

# Silkmoth Chorion Proteins

THEIR DIVERSITY, AMINO ACID COMPOSITION, AND THE NH<sub>2</sub>-TERMINAL SEQUENCE OF ONE COMPONENT\*

(Received for publication, July 5, 1977)

JEROME C. REGIER AND FOTIS C. KAFATOS

*From Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138*

KARL J. KRAMER

*From the United States Grain Marketing Research Laboratory, Agricultural Research Service, United States Department of Agriculture, Manhattan, Kansas 66502*

ROBERT L. HEINRIKSON AND PAMELA S. KEIM

*From the Department of Biochemistry, University of Chicago, Chicago, Illinois 60637*

Silkmoth eggshell (chorion) proteins have been characterized by electrophoresis on sodium dodecyl sulfate and isoelectric focusing polyacrylamide gels; up to 33 and 41 components, respectively, were detected from a single chorion. Some of these components are polymorphic, being absent from chorions of certain animals. A system of nomenclature for all chorion proteins is presented, based on their separation on sodium dodecyl sulfate and isoelectric focusing gels.

The chorion is enriched in glycine, alanine, cysteine, and tyrosine and poor in methionine and histidine. The proteins were fractionated into four partially overlapping groups; all four are enriched in the above amino acids, although significant differences exist.

Further fractionation by isoelectric focusing of one of the above groups, s/s, yielded seven components, two of which are homogeneous both on sodium dodecyl sulfate and isoelectric focusing gels. The amino acid compositions, molecular weights, and solubility properties of the components share certain features which distinguish s/s as a group from the other three groups.

The sequence of the first 67 NH<sub>2</sub>-terminal residues of a homogeneous protein purified from s/s has been determined. The protein contains a cysteine-rich tail (3 cysteines in the first 18 residues) followed by a 49-residue segment which contains only a single cysteine residue. This latter segment also contains two different tetrapeptide sequences which are each repeated, one twice and the other four times.

muscle fibers, microfilaments and microtubules, flagella and cilia, collagen networks, etc. A quite fruitful approach to a molecular understanding of the assembly process has been to isolate the components, to characterize them individually, and to recombine the components *in vitro*. For example, the assembly of collagen molecules *in vitro* into a variety of supramolecular forms under well defined conditions, combined with extensive sequence and x-ray diffraction data, has resulted in detailed models of self-assembly (1). More complicated types of assembly have been studied in systems where several types of protein interact (*e.g.* muscle (2)), where they must be added in a specific order (*e.g.* ribosomes (3)), where non-protein molecules are involved in the assembly (*e.g.* membranes (4)), or where other proteins aid in the construction, but are not part of the structure itself (*e.g.* bacteriophage P22 (5)).

As a long term project we have chosen the eggshell (chorion) of silkmths as a system to analyze the assembly of structural proteins. Future studies on the *in vitro* assembly of chorion proteins and their localization within the supramolecular structure by the use of antibodies will require purified proteins. In this report we describe the isolation and biochemical characterization of some chorion components.

The chorion is large (about 3 mm in diameter) and composed almost entirely of protein. The proteins are small, have unusual amino acid compositions, and can be conveniently solubilized (6, 7). The proteins are predominantly organized as fibers embedded in a matrix, suggesting an analogy with vertebrate keratins and other fiber-matrix systems (8). In addition, the helicoidal orientation of these fibers suggests similarities with arthropod cuticle and many other structural systems (9, 10). Thus, we assume that the principles manifested by chorion assembly will have some general relevance.

As this report and previous ones detail, the chorion is biochemically complex (7, 11). More than 40 distinct proteins are present in a single chorion. The work reported here lays the groundwork for detailed biochemical characterization of individual members of this protein family. A system of nomen-

Structural proteins interact to form a wide variety of supramolecular complexes, including cell membranes, ribosomes,

\* This work was supported predominantly by grants from National Science Foundation and National Institutes of Health to F. C. K. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

clature is established, taking into account fractionation by size and charge, and the existence of polymorphism for some of the proteins. Purification methods are established, and the partial or extensive purification of several proteins is described. These proteins are characterized by isoelectric focusing, SDS-gel electrophoresis, and amino acid composition. The proteins isolated are closely related to each other in terms of molecular weight, solubility, and amino acid composition. This suggests that some of the biochemical complexity may be the result of ancestral gene duplications followed by sequence divergence. The NH<sub>2</sub>-terminal sequence of two-thirds of one of these proteins is presented. The high glycine content and the presence of short, repeated segments make this protein similar to a number of other structural proteins, although its particular sequence is unique.

#### MATERIALS AND METHODS

**Collection of Silkmoth Chorion**—Pupae of the American silkmoth, *Antheraea polyphemus*, were purchased commercially. Laid eggs were either pooled from individual moths to document protein polymorphism or combined randomly for preparative purposes. Pure eggshell (chorion) protein was obtained by washing the eggs briefly in 0.01 M Tris/HCl (pH 8.4), cracking the eggs with a mortar and pestle, and washing away the oocyte with distilled water.

**Solubilization and Fractionation**—Solubilization of the chorion (except for a very minor insoluble fraction) occurs in a buffer containing a denaturing agent (guanidine-HCl or urea) and a reducing agent (dithiothreitol or mercaptoethanol). Solubilization occurs more rapidly at basic pH (7).

For electrophoresis, chorion proteins were solubilized at room temperature in 8 M urea, 0.36 M Tris/HCl (pH 8.4), and 0.03 M dithiothreitol to a concentration of 7 mg/ml. An aqueous solution of iodoacetamide (0.7 M) was added in subdued light at 15% molar excess over sulfhydryl groups, and after 15 min, excess mercaptoethanol was added. The carboxamidomethylated protein solution was applied directly to isoelectric focusing gels. For electrophoresis on SDS gels it was diluted 9-fold with a solution containing 2% SDS (BDH), 0.0626 M Tris/HCl (pH 6.8), 10% glycerol, 0.143 M 2-mercaptoethanol, and 0.002% bromphenol blue (12). Samples for SDS gels were boiled briefly immediately before electrophoresis.

For preparative isolation, proteins were solubilized in 3.0 M guanidine HCl, 0.36 M Tris/HCl (pH 8.4), 1.5 mM Na<sub>2</sub>EDTA, and 0.03 M dithiothreitol to a protein concentration of 25 mg/ml. The solution was dialyzed overnight at room temperature against 26 volumes of 14.3 mM 2-mercaptoethanol, using dialysis tubing which retained peptides larger than 3500 daltons (Spectrapor 3, Spectrum Medical Industries). The precipitate (p/-) and the supernatant (s/-) were separated by centrifugation.

The supernatant (s/-) (55% of starting material) was dialyzed against distilled water, lyophilized, and dissolved in 6 M guanidine HCl, 0.36 M Tris/HCl (pH 8.4), 1.5 mM Na<sub>2</sub>EDTA, and 0.03 M dithiothreitol to a protein concentration of 10 mg/ml. An aqueous solution of iodo[<sup>14</sup>C]acetamide (approximately 10 mCi/mmol, New England Nuclear; 5.4 mM; 50 μCi of <sup>14</sup>C/g of protein) was added. After 10 min the dithiothreitol concentration was increased to 0.06 M, and after 15 min a 15% molar excess of nonradioactive iodoacetamide over sulfhydryl groups was added in a 0.5 M solution. After 15 min more 2-mercaptoethanol and acetic acid were added to 1 and 7% (v/v), respectively. The carboxamidomethylated s/- was dialyzed twice in the dark, each time against 20 volumes of 0.3 M guanidine HCl, 14.3 mM 2-mercaptoethanol, and 5 mM acetic acid, for a total of 20 h. The supernatant (s/s) and precipitate (s/p) were separated by centrifugation, dialyzed against distilled water, and lyophilized. The specific radioactivity of s/s was typically about 50 cpm/μg.

