N-Terminal Nonrepetitive Domain Common to Dragline, Flagelliform, and Cylindriform Spider Silk Proteins

Anna Rising,[†] Göran Hjälm,[‡,§] Wilhelm Engström,[†] and Jan Johansson*[,‡]

Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden, and Department of Molecular Biosciences, Veterinary Medical Biochemistry, Biomedical Centre, Swedish University of Agricultural Sciences, SE-751 23 Uppsala, Sweden

Received July 17, 2006; Revised Manuscript Received September 6, 2006

Spider silk has been extensively studied for its outstanding mechanical properties. Partial intermediate and C-terminal sequences of different spider silk proteins have been determined, and during the past decade also N-terminal domains have been characterized. However, only some of these N-terminal domains have been reported to contain signal peptides, leaving the mechanism whereby they enter the secretory pathway open to speculation. Here we present the sequence of a 394-residue N-terminal region of the *Euprosthenops australis* major ampullate spidroin 1 (MaSp1). A close comparison with published sequences from other species revealed the presence of N-terminal signal peptides followed by an approximately 130-residue nonrepetitive domain. From secondary structure predictions, helical wheel analysis, and circular dichroism spectroscopy this domain is concluded to contain five α-helices and is a conserved constituent of hitherto analyzed dragline, flagelliform, and cylindriform spider silk proteins.

Introduction

Spider dragline silk is produced in the major ampullate gland and used for construction of the framework of the web, as well as a lifeline. It is one of nature's best performing materials. Its mechanical properties are superior to manmade materials such as Kevlar and high tensile steel. In addition, some spiders can produce up to six other types of silk in specialized glands.[1] Such silks include flagelliform and cylindriform (tubuliform) silks that are produced in glands with corresponding names.[1,2] The flagelliform gland produces silk used in the capture spiral of the orb web and the cylindriform glands produce egg case silk.[1] Spider silk proteins, spidroins, are generally composed of sequence repeats, e.g., Ala- or Gly-rich blocks, flanked by conserved nonrepetitive C- and N-terminal regions. Major ampullate spidroin (MaSp) 1 and MaSp2 are the two main proteins of the dragline silk.[3,4] The spider stores these proteins as a dope in liquid crystalline form until they undergo phase transition into a solid fiber in the spinning duct.[5,6] The dope is produced by epithelial cells lining the interior of the major ampullate gland. These cells have an extensive endoplasmic reticulum (ER) and a large number of secretory vesicles, two morphological characteristics that are compatible with high rate synthesis of secretory proteins.[7,8] To enter the secretory pathway, a signal peptide that directs the protein to the ER is required. The structural features of signal peptides are conserved between different eukaryotic organisms. The desire to identify putative signal sequences has led to the development of software, e.g., SignalP,[9] that predicts whether a particular sequence is likely to function as a signal peptide or not. Signal peptides of secretory proteins are located in the extreme N-terminal region and consists of three parts; a hydrophilic n-region usually with a positive net charge (1−5 residues long), followed by a hydrophobic h-region (7−15 residues), and finally a more polar c-region containing the signal peptidase cleavage site (3−7 residues).[10]

Most published spider silk sequences lack the 5′-end of the messenger RNA. This is probably due to technical cloning circumstances that favor amplification of 3′-regions. Furthermore, large transcript sizes that exceed 10 kb, as judged by Northern analysis,[11−13] have made it difficult to obtain full length sequences. Despite these difficulties, two N-terminal sequences from flagelliform silk proteins and four from cylindriform silk proteins have been described.[14−18] Included in the cylindriform group is, except the cylindriform spidroin (CySp) 1 and CySp2 from *Argiope bruennichi* and CySp1 from *Nephila clavata*, also the tubuliform spidroin (TuSp) 1 from *Latrodectus hesperus*. Among the flagelliform and cylindriform silk proteins, signal peptides were only identified in both CySp1 proteins and the TuSp1.[14−17,19] Recently, the first three N-terminal sequences from dragline silk proteins, originating from MaSp2 of *Argiope trifasciata*, MaSp2 of *Nephila inaurata madagascariensis*, and MaSp1 from *Latrodectus geometricus* were published.[13] Several tentative start codons were found in the MaSp sequences, and two possible isoforms derived from different translational start sites were suggested. The shorter isoform was proposed to contain the most likely translational start site, due to a positionally conserved methionine codon and apparent similarity in N-terminal domain size compared to flagelliform silk, but signal peptides were not unequivocally identified. In addition to these N-terminal domains, two N-terminal sequences, encoding egg case protein 1 and 2 (ECP-1 and -2) have been reported.[14,20] However, these are clearly different from other spider silk proteins and are therefore not further discussed herein.

