for Δ_{ST} as a correction to a strict additivity estimate of $\Delta H°_{f,0}$[singlet carbene or biradical].

We make these ionization potential measurements to derive thermochemical data that quantify a conceptual picture of the bonding in radicals, carbenes, and biradicals. Much more than isolated bits of information, the measurements are solid experimental support for the picture in which the triplet state of a carbene or biradical is, for practical purposes, the thermochemical reference molecule with "noninteracting" electrons. Strict bond additivity estimates of ΔH_f[carbene or biradical] give the heat of formation for the triplet carbene or biradical. For singlet ground-state carbenes or biradicals, the singlet-triplet gap is therefore the correction to additivity. It is also the energetic measure of the interaction between the two electrons left by consecutive scission of two bonds in a molecule. We believe that the thermochemical cycles proposed in this picture connect the organic chemists' intuitive notions of extra "effects" (that stabilize or destabilize a molecule relative to some reference state) to a well-defined, measurable property of a molecule.

Conclusions

The ready availability of radicals, carbenes, and biradicals under spectroscopically convenient conditions promises to be a major tool for the further elaboration of models for their thermochemistry and structure. The identification of the triplet carbene or biradical with the "noninteracting" reference state of additivity and the singlet-triplet gap as the correction to additivity estimates of ΔH_f is a simple intuitive picture for which credibility could only have been supplied, in light of the severe assumptions, by rigorous experimental testing. Systematic structural variation and design of organic molecules for which spectroscopic measurements can quantify conceptual models will go a long way toward putting intuitive notions of stability and bonding on firm footing.

Support from the National Science Foundation, the Department of Energy, the donors of the Petroleum Research Fund, administered by the American Chemical Society, the David and Lucile Packard Foundation, the Camille and Henry Dreyfus Foundation, and the Alfred P. Sloan Foundation is acknowledged.

Spider Silk: The Unraveling of a Mystery

RANDOLPH V. LEWIS

Department of Molecular Biology, Box 3944, University of Wyoming, Laramie, Wyoming 82071-3944

Received February 13, 1992

Humans see spiders with both dread and delight. The delight is based on the beauty and precision of the classic spider web. There are few who would disagree that nature's beauty is truly reflected in the glistening morning dew on a spider web. This production of silk by glands located in the abdomen is the key feature making spiders unique creatures. Spiders do not use silk just for a single event in their lifespan as do the silkworm and some other insects. They trail dragline silk behind themselves as a lifeline and use different silks for their egg cocoons and for the sticky threads of their web. Thus silk forms an integral part of the spider's method of living.

Spiders and their webs have undoubtedly been observed for millennia, but the first publications using a scientific approach to spider webs appeared in 1830.[1] Additional publications appeared which focused on the biology of spiders and their webs, but the study of spider silks was not explored. One of the earliest published works on spider silk described some of its physical properties in 1907.[2] Later that same year, Fischer was able to demonstrate the protein nature of spider silk fibers by showing that the predominant substances found in the silk were amino acids.[3] For the most part further studies on the chemical nature of spider silks languished except for brief flurries of activity prior to World War II and during the late 1950s. In comparison to silkworm silk, little additional progress was achieved. Starting in the 1970s, a revived interest in spider silk emerged with several papers from the laboratories of Gosline, Work, and Tillinghast describing detailed physical and mechanical properties with some additional chemical information. Despite this increased effort, the structures of the protein(s) making up the spider silk fibers were unknown.

Table I[a]

material	strength (N m^{-2})	energy to break (J kg^{-1})
dragline silk	1×10^9	1×10^5
KEVLAR	4×10^9	3×10^4
rubber	1×10^6	8×10^4
tendon	1×10^9	5×10^3

[a] Data derived from Gosline et al.[4]

Mechanical Properties of Silk

The major reason for studying spider silk, beyond inherent curiosity, is its unique mechanical properties. In order to survive, the spider must use a minimum

Dr. Lewis was born in 1950 in Powell, Wyoming. He received his B.S. degree in chemistry from the California Institute of Technology in 1968 and his M.S. and Ph.D. degrees in chemistry from the University of California at San Diego in 1974 and 1978. Following a postdoctoral period at the Roche Institute of Molecular Biology (Nutley, NJ), he joined the faculty at the University of Wyoming in 1980. He was Chair of the Molecular Biology Department from 1986 to 1991 and is now Professor of Molecular Biology. In addition to spider silk his laboratory is studying opioid peptide hormones secreted by the adrenal gland.

(1) Blackwell, J. *Zool. J.* **1830**, *5*, 181-188.
(2) Benton, J. R. *Am. J. Sci.* **1907**, *24*, 75-78.
(3) Fischer, E. *Hoppe-Seyler's Z. Physiol. Chem.* **1907**, *53*, 440-450.