The precipitate (p/-) (45% of the starting material) was redissolved and carboxamidomethylated as above, except that only non-radioactive iodoacetamide was used, and was fractionated by dialysis as for s/- to yield two fractions, p/s and p/p.

**Electrophoresis**—Preparative scale isoelectric focusing was performed in a 440-ml water-jacketed column (LKB, Sweden) at 4°. Approximately 400 mg of s/s were fractionated per column, in a 0 to 35% (w/v) sucrose gradient containing 6 M urea (deionized) and

typically 1% (w/v) pH 4 to 6 Ampholine carrier ampholytes (LKB) plus 0.1% (w/v) pH 3.5 to 10 ampholytes (LKB). Electrofocusing occurred at 800 V for several days, and fractions were collected over 8 h. Aliquots of each fraction were counted by liquid scintillation after addition of ScintiVerse (Fisher). Selected fractions were pooled and electrofocused again over a narrower pH range using prefractionated ampholytes, prepared as described in the instruction manual for the column. The narrowest range used was 0.5 pH unit. After electrofocusing, selected fractions were pooled, dialyzed, and lyophilized.

Analytical scale isoelectric focusing in a polyacrylamide slab gel was performed as described (13), using pH 4 to 6 Ampholine carrier ampholytes (2%, w/v, LKB). Electrofocusing occurred either in the long dimension of the gel at 7° and 1000 V for 15 h or in the short dimension at 800 V for 4 h. Approximately 120 and 60 μg of unfractionated chorion protein yielded the optimal staining pattern in the long and short dimensions, respectively.

Analytical scale SDS-polyacrylamide slab gel electrophoresis was performed by a modification<sup>2</sup> of Laemmli's procedure (described in Ref. 12). The separating gel contained 9.4% acrylamide (w/v), 0.6% N,N'-methylenebisacrylamide (w/v), 7 M urea (deionized), 0.375 M Tris/HCl (pH 8.9), and 0.5% SDS (w/v). The stacking gel contained 3.0% acrylamide, 0.08% N,N'-methylenebisacrylamide, 0.125 M Tris/HCl (pH 6.8), and 0.5% SDS. The glass plates containing the slab gel were clamped onto a cooling plate through which 15° water circulated. Electrophoresis was at 8.4 mA/cm<sup>2</sup> for 7 h. Approximately 15 μg of unfractionated chorion protein yielded the optimal staining pattern with Coomassie brilliant blue. Autoradiography of <sup>14</sup>C-labeled gels has been described elsewhere (13).

**Amino Acid Analysis**—Carboxamidomethylated chorion proteins were hydrolyzed for 24 and 72 h (also for 48 h for Table I) at 110° in 4 M methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole (14). Unmodified cysteine was never detected, only carboxymethylcysteine. For determination of glucosamine and galactosamine, proteins were also hydrolyzed for 6 h. Recoveries of tryptophan, threonine, serine, and galactosamine after 24 h of hydrolysis were 97%, 98%, 95%, and 78%, as determined from recoveries of known amounts of free amino acids and amino sugars after hydrolysis; all values listed are corrected. Values for isoleucine and valine continue to increase after 24 h of hydrolysis (approximately 5%), and were measured after 72 (or 48 and 72) h of hydrolysis. Analyses were performed on a Beckman 121 M amino acid analyzer using a single column. Glucosamine and galactosamine were determined with a separate program; they eluted separately between methionine and isoleucine using AA-20 resin (Beckman Instruments) and 0.20 M sodium citrate (pH 3.23, Beckman Instruments).

**Sequence Analysis**—Automated Edman degradation of [<sup>14</sup>C]carboxamidomethylated A4-cl was performed on two samples (400 and 500 nmol) in Beckman Sequencers (models 890B and 890C, the former updated). The first time Beckman's 1.0 M Quadrol program MKII-8 for residues 1 to 20, Beckman's fast protein DMAA program 041073 with double cleavage for residues 21 to 50 and Beckman's fast peptide DMAA program 071872 for residues 51 to 70 were used. The second time residues 1 to 86 were sequenced using a modified 0.1 M Quadrol program (15) as described by Terhorst *et al.* (16). Residues were identified after conversion in 1.0 M HCl at 80° for 10 min by combination of gas chromatography (with and without silylation, using N,O-bis(trimethylsilyl)acetamide, Pierce (17)), thin layer chromatography (18-20), liquid scintillation counting to detect [<sup>14</sup>C]carboxamidomethylated cysteine, conversion of the phenylthiohydantoin derivative to the free amino acid followed by amino acid analysis (21, 22) and a spot test for arginine (23). All methods except the last were performed on all residues. Subsequent tryptic peptide analysis has confirmed that the first 67 residues described here include only 1 arginine residue and 1 tryptophan.

The first 67 residues were positively identified by at least two of the above methods, except tryptophan at position 42, which was identified only by thin layer chromatography. There was at least a 2-fold molar increase in all residues identified, except at position 52 where the increase in glycine was only 50% due to the glycine residue at position 51. For one of the sequenator runs the recovery of valine at positions 1, 58, and 66 was 300, 26, and 15 nmol. For leucine the recoveries at positions 6, 21, 36, and 64 were 225, 110, 88, and 15 nmol. By residue 65 the overlap was approximately 90%. Identification was still possible, however, because no identical residues were adjacent to each other after position 52. For the other

<sup>1</sup> The abbreviation used is: SDS, sodium dodecyl sulfate.

<sup>2</sup> M. Koehler, unpublished results.

sequenator run the recovery of leucine at positions 6, 36, and 64 was 120, 20, and 9 nmol. Of the sequence determined in these runs, only that which is completely unambiguous is presented here. Positions 51 through 67 have been confirmed independently by automatic sequencing of a peptide fragment.<sup>3</sup>

## RESULTS

**Multiplicity of Chorion Proteins**—The chorion proteins of *Antheraea polyphemus* have previously been resolved on SDS-polyacrylamide gels into three major classes, called A, B, and C, in order of increasing molecular weight (7). The molecular weights of the proteins in these classes range from approximately 7,000 to 18,000. The proteins have been further divided into subclasses, A1, A2...B1, B2, etc., also in order of increasing molecular weight.

Improvements in the gel system have resolved more components, up to 33 from a single chorion (Fig. 1, Slot 1). A representative autoradiogram of *in vitro* labeled chorion proteins from different moths is shown in Fig. 1, Slots 2 to 8; subclasses are indicated for only Class A proteins, which are further characterized in this report. All five A subclasses contain multiple components, at least for some chorions.

The number of resolvable components is greater on isoelectric focusing gels, if a properly narrow pH range of carrier ampholytes is used. Approximately 80% by mass of the *A. polyphemus* proteins have isoelectric points between pH 4 and 6, and the remainder near pH 7. Representative isoelectric focusing electrophoretograms of chorion proteins from different females are shown in Fig. 2. The resolved components have been divided into 10 groups, labeled a, b,...j beginning at the basic end of the gel (Fig. 2). Within each group individual components have been numbered (e.g. d1, d2, d3, etc.) starting from the basic end. As many as 41 components have been resolved in a single chorion. As shown in Figs. 1 and 2 and discussed in the next section, some of the chorion proteins are polymorphic.

A wide variety of evidence indicates that the multiplicity of chorion proteins is due to multiplicity of structural genes, rather than post-translational modifications of a small number of primary products. Although this evidence is presented in detail elsewhere, it should be summarized here, in order of increasing directness, because of its importance in establishing the significance of this study.

1. Individual chorion proteins are synthesized *in vivo* during distinct development stages (7, 11, 26). Fig. 3A shows the newly synthesized proteins of the follicular cells which produce the chorion, at a few selected stages of choriogenesis. For much of this period, the proteins synthesized are almost exclusively chorion. It is clear that the synthesis of individual bands is temporally limited and distinct.