* Corresponding author. E-mail: jan.johansson@vmk.slu.se. Fax: +46-18-550762.

[†] Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences.

[‡] Department of Molecular Biosciences, Veterinary Medical Biochemistry, Biomedical Centre, Swedish University of Agricultural Sciences.

[§] Present address: Department of Medical Biochemistry and Microbiology, Biomedical Centre, Uppsala University, SE-751 23 Uppsala, Sweden.

10.1021/bm060693x CCC: $33.50
Published on Web 10/13/2006

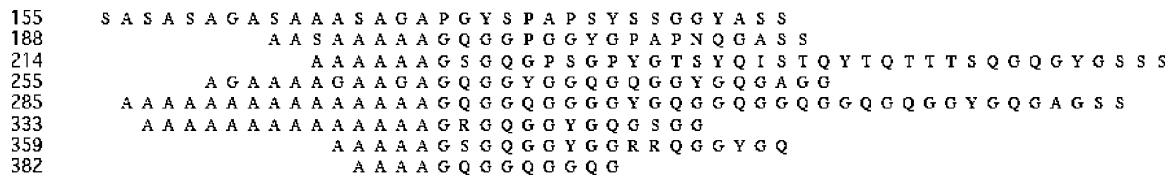

Figure 1. Amino acid sequence of the repetitive part translated from clone 55 (residues 155−394). The sequence repeats were aligned with respect to the last Ala in each Ala-rich region. The first residue in each row is numbered according to the position in the translated protein of clone 55 (EMBL accession no. AM259067).

On the basis of cDNA cloning and sequencing results, we have deduced the amino acid sequence of the N-terminal region of MaSp1 from *Euprosthenops australis*. When this sequence was compared to published sequences of dragline, flagelliform, and cylindriform silk proteins from other spider species, a general characterization of N-terminal signal peptides and a common nonrepetitive domain could be performed.

Materials and Methods

The major ampullate glands from approximately 100 adult female nursery web spiders (*E. australis*) were used for construction of a random primed custom pDONR222-based CloneMiner cDNA library (Invitrogen, Paisly, U.K.). Clones encoding MaSp were obtained by screening with a 1.4 kb cDNA fragment encoding Ala- and Gly-rich repeats, isolated from an oligo-dT primed library from *E. australis*.[21] Colony blotting and detection were performed using an ECL direct labeling and detection system (GE Healthcare, Uppsala, Sweden) according to the manufacturer's instructions. A single clone, denoted clone 55 (deposited in the EMBL databank with accession no. AM259067), was sequenced using the sense primer 5′-GTAAAAC-GACGGCCAGTCTTAAG-3′ and the antisense primer 5′-GCCAG-GAAACAGCTATGACCAT-3′, complementary to vector sequences, utilizing a MegaBase 1000 instrument (Amersham Biosciences). The sequences were analyzed using MacVector 7 (Accelrys, Cambridge, U.K.), PHDseq,[22,23] and SignalP 3.0.[9] Protein sequence similarity searches were performed using blastp and the NCBI databank. The Expasy tools (www.expasy.org) were applied for analysis of sequence motifs. For the phylogenetic analysis, a Neighbor-Joining tree was created using ClustalX 1.81 with correction for multiple substitutions and with gaps excluded.

