SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION (U)	1b RESTRICTIVE MARKINGS NA
2a SECURITY CLASSIFICATION AUTHORITY NA	3 DISTRIBUTION/AVAILABILITY OF REPORT Unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NA	5 MONITORING ORGANIZATION REPORT NUMBER(S) NA
6a NAME OF PERFORMING ORGANIZATION University of Wyoming	6b OFFICE SYMBOL (If applicable) NA
7a NAME OF MONITORING ORGANIZATION Office of Naval Research	
6c ADDRESS (City, State, and ZIP Code) Box 3944, University Station Laramie, WY 82071	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research	8b OFFICE SYMBOL (If applicable) ONR
9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-89-J-1564	
8c ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO N00014	WORK UNIT ACCESSION NO
61153N	PR04106	89-J-1564	NA

11 TITLE (Include Security Classification)

Cloning and Structure of Different Types of Spider Silk

12 PERSONAL AUTHOR(S)
Randolph V. Lewis

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 12/1/88 TO 11/10/91	11/10/91	7

16 SUPPLEMENTARY NOTATION
NA

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
06	03		protein, silk, elasticity

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Based on both physical and chemical properties it is clear that spider silks are very different than silk worm silk. In particular, the amino acid composition of the various spider silks are different with the variations being in Glx and Pro which are much higher than in worm silk. Studies by others using X-ray diffraction and by us with FTIR demonstrate that the dragline silk is predominantly β-sheet. In the past contract period we constructed a cDNA library in a plasmid vector from the silk glands (dragline) of one species of spider. We isolated several clones which are positive in hybridizing to our DNA probes based on the protein sequence. We sequenced two clones which encoded different proteins both of which are components of major ampullate silk. We have obtained clones which appear to encode proteins for other spider silks as well.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	(U)

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
Dr. M. Marron	(202)696-4760	ONR

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

FINAL REPORT on N00014-89-J-1564 for December 1, 1990

PRINCIPAL INVESTIGATOR: Randolph V. Lewis
CONTRACTOR: University of Wyoming
CONTRACT TITLE: Cloning and Structure of Spider Silk
START DATE: December 1, 1988

A-1

RESEARCH OBJECTIVES: Clone, sequence and express dragline silk protein from Nephila Clavipes and compare the sequence to clones of the same protein from Areneus gemmoides.

PROGRESS:

The results described here were accomplished with 2 graduate students(1 PhD and 1 MS), 1 postdoctoral fellow and 1 Research Asst.

To describe the recent results briefly we have: 1) published two papers describing the results of our sequencing of the major ampullate silk and the biophysical studies on the silk fibers and peptides(copies in the Appendix), 2) demonstrated expression of the major ampullate silk protein in vivo and in vitro, 3) identified a second major ampullate silk protein, cloned and have sequence data for a partial clone(it is significantly different in much of the sequence), 4)begun screening silk gland cDNA libraries for Areneus silk proteins, 5) developed methods to produce multimers of our repeat peptides and 6) obtained excellent 1-D NMR data and have started the 2-D work.

Pure silk. The first requirement for these studies was to obtain sufficient quantities of a single pure type of silk. Spiders were obtained from Florida(Nephila clavipes) or collected locally(Areneus gemmoides). Silk was obtained following a modified procedure of forced silking described by Work(12) for the major ampullate silk. It basically immobilizes the spider with carbon dioxide. A single silk strand is attached to a spool on a mechanical device for winding, in our case a variable speed drill. Using our procedure we could collect both major and minor ampullate silk simultaneously.

Amino acid analyses were done on each of the silks and on each different batch of silk. The compositions were similar in every case to those published for that species(Nephila) or for a closely related species(Areneus).

Peptides for sequencing. One of the keys to our success in these studies compared to other groups who have worked on it was the ability to obtain peptides for sequencing. Initially we attempted to directly sequence the silk itself to no avail. The N-terminus was clearly "ragged" and, in fact, the PTH analysis at each residue looked more like the amino acid composition of the silk. We used nearly every protease commercially available to try to cleave the silk fibers but without success. This is probably due to its insolubility and lack of accessibility by the enzymes. Various chemical methods were

tried as well but none were successful enough to be useful.