2. Labeling experiments of even very brief duration reveal a high multiplicity of chorion proteins, comparable to that shown by the accumulated proteins of the mature chorion. Fig. 3B shows such data from an experiment in which proteins were labeled for 55 s and chased for 1 min (a total time equal to that required for synthesis of individual chorion protein chains (7)). No change in the electrophoretic pattern of *in vivo* pulse-labeled proteins is detected on SDS-polyacrylamide gels during chase for periods extending up to 4 h (7, 26).

3. Silkmoth chorions are rendered insoluble by disulfide and noncovalent bonds; and nearly complete dissolution occurs upon addition of a denaturant and a mild reducing agent (6,

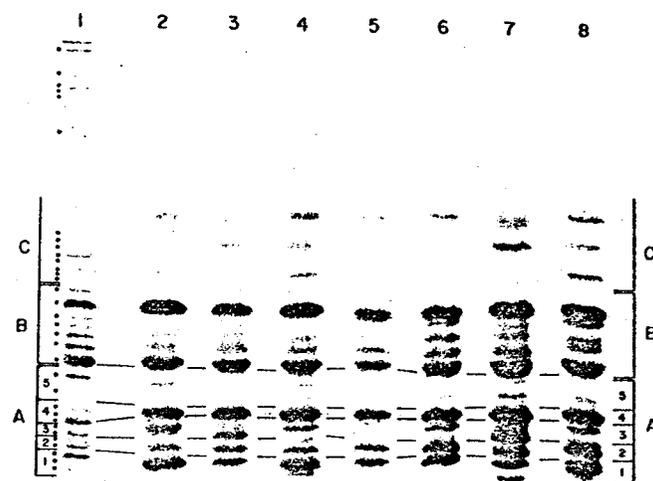


FIG. 1. Electrophoretic polymorphism on SDS-polyacrylamide gels of chorion proteins from different individuals of *Antheraea polyphemus*. The electrophoretograms are representative of the polymorphism shown by 40 different moths, and were obtained in two different experiments (Slots 1 and 2 to 8). Chorions from an almost mature egg (Slot 1) or from ovulated, unlaidd eggs (Slots 2 to 8) were dissolved, carboxamidomethylated (with iodo[<sup>14</sup>C]acetamide for Slots 2 to 8), electrophoresed on SDS-polyacrylamide slab gels, and stained as described under "Materials and Methods." Slot 1 shows the stained proteins from an aliquot of the entire eggshell. Slots 2 to 8 contain proteins from the aeropyle chimney region (26), which appears not to lack proteins found elsewhere in the chorion<sup>2</sup>; the proteins were detected by autoradiography. Bands are divided into classes and subclasses according to molecular weight (7). Solid horizontal lines between slots serve to define the subclasses of Class A proteins (which were indicated tentatively and incompletely in previous publications). Dots indicate all proteins resolved in the first slot; the high molecular weight components are consistently observed in eggshell preparations.

7). Although some insect eggshells are rendered insoluble by di- and trityrosine cross-links, no such cross-links are found in the solubilized silkmoth chorion proteins (46).

4. The chorion mRNAs have been purified and shown to be of a size expected for monocistronic messages encoding proteins in the chorion size range (43). These mRNAs have been resolved into multiple bands by gel electrophoresis, following removal of the variable length poly(A) tails (44).

5. Cell-free translation of purified chorion mRNAs in a heterologous wheat germ system has yielded specifically immunoprecipitable products of high multiplicity, comparable to that of *in vivo* labeled and bulk chorion proteins.<sup>4</sup> The products have electrophoretic patterns which can be related directly to those of authentic chorion proteins, although they are shifted to slightly higher molecular weight and more basic isoelectric point, apparently because of the presence of NH<sub>2</sub>-terminal "signal" peptides such as are generally found in secretory proteins.

6. Double-stranded chorion cDNAs have been synthesized using the mRNAs as template, and have been cloned in bacteria using a plasmid vector (45). Characterization of the multiple clones generated has led thus far to the identification of 19 distinct clones, each corresponding to a different mRNA sequence as shown by several criteria. The chorion insertion of two of these clones has been partially sequenced; the coding region of each specifies a different protein comparable in size

<sup>4</sup> G. Thireos, M. Nadel, and A. Efstratiadis, personal communication.

<sup>3</sup> J. C. Regier, unpublished results.

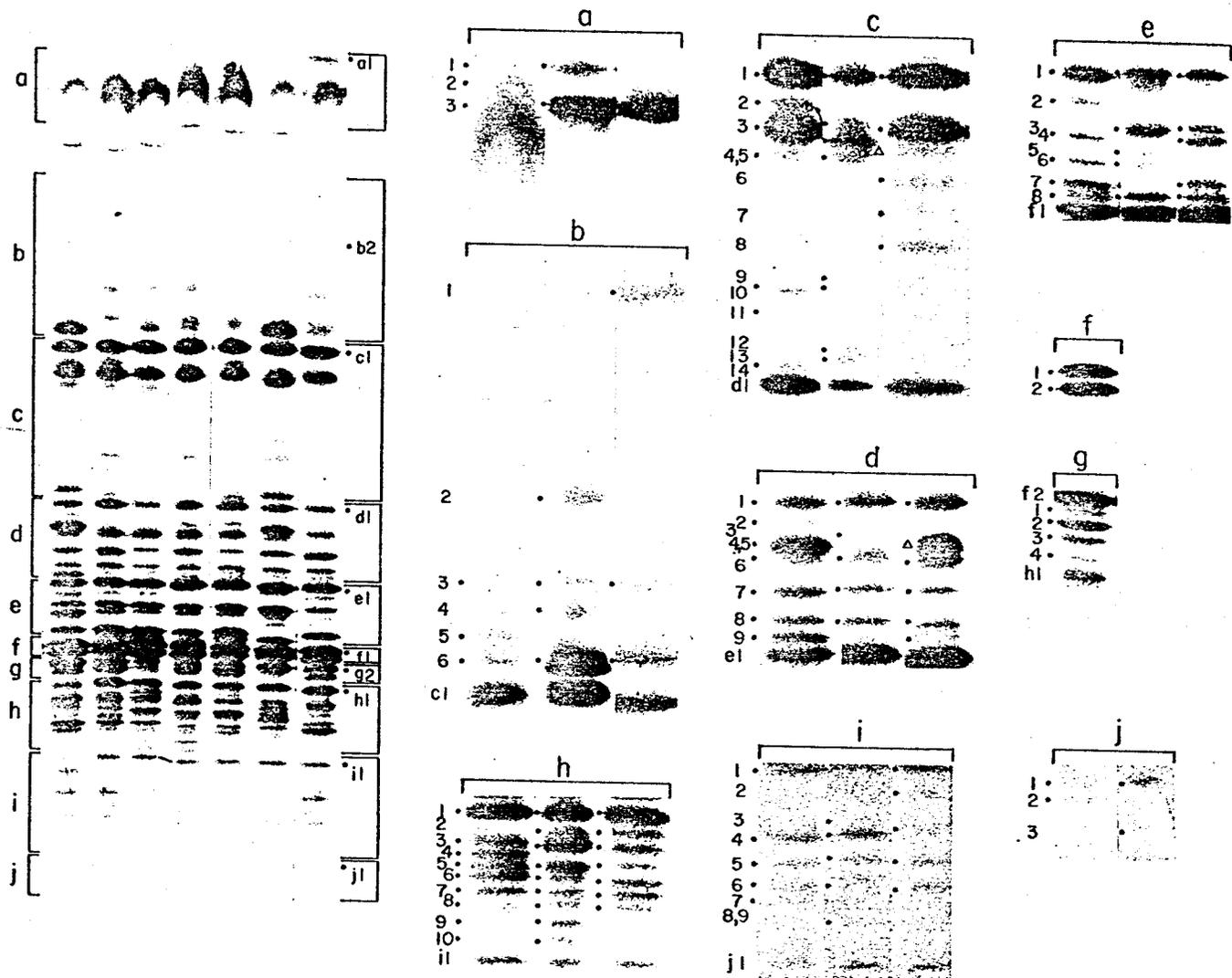


FIG. 2. Electrophoretic polymorphism on isoelectric focusing gels of chorion proteins from different individuals of *Antheraea polyphemus*. Chorions from ovulated laid and unlaid eggs were dissolved, carboxamidomethylated, electrophoresed on isoelectric focusing gels (ampholyte range pH 4 to 6), and stained as described under "Materials and Methods." Representative electrophoretograms of total chorion proteins from unlaid eggs are shown on the left. The proteins have been divided into 10 arbitrary groups (a to j, from basic to acidic end). Except for a and b, all groups were defined so as to begin with a locally prominent and invariant band. On the

right are shown individual groups at optimum resolution; all resolved proteins in each group are numbered, and those observable in each case are indicated by dots at the left of each slot. c4 and d4 are indicated by triangles, and are distinguished from c5 and d5 by their pink rather than blue color. i8 and i9 were resolved on a gel not displayed. A few of the very minor proteins may be localized in the vitelline membrane or the glue surrounding the laid eggshells, although preliminary fractionation of these two components has not revealed bands in the chorion region.

to authentic chorion proteins and characterized by the same unusual amino acid composition and distinctive repeating peptides as documented in the present report.

