Mommaerts, W. F. H. M. (1952), *J. Biol. Chem. 198,* 445.
Mommaerts, W. F. H. M., and Parrish, R. G. (1951), *J. Biol. Chem. 198,* 445.
Perry, S. V., Cole, H. A., Head, J. F., and Wilson, F. J. (1972), *Cold Spring Harbor Symp. Quant. Biol. 37,* 251.
Potter, J. D. (1974), *Arch. Biochem. Biophys. 162,* 436.
Potter, J. D., Seidel, J. C., Leavis, P. C., Lehrer, S. S., and Gergely, J. (1974), *in* Symposium on Calcium Binding Proteins, Drabikowski, W., and Carafoli, E., Ed., Amsterdam, Elsevier (in press).
Regenstein, J. M. (1972), Ph.D. Thesis, Brandeis University, Waltham, Mass.
Regenstein, J. M., and Szent-Gyorgyi, A. G. (1973), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 32,* 570.
Schaub, M. C., Perry, S. V., and Hacker, W. (1972), *Biochem. J. 126,* 237.
Straub, F. B. (1942), *Studies Inst. Med. Chem. Univ., Szeged 2,* 3.
Szent-Gyorgyi, A. (1951), Chemistry of Muscular Contraction, 2nd ed, New York, N.Y., Academic Press.
Tilney, L. G., Hapano, S., Ishikawa, H., and Mooseker, M. S. (1973), *J. Cell Biol. 59,* 109.
Weber, A., and Murray, J. M. (1973). *Physiol. Rev. 53,* 612.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.
Wilkinson, J. M., Perry, S. V., Cole, H. A., and Trayer, I. P. (1972), *Biochem J. 127,* 215.

The *Bombyx mori* Silk Proteins: Characterization of Large Polypeptides†

Karen U. Sprague

ABSTRACT: Proteins taken directly from the *Bombyx mori* silk gland have been separated and identified as either fibroin or sericin on the basis of their location within the gland and their amino acid composition. Molecular weights of these polypeptides have been determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis, and by agarose-guanidine chromatography. Fibroin consists of approximately equimolar amounts of two large (350,000) polypeptide chains. These may be the products of distinct fibroin alleles present in hybrid silkworm strains. Sericin, on the other hand, is composed of at least the three largest polypeptides (130,000–220,000) present in a mixture of proteins ranging in size from about 20,000 to 220,000.

Physical characterization of the silk proteins produced by *Bombyx mori,* the commercial silkworm, is of current interest because recent experiments have revealed some structural features of the gene and the mRNA molecule coding for one of these proteins, fibroin (Suzuki and Brown, 1972; Suzuki *et al.,* 1972; Lizardi and Brown, 1973).

Historically, two kinds of silk proteins have been distinguished: fibroin, a fibrous protein which makes up the core of the silk filament; and sericin, a poorly characterized protein (or proteins) which surrounds the spun fiber and functions as an adhesive (Lucas *et al.,* 1958). Both types of protein accumulate in the middle silk gland during the fifth larval instar. Classical experiments in which specific portions of the gland were removed from living animals establish that fibroin is actually synthesized by the posterior silk gland, whereas sericin is produced by the cells of the middle silk gland (Machida, 1927).

Although fibroin from *Bombyx mori* has been studied for over a century (Cramer, 1865), accurate determination of its size and possible subunit structure has been complicated by the difficulty of preparing solutions in which the molecules are disaggregated yet intact. Hence, molecular weight estimates for fibroin, which have been obtained primarily by sedimentation analysis, range from about 2×10^4 to over 10^6 (Rao and Pandit, 1965; Lucas, 1966; Tashiro and Otsuki, 1970a,b; Tashiro *et al.,* 1972; Tokutake and Okuyama, 1972; Sridhara *et al.,* 1973; Sasaki and Noda, 1973a,b). It is not clear whether more than one type of polypeptide chain is involved. The inability to distinguish clearly between fibroin and sericin components has contributed further uncertainty to the characterization of these molecules.

In this paper, I present evidence that fibroin consists of approximately equimolar amounts of two large (3.5×10^5) polypeptide chains which are distinct from the sericin proteins. Rather than attempting solubilization of spun silk fibers (a procedure which frequently leads to degradation), I have solubilized fibroin and sericin taken directly from the middle silk gland of mature *Bombyx* larvae. Fractionation and characterization of the silk proteins were achieved by molecular sieve techniques (SDS[1] polyacrylamide gel electrophoresis and guanidine-agarose chromatography).

Materials and Methods

Rearing of Silkworms. Silkworm larvae were hatched from a laboratory stock of eggs (originally supplied by O. Yamashita) and were reared either on fresh mulberry leaves or on the artificial diet described by Suzuki and Brown (1972). In both cases, the larvae weighed between 4

† From the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510. *Received July 15, 1974.* This work was supported by a grant from the National Science Foundation (GB-41926) awarded to J. A. Steitz and by the Sessel Anonymous Fund.

[1] Abbreviations used are: SDS, sodium dodecyl sulfate; Gdn · HCl, guanidine hydrochloride.