0001-4842/92/0125-0392$03.00/0

Table II

silk gland	use	spinneret
major ampullate	dragline, frame threads	anterior
minor ampullate	reinforces dragline	median
piriform	attachment disk	anterior
aciniform	swathing silk	median, posterior
cylindrical (tubuliform)	cocoon silk	median, posterior
aggregate	sticky silk glue	posterior
flagiliform (coronate)	thread for sticky silk	posterior

amount of silk in its web to capture prey, and yet the web must be able to stop and capture an insect flying at high velocity. To do this the web must absorb the energy of the insect without creating rebound which would trampoline the prey off the web. A recent study[4] has concluded that the web and the spider silk used to construct it are nearly optimally designed for each other. The crucial factors in this optimization are tensile strength and elasticity. As shown in Table I, dragline silk is a truly unique biomaterial. Even though it is not as strong as some fibers, it is much more elastic. This allows it to absorb more energy prior to breaking than any commonly used material.

Another fascinating evolutionary adaptation of dragline silk is its ability to supercontract. The fiber will contract to less than 60% of its original length when it is wetted. This results in nearly a 1000-fold decrease in the elastic modulus and an increased extensibility.[5] The practical application for the spider is that the web will tighten each day when it is wetted with dew, thus maintaining its shape and tension. There are several polymers which exhibit supercontraction in organic solvents but virtually none which will supercontract in water alone.[6,7] This supercontraction is reversible and repeatable and can do mechanical work such as lifting a weight. An important feature in conjunction with supercontraction is the insolubility of the silks. Extreme chaotropic agents are required to solubilize any of the silks.

Biological Aspects

In order to fully appreciate the uniqueness of spider silk, it is important to understand the biological aspects of its production. The spider web is constructed of several different silks, each of which is produced in a different gland. The different glands, the silks they produce, and the uses for each are shown in Table II. Although each gland has a distinctive shape and size, they all function in a similar manner. The major ampullate glands have been the most studied of the different silk glands due to their large size and the ease with which the major ampullate (dragline silk) can be obtained. Although most of the current information about the synthesis of the silk proteins is from the study of this gland, morphological and histochemical studies on other glands support the data from the major ampullate gland.

The silk protein(s) is synthesized in specialized columnar epithelial cells located in the distal portion of the gland.[8] There appear to be two different cell types producting silk proteins, which agrees with our biochemical data.[9] The newly synthesized proteins appear as separate droplets inside the cell, which are then secreted into the lumen of the gland. The majority of the gland serves as a storage reservoir for the silk protein in this soluble state. The form of the proteins during storage is not known. It is likely that a combination of the protein's structure and concentration and possibly other chemical factors are responsible for its solubility inside the gland.

It is important to note that these silk fibers are pulled from the spider, not forced out by pressure. The silk fiber is formed as it transits down a tubular duct leading from the gland to the exit valve. The key chemical and physical events which change the soluble protein(s) into a solid fiber take place during this journey. The formation of the fiber can be followed by observing its birefringence, which indicates that the protein molecules are aligning into a regular array.[10] On the basis of these data, the formation of the fiber appears to be nearly complete after traveling two-thirds of the way down the duct. The factors needed to form the fiber are not well understood at this point, but it appears that frictional and mechanical alignment of the protein molecules may be sufficient. In our laboratory fibers could be drawn directly from the lumen of the major, minor, and cylindrical glands by pulling the liquid rapidly with tweezers (unpublished data), suggesting that no chemical alterations are required for fiber formation. Iizuka[11] has proposed a similar mechanism for the formation of silkworm silk fibers.

Chemical and Physical Studies

As mentioned above, the chemical composition of spider silks has been the object of study since the beginning of this century.[3] Except for the capture thread or sticky spiral, the silks are virtually all protein with less than 1% of any other substance. The sticky spiral, in contrast, contains a number of water-soluble compounds which keep the fiber wetted, maintaining its ability to stretch and entangle prey that hit the web.[12] Each different silk has a unique amino acid composition, and the silks from different spider species have some differences in composition as well.

The compositions of different silks from three spider species and silkworm silk are shown in Table III. All of the spider silks show a predominance of glycine, alanine, serine, and glutamine (shown as glutamic acid due to hydrolysis). The first three are the major constituents of silkworm silk and are the three smallest amino acids. There are large amounts of glutamine and proline in many of the silks, which is a clear departure from silkworm silk. This difference indicates that the standard structure for that silk is unlikely to apply to the spider silks. There are also substantial amounts of leucine and tyrosine, which are among the larger amino acids. Of interest as well is the difference in the three species for major ampullate silk in the proline and leucine contents. As will be discussed later, this can be accounted for with our proposed model for this silk. The minor ampullate silks are clearly delineated from

(4) Gosline, J. M.; DeMont, M. E.; Denny, M. W. *Endeavor* **1986**, *10*, 37–43.
(5) Gosline, J. M.; Denny, M. W.; DeMont, M. E. *Nature* **1984**, *309*, 551–552.
(6) Work, R. W. *Trans. Am. Microsc. Soc.* **1977**, *100*, 1–20.
(7) Work, R. W. *J. Exp. Biol.* **1977**, *118*, 379–404.
(8) Bell, A. L.; Peakall, D. B. *J. Cell Biol.* **1969**, *42*, 284–295.
(9) Kovoor, J. *Ann. Sci. Nat., Zool. Biol. Anim.* **1972**, *14*, 1–40.
(10) Work, R. W. *Text. Res. J.* **1977**, *47*, 650–662.
(11) Iiazuka, E. *Experientia* **1983**, *39*, 449–454.
(12) Anderson, C. M.; Tillinghast, E. K. *Physiol. Entomol.* **1980**, *5*, 101–106.