7. Partial sequences from 17 purified chorion proteins are available to date, beginning at the  $\text{NH}_2$  terminus for A proteins and at the  $\text{NH}_2$ -proximal methionine for B proteins (29).<sup>5</sup> All but one of these sequences are distinguished by unique primary structure, indicating that these proteins could not have arisen from each other through post-translational modification.

**Polymorphism of Chorion Proteins**—Extensive polymorphism of chorion proteins resolved by size is apparent in Fig. 1. Of the Class A proteins, A1 and A3 show the greatest variation. A3 proteins are almost completely missing in one

<sup>5</sup> G. C. Rodakis, J. C. Regier, and F. C. Kafatos, unpublished results.

chorion and seem quantitatively variable in the others. It is possible that quantitative differences reflect the complete absence of certain unresolved components, especially considering the greater multiplicity of proteins revealed by isoelectric focusing gels (Fig. 2). A4 is present in all chorions; under certain electrophoretic conditions it can be separated into three bands. The components in A5 are found in low and variable amounts but they appear to be present in all chorions.

Protein polymorphism has been observed only in chorions from different females. All chorions of the same moth appear identical. Class B and C proteins are also polymorphic, although the two major class B bands are present in all females. The patterns are completely repeatable in separate experiments, indicating that polymorphism and multiplicity are not generated by experimental variables.

Charge polymorphism is also apparent (Fig. 2). Although up to 41 proteins from a single chorion are evident on isoelec-

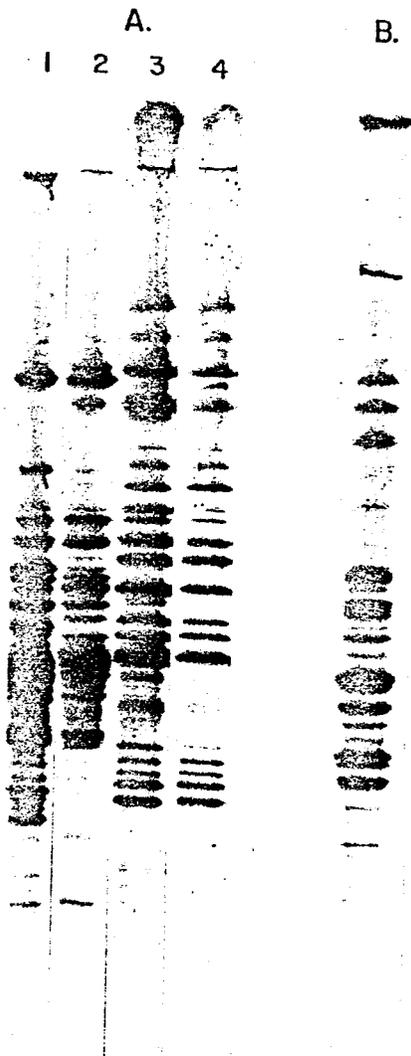


FIG. 3. A, multiplicity of chorion proteins, as shown by developmental changes in the synthesis of specific proteins. Follicles of different developmental stages (see Ref. 11 for definitions of stages) were labeled for 2 h in Grace's medium containing [ $^{14}$ C]leucine (14.3  $\mu$ Ci/ml, 0.29  $\mu$ mol/ml) followed by 1/2 h in Grace's medium containing nonradioactive leucine (0.57  $\mu$ mol/ml). After washing out the oocyte, follicles were dissolved, carboxamidomethylated, and electrophoresed on an isoelectric focusing gel (ampholyte range, pH 4 to 6). The gel was dried and placed over x-ray film. An autoradiogram of the resolved proteins from selected developmental stages is shown. 1, synthetic stage  $I_c^+$ ; 2, synthetic Stage III; 3, synthetic stage  $X_b$ ; 4, synthetic stage  $X_c$ . B, multiplicity of rapidly labeled proteins. A follicle of synthetic Stage V was labeled for 55 s in Grace's medium containing [ $^{14}$ C]leucine (50  $\mu$ Ci/ml, 0.16  $\mu$ mol/ml), [ $^{14}$ C]tyrosine (50  $\mu$ Ci/ml, 0.11  $\mu$ mol/ml), and [ $^{14}$ C]arginine (50  $\mu$ Ci/ml, 0.18  $\mu$ mol/ml) followed by 1 min in Grace's medium containing nonradioactive amino acids. The follicle was processed as for Part A.

tronic focusing gels, a total of 68 distinct components has been detected in chorions from 66 different females. All groups except for f contain polymorphic proteins. The most abundant components (a3, c1, e1, f1, f2) are not polymorphic, although some prominent components (e.g. e7 and i1) are.

The numbering systems presented in Figs. 1 and 2 will be used for identification of fractionated chorion proteins. Pure and fully characterized proteins will be named by a double code identifying them, in sequence, on SDS and isoelectric focusing gels (e.g. A4-c1; see below). In this code, capital

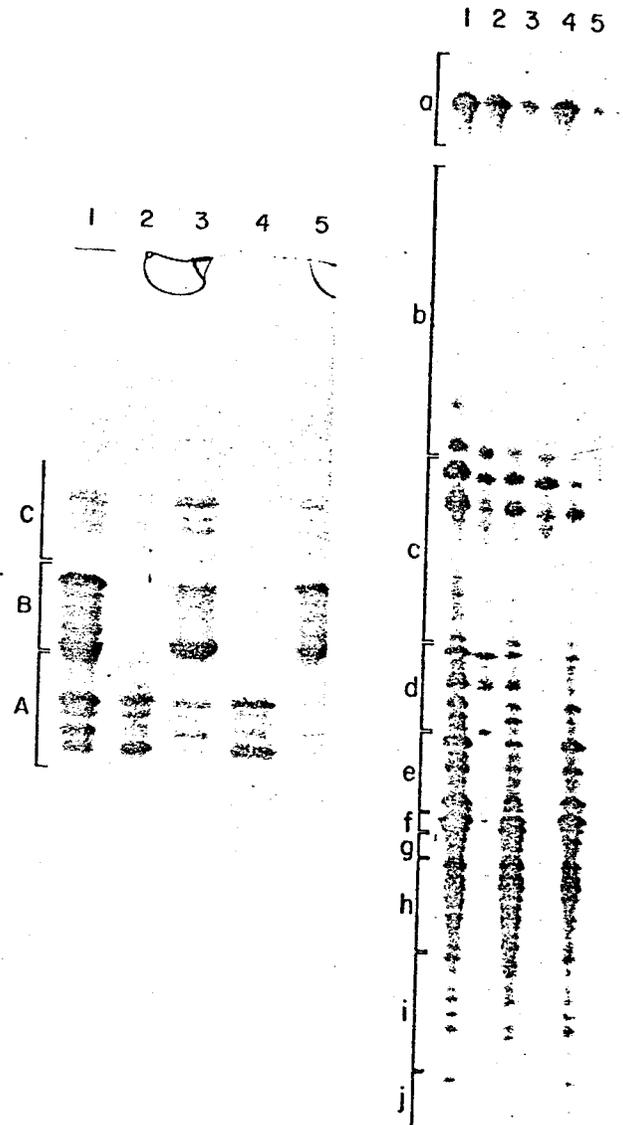


FIG. 4. Electrophoretograms of chorion proteins fractionated on the basis of differential solubility. See "Methods" for details. A, fractions analyzed on SDS-polyacrylamide gels. B, fractions analyzed on isoelectric focusing polyacrylamide gels. 1, unfractionated; 2, Fraction s/s; 3, Fraction s/p; 4, Fraction p/s; 5, Fraction p/p. Fractions s/s, s/p, p/s, and p/p contained 35%, 20%, 5%, and 40%, respectively, of the mass of the unfractionated starting material. Aliquots for electrophoresis were chosen to yield optimal staining (10 to 15  $\mu$ g on SDS gels and 60 of 120  $\mu$ g on isoelectric focusing gels).