Table I: Column and Gel Standards.

Protein	Source	Molecular Weight	Ref
Rabbit muscle myosin	Prepared by the method of Szent-Gyorgyi (1951)	470,000	*a*
Core		420,000	
Subunit		212,000	
Human γ-globulin	A gift of W. H. Konigsberg	154,000	*b*
Heavy chain		51,600	
Light chain		23,500	
E. coli RNA polymerase β, β′ subunits	A gift of R. Ludwig	155,000–165,000	*c*
E. coli β-galactosidase subunit	Sigma	130,000	*d*
Rabbit muscle phosphorylase subunit	Worthington	92,500	*e*
Bovine serum albumin	Miles, fraction V	69,000	*f*
E. coli alkaline phosphatase subunit	Prepared by the method of Torriani (1966)	43,000	*g*
T_7 bacteriophage proteins	Prepared by SDS solubilization of whole phage particles or phage infected *E. coli*, labeled with [^{35}S]methionine.	40,000-150,000	*h*

[a] Gershman *et al.* (1969). [b] Edelman *et al.* (1968). [c] Burgess (1969). [d] Weber *et al.* (1964). [e] Seery *et al.* (1967). [f] Tanford *et al.* (1967). [g] Schlesinger (1964). [h] Studier (1972).

and 5 g when sacrificed near the end of the fifth instar.

Preparation of Fibroin. Pure fibroin was dissected as a translucent column from the posterior third of the middle silk gland as described by Tashiro and Otsuki (1970a). It was solubilized directly in either 8 M urea or 6 M Gdn · HCl (both made pH 7.5 with 0.01 M Tris-HCl). After occasional gentle shaking for 18–24 hr at room temperature, no undissolved material remained. In this manner, 1 or 2 ml of protein solution at a concentration of 10–15 mg/ml were routinely obtained from a single larva. Alternatively, I was able to obtain fibroin from the posterior silk gland by exposing small (3–4 mm) pieces of the gland to distilled water. This procedure causes extrusion of the fibroin which can then be collected in a capillary tube, yielding about 0.5 mg of protein/larva.

Preparation of Sericin. To prepare solutions of sericin proteins, the fluid material at the anterior tip of the middle silk gland was removed with a drawn-out capillary and dissolved in approximately 0.1 ml of 10% SDS at room temperature. The yield of unfractionated sericin proteins was about 0.5 mg/larva.

Gel Electrophoresis. Discontinuous SDS polyacrylamide gel electrophoresis was performed according to Laemmli (1970). Gels were formed either in tubes (5 × 127 mm) or in the slab gel apparatus described by Studier (1973). The stacking gel was omitted. Proteins were generally boiled for 3 min in sample buffer containing β-mercaptoethanol (Laemmli, 1970) and then layered directly on the running gel (3–5% acrylamide, 0.08–0.12% bisacrylamide) in volumes of 20–50 μl at a concentration of 0.2 μg protein/mm^2 gel surface. After electrophoresis, proteins were stained with Coomassie Blue, and the gels were destained in 7% acetic acid as described by Maizel (1971).

Acrylamide (Eastman) was recrystallized from chloroform (Loening, 1967). SDS (Matheson Coleman and Bell) was recrystallized from hot ethanol before use either in polyacrylamide gels, or in SDS–protein binding experiments (Burgess, 1969).

Gdn · HCl-Agarose Chromatography. Columns (1.8 × 75 cm) of 2% agarose equilibrated with 6 M Gdn · HCl (Schwartz/Mann, ultra pure), 10 mM Tris-HCl (pH 7.5), and 1 mM EDTA were prepared essentially according to Fish *et al.* (1969). Samples containing 3–4 mg of the protein to be analyzed, in addition to void volume and included volume markers, were taken up in the column solvent and made 7–8 M in Gdn · HCl or 5% in sucrose. After a small volume (0.3–0.5 ml) of this solution had been layered under the solvent at the top, the column was operated at a hydrostatic pressure differential of 15 cm. The optical density of the column effluent was monitored with an LKB Uvicord, and radioactivity was assayed by scintillation counting.

Data from different column runs were normalized using the distribution coefficient, K_d, to express the elution volume (Gelotte, 1960). K_d is defined as

$$K_d = (V_e - V_0)/(V_i - V_0)$$

where V_e is the elution volume of the protein being analyzed, V_0 is the excluded volume of the column, and V_i is the included volume.

Column and Gel Standards (Marker Proteins). The proteins used as molecular weight markers and their sources are listed in Table I.

Column Standards (Included and Excluded Volume Markers). Because it contains only a small fraction of appropriately large polymers, Blue Dextran 2000 (Pharmacia) is a poor void volume marker for 2% agarose. Therefore, bacteriophage T_7 DNA (single stranded molecular weight 12 × 10^6 [Studier, 1972]) was used for this purpose. DNA was prepared by KCl-SDS extraction of purified phage particles (Chessin and Summers, 1970), and was

checked for intactness by sedimentation through alkaline sucrose (Masamune and Richardson, 1968); 1–2 OD units were utilized in each column run.

