-spidroins with flanking NTD and CTD seems easy to be readily soluble. Dramatically, the highly concentrated, purified proteins appear to retain water and remain as a gelatin-like fluid instead of powdered form after lyophilization. The NTD and CTD have been speculated to enhance solubility of spidroin proteins (Gao et al. 2013) (Huemmerich et al. 2004a, b), and our observations are consistent with that hypothesis.

Table 2 Mini-spidroin yield

	Crude extract		Purified mini-Spidroin		Yield (% TSP)
	TSP (mg/ml)	Volume (ml)	TSP (mg/ml)	Volume (ml)	
rMaSp1R8	0.61	1900	2.88	2.97	0.74
rMaSp2R8	0.254	950	3.66	1.25	1.89

TSP total soluble protein

The availability of both MaSp1 and MaSp2 minispidroins with variable copies of the repeat domains enables us to evaluate their assembly into fibers as individual proteins and in various combinations. Thus, we can examine the distinct features of MaSp1 and MaSp2 proteins and the effect of repeat domain numbers on the mechanical properties of a broad range of biomaterials including fibers, hydrogels, coating materials, etc. These materials are likely to have many applications in medical and engineering related fields.

Acknowledgments The authors would like to thank Dr. Indu Maiti at the Kentucky Tobacco Research and Development Center for the tobacco expression plasmids and Allison Nelson for technical support. This work was funded by grant 1R15EB007403-01 from the National Institute of Biomedical Imaging and Bioengineering to WRM.

References

- Allmeling C, Jokuszies A, Reimers K, Kall S, Choi CY, Brandes G, Kasper C, Scheper T, Guggenheim M, Vogt PM (2008) Spider silk fibres in artificial nerve constructs promote peripheral nerve regeneration. *Cell Prolif* 41:408–420
- Altman GH, Diaz F, Jakuba C, Calabro T, Horan RL, Chen J, Lu H, Richmond J, Kaplan DL (2003) Silk-based biomaterials. *Biomaterials* 24:401–416
- Arcidiacono S, Mello C, Kaplan D, Cheley S, Bayley H (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. *Appl Microbiol Biotechnol* 49:31–38
- 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*. doi:10.1371/journal.pone.0000514
- Beckwitt R, Arcidiacono S (1994) Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentarius* (Araneidae). *J Biol Chem* 269:6661–6663
- Bini E, Knight DP, Kaplan DL (2004) Mapping domain structures in silks from insects and spiders related to protein assembly. *J Mol Biol* 335:27–40
- Dey N, Maiti I (1999) Structure and promoter/leader deletion analysis of mirabilis mosaic virus (MMV) full-length transcript promoter in transgenic plants. *Plant Mol Biol* 40:771–782
- Dicko C, Vollrath F, Kenney JM (2004) Spider silk protein refolding is controlled by changing pH. *Biomacromolecules* 5:704–710
- Evans TC, Benner J, Xu M-Q (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J Biol Chem* 274:18359–18363
- Fahnestock SR, Bedzyk LA (1997) Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol* 47:33–39