A DNA fragment corresponding to nucleotide positions 84−524 of clone 55 (EMBL accession No. AM259067) and thus to amino acid residues 24−170 of *E. australis* MaSp1 (Figures 1 and 2A), was amplified by PCR with *LA Taq* (TaKaRa Bio, Saint-Germain-en-Laye, France) introducing a 5′ end *Eco*RV site and 3′ end stop and *Hind*III sites. Compared to the N-terminal domain of *E. australis* MaSp1 (Figure 2A), the recombinantly expressed protein lacks the predicted signal peptide and includes the first 16 amino acid residues of the repetitive part (Figure 1). The DNA fragment was subcloned into a modified pET32 vector (Merck Biosciences, Darmstadt, Germany) yielding a fusion protein comprising thioredoxin/His-tag/S-tag/thrombin cleavage site/*E. australis* MaSp1 N-terminal protein. The construct was subjected to sequencing to verify the correct sequence of the inserted fragment. *E. coli* BL21(DE3) cells (Merck Biosciences) were transformed and grown at 37 °C in Luria−Bertani medium containing ampicillin to an OD_{600} of 1, induced with IPTG, and further incubated for 4 h at 25 °C. The cells were harvested by centrifugation, disrupted by sonication (Vibra cell sonicator, Sonics and materials Inc., Danbury, U.S.A.) and further purified on Ni−NTA agarose (Qiagen, West Sussex, U.K.). Bound proteins were eluted from the column with 100 mM imidazole in buffer containing 20 mM Tris-HCl (pH 8.0) and dialyzed against 20 mM Tris-HCl (pH 8.0). Correct size and >90% purity of the fusion protein was confirmed by SDS−PAGE using 10% Bis−Tris gels (Invitrogen Inc., Carlsbad, CA) under reducing conditions. The *E. australis* MaSp1 N-terminal domain was released from the thioredoxin/His-tag/S-tag part of the fusion protein using a thrombin:fusion protein ratio of 1:1000 (w/w) at room temperature for 2 h, and isolated using a Ni−NTA column. Circular dichroism (CD) spectra of the recombinant *E. australis* N-terminal domain were recorded from 260 to 190 nm at 20 °C in a 0.1 cm path length quartz cuvette using a J-810 spectropolarimeter (Jasco, Tokyo, Japan). The scan speed was 50 nm/min, response time 2 s, acquisition interval 0.1 nm, and bandwidth 1 nm. The spectrum shown is an average of three consecutive scans.

Results and Discussion

A number of clones were isolated from an *E. australis* major ampullate gland cDNA library, after screening with a probe containing a MaSp1 Ala- and Gly-rich repetitive region. One of these, clone 55, contained a cDNA insert of 1.2 kb, and the translated sequence revealed a nonrepetitive region of 154 amino acid residues. Analysis of this sequence using blastp and the NCBI databank revealed no significant similarity except to N-terminal domains of spider silk proteins. The nonrepetitive region was followed by a 240-residue long repetitive part, consisting of segments of 4−15 Ala residues, interspersed with Gly-rich segments of variable length (Figure 1). These segments, in particular the Ala-rich segments, are more heterogeneous than the Ala-rich (12−15 Ala) and Gly-rich (consensus sequence of the dominant repeat GQGGQGGQGGLGQGGYGQGQGSS) segments of the downstream part of MaSp1 of *E. australis*.[21] However, the presence of poly-Ala segments sandwiched between Gly-rich repeats and the lack of repeated GPGXX motifs (which are characteristic of MaSp2 proteins) indicate that this clone is derived from a MaSp1 gene.[3,4,19]

The existence of several isoforms has previously been suggested for some spider silk proteins, since alternative translational starts have been identified.[13] Bearing this in mind, we analyzed the *E. australis* MaSp1 sequence for possible initiation sites. The nonrepetitive N-terminal domain of *E. australis* MaSp1 has a Leu residue at the position which corresponds to the translational start site proposed for flagelliform silk proteins and the short isoform of dragline silk[13,17,18] (position 81 in Figure 2A). *E. australis* MaSp1 thus cannot be translated into any corresponding short isoform. The N-terminal domain of *E. australis* MaSp1 contains 11 Met residues. One would expect a translational start site to be followed by a signal peptide. Each of the 11 potential start sites was therefore analyzed with SignalP for identification of signal peptides. As a result, three predicted signal peptides, corresponding to translational start sites at Met1, Met152, and Met158 (Figure 2A), were identified. The last two of these are located at the end of the N-terminal nonrepetitive domain, and none of the other N-terminal spidroin sequences has a Met at the position corresponding to Met158 in *E. australis* MaSp1. Therefore, it is unlikely that Met152 or Met158 represents translational start sites. This conclusion was supported by comparatively low signal peptide probability scores (0.600 and 0.664, respectively). In contrast, a maximum signal peptide probability score (1.000) was obtained for the 23-residue segment starting with Met1

