J Mol Evol (1983) 19:322–332

Journal of Molecular Evolution

The B Multigene Family of Chorion Proteins in Saturniid Silkmoths

George C. Rodakis[1], Nikos K. Moschonas[1], Jerome C. Regier[2], and Fotis C. Kafatos[1]

1 Cellular and Developmental Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
Department of Biology, University of Athens, Panepistemiopolis, Kouponia, Athens 621, Greece
2 Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, IL 60201, USA

Summary The main features of the B family of chorion proteins in saturniid moths were examined by partial sequencing of representative B proteins, seven from *Antheraea polyphemus* and two from *A. pernyi*. Comparisons were made to sequences derived from seven recombinant DNA clones representing three types of B family proteins of *A. polyphemus*. The central regions of the sequences are conservative, both within and between moth species, and differ largely by a few amino acid replacements, rather than deletions or insertions. By contrast, the amino-terminal third varies more substantially, in a manner which defines two protein subfamilies: within each subfamily sequences are similar, but the subfamilies differ by at least two multiresidue deletions as well as by amino acid replacements. These properties are analogous to features of the A family of chorion proteins.

Key words: Eggshell – Protein domains – Repetitive sequences – Gene families – Chorion – Silkmoths

Introduction

It is now clear from sequencing studies that silkmoth chorion proteins are encoded by families of evolutionarily related genes, i.e. multigene families (Regier et al. 1978a; Rodakis et al. 1982; Jones and Kafatos 1982). Two-dimensional electropherograms of pure chorion from *Antheraea polyphemus* have revealed a large number of polypeptide spots, of which 26 have been assigned to the A family, 42 to the B family, and 41 to the C family (Regier et al. 1980). The A and B families are quantitatively major, together accounting for 88% of the protein mass (Kafatos et al. 1977). Analysis of cloned chorion cDNAs and genes has indicated that the families may be even more complex than is apparent at the protein level. For example (Jones and Kafatos 1980a), approximately 15 "copies" per genome exist for each of two gene "types", named 401 (B family) and 18 (A family). These multiple "copies" are highly similar although not identical (less than 5% mismatched; Jones and Kafatos, 1980b); many of them undoubtedly correspond to the same protein spot resolved in two dimensional gels, although some, like different gene types, might encode different protein spots.

The characteristic features of the A protein family have been established by both DNA and protein sequencing (Regier et al. 1978a; Tsitilou et al. 1980; Rodakis et al. 1982; Jones and Kafatos 1982). By contrast, the only B family sequences published include one partial protein sequence (Jones et al. 1979) and complete sequences of three types of genes (a total of 7 copies; Jones and Kafatos 1980b, 1982). Since the B family encodes 42 electrophoretically distinguishable components (Regier et al. 1980), it would be desirable to have sequence information for a more representative group of B proteins. Here we report proteins of the B family in *A. polyphemus*, representative of the range of molecular weights and isoelectric points encountered in the family. We also present corresponding information on two components from *A. pernyi*, permitting preliminary interspecies comparisons. Part of this work was presented earlier in Greek as doctoral dissertations (Rodakis 1978; Moschonas 1980).

Materials and Methods

Materials and Buffers

Two buffers were used in protein isolation and characterization: G-buffer (6 M guanidine-HCl, 0.36 M Tris-HCl pH 8.4, 1 mM

Offprint requests to: Fotis C. Kafatos

Na_2EDTA) and I-buffer (8 M urea, 0.05 M Tris-HCl pH 9, 1.5 mM Na_2EDTA and 1.5 mM lysine). Use of dialysis tubing with small pores (3500 dalton cutoff, Spectrum Medical Industries) was found to be important for retaining the smaller chorion proteins. Pyroglutamate aminopeptidase was obtained from Boehringer Mannheim. Chemicals for amino acid analysis and protein sequencing were obtained from Beckman and Pierce Chemical Co.

Pupae of *A. polyphemus* and *A. pernyi* were obtained from commercial U.S. growers. *A. polyphemus* is a semi-wild population with considerable polymorphism (Regier et al. 1978b). *A pernyi* is a moderately inbred population originally obtained from the Japanese sericulture industry, and shows limited polymorphism: in an isoelectric focusing study of chorion proteins from 100 moths only three distinct patterns were observed (Moschonas 1980). Pupae were allowed to undergo adult development upon removal from 4°C and chorions were purified from ovulated follicles as previously described (Paul et al. 1972).