letters indicate identification of SDS gels and lower case letters identification on isoelectric focusing gels. For fractions containing more than one protein, commas will be used to set the identifiable components apart (e.g. A3,5-d9 will indicate a fraction which corresponds to a single isoelectric focusing band, d9, but includes two proteins resolved by SDS-polyacrylamide gel electrophoresis, A3 and A5). Very minor contaminants will be omitted for simplicity, and an incomplete code will be used for partially characterized fractions (e.g. A--e7 will indicate a fraction corresponding to isoelectric focusing component e7 and including unidentified A size proteins). Finally, limited uncertainty of identification will be indicated by a short dash within either half of the code (e.g. A--e6-7 will indicate a fraction which contains A proteins and corre-

sponds to either e6 or e7 on isoelectric focusing gels). Limited uncertainty will remain in some cases, especially for identifications by isoelectric focusing, because of the close proximity of some components (Fig. 2), their absence in some animals, and the slight variations in electrophoretic conditions. Such uncertainty might be corrected by the use of additional criteria

TABLE I

Amino acid compositions of *Antheraea polyphemus* chorion protein fractions (residues/100 residues)

Amino acids	Unfractionated	Fraction s/s	Fraction p/s	Fraction s/p	Fraction p/p
Cysteine <sup>a</sup>	6.4	8.4	8.0	5.8	5.6
Aspartic acid <sup>b</sup>	3.7	2.5	2.4	4.5	4.0
Threonine	3.0	3.4	3.3	3.4	3.0
Serine	3.7	3.0	2.7	4.1	3.8
Glutamic acid <sup>b</sup>	4.5	4.2	4.2	4.4	4.7
Proline	4.4	4.1	3.6	5.2	4.5
Glycine	32.6	32.4	35.0	28.2	32.2
Alanine	12.1	13.8	12.6	13.6	11.4
Valine	6.5	7.2	7.1	6.5	6.1
Methionine	0.4	0.1	0.1	1.0	0.5
Isoleucine	3.8	3.8	3.7	4.7	3.6
Leucine	7.6	6.4	6.4	7.4	7.9
Tyrosine	6.4	6.2	6.7	5.8	6.9
Phenylalanine	1.4	1.0	1.0	1.6	1.6
Histidine	0.0	0.0	0.0	0.1	0.0
Lysine	0.5	1.0	0.8	0.6	0.4
Tryptophan	1.0	0.5	0.4	1.0	1.2
Arginine	2.3	1.8	2.0	2.2	2.4
Glucosamine	0.0	0.0			
Galactosamine	0.0	0.0			

<sup>a</sup> Determined as carboxymethylcysteine.

<sup>b</sup> Includes the amidic forms.

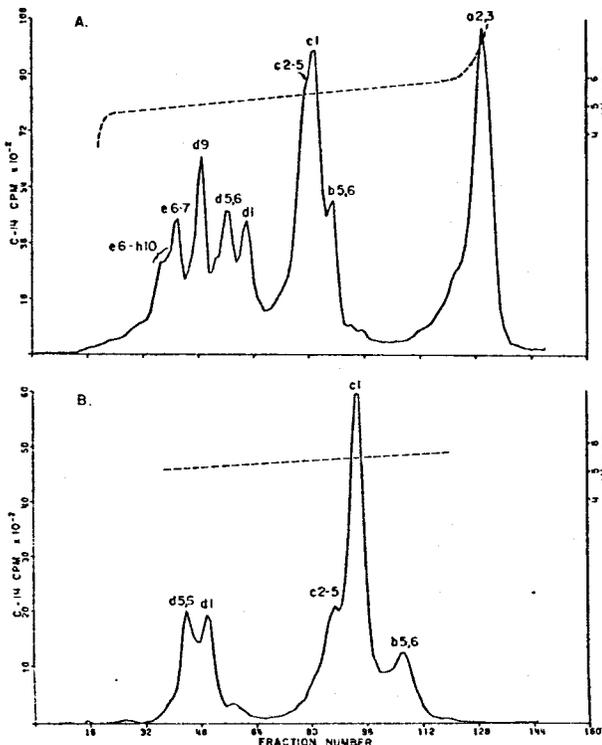


FIG. 5. Preparative scale isoelectric focusing of fractionated chorion proteins. Fraction s/s (Fig. 4) was electrophoresed in a sucrose gradient containing pH 4 to 6 carrier ampholytes (Panel A). Selected fractions, corresponding to Peaks c1 and d1, were pooled and electrophoresed over a narrower pH range (Panel B). Protein was detected by liquid scintillation counting of aliquots.

of identification, e.g. the developmental time of synthesis (11).

**Fractionation and Characterization of Chorion Proteins.**—As a first step toward purifying several chorion proteins, eggshells from a number of moths were combined, solubilized, and fractionated on the basis of differential solubility before and after carboxamidomethylation of the cysteine residues (see "Materials and Methods"). Four fractions (labeled s/s, s/p, p/s, and p/p) are displayed on SDS gels in Fig. 4A. Fraction s/s contains 90% of the Class A proteins by mass, no detectable Class B proteins, and only a very minor amount of Class C proteins. Fraction p/s is very similar to s/s, although it contains less than 10% of the total Class A proteins. Fractions s/p and p/p contain almost all of the Class B and C proteins. Fraction s/p is enriched for the major low molecular weight Class B protein, but contains only one-third of its total; s/p is also enriched for higher molecular weight Class C proteins.

The same four fractions were also displayed on isoelectric focusing gels (Fig. 4B). On the average, the proteins in s/s and p/s are more basic than those in s/p and p/p. This indicates that in *A. polyphemus* the A proteins are generally more basic than the prominent B proteins; the latter generally can be identified as major bands which are present in p/p but

