Diverse Formulas for Spider Dragline Fibers Demonstrated by Molecular and Mechanical Characterization of Spitting Spider Silk

ARTICLE *in* BIOMACROMOLECULES · OCTOBER 2014

Impact Factor: 5.75 · DOI: 10.1021/bm501409n · Source: PubMed

2 AUTHORS, INCLUDING:

Sandra Correa-Garhwal

University of California, Riverside

4 PUBLICATIONS **6** CITATIONS

Article

pubs.acs.org/Biomac

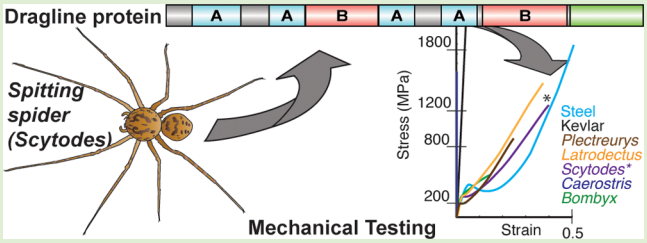
Diverse Formulas for Spider Dragline Fibers Demonstrated by Molecular and Mechanical Characterization of Spitting Spider Silk

Sandra M. Correa-Garhwal*,† and Jessica E. Garb*

Department of Biological Sciences, University of Massachusetts—Lowell, Lowell, Massachusetts 01854, United States

Supporting Information

ABSTRACT: Spider silks have outstanding mechanical properties. Most research has focused on dragline silk proteins (major ampullate spidroins, MaSps) from orb-weaving spiders. Using silk gland expression libraries from the haplogyne spider *Scytodes thoracica*, we discovered two novel spidroins (*S. thoracica* fibroin 1 and 2). The amino acid composition of *S. thoracica* silk glands and dragline fibers suggest that fibroin 1 is the major component of *S. thoracica* dragline silk. Fibroin 1 is dominated by glycine-alanine motifs, and lacks sequence motifs associated with orb-weaver MaSps. We hypothesize fibroin 2 is a piriform or aciniform silk protein, based on amino acid composition, spigot morphology, and phylogenetic analyses. *S. thoracica*'s dragline silk is less tough than previously reported, but is still comparable to other dragline silks. Our analyses suggest that dragline silk proteins evolved multiple times. This demonstrates that spider dragline silk is more diverse than previously understood, providing alternative high performance silk designs.



INTRODUCTION

Spider silks are renowned for their impressive mechanical properties. For example, some spider silks are tougher than high-performance, man-made materials such as steel and Kevlar.[1,2] Spider silks are also prized because they are lightweight and nonimmunogenic.[3] Thus, spider silks have been studied extensively for use in applications ranging from the production of medical devices to high-performance textiles.[4] Most applied spider silk research has focused on replicating the properties of major ampullate (dragline) silk from araneoid species (ecribellate orb-weavers) because of its high toughness and tensile strength in comparison to other silks.[2−4] Characterization of the molecular and mechanical diversity of major ampullate and other silk types from distantly related spider species has been relatively limited, and these neglected materials present novel designs for silk-based applications.

Spider silks are largely composed of proteins that are expressed in abdominal silk glands. The stored silk proteins travel through glandular ducts that connect to external spigots on the spinnerets, from which fibers are extruded.[5,6] The different types of silks a spider can spin are each produced by sets of morphologically distinct silk glands, and the number and morphology of these glands varies tremendously across species. An individual araneoid spider synthesizes up to seven functionally different silk types from the following glands: (1) major ampullate, (2) minor ampullate, (3) flagelliform, (4) tubuliform, (5) aciniform, (6) piriform, and (7) aggregate.[6] By contrast, the distantly related Mygalomorphae (tarantulas and their kin) only have one or two types of silk glands.[7]

Spider silk is predominantly composed of spidroins, a unique protein family with members that are differentially expressed in the various types of silk glands.[8,9] Spidroins are large proteins (2000−6000 amino acids[10−13]) and have a primary structure that largely consists of repetitive sequences (tandem repeats) flanked by short (∼100−150 amino acids), conserved nonrepetitive amino and carboxyl terminal domains (Figure 1A).[14,15] Tandem repeats for orb-weaver MaSps (major ampullate spidroins) are ensembles of simple amino acid sequence motifs, such as GGX (where X is a small subset of amino acids), GPGXX, polyA (uninterrupted strings of A), polyGA (strings of GA couplets), and nonrepetitive spacers (Figure 1B).[11,16−18] Different spidroins, which make up distinct silk types, are composed of varying combinations of these amino acid motifs.[9,13,19] These differences in motif composition are thought to contribute to the mechanical properties of various silk fibers, which differ considerably in extensibility and strength.[20,21] For example, polyA and polyGA motifs are hypothesized to form β-sheets that confer fiber strength, whereas polyGGX may adopt 3_1 helical conformations, and polyGPGXX motifs are postulated to form β-turns that contribute to fiber extensibility.[20,22,23]

Spidroins have largely been characterized from species in the Araneoidea (ecribellate orb-weavers and their descendants), a derived lineage within the Araneomorphae (true spiders; Figure 2).[14,15,24] Recent studies of the distantly related Mygalomorphae (the sister group to Araneomorphae) and Mesothelae

Received: September 21, 2014
Revised: October 22, 2014
Published: October 23, 2014

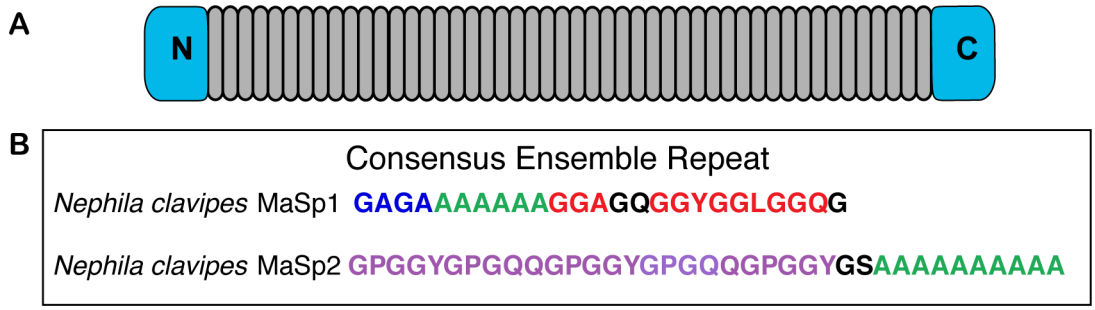


Figure 1. Spider silk proteins (spidroins) are highly repetitive structural polymers. (A) Schematic of spidroin protein showing iterations of tandem repeat sequences and nonrepetitive N- and C-terminal domains. (B) Consensus ensemble repeats for dragline silk spidroins MaSp1[16] and MaSp2[18] from the orb-weaver *Nephila clavipes*; amino acid motifs color-coded as poly alanine (A_n) in green, polyglycine-alanine [$(GA)_n$] in blue, [$(GGX)_n$] in red, and $(GPGXX)_n$ in purple. Amino acid motifs are hypothesized to form structural modules, for example, A_n and $(GA)_n$ may form β-sheets, polyGGX may form 3_1 helices and $(GPGXX)_n$ may form β-turns.[20,22,23]