The included volume of the column was determined by monitoring the elution of approximately 1×10^6 cpm of a ^{14}C- or ^{35}S-labeled amino acid.

Carboxymethylation. Proteins were carboxymethylated with iodoacetic acid as described by Fish *et al.* (1969).

Measurement of SDS Binding to Protein. The amount of SDS bound to fibroin and to a control protein, bovine serum albumin (from Miles), was determined after equilibrium dialysis according to a modification of the method described by Reynolds and Tanford (1970a). SDS binding was achieved by boiling each protein for 5 min at a concentration of 1 mg/ml in 0.2% SDS, 0.01 M Tris-HCl (pH 7.5), and 1% β-mercaptoethanol. Urea (derived from the original solvent in the case of fibroin) was present at 0.2–1 M. Samples were then dialyzed for 6 days against 0.1% SDS and 0.01 M Tris-HCl (pH 7.5), containing 1 M urea in some cases. The SDS concentrations inside and outside the dialysis bags were measured daily to determine when equilibrium was reached. SDS was assayed by complex formation with Methylene Blue according to Reynolds and Tanford (1970a), with the following reduction in scale: samples containing 0.5–8 μg of SDS were assayed in the presence of 0.5 ml of Methylene Blue (0.1 mg/ml) and 2 ml of chloroform. Protein concentrations were determined spectrophotometrically, using published extinction coefficients: fibroin, $E_{1cm}(1\%)$ at 276 nm = 11.3 (Tashiro *et al.*, 1972; and K. U. Sprague, unpublished observations); bovine serum albumin, $E_{1cm}(1\%)$ at 280 nm = 6.6 (Wetlaufer, 1962).

Amino Acid Analysis. The amino acid composition of unfractionated fibroin taken from the posterior third of the middle silk gland was determined with a Beckman 120 B amino acid analyzer after hydrolysis in 6 N HCl for 24, 48, and 72 hr.

Amino acid analysis of fractionated fibroin and sericin polypeptides was carried out after direct hydrolysis of the Coomassie Blue stained proteins within gel slices (Maizel, 1971).

Specific Staining of Fibroin and Sericin Proteins. Aliquots of fibroin and sericin were precipitated on Whatman 540 paper with ice-cold 10% Cl_3CCOOH, rinsed with ethanol, dried, and stained with a mixture of acid fuchsin, Methyl Green, and picric acid (Machida, 1927).

Results

(1) *Fibroin Molecular Weight Determination by SDS Gel Electrophoresis.* Electrophoretic analysis on gels of various polyacrylamide concentrations ranging from 3 to 5% (representative data from a 4% gel is shown in Figure 1) suggests that SDS-denatured fibroin obtained from the middle silk gland has a monomer molecular weight of about 350,000. This estimate is based on the average relative mobility of fibroin in ten gel runs, using different fibroin preparations and including bacteriophage T_7 proteins in addition to those markers shown in Figure 1. It should be noted that most of the marker proteins fall well below fibroin on the log molecular weight *vs.* mobility plot, and that the molecular weight estimate for fibroin thus depends heavily upon extrapolation. The single point above fibroin represents undissociated myosin, which was identified by the fact that the corresponding gel band disappears when myosin samples are dissociated more vigorously than usual (*e.g.*, by boiling for 5 min or more). The validity of this molecular

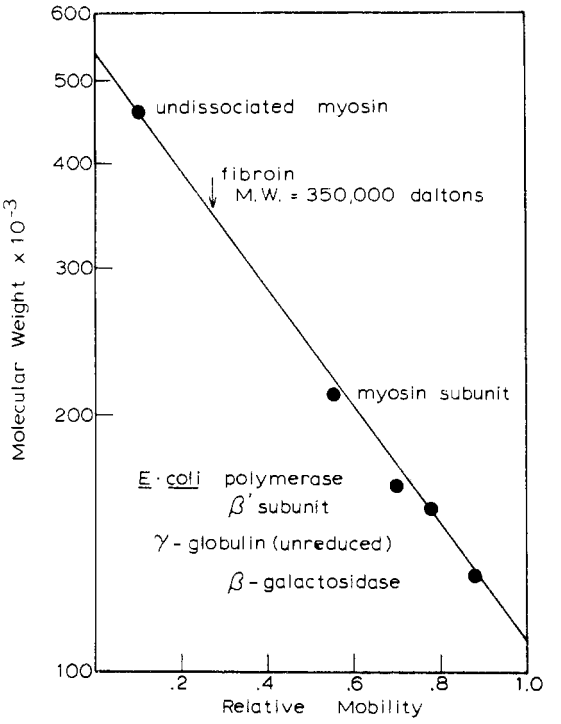

FIGURE 1: Determination of the molecular weight of fibroin by SDS polyacrylamide gel electrophoresis. Analysis was carried out using 5 μg of each protein in cylindrical gels containing 4% acrylamide and 0.1% bisacrylamide. The molecular weights of the marker proteins were taken from the data in Table I.

weight estimate, however, is supported by the fact that the mobility of fibroin relative to marker proteins is independent of gel concentration over a range of 3–5% acrylamide.