Protein Isolation

As detailed below, chorion proteins were pre-fractionated by a variety of procedures and finally isolated by isoelectric focusing in gels. Their purity was assayed by SDS-polyacrylamide gel electrophoresis (Efstratiadis and Kafatos 1976). According to our standard convention (Regier et al. 1978b), proteins are identified by a compound code of letters and numbers separated by two dashes, e.g. B6--f1: the first half of the code corresponds to size fractionation (B being the family and 1 to 6 subgroups from lower to higher molecular weight), and the second half of the code corresponds to isoelectric focusing (a to j being charge subgroups and 1 to 11 indivudual components in each subgroup, from basic to acidic, respectively). Routinely, the sulfhydryl groups were blocked and the proteins labelled by S-carboxamidomethylation with [^{14}C] iodoacetamide (for conditions, see Regier et al. 1978b and Rodakis 1978).

Series 0 (Butanol Fractionation). A. polyphemus chorions (303 mg) were dissolved at a concentration of 3 mg/ml in G-buffer containing 0.58 M dithiothreitol (DTT). The pH was adjusted to 6.5 with acetic acid and an equal volume of *n*-butanol was added at 5 to 7°C. The mixture was vortexed, allowed to stand overnight at room temperature, and the phases were separated by centrifugation. The organic phase was discarded, and the aqueous phase was separated from the protein precipitate at the interphase. The aqueous phase contained 25% (by mass) of the protein and was highly enriched in A's, most of the C's, and some low molecular weight B's; the interphase contained the remaining 75% and was enriched in most B's and a few C's. The interphase was redissolved in I-buffer containing DTT and was carboxamidomethylated.

The CAM-proteins were fractionated on a preparative isoelectric focusing gel (16 x 27 x 0.15 cm), containing 5% acrylamide, 0,3% N', N'-methyenebisacrylamide, 6 M urea and 1.5 to 2% ampholine carrier ampholytes (LKB, pH range 4 to 6). Approximately 3 mg of protein in I-buffer was applied to a 1 x 15 cm strip of Whatman 3 MM paper positioned across the short dimension of the gel 1 to 2 cm from the negative electrode strip. Electrophoresis was at 25 mA until the voltage reached 750 V (approximately 1 hr), was continued at constant 750 V for 16 hr and was terminated after 1 hr at 1,000 V. Protein bands were visualized with oblique transillumination after brief immersion in 40% saturated ammonium sulfate, a concentration high enough to precipitate chorion proteins but low enough to keep the background clear and prevent gel shrinkage. Selected bands were excised with a razor blade. At room temperature, the excised band was extensively washed with 40% saturated ammonium sulfate to remove ampholytes, rinsed with distilled water, homogenized, mixed with an equal volume of G-buffer, vortexed for 10 min and centrifuged at 27,000 g. The supernatant was collected and the gel particles re-extracted at least 3 times. Pilot experiments with radioactive protein showed over 90% recovery under these conditions (Rodakis 1978). The proteins were precipitated from the combined supernatants with 90% saturated ammonium sulfate and pelleted at 7,700 g, presumably leaving gel contaminants in suspension. The pellet was dissolved in G-buffer, extensively dialyzed against distilled water, and lyophilized. No significant contaminants were detected by subsequent amino acid analysis or sequence determination.

Series II (Fractionation by Dialysis at Low Ionic Strength). Chorions were prepared from 10 *A. polyphemus* moths selected according to chorion protein electropherograms so as to minimize polymorphism in the bands of interest (Regier et al. 1978b). Chorions totaling 1 g were dissolved, carboxamidomethylated, dialyzed against water, lyophilized and redissolved to 3.33 mg/ml in G-buffer. This sample was then redialyzed against water, and the precipitate (P, 75% of the protein) was separated from the supernatant by centrifugation. The protein and solute concentrations during dialysis were found to be crucial for repeatable precipitation results, thus necessitating the second dialysis step. The supernatant (S) was lyophilized, redissolved in G-buffer at 15 mg/ml, and refractionated by adding ammonium sulfate to 10% of saturation at room temperature; the new precipitate (S/P) contained 11% and the new supernatant (S/S) 14% of the original protein. All three samples were subjected to fractionation on preparative isoelectric focusing gels, and selected bands were excised and processed as described above.

Series IV (Fractionation on a BioGel Column). Chorions were prepared from unselected *A. polyphemus* individuals, washed thoroughly with 4 M guanidine-HCl, 50 to 100 mM Tris-HCl (pH 7.5) buffer, dissolved and carboxamidomethylated. The sample (0.7 g CAM-protein) was fractionated on a Pharmacia 5 x 90 cm column packed with BioGel-P150 (100–200 mesh, Bio-Rad Laboratories) for 4 to 6 days under 25 to 30 cm hydrostatic pressure using 7 M guanidine-HCl, 50 mM Tris-HCl (pH 9.0) as running buffer. Selected fractions were combined, dialyzed against water, lyophilized, fractionated and recovered from preparative isoelectric focusing gels as above.

Proteins from A. pernyi. Chorions from individuals with the predominant protein pattern were combined, dissolved and carboxamidomethylated. Two size fractionation experiments similar to series IV were performed; details have been reported previously (Rodakis et al. 1982). A third fractionation experiment, analogous to series II, used a chorion protein concentration of 3 mg/ml, and yielded the following fractions: P, 86%; S, 14%.