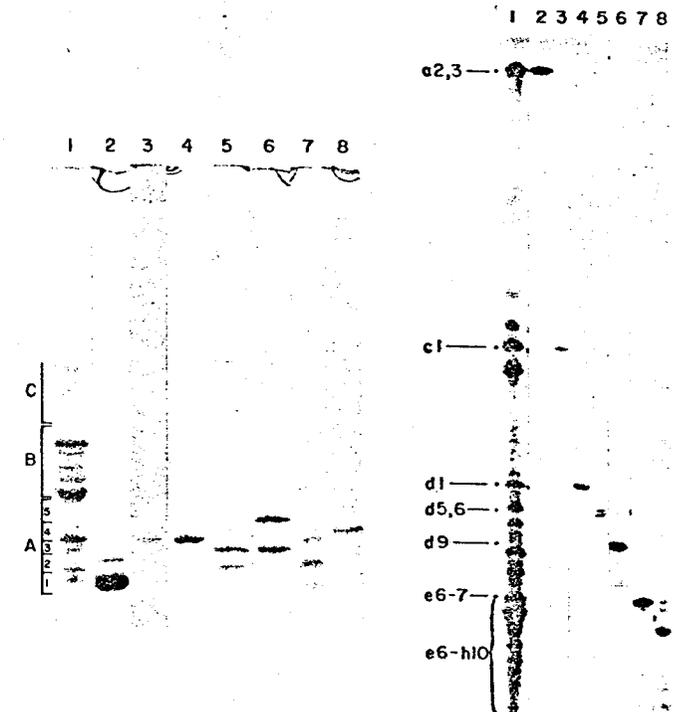


FIG. 6. Electrophoretograms of chorion proteins fractionated on the basis of differential solubility and isoelectric point. Protein fractions similar to those shown in Fig. 5B were electrophoresed on SDS-polyacrylamide gels (Panel A) and on isoelectric focusing gels (Panel B). Slot 1 shows unfractionated chorion. Slots 2 to 8 show fractions selected from a sucrose gradient isoelectric focusing column (basic to acidic end), after two focusing cycles; slots with the same number in the two panels correspond to the same sample. Slot 2, Fractions A1,1-a2; Slot 3, Fraction A4-c1; Slot 4, Fraction A4-d1; Slot 5, Fraction A2,3-d5,6; Slot 6, Fraction A3,5-d9; Slot 7, Fraction A1,2,3-e6-7; Slot 8, Fraction A3,4,5,B,C-e6-h10.

absent in s/s (cf. Fig. 4B).

Amino acid compositions have been determined (Table I) for the four fractions and for unfractionated chorion (minus a very minor insoluble fraction (6)). Unfractionated chorion proteins are unusually high in glycine (32.6%), alanine (12.1%), cysteine (6.4%), and tyrosine (6.4%). They are very low in methionine (0.4%) and histidine (<0.1%), and have undetectable amounts of the amino sugars, glucosamine and galactosamine. The fractions all share these unusual compositional features, but also are distinguished by some significant differences. Fractions s/s and p/s, which are quite similar to each other, are enriched for cysteine and depleted in methionine and tryptophan, relative to s/p and p/p. The remaining two fractions are more similar to each other than to s/s and p/s.

Fraction s/s was chosen for further fractionation by preparative isoelectric focusing because it is the most soluble fraction, its proteins are the smallest, and it contains in good yield a limited number of components which are well resolved by isoelectric focusing (Fig. 4B). A typical fractionation is shown in Fig. 5A. The partially resolved peaks were identified by comparison with Fig. 4B, and selected portions of the column were then pooled and refocused over a narrower pH range, and fractions were collected. A typical profile after refocusing is shown in Fig. 5B. Fractions were collected, their purity was assessed, and their identification established by electrophoresis on SDS and isoelectric focusing gels (Fig. 6). Two of the fractions appeared homogeneous (A4-c1 and A4-d1). One fraction contained two proteins of subclass A1, not resolved in Fig. 6A (A1,1-a2,3). Two of the fractions were homogeneous by isoelectric focusing but were resolved into two or three proteins by SDS-gel electrophoresis (A3,5-d9 and A1,2,3-e6-7). One fraction contained two proteins partially resolved on isoelectric focusing gels (A2,3-d5,6). Fi-

nally, one fraction contained predominantly one component (g2) by isoelectric focusing, and at least six others, which were apparently outside the fractionation range of the carrier ampholytes; on SDS gels this fraction contained predominantly one component (A4) although others were faintly visible (A3,4,5,B,C-e6-h10). The legend to Table II lists additional information on these fractions, including trace contaminants.

Amino acid compositions of the purified fractions are shown in Table II. These proteins are typical of unfractionated chorion proteins in being enriched for glycine, alanine, cysteine and tyrosine and poor in histidine and methionine. In addition, they share some characteristics of the total fraction s/s, being relatively enriched in cysteine and depleted in methionine (see Table I). However, each fraction is also distinct in composition. Significant differences in glycine content exist, largely correlated inversely with differences in alanine content. A1,1-a2,3 has high glycine, low alanine, aspartic acid + asparagine, serine, proline, leucine, and no tryptophan. A4-c1 is high in arginine and leucine. A4-d1 is typical of s/s. A2,3-d5,6 contains a small amount of histidine and methionine. A3,5-d9 is low in arginine. A1,2,3-e6-7 contains no arginine or tryptophan, is low in glycine and enriched in alanine, proline, and methionine. A3,4,5,B,C-e6-h10 is the most unusual. It contains galactosamine and a high amount of serine; it is also enriched in threonine, proline, methionine, and histidine; it completely lacks tryptophan and is low in arginine and in valine + isoleucine + leucine. The protein solution containing A3,4,5,B,C-e6-h10 is also light brown in color, while all others are transparent.

*NH<sub>2</sub>-terminal Sequence of Purified Chorion Protein*—The *NH<sub>2</sub>*-terminal 67 residues of A4-c1 are shown in Table III. SDS-gel electrophoresis, amino acid analysis, and preliminary sequencing of the COOH-terminal portion indicate that the

TABLE II  
Amino acid compositions of *Antheraea polyphemus* chorion proteins (residues/100 residues)

The proteins included in the seven fractions were purified and identified as discussed in the text; they are displayed in Fig. 5, Slots 2 to 8. In addition to the proteins indicated by the code of each fraction, very minor contaminants were present as follows: one A2 protein in A1,1-a2,3; one A4 and one Class C protein in A2,3-d5,6;

one A4 protein in A3,5-d9. Fraction A3,4,5,B,C-e6-h10 contained predominantly a g2 protein on isoelectric focusing gels and an A4 protein on SDS gels, although significant amounts of at least six other proteins were present on SDS gels.

	A1,1-a2	A4-c1	A4-d1	A2,3-d5,6	A3,5-d9	A1,2,3-e6-7	A3,4,5,B,C-e6-h10
Cysteine <sup>a</sup>	9.8	9.4	10.0	8.4	10.3	7.8	6.1
Aspartic acid <sup>b</sup>	1.3	3.6	2.1	3.3	2.9	2.6	3.2
Threonine	3.4	2.7	3.1	3.0	3.8	3.6	5.6
Serine	1.2	3.6	3.5	4.7	3.4	3.7	13.5
Glutamic acid <sup>b</sup>	4.0	3.6	4.5	4.3	3.1	4.1	3.2
Proline	1.9	3.6	4.0	4.4	4.8	8.0	7.9
Glycine	39.4	31.5	30.9	28.6	29.7	24.9	27.7
Alanine	10.8	12.9	14.1	14.7	15.3	18.1	11.5
Valine	7.3	7.1	7.3	7.3	6.9	7.7	6.0
Methionine	0.0	0.0	0.0	0.2	0.0	0.4	0.7
Isoleucine	4.4	3.5	4.0	4.4	3.8	3.8	2.8
Leucine	5.6	7.4	6.8	6.3	6.1	7.3	5.2
Tyrosine	6.9	6.1	5.4	5.5	6.2	5.9	5.0
Phenylalanine	1.0	0.9	0.9	1.2	1.0	1.1	0.6
Histidine	0.0	0.0	0.0	0.1	0.0	0.0	0.1
Lysine	0.9	0.9	1.0	0.7	0.9	1.0	0.7
Tryptophan	0.0	0.8	0.8	0.8	0.8	0.0	0.0
Arginine	2.1	2.5	1.7	2.0	1.0	0.0	0.3
Glucosamine							0.0
Galactosamine <sup>c</sup>	0.0	0.0	0.0	0.0	0.0	0.0	4.1

<sup>a</sup> Determined as carboxymethylcysteine.

<sup>b</sup> Includes the amidic forms.

<sup>c</sup> Expressed as molar per cent of total amino acids.

TABLE III

*NH<sub>2</sub>-terminal sequence of Antheraea polyphemus chorion protein A4-c1*

Solid and dashed lines indicate two types of tetrapeptide repeats, whereas dotted lines indicate the dipeptide repeats Val-Ala and Gly-Leu (the latter of which is also present in the three repeats of Gly-Leu-Gly-Tyr).