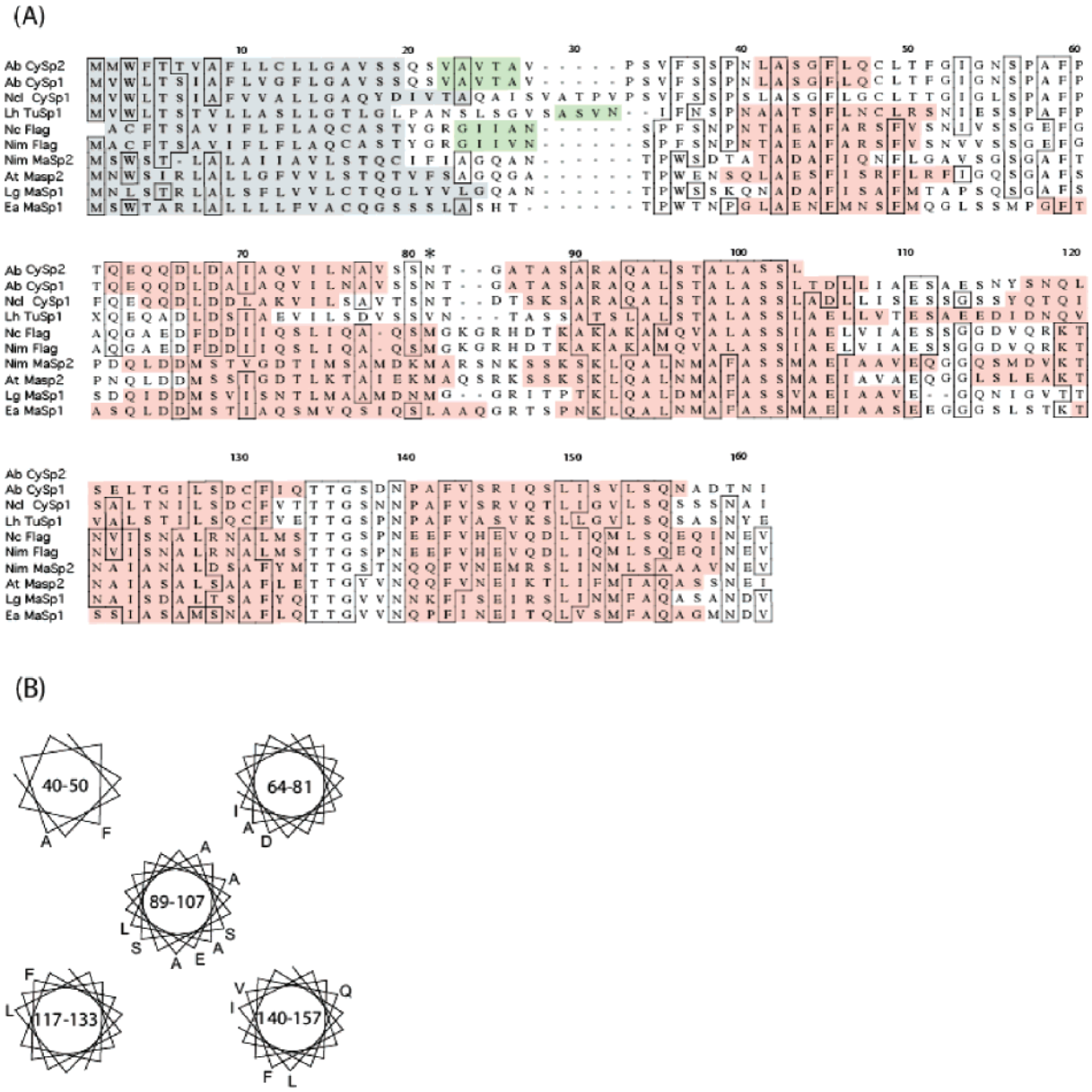

Figure 2. (A) Sequence alignment of all presently known nonrepetitive N-terminal regions of dragline, flagelliform, and cylindriform silk proteins. Signal peptides (probability scores >0.976 when analyzed with SignalP) are shaded in gray. Positions with identical residues in a majority of the proteins are boxed. Regions with predicted helical structures are shaded in red, and green shading indicates segments with β-sheet conformation (PHDseq probability scores >4 for at least 4 consecutive residues). The asterisk indicates the translational start position of a previously suggested short isoform.[13] The sequences shown are the following: Ab CySp2, *Argiope bruennichi* cylindriform spidroin 2, after translation of the nucleotide sequence as shown in Figure 1A in the Supporting Information, and described in detail in the text (GenBank accession no. AB242145); Ab CySp1, *Argiope bruennichi* cylindriform spidroin 1, residues 1−154 (GenBank accession no. BAE86855); Ncl CySp1, *Nephila clavata* cylindriform spidroin 1, residues 1−159 (GenBank accession no. BAE54451); Lh TuSp1, *Latrodectus hesperus* tubuliform spidroin, residues 1−156;[14] Nc Flag, *Nephila clavipes* flagelliform silk protein, residues 2−154 after editing of the nucleotide sequence as shown in Figure 1C in the Supporting Information, and described in detail in the text (GenBank accession no. AF027972); Nim Flag, *Nephila inaurata madagascariensis* flagelliform silk protein, residues 1−154 (translation of GenBank accession no. AF218623, starting with the first Met in frame); Nim MaSp2, *Nephila inaurata* madagascariensis major ampullate spidroin 2, residues 1−154 (GenBank accession no. AAZ15322); At MaSp2, *Argiope trifasciata* major ampullate spidroin 2, residues 1−155 (GenBank accession no. AAZ15371); Lg MaSp1, *Latrodectus geometricus* major ampullate spidroin 1, residues 1−151 after inserting a G in position 43 in the nucleotide sequence as shown in Figure 1B in the Supporting Information, and described in detail in the text (GenBank accession no. DQ059133, nucleotides 2012−2472); Ea MaSp1, *Euprosthenops australis* major ampullate spidroin 1, residues 1−154, as derived from the presently analyzed clone 55 (EMBL accession no. AM259067). (B) Helical wheel presentations of the predicted α-helices (positions 40−50, 64−81, 89−107, 117−133, and 140−157 in part A) in the N-terminal domain of spider silk proteins. Residues that are conserved in at least 7 of the N-terminal sequences and are predicted to adopt an α-helical structure in at least 5 of the sequences in part A are identified (omitting the incomplete *A. bruennichi* CySp2). Position one of each helix is plotted at the top of each wheel.

(Figure 2A), which indicates that it represents the true translational start site of *E.* australis MaSp1. This notion is corroborated by previous findings which have shown that the N-terminal regions of *N. inaurata madagascariensis* and *A. trifasciata* MaSp2 proteins,[13] the flagelliform silk protein of *N. inaurata madagascariensis*,[18] and the cylindriform silk proteins of *N. clavata* (CySp1),[16] *L. hesperus* (TuSp1),[14] and *A. bruennichi* (CySp1)[15] have Met residues in frame at the positions corresponding to Met1 of *E. australis* MaSp1 (Figure 2A). Moreover, signal peptide predictions indicate that, like in *E. australis* MaSp1, these Met residues are followed by typical signal peptides (varying in length between 18 and 23 residues