Table III[a]

amino acid	gland (species) major (Ad)	major (Ag)	major (N)	minor (Ad)	minor (Ag)	minor (N)	cylind (Ag)	cylind (N)	acinifrm (Ad)	pyrifrm (Ad)	aggregate (Ad)	coronate (Ad)	silk-worm
Asp	1.0	2.4	1.2	1.9	1.2	2.8	5.9	3.6	8.3	10.5	9.2	2.7	1.3
Thr	0.9	0.8	0.1	1.4	0.3	0.3	3.6	5.2	9.8	4.4	7.6	2.5	0.9
Ser	7.4	10.3	7.4	5.1	5.2	10.9	28.1	23.4	15.7	14.8	6.8	3.1	12.1
Glu	11.5	9.2	10.9	1.6	8.0	4.5	10.0	9.3	7.5	10.4	9.8	2.9	1
Pro	15.8	10.1	3.5	0.1	0.5	0.6	0.4	0.4	3.3	7.8	10.8	20.5	0.4
Gly	37.2	42.0	45.8	42.8	51.6	42.6	9.3	7.3	13.6	7.8	14.5	44.2	44.5
Ala	17.6	18.0	22.2	36.6	24.1	20.0	23.7	30.0	11.0	9.9	6.2	8.3	29.3
Val	1.2	2.2	0.8	1.7	1.2	2.5	4.9	1.8	7.1	5.4	5.8	6.7	2.2
Cys													0.1
Met													0.1
Ile	0.6	0.2	0.1	0.7	0.4	0.4	1.8	1.9	4.3	3.7	4.7	1.0	0.7
Leu	1.3	1.5	3.5	1.0	1.5	2.2	6.6	6.9	9.1	5.4	5.5	1.4	0.5
Tyr	3.9	3.9	4.3	4.7	7.9	8.4	1.5	2.1	1.7	2.2	2.2	2.6	5.2
Phe	0.5	1.0	0.3	0.4	1.0	0.9	3.2	5.2	2.4	2.3	3.8	1.1	0.6
Lys	0.5	0.5	0.3	0.4	0.5	0.5	2.0	1.0	2.2	9.0	7.4	1.4	0.3
His									0.5	2.8	2.4	0.7	0.2
Arg	0.6	1.9	1.7	1.7	2.8	2.9	1.7	3.2	4.0	3.6	3.4	1.1	0.5

[a] The data were derived from Anderson,[25] Work and Young,[26] and Lewis (unpublished). Ad is *Areneus diadematus*, Ag is *Areneus gemmoides*, and N is *nephila clavipes*.

the major ampullate silks by the very low proline content. The cocoon silk from the cylindrical (or tubuliform) glands is radically different from any of the other silks. The glycine content is greatly reduced and serine is increased proportionately. Coronate gland silk is characterized by a high proportion of proline relative to the other amino acids. Our proposed model could account for the high proline content and the accompanying elasticity of this silk. The other three silks show smaller variations and a more uniform distribution of amino acids.

The chemical studies of spider silk shed very little light on the possible protein structures. This led to a series of biophysical studies, the vast majority of which were conducted on dragline (major ampullate) silk due to the ease of obtaining large quantities and its unique mechanical properties. The large variations in properties and amino acid composition probably make it unwise to extrapolate the data on dragline silk to any of the other silks. As methods develop to obtain sufficient quantities of the other silks, biophysical studies on them are likely to be initiated.

The predominant method of studying the silk structure was the use of X-ray diffraction on either fibers or powdered silk. These studies led to the classification of dragline silk into the standard β silk groups 3, 4, or 5.[13] The predominant structure was β-sheet with the different groups distinguished by the intersheet distances, the higher numbers indicating a greater spacing. The presence of the large bulky amino acids which could not be accommodated in the β-sheet and the X-ray diffraction patterns both indicated that the silk was also composed of substantial regions of undefined structure.