Val-Cys-Arg-Gly-Gly-Leu-Gly-Leu-Lys-Gly-Leu-Ala-Ala-Pro-Ala-Cys-Gly-Cys-Gly-Gly-Leu-Gly-Tyr-Glu-Gly-Leu-Gly-Tyr-Gly-Ala-Leu-Gly-	5	10	15	20	25	30
Tyr-Asp-Gly-Leu-Gly-Tyr-Gly-Ala-Gly-Trp-Ala-Gly-Pro-Ala-Cys-Gly-Tyr-Gly-Gly-Glu-Gly-Ile-Gly-Asn-Val-Ala-Val-Ala-Gly-Glu-Leu-Pro-	35	40	45	50	55	60
Val-Ala	65					

protein contains 112 residues. The sequence shown in Table III thus represents about three-fifths of the total.

There are several interesting features of this sequence. About 80% is composed of nonpolar amino acids (including glycine). Of the known residues, 51% are found in repeating sequences. The tetrapeptide sequence Gly-Leu-Gly-Tyr or the closely related tetrapeptide sequence Ala-Leu-Gly-Tyr is present in four tandem repeats, separated by 1 spacer residue (acidic in two out of three cases). In addition, a second tetrapeptide sequence Pro-Ala-Cys-Gly is found twice. Val-Ala occurs three times. Of the 4 basic residues of the molecule, two are found in the first 9 residues, none in the next 58. Of the 9 cysteine residues, three are in the first 18 residues, only one is found in the next 49 residues, and five occur in the COOH-terminal third of the protein. Eight leucine residues are within the sequenced segment, and of these six occur as Gly-Leu and one as Ala-Leu. Asymmetric distributions of several other types of residue can be inferred from the partial sequence and the amino acid composition of the protein (Table II).

## DISCUSSION

*Chorion Protein Multiplicity and Polymorphism*—In previous studies we have shown that silkmoth eggshells consist of multiple proteins (7, 11, 24). In this report we document further the considerable heterogeneity of the proteins on SDS-polyacrylamide gels and establish their even greater heterogeneity on isoelectric focusing gels. Moreover, we show that bands which appear single on isoelectric focusing gels often contain more than one component resolvable by SDS electrophoresis (e.g. d9), and *vice versa* (e.g. A4). Although the number of proteins present in an individual chorion cannot yet be determined exactly, a reasonable estimate would be of the order of 100.

The present report further establishes that many of the chorion proteins of *A. polyphemus* are polymorphic. In some cases this polymorphism appears to be quantitative, *i.e.* to correspond to significant differences in the abundance of certain proteins, whereas in other cases it appears to be truly qualitative. Of course, the distinction between quantitative and qualitative polymorphism is difficult to establish with certainty, especially since the protein profiles are so complex. For example, isoelectric focusing components which consist of multiple proteins would appear to be qualitatively polymorphic only if all constituent proteins were sometimes missing in the same chorion. This appears to be the case for A3,5-d9 and A1,2,3-e6-7 (Fig. 2). Two-dimensional gels will be necessary for further studies of polymorphism. With this qualification, it should be noted that many of the major bands resolved on isoelectric focusing gels (as well as on SDS gels) appear to be invariably present in the chorion.

From studies on enzymatically active proteins, it is now clear that a large proportion of eukaryotic genes are polymorphic (25). The observation of polymorphism in the chorion proteins is of special interest, because little is known as yet about the extent of polymorphism in structural proteins. At the same time, the polymorphism together with the multiplicity of proteins in any one chorion raise the issue of the molecular basis of the difference between resolvable chorion protein bands.

As outlined under "Results," it is clear that chorion proteins are encoded by a large number of distinct structural genes. On the basis of the evidence outlined, we believe that the number of genes is comparable to the number of protein bands resolved from mature chorion, even though the proteins are apparently modified post-translationally. One type of modification, which is probably general for all chorion components, appears to be the removal of  $NH_2$ -terminal signal peptides.<sup>4</sup> An additional modification affecting the charge has been detected for some proteins, including at least the major B proteins, but not for others, including the major A proteins (28). The modification is relatively slow ( $t_{1/2}$  approximately 17 min) and appears to shift the isoelectric point of a group of bands which are multiple from the time of synthesis, rather than to generate that multiplicity (28). The shift does not significantly alter the size of the affected proteins (7, 26), and may correspond to blocking of the  $NH_2$  terminus, as suggested by the failure of the Edman procedure applied to intact B proteins.

Irrespective of the origin of chorion protein multiplicity and polymorphism, the nomenclature established here is valuable in permitting identification of individual chorion components, with only limited ambiguity. The nomenclature for A proteins is fully established, whereas for B and C proteins the molecular weight subclasses remain to be defined, when the need arises.

*Chorion Protein Purification*—The large number of chorion proteins combined with their similar molecular weights and isoelectric points has made preparative scale purification by conventional means difficult. Even isoelectric focusing in a sucrose gradient using a narrow pH range of ampholytes has yielded only two fractions which are homogeneous both on isoelectric focusing and SDS gels. The methods used here, especially differential precipitation through dialysis, are useful for initial fractionation; moreover, preliminary characterization and sequencing can be performed effectively on mixtures of two proteins which can be isolated relatively easily by these methods (29). Considerable further purification can be attained by preparative isoelectric focusing on polyacrylamide gels.<sup>4</sup>

<sup>4</sup> G. Rodakis, personal communication.

**Protein Families**—A group of proteins has been isolated from fraction s/s; they have similar solubility properties (Fig. 3), molecular weights (Fig. 5A) and amino acid compositions (Table II). On this basis the proteins appear to be more similar to each other than to proteins in other chorion protein groups. This conclusion is reinforced by the extensive homology of all the proteins in this group which have thus far been sequenced partially (29). However, some chorion proteins appear to be substantially different than others, as shown by the unusual composition of A3,4,5,B,C-e6-h10 (Table II). In summary, most of the proteins in s/s appear to constitute a family of proteins which are closely related evolutionarily and may have similar functions. Such protein families are well established for hemoglobins (30), keratins (8, 31), bovine dental enamel proteins (32) and antibodies (33), and may be of quite general occurrence (33).

**Sequence of A4-c1**—Protein A4-c1 is an invariant and quantitatively major component of the chorion in *A. polyphemus* (Fig. 2). Its partial sequence shows certain features of considerable interest. One is the occurrence of tetrapeptide repeats. The presence of a single spacer residue between each (Gly/Ala)-Leu-Gly-Tyr repeat may suggest that the original repeating unit was a pentapeptide, subsequently modified by point mutations to yield the apparent tetrapeptide repeat. Short internal repeating peptides have been found in many proteins, including silk fibroin, collagen, keratins, "freezing point depressing" proteins, and protamines (34). The two repeating sequences in the first chorion protein sequenced (Table III), Gly-Leu-Gly-Tyr and Pro-Ala-Cys-Gly, have not been reported as repeating units in other proteins, although Gly-Ala-Gly-Tyr and Ala-Ala-Gly-Tyr repeats are found in *Bombyx mori* silk fibroin (35, 36). The significance of these repeats in chorion is unknown, but as in silk fibroin, they may result in the formation of a repeating three-dimensional structure important for fiber formation. The sequences Ile-Gly-Gly-Tyr and Gly-Leu-Ile-Tyr are found in a bacteriophage coat protein; these tetrapeptides are nonrepeating but appear to form part of an extended  $\alpha$ -helical region which probably participates in filament formation (37). The reiteration of certain dipeptides (Gly-Leu, Val-Ala) in protein A4-c1 is also notable. Finally, certain amino acids show strikingly asymmetric distribution: the NH<sub>2</sub> terminus is relatively enriched in cysteine and the basic amino acids, arginine and lysine, whereas the internal region of the sequence is more enriched in hydrophobic and acidic amino acids.