- Fisher D, Gultinan M (1995) Rapid, efficient production of homozygous transgenic tobacco plants with *Agrobacterium tumefaciens*: a seed-to-seed protocol. *Plant Mol Biol Rep* 13:278–289
- Gaines IV, Marcotte WR Jr (2008) Identification and characterization of multiple Spidroin 1 genes encoding major ampullate silk proteins in *Nephila clavipes*. *Insect Mol Biol*. doi:10.1111/j.1365-2583.2008.00828.x
- Gaines WA, Marcotte WR Jr (2011) Recombinant dragline silk-like proteins-expression and purification. *AATCC Rev* 11:75–79
- Gaines WA, Sehorn MG, Marcotte WR Jr (2010) Spidroin N-terminal domain promotes a pH-dependent association of silk proteins during self-assembly. *J Biol Chem* 285:40745–40753
- Gao Z, Lin Z, Huang W, Lai CC, Fan JS, Yang D (2013) Structural characterization of minor ampullate spidroin domains and their distinct roles in fibroin solubility and fiber formation. *PLoS One*. doi:10.1371/journal.pone.0056142
- Garb JE, Ayoub NA, Hayashi CY (2010) Untangling spider silk evolution with spidroin terminal domains. *BMC Evol Biol*. doi:10.1186/1471-2148-10-243
- Gosline JM, DeMont ME, Denny MW (1986) The structure and properties of spider silk. *Endeavour* 10:37–43
- Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function. *J Exp Biol* 202:3295–3303
- Gronau G, Qin Z, Buehler MJ (2013) Effect of sodium chloride on the structure and stability of spider silk's N-terminal protein domain. *Biomater Sci* 1:276–284
- Guehrs KH, Schlott B, Grosse F, Weisshart K (2008) Environmental conditions impinge on dragline silk protein composition. *Insect Mol Biol* 17:553–564
- Hagn F, Eisoldt L, Hardy JG, Vendrely C, Coles M, Scheibel T, Kessler H (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* 465:239–242
- Hauptmann V, Weichert N, Rakhimova M, Conrad U (2013a) Spider silks from plants—a challenge to create native-sized spidroins. *Biotechnol J* 8:1183–1192
- Hauptmann V, Weichert N, Menzel M, Knoch D, Paegle N, Scheller J, Spohn U, Conrad U, Gils M (2013b) Native-sized spider silk proteins synthesized in planta via intein-based multimerization. *Transgenic Res* 22:369–377
- Hauptmann V, Menzel M, Weichert N, Reimers K, Spohn U, Conrad U (2015) In planta production of ELPylated spidroin-based proteins results in non-cytotoxic biopolymers. *BMC Biotechnol* 15:9. doi:10.1186/s12896-015-0123-2
- Hedhammar M, Bramfeldt H, Baris T, Widhe M, Askarieh G, Nordling K, Aulock S, Johansson J (2010) Sterilized recombinant spider silk fibers of low pyrogenicity. *Biomacromolecules* 11:953–959
- Hinman MB, Lewis RV (1992) Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J Biol Chem* 267:19320–19324
- Huang W, Lin Z, Sin YM, Li D, Gong Z, Yang D (2006) Characterization and expression of a cDNA encoding a tubuliform silk protein of the golden web spider *Nephila antipodiana*. *Biochimie* 88:849–858