We then resorted to partial acid hydrolysis as was used in the "good old days" of protein sequencing. Even this was less than an easy route. Apparently due to the repetitive nature of the silk protein there is only a very small variation in the peptide bond strengths. Thus the protein goes from an insoluble mass to hydrolyzed amino acids in the space of 2-3 min. at 125°C in 6N HCl. We tried other temperatures, concentrations and acids but no real advantages were seen over this system.

For those silks which can be dissolved in concentrated formic acid we found recently that cleavage with N-bromosuccinimide will work. A four to one ratio of NBS to Tyr at elevated temperatures(45-65°C) overnight is used. Due to the absence of Trp residues cleavage only occurs at Tyr. Even this method is highly variable but we do get substantial cleavage to yield peptides which can be isolated.

Peptides were separated by reverse phase HPLC on a C18 column. They all eluted early in the gradient and thus a very shallow gradient was used for the next steps which were a repeat run on a different manufacturer's column which we knew to be more retentive and then on a diphenyl column.

Peptide Sequences: The sequences of the peptides we obtained are shown below. These are the longest ones we found. In addition to these we found several shorter peptides but their sequences were all contained in the peptides shown here except for an AAA and an AAAA peptide. The similarity between the sequence for Areneus and for Nephila (from our DNA sequence shown later) indicates that the two major ampullate silks share substantial homology. It should also be noted that the minor ampullate silk also resembles major ampullate silk but no poly-Ala stretches were detected in the cleaved peptides.

Nephila major ampullate	Areneus major ampullate
GYGPG	GPYGPGQQGP
GQGAG	GAGAAAAAA
GAGQG	
GYGGLG	
Nephila minor ampullate	
GGAGQGG(Y)	

Protein size. Due to the difficulties in ascertaining the molecular weight of the silk protein we have developed methods to block translation of the specific mRNA using our probes in an in vitro translation assay. This allows us to detect small amounts of protein in the SDS gel and determine which ones are missing when the complementary DNA is present. Based on very preliminary data the major ampullate silk shows two bands missing at 75 and 105 kdal.

Cloning and sequencing of Nephila dragline silk protein I. (see PNAS paper in

Appendix)

The dragline silk protein I has predominantly a G-G-X(X=Q,Y,L and A) repeating unit with polyalanine stretches interrupting them. The overall repeat is 34 aa maximum. However, the repeats are not nearly as exact as those found in silkworm silk(13). There seem to be three segments to the repeating unit. The first, comprising the initial 9 amino acids, is variable to the point that it is eliminated in some repeats. However, the sequence involves only the same three possible triplets and missing segments are all triplets. The second region is the poly Ala region starting with the G at amino acid 10 through the A at 19. This region is highly conserved and the two differences in this region need to be checked by sequencing again. The last segment is the most highly conserved of the three averaging 92% identity for the segments shown. It is also clear from this data that a simple interpretation of the secondary and tertiary structure of these sequences is not possible. The secondary structure predictions are shown in Fig. 1 in the Appendix(note the differences between the different methods). However, we have some indications on this from the peptide work we are doing which is described later.

A short word on codon usage in this sequence is needed(see PNAS paper). Probably not surprisingly C and G are rarely used in the third position. In fact, G is virtually never used and C less than 10% of the time. Thus, there is a very strong codon bias which was not observed to this extent in the limited silkworm silk sequence published to date.

Studies of the gene suggest a single copy gene which is repetitive for nearly its entire length as the only restriction sites found with 12 different 4 cutters were those present in the cDNA clone we have. The data also strongly suggests there are no introns present which is in great contrast to the genes for the collagen proteins.

Sequencing the clones has been a major problem due to two factors. The first is the repetitive nature of the sequence which requires significant regions of overlap to align the sequence fragments and the second is that the clones are unstable in the different plasmid vectors we have used. "Sure" cells help as does a new broth we have developed but do not completely eliminate the problem. As a result of these problems sequencing progress is much slower than with the standard proteins. However, we completed the sequence of one strand of the Protein 2 in less than 2 months. So the modifications do help substantially.