Using polarized Fourier transform infrared (FTIR) spectroscopy, we recently examined dragline silk fibers in both the relaxed and extended states using a microscope to focus the beam on the fiber to increase the signal to noise ratio.[14] These data confirmed the presence of substantial regions of β-sheet in both the relaxed and extended forms and further indicated that virtually no changes occurred in these structures when the fiber was stretched. These β-sheet regions are predominantly oriented parallel to the fiber axis. When the fibers were extended, an unexpected change was observed with the apparently random structure regions shifting to a helical structure. The extent of helical formation was correlated with the degree of fiber extension. When the fiber was allowed to relax, the spectra returned to the original form indicating the change was completely reversible. Polarization indicated that the helices were also oriented parallel to the fiber axis. This change in structure was observed for dragline silks from both *Nephila clavipes* and *Araneus gemmoides*. Minor ampullate silk from both species, which exhibits very limited elasticity, showed no helix formation when under tension and, instead, showed breakdown of the β-sheet region as the fiber started to break.

In order to further study the secondary structure of the dragline silk proteins, the silk was dissolved in 4.7 M $LiClO_4$. This harsh chaotropic reagent is necessary to dissolve the silk as it is insoluble in less vigorous reagents. The dissolved silk was then dialyzed against water with samples removed at various times to assess the structure of the proteins. The goal was to follow the reformation of the structure of the protein. It was thought that it would be possible to observe the protein structure as it reformed in solution before the silk became insoluble. The analysis was performed using UV circular dichroism (CD) on the soluble portion of each sample.

The data clearly indicated that in 4.7 M $LiClO_4$ the proteins were in a random structure (minimum at 195 nm). As the dialysis proceeded, the proteins began to assume a β-sheet-like structure with a strong minimum observed at 215 nm. The progression indicated that it was only a two-state transition, either random or sheet. When the dialysis was done into 0.1 or 1.0 M NaCl, the structure was random at low temperatures but showed an inverse temperature transition to the β-sheet form between 40 and 80 °C. The sample in 0.1 M NaCl showed nearly complete formation of β-sheet, whereas the sample in 1.0 M NaCl demonstrated only about

(13) Fraser, R. D. B.; MacRea, T. P. *Conformation in Fibrous Proteins*; Academic Press: New York, 1990; pp 293–343.
(14) Dong, Z.; Lewis, R. V.; Middaugh, C. R. *Arch. Biochem. Biophys.* 1991, *284*, 53–57.

40% sheet formation. Interestingly, an inverse temperature formation has also been observed for a peptide model of elastin, which is the other major elastic protein.[15]

When the dialyzed silk proteins were placed in 50% trifluoroethanol, which is known to promote helix formation, the sample showed a classic helix spectrum with minima at 204 and 220 nm. These data agree with the FTIR data described above which indicate the presence of substantial β-sheet in the fiber and suggest the presence of regions which can form helical structures as well.

Studies of the supercontracted form of dragline silk also provided key data concerning both the structure and energetic basis for elasticity. X-ray diffraction data from this form of silk indicated that rotation of the β-sheet regions might be responsible for the contraction of the fiber.[16] FTIR data from our laboratory have indicated some change in the sheet region but have not been sensitive enough to determine if it is due to a rotation of the sheet. Whatever the exact details, water is clearly causing changes in the β-sheet regions relative to the other regions. Study of the energetics of elasticity of the supercontracted silk showed that the predominant force for retraction was entropic, accounting for about 85% of the force.[5] A later calculation estimated the size of the regions involved to be 15 amino acids.[4] However, it is unclear how this data will apply to the normal dry form of dragline silk since the wet supercontracted silk has significant structural differences.

A distillation of the biophysical data suggests a model of a protein fiber composed of pseudocrystalline regions in a β-sheet conformation interspersed with regions of an unknown but presumably random conformation. The noncrystalline random structure regions form an extended conformation when the fiber is stretched. Retraction of the fiber is mainly due to unfavorable entropic changes occurring in the random regions when they stretch.

Cloning and Sequencing

In order to achieve the results described below, several new methods had to be developed. These are briefly detailed here. Despite the fact that much research had been done on the dragline silk obtained by taking the trailing thread left by the spider, we found that frequently two different fibers were present. Thus we developed a method to obtain only a single fiber of dragline silk. This was based on a paper by Work and Emerson[17] in which they "forcibly silked" the spiders. We designed a simple apparatus to do this using a variable speed drill controlled by a sewing machine foot pedal. The spider is anaesthetized with CO_2 and gently taped to the platform under a microscope. A single silk fiber is taken with forceps and wrapped around a spool attached to the drill. The foot pedal is used to control the speed as the fiber is pulled out onto the spool. It is possible to obtain 100 m or about 1 mg at a single silking. Minor ampullate silk can be spooled this way as well.

Unfortunately, it is not possible to use this method for the other silks. Both cocoon and swathing silk must be teased from the egg cocoon and swathed prey, respectively. These are removed by gently separating a fiber and teasing it out of the rest of the silk. It is possible to obtain up to 1 mg with this method also. A white hydrophobic protein secretion is frequently observed on the egg cocoon, and care must be taken to avoid silk with that coating as it seriously alters the amino acid composition and the peptides obtained from fragmentation.