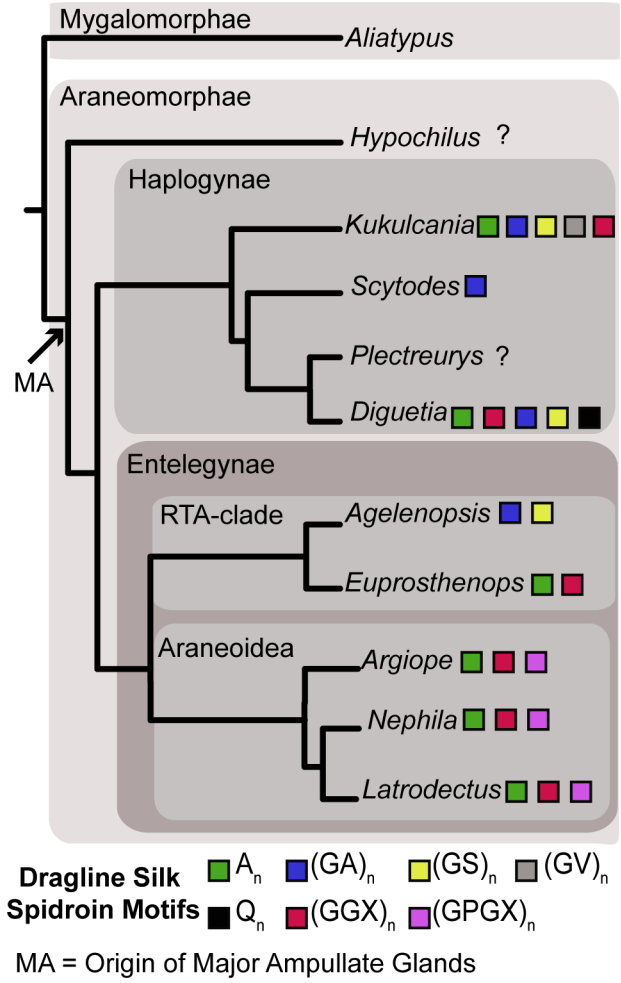


Figure 2. Dragline silk proteins exhibit sequence diversity across spider phylogeny. Phylogenetic relationships of lineages discussed in this study showing placement of *Scytodes* and indicating amino acid motifs found in dragline silk proteins of species where sequences have been cloned from major ampullate glands or sequences identified with characteristic MaSp C-terminal domains.[13−17,25,26,28,29] Question marks indicate species with major ampullate glands from which multiple spidroins were cloned from total silk tissues and where it is unknown which sequences encode major ampullate proteins.

(sister group to Araneomorphae and Mygalomorphae) indicate species in these two clades have fewer numbers of spidroins that are also more uniform in amino acid composition and repeat structure in comparison to araneoid spidroins.[14,25,26] This parallels the morphological uniformity of mygalomorph and mesothele silk glands, which most closely resemble araneoid piriform or aciniform glands.[7] Moreover, mygalomorphs and mesotheles lack major ampullate glands and do not spin draglines; draglines are only produced by araneomorphs because major ampullate glands are unique to this lineage (Figure 2).[7,27] However, the evolution of dragline silk proteins is poorly understood because few spidroins have been described from nonaraneoid araneomorphs, and are particularly lacking from members of the Haplogynae (Figure 2).[14,15,26,28,29]

To further characterize the molecular, mechanical, and functional diversity of spider silks and particularly of dragline fibers from nonaraneoid araneomorph spiders, this study focused on silk from the haplogyne species *Scytodes thoracica* (Scytodidae; Figure 2). Spiders in this family are known as spitting spiders because they capture their prey from a distance by ejecting an adhesive silk-like venom from their fangs onto prey.[30−32] Similar to other spiders, *Scytodes* also produce silk fibers from abdominal spinnerets. A previous study on the mechanical properties of *Scytodes* dragline silk showed it has high toughness and exceptional extensibility[33] that are second only to the toughest known biological material, dragline silk from *Caerostris darwini* (Darwin's bark spider, another araneoid species).[2] Here we used a combination of *S. thoracica* silk gland gene expression data, amino acid composition analyses, and dragline silk mechanical testing to better understand the functional diversity of dragline silk.

MATERIALS AND METHODS

Silk Gland cDNA Library Construction. Adult female *S. thoracica* were obtained from Oxford, Mississippi, and were individually caged at room temperature, sprayed with water, and fed crickets. Each spider was anesthetized with CO_2 and then total silk gland tissue was dissected, frozen in liquid nitrogen, and stored at −80 °C. Total RNA was extracted from silk gland tissues by homogenization in TRIzol (Invitrogen), followed by purification using an RNeasy Mini Kit (Qiagen). mRNA was isolated from total RNA with Oligo- $(dT)_{25}$-tagged magnetic beads (Invitrogen). cDNA was synthesized with the SuperScript III choice protocol (Invitrogen) using an oligo-anchored $(dT)_{18}$ V (V = A, C, or G) primer. cDNA fragments were size-selected (≥1000 base pairs) using ChromaSpin 1000 columns (Clontech). Size-selected cDNA fragments were ligated into EcoRV-digested pZErO-2 plasmids and transformed into TOP10 *Escherichia coli* cells (Invitrogen) by electroporation. A total of 12 microplates (96-well) were inoculated with positive colonies, and the library of 1152 cDNA clones was stored at −80 °C. The silk gland cDNA library was