The observed molecular weight of fibroin was unaffected by varying the conditions under which the protein was denatured. Neither omission or inclusion of reducing agents (up to 5% β-mercaptoethanol or 0.1 M dithiothreitol), nor carboxymethylation of the reduced protein altered its relative mobility on gels. The means by which SDS binding was achieved likewise caused no detectable change: comparable results were obtained whether the protein was complexed with SDS by boiling for 3 min (Laemmli, 1970) or by incubation at 37° for 2 hr (Weber and Osborn, 1969).

In experiments where excellent resolution was achieved (either by exercising care not to overload the sample, or by including 3 M urea in the gel), fibroin was seen to fractionate into two bands of equal staining intensity (Figure 2). The molecular weights of both of these polypeptides are in the range of 350,000; they probably differ from each other by no more than 5%. The appearance of the two bands is correlated neither with the state of reduction of the protein nor with the use of a particular region of the silk gland as the fibroin source.

Analysis of fibroin on gels of smaller pore size (up to 15% acrylamide) does not reveal other major protein components. A faint band corresponding to a 25,000 molecular weight protein is sometimes seen on severely overloaded gels but amounts to only about 1% of the total stained material. It has been observed in samples which were not exposed to β-mercaptoethanol, and may be a contaminant, rather than an integral part of native fibroin.

(2) *SDS Binding to Fibroin.* Because fibroin taken from the silk gland does not readily dissolve in solutions of SDS alone, anomalous behavior on SDS polyacrylamide gels (due, for instance, to low affinity for SDS) must be ruled

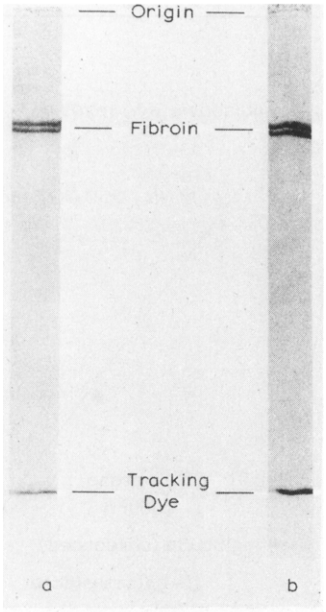

FIGURE 2: Separation of the polypeptide chains of fibroin by SDS polyacrylamide gel electrophoresis. The slab gel was 3% acrylamide, 0.08% bisacrylamide, and contained 3 M urea, in addition to the standard components. Samples consisted of 3 μg of fibroin from either (a) the middle or (b) the posterior silk gland of a single larva.

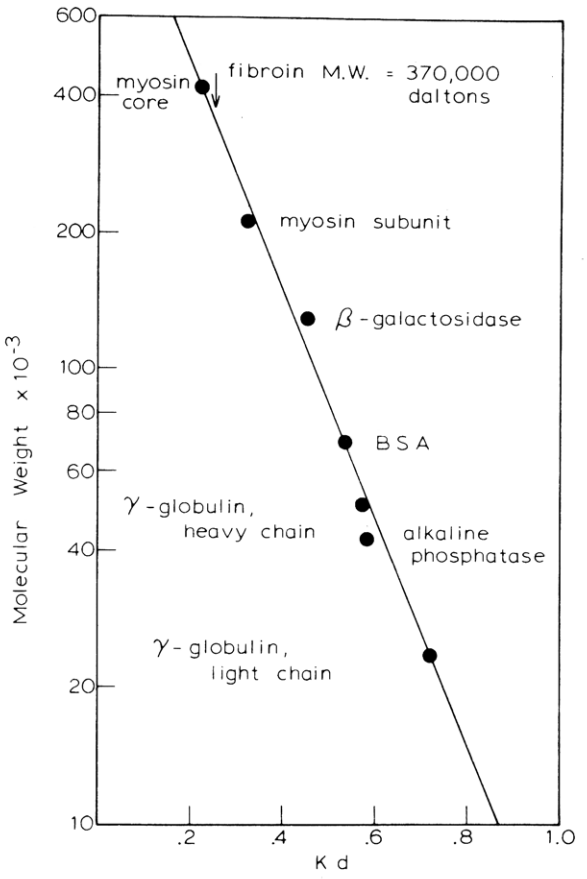

FIGURE 3: Determination of the molecular weight of fibroin by chromatography on 2% agarose equilibrated with 6 M Gdn · HCl. Before analysis, all proteins except myosin core were reduced and carboxymethylated as indicated under Materials and Methods. Myosin core was treated overnight at 4° with a tenfold molar excess (with respect to -SH groups) of *N*-ethylmaleimide, and was chromatographed in the presence of reduced amounts of guanidine HCl (3–4 M).

out. The possibility of an aberrant fibroin–SDS binding ratio was therefore tested directly.

