J. Biochem. **138**, 593–604 (2005)
doi:10.1093/jbchem/mvi155

Unique Molecular Architecture of Egg Case Silk Protein in a Spider, *Nephila clavata*

Aichun Zhao[1], Tianfu Zhao[1], Yanghu SiMa[2], Yuansong Zhang[1], Koichi Nakagaki[1], Yungen Miao[3], Kunihiro Shiomi[1], Zenta Kajiura[1], Yoko Nagata[4] and Masao Nakagaki[1,*]

[1]Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda 386-8567; [2]Faculty of Agricultural Science and Technology, Suzhou University, Su Zhou 215006, China; [3]College of Animal Sciences, Zhejiang University, Hang Zhou 310029, China; and [4]College of Science and Technology, Nihon University, Tokyo 101-8308

Received July 13, 2005; accepted August 17, 2005

We describe a unique silk protein secreted from the cylindrical silk glands of the spider *Nephila clavata*. This silk is primarily composed of three proteins, whose transcripts of approximately 16.0, 14.5 and 13.0 kb are homologous to one another in two termini and repetitive units, as determined on Northern blotting. Its overall organization shows that it is similar to other characterized silk proteins, including in the mainly central repetitive region as well as the non-repetitive N-terminal (166 residues) and C-terminal (176 residues) parts. However, up to 90% of the protein consists of highly ordered repetitive structures that are not found in other silks. The repetitive region mainly consists of several types of complexes and remarkably conserved polypeptide repeats. The assembled repeat units (A_1B_1) contain a high proportion of Ala (30.41%), Ser (25.15%), and residues with hydrophobic side chains (22.22% for Gly, Leu, Ile, Val and Phe combined). The presence of Ser-rich and GVGAGASA motifs suggests the formation of a β-sheet. The repetitive region is characterized by alternating arrays of hydrophobic and hydrophilic blocks. The results suggested that this egg case silk is an exceptional protein when compared with previously investigated spider silks.

Key words: cylindrical gland, egg case silk fibroin, *Nephila clavata*, spider, repetitive structure.

Abbreviations: MA, major ampullate silk gland; MI, minor ampullate silk gland; FL, flagelliform silk gland; CY, cylindrincal silk gland; Sp, silk protein.

Silk is a proteinaceous polymer secreted from specialized exocrine glands of several groups of arthropods. The orb-web spiders have seven sets of silk-producing glands, each synthesizing a different type of silk (*1*). The different and unique mechanical properties of these silks enable them to be used for a variety of diverse purposes ranging from the arresting prey to protection of their offspring enclosed in the egg case. However, the common features of the silks characterized are that they are composed almost solely of proteins having a predominance of Ala, Gly and Ser in their amino acid compositions, and that they undergo essentially irreversible transformation from soluble proteins into insoluble fibers. The exceptional mechanical properties (*2*) and the number of potential applications of spider silks have made them the focus of a number of molecular and structural studies (*3–6*). To date, the partial and full primary structures of several types of spider silk, such as major ampullate silk, minor ampullate silk, flagelliform silk and aciniform silk, have been elucidated (*7–11*). The ADF-2 clone obtained by Guerette *et al*. (*12*) from a cylindrical gland complementary DNA (cDNA) library of *Araneus diadematus* and ECP-1 reported by Hu *et al*. (*13*) are likely to be two constituents of the egg case, but their predicted amino acid compositions differ markedly from that reported for this silk. The primary structure of the cylindrical silk protein that primarily comprises the spider egg case has not yet been fully characterized.

Spider egg case silk is secreted by the cylindrical glands (*1*, *14*, *15*), which differ from all other silk glands of spiders in that they are not used daily. Egg cocoons are normally produced only once or twice in the lifetime of a spider (*16*). The egg case of a spider serves to protect the enclosed offspring during the early stages of their lives. It must be sufficiently robust to resist such threats as predator/parasitoid invasion and temperature fluctuations (*17–19*). Analysis of the amino acid compositions of the silks of several web-orb spiders in previous investigations (*20–22*) showed that Gly is the most abundant component of major ampullate silk, minor ampullate silk and flagelliform silk, while that of cylindrical silk is Ala or Ser (Table 1).

In order to elucidate the structure and function of spider egg case silk, we investigated and cloned the genes of these silk proteins using a novel strategy. Here, we report the first description of a substantial cylindrical silk protein and cDNA sequences. Our results provide an insight into the relationship between protein structure and function.

MATERIALS AND METHODS

Spider Silk Gland Preparation—Mature female *N. clavata* spiders, in which the cylindrical glands are very prominent, were collected from October to November.

*To whom correspondence should be addressed. Tel: +81-268-21-5335, Fax: +81-268-21-5331; E-mail: nakagak@giptc.shinshu-u.ac.jp

Table 1. **Amino acid compositions (mole%) of luminal contents or fibers of glands of various spiders.**

	Silk glands		Cylindrical			Major ampullate	Minor ampullate	Flagelliform
Species source	*N. clavata* luminal contents	*A. aurantia* extracted egg case outer cover[b]	*L. hesperus* egg case[c]	*N. edulis* luminal contents[d]	*A .diadematus* luminal contents[a]	*A. diadematus*[a]		
Ala	28.28	27.13	23.3	27.31	24.44	17.6	36.75	8.29
Arg	3.01	1.95	1.3	2.18	1.49	0.57	1.69	1.13
Asp/Asn	3.25	6.96	6.3	3.87	6.26	1.04	1.91	2.68
Cys	0	—[e]	—	0.09	—	—	—	—
Gln/Glu	8.27	7.67	11	8.59	8.22	11.49	1.59	2.89
Gly	8.74	8.76	7.7	8.61	8.63	37.24	42.77	44.16
His	0	0.19	—	0.32	trace	trace	trace	0.69
Ile	2.14	1.7	2	2.79	1.69	0.63	0.67	1.01
Leu	6.65	6.76	5	7.41	5.73	1.27	0.96	1.4
Lys	0.61	0.27	—	1.64	1.76	0.54	0.39	1.35
Met	0	0	—	1.44	—	—	—	—
Phe	3.73	4.23	—	3.85	3.22	0.45	0.41	1.08
Pro	0	0.6	2	—	0.59	15.77	trace	20.54
Ser	23.81	24.17	25.7	20.18	27.61	7.41	5.08	3.08
Thr	6.52	3.64	5	5.89	3.44	0.91	1.35	2.48
Tyr	1.02	1.25	1.7	2.05	0.95	3.92	4.71	2.56

[a]Data from Andersen (*20*), [b]Data from Foradori, M.J. *et al*. (*24*), [c]Data from Casem, M.L. *et al*. (*22*), [d]Data from Dicko, C. (*25*), [e]Suggested that this amino acid was not detected or shown in original references. *A. diadematus* = *Araneus diadematus*; *A. aurantia* = *Argiope aurantia*; *L. hesperus* = *Latrodectus hesperus*; *N. clavata* = *Nephila clavata*.

Cylindrical glands from selected spiders were harvested under a dissection microscope. Part of the harvested cylindrical glands was dissolved in SDS-PAGE loading buffer for SDS-PAGE analysis, the remainder being used for RNA isolation and amino acid compositions analysis.

Amino Acid Composition Analysis—Percent amino acid composition analysis of the luminal contents of the individual cylindrical glands was accomplished in collaboration with the College of Science and Technology of Nihon University. Protein samples were subjected to acid hydrolysis in preparation for amino acid analysis with an L-8800 amino acid analyzer (Hitachi, Tokyo).