Amino Acid Analysis and Protein Sequencing

Amino acid and sequence analyses were generally performed as previously described (Regier et al. 1978b; Jones et al. 1979; Rodakis et al. 1982). Lyophilized, approximately 5nmole protein examples were hydrolyzed under vacuum in 5.7 N HCl containing 0.2% mercaptoacetic acid, for 24 hr at 110°C. For estimations of tryptophan and CAM-cysteine, hydrolysis was in the presence of 4 M mercaptoethanesulphonic acid and 0.2% tryptamine. Compositions were determined in a Beckman 121 M amino acid analyzer, using 2 hr dual column programs and/or 3 hr single column programs, according to the Spinco DS-475 manual. Values after 24 hr methanesulfonic acid hydrolyses were corrected for destruction of Trp, Thr and Ser (3%, 2% and 5% respectively), and for incomplete hydrolysis of Ile and Val (5% each). Values presented are averages of at least two runs.

Sequences were determined with a Beckman 890 B sequencer, using 100–400 nmole protein samples and the modified 0.1 M Quadrol program as described by Terhorst et al. (1976). Residues were identified as previously described, after conversion, thin layer chromatography, gas chromatography, and high pressure liquid chromatography. For selected samples, back hydrolysis, spot tests for arginine and histidine, and measurements of radioactivity (for ^{14}C-CAM-cysteine) were used. Prior to sequencing, B proteins were digested with cyanogen bromide (Gross 1967; Witkop 1968; Spandle et al. 1970).

Additional sequencing runs of B6–f1 (series I and III) have been published previously (Jones et al. 1979), as has the series II sequence of B6–f1 which is reported here with corrections at positions 74 and 100.

Peptide Analysis

Cyanogen bromide fragments or tryptic peptides were analyzed by electrophoresis on 10% polyacrylamide, 1% N, N'-methylenebisacrylamide slab gels (1 mm thick), containing 0.1% SDS, 8 M urea and 0.1 M Tris-phosphate buffer pH 6.6, according to Swank and Munkres (1971). The same gels were used for analysis of Tryptic peptides, prepared by digesting chorion proteins with trypsin (Worthington Biochemicals, TRTPCK grade) in a 100:1 molar ratio, 80 μg/ml, for 16 hr at 25°C, in 6 M urea, 0.1 M Tris-HCl pH 8.5. For preparative isolation of cyanogen bromide fragments, to be used in testing for the presence of a free amino terminus, similar gels of 5 mm thickness were used. Cyanogen bromide fragments of sperm whale myoglobin were used as molecular weight standards.

Results

Protein Fractionation in A. polyphemus

We chose to characterize highly enriched rather than homogeneous protein fractions--a strategy consistent with our goals and imposed by the large number of chorion components and their similarities in size and charge. Fractions were obtained by an initial coarse fractionation step followed by high resolution isoelectric focusing on preparative polyacrylamide gels. Most of the proteins were sequenced twice, and to reduce the effects of contaminants, different coarse fractionation procedures were used for each repeat. After extensive pilot experiments (Rodakis 1978), three such procedures were chosen: differential precipitation by butanol, differential precipitation by dialysis to low ionic strength and by addition of ammonium sulfate, and size fractionation on a BioGel-P150 column.

Figure 1 presents a typical coarse fractionation by dialysis and ammonium sulfate precipitation. Each of the three fractions obtained by this procedure is distinct from each other and simpler than total chorion in terms of isoelectric focusing (Fig. 1a) and SDS-elctrophoresis (Fig. 1b), improving the purity of the bands ultimately excised from each fraction. Excised components belonging to the B family are indicated in Figure 1a and characterised, by SDS-polyacrylamide gel electrophoresis, in Fig. 1c. Seven B-family components were deemed sufficiently pure to be subjected to amino acid and/or sequence analysis, although some contaminants were evident in most cases (especially B2-- g1, B3--e1, B4--e7 and B6--f1). Comparable degrees of purity were attained for various A and C family components (data not shown; Rodakis 1978).

Figure 2 presents typical results of coarse fractionation by BioGel-P150 column chromatography. Again, three of the fractions obtained (II, III and IV; Fig. 2a) were quite distinct by isoelectric focusing (Fig. 2c) and corresponded largely to the C, B and A families, respectively (Fig. 2b). Individual components (identified in Fig. 2c) were recovered by isoelectric focusing at a purity comparable to that of Figure 1c (Rodakis 1978).

Figure 3a tentatively identifies on a two dimensional gel nine *A. polyphemus* family components which were studied further. Although the identification is tentative (since it is not based on co-electrophoresis of purified components with the total sample, but rather on deductions from separate isoelectric focusing and SDS-electrophoretic analyses), it is sufficiently precise to indicate the range of sizes and isoelectric points covered by the components analyzed. Clearly, our sample of nine B proteins is reasonably representative of at least the major components of the family.