The amount of SDS bound to fibroin was compared to that bound to bovine serum albumin under identical conditions. After equilibrium dialysis against 0.1% SDS, bovine serum albumin bound 0.45 mg of SDS/mg of protein, and fibroin bound 0.49 mg/mg of protein. These values are in good agreement with those reported for a variety of other proteins (Reynolds and Tanford, 1970a; Maizel, 1971). To control for the possible effect of residual urea in the fibroin preparation, SDS binding was assayed in the presence and absence of 1 M urea. The addition of urea did not alter SDS binding to either fibroin or bovine serum albumin. Furthermore, boiling in SDS prevents the precipitation of fibroin which otherwise occurs when urea is completely removed by dialysis.

(3) *Fibroin Molecular Weight Determination by Gel Filtration in Gdn · HCl.* To obtain an independent estimate of its molecular weight, fibroin was subjected to chromatographic analysis on 2% agarose in the presence of 6 M Gdn · HCl. Figure 3 shows that the molecular weight determined by this technique is 370,000, in agreement with the estimate of 350,000 from polyacrylamide gel electrophoresis of SDS denatured protein. This method does not, however, permit resolution of the two polypeptide chains which can be distinguished on gels.

(4) *Distinction between Fibroin and Sericin.* In addition to fibroin, I have prepared stable solutions of another class of silk proteins which I believe to be sericin. Unlike fibroin, the sericin proteins precipitate immediately upon exposure to cold water. Although they can be dissolved in urea or Gdn · HCl, the most stable solutions are obtained in 10% SDS at room temperature. I have used the following criteria to identify the sericin proteins and distinguish them from fibroin:

(a) LOCATION WITHIN THE SILK GLAND. Fibroin and sericin were taken from two quite distinct parts of the middle silk gland: fibroin from the posterior portion, and sericin from the tip of the anterior region. Differential histological staining has been used previously to assign fibroin and sericin to these locations within the gland (Yamanouchi, 1922; Machida, 1927). I have found that isolated fibroin and sericin after Cl_3CCOOH precipitation possess staining properties which are indistinguishable from those described for the silk proteins in tissue sections (Machida, 1927).

Conclusive identification of fibroin can be made by examining the protein present in the lumen of the posterior silk gland since fibroin is the only major silk protein synthesized in this part of the gland (Machida, 1927). Samples of such protein do, in fact, give gel patterns identical with those obtained with protein from the posterior third of the middle silk gland (Figure 2).

(b) MIGRATION IN SDS GELS. Comparison of Figures 3 and 4 shows that fibroin and sericin consist of polypeptide chains of quite different molecular weights. Unlike fibroin, sericin is a fairly complex mixture of polypeptides (at least 15), ranging in size from about 20,000 to 220,000. Thus, on the basis of size alone, the sericin proteins are readily distinguishable from fibroin.

(c) AMINO ACID ANALYSIS. The amino acid composition of the major silk proteins is presented in Table II. The values for unfractionated fibroin from the middle silk gland are in close agreement with the published data for fibroin isolated from cocoons. Preliminary analysis of the two separated fibroin polypeptides indicates that the same propor-

Table II: Amino Acid Composition of Fractionated[a] and Unfractionated Fibroin and Sericin. Per Cent of Total Amino Acid Residues.

	Fibroin				Sericin			
		This Work			Qualitative	This Work		
Amino Acid	Data of Lucas *et al.* (1958) Total	Total	350,000 dalton Proteins Slow	Fast	Data of Lucas *et al.* (1958) Total	220,000 dalton Protein	210,000 dalton Protein	130,000 dalton Protein
Gly	44.5	43.3	48	48	+++	7.8	8.7	8.7
Ala	29.3	34.7	37	37	++	4.0	4.5	2.9
Ser	12.1	10.3	11	10	+++	15	15	11
Tyr[b]	5.17	4.79			++			
Val	2.2	1.60	2.1	2.0	++			
Asp	1.3	1.39	0.64	0.59	+++	32	26	23
Glu	1.02	1.06	0.60	0.54	++	20	28	17
Thr	0.91	0.72	0.64	0.55	++	4.2	4.4	5.2
Ile	0.66	0.46	0.14	0.12	++			
Phe[b]	0.63	0.68			+			
Leu	0.53	0.24			++	1.7	1.5	3.6
Arg[c]	0.47	0.45			++			
Pro	0.36	0.35						
Lys	0.32	0.34	0.20	0.10	+	16	12	29
His	0.14	0.19						
Trp	0.11				++			
Met								

[a] Amino acid analyses on fractionated silk proteins within gel slices were carried out on small amounts of material: fibroin, 400 nmol of total amino acids per polypeptide; sericin, 50 nmol of total amino acids per polypeptide. Blank gel slices were analyzed to correct for contamination. [b] Neither tyrosine nor phenylalanine was detected in proteins hydrolyzed within gel slices. These amino acids may have been destroyed during hydrolysis, or obscured by unidentified material eluting in the same region. [c] Arginine in proteins hydrolyzed within gel slices was obscured by large amounts of ammonia derived from acrylamide.

tions of the major fibroin amino acids (glycine, alanine, and serine) appear in both proteins. The three largest sericin proteins are also rich in glycine, alanine, and serine, but differ from fibroin in having large amounts of aspartic acid and glutamic acid as well. A qualitative amino acid analysis of total sericin from cocoons (Lucas *et al.*, 1958) is included in Table II for comparison.