RNA Isolation and cDNA Library Construction—Cylindrical glands (50–100 mg) were dissected from euthanized *N. clavata* and then rapidly homogenized in 1 mL of ISOGEN. Total RNA and mRNA were extracted from the glands and cDNA was synthesized using the protocols described in the manuals for ISOGEN, MicroFastTrack™ 2.0 and cDNA synthesis kits (Nippon Gene, Invitrogen and Takara, Japan, respectively). Two different primers were used for first strand synthesis for construction of two cDNA libraries. One oligo-d(T) primer was utilized in the initial process of cDNA synthesis and the other specific primer, whose sequence is based on the sequence of NcCy-21, was then used to synthesize cDNA. Library construction was completed by blunt-end ligating the synthesized cDNAs into pGEM-3zf(+), which was digested with the restriction nucleotidase *Sma*I and then treated with CIAP (calf intestine alkaline phosphatase).

cDNA Library Screening and DNA Sequencing—XL1-Blue cells were transformed, and the first library of more than 1,000 recombinant colonies identified on α-complementation, as described in a laboratory manual for molecular cloning written by Sambrook *et al*. (*23*), was constructed. The novel method used to screen the cDNA library was based on the hypothesis that the proportion of egg case fibroin gene mRNA is perhaps the highest among mRNAs from the mature cylindrical gland. The partial sequences of more than 100 recombinant colonies were sequenced and analyzed using an ABI Prism® genetic analyzer 373, 310 or 3100. Full sequencing of the longest insert fragments among these clones was completed in detail. The sequence of the longest insert was determined by producing a series of nested deletions with a deletion kit containing exonuclease III. The sequence data have been deposited with DDBJ, the accession numbers being AB218973 for Cy3′ and AB218974 for Cy5′.

Northern Blotting—Aliquots of about 5 μg of total RNA were prepared from the cylindrical glands, major ampullate glands, minor ampullate glands and flagelliform glands of five spiders. Total RNAs together with RNA size markers were electrophoresed through a 1% agarose gel (1× MOPS, 0.66 M formaldehyde) (*23*) and then blotted. Using the hybridization conditions given in the manual for the DIG nucleic acid detection kit (Boehringer Mannheim), Northern blots were probed with a Digoxygenein-labeled probe. Relative positions were evaluated with RNA size markers. Three different probes were used in three Northern blotting experiments. The 5′-terminal region probe was generated using the following primers: Cy5′ (752–731) (5′-TGTTTGATCCAGTTGTTGTGAC-3′) and Cy5′ (104–125) (5′-TTTTACTATGGATTCTGGGCTC-3′). Cy5′ recombinant plasmid DNA (20 ng) was used as a PCR template. The resulting PCR product is a 649-bp 5′-terminal region composed of a non-repetitive sequence. The repeat region probe was generated using the following primers: NcCy-21M13F (393–410) (5′-TCTGCATCTACCTACGCT-3′) and pUC/M13 Reverse Primer. Cy5′ recombinant plasmid DNA (20 ng) was also used as a template. The 485-bp PCR product was purified from the agarose gel, and then digested with *Xba*I in order to remove the vector sequence. A 399-bp DNA fragment

was recovered from the gel after electrophoresis. The resulting DNA fragment was composed of a 390-bp repetitive region and 9-bp vector sequences. The 3′-terminal probe was obtained by digesting clone NcCy-21 with *Sna*BI and *Hin*dIII. The resulting 268-bp band was purified from the gel. All the obtained fragments were labeled with DIG-HIGH-Primer reagent.

Comparative Analysis—Computer analysis of DNA and amino acid sequences was conducted using the Genetyx package (Windows version, Genetyx, Inc) and a Sequencer 4.1.4 (Demo version), respectively. Comparisons were performed through the Internet using the new cylindrical gland gene sequences and previously published spider silk genes.

RESULTS

Amino Acid Composition and SDS-PAGE—Samples of the luminal contents of cylindrical glands were collected and processed for amino acid composition analysis (Table 1). For comparison, the amino acid compositions of the major ampullate, minor ampullate, flagelliform and cylindrical silks from several other spiders are also presented. In major ampullate, minor ampullate and flagelliform silks, Gly is the most abundant amino acid present. In contrast to these silks, the protein from the cylindrical glands had a higher Ser content with a concomitant loss of Gly. Ala was still a significant component of this silk. The overall percent amino acid composition of *N. clavata* corresponds well to the values for silk proteins from cylindrical glands reported previously (*20*, *22*, *24*, *25*). Cys and Met residues were not detected in any of the silks.

Figure 1 shows an SDS-PAGE gel with two apparent bands obtained for the solubilized contents of the cylindrical glands (lanes 2–6). The molecular weights of the two materials are approximately 342 and 303 kDa.

Isolation of the Cylindrical Glands cDNA Clones—More than 200 recombinant plasmid DNAs were purified from cultures. The two sides of insert fragments (of about 800 bp) of more than 100 recombinant plasmids were sequenced. Homology analysis of the sequences, which are characterized by ones containing a polyA tail and polyadenylation signal AATAAA sequences, was conducted using Sequencher and Genetyx. Of the 100 sequences, 17 containing polyA tails from different recombinant plasmids can be grouped together. Further analysis revealed that the sequences are identical to one another, except in size. Among the 17 recombinant plasmids, one recombinant plasmid with the longest insert was used for further research and was designated as NcCy-21. The NcCy-21 insert is a 1,685-bp cDNA fragment that includes the 3′-terminus and part of the repeat sequence. A primer (5′-ACCCAGGAACACCGACAGAAAT-3′) was designed from the 3′-non-repetitive sequence of NcCy-21 and then used for first-strand cDNA synthesis with template mRNA from mature cylindrical glands. The cDNAs were then ligated into pGEM-3zf(+). XL1-Blue cells were transformed and a library of approximately 500 recombinant colonies was obtained. Twenty-two of these recombinant plasmids, in which the inserts are longer than about 1,000 bp, were partially sequenced. Most of the sequenced inserts overlapped with NcCy-21. A recombinant clone containing a 3.5-kb insert, which fully

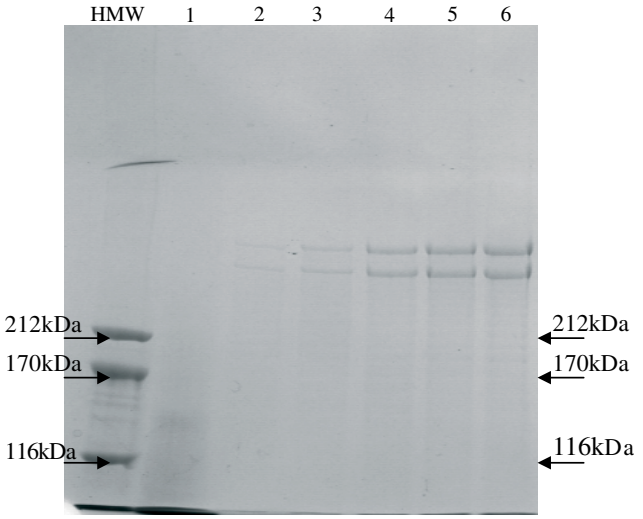

Fig. 1. **SDS-PAGE of luminal proteins of cylindrical glands under reducing conditions (the reducing reagent was 2-mercaptoethol).** The analytical gel was 6% polyacrylamide gel. HMW indicates high molecular weight markers, and the bands at 212 kDa, 170 kDa and 116 kDa are myosin, α_2-macroglobulin and β-galactosidase, respectively. Lane 1 is a blank control. Lanes 2–6 (2, 4, 6, 8 and 10 μl of sample were loaded, respectively) contain cylindrical gland proteins in different sample loading volumes. molecular size determination was based on the relative mobilities of markers and experimental samples. These were plotted against the log10 of known molecular weights, there being a linear relationship between the two (*26*).

overlapped with NcCy-21, was fully sequenced as described under Materials and Methods. The two sequences of this clone and NcCy-21 were combined into Cy3′, which represents the C-terminus and preceding repetitive regions. Four clones including a putative N-terminal sequence followed by the same repetitive unit as in Cy3′ were obtained. The sequence of one of the four clones having a 2.6-kb insert was designated as Cy5′. The translated repetitive region sequences were consistent with the amino acid composition of the cylindrical gland protein, most notably in that Ser and Ala accounted for over 55% of the residues (data shown in Fig. 5). Northern analyses indicated three large transcripts of approximately 16.0, 14.5 and 13.0 kb. The lengths of these transcripts are comparable to the 12.5-kb MaSp1 (*7*), 14.5- and 15.5-kb Flag mRNAs (*10*), but are longer than the 9.5- and 7.5-kb MiSp mRNAs (*9*). The smallest transcript is consistent with the size of the obtained molecular masses of electrophoresed cylindrical gland proteins in this study. Identical bands of different density were detected on Northern blotting with three different probes (for the repeat sequence, and the 3′- and 5′-termini), which suggests that the three transcripts are homologous to one another in overall structure (Fig. 2). In addition, the Northern blot results apparently suggest that the amount of the smallest transcript is much greater than those of the two larger transcripts.