dx.doi.org/10.1021/bm501409n | *Biomacromolecules* 2014, 15, 4598−4605

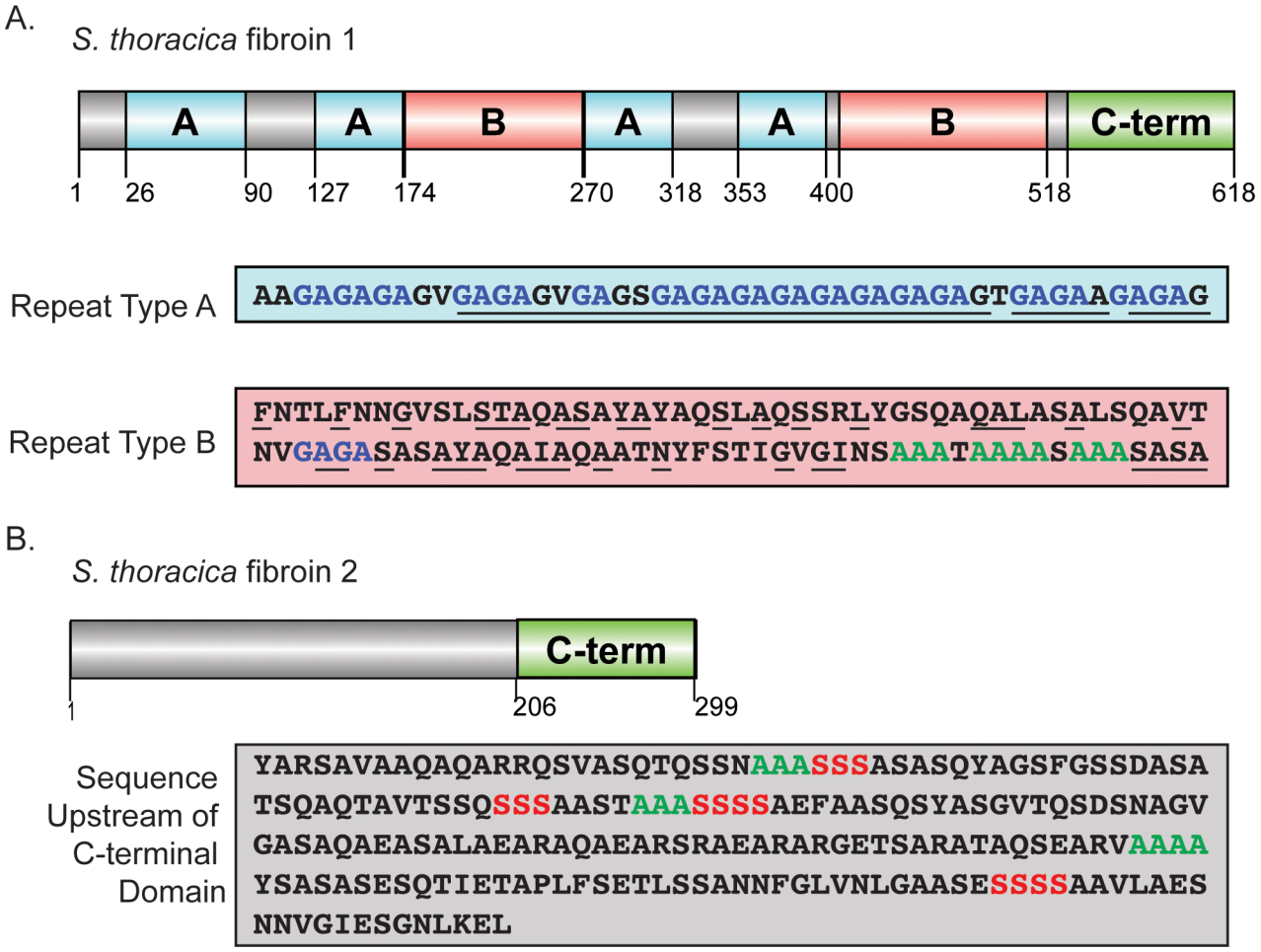


Figure 3. Sequence characteristics of the two spidroins obtained from *S. thoracica*. (A) Top: schematic of *S. thoracica* fibroin 1, showing arrangement of repeat types A and B relative to the C-terminal region. Boxes below show amino acid sequence of a representative from repeats A and B. Underlined amino acids occur in at least 50% of aligned repeats of each type. PolyGA motifs are indicated in blue text, poly alanine motifs (three or more consecutive alanines) in green text. Sequence outside of repeat regions (identified by RADAR analyses) shown in gray boxes, the majority of which (positions 91−126 and 319−352) is composed of GA repeats. (B) Top: schematic of *S. thoracica* fibroin 2, with box below showing full sequence upstream of C-terminal domain with polyS (serine) motifs in red.

screened for insert size and clones with inserts of at least 700 bp were sequenced with T7 or Sp6 universal primers.

Sequence Analysis. Sequences were edited using SEQUENCHER 4.8 (Gene Codes) and submitted to BLASTX searches against the NCBI protein (nr) database to infer protein translations in each cDNA and identify putative spidroins. Sequences were analyzed with BLASTclust to identify groups of similar sequences, and all spidroin cDNAs were sized by restriction digests. One long clone was sequenced completely using the transposon-based EZ−Tn5⟨TET-1⟩ Insertion Kit (Epicentre). Transposon insertion clones were sequenced and assembled into a complete contig based on restriction digest maps. To determine the amino acid sequence repeats in newly characterized spidroins, inferred translated sequences were analyzed with the RADAR program (http://www.ebi.ac.uk/Tools/pfa/radar/).[34] The University of Virginia FASTA server (http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1) was used to predict protein secondary structure using the Chou-Fasman[35] and Garnier[36] methods and protein hydrophilicity using Kyte-Doolittle[37] predictions.

Amino Acid Composition Analyses. Extractions of proteins from total silk gland tissue was done by grinding the tissue in ddH_2O followed by centrifuging the material and rinsing the pellet twice with acetone. Silk gland extractions and dragline fibers were sent to the UC Davis Molecular Structure Facility, Davis, CA, and hydrolyzed with 6 N HCl for 24 h at 110 °C. An L-8800 Hitachi analyzer was used for ion-exchange chromatography to separate amino acids followed by a "postcolumn" ninhydrin reaction detection system. Results from these composition analyses were compared to the amino acid compositions of identified silk *S. thoracica* cDNAs and published full-length araneoid MaSps and MiSps predicted using ProtParam (http://web.expasy.org/protparam).