Discussion

Proteins of the B. mori Silk Gland. Silk proteins derived from different regions of the *B. mori* silk gland have been solubilized and characterized with respect to their size and amino acid composition. A distinction has been made between the two classes of silk proteins, fibroin and sericin.

Fibroin appears to consist of two very large polypeptide chains with slightly different mobilities in SDS gels. Both have molecular weights of about 350,000, and are thus the largest single polypeptide chains which have been characterized to date. Amino acid analysis shows the similarity of these two polypeptides and establishes their identity with fibroin prepared in the classical manner from cocoons.

Sericin, in contrast, is a mixture of at least 15 different polypeptide chains, ranging in size from about 20,000 to 220,000. No attempt has been made to determine whether all of the sericin proteins isolated from the silk gland actually appear in the spun silk filament. However, since the amino acid composition of the three largest sericin polypeptides closely resembles that of sericin derived from boiled cocoons, these species, at least, are likely to contribute to the sericin layer of the spun fiber.

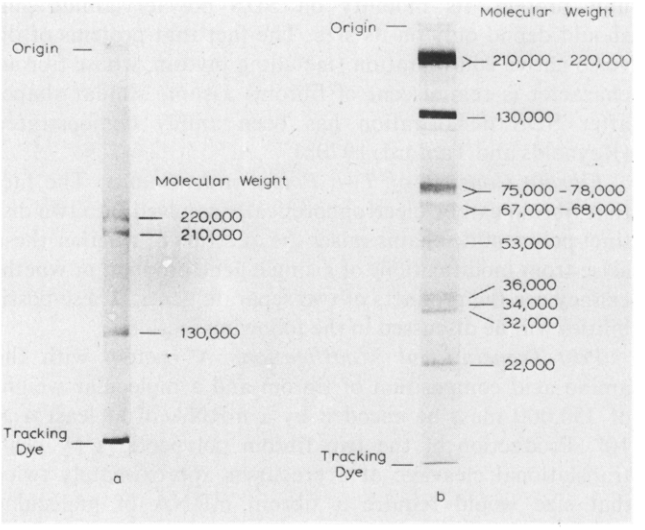

FIGURE 4: Separation of the protein components of sericin by SDS polyacrylamide gel electrophoresis on slabs of (a) 4% acrylamide, 0.1% bisacrylamide; (b) 8% acrylamide, 0.2% bisacrylamide.

The Molecular Weight of Fibroin. The value of 350,000 daltons obtained by molecular sieve techniques for the size of fibroin is not far from some of the previously reported estimates derived from sedimentation data (Lucas (1966), 400,000; Tashiro *et al.* (1972), 310,000). It seems likely that values differing considerably from 350,000 were obtained with either degraded or aggregated samples of protein. Such phenomena were in fact observed during the

course of this work. Gel electrophoretic analysis indicated that one of the traditional methods for solubilizing fibroin from cocoons (treatment with 2% Na_2CO_3, 100°, for up to 3 hr; Pandit *et al.*, (1972)) leads to severe degradation after only 30 min. Conversely, fibroin dissolved in water (Tashiro and Otsuki, 1970a) frequently undergoes irreversible aggregation within a few hours after preparation. Furthermore, I observed that when fibroin samples are complexed with SDS in the absence of β-mercaptoethanol, electrophoretic analysis sometimes reveals additional bands near the top of the gel. Because they appear in low and variable amounts, they probably represent aggregates or molecules which have bound inadequate amounts of SDS. Their existence suggests that previous results demonstrating an effect of reducing agents on the size of fibroin may have been artifactual. I have observed no effect of reducing agents on either the molecular weight of fibroin or the presence of two fibroin polypeptides.

A weakness of the fibroin molecular weight estimate derived here is its dependence on extrapolation from marker proteins with lower molecular weights. Even the use of incompletely dissociated protein to define the mobility *vs.* log molecular weight curves on the heavy side of fibroin is open to criticism because such proteins may complex unusual amounts of SDS and may not behave as random coils in concentrated guanidine solutions. However, the scarcity of appropriately large protein markers makes such procedures necessary.

Two types of data support the legitimacy of extrapolation in this case. (1) Relative to marker proteins (including undissociated myosin), the observed molecular weight of fibroin was unaffected by changes in gel concentration. Thus, fibroin size estimates based on polyacrylamide gel electrophoresis fall within the linear range of the gel system used here. (2) Since fibroin binds a normal amount of SDS per unit protein, its mobility on SDS polyacrylamide gels should depnd only on its size. The fact that proteins of diverse native configuration (including myosin, whose fibrous character is reminiscent of fibroin) assume similar shapes after SDS denaturation has been amply demonstrated (Reynolds and Tanford, 1970b).