The protein sequencing efforts on the purified silk fibers led to both erratic and uninterpretable data. Since we already knew the silk could only be dissolved with the harshest chaotropic agents, we decided to fragment the proteins instead of trying to purify them. We initially used every protease we could beg, borrow, or buy, all to no avail. None of them was able to digest the protein in the solid state. We were also unable to find conditions in which both the silk proteins were soluble and the enzyme was active. At the beginning the only chemical cleavage method we found to work was partial acid hydrolysis. Even this method was difficult because the repetitive nature of the proteins led to very little difference in the strength of various peptide bonds. Thus we only found penta- and heptapeptides in our digests. These were sequenced and used as the basis for our cloning experiments.

Finally, we had to develop methods to help our DNA sequencing which, despite being a well-established set of procedures, was not designed for repetitive proteins of high guanine/cytosine (G/C) content. The repetitive nature of the DNA prohibited the use of random fragmentation of the clones, sequencing of fragments, and overlapping of fragments to align the total sequence. We chose instead to use the nested deletion method in which an exonuclease is used to remove bases from one end of the clone in a time-dependent manner. With this procedure it is possible to obtain specific size clones which are then sequenced and easily aligned. However, due to the repetitive nature of the DNA and apparent secondary structure from it, the time course for deletion was not linear. This led to difficulties in finding clones of the correct size. To aid in screening, a rapid high-volume procedure for the isolation of plasmids was created. This allowed us to screen a large number of plasmids to identify the ones with the correct size.

Spidroin 1

The first dragline silk protein DNA was initially cloned in 1989, but the DNA sequencing required over 9 months to complete as we had not developed the new techniques at that time.[18] The clone we obtained was 2.4 kb in length and contained the carboxy end of the protein including the 3′ untranslated region into the poly A tail. We subsequently determined that the complete mRNA was approximately 5.6 kb. From restriction digests of genomic clones, we believe the rest of the sequence is very similar to that portion we have already sequenced.

This protein, called spidroin 1, had a unique protein sequence (Figure 1). The 24 repeats shown in this figure are followed by an additional 96 amino acids which show very little sequence similarity to the repeats at either the protein or DNA levels. Careful examina-

(15) Urry, D. W. *J. Protein Chem.* **1988**, *7*, 1–34.
(16) Fornes, R. E.; Work, R. W.; Morosoff, N. *J. Polym. Soc.* **1983**, *21*, 1163–1172.
(17) Work, R. W.; Emerson, P. D. *J. Arachnol.* **1982**, *10*, 1–10.
(18) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120–7124.

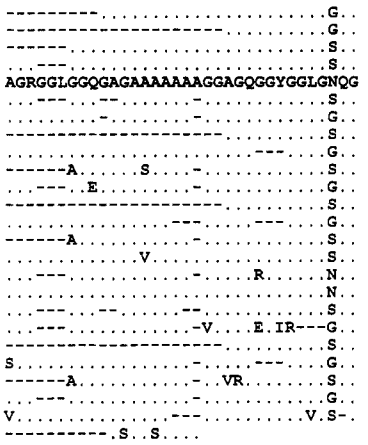


Figure 1. Spidroin 1 repeat sequences. The repeating amino acid sequences of spidroin are shown with maximum alignment. The single letter amino acid abbreviations are used.

tion of the repeating sequence reveals three distinct regions within it. The first nine amino acids are AGRGGLGGQ, but this region is clearly not well conserved with either partial or complete deletion in several cases. Interestingly, the deletions are all in multiples of three amino acids. The second region is the $GAG(A)_n$ region with n between 4 and 7, but predominantly 6. This region is highly conserved with very few substitutions. The third region is the final 15 amino acids, which generally show a GGX type of three amino acid repeat. This region is also highly conserved except for one position which varies among G, S, and N. There are very few substitutions with the few deletions only in three amino acid groups as in the first region. Overall, the repeat sequence can be thought of as $(GGX)_n(A)_n$ or $(GXG)_n(A)_n$ with X being Gln, Tyr, or Leu. A search of the protein database failed to find any protein with an identical sequence to any six amino acids except the poly-Ala region. Thus the sequence and structure of the GGX regions would seem to be unique in the protein world to date.

Spidroin 2

Following the completion of the sequencing of this clone, it was apparent that spidroin 1 could not account for all of the dragline silk protein. This was due to two factors: (1) one of the major peptides sequenced from the acid-hydrolyzed fragments, which contained a Pro residue, was absent from spidroin 1, and (2) the dragline silk fiber contains 3.5% Pro whereas spidroin 1 is devoid of Pro. On the basis of the sequence of the missing peptide, we synthesized a new DNA probe and rescreened the original library. We quickly identified a new clone which was different from spidroin 1.[19] This clone of 2.4 kb was sequenced in a much shorter time due to the advances in techniques described above. However, that sequence presented its own set of difficulties as the first strand sequence had numerous areas where the secondary structure of the DNA prevented accurate determination of the sequence. Thus we were forced to sequence the second strand in order to insure the sequence was entirely correct. This clone was also lacking the complete sequence but contained the carboxy terminal region including the 3′ untranslated region and the poly A tail. The mRNA for this clone is 3.5 kb,