Silk Collection and Tensile Testing. Dragline silk was collected by allowing spiders to walk on a plastic container containing black cardboard. Dragline was then removed from the cardboard and mounted onto C-shaped cards with a 1 cm gaps to which they were secured with cyanoacrylate.[38] Given the collection method, mostly double-stranded fibers were recovered because spiders use both of their major ampullate spigots when laying down draglines.[5] Fibers were then visually inspected with a stereomicroscope to establish the number of fibers across the gap. Cards containing fibers that were too loose due to lack of tension or double-stranded fibers were discarded. The diameter of each fiber was then measured using polarized light microscopy as outlined in Blackledge et al.[38] Tensile testing was accomplished using a Nano Bionix tensile tester (Agilent Technologies, Oakridge, TN, U.S.A.), as described in Collin et al.[39] and Swanson et al.[33] In short, cards containing fibers were attached to the clamps of the Nano Bionix tensile tester. Connecting cardboard material was removed so that the tester only pulled on the silk sample between the clamps. All fibers were extended at a constant rate of 1% strain/sec until failure. The testing environment ranged from 23 to 26

dx.doi.org/10.1021/bm501409n | *Biomacromolecules* 2014, 15, 4598−4605

°C and a relative humidity of 28 to 42%. Four different properties were measured that assess fiber performance. The first was the ultimate tensile strength, defined as the breaking force (in Newtons, N) divided by the cross sectional area. The second was extensibility (true breaking strain), defined as the natural log of the breaking length divided by original length. The third was Young's modulus, or stiffness of the fiber, or the amount of stress required to strain the sample a given amount.[38] The fourth property measured was toughness, which is the energy absorbed per volume before the fiber breaks.[33,39]

Phylogenetic Analyses. We estimated the phylogenetic relationships of newly identified *S. thoracica* silk spidroins relative to previously published spidroins using parsimony and Bayesian methods. Trees were based on the nonrepetitive C-terminal domains (~100 amino acids long) that are conserved across spidroins and are not as problematic to align as the repetitive sequences.[14] Inferred translations of spidroin cDNA C-termini were combined with selected published spidroin C-termini sampled to represent a diversity of spidroin type and included all available basal araneomorph spidroins. The data set included *Argiope trifasciata* AcSp1 (AY426339) and Flag (AF350264); *Nephila clavipes* Flag (AF027973), PySp1 (GQ980330), MaSp1a (AY654292), MaSp2 (AY654297), and MiSp1 (AF027735); *Nephila clavata* TuSp1 (AB218973); *Latrodectus hesperus* MaSp1 (AY953074), MaSp2 (AY953075), MiSp (HM752571), PySp1 (ACV41934), AcSp1 (EU025854), and TuSp1 (AY953070); *Agelenopsis aperta* MaSp (HM752573) and TuSp1 (HM752576); *Euprosthenops australis* MaSp (AJ973155); *Plectreurys tristis* fibroins 1−4 (AF350281−AF350284); *Diguetia canities* MaSp (HM752565) and MaSp-like (HM752566); *Hypochilus thorelli* fibroin 1 (JX102555) and fibroin 2 (JX102556); and *Aliatypus gulosus* fibroin 1 (EU117159).

The selected sequences were aligned using the program MUSCLE.[40] Trees were generated with parsimony analysis using heuristic searches in PAUP*4.0b10[41] with 1000 random taxon addition replicates (RTA), treating alignment gaps as missing data, and nodal support was calculated by 1000 bootstrap (BS) replicates. The tree was rooted with the single included mygalomorph C-terminus. Bayesian analyses were completed with MrBayes 3.2[42] using a "mixed" amino acid substitution model and gamma distribution, allowing for among site rate variation. This analysis included 5×10^6 generations with four MCMC chains, sampling a tree every 1000 generations. The first 25% of sampled trees were discarded as burnin and the remaining trees were summarized in a 50% majority-rule consensus to determine clade posterior probability (CPP) support values.

RESULTS AND DISCUSSION

Identified *Scytodes* Spidroins. We characterized two distinct spidroin sequences from the silk gland cDNA library constructed from the spitting spider *S. thoracica*. BLASTx results revealed that sequenced cDNAs included 24 spidroin transcripts (~9% of the sequenced clones), which clustered into two groups. Inferred translation of these sequences included conserved C-terminal domains that aligned to published spidroin C-termini. The longest cDNA clone in the first spidroin cluster (*S. thoracica* fibroin 1) was 1948 bp in length (encoding 618 amino acids) and the longest cDNA clone in the second spidroin cluster (*S. thoracica* fibroin 2) was 1017 bp in length (encoding 299 amino acids). *S. thoracica* fibroin 1 was sequenced with transposon mapping to span the repetitive region; the assembled sequences were deposited at NCBI (accession numbers KM987233 and KM987234). BLASTx searches of the C-terminus of *S. thoracica* fibroin 1 showed the top hit to be major ampullate spidroin 2 (MaSp2) from the spider *Argiope amoena* (accession number AAR13813; *E*-value 4e-10; 41% similarity). The repetitive region of *S. thoracica* fibroin 1 had a top BLAST hit to the partial MiSp-like sequence from *Parawixia bistriata* (ADG57595; *E*-value 2e-105; 46% similarity). The C-terminus of *S. thoracica* fibroin 2 had a top BLAST hit to MaSp2 from *Latrodectus hesperus* (AAY28936; *E*-value 2e-04; 36% similarity) and the sequence upstream of its C-terminus had a top BLAST hit to the PySp1 (piriform spidroin 1) from *Latrodectus hesperus* (ACV41934; *E*-value 6e-19; 50% similarity).

Analysis of *S. thoracica* fibroin 1 using RADAR found two distinct repeat types (A and B) arranged in an alternating configuration of AABAAB upstream of the C-terminus (Figure 3A). The first repeat type (type A) in *S. thoracica* fibroin 1 is largely composed of alternating glycine and alanine (GA) motifs and includes one unit of 65 amino acids, and three more highly similar units that are 48 amino acids. The two B repeat types of *S. thoracica* fibroin 1 are 96 and 112 amino acids long and have almost no GA motifs (Figure 3A). Additional sequence between both sets of adjacent A type repeats (Figure 3A, positions 91−126 and 319−352) were largely composed of GA couplets, but were not identified by RADAR as part of either repeat type. *S. thoracica* fibroin 1 contains a small number of polyA motifs but no GGX or GPGXX motifs. No repeats were found in *S. thoracica* fibroin 2 using RADAR, likely because this relatively short sequence (299 amino acids including ~100 aa C-terminal domain) did not contain two entire repeats (a repeat can exceed 100 amino acids in length[26]). The sequence upstream of *S. thoracica* fibroin 2's C-terminal region is glycine-poor but contains serine (S) repeated in tandem (Figure 3B), which is not characteristic of araneoid MaSps, but are found in araneomorph TuSp1[43−45] and mygalomorph spidroins.[25,26] The average of Garnier and Chou-Fasman analyses predicted a secondary structure for *S. thoracica* fibroin 1 with greater formation of alpha helices (48%) than β-sheets (37%). Similarly, *S. thoracica* fibroin 2 was predicted to form more alpha helices (58%) than β-sheets (20%). Kyte−Doolittle predictions showed both sequences having alternating regions of hydrophobicity and hydrophilicity (Supporting Information, Figure S1).