Fibroin Consists of Two Polypeptide Chains. The fact that fibroin can be electrophoretically resolved into two distinct polypeptide chains raises the question of whether these arise from modifications of a single gene product, or whether they are the products of two separate genes. These possibilities will be discussed in the following sections.

Post-Translational Modification? A protein with the amino acid composition of fibroin and a molecular weight of 350,000 must be encoded by a mRNA of at least 4×10^6. Production of the two fibroin polypeptides by post-translational cleavage of a precursor approximately twice that size would require a fibroin mRNA of molecular weight at least 8×10^6. Since recent estimates based on electron microscopy and on gel electrophoresis in formamide place the molecular weight of fibroin mRNA at $5.5\text{–}6 \times 10^6$ (Lizardi and Brown, 1973), this model for the generation of two polypeptide chains seems unlikely. The fact that the fibroin mRNA molecule exceeds the length needed to encode a single fibroin polypeptide is not unexpected since similar relationships between mRNA and protein have been observed in both prokaryotic and eukaryotic systems: *Escherichia coli* tryptophan operon mRNA (Bronson *et al.*, 1973), immunoglobulin light chain mRNA (Brownlee *et al.*, 1973), and globin mRNA (Gaskell and Kabat, 1971) all contain substantial untranslated regions.

On the other hand, one of the fibroin polypeptides might arise through cleavage of the other or addition of a nonprotein moiety. If this occurs, the process must be restricted to half of the fibroin molecules and is not obviously correlated with stages in fibroin transport or storage. Both fibroin polypeptides appear on stained gels in approximately equal amounts which do not change when fibroin is obtained from functionally different regions of the gland. It is unlikely that the two polypeptides are generated by protease attack during sample preparation, since inclusion of a serine protease inhibitor (10^{-5} M phenylmethanesulfonyl fluoride) in all solutions which came in contact with fibroin did not alter the band pattern on gels.

Furthermore, post-translational addition of either a large carbohydrate or lipid seems unlikely. A maximum of one residue each of glucosamine and mannose per 1000 amino acid residues has been found in cocoon fibroin (Sinohara and Asano, 1967). Fibroin obtained from silk glands contains no detectable lipid (Tashiro and Otsuki, 1970a).

Two Fibroin Genes per Haploid Genome? This hypothesis has already been tested directly by measuring the amount of *B. mori* DNA coding for fibroin. The fibroin polypeptides should be encoded by at least 8×10^6 daltons of double-stranded DNA if there is one gene per haploid genome, or 16×10^6 if there are two genes. Data from hybridization saturation experiments (Suzuki *et al.*, 1972; Gage, personal communication) and from buoyant density analysis of DNA fragments containing variable proportions of high G-C fibroin genes (Lizardi and Brown, 1973) give double-stranded gene sizes of 13×10^6 and 6×10^6, respectively, and suggest that there is only one fibroin gene per haploid genome.

Multiple Alleles? The most likely explanation for the appearance of two fibroin polypeptides seems to be that there is one fibroin gene per haploid genome, but that it occurs in multiple allelic forms. It is quite possible that the original *Bombyx* strain I received from Japan was heterozygous for the fibroin locus, since most commercially available strains are hybrids. Thus, the two fibroin polypeptides may be the products of codominant fibroin alleles differing slightly in size. Preliminary electrophoretic analysis of the progeny of presumed heterozygotes reveals segregation of the alleles. The existence of fibroin variants may permit genetic mapping of the fibroin structural locus. Experiments in this direction are currently in progress.

Acknowledgments

I would like to thank J. A. Steitz for helpful discussions during the course of this work, and for her comments on the manuscript. I also thank J. L. Mouning, F. M. Richards, Y. Nakashima, and W. H. Konigsberg for their assistance with the amino acid analyses.

References

Bronson, M. J., Squires, C., and Yanofsky, C. (1973), *Proc. Nat. Acad. Sci. U.S. 70,* 2335.

Brownlee, G. G., Cartwright, E. M., Cowan, N. J. Jarvis, J. M., and Milstein, C. (1973), *Nature (London), New Biol. 244,* 236.

Burgess, R. R. (1969), *J. Biol. Chem. 244,* 6168.

Chessin, H., and Summers, W. C. (1970), *Biochem. Biophys. Res. Commun. 38,* 40.

Cramer, E. (1865), *J. Prakt. Chem. 96,* 76.