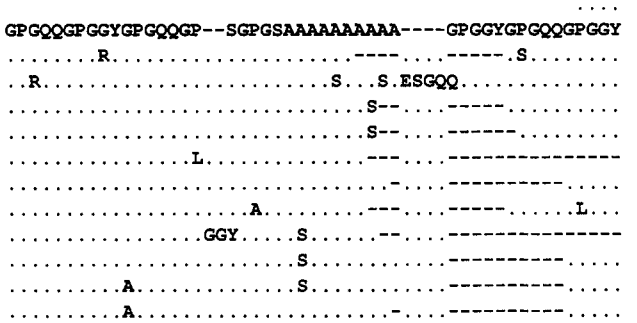


Figure 2. Spidroin 2 repeat sequences. Data as in Figure 1.

indicating that we have a high proportion of the sequence represented.

The amino acid sequence of spidroin 2 is shown in Figure 2. The repeating sequence motif is easily seen as in spidroin 1. Interestingly, there are also three regions in the repeat sequence of spidroin 2. The first 15 amino acids are very highly conserved and consist of three 5 amino acid repeats of GPGXY with XY being QQ or GY. The second region mimics that of spidroin 1 with a heptapeptide GPSGPGS leading into a poly-Ala region of 6–10 Ala residues. This is also highly conserved with mainly conservative Ser substitutions occurring. The final region is like the first region with pentapeptide units, but there is a great deal of diversity with deletions of one, two, or all of the units. This is also the same as found in the first region of spidroin 1. When the database was searched with this repeat segment, no identical matches were found. However, there were proteins with repetitive sequences containing Gly and Pro present. These are described more fully later.

Computer Analysis of Spidroin Sequences

Another very interesting common feature of these two proteins is the codon usage in their genes. There is an incredible codon bias toward not using G or C in the third or wobble base of the codon. This bias results in 86.5% of the codons using A or T in the wobble base or nearly a 7-fold excess of A/T over G/C. There are no other spider genes in the databases for comparison; however, we isolated and partially sequenced an actin gene from the major ampullate silk gland which did not exhibit this codon bias. The bias against G or C in that position may not be overly surprisingly as the first two bases for the majority of the amino acids in the silk proteins are C or G. Since excessive formation of secondary structure would prevent translation into protein, the evolutionary pressure appears to keep A and T as the wobble base.

When computer secondary structure prediction methods were used on these two proteins, a number of interesting possibilities were suggested. The first was that the poly-Ala regions were predicted to form helices. This is because Ala is one of the strongest helix-forming amino acids as determined by virtually all methods. For spidroin 1 the GGX sequence gave no clear predicted secondary structure, but it clearly was not β-sheet. In fact, evidence from peptide studies strongly suggests that this type of sequence cannot form β-sheets.[20] Spidroin 2, in contrast, was predicted to have numerous β-turns each centered at the pentapeptide unit. In fact, by some prediction schemes,[21] the GPGQQ sequence

(19) Hinman, M.; Lewis, R. V. *J. Biol. Chem.*, in press.

(20) Lotz, B.; Brack, A.; Spach, G. *J. Mol. Biol.* **1974**, *87*, 193–203.

had the second highest possible predictive value for a type II turn, which is one of several possible turn structures observed in proteins. The GPGGY repeat also has a very high predictive value for this turn.

Peptide Studies

The low solubility of the spider silk proteins precluded detailed studies of their structure as was described earlier. To circumvent that problem, peptides based on the repeating sequences of each protein were synthesized. P30 represented the consensus sequence of spidroin 1, with P15 being the conserved last 15 amino acids ($(GGX)_n$) of that repeat. P47 is the complete consensus repeat of spidroin 2. Using CD, these peptides were studied at various temperatures, concentrations, and organic solvents. FTIR was also used to study the dried peptides in a film.

P30 and P15 showed only random structure below 1 mg/mL. Above that concentration there was a marked temperature dependence on structure. At low temperatures (<20 °C) based on CD, the structure appeared to be a triple-stranded helical coil, which Gly-Ala polymers seem to form.[22] Between 20 and 40 °C the structure had changed but fit no clear pattern. The best fit was similar to a predominance of turns. As the temperature increased above 40 °C, the structure strongly resembled a β-sheet. This is the same type of inverse temperature transition observed with the dialyzed silk proteins from the fiber. An interesting sidelight of these studies was that above 10 mg/mL the P30 peptide solution started to precipitate with the precipitate being very insoluble and FTIR spectroscopy showing a β-sheet structure.

When the peptides were placed in solutions containing trifluoroethanol (TFE), which can induce helix formation, P30 appeared to form a classic poly-Ala helix with minima at 208 and 220 nm with a ratio nearly identical to that of poly-Ala. P15 showed no tendency to form a helix in these solutions but remained in a random structure. P47 demonstrated no clear structure under any of the conditions used. The fact that the predicted turns were not observed with P47 was not unforeseen as turns are difficult to detect with virtually all spectroscopic methods.