Dragline Fiber and Silk Gland Amino Acid Composition Analyses. The most abundant amino acids in *S. thoracica* dragline fiber are alanine (35.4% on average) and glycine (22.3%) followed by serine (10.7%; Supporting Information, Table S1), and these values paralleled the amino acid composition of combined silk gland tissue from *S. thoracica* (alanine = 31.1%; glycine = 22.5%; serine = 9.7% on average). Moreover, the composition of the three most abundant amino acids predicted from the *S. thoracica* fibroin 1 sequence (alanine = 32.4%; glycine = 20.7%; serine = 14.2%), along with all other residues, closely match the composition of *Scytodes* dragline (Supporting Information, Table S1). In contrast to *S. thoracica* fibroin 1, fibroin 2 has a much lower amount of glycine (4.7%), a higher percentage of serine (~24%), and glutamine/glutamic acid (~12%), and does not closely match the composition of the dragline or total silk gland tissue. These results suggest that *S. thoracica* fibroin 1 represents the major protein in *S. thoracica* dragline silk and are also consistent with the higher abundance of fibroin 1 in total silk gland tissue relative to *S. thoracica* fibroin 2 (20 vs 4 cDNA clones, respectively).

Though *S. thoracica* fibroin 1 appears to be the major protein constituting *Scytodes thoracica* dragline silk, the amino acid composition and repetitive architecture of *S. thoracica* fibroin 1 is markedly different from previously described araneoid MaSps (Supporting Information, Table S1).[11,16,17,29,46] Specifically, MaSp1 repeats are dominated by GGX, polyA, and fewer polyGA motifs, whereas MaSp2 has abundant GPGXX and polyA motifs, and many fewer GGX motifs. However, only

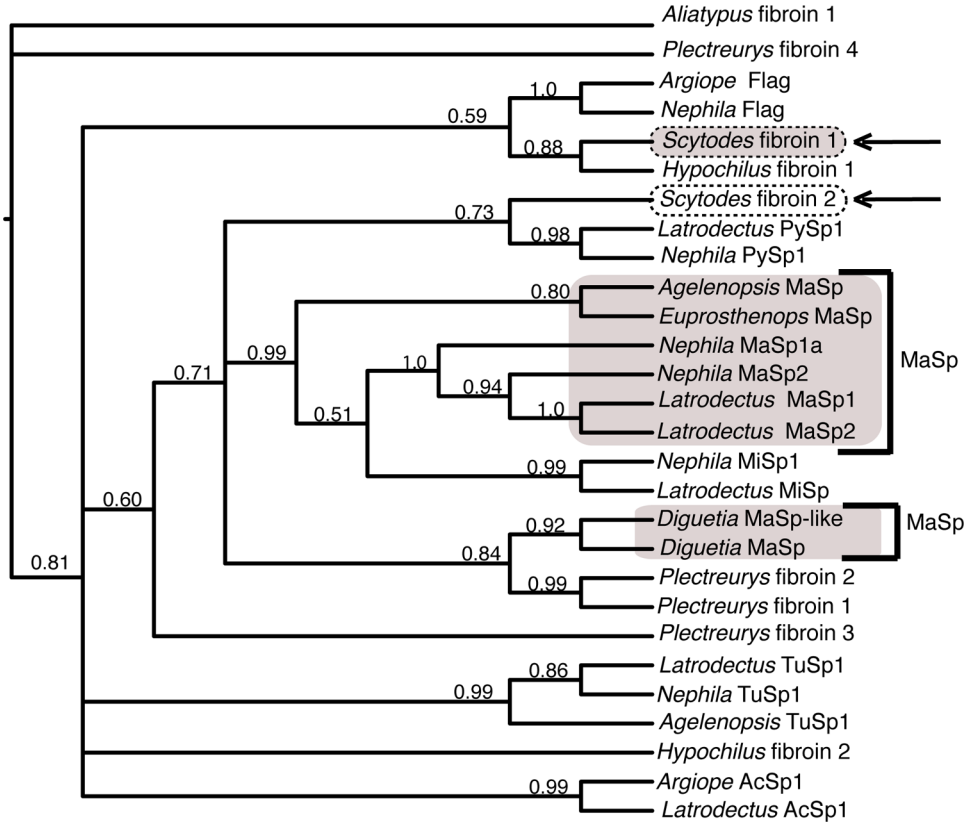


Figure 4. Bayesian phylogenetic tree of spidroin C-terminal domain sequences. The 50% majority-rule consensus after removal of 25% burnin trees, the number above branches, indicates clade posterior probability support values. *Scytodes* fibroins 1 and 2 are indicated by arrows and circled with dashed lines, MaSp clades are indicated by brackets, and dragline silk spidroins are indicated by gray shading.

polyGA motifs were abundant in *S. thoracica* fibroin 1, which lacked GGX and GPGXX motifs and had a few polyA motifs. The amino acid composition of *Scytodes* dragline fiber is dominated by alanine and glycine, consistent with the abundant GA motifs of *S. thoracica* fibroin 1, but contains less than 2% proline, unlike the proline content of araneoid dragline fibers, which can range between 3 and 16%.[47,48] Moreover, *S. thoracica* fibroin 1 and *Scytodes* dragline contain similar amounts of other residues (e.g., valine, leucine, threonine), which are nearly absent from well-characterized araneoid MaSps (Supporting Information, Table S1).

Our discovery of only two spidroins from the *Scytodes* silk gland library is consistent with the relatively simple morphology of *Scytodes* silk glands. Specifically, Platnick et al.[49] described *Scytodes* spigots as including major ampullate spigots and piriform spigots on the anterior lateral spinnerets but with reduced posterior lateral and posterior median spinnerets that may only have aciniform spigots. Our observations of *S. thoracica*'s silk glands indicate they include one pair of long sac-like glands (presumed to be homologous to major ampullate glands) and numerous tiny globular glands close to the spinnerets (likely including piriform and aciniform glands; see Supporting Information, Figure S2). Thus, while *S. thoracica* fibroin 1 appears to be a component of *Scytodes* dragline silk, the distinctly different amino acid composition of *S. thoracica* fibroin 2 in comparison to dragline fibers suggests fibroin 2 is a component of aciniform or piriform silk. *Scytodes* use silk for a limited number of functions, as they do not build foraging webs and are largely wandering hunters.[31] However, scytodids likely use piriform silk to secure their dragline silk to a substrate, whereas aciniform silk may play a role in reproduction, as it does in other species.[27]