Edelman, G. M., Gall W. E., Waxdall, M. J., and Konigs-

berg, W. H. (1968), *Biochemistry 7,* 1950.
Fish, W. W., Mann, K. G., and Tanford, C. (1969), *J. Biol. Chem. 244,* 4989.
Gaskell, P., and Kabat D. (1971), *Proc. Nat. Acad. Sci. U.S. 68,* 72.
Gelotte, B. J. (1960), *J. Chromatogr. 3,* 330.
Gershman, L. C., Stracher, A., and Dreizen, P. (1969), *J. Biol. Chem. 244,* 2726.
Laemmli, U. K. (1970), *Nature (London) 227, 680.*
Lizardi, P. M., and Brown, D. D. (1973), *Cold Spring Harbor Symp. Quant. Biol. 38,* 701.
Loening, U. E. (1967), *Biochem. J. 102,* 251.
Lucas, F. (1966), *Nature (London) 210,* 952.
Lucas, F., Shaw, J. T. B., and Smith, S. G. (1958), *Advan. Protein Chem. 13,* 107.
Machida, J. (1927), *J. Coll. Agric., Tokyo Imp. Univ. 9,* 119.
Maizel, J. V. (1971), *Methods Virol. 5,* 179.
Masamune, Y., and Richardson, C. C. (1968), *Proc. Nat. Acad. Sci. U.S. 61,* 1328.
Rao, M. S. N., and Pandit, M. W. (1965), *Biochim. Biophys. Acta 94,* 238.
Reynolds, J. A., and Tanford, C. (1970a), *Proc. Nat. Acad. Sci. U.S. 66,* 1002.
Reynolds, J. A., and Tanford, C. (1970b), *J. Biol. Chem. 245,* 5161.
Sasaki, T., and Noda, H. (1973a), *Biochim. Biophys. Acta 310,* 76.
Sasaki, T., and Noda, H. (1973b), *Biochim. Biophys. Acta 310,* 91.
Schlesinger, M. J. (1964), *Brookhaven Symp. Biol. 17,* 66.
Seery, V. L., Fischer, E. H., and Teller, D. C. (1967), *Biochemistry 6,* 3315.
Sinohara, H., and Asano, Y. (1967), *J. Biochem. 62,* 129.
Sridhara, S., Prudhomme, J. C., and Daillie, J. (1973), *Arch. Biochem. Biophys. 156,* 168.
Studier, F. W. (1972), *Science 176,* 367.
Studier, F. W. (1973), *J. Mol. Biol. 79,* 237.
Suzuki, Y., and Brown, D. D. (1972), *J. Mol. Biol. 63,* 409.
Suzuki, Y., Gage, L. P., and Brown, D. D. (1972), *J. Mol. Biol. 70,* 637.
Szent-Györgyi, A. (1951), The Chemistry of Muscular Contraction, 2nd ed, New York, N.Y., Academic Press.
Tanford, C., Kawahara, K., and Lapanje, S. (1967), *J. Amer. Chem. Soc. 89,* 729.
Tashiro, Y., and Otsuki, E. (1970a), *J. Cell Biol. 46,* 1.
Tashiro, Y., and Otsuki, E. (1970b), *Biochim. Biophys. Acta 214,* 265.
Tashiro, Y., Otsuki, E., and Shimadzu, T. (1972), *Biochim. Biophys. Acta 257,* 198.
Tokutake, S., and Okuyama, T. (1972), *J. Biochem. 71,* 737.
Torriani, A. (1966), *Procedures Nucleic Acid Res. 1,* 224.
Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244,* 4406.
Weber, K., Sund, H., and Wallenfels, K. (1964), *Biochem. Z. 339,* 498.
Wetlaufer, D. B. (1962), *Advan. Protein Chem. 17,* 378.
Yamanouchi, M. (1922), *J. Coll. Agric. Hokkaido Imp. Univ. 10,* 1.

Analysis of Cooperativity in Hemoglobin. Valency Hybrids, Oxidation, and Methemoglobin Replacement Reactions†

Attila Szabo and Martin Karplus*

ABSTRACT: An allosteric model proposed previously for structure–function relations in hemoglobin is applied to the analysis of low- and high-spin valency hybrids. By assuming that the low-spin oxidized chains have the tertiary structure of oxygenated chains while the high-spin oxidized chains have a tertiary structure intermediate between that of deoxygenated and oxygenated chains, the model parameters associated with the different valency hybrids can be obtained, and their equilibrium properties can be estimated. The hybrid results are used also to provide an interpretation of methemoglobin and its ligand replacement reactions and of the oxidation–reduction equilibrium of normal hemoglobin. For the various systems studied, it is found that the effects of pH and 2,3-diphosphoglycerate are in agreement with the model.

To understand the mechanism of cooperative ligand binding by the hemoglobin tetramer, it is not sufficient to know the structure and properties of the completely deoxygenated (Hb) and fully oxygenated ($Hb(O_2)_4$) species. Information about the intermediates ($Hb(O_2)$, $Hb(O_2)_2$, $Hb(O_2)_3$) that occur in the course of the oxygenation reaction is required. Such knowledge is difficult to obtain in a highly cooperative system like hemoglobin because the equilibrium concentra-

† From the Institut de Biologie Physico-Chimique, Université de Paris VI, Paris 5e, France, the MRC Laboratory of Molecular Biology, Cambridge CG 2 2QH, England, and the Laboratoire de Chimie Théorique, Université de Paris VII, Paris 5e, France. *Received June 25, 1974.* Supported in part by grants from the National Science Foundation (GP36104X) and the National Institutes of Health (EY00062). A. Szabo was supported by a fellowship from the National Research Council of Canada.

* Address correspondence to Department of Chemistry, Harvard University, Cambridge, Mass. 02138.