The conclusions of these studies were that P30 appeared to be flexible enough to form a variety of secondary structures which depended on the conditions. The presence of the poly-Ala sequence in P30 was necessary and sufficient to induce helix formation in TFE. However, the structure of P47 could not be induced into a helical conformation despite the presence of an even longer poly-Ala sequence. P47 gave no clear evidence of any strong secondary structure, although the detection of turns would have been difficult.

Protein Structures

At this point our original structural hypothesis based on spidroin 1 was clearly invalid. This hypothesis was that the poly-Ala regions were the "random" structure which formed the helical structure when stretched and that the Gly-rich regions were the β-sheet. In addition to devising a new protein structure, we had to fit both proteins into a structure that would form a fiber with the properties of dragline silk.

An important piece of data was the fiber X-ray diffraction study which clearly showed a 5.3-Å intersheet spacing. This spacing is exactly that of the Ala β-sheet observed with small poly-Ala peptides[23] and totally inconsistent with any other amino acid sequence in either protein. Another important experiment was the determination of the water content of the silk fibers (Matsuno and Lewis, unpublished data). These showed less than 6% water by weight, which corresponds to less than one water molecule per three amino acid residues. This very low level of water is exactly what was required to form β-sheets with the small poly-Ala peptides.

Thus the poly-Ala segments must be forming the β-sheet regions seen in the fibers. By forming interprotein sheets, these poly-Ala pseudocrystalline regions would serve to interlock the two different proteins in phase where the nonsheet regions would be aligned as well. Due to the high number of hydrogen bonds in the β-sheet and the hydrophobic interactions of the Ala chains in the absence of water, these regions would provide the tensile strength observed in the fibers.

The assignment of the poly-Ala regions to β-sheet forced the Gly-rich regions to be responsible for elasticity since the β-sheet regions were unaffected by stretching. The very high probability that the pentapeptide segments formed β-turns indicated a similarity to gluten[24] and elastin,[14] both of which have elasticity based on linked β-turns which form a β-spiral. At this point it is unclear whether these linked β-turns form a planar structure or a spiral. Either form can be extended sufficiently to account for the elasticity of the fiber. Computer modeling suggests that both are viable options, although detailed energy minimizations for both the retracted and extended forms have not yet been done. The question of the retractive force still remains as both hydrogen bonding and cis–trans conversions at the Pro residues could play important roles.

This model differs greatly from the traditional thinking about spider silk and its elasticity and strength. Yet the actual protein sequences indicate the traditional view cannot be valid. Only further structural and biophysical studies can resolve which of the models above is correct or whether another model must be devised.

Concluding Remarks

The elucidation of the sequences of these two spider silk proteins opens up research into both the basic understanding of the protein structure/function relationships and possible commercial application of these proteins as biomaterials. The basic understanding of how these proteins provide strength and elasticity simultaneously will require the use of techniques adapted to the solid phase in which they exist. This is in sharp contrast to virtually all current protein studies which deal with soluble proteins. The availability of other types of spider silks and a wide variety of species will help identify key areas of the protein which are needed for various properties. Methods such as solid-state NMR and FTIR spectroscopy which can utilize small samples may give detailed structures of the proteins in

(21) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221–232.
(22) Rippon, W. B.; Walton, A. G. *Biopolymers* **1971**, *10*, 1207–1212.
(23) Arnot, S.; Dover, S. D.; Elliott, A. *J. Mol. Biol.* **1967**, *30*, 201–208.
(24) Matsushita, N.; Creutz, C. E.; Kretsinger, R. H. *Proteins: Structure, Function and Genetics* **1990**, *7*, 125–155.
(25) Anderson, S. O. *Comp. Biochem. Physiol.* **1970**, *35*, 705–711.
(26) Work, R. W.; Young, C. T. *J. Arachnol.* **1987**, *15*, 65–80.

the fiber state. The combination of these and other approaches will be needed to determine the details of how proteins can function as fibers.

Commercial application of spider silk fibers will depend on the availability of a sufficient amount of material. Clearly the most likely source will be to use microorganisms to produce these proteins, which can then be spun into fibers. Our efforts in this direction have shown that it is feasible to produce these proteins in bacteria, but the yields have been very low. With this hurdle surmounted the key factor will become cost. At this point it is likely the first applications will be in medical products where cost is less of a factor, although other high-technology applications may be cost effective as well.

Although it is unlikely that either a complete understanding of the mechanism of spider silk strength and elasticity or commercial production of spider silk proteins will be achieved in the short term, the mystery of spider silk is unraveling.

Hopanoids. 1. Geohopanoids: The Most Abundant Natural Products on Earth?