and probability scores of 0.989, 0.995, 0.999, 0.999, 0.999, and 0.999, respectively; Figure 2A).

It is noteworthy that the CySp2 protein from *A. bruennichi* was reported to differ in the extreme N-terminal part relative to CySp1, and also lacks any apparent signal peptide.[15] By aligning these two CySp cDNA sequences (GenBank accession nos. AB242144 and AB242145) it is apparent that CySp2 is lacking 352 nucleotides compared to CySp1 (Supporting Information, Figure 1A). The absence of these 352 nucleotides could, for instance, be the result of recombination events in the plasmid-harboring bacteria, a phenomenon which is known to occur albeit at low frequency. Possibly, the published sequence of CySp2[15] could be a splice variant, but this is less likely since it translates into a nonsecretory protein in which the N-terminal domain differs completely from other silk proteins. If the first 318 nucleotides of the CySp2 cDNA are translated, the resulting protein sequence is nearly identical to CySp1 (Figure 2A), and SignalP gives a signal peptide prediction value of 1.000. Also, analysis of the published sequence of the long isoform of *L. geometricus* MaSp1[13] predicts no signal peptide (zero signal peptide probability, data not shown). However, by aligning the DNA sequences encoding the different MaSp N-terminal domains, it becomes evident that the MaSp1 sequence of *L. geometricus* appears to lack a single nucleotide at position 43 (Supporting Information, Figure 1B). By inserting one nucleotide at this position, the *L. geometricus* MaSp1 DNA sequence translates into a protein that aligns over its entire length with other long MaSp1 proteins (see the edited sequence shown in Figure 2A). The edited *L. geometricus* MaSp1 sequence contains a 24-residue signal peptide starting with Met1 with a probability score of 0.976 (Figure 2A). This observation made us look closer at the published cDNA sequence of *N. clavipes* flagelliform silk,[17] since it translates into a protein where the proposed initiation Met residue (corresponding to position 81 in the alignment of Figure 2A) is localized downstream of the start site now proposed for *N. inaurata madagascariensis* flagelliform silk protein (see above). Again, by aligning the genomic DNA sequence of *N. inaurata madagascariensis* flagelliform silk with the *N. clavipes* cDNA sequence upstream of the proposed start codon, it becomes clear that the two sequences are very similar apart from the fact that the latter has a single nucleotide insertion at position 115 (Supporting Information, Figure 1C). Removal of this nucleotide makes the translated sequences of *N. clavipes* and *N.* inaurata madagascariensis flagelliform silk almost identical (Figure 2A), and generates a predicted signal peptide with a probability score 0.998 for the *N. clavipes* flagelliform silk protein. These single nucleotide differences can be the result of analysis misinterpretations of the DNA chromatogram traces or could have occurred, for instance, at cDNA synthesis during the library construction. The identity of the codon corresponding to Met1 in *N. clavipes* flagelliform silk cannot be determined at the present, since it is lacking in the published cDNA sequence (Supporting Information, Figure 1C).