Amino Acid Analyses

The purified B proteins were subjected to amino acid analyses (two or more samples from independent preparations in each case). Results are tabulated in Table 1. All nine are quite similar in amino acid content, suggesting substantial sequence homologies. They are rich in glycine and alanine and total non-polar amino acids, as well as relatively rich in cysteine. Some minor differences are apparent, nevertheless: for example, the B2--f2, B2--g1 and B4--e7 proteins lack methionine, and are slightly richer in alanine than most of the rest. Greater differences are apparent in inter-family comparisons (Regier et al. 1978b). Furthermore, the amino acid compositions within the B family tend to be more uniform than those within either the A or the C family (Regier et al. 1978b, 1983). This is consistent with the greater clustering of the B proteins on two- dimensional gels (Fig. 3a), and may suggest that B proteins are more conservative than the rest.

Sequence Analysis: Central Region of the B Family

Since B proteins have blocked amino termini, consisting of pyrrolidonecarboxylate (Regier and Kafatos 1981), sequencing was performed from a single internal methionine present in many members of this family. Five methionine-positive components which were available in subtstantial amounts (1.8 to 61 mg) were cleaved with cyanogen bromide and then subjected to sequence

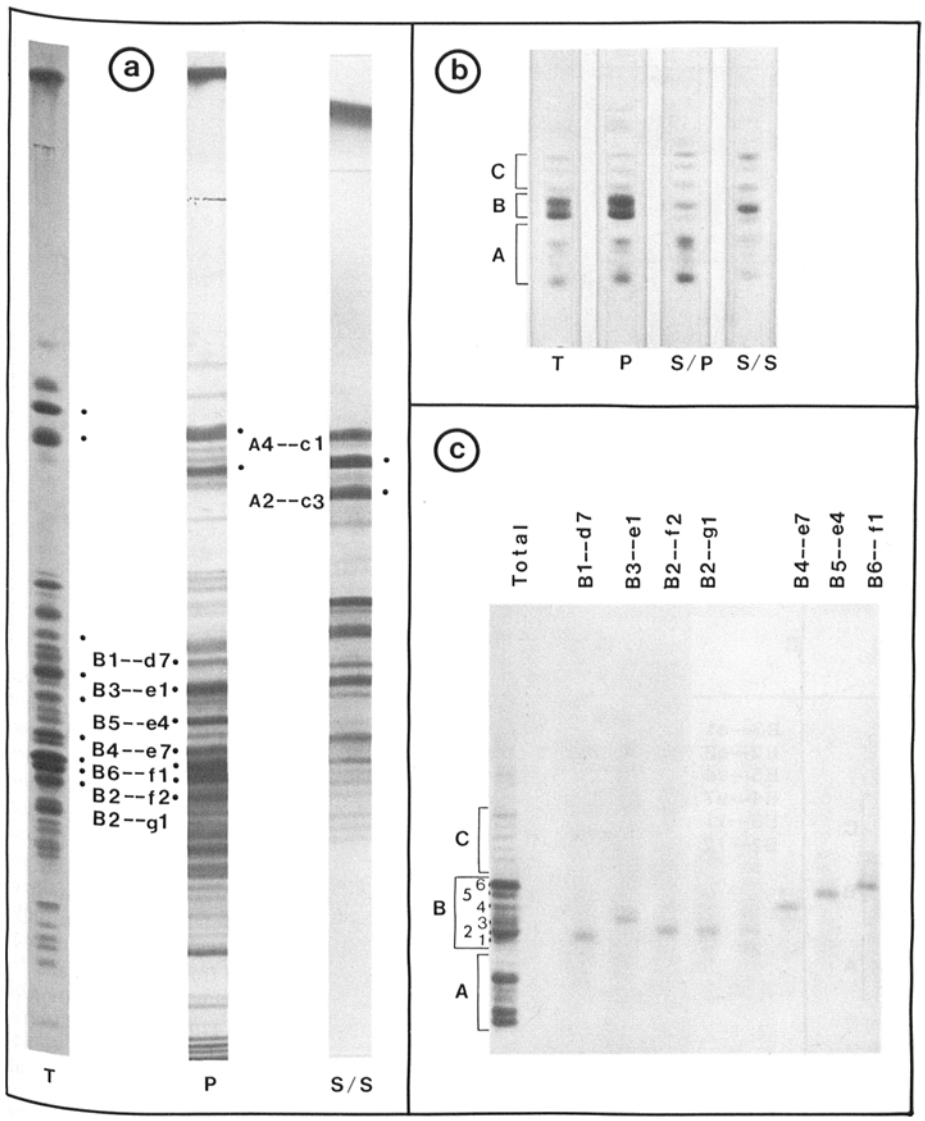

Fig. 1. Chorion protein isolation after coarse fractionation by dialysis at low ionic strength (Series II). a (left) shows isoelectric focusing of total chorion protein, the precipitate of exhaustive dialysis against water (P), and the supernatant after removal of proteins precipitable by ammonium sulfate at 10% saturation (S/S). Components identified include two major A family components which were enriched in fraction (S/S), and the seven B family components which were isolated from fraction (P). Lane (T) is from an analytical gel, lanes (P) and (S/S) from higher resolution preparative gels. To increase resolution of certain bands, the conditions for (P) were nonstandard (1.5% ampholytes, 12 h at 1,200 V). b (upper right) shows SDS-electropherograms in cylindrical gels (Paul et al. 1972) of total chorion (T), precipitate (P) and supernatant after further fractionation into fractions precipitable (S/P) and soluble (S/S) in 10% ammonium sulfate. The major classes of chorion proteins, A, B and C, are identified. c (lower right) shows SDS-slab gel electropherograms of total chorion and bands isolated from fraction (P) in Fig. 1a. The size subclasses of B protein are identified in the total chorion profile

analysis without further purification. Despite a higher than usual background, the sequences were unambiguous except as discussed below.

Figure 4 summarizes these sequences, and compares them to the corresponding region of proteins encoded by recombinant chorion DNA clones (Jones and Kafatos 1982). From the clones we know that this region is approximately in the center of the molecules: the sequences are numbered in Fig. 4 according to the mature protein encoded by clone pc401, in which methionine is residue 69 out of 153 (Jones et al. 1979; Regier and Kafatos 1981). Clearly, the B family of *A. polyphemus* is highly conserved in this region: of the 36 residues included in Fig. 4, 20 (possibly 21) are invariant in the sequences shown. Nearly all differences between sequences are amino acid replacements; a one-residue gap only exists at position 93 of B5--e4.

In three cases, different residues were identified at the same position in different samples of the same protein (position73 of B5--e4 and B2--e2, and position 90 of B3--e1). These samples were obtained from different batches of animals of a semi-wild population, in different years. We interpret these differences as indicative of polymorphism, which is also suggested