Diversity of Spider Dragline Silk Mechanical Properties. A 2006 study by Swanson et al.[33] found that *Scytodes* sp. dragline silk was the toughest of the silks tested in their study, ranking it as number one among known silks in toughness. In 2010, *Scytodes* dragline moved to second place with the discovery of *C. darwini* as the toughest known silk.[2] From the 18 *S. thoracica* silk samples we tested, the generated stress–strain curves showed an initial stiff modulus of 6–10 GPa until a yield point. After the yield point, the silk fiber stretches in a linear and continuous manner until failure (Supporting Information, Figure S3). Tensile testing of *S. thoracica* dragline silk also showed high toughness (144.61 ± 41.70 MJ/m^3) and exceptional extensibility (i.e., 36% extension; Supporting Information, Table S2). However, it was not as tough as reported by Swanson et al. (230.02 ± 84.53 MJ/m^3).[33] Nevertheless, *S. thoracica* dragline silk was still substantially tougher than silkworm (*Bombyx mori*) silk, Kevlar, and steel[50,51] and was similar to toughness values from dragline silk of other species of Haplogynae (*Plectreurys* and *Kukulcania*), as well as araneoid species such as *Latrodectus*, *Nephila*, and *Araneus* (Table S2).[2,33,52,53]

The difference between toughness values for *Scytodes* dragline obtained in this study and Swanson et al.[33] could be due to variation in species, spider condition, and silk collection methods, which can change the mechanical properties of silks.[54,55] In this study, *Scytodes thoracica* dragline silk was collected by allowing spiders to walk on cardboard and collected silk fibers were subsequently transferred onto cards,

whereas Swanson et al.[33] collected *Syctodes* sp. dragline by forcible silking. When forcibly silked, spiders can alter the tension of fibers as they are drawn and this can substantially increase the silk's tensile strength.[56] Forcible silking of dragline may favor greater formation of β-sheet crystals and silk protein alignment, leading to greater fiber strength and stiffness,[56] and this is consistent with the greater strength and stiffness of *Scytodes* silk measured by Swanson et al.[33]

Recently, Blackledge et al.[55] compared the mechanical properties of dragline silks from a broad range of species (including *Scytodes* sp.) using supercontracted silk to minimize variation in silk spinning effects. Supercontraction involves the hydration of collected silk to break hydrogen bonds aligning silk proteins, and allows for the rearrangement of noncrystalline regions in the absence of forces applied by the spider. Though results from Blackledge et al.[55] are not directly comparable to our mechanical data, they showed that *Scytodes* draglines (and other haplogyne draglines) have lower tensile strength and are less tough than those produced by species in the RTA clade, and are substantially less tough than araneoid orb-weaver dragline silks. Our characterization of the dragline silk protein *S. thoracica* fibroin 1 shows a lack of GGX and GPGXX motifs that are abundant in orb-weaver dragline silk proteins (MaSp1 and MaSp2, respectively) and are hypothesized to confer fiber extensibility and toughness,[20] which may explain the greater toughness of araneoid dragline fibers relative to those from *Scytodes* and other Haplogynae, as found by Blackledge et al.[55] Unlike araneoid MaSps, *S. thoracica* fibroin 1 is dominated by GA repeats, which are hypothesized to form β-sheets[10,20,57] (β-sheets are predicted in ~37% of fibroin 1). The abundant GA motifs of *S. thoracica* fibroin 1 are also found in high frequency in araneoid minor ampullate silk spidroins (MiSps) and in the *Bombyx* fibroin heavy chain protein that largely contains repeating $GSGAGA_n$ motifs.[10] Guinea et al.[58] recently compared the mechanical properties of araneoid minor ampullate and dragline silk (under forcibly silked and supercontracted states) and found that unlike dragline silk, minor ampullate silk is similar to *Bombyx* silk in not exhibiting a supercontraction effect (e.g., greater extensibility). However, Guinea et al.[58] also found that araneoid minor ampullate silk is more similar to araneoid dragline silk than either is to *Bombyx* silk in terms of β-sheet parameters and percent crystallinity. Similarly, future structural studies of *Scytodes* dragline silk may be helpful in determining the role of β-sheet structure and abundance in fiber mechanical performance.

Complex Evolution of Dragline Silk. Phylogenetic analyses of our *S. thoracica* and published spidroins were done using an amino acid alignment of the nonrepetitive C-terminal domain that had 114 characters. The Bayesian consensus tree placed *S. thoracica* fibroin 1 and *Hypochilus* fibroin 1 from a basal araneomorph spider together (CPP = 0.88), and this group was sister to the Flag (flagelliform) spidroins, albeit with poor support (CPP = 0.59; Figure 4). *S. thoracica* fibroin 2 was placed as sister to PySp (piriform) spidroins with moderate support (CPP = 0.73), consistent with the possibility of *S. thoracica* fibroin 2 representing a piriform spidroin. Monophyly of flagelliform (Flag), piriform (PySp), aciniform (AcSp1), tubuliform (TuSp), and minor ampullate (MiSp) spidroins were recovered with significant clade support (CPP $\geq$ 0.99). However, the MiSp clade was contained within a well-supported clade (CPP = 0.99) of major ampullate spidroins (MaSps) from araneoid (*Nephila* and *Latrodectus*) and RTA-clade spiders (*Agelenopsis* and *Euprosthenops*). Moreover, the MaSp and MaSp-like spidroins from *Diguetia canities*, a member of the Haplogynae, appeared more closely related to uncharacterized spidroins from *Plectreurys tristis*, also from the Haplogynae clade.

The parsimony and Bayesian consensus trees disagreed at several nodes (Figure 4, Supporting Information, Figure S4). Heuristic searches with gaps treated as missing data retrieved one single most parsimonious tree (Figure S4). Parsimony also grouped *S. thoracica* fibroin 1 with the spidroin *Hypochilus* fibroin 1, but this relationship was not well supported by bootstrap analysis. This group was shown as most closely related to MaSps and MiSps, rather than to Flag as in the Bayesian tree. Instead, *S. thoracica* fibroin 2 was positioned as sister to the Flag sequences, but this relationship was also not supported with high bootstrap values.

In general, many previous phylogenetic analyses of spidroins show paralogs grouping into clades according to their glandular origin.[6,13,15] This pattern is largely repeated in our analyses; for example, TuSp1 from *Agelenopsis* (RTA clade) grouped with TuSp1s from *Latrodectus* and *Nephila* (Araneoidea). However, Haplogynae spidroins *Diguetia* MaSp and MaSp-like did not group with the MaSps from araneoid (*Nephila* and *Latrodectus*) and RTA-clade spiders (*Agelenopsis* and *Euprosthenops*). Moreover, *S. thoracica* fibroin 1 did not group with either the araneoid and RTA MaSps, or with *Diguetia* MaSp, suggesting the possible independent evolution of the proteins composing spider dragline silk. This scenario was proposed by Garb et al.[15] and may be attributed to the rapid evolution of the spidroin gene family, involving numerous gene duplications, extinctions, and replacement of the dominantly expressed spidroin with less closely related paralogs. Previous studies have shown that spidroins from basal mygalomorphs (*Aphonopelma* and *Sphodros*) form a clade sister to spidroins from the haplogyne spiders *Plectreurys* and *Diguetia*.[26] These results, in which gene trees do not closely mirror species trees, can similarly be attributed to frequent gene duplication events and losses and changes in expression.