**Structure-Function Analysis**—Chorion proteins from a variety of silkmooths appear to be related (6, 24, 38). They are small, having an average molecular weight of 11,500 in *A. polyphemus* (7). They are quite nonpolar, having an average of approximately 69% nonpolar residues in *A. polyphemus* (Table I); and they show higher frictional coefficient ratios than a number of globular proteins (6, 39), consistent with the fibrous ultrastructure of the chorion (40).

We have chosen to study in detail the structure of the chorion of the silkmooth *A. polyphemus*. Through a combination of protein sequencing, protein localization, self-assembly, and ultrastructural studies currently in progress, we plan to study how much of the ultrastructure of the morphologically complex chorion (26) can be explained in terms of the primary structure of its components.

The large number of distinct proteins in a single chorion raises an interesting question as to whether all proteins have distinct functions, as for bacteriophage T4 structural proteins (41); whether the proteins have partially overlapping func-

tions, as for antibodies (33); or whether some of the proteins have identical functions, as may be the case for wool keratin proteins (8, 31). The analogy with wool keratins is at present quite striking. Both wool and chorion serve as moderating barriers between the environment, on the one side, and an organism on the other. Both are composed of a large number of proteins which have similarly small molecular weights. Both are composed predominantly of fibers embedded in a matrix. And both contain relatively cysteine-rich proteins.

The matrix of wool is composed of exceedingly cysteine-rich (25 molar per cent) proteins. The matrix of the chorion of the commercial silkmooth *Bombyx mori* also includes a group of cysteine-rich proteins, which contain up to 37 molar per cent cysteine (6).<sup>3</sup> However, *A. polyphemus* only appears to contain proteins with up to approximately 10 molar per cent cysteine. In this species, the moderately cysteine-rich proteins belong to the molecular weight Class A and are about 50% richer in cysteine than the remaining, mostly class B proteins. By analogy with wool, it may be suggested that A proteins may form the matrix. Alternatively, in *A. polyphemus* chorion different portions of the same molecule may form matrix and fiber, as appears to be the case for an avian feather keratin (42). The localization of 3 cysteines in the first 18 residues of A4-c1 and of only 1 cysteine in the next 49 residues suggests that this may be the case.

**Acknowledgments**—We are grateful to M. Tullis for technical assistance and to M. Koehler and M. Randell for help with preparation of the manuscript.

#### REFERENCES

1. Doyle, B. B. (1975) *J. Mol. Biol.* 91, 79-99
2. Weber, A., and Murray, J. M. (1973) *Physiol. Rev.* 53, 612-673
3. Nomura, M. (1973) *Science* 179, 864-873
4. Bretscher, M. S., and Raff, M. C. (1975) *Nature* 258, 43-49
5. Casjens, S., and King, J. (1974) *J. Supramol. Struct.* 2, 202-224
6. Kawasaki, H., Sato, H., and Suzuki, M. (1971) *Insect Biochem.* 1, 130-148
7. Paul, M., Goldsmith, M. R., Hunsley, J. R., and Kafatos, F. C. (1972) *J. Cell Biol.* 55, 653-680
8. Fraser, R. D. B., MacRae, T. P., and Rogers, G. E. (1972) *Keratins. Their Composition, Structure and Biosynthesis*, Charles C Thomas, Springfield, Ill.
9. Neville, A. C. (1975) *Biology of the Arthropod Cuticle*, Springer-Verlag, New York
10. Bouligand, Y. (1972) *Tissue & Cell* 4, 189-217
11. Paul, M., and Kafatos, F. C. (1975) *Dev. Biol.* 42, 141-159
12. Maizel, J. V., Jr. (1972) *Methods Virol.* 5, 180-246
13. Efstratiadis, A., and Kafatos, F. C. (1976) *Methods Mol. Biol.* 8, 1-124
14. Simpson, R. J., Neuberger, M. R., and Liu, T.-Y. (1976) *J. Biol. Chem.* 251, 1936-1940
15. Brauer, A. W., Margolies, M. N., and Haber, E. (1975) *Biochemistry* 14, 3029-3035
16. Terhorst, C., Parham, P., Mann, D. L., and Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. U. S. A.* 73, 910-914
17. Pisano, J. J., and Bronzert, T. J. (1969) *J. Biol. Chem.* 244, 5597-5607
18. Jeppsson, J.-O., and Sjöquist, J. (1967) *Anal. Biochem.* 18, 264-269
19. Summers, M. R., Smythers, G. W., and Oroszlan, S. (1975) *Anal. Biochem.* 53, 624-628
20. Kulbe, K. D. (1974) *Anal. Biochem.* 59, 564-573
21. Smithies, O., Gibson, D., Fanning, E. M., Goodflesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971) *Biochemistry* 10, 4912-4921
22. Mendez, E. and Lai, C. Y. (1975) *Anal. Biochem.* 68, 47-53
23. Itano, H. A., and Yamada, S. (1972) *Anal. Biochem.* 48, 483-490
24. Paul, M., Kafatos, F. C., and Regier, J. C. (1972) *J. Supramol. Struct.* 1, 60-65
25. Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary*

- Change, Columbia University Press, New York
26. Kafatos, F. C., Regier, J. C., Mazur, G. D., Nadel, M. R., Blau, H. M., Petri, W. H., Wyman, A. R., Gelinas, R. E., Moore, P. B., Paul, M., Efstratiadis, A., Vournakis, J. N., Goldsmith, M. R., Hunsley, J. R., Baker, B. K., Nardi, J., and Koehler, M. (1977) in *Results and Problems in Cell Differentiation* (Beermann, W., ed) Vol. 8, pp. 45-145, Springer-Verlag, Berlin
  27. Blobel, G., and Dobberstein, B. (1975) *J. Cell Biol.* 67, 835-851
  28. Regier, J. C. (1975) Ph.D. thesis, Harvard University
  29. Regier, J. C., Kafatos, F. C., Goodflesh, R., and Hood, L. (1978) *Proc. Natl. Acad. Sci. U. S. A.*, in press
  30. Kabat, D. (1972) *Science* 175, 134-140
  31. Swart, L. S. (1973) *Nature New Biol.* 243, 27-29
  32. Eggert, F. M., Allen, G. A., and Burgess, R. C. (1973) *Biochem. J.* 131, 471-484
  33. Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975) *Annu. Rev. Genet.* 9, 305-353
  34. Ycas, M. (1972) *J. Mol. Evol.* 2, 17-27
  35. Lucas, F., Shaw, J. T. B., and Smith, S. G. (1957) *Biochem. J.* 66, 468-479
  36. Lucas, F., Shaw, J. T. B., and Smith, S. G. (1962) *Biochem. J.* 83, 164-171
  37. Nakashima, Y., Wiseman, R. L., Konigsberg, W., and Marvin, D. A. (1975) *Nature* 253, 68-71
  38. Kawasaki, H., Sato, H., and Suzuki, M. (1972) *Insect Biochem.* 2, 53-57
  39. Tanford, C. (1961) *Physical Biochemistry of Macromolecules*, p. 359, John Wiley and Sons, Inc., New York
  40. Smith, D., Telfer, W. H., and Neville, A. C. (1971) *Tissue & Cell* 3, 477-498
  41. Wood, W. B., Edgar, R. S., King, J., Lielausis, I., and Henninger, M. (1968) *Fed. Proc.* 27, 1160-1166
  42. O'Donnell, I. J. O. (1973) *Aust. J. Biol. Sci.* 26, 415-437
  43. Gelinas, R. E., and Kafatos, F. C. (1973) *Proc. Natl. Acad. Sci. U. S. A.* 70, 3764-3768
  44. Vournakis, J. N., Efstratiadis, A., and Kafatos, F. C. (1975) *Proc. Natl. Acad. Sci. U. S. A.* 72, 2959-2963
  45. Sim, G. K., Efstratiadis, A., Jones, W. C., Kafatos, F. C., Kronenberg, H. M., Koehler, M., Maniatis, T., Regier, J. C., Roberts, B. F., and Rosenthal, N. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, in press
  46. Kawasaki, H., Sato, H., and Suzuki, M. (1974) *Insect Biochem.* 4, 99-111