Our data indicate that the localization of translational start sites and signal peptides in spider silk proteins has been conserved throughout the evolution of dragline, flagelliform, and cylindriform silk. The common site of translational start (Figure 2A) corresponds to the N-terminus of the recently identified long isoform.[13] In contrast, the presence of previously suggested short isoforms[13] is not a generally observed phenomenon in spidroins (Figure 2A). In four spidroins (MaSp2 of *N. inaurata madagascariensis* and *A. trifasciata* and flagelliform silk protein of *N. clavipes* and *N. inaurata madagascariensis*), the sequences following the initiation Met of the short isoform

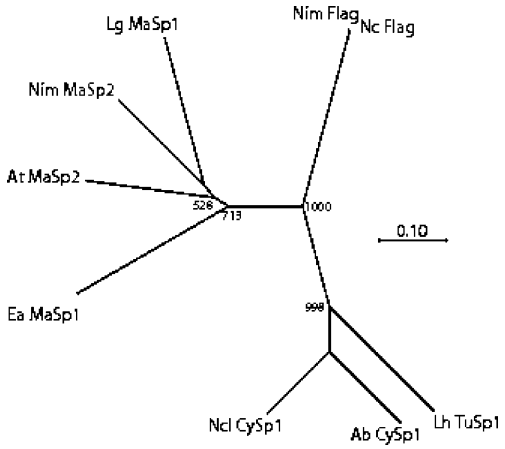

Figure 3. Phylogenetic tree of spider silk N-terminal domains (positions 34−161 in Figure 2A) constructed using the Neighbor-Joining method (ClustalX 1.81) with correction for multiple substitutions and with gaps being excluded. The number of recoveries in 1000 bootstrap replicates is given at each node. The bar represents a PAM value (percent accepted point mutations) of 10%.

(position 81 in Figure 2A) yield signal peptide probability scores above threshold (although comparatively low, ranging from 0.513 to 0.786). The other sequences either lack a Met at this position (MaSp1 of *E. australis* and cylindriform silk protein of *N. clavata*, *A. bruennichi*, and *L. hesperus*) or are not predicted to contain a signal peptide (*L. geometricus* MaSp1). It is possible that the short isoform of certain spidroins is produced as an alternative in case the translational start of the long isoform is missed.[13,24] However, this possibility appears less likely since the amino acid sequences are conserved between all analyzed dragline, flagelliform, and cylindriform silk proteins also N-terminally of the tentative cleavage sites of the signal peptides of the short isoform (Figure 2A; see further below). Thus, alignment of known spider silk N-terminal regions shows that in all cases the most likely start codon corresponds to the first Met. This is in line with the observation that the most proximal AUG triplet serves as the initiator codon in a majority of eukaryotic mRNAs.[25,26] Another common feature of start codons is the presence of a purine (most often adenine) in the −3 position relative to the AUG triplet.[25,26] In cases whenever the DNA sequences are available, all now proposed N-terminals share this feature.