Though *Hypochilus* fibroin 1 and *Scytodes* fibroin 1 appear most closely related to each other based on their C-termini (Figure 4), their repetitive regions differ significantly. The repetitive region of *Hypochilus* fibroin 1 contains polyalanine (A_n), polyglycine (G_n), and polyserine (S_n),[26] whereas *S. thoracica* fibroin 1 largely has alternating glycine and alanine couplets [$(GA)_n$], although it is unknown whether *Hypochilus* fibroin 1 is a constituent of *Hypochilus* dragline silk.

The extensive polyA motifs in *Hypochilus* fibroin 1, are also abundant in phylogenetically distant MaSp silk proteins from both RTA and araneoid species, and other haplogyne species (e.g., *Diguetia*[15]), but are largely absent from *Scytodes* fibroin 1. It is estimated that araneomorphs originated at least ~240 million years ago,[59] and hence, the polyA amino acid motif could have been retained in spidroins since that time and was lost in some lineages or evolved repeatedly in different spidroins synthesized by different lineages.

The few previous studies of Haplogynae spidroins also suggest a greater diversity of spider dragline silk proteins.[14,15,28] For example, of four spidroins from *Plectreurys tristis*, some have (A_n) $(GA)_n$ and $(QA)_n$ motifs, but no GPGXX motifs and very few GGX motifs. Yet is also unknown which of the four spidroins are used in *Plectreurys* dragline.[14] Spidroins isolated from ampullate-shaped glands of the haplogyne *Diguetia canities* (termed MaSp and MaSp-like) contain polyA and GGX motifs similar to those in araneoid MaSps, as well as uncommon

polyQ motifs, but no GPGXX motifs (Figure 2).[15] Similarly, spidroins expressed in the major ampullate silk glands of the haplogyne species *Kukulcania hibernalis* contain polyA, GGX, and polyGA motifs, but they also contain GS and GV couplets that are not seen in araneoid MaSps, and no GPGXX motifs (Figure 2).[28] *Scytodes* fibroin 1 adds to the few examples of Haplogynae dragline spidroins in clearly demonstrating that the composition of spider dragline silk is highly variable at the molecular level and has significantly changed over time. Accordingly, our results show that species from basal araneomorph clades express a diversity of dragline silk proteins, with varying assemblages of repetitive sequence motifs that are distinctly different from araneoid MaSp sequences. More sampling of dragline silk cDNAs from poorly studied araneomorph lineages is clearly needed to further trace the dynamics of spidroin amino acid motif evolution and its impact on generating biomechanically diverse fibers.

CONCLUSIONS

We were able to identify two distinct spidroins from the spitting spider *S. thoracica*, one of which (*Scytodes* fibroin 1) appears to be the principal component of *Scytodes* dragline silk. The molecular organization of this spidroin is different from derived dragline silk proteins (MaSp1 and MaSp2), in lacking several of the structural motifs that correlate with the superior mechanical properties of araneoid dragline silk. Our phylogenetic analyses also suggest that major ampullate spidroins may not have a single evolutionary origin. Rather, they have a complex evolutionary history involving the repeated evolution of dragline silk proteins, associated with the gain and loss of key structural motifs. Our findings have significance for bioengineering applications since spider silks with different molecular compositions conferring different mechanical properties could be utilized for innovative biomimetic designs.

ASSOCIATED CONTENT

Supporting Information

Kyte and Doolittle hydropathy plots; Amino acid composition of *S. thoracica* silk glands, fibers, and fibroins; *S. thoracica* silk gland morphology; *S. thoracica* dragline silk stress versus strain curve; Tensile properties of *S. thoracica* dragline silk; Parsimony phylogeny of spidroin C-terminal domain sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: scorr006@ucr.edu.

*E-mail: jessica_garb@uml.edu.

Present Address

[†]Department of Biology, University of California Riverside, Riverside California (S.M.C.-G.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Bob Suter, Gail Stratton, Kanaka Varun Bhere, Rujuta Gadgil, Matthew Collin, Cheryl Hayashi, and Crystal Chaw for assistance with this study. Tensile tests were conducted at the University of California, Riverside, with support from the Army Research Office to C. Hayashi. This work was funded by Grants-in-Aid of Research Programs by Sigma Xi and the Society for Integrative and Comparative Biology (SICB) to S.M.C.-G. and a UML Seed Grant for Advancing Research Scholarship and Creative Work to J.E.G.

REFERENCES

(1) Gosline, J. M.; DeMont, M. E.; Denny, M. W. *Endeavour* **1986**, *10*, 37−43.

(2) Agnarsson, I.; Kuntner, M.; Blackledge, T. A. *PloS One* **2010**, *5*, e11234.

(3) Hu, X.; Vasanthavada, K.; Kohler, K.; McNary, S.; Moore, A. M. F.; Vierra, C. A. *Cell. Mol. Life Sci.* **2006**, *63*, 1986−1999.

(4) Kluge, J. A.; Rabotyagova, O.; Leisk, G. G.; Kaplan, D. L. *Trends Biotechnol.* **2008**, *26*, 244−251.

(5) Coddington, J. A. *J. Arachnol.* **1989**, *17*, 71−95.

(6) Hayashi, C. Y. In *Molecular Systematics and Evolution: Theory and Practice*; DeSalle, R., Giribet, G., Wheeler, W., Eds.; Birkauser: Berlin, 2002; pp 209−224.

(7) Palmer, J. M.; Coyle, F. A.; Harrison, F. W. *J. Morphol.* **1982**, *174*, 269−274.

(8) Guerette, P. A.; Ginzinger, D. G.; Weber, B. H. F.; Gosline, J. M. *Science* **1996**, *272*, 112−115.

(9) Hayashi, C. Y.; Blackledge, T. A.; Lewis, R. V. *Mol. Biol. Evol.* **2004**, *21*, 1950−1959.

(10) Chen, G.; Liu, X.; Zhang, Y.; Lin, S.; Yang, Z.; Johansson, J.; Rising, A.; Meng, Q. *PLoS One* **2012**, *7*, e52293.

(11) Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y. *PloS One* **2007**, *2*, e514.

(12) Hayashi, C. Y.; Lewis, R. V. *Science* **2000**, *287*, 1477−1479.

(13) Ayoub, N. A.; Garb, J. E.; Kuelbs, A.; Hayashi, C. Y. *Mol. Biol. Evol.* **2013**, *30*, 589−601.