Status of cloning other silk proteins. We have recently cloned a second dragline silk protein. As can be seen from the peptides we isolated(above) there is a proline containing peptide which was not present in the clone we have. When the data from the gene indicated the repeat was the same for the complete length we used a DNA probe for that peptide and isolated a different set of clones. We have completed the sequence of one strand of the complete clone. The consensus repeat is G-P-G-G-Y-G-P-G-Q-Q repeated 2 or three times followed by G-P-Y-G-P-G and 7-10 alanine residues with an occasional substitution of serine. The sequence and secondary structure predictions are shown in Fig. 2 in the Appendix. This repeat resembles an unusual wheat storage

protein believed responsible for the elasticity of dough. The repeats of that protein are P-G-Q-G-Q-Q and it is believed to be a series of linked reverse turns(20).

Cloning of the Areneus dragline silk protein was at the sequencing stage previously. We had identified clones with inserts hybridizing to our probe based on the protein sequence presented above. One insert is 4.5 kb, also in the pBluescript vector. Unfortunately this clone did not have the correct protein sequence. It was from actin and the probe sequence was found. We have constructed a new cDNA library from the major ampullate glands and are in the process of screening that library. We are using probes based on the peptide sequence we obtained from the NBS cleavage which showed the presence of the poly-Ala region and on codon usage or rather the lack of codons with G or C in the third position.

Protein expression in bacteria. We have moved our largest insert(2.5 kb) of the Nephila dragline silk protein clones from pBluescript into several different expression vectors. We have sequenced these to determine which ones are in the correct reading frame. We have been transforming E. Coli and assessing the levels of protein production. Production has been low as determined by SDS PAGE and silver staining. However, we could detect a new protein when IPTG is included. Recent changes in the vector and conditions have greatly improved the level of production and we expect to obtain sufficient protein for chemical characterization in the next few weeks.

Clearly we may need to increase the level of expression and thus are turning to baculovirus expression. This methodology is already in use in one laboratory in the building and thus the expertise is available to assist us. It would be ironic to use this approach to produce spider silk in silkworms. The various proteins will be studied using the methodologies described below. Ultimately we will design various synthetic proteins to test our hypotheses concerning the mechanical and biophysical properties seen in the native proteins.

Biophysical studies. A description of our previous work is presented in the Appendix in the ABB paper. The use of polarized IR to study silk has an historical precedent(21) and FTIR via microscope is now a well accepted approach for synthetic fibers(22). The bottom line is that using FTIR we can detect protein conformational changes when Nephila and Areneus dragline silk is stretched. The effect is proportional to the weight applied. The changes do not affect the ß-sheet region but appear to be a shift to a helical conformation in the unstructured regions of the protein. These changes are reversible because when the weight is removed the original FTIR pattern is seen. No such change is seen in the minor ampullate silk which possesses very little elasticity. As mentioned in the paper it is possible that the helical conformation seen is due to alignment of previously existing helices but we have been unable to detect them in silk which is not under tension by either polarized or non-polarized light in the FTIR.

The CD spectral work has not been published yet as we are waiting to include the new peptide from silk protein 2. The data are important as they suggest we have a method to study the formation of ß-sheets and turns and then aggregation at high protein concentrations. This will verify the data we get from the peptide studies.

Historical attempts to use X-ray diffraction to elucidate more about the protein structure of these spider silks have been unsuccessful to date owing to the narrow fibers and the inability to focus the beam tightly enough. We are now collaborating with the crystallography group at Smith Kline Beecham to obtain the crystal structures of our long peptides and with Eric Eickenberry(R.W. J. Med. School) to use fiber X-ray diffraction utilizing the latest equipment which he believes will be successful.