Cylindrical Fibroin Gene Structure—N-terminal region. The identified Cy5′ sequence encodes a putative peptide of 757 amino acid residues. The expanse of the first 166 residues exhibits no obvious sequence iterations (Figs. 3 and 6b), but the remainder consists of regular repeats. The first 166 residues of Cy5′ comprise a potential

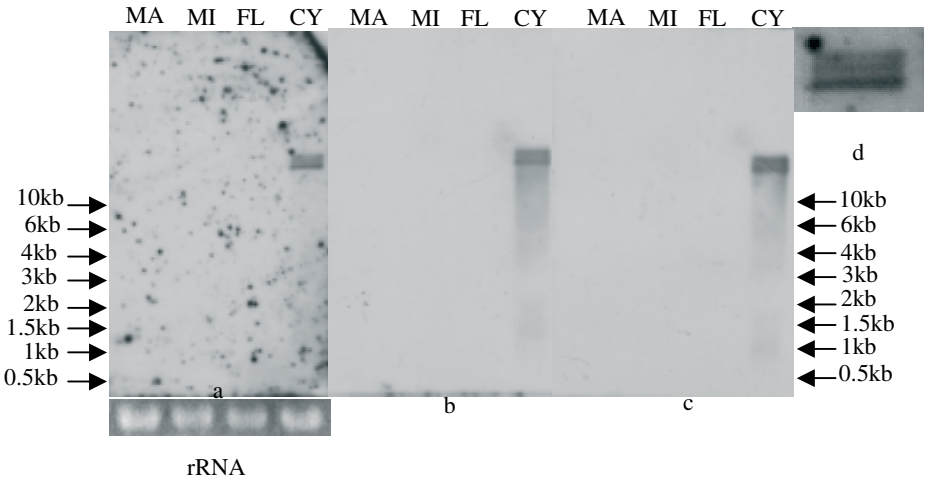

Fig. 2. **Northern blot analysis of the silk gland specific distributions of Cy5′ and Cy3′.** Total RNA (5 μg) from major ampullate glands (MA), minor ampullate glands (MI), flagelliform glands (FL), and cylindrical glands (CY) was probed with (a) a 649-bp 5′-terminal probe, (b) a probe containing the 390-bp repeat sequence and (c) a 268-bp 3′-terminal probe, respectively. d shows an amplified photo of the bands in a. Ribosomal RNA stained with ethidium bromide was used to confirm RNA integrity and relative quantity. The sizes of RNA markers are shown by arrows.

```
1   CCGTGATCGTCTGCGCTTTTCTGCTGGCTATTTATTTTAAGCGGGAAACTGTTATGACTTGTTCCACAAGCCGAAGTTCAAAAGTGCATT
91  AGATGTTCTGCGTTTTTACTATGGATTCTGGGCTCTGTACAGAGGTCTGAGGTATAAATAGAAACAATATTATTATTATTGAAGAATTTT
181 TCACCAAAAGATAATGCTTTGAAAATAGATAAATTCTGAATCACGAACTTCTATAATTAAAGATGGGACTGTCTTGTGGAGCAAAATGTG
271 CCAAAGGGTTTCTGGTCACTTTTAATTTAATATTTTGGATTTCTGTCGGTGTTCCTGGCTGATCGAAACAGCAATGGTTTGGCTGACAAG
                                                                    M  V  W  L  T  S
361 CATAGCGTTTGTTGTGGCTCTTTTAGGAGCACAATACGACATCGTGACTGCTCAGGCAATTTCAGTTGCAACTCCTGTCCCATCAGTGTT
    I  A  F  V  V  A  L  L  G  A  Q  Y  D  I  V  T  A  Q  A  I  S  V  A  T  P  V  P  S  V  F
451 CAGTAGCCCTAGCCTTGCGAGTGGTTTCCTTGGATGCCTCACAACTGGTATTGGACTATCTCCAGCTTTCCCGTTTCAAGAACAACAAGA
    S  S  P  S  L  A  S  G  F  L  G  C  L  T  T  G  I  G  L  S  P  A  F  P  F  Q  E  Q  Q  D
541 TTTAGATGACTTAGCCAAGGTAATTCTCTCCGCAGTAACCAGTAATACTGACACCTCAAAGTCAGCGAGAGCACAAGCCTTGAGCACTGC
    L  D  D  L  A  K  V  I  L  S  A  V  T  S  N  T  D  T  S  K  S  A  R  A  Q  A  L  S  T  A
631 ATTAGCATCTTCCTTAGCCGACCTACTGATATCCGAATCAAGTGGAAGCAGCTACCAAACTCAAATATCTGCCCTCACTAATATCCTATC
    L  A  S  S  L  A  D  L  L  I  S  E  S  S  G  S  S  Y  Q  T  Q  I  S  A  L  T  N  I  L  S
721 CGATTGTTTTGTCACAACAACTGGATCAAACAATCCTGCATTTGTATCAAGAGTTCAAACACTTATAGGAGTGCTTTCTCAAAGCAGCAG
    D  C  F  V  T  T  T  G  S  N  N  P  A  F  V  S  R  V  Q  T  L  I  G  V  L  S  Q  S  S  S
811 TAATGCAATTTCAGGAGCAACAGGTGGCTCC
    N  A  I  S  G  A  T  G  G  S
```

Fig. 3. **Cy5′ includes a putative N-terminal sequence.** The coding sequence of the most proximal Met is highlighted with an underline. The presumed signal peptide, whose cleavage site is indicated by an arrow, is shown as bold italics. The reading frame was unambiguously determined from the lack of stop codons throughout the Cy5′ sequence.

N-terminal sequence for the spider silk, which to date has only been reported in the flagelliform fibroin gene sequence. The region has a different amino acid composition and sequence from the repetitive portions of both Cy5′ and Cy3′. Following this N-terminal region, the remaining Cy5′ sequence proceeds without interruption to the repeats found in the rest of the protein (Fig. 6b). The only in-frame start codon is at position 344. There is no ATG codon upstream of position 344 that allows continuous translation through the rest of the sequence. This suggests that the transcription of the cylindrical fibroin gene starts at position 344.

In addition to the start codon, the 5′ mRNA sequence is also likely to encode a signal peptide. Secretory signals are key elements of silk proteins, like spider silks and silkworm silk, which must be transported across the endoplasmic reticulum and secreted. A signal peptide typically consists of three regions: (i) a positively charged, short N-region; (ii) a hydrophobic, 7–15 residue H-region; and (iii) a relatively polar, 3–7 residue C-region that contains the signal peptidase cleavage site (*27*, *28*). The translated amino acid sequence was analyzed with signalP 3.0 software (eukaryotic option) (*29*) in order to determine whether a likely signal peptide was present. The analysis suggested that the three regions described above are present in the Cy5′ sequence. The most likely cleavage site exists between the Ala coded at positions 411 to 413 and the Gln coded at positions 414 to 416 (Fig. 3).

```
N.clavipes Flag          ------------------------------------------------------------
N.madagascariensis Flag  ------------------------------------------------------------
N.clavata CySp1          MVWLTSIAFVVALLGAQYDIVTAQAISVATPVPSVFSSPSLASGFLGCLTTGIGLSPAFP

N.clavipes Flag          ----------------MGKGRHDTKAKAKAMQVALASSIAELVIAESSGGDVQRKTNV
N.madagascariensis Flag  ----------------MGKGRHDTKAKAKAMQVALASSIAELVIAESSGGDVQRKTNV
N.clavata CySp1          FQEQQDLDDLAKVILSAVTSNTDTSKSARAQALSTALASSLADLLISESSGSSYQTQISA
                                 .:   ...*:*:*:..*****:*:*:*:****.. * : ..

N.clavipes Flag          ISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEVDTSGPGQYYRSSSSGGGGGGQ
N.madagascariensis Flag  ISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEVDTSGPGQYYRSSSSGGGGGGG
N.clavata CySp1          LTNILSDCFVTTTGSNNPAFVSRVQTLIGVLSQSSSNAISGATGGS----------
                         ::* * :.:::**** *  ** .** ** :***.. * :. :  *.

N.clavipes Flag          GGPVVTETLTVTV
N.madagascariensis Flag  GGPVITETLTVTV
N.clavata CySp1          -------------
```

Fig. 4. **CLUSTAL W alignment of the N-terminal domains of two flagelliform silk proteins and the cylindrical silk protein.** At the bottom of the alignment, asterisks represent identical amino acids, one dot (.) represents conserved substitutions, and two dots (:) represent semi-conserved substitutions. Amino acids are indicated by one-letter abbreviations. *N.* in this figure is the abbreviation of *Nephila*. The accession numbers are AAC38846 for *N. clavipes* Flag, and AAF36091 for *N. madagascariensis* Flag.