- Huemmerich D, Helsen CW, Quedzuweit S, Oschmann J, Rudolph R, Scheibel T (2004a) Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43:13604–13612
- Huemmerich D, Scheibel T, Vollrath F, Cohen S, Gat U, Ittah S (2004b) Novel assembly properties of recombinant spider dragline silk proteins. *Curr Biol* 14:2070–2074
- Ittah S, Cohen S, Garty S, Cohn D, Gat U (2006) An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. *Biomacromolecules* 7:1790–1795
- Ittah S, Michaeli A, Goldblum A, Gat U (2007) A model for the structure of the C-terminal domain of dragline spider silk and the role of its conserved cysteine. *Biomacromolecules* 8:2768–2773
- Lazaris A, Arcidiacono S, Huang Y, Zhou JF, Duguay F, Chretien N, Welsh EA, Soares JW, Karatzas CN (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* 295:472–476
- Lewis RV, Hinman M, Kothakota S, Fournier MJ (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. *Protein Expr Purif* 7:400–406
- Maiti IB, Shepherd RJ (1998) Isolation and expression analysis of peanut chlorotic streak Caulimovirus (PCISV) full-length transcript (FLt) promoter in transgenic plants. *BBRC* 244:440–444
- Menassa R, Zhu H, Karatzas CN, Lazaris A, Richman A, Brandle J (2004) Spider dragline silk proteins in transgenic tobacco leaves: accumulation and field production. *Plant Biotechnol J* 2:431–438
- Miao Y, Zhang Y, Nakagaki K, Zhao T, Zhao A, Meng Y, Nakagaki M, Park EY, Maenaka K (2006) Expression of spider flagelliform silk protein in *Bombyx mori* cell line by a novel Bac-to-Bac/BmNPV baculovirus expression system. *Appl Microbiol Biotechnol* 71:192–199
- Motriuk-Smith D, Smith A, Hayashi CY, Lewis RV (2005) Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules* 6:3152–3159
- Patel J, Zhu H, Menassa R, Gyenis L, Richman A, Brandle J (2007) Elastin-like polypeptide fusions enhance the accumulation of recombinant proteins in tobacco leaves. *Transgenic Res* 16:239–249
- Rammensee S, Slotta U, Scheibel T, Bausch AR (2008) Assembly mechanism of recombinant spider silk proteins. *Proc Natl Acad Sci USA* 105:6590–6595
- Schacht K, Scheibel T (2011) Controlled hydrogel formation of a recombinant spider silk protein. *Biomacromolecules* 12:2488–2495
- Scheller J, Guhrs KH, Grosse F, Conrad U (2001) Production of spider silk proteins in tobacco and potato. *Nat Biotechnol* 19:573–577
- Seidel A, Liivak O, Jelinski LW (1998) Artificial spinning of spider silk. *Macromolecules* 31:6733–6736
- Service, R.F (2002) Materials science. Mammalian cells spin a spidery new yarn. *Science* 295:419–421
- Sponner A, Unger E, Grosse F, Weisshart K (2004) Conserved C-termini of spidroins are secreted by the major ampullate glands and retained in the silk thread. *Biomacromolecules* 5:840–845
- Sponner A, Schlott B, Vollrath F, Unger E, Grosse F, Weisshart K (2005) Characterization of the protein components of *Nephila clavipes* dragline silk. *Biochemistry* 44:4727–4736
- Stam M, Mol JNM, Kooter JM (1997) Review article: the silence of genes in transgenic plants. *Ann Bot* 79:3–12
- Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WR, Perler FB (1997) The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. *J Bacteriol* 179:6378–6382
- Teulé F, Miao Y-G, Sohn B-H, Kim Y-S, Hull JJ, Fraser MJ, Lewis RV, Jarvis DL (2011) Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. *Proc Natl Acad Sci USA* 109:923–928
- Watanabe T, Ito Y, Yamada T, Hashimoto M, Sekine S, Tanaka H (1994) The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. *J Bacteriol* 176:4465–4472
- Weichert N, Hauptmann V, Menzel M, Schallau K, Gunkel P, Hertel TC, Pietzsch M, Spohn U, Conrad U (2014) Transglutamination allows production and characterization of native-sized ELPylated spider silk proteins from transgenic plants. *Plant Biotechnol J* 12:265–275
- Widmaier DM, Tullman-Ercek D, Mirsky EA, Hill R, Govindarajan S, Minshull J, Voigt CA (2009) Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol Syst Biol*. doi:10.1038/msb.2009.62
- Xia XX, Qian ZG, Ki CS, Park YH, Kaplan DL, Lee SY (2010) Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proc Natl Acad Sci USA* 107:14059–14063
- Xu M, Lewis RV (1990) Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci USA* 87:7120–7124
- Xu HT, Fan BL, Yu SY, Huang YH, Zhao ZH, Lian ZX, Dai YP, Wang LL, Liu ZL, Fei J et al (2007) Construct synthetic gene encoding artificial spider dragline silk protein and its expression in milk of transgenic mice. *Anim Biotechnol* 18:1–12
- Yang J, Barr LA, Fahnstock SR, Liu ZB (2005) High yield recombinant silk-like protein production in transgenic plants through protein targeting. *Transgenic Res* 14:313–324
- Zhang Y, Hu J, Miao Y, Zhao A, Zhao T, Wu D, Liang L, Miikura A, Shiomi K, Kajijura Z et al (2008) Expression of EGFP-spider dragline silk fusion protein in BmN cells and larvae of silkworm showed the solubility is primary limit for dragline proteins yield. *Mol Biol Rep* 35:329–335