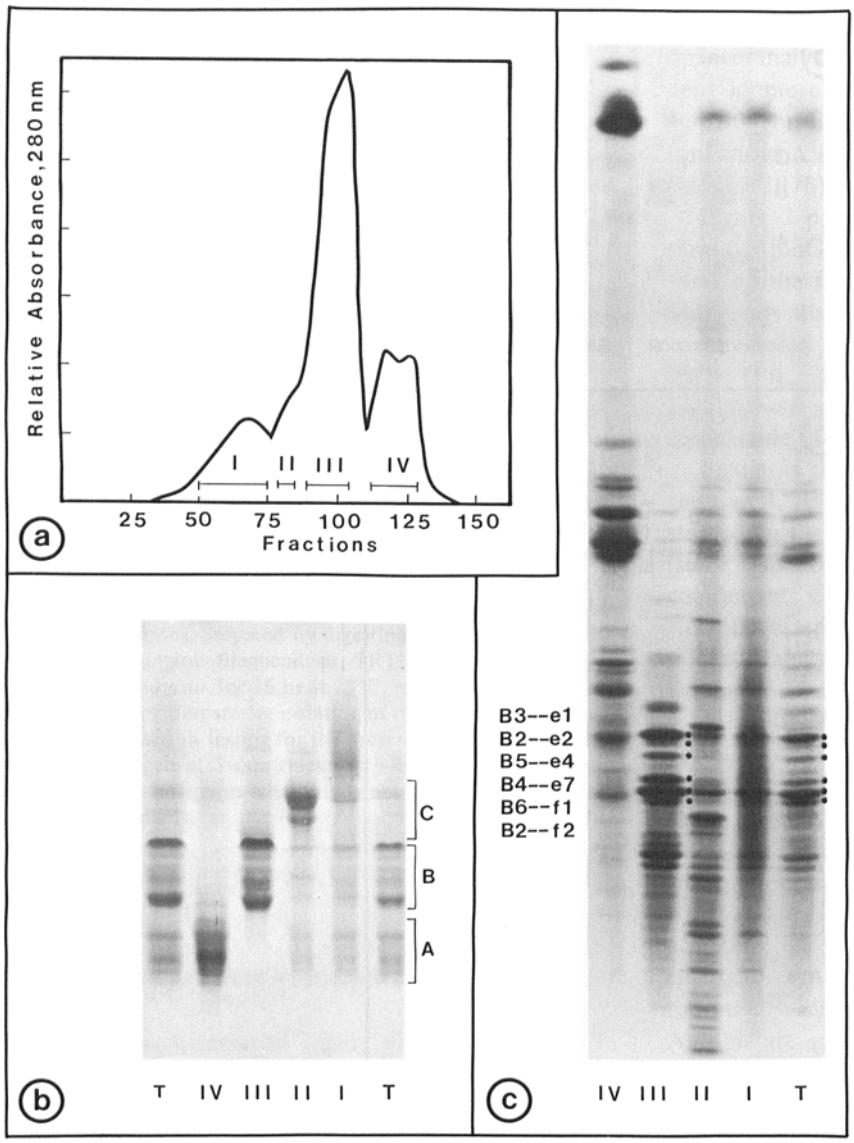

Fig. 2. Chorion protein fractionation on a BioGel-P150 column (series IV). a (upper left) shows the elution profile and the combined fractions collected. b (bottom left) shows characterization of these fractions and of total chorion (T) by SDS-slab gel electrophoresis. Note that fraction III contains the bulk of the B proteins, with only minor contaminants from classes C and D. c (right) shows characterization of the same samples by isoelectric focusing. The components isolated from fraction III are indicated by dots and identified; their purity was comparable to that shown in Fig. 1c

Table 1. Amino acid composition of B proteins (mol %)

Amino acid	*A. polyphemus*									*A. pernyi*		
	B1–d7	B3–e1	B2–e2	B5–e4	B4–e7	B6–f1	B2–f2	B2–g1	B2–g2	B1–c1	B1–c4	B1–c12
CAM-Cysteine	7.1	6.6	6.2	6.3	5.5	6.2	5.1	5.0	5.0	6.3	7.6	6.5
Aspartate (+Asparagine)	3.9	3.1	3.2	3.0	3.9	2.9	4.4	5.0	5.2	3.0	2.9	2.5
Threonine	2.0	2.0	2.2	2.3	2.3	2.2	2.5	2.6	2.3	1.5	1.2	1.8
Serine	3.7	3.0	3.2	2.9	2.8	3.2	3.3	3.2	3.2	3.4	3.2	2.9
Glutamate (+Glutamine)	4.3	5.3	5.0	4.9	4.8	5.2	4.9	4.4	5.2	4.5	4.0	4.3
Proline	3.6	3.4	3.1	3.4	3.2	3.1	3.6	3.4	3.3	3.0	2.0	2.9
Glycine	34.1	36.5	36.6	34.2	34.5	37.7	33.1	32.0	31.1	36.4	37.6	32.7
Alanine	9.3	8.3	9.9	8.9	11.0	8.4	12.3	12.9	12.7	9.6	9.1	10.5
Valine	7.9	6.5	6.8	7.3	7.9	7.0	5.8	6.8	6.7	7.1	8.1	8.2
Methionine	0.6	0.6	0.5	0.8	0.0	0.6	0.0	0.0	0.4	0.7	0.1	0.7
Isoleucine	3.6	3.4	3.2	3.1	2.9	2.8	4.2	4.1	4.6	3.6	3.4	3.4
Leucine	8.3	8.3	8.4	10.9	8.9	8.8	8.5	8.1	8.0	8.2	8.1	8.8
Tyrosine	7.1	7.1	7.0	6.8	7.1	7.1	6.5	6.7	6.8	7.5	6.6	9.9
Phenylalanine	0.8	1.5	1.1	1.4	1.6	1.3	1.3	1.7	1.7	1.5	1.4	1.4
Histidine	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Lysine	0.6	0.6	0.2	0.0	0.0	0.0	0.7	0.6	0.5	0.0	0.8	0.1
Tryptophan	0.0	0.7	0.8	1.2	0.8	1.1	1.2	1.0	0.5	0.0	0.0	0.0
Arginine	3.1	3.1	2.6	2.6	2.8	2.4	2.6	2.5	2.6	3.6	3.8	3.4