GUY OURISSON* and PIERRE ALBRECHT

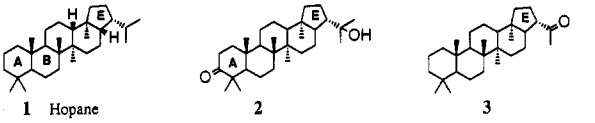
Laboratoire de Chimie des Substances Naturelles, CNRS, Université Louis Pasteur, F-67084 Strasbourg, France

Received January 29, 1992

Introduction

This Account introduces the reader to a most abundant family of complex organic substances on Earth and to their significance in geochemistry, as well as in fields from petroleum exploration to archaeology. These "geohopanoids" are molecular fossils, whose identification has exposed their precursors, the "biohopanoids", a novel family of bacterial lipids presented in the following Account in this issue.

Hopane, the name of the parent hydrocarbon 1, is derived neither from that of one of the cardinal virtues nor from that of *Humulus lupulus* (though "hops" is indeed a slang name of hopanoids used by American petroleum geochemists). It recalls rather the lesser 18th century British botanist John Hope, in whose honor his colleague Roxbury called a genus of huge Burmese trees *Hopea*, lavish producers of a resin used in the varnish trade called "dammar"; in 1955, at the British Museum, Mills and Werner isolated from such a *Hopea* dammar the hydroxy ketone 2, which they called "hydroxyhopanone".[1]



1 Hopane 2 3

Since then, C_{30} (or C_{29}) derivatives of hopane 1 have been found in scattered plants, some in higher plants (then, always with an oxygen function at C-3) or, without such a function, some in lichens and many ferns; these "phytohopanoids" have remained rare and few, and constitute only one of the many minor families of plant triterpenes present in secretions, barks, and cuticles.

It was therefore a major surprise some 20 years ago, soon after we initiated our study of sedimentary organic substances,[2] that Ted Whitehead, at British Petroleum, would identify hopane derivatives in a crude oil and we in several sediments.[3] The surprise became exhilaration when we realized that we were dealing with a major finding, as geohopanoids are very varied (more than 200 structures identified so far), more abundant globally than any other group of natural products, ubiquitous in sediments, useful in many ways, and biologically important. The study of organic geochemical problems in the Strasbourg group and elsewhere has provided many important results, but the "hopanoid story" remains probably the most extraordinary one.

The Prehistory of Geohopanoids

Our identification of these geohopanoids was markedly eased by our unplanned preparedness. Firstly, with Takeioshi Takahashi, we had studied the "gurjun" balsams of *Dipterocarpus*, trees closely allied to *Hopea* and containing polyterpenes closely related to those of

Guy Ourisson, born in 1926, was educated in Paris (Ecole Normale Supérieure; Dr.Sc., G. Dupont, 1954) and at Harvard University (Ph.D., L. F. Fieser, 1952). In 1955, he was appointed to the University of Strasbourg, where he has since been a Professor. From 1971 to 1975, he was Founding President of the new Université Louis Pasteur. From 1985 to 1989, he was Director of the CNRS Institut de Chimie des Substances Naturelles in Gif-sur-Yvette, France. G.O. has received prizes from the French, Belgian, British, American, and German Chemical Societies, as well as the Roussel Prize which he shared with P. A., the Heinrich Wieland Award, and the A. von Humboldt Forschungspreis. He is an Honorary Member of the Chemical Societies of Belgium, Great Britain, and Switzerland and a Foreign Member of the Science Academies of Sweden, Denmark, India, Rhineland-Westfalia, Luxembourg, Serbia, Leopoldina (Halle), and Academia Europaea; he is also the Foreign Secretary of the French Académie des Sciences. P.A. has twice received the Best Paper Award (once shared with G.O.) and the Alfred Treibs Award from the American Geochemical Society and prizes from CNRS and the French Chemical Society. Along with structural studies on many natural products, G.O. has launched a variety of new lines of research in Strasbourg, always at the interface between chemistry and biology or other fields and often pursued by his former students.

Pierre Albrecht, born in 1941, was educated in Strasbourg and obtained his doctorate in 1969 in G.O.'s group, in which he initiated geochemical studies. After a postdoctoral year at the University of California at Berkeley with W. G. Dauben, he returned to Strasbourg where he runs the Organic Geochemistry Laboratory. A Research Director in the CNRS, he has succeeded G.O. as Director of the Laboratoire de Chimie Organique des Substances Naturelles, a federation of teams of researchers at Université Louis Pasteur, associated with CNRS.

(1) Mills, J. S.; Werner, A. E. A. *J. Chem. Soc.* **1955**, 3132–3140.
(2) Albrecht, P.; Ourisson, G. *Geochim. Cosmochim. Acta* **1969**, 227–230. Albrecht, P.; Ourisson, G. *Angew. Chem., Int. Ed. Engl.* **1971**, *41*, 209–225.
(3) Whitehead, E. V. *Chem. Ind.* **1971**, 1116–1118. Ensminger, A.; Albrecht, P.; Ourisson, G.; Kimble, B. J.; Maxwell, J. R.; Eglinton, G. *Tetrahedron Lett.* **1972**, 3861–3864.

0001-4842/92/0125-0398$03.00/0