(14) Gatesy, J.; Hayashi, C.; Motriuk, D.; Woods, J.; Lewis, R. *Science* **2001**, *291*, 2603−2605.

(15) Garb, J. E.; Ayoub, N. A.; Hayashi, C. Y. *BMC Evol. Biol.* **2010**, *10*, 243.

(16) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120−7124.

(17) Zhang, Y.; Zhao, A.-C.; Sima, Y.-H.; Lu, C.; Xiang, Z.-H.; Nakagaki, M. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **2013**, *164*, 151−158.

(18) Hinman, M. B.; Lewis, R. V. *J. Biol. Chem.* **1992**, *267*, 19320−19324.

(19) Hayashi, C. Y.; Lewis, R. V. *J. Mol. Biol.* **1998**, *275*, 773−784.

(20) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *Int. J. Biol. Macromol.* **1999**, *24*, 271−275.

(21) Blackledge, T. A.; Hayashi, C. Y. *J. Exp. Biol.* **2006**, *209*, 2452−2461.

(22) van Beek, J. D.; Hess, S.; Vollrath, F.; Meier, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10266−10271.

(23) Becker, N.; Oroudjev, E.; Mutz, S.; Cleveland, J. P.; Hansma, P. K.; Hayashi, C. Y.; Makarov, D. E.; Hansma, H. G. *Nat. Mater.* **2003**, *2*, 278−283.

(24) Coddington, J. A.; Giribet, G.; Harvey, M. S.; Prendini, L.; Walter, D. E. In *Assembling the Tree of Life*; Cracraft, J., Donoghue, M., Eds.; Oxford University Press: U.K., 2004; pp 296−318.

(25) Garb, J. E.; DiMauro, T.; Lewis, R. V.; Hayashi, C. Y. *Mol. Biol. Evol.* **2007**, *24*, 2454−2464.

(26) Starrett, J.; Garb, J. E.; Kuelbs, A.; Azubuike, U. O.; Hayashi, C. Y. *PloS One* **2012**, *7*, e38084.

(27) Shultz, J. W. *Biol. Rev.* **1987**, *62*, 89−113.

(28) Tian, M.; Liu, C.; Lewis, R. *Biomacromolecules* **2004**, *5*, 657−660.

(29) Rising, A.; Johansson, J.; Larson, G.; Bongcam-Rudloff, E.; Engström, W.; Hjälm, G. *Insect Mol. Biol.* **2007**, *16*, 551−561.

(30) Clements, R.; Li, D. *Ethology* **2005**, *111*, 311−321.

(31) Suter, R. B.; Stratton, G. E. *J. Insect Sci.* **2009**, *9*, 62.

(32) Zobel-Thropp, P. A.; Correa, S. M.; Garb, J. E.; Binford, G. J. *J. Proteome Res.* **2014**, *13*, 817−835.

(33) Swanson, B. O.; Blackledge, T. A.; Summers, A. P.; Hayashi, C. Y. *Evolution* **2006**, *60*, 2539−2551.
(34) Heger, A.; Holm, L. *Proteins Struct. Funct. Bioinform.* **2000**, *41*, 224−237.
(35) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211−222.
(36) Garnier, J.; Osguthorpe, D. J.; Robson, B. *J. Mol. Biol.* **1978**, *120*, 97−120.
(37) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105−132.
(38) Blackledge, T. A.; Cardullo, R. A.; Hayashi, C. Y. *Invertebr. Biol.* **2005**, *124*, 165−173.
(39) Collin, M. A.; Garb, J. E.; Edgerly, J. S.; Hayashi, C. Y. *Insect Biochem. Mol. Biol.* **2009**, *39*, 75−82.
(40) Edgar, R. C. *Nucleic Acids Res.* **2004**, *32*, 1792−1797.
(41) Swofford, D. L. *PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4; Sinauer Associates: Sunderland, MA, 2003.
(42) Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D. L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M. A.; Huelsenbeck, J. P. *Syst. Biol.* **2012**, *61*, 539−542.
(43) Tian, M.; Lewis, R. V. *Biochemistry* **2005**, *44*, 8006−8012.
(44) Garb, J. E.; Hayashi, C. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11379−11384.
(45) Hu, X.; Lawrence, B.; Kohler, K.; Falick, A. M.; Moore, A. M. F.; McMullen, E.; Jones, P. R.; Vierra, C. *Biochemistry* **2005**, *44*, 10020−10027.
(46) Garb, J. E.; Dimauro, T.; Vo, V.; Hayashi, C. Y. *Science* **2006**, *312*, 1762.
(47) Blackledge, T. A.; Kuntner, M.; Marhabaie, M.; Leeper, T. C.; Agnarsson, I. *Sci. Rep.* **2012**, *2*, 833.
(48) Marhabaie, M.; Leeper, T. C.; Blackledge, T. A. *Biomacromolecules* **2014**, *15*, 20−29.
(49) Platnick, N. I.; Coddington, J. A.; Forster, R. R.; Griswold, C. E. *Am. Mus. Novit.* **1991**, *3016*, 1−73.
(50) Porter, D.; Vollrath, F.; Shao, Z. *Eur. Phys. J. E* **2005**, *16*, 199−206.
(51) Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N. *J. Exp. Biol.* **1999**, *202*, 3295−3303.
(52) Swanson, B. O.; Blackledge, T. a.; Beltrán, J.; Hayashi, C. Y. *Appl. Phys. A: Mater. Sci. Process.* **2006**, *82*, 213−218.
(53) Sensenig, A.; Agnarsson, I.; Blackledge, T. A. *J. Evol. Biol.* **2010**, *23*, 1839−1856.
(54) Garrido, M. A.; Elices, M.; Viney, C.; Pérez-Rigueiro, J. *Polymer* **2002**, *43*, 1537−1540.
(55) Blackledge, T. A.; Pérez-Rigueiro, J.; Plaza, G. R.; Perea, B.; Navarro, A.; Guinea, G. V.; Elices, M. *Sci. Rep.* **2012**, *2*, 782.
(56) Ortlepp, C. S.; Gosline, J. M. *Biomacromolecules* **2004**, *5*, 727−731.
(57) Colgin, M. A.; Lewis, R. V. *Protein Sci.* **1998**, *7*, 667−672.
(58) Guinea, G. V.; Elices, M.; Plaza, G. R.; Perea, G. B.; Daza, R.; Riekel, C.; Agulló-Rueda, F.; Hayashi, C.; Zhao, Y.; Pérez-Rigueiro, J. *Biomacromolecules* **2012**, *13*, 2087−2098.
(59) Selden, P. A.; Shear, W. A.; Bonamo, P. M. *Paleontology* **1991**, *34*, 241−281.