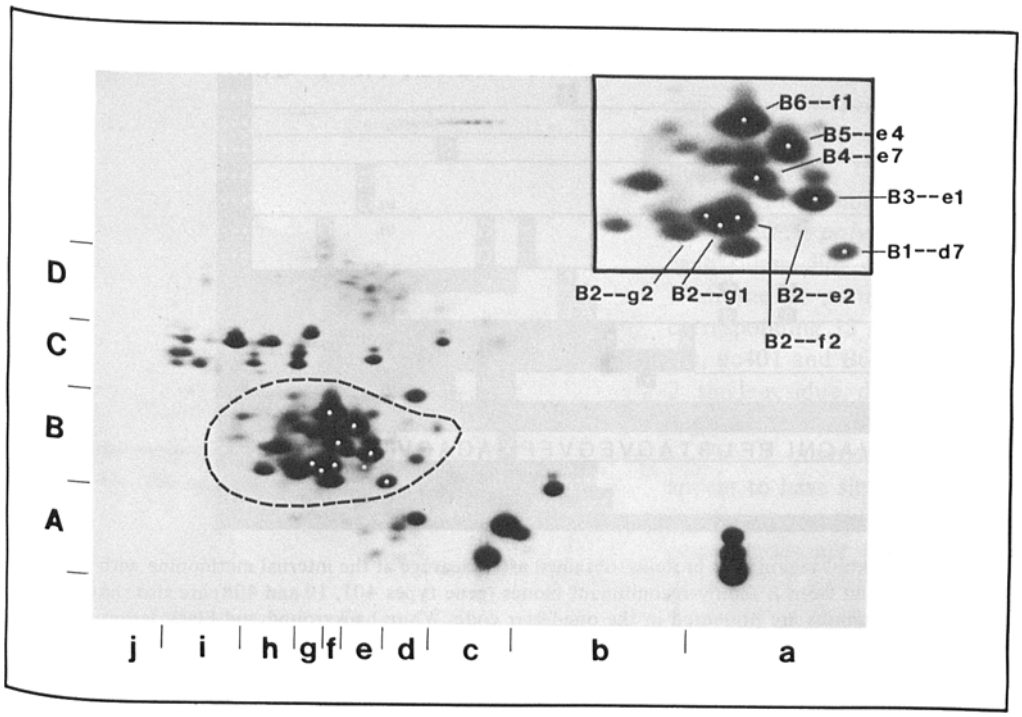

Fig. 3a

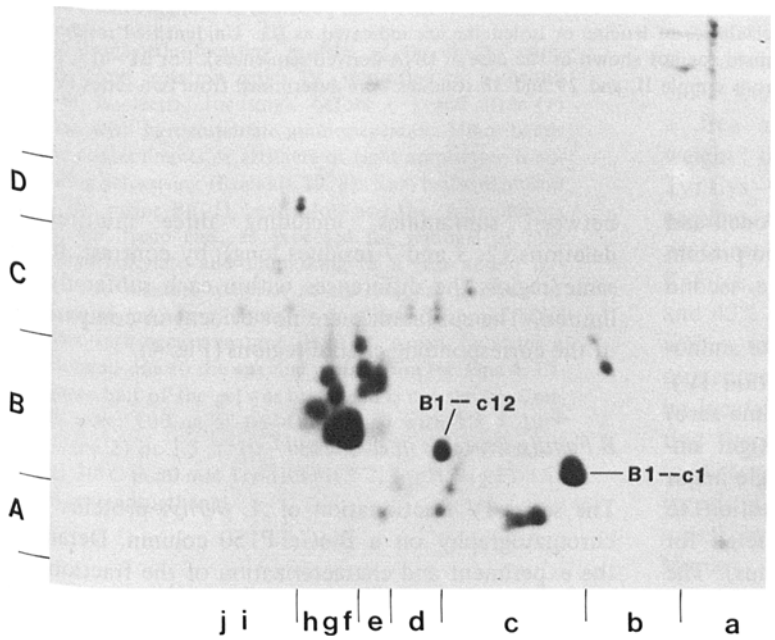

Fig. 3. Identification of the B proteins studied in this report on two-dimensional gel electropherograms. a (top) shows an autoradiogram of total chorion proteins of *A. polyphemus* from the aeropyle crown region (Regier et al. 1980). Size and charge classes are indicated in the two axes. The B family components are surrounded by a dashed line, and those studied in this report are indicated by white dots. The magnified inset is a shorter exposure of the autoradiogram in the B family region, and the components studied are identified by their code. b (bottom) shows an autoradiogram of total chorion proteins of *A. pernyi*, with the major classes delineated and the two sequenced B proteins identified

by sequence differences between cloned copies of chorion DNA (e.g. Fig. 4 shows that residue 91 varies between copies of gene type 401; Jones and Kafatos 1982). Indeed, many of the B components studied here appear to be polymorphic in total chorion protein profiles (Regier et al. 1978b). Since some chorion proteins are encoded by multiple gene copies, this "polymorphism" may refer to allelic sequences found at equivalent locations of homologous chromosomes, or to pseudoallelic gene copies from either identical or homologous chromosomes.

In a total of five positions (71, 73, 80, 83, 87), a second residue was detected simultaneously, clearly above background, in sequencing runs of proteins B3--e1 and B1--d7; in five additional positions (74, 76, 77, 98, 99), second or third residues were detected simultaneously but barely above background in sequencing runs of B6--f1 which were reported elsewhere (Jones et al. 1979). Such multiple residues might also be due to allelic or pseudoallelic polymorphisms, or may be caused by contaminating proteins (see Fig. 1 and 2).