The N-terminal sequences of only two spider sequences, those of the flagelliform silks of *Nephila clavipes* (AAC38846) and *Nephila madagascariensis* (AAF36091), have been reported. This may reflect the challenges posed by the cloning and sequencing of silk sequences, as truncation of longer cDNA partial clones from the spider *N. clavipes* is routinely observed in *Escherichia coli* (*30*). The two potential N-terminal sequences (115aa), which are 98% identical, were aligned with that (166aa) of the cylindrical silk found in our study (Fig. 4). Alignment analysis showed the N-terminal sequence of the cylindrical gland protein from residue 78 to 166 is considerably homologous (up to 80%) to those of the others. Up to 44% of the region from residue 87 to 165 in the cylindrical gland protein is identical to that from residue 7 to 79 of the two flagelliform proteins.

Finally, we analyzed the two flagelliform sequences using signalP 3.0 and 1.1 software, respectively. The results suggested that the two flagelliform sequences do not apparently retain a signal peptide, which is consistent with the results obtained by Bini *et al*. (*31*), although the same analysis suggested that the potential N-terminus of the cylindrical silk apparently retains a signal peptide.

Repetitive region: The majority of the translated cylindrical protein sequence is composed of Ala and Ser, Gln and Gly being the next most abundant. Comparison of the Cy5′ and Cy3′ repetitive sequences showed that they are nearly indistinguishable. The repeats found in Cy5′ and Cy3′ are identical, indicating that they are iterated throughout the central gene region between the two ends (Fig. 6a). Northern analysis showed that the cylindrical silk proteins contain three transcripts having identical central repeat units and similar sizes, which encode about 4,600 amino acid residues. The deduced molecular mass of *N. clavata* cylindrical fibroin is thus close to 418 kDa, and over 90% of the molecule consists of repeats. They are built up of predominantly hydrophobic and neutral residues, and only about 6% of the residues are hydrophilic. The repeat composition is decisive for the amino acid representation in the whole fibroin. The relative representation of amino acid residues in the deduced translation products of an assembled repeat, A_1B_1, from Cy5′ or Cy3′ is consistent with the results of direct amino acid analysis of the luminal contents of cylindrical silk glands (Fig. 5).

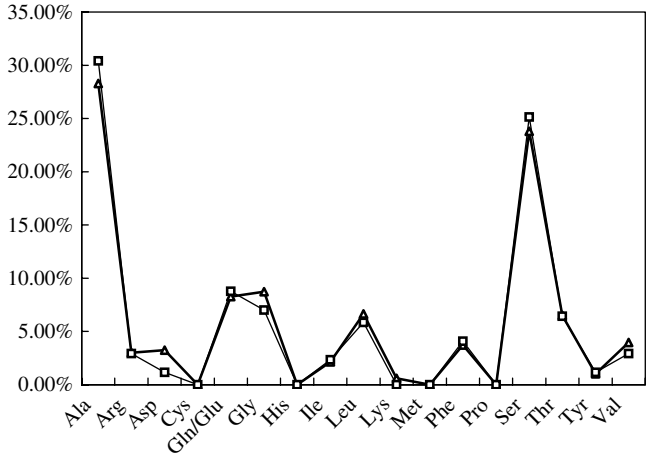

Fig. 5. **Comparison of the amino acid composition (mole %) of an assembled repeat unit to that of the luminal contents.** The composition of the assembled repeat unit is represented by the dimetric line, and that of the luminal contents is represented by the triangular line.

Northern analyses with probes specific to sequences found in Cy5′ (5′-probe), but not Cy3′, and *vice versa* resulted in identical-sized bands (Fig. 2). While it is possible that the two cylindrical protein sequences are derived from different genes with similar functions and sequences, it is likely that Cy5′ and Cy3′ represent the

proximal region and distal ends of the same gene. Future characterization of genomic sequences will be necessary to examine between these possibilities. All analyses of the repetitive region were based on the combined Cy5′ and Cy3′ sequences. Identified short motifs, such as GPGGX, GP(S, Y, G), $(GA)_n/A_n$ and GGX, which apparently exist in the sequences of several other spider silks (*7–10*), were not found in the cylindrical silk. Beginning at predicted amino acid residue 166, the sequence of the *N. clavata* cylindrical fibroin continues as a regular repetitive arrangement of four types of complex repeat (Fig. 6b). Repeats A_1 and A_2 consist of about 110 and 98 residues, and repeats B_1 and B_0 of 61 and 48 amino acid residues, respectively. Repeat B_0 only represents the beginning of the repeat region and is part of repeat B_1, while repeat A_2 only represents the end of the repeat region and is also a major portion of repeat A_1. These repeats are combined in higher order repetition A_1B_1, and are arranged into a large assembly $[B_0(A_1B_1)_nA_2]$. We estimate that the *n* value is about 24. Both the absence of any variability in the length of repeats and the high conservation of their sequences are striking (Fig. 6, a and b). For example, the nucleotide sequences of various A_1 repeats differ by only 6 bases from one another, while various B_1 repeats differ by only 4 bases. Of the 6 changes in repeat A_1, all are silent substitutions and the amino acid sequences are completely identical to one another; however, all 4 changes in repeat B_1 are non-silent substitutions. The repetitive nature of the cylindrical silk sequence with alternating hydrophilic and hydrophobic blocks can be apparently visualized using hydropathy plots (Fig. 7). In the Kyte-Doolittle plot, hydrophobic regions are represented by peaks above 0 and hydrophilic regions by negative values. In Fig. 7a, the downward spikes represent the hydrophilic B_1 repeat between the larger hydrophobic A_1 repeats. The proportion of hydrophobic residues in repeat A_1 is 63%, while that of hydrophilic and neutral residues in repeat B_1 is more than 67%. In Fig. 7, the spikes for repeat B_1 are largely below the abscissa axis, while those for repeat A_1 are largely above.

C-Terminal region: The obtained Cy3′ sequence, which comprises 3,893 bp and includes a 113-bp 3′-untranslated region containing polyadenylation signal AATAAA (*32*) localized 17 bp upstream from the polyadenylated tail, encodes a putative peptide of 1259 amino acids residues. The C-terminal region (about 176 residues) is distinct from the core repetitive sequence, as it exhibits no obvious sequence iterations (Fig. 6b).

The C-terminal amino acid sequences of previously published spider silks from *Nephila* and the cylindrical silk in this study were aligned and systematically analyzed. Upon alignment, these sequences almost fall into separate groups, depending on the type of silk. The C-termini of the major ampullate, minor ampullate and flagelliform silks of several different spiders are well conserved, but the C-termini of cylindrical silks exhibit a low degree of similarity with other silks (Fig. 8a). The NJ tree obtained shows that CySp1 cannot cluster robustly with other silk proteins, while several silk proteins from different spiders can form two groups (Fig. 8b). The above results suggest that CySp1 may be a new member of the spider silk gene family or an orthologous gene.