The alignment in Figure 2A shows that the N-terminal domain beyond the signal peptide region is well conserved. For residues 34−161 dragline, flagelliform, and cylindriform spidroins have 13% identity. There are 35% identical residues between the four dragline (MaSp) spidroins analyzed, and the corresponding pairwise sequence identities range between 47% and 63%. As expected and discussed above, the two *Nephila* flagelliform proteins are nearly identical, with only one amino acid residue replacement for positions 34−161. The three cylindriform silk proteins, where a full-length N-terminal domain is available, have 51% identical residues, with 56−72% identity in the pairwise comparisons. Phylogenetic analysis shows that the proteins cluster according to their glandular origin (Figure 3). This indicates that the N-terminal domains of the major ampullate, cylindriform, and flagelliform silk proteins share a common origin dated before the divergence of *Nephila*, *Argiope*, *Latrodectus*, and *Euprosthenops*, which occurred some 125−240 million years ago.[27]

Secondary structure analysis of the N-terminal domain sequences using the PHD algorithm[22,23] yields a prediction of five α-helices covering the regions corresponding to positions

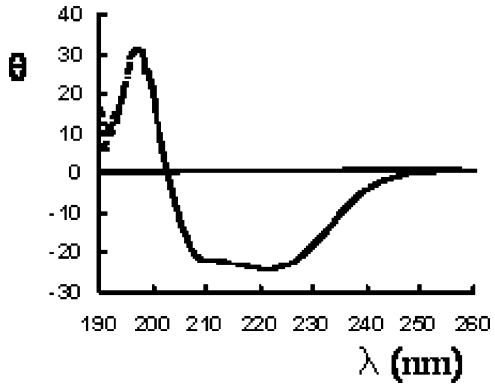

Figure 4. Far-UV CD spectrum of a recombinant *E. australis* MaSp1 N-terminal domain residues 24−170. The residual molar ellipticity, θ, is expressed in kdeg cm^2 dmol^{-1}.

40−50, 64−81, 89−107, 117−133, and 140−157 (Figure 2A). The nearly uniform presence of helices in these regions in all hitherto analyzed sequences supports a conserved structure for the N-terminal domain of spider silk. In all five helical regions, certain residues are strongly conserved between all species analyzed, and a majority of these residues have nonpolar side-chains. Plotting the location of these residues in helical wheel presentations makes it evident that, in all helices, with the exception of the last helix, the most conserved positions cluster at one face of each helix (Figure 2B). This is compatible with the burial of residues at these positions in a folded structure, while residues located at the less conserved helical faces are oriented toward the solvent. A possible fold for the nonrepetitive N-terminal domain of spider silk is thus a bundle of five helices connected by relatively short loops. To analyze the overall secondary structure of the N-terminal domain of *E. australis* MaSp1, a fragment corresponding to residues 24−170 (i.e., without the signal peptide and with a C-terminal overhang of 16 residues) was expressed in *E. coli*. The CD spectrum of the recombinant protein (Figure 4) contains minima at 208 and 222 nm and a maximum around 195 nm, and is thus typical for α-helical proteins. Estimation of the helical content from the residual molar ellipticity values at 208 and 222 nm[28] results in about 60% helices. This is in good agreement with the fact that the predicted five helices of *E. australis* MaSp1 N-terminal domain (Figure 2A) correspond to about 70% of the domain without the signal peptide. An intriguing observation is that within the segment connecting helices 4 and 5 (residues 134−139) a TTGXXN motif is strictly conserved (Figure 2A). When this motif is compared with known structural and functional motifs using the Expasy tools, no hits are found. Taken together, our data indicate that the N-terminal domain is conserved due to a general role in the structural and/or functional properties of dragline, cylindriform, and flagelliform spider silk proteins.

In conclusion, the characterization of the 5′ end of the *E. australis* MaSp1 gene has enabled a further analysis of common denominators for the nonrepetitive N-terminal domain of spider silk proteins, including the enigmatic signal peptide part. The results may prove useful for further understanding of the structural basis for the unique properties of spider silk.

Acknowledgment. We thank Professor Hans Ronne for useful advice regarding the phylogenetic analysis and Ulla Gustafsson for help with the DNA sequencing. This work was supported by the EU Commission grant "Spiderman" contract no G5RD-CT-2002-00738 and the Swedish Research Council. G.H. was funded by the Agricultural Functional Genomics program, Swedish University of Agricultural Sciences, Uppsala.