Sequence Analysis: Amino-Terminal Region

The enzyme, pyroglutamate aminopeptidase (EC 3.4.11.8), cleaves off cyclized NH_2-terminal glutamine

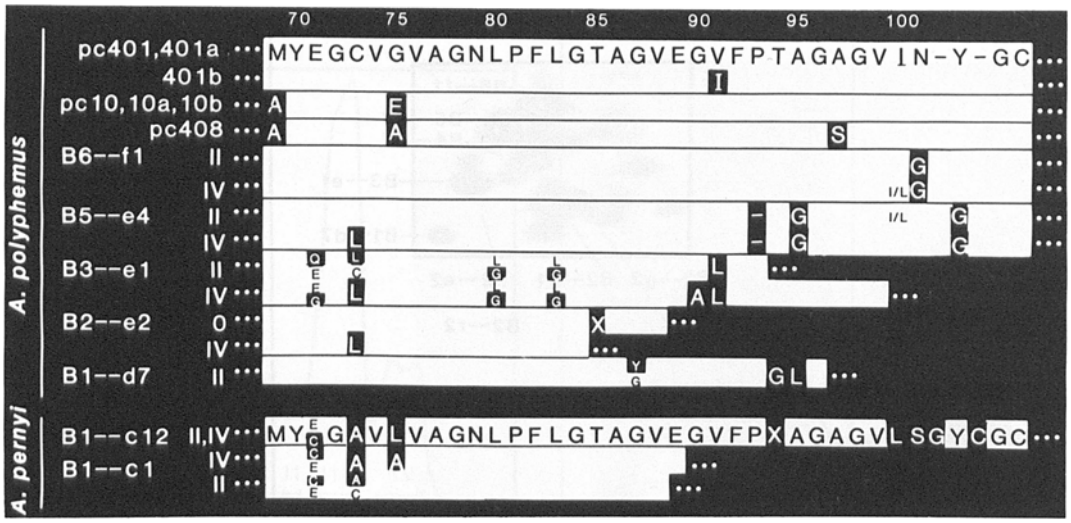

Fig. 4. Summary of sequences from the central region of B proteins, obtained after cleavage at the internal methionine with cyanogen bromide. For comparison, sequences deduced from B family recombinant clones (gene types 401, 10 and 408) are also shown (Jones et al. 1979; Jones and Kafatos 1982). Sequences are presented in the one-letter code. White background and black letters represent residues identical with pc401 and 401a, black background and white letters indicate amino acid replacements or deletions (dashes). Sequences are numbered from the mature protein amino terminus of pc401 (Regier and Kafatos 1981). Each type of protein or gene is separated from the rest by a thin line, and *A. polyphemus* and *A. pernyi* sequences are separated by a wider gap; non-identical sequences of gene copies (pc401, 401a vs. 401b) and of independent protein samples are presented separately (the latter identified according to the isolation series, as in Materials and Methods). Double residues identified at individual positions of a single run are stacked vertically within the corresponding entry, while uncertainties of leucine or isoleucine are indicated as I/L. Unidentified residues are indicated as X. Dots symbolize sequences not determined (or not shown in the case of DNA-derived sequences). For B1–c12, twelve residues following the methionine were determined from sample II, and 29 and 38 residues were determined from two series IV samples, respectively

(pyrrolidonecarboxylic acid; Doolittle 1970; Podell and Abraham 1978). The sequence of the unblocked protein can then be determined, beginning with the second residue (see also Regier and Kafatos 1981).

Figure 5 shows the results of pyroglutamate aminopeptidase action on protein B6--f1 (preparation IV). Several minor bands evident in the control (time zero) sample may be contaminants or artifacts of tight ampholyte binding (Rodakis 1978). Both the single major and the minor bands are shifted upon digestion to slightly more basic isoelectric point (as expected for unmasking of a new, unblocked amino terminus). The profile is the same for samples digested with tenfold different amounts of enzyme, indicating that digestion is complete, and no side-reactions occur.

Three proteins (from preparation IV) were sequenced by this procedure, including two that lack methionine. These three components are major and span most of the molecular weight range of the B family: B2--f2, B4--e7, and B6--f1. In this case the background was low, and each sequence was completely unambiguous from a single run. Figure 6 compares these sequences to each other and to similar sequences determined from recombinant chorion DNA clones (Jones and Kafatos 1982). The variability seen is greater than in the central region (Fig. 4) and, interestingly, it defines two subfamilies of B sequences. In the region defined by residues 4 through 42 of gene type 401, major differences exist between subfamilies, including three multiresidue deletions (2, 3 and 7 residues long); by contrast, in the same region the differences within each subfamily are limited. The subfamilies are not evident in comparisons of the corresponding central regions (Fig. 4).

B Family Proteins in A. pernyi

The series IV fractionation of *A. pernyi* proteins used chromatography on a BioGel-P150 column. Details of the experiment and characterization of the fractions by isoelectric focusing have been published previously (see Fig. 2 in Rodakis et al. 1982). The B and A proteins were effectively separated into different fractions (Fig. 7a), and