DISCUSSION

Screening of a cDNA Library—The novel method used for screening egg case fibroin cDNA was based on the hypothesis that the proportion of its mRNA would be the highest in transcripts from mature cylindrical glands. Because the amino acid and nucleotide sequences were unknown before cloning of the cylindrical fibroin gene, we had to resort to a screening method. Our research showed that a screening method based on a statistical approach is feasible when cloning a new gene of unknown sequence.

Cylindrical Silk Protein—The electrophoresis of the *N. clavata* cylindrical silk protein in a polyacrylamide gel indicated two bands of about 342 and 303 kDa, which are little less than the 369 kDa deduced from the smallest transcript based on the results of Northern blotting. However, size estimation on SDS-PAGE gel can be inaccurate, particularly in the case of large hydrophobic molecules, such as spider silk proteins. In addition, partial degradation and cleavage of the analyzed cylindrical luminal contents during solubilization in sample loading buffer could not be excluded. Size assessments based on gene analysis are more accurate, and the cylindrical silk protein of *N. clavata* is larger or similar to several other silk proteins from spiders examined to date. Sizes of about 500, 420, 320, and 250 kDa were computed for Flag (*10*), MaSp1 (*7*), MiSp1 (*9*), and MiSp2 (*9*), respectively. The diversity of these silk protein sizes is due to differences in the large central protein region composed of repeats, while the terminal non-repetitive sequences have little effect on the sizes of these silk proteins.

In addition, the cylindrical silk protein apparently exhibits two protein masses and three large transcripts, as suggested by the results of SDS-PAGE and Northern blotting. This differs from the results of SDS-PAGE analysis of the luminal protein of *N. clavipes* by Candelas *et al*. (*6*) and the egg case silk protein of *L. Hesperus* by Hu *et al*. (*13*), *i.e.*, only one protein mass of approximately 380 and 100 kDa, respectively. Finally, compared with the egg case silk proteins from *L. Hesperus* named ECP-1 (*13*), which do not have several common characteristics of published traditional silk proteins, CySp1 is derived from a much larger transcript and has a similar overall structure to silk proteins characterized to date. These differences may be due to the different source species.

Organization of the Repetitive Cylindrical Silk Protein Core—The most abundant amino acids in the repetitive region of the cylindrical silk protein of *N. clavata* are Ala and Ser. The striking conservation in the A_1 and B_1 repeats, and the regularity in their alternating arrangement are characteristic features of the cylindrical silk protein of *N. clavata*. The basic repeats A_1 and B_1 are longer and more complex than the elementary repetitive units of other spider and lepidopteran silks. The central region of the protein is composed of approximately 24 tandemly arranged assembly repeats (A_1B_1). The number of assembly repeats is about twice that of Flag of spiders and H-fibroin of Lepidoptera, which have 13 and 12 assembled repeats, respectively (*33*, *34*). Furthermore, the sizes of the large assembled repeats of Flag and H-fibroin vary due to the irregular numbers of basic repeat units they contain.

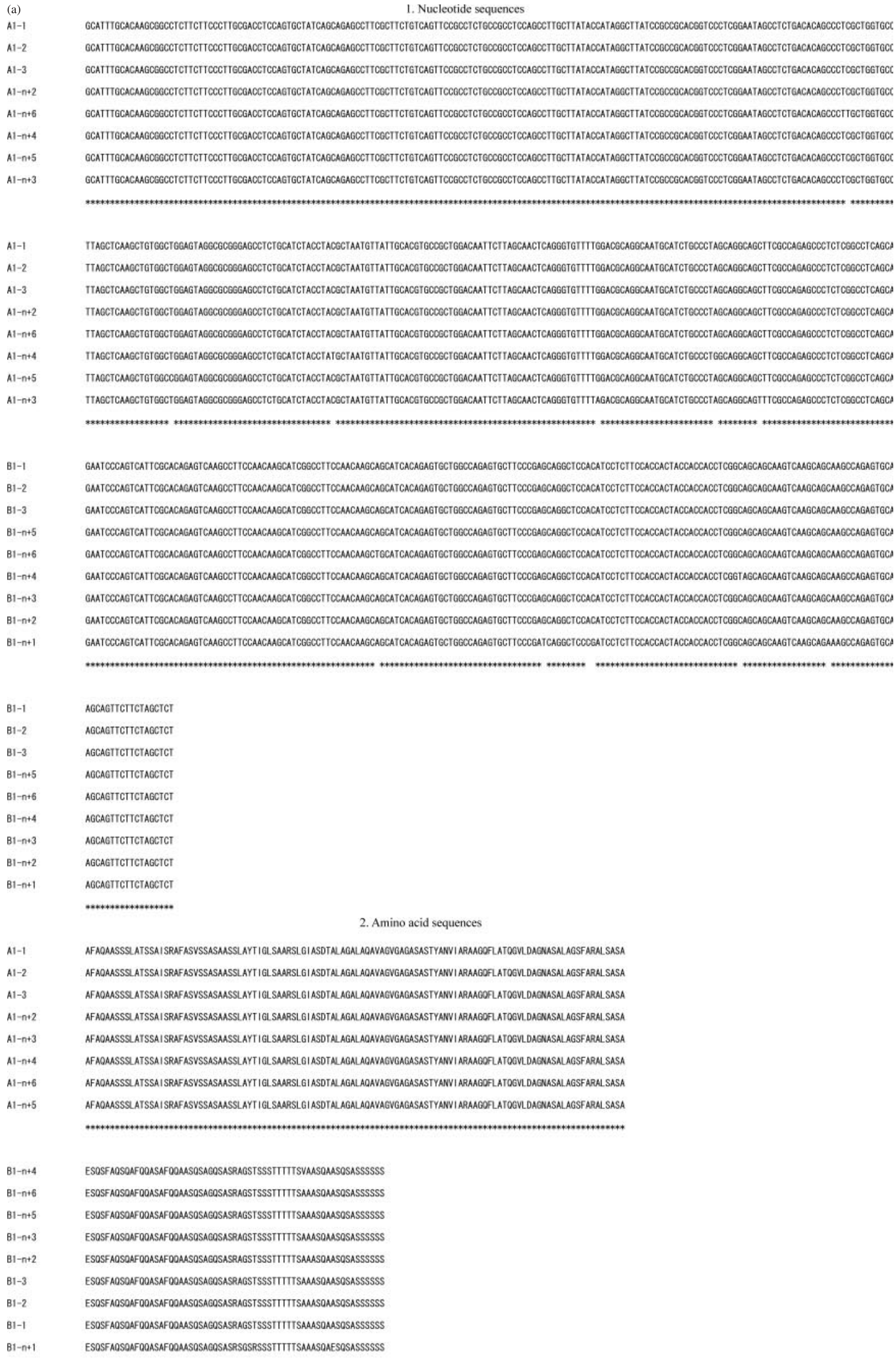

Fig. 6. **Continued.**

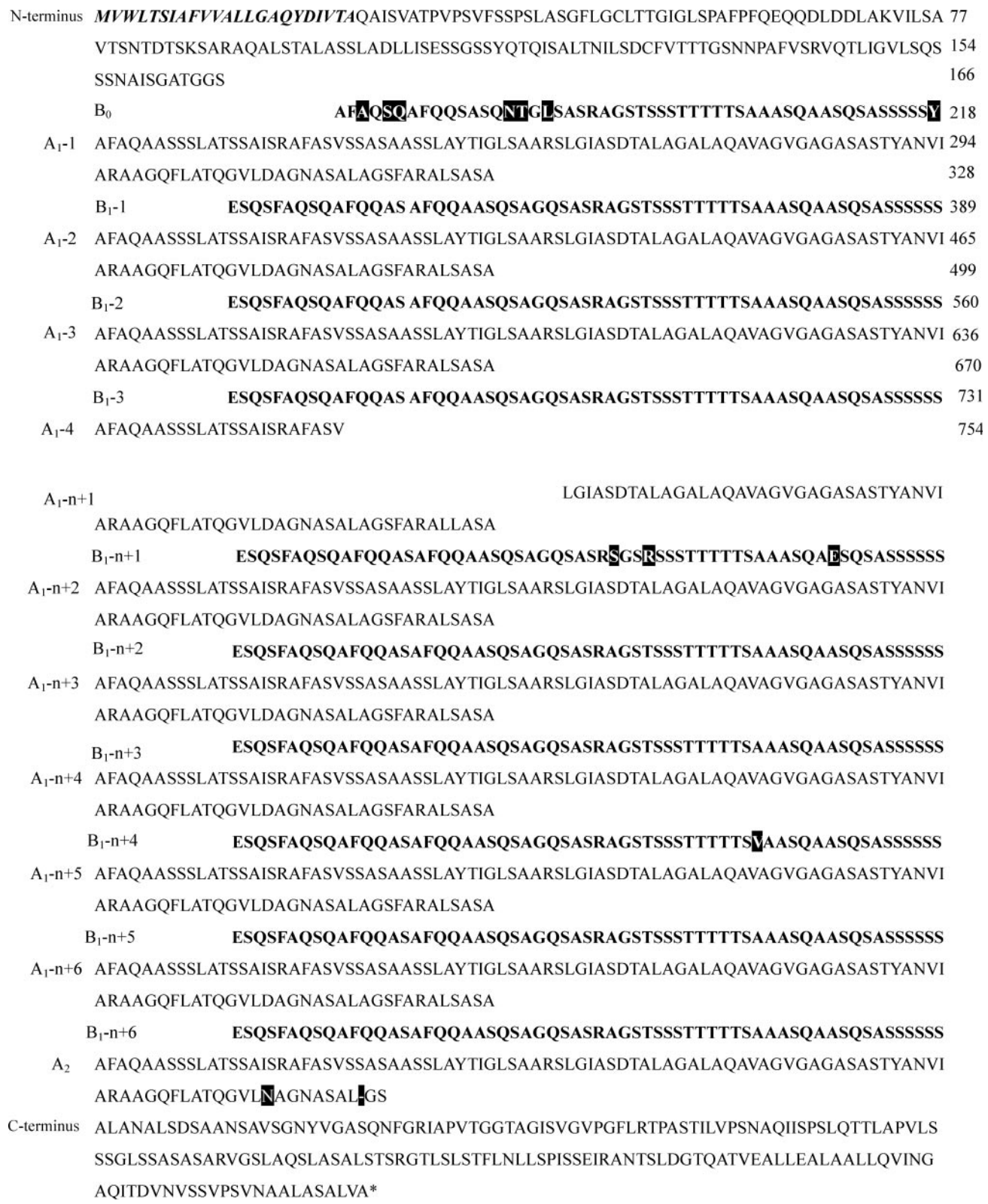

Fig. 6. **Cylindrical silk protein structure.** a, CLUSTAL W alignment of nucleotide (a-1) and amino acid (a-2) sequences of cylindrical fibroin gene repeats. b, amino acid sequences deduced from the gene sequences of Cy5′ and Cy3′. The amino acid sequences consist of a non-repetitive N-terminus (166 residues numbered from the translation start, with the presumed signal peptide in italics), repetitive central parts, and non-repetitive C-terminus. The repetitive central region is arranged in the repeats B_0 (bold), A_1, B_1 (bold), and A_2, which are numbered according to their positions in the Cy5′ sequence (numbering begins with 1) or in the Cy3′ sequence (numbering starts with $n + 1$). Amino acid residue substitutions in the central repeat region are highlighted in black block. * indicates a stop codon.

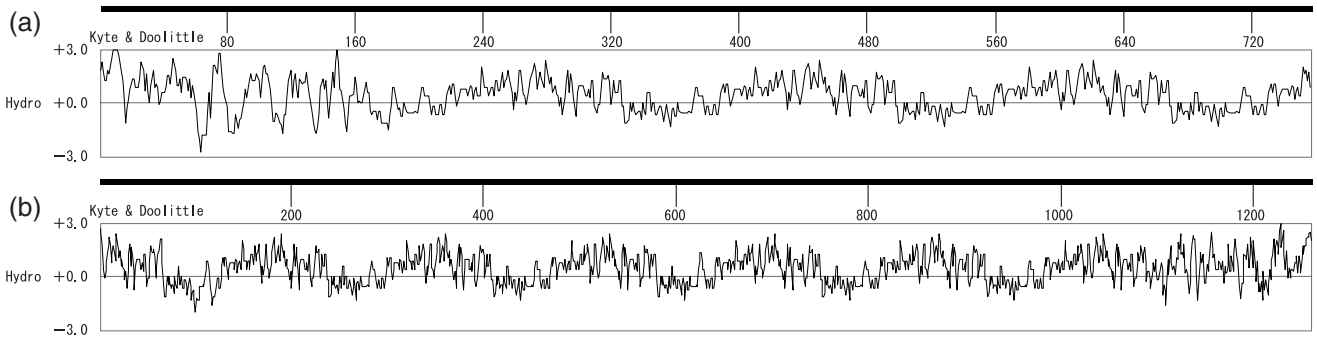

Fig. 7. **Kyte-Doolittle hydropathy plots of amino acid sequences of cylindrical silk proteins.** The numbers on the axis above the plots indicate amino acid positions. The peaks above the abscissa indicate hydrophobicity, while the scores below the abscissa indicate hydrohilicity. a, amino acid sequences deduced from Cy5′. b, amino acid sequences deduced from Cy3′.

(a)

```
N.c MaSp2     SRLASPDSGARVASAVSNLVSSGPTSSAALSS----VISNAVSQIGASNPGLSGCDVLIQ
N.s MaSp2     SRLASPDSGAXVASAVSNLVSSGPTSSAALSS----VIXNAVSQIGASNPGLSGCDVLIX
N.c MaSp1     SRLSSPQASSRVSSAVSNLVASGPTNSAALSS----TISNVVSQIGASNPGLSGCDVLIQ
N.s MaSp1     SRLSSPEASSRVSSAVSNLVSSGPTNSAALSS----TISNVVSQIGASNPGLSGCDVLIQ
N.c MiSp1     SRLSSAEASSRISSAASTLVSGGYLNTAALPS----VISDLFAQVGASSPGVSDSEVLIQ
N.c MiSp2     SRLSSAEACSRISAAASTLVSG-SLNTAALPS----VISDLFAQVSASSPGVSGNEVLIQ
N.c Flag      -YYPSSRVPDMVNGIMSAMQGS-GFNYQMFGN----MLSQYSSGSGTCNP--NNVNVLMD
N.m Flag      -YYPSSRVPDMVNGIMSAMQGS-GFNYQMFGN----MLSQYSSGSGSCNP--NNVNVLMD
N.c CySp1     SGLSSASASARVGSLAQSLASALSTSRGTLSLSTFLNLLSPISSEIRANTSLDGTQATVE
                .*.     :. . :..  .   :     :. :   ...  .. :. :

N.c Masp2     ALLEIVSACVTILSSSSIGQVNYGAASQFAQVVGQSVLSAF-
N.s MaSp2     ALLEIVSACVTILSSSSIGQVNYGAA----------------
N.c Masp1     ALLEVVSALIQILGSSSIGQVNYGSAGQATQIVGQSVYQALG
N.s MaSp1     ALLEVVSALVHILGSSSIGQVNYGSAGQATQ-----------
N.c MiSp1     VLLEIVSSLIHILSSSSVGQVDFSSVGSSAAAVGQSMQVVMG
N.c MiSp2     VLLEIVSSLIHILSSSSVGQVDFSSVGSSAAAVGQSMQVVMG
N.c Flag      ALLAALHCLSNHGSSSFAPSPTPAAMSAYSNSVGRMFAY---
N.m Flag      ALLAALHCLSNHGSSSFAPSPTPAAMSAYSNSVGRMFAY---
N.c CySp1     ALLEALAALLQVINGAQITDVNVSSVPSVNAALASALVA---
                .**  :.      ..:   .  .:
```

(b)

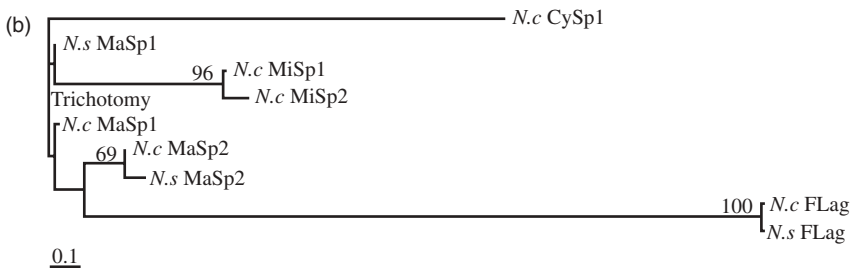

Fig. 8. **Comparative analyses of the C-terminal regions of several spider silk fibroins from the *Nephila* genus**. (a) CLUSTAL W alignment was performed with ClustalW (version 1.83; default setting) at the DDBJ site. At the bottom of the alignment, an asterisk represents an identical amino acid, one dot (.) represents conserved substitutions, and two dots(:) represent semi-conserved substitutions. (b) Neighbor-joining (NJ) tree for the aligned C-terminal amino acid sequences. The scale bar equals 0.1 and is a unit of the distance of branch lengths. Bootstrap percentages of greater than 50% are shown above internodes. Naming adopted in this Figure: first two letters represent the initials of the genus and species. The nomenclature refers to the gland followed by Sp (Ma = major ampullate; Mi = minor ampullate; Flag = flagelliform; Cy = cylindrical; Sp = spidroin or silk protein). Genebank accession numbers: *N.c* MaSp1, AAA29380; *N.c* MaSp2, AAA29381; *N.c* MiSp1, AAC14589; *N.c* MiSp2, AAC14590; *N.c* Flag, AAC38847; *N.m* Flag, AAF36092; *N.s* MaSp1, AAK30608; *N.s* MaSp2, AAK30609. *N.c* is an abbreviation of *Nephila clavipes,* except for *N.c* CySp1, which is an abbreviation of *Nephila clavata. N.m* = *Nephila madagascariensis*; *N.s* = *Nephila senegalensis*.

Maintenance of Uniform Repeats—All four types of repeat in the *N. clavata* cylindrical silk protein are remarkably conserved throughout the sequenced part of the gene. Their uniformity at the DNA level is strengthened by the preferential use of certain isocodons (Table 2). The high content of codons for Ala (rich in G and C at the first two positions) is compensated for by the high frequency of isocodons ending with T and A. Similar compensation occurs in various DNA sequences (*35*). In the cylindrical silk protein genes, however, the rate of recurrence of T or A at the third position is actually not random, and the frequency of isocodons ending with T and A is determined by the required degree of compensation. There is a striking abundance of G/C at the first two codon positions because

of the prevalence of encoding Gly and Ala or Pro residues in several other spider silks (Fig. 9). Compensating for this skewed nucleotide composition is a strong preference for A/T at the third positions. The Gly content is significantly decreased in the cylindrical silk protein, but Ala remains the most abundant component. In order to compensate for this Ala-skewed nucleotide composition, there is an apparent preference for A/T at the third positions of codons encoding Ala and Gln, and a small degree of preference for A/T at the third positions of codons encoding in the next most abundant amino acid residue, Ser (Table 2). The G/C content of the repeat region of these spider silk genes remains at around 60%. The shared base composition patterns among spider silk protein genes indicates that the codon choice pattern in silk proteins is perhaps for the most stable conformation of chromatin or encoded mRNA (*36*). We propose that this codon bias may also be related to the stability of reiterated motifs. The concatenations of short motifs, such as the microsatellites, are prone to replication slippage, unequal sister chromatid exchange and unequal allelic recombination that generate variations in the length of concatenations (*37*). Allelic divergence in spider silk genes (*38*) and the length variation of *Bombyx mori* H-fibroin (*39*) show the importance of this mechanism in silk genes with short repetitive motifs. By contrast, larger DNA blocks without intrinsic repeats are less likely to be internally misaligned during replication and crossing over. We presume that the lack of short repetitive motifs within the large A_1, B_1, A_2 and B_0 DNA blocks is a prerequisite for maintaining length. The unique parts of some other silk genes, for example, introns and "spacers" in the repetitive coding regions of the spider flagelliform silk gene (*33*) and amorphous linkers joining the repetitive domains of *B. mori* H-fibroin (*34*), are also characterized by length conservation.

Table 2. **Codon usage for the most abundant amino acids in assembled repeats (A1B1) of the cylindrical silk protein.**

Residue	Occurrence/percentage	Codon/pencentage
Ala	52/30.41%	gcg/5.77%
		gcc/32.69%
		gct/21.53%
		gca/40.38%
Ser	43/25.15%	agc/13.95%
		agt/18.60%
		tcg/7%
		tcc/25.58%
		tct/7%
		tca/27.91%
Gln	15/8.77%	cag/40%
		caa/60%

The crossover and replication slippage between the repetitive DNA blocks drive their concerted evolution and lead to DNA block homogenization (*40*, *41*). Homogenization of DNA repeats has been reported for microsatellites, repetitive regions of mitochondrial DNA, tandem gene arrays and repetitive sequences within large genes. The process of sequence homogenization constrains but facilitates the rapid spreading of base changes among the repeats. The distribution of mutative codons is initially stochastic, but the function of the encoded protein imposes constraints on their maintenance and propagation (*42*). Mutations improving function are favored by natural selection, which fosters the spreading of mutative nucleotides to other repeats.

Molecular Conformation of the Cylindrical Silk Protein—The repetitive nature of the cylindrical silk sequences with alternating hydrophilic and hydrophobic blocks can be visualized, as seen in other silk sequences, using hydrophathy plots, but the size of the internal hydrophilic blocks within the central core repetitive region of the silk sequences (up to 61 residues) is much larger than that of other silk sequences (the size of hydrophilic blocks varies from 6 to 33; the ratio of size of hydrophilic blocks/size of hydrophobic blocks is less than 1/10) (*31*). The size, number and hydrophilicity of the internal hydrophilic blocks may correlate with the microenvironment in which the fibrous silk is designed to operate (*31*).

The overall structure of the cylindrical silk protein is similar to that of characterized silk proteins to date, and includes a large core repetitive region, a high content of specific amino acid residues, and alternating hydrophobic

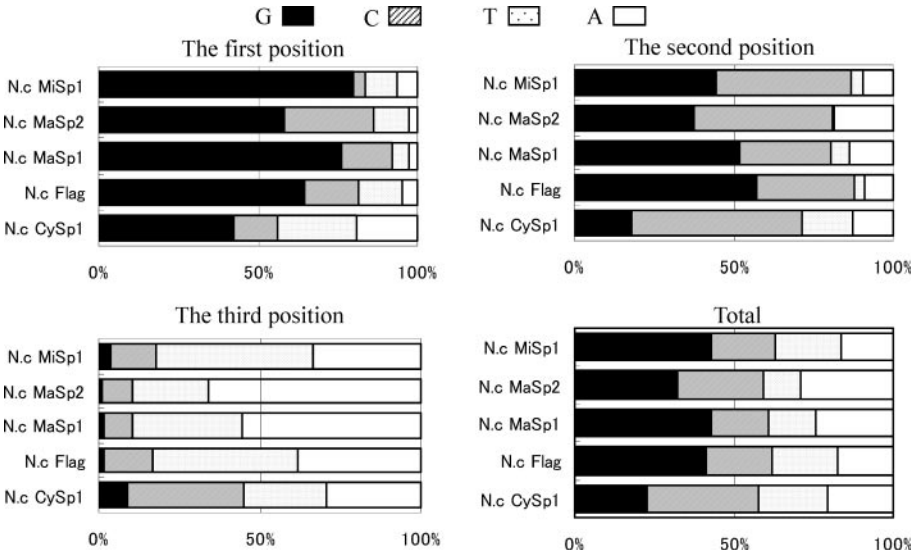

Fig. 9. **Nucleotide base compositions (%) of the repetitive regions of several spider silk genes, and the assembled repeats of Cy5′ or Cy3′ were calculated based on codon position.** The base is indicated as a single letter abbreviation, and different bases are shown using different patterns in the graph, as indicated above the graph. Genebank accession numbers: *N.c* MaSp1, AAA29380; *N.c* MaSp2, AAA29381; *N.c* MiSp1, AAC14589; *N.c* Flag, AAC38847. N.c is an abbreviation of *Nephila clavipes,* except for *N.c* CySp1, which is an abbreviation of *Nephila clavata.*

and hydrophilic blocks. However, the cylindrical silk protein has a unique molecular architecture, such as the lack of small iterated motifs confirmed in other silk proteins and the invariable size of assembled repeats, when compared with other spider and silkworm silk proteins. The unique molecular architecture could be the result of adaption for the function of the egg case, the silk for which differs from all the other silks used daily for making webs and preying on other organisms, in that it only serves to protect the enclosed offspring during the early stages of their lives. It will provide a new biomaterial of medical applications and a scaffold for tissue engineering because it is characterized by good air permeability, heat preservation and moisture retention.

The nucleotide sequence data reported are available in the DDBJ databases under accession numbers AB218973 for Cy3′ and AB218974 for Cy5′.

REFERENCES

1. Lucas, F. (1964) Spiders and their silk. *Discovery* **25**, 20–26
2. Gosline, J.M., DeMont, M.E., and Denny, M.W. (1986) The structure and properties of spider silk. *Endeavor* **10**, 37–43
3. Candelas, G.C., Candelas, T., Ortiz, A., and Rodriguez, O. (1983) Translational pauses during a spider fibroin synthesis. *Biochem. Biophys. Res. Commun.* **116**, 1033–1038
4. Candelas, G.C. and Cintron, J. (1981) A spider fibroin and its synthesis. *J. Exp. Zool.* **216**, 1–6
5. Candelas, G.C. and Lopez, F. (1983) Synthesis of fibroin in the cultured glands of *Nephila clavipes*. *Comp. Biochem. Physiol.* **74B**, 637–641
6. Candelas, G.C., Ortiz, A., and Molina, C. (1986) The cylindrical or tubiliform glands of *Nephila clavipes*. *J. Exp. Zool.* **237**, 281–285
7. Xu, M. and Lewis, R. (1990) Structure of a protein superfiber: spider dragline silk. *Proc. Natl Acad. Sci.USA* **87**, 7120–7124
8. Hinman, M. and Lewis, R. (1992) Isolation of a clone encoding a second dragline silk fibroin. *J. Biol. Chem.* **267**, 19320–19324
9. Colgin, M. and Lewis, R. (1998) Spider minor ampullte silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions". *Protein Sci.* **7**, 667–672
10. Hayashi, C.Y. and Lewis, R.V. (1998) Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J. Mol. Biol.* **275**, 773–784
11. Hayashi, C.Y., Blackledge, T.A., and Lewis, R.V. (2004) Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol Biol Evol.* **21**, 1950–1959
12. Guerette, P., Ginzinger, D., Weber, B., and Gosline, J. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* **272**, 112–115
13. Hu, X., Kohler, K., Falick, A.M., Moore, A.M., Jones, P.R., Sparkman, O.D., and Vierra, C. (2005) Egg case protein-1. A new class of silk proteins with fibroin-like properties from the spider *Latrodectus hesperus*. *J. Biol. Chem.* **280**, 21220–21230
14. Braunitzer, V.G. and Wolff, D. (1955) Vergleichende chemische unterschungen uber die fibroin von *Bombyx mori* and *Nephila madagascariensis*. *Z. Naturforsch.* **10b**, 404–412
15. Peters, V.H.M. (1955) Uber der spinneraparat von *Nephila madagascariensis*. *Z. Naturforsch.* **10b**, 395–404
16. Comstock, J.H. (1948) *The Spider Book*, Ithaca, New York
17. Austin, A.D. (1985) The function of spider egg sacs in relation to parasitoids and predators, with special reference to the Australian fauna. *J. Nat. His.* **19**, 359–376
18. Hieber, C. (1985) The "insulation" layer in the cocoons of *Argiope aurantia* (Araneae: Araneidae). *J. Therm. Biol.* **10**, 171–175
19. Hieber, C. (1992) Spider cocoons and their suspension systems as barriers to generalist and specialist predators. *Oecologia* **91**, 530–535
20. Andersen, S. (1970) Amino acid composition of spider silks. *Comp. Biochem. Physiol.* **35**, 705–711
21. Peakall, D.B. (1964) Composition, function and glandular origin of the silk fibroins of the spider *Araneus diadematus* CL. *J. Exp. Zool.* **156**, 345–352
22. Casem, M.L., Turner, D., and Houchin, K. (1999) Protein and amino acid composition of silks from the cob weaver, *Latrodectus hesperus* (black widow). *Int. J. Biol. Macromol.* **24**, 103–108
23. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecule Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York
24. Foradori, M.J., Kovoor, J., Moon, M.J., and Tillinghast, E.K. (2002) Relation between the outer cover of the egg case of *Argiope aurantia* (Araneae: Araneidae) and the emergence of its spiderlings. *J. Morphol.* **252**, 218–226
25. Dicko C., Knight D., Kenney J.M., and Vollrath F. (2004) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. *Biomacromolecules* **5**, 2105–2115
26. Weber, K., Pringle, J.R., and Osborn, M. (1972) Measurement of molecular weights by electrophoresis on SDS-acrylamide gel. *Methods Enzymol.* **26 PtC**, 3–27
27. Von Heijne, G. (1985) Signal sequences: The limit of variation. *J. Mol. Biol.* **184**, 99–105
28. Claros, M.G., Brunak, S., and Von Heijne, G. (1997) Prediction of N-terminal protein sorting signals. *Curr. Opin. Struct. Biol.* **7**, 394–398
29. Bendtsen, J.D., Nielsen, H., Heijne, G.V., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/). *J. Mol. Biol.* **340**, 783–795.
30. Arcidiacono, S., Mello, C., Kaplan, D., Cheley, S., and Bayley, H. (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **49**, 31–38
31. Bini, E., Knight, D.P., and Kaplan, D.L. (2004) Mapping domain structures in silks from insects and spiders related to protein assembly. *J. Mol. Biol.* **335**, 27–40
32. Proudfoot, N.J. and Brownlee, G.G. (1976) 3′ non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211–214
33. Hayashi, C.Y. and Lewis, R.V. (2000) Molecular architecture and evolution of a modular spider silk protein gene. *Science* **287**, 1477–1479
34. Zhou, C.Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.G., and Janin, J. (2001) Silk fibroin: structural implications of a remarkable amino acid sequence. *Proteins* **44**, 119–122
35. Nakamura, T., Suyama, A., and Wada,A. (1991) Two types of linkage between codon usage and gene-expression levels. *FEBS Lett.* **289**, 123–125
36. Mita, K., Ichimura, S., Zama, M., and James, T.C. (1988) Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA. *J. Mol. Biol.* **203**, 917–925
37. Jeffreys, A.J., Monckton, D.G., Tamaki, K., Neil, D.L., Armour, J.A., MacLeod, A., Collick, A., Allen, M., and Jobling, M. (1993) Minisatellite variant repeat mapping: application to DNA typing and mutation analysis. *EXS* **67**, 125–139

38. Beckwitt, R., Arcidiacono, S., and Stote, R. (1998) Evolution of repetitive proteins: spider silks from Nephila clavipes (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *Insect Biochem. Mol. Biol.* **28**, 121–130
39. Manning, R.F. and Gage, L.P. (1980) Internal structure of the silk fibroin gene of *Bombyx mori*. II. Remarkable polymorphism of the organization of crystalline and amorphous coding sequences. *J. Biol. Chem.* **255**, 9451–9457
40. Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* **191**, 528–535
41. Dover, G. (1982) Molecular drive: a cohesive mode of species evolution. *Nature* **299**, 111–117
42. Meeds, T., Lockard, E., and Livingston, B.T. (2001) Special evolutionary properties of genes encoding a protein with a simple amino acid repeat. *J. Mol. Evol.* **53**, 180–190