Supporting Information Available. Alignment of the sequences coding for: CySp1 and CySp2 from *A. bruennichi* (Figure 1A); the major ampullate spidroins from *N. inaurata madagascariensis, A. trifasciata, L. geometricus*, and *E. australis* (Figure 1B); the flagelliform proteins from *N. inaurata madagascariensis* and *N. clavipes* (Figure 1C). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Gosline, J. M.; DeMont, M. E.; Denny, M. W. *Endeavour* **1986**, *10*, 37−43.
(2) Candelas, G. C.; Ortiz, A.; Molina, C. *J. Exp. Zool.* **1986**, *237*, 281−5.
(3) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120−4.
(4) Hinman, M. B.; Lewis, R. V. *J. Biol. Chem.* **1992**, *267*, 19320−4.
(5) Kerkam, K.; Viney, C.; Kaplan, D.; Lombardi, S. *Nature* **1991**, *349*, 596−598.
(6) Knight, D. P.; Vollrath, F. *Proc. R. Soc. London, Ser. B* **1999**, *266*, 519−523.
(7) Bell, A. L.; Peakall, D. B. *J. Cell Biol.* **1969**, *42*, 284−95.
(8) Plazaola, A.; Candelas, G. C. *Tissue Cell* **1991**, *23*, 277−84.
(9) Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S. *J. Mol. Biol.* **2004**, *340*, 783−95.
(10) von Heijne, G. *J. Membr. Biol.* **1990**, *115*, 195−201.
(11) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *Int. J. Biol. Macromol.* **1999**, *24*, 271−5.
(12) Pouchkina-Stantcheva, N. N.; McQueen-Mason, S. J. *Comp. Biochem. Physiol., B* **2004**, *138*, 371−6.
(13) Motriuk-Smith, D.; Smith, A.; Hayashi, C. Y.; Lewis, R. V. *Biomacromolecules* **2005**, *6*, 3152−9.
(14) Hu, X.; Kohler, K.; Falick, A. M.; Moore, A. M.; Jones, P. R.; Vierra, C. *Biochemistry* **2006**, *45*, 3506−3516.
(15) Zhao, A. C.; Zhao, T. F.; Nakagaki, K.; Zhang, Y. S.; Sima, Y. H.; Miao, Y. G.; Shiomi, K.; Kajiura, Z.; Nagata, Y.; Takadera, M.; Nakagaki, M. *Biochemistry* **2006**, *45*, 3348−3356.
(16) Zhao, A.; Zhao, T.; Sima, Y.; Zhang, Y.; Nakagaki, K.; Miao, Y.; Shiomi, K.; Kajiura, Z.; Nagata, Y.; Nakagaki, M. *J. Biochem. (Tokyo)* **2005**, *138*, 593−604.
(17) Hayashi, C. Y.; Lewis, R. V. *J. Mol. Biol.* **1998**, *275*, 773−84.
(18) Hayashi, C. Y.; Lewis, R. V. *Science* **2000**, *287*, 1477−9.
(19) Bini, E.; Knight, D. P.; Kaplan, D. L. *J. Mol. Biol.* **2004**, *335*, 27−40.
(20) Hu, X.; Kohler, K.; Falick, A. M.; Moore, A. M.; Jones, P. R.; Sparkman, O. D.; Vierra, C. *J. Biol. Chem.* **2005**, *280*, 21220−30.
(21) Stark, M.; Grip, S.; Rising, A.; Hedhammar, M.; Engström, W.; Hjälm, G.; Johansson, J. Submitted.
(22) Rost, B.; Sander, C. *J. Mol. Biol.* **1993**, *232*, 584−99.
(23) Rost, B.; Sander, C. *Proteins* **1994**, *19*, 55−72.
(24) Su, L. K.; Qi, Y. *Genomics* **2001**, *71*, 142−9.
(25) Kozak, M. *Nucleic Acids Res.* **1984**, *12*, 857−72.
(26) Peri, S.; Pandey, A. *Trends Genet.* **2001**, *17*, 685−7.
(27) Gatesy, J.; Hayashi, C.; Motriuk, D.; Woods, J.; Lewis, R. *Science* **2001**, *291*, 2603−5.
(28) Barrow, C. J.; Yasuda, A.; Kenny, P. T. M.; Zagorski, M. G. *J. Mol. Biol.* **1992**, *225*, 1075−93.

BM060693X