Peptide synthesis and characterization This is a major new approach we have developed since the first proposal was submitted. We synthesized two peptides based on the complete silk protein 1 thirty amino acid consensus repeat(P30) and the 15 amino acid highly conserved third segment of the repeat(P15). These two peptides were made to test the effect of the poly-Ala region and whether it was directly involved in the elastic properties of the silk protein. Both were synthesized by the Fmoc method and purified by gel filtration after cleavage from the resin. HPLC was then used for final purification. Amino acid analysis was done on the purified peptide and confirmed its identity. The peptides are both freely soluble in water up to 5 mg/mL. At 10mg/mL P30 forms a time dependent insoluble precipitate which dissolves only under conditions similar to those used for the silk itself.

The peptides were examined at various concentrations and temperatures by CD(a brief part of these studies is in the ABB paper in the Appendix). At 0.5 mg/mL P30 showed a broad positive peak at 212 nm and a sharp negative one at 202 nm. As the temperature was increased from 5° to 80°C the positive peak decreased and the negative peak red shifted to 204 nm. At concentrations above 2 mg/mL the broad positive peak was at 214 nm and the negative peak at 206 nm. As the temperature was increased it became clear that the positive peak was composed of two peaks with the second at 224 nm. This second peak is likely to be due to the Tyr residues. With further increases in temperature the positive peak disappeared and became a broad negative peak at 212-217 nm. Although the peaks are slightly shifted, the low temperature profile resembles that of a three strand helix(17). The high temperature profile appears to be a multi-strand ß-sheet. The profile of the CD spectra in the intermediate, 20-40°C, range resembles those found for the linked turn wheat seed proteins(23). Discussions with Dr. Robert Woody at Colorado State led to and confirmed our ideas on these data. At both high and low pH values these differing profiles are not seen. Nor are they seen in the presence of trifluoroethanol. The profile for P15 shows a similar temperature dependence except the negative peak in the 206 nm region is missing and the two positive peaks are clearly present. This may be due to the increased tyrosine content of P15. Interestingly a very similar temperature dependent CD spectra was obtained from the silk after dissolving in GnHCl.

We are currently in the process of obtaining 2-D NMR data on these peptides at the various temperatures to confirm our structural model. This is being done in the Chemistry Department here at UW and in collaboration with Dr. Art Pardi at the U. of Colorado. We have also synthesized and purified the 47 aa consensus peptide from Protein 2 and are conducting the same studies described above.

In order to more nearly mimic the protein we are polymerizing the peptides using HBTU and TBTU to study them in a higher molecular weight form. Since there are no reactive functional groups other than the N- and C- termini in these peptides this approach can be used. We will utilize a solid phase approach by attaching our peptide via the C-terminus to the resin and blocking the N-terminus with F-moc. Following deblocking the additional block units will be added sequentially. We can then remove resin at various stages to provide polymers of various sizes. We plan to polymerize the peptide to generate peptides of up to 30 kdal and repeat these various studies. If these are insoluble we will select a size which will remain soluble under reasonable conditions. We will also be studying the insoluble form of the peptide which occurs at high protein concentrations. Attempts will be made to create fibers using methods such as pushing a concentrated peptide solution through a very narrow needle into an acidic solution. We will be using CD, NMR, FTIR and front surface fluorescence with all of these peptides in their various forms. Thus we believe these peptides will provide a key method to understanding the protein structure and structural transitions it undergoes.

Summary. In the past two and a half years we believe we have made major advances in the knowledge of the structure of one major spider silk protein and are in a position in the next three years to make the major contribution in understanding these novel proteins and how they function.

INVENTIONS: We have filed a patent application on both proteins and are awaiting the Patent Office review of it.

PUBLICATIONS:

1. Dong, Lewis, and Middaugh(1991) The Molecular Basis of Spider Silk Elasticity. ABB 284:53-58.

2. Xu and Lewis(1990) Structure of a Protein Superfiber: Spider Dragline Silk. PNAS 87:7120-7124.

3. Kadokami and Lewis(1990) Membrane Bound PCR Nucl. Acid. Res. 18:3082.

TRAINING ACTIVITIES: Two graduate students have been working on this project as well as one undergraduate who is now a graduate student with me. One student finished his PhD this and has a postdoctoral position in Montreal. In addition, two research assistant professors are currently working on it as well.

Women and minorities	1
Non-citizens	2 (both from China)

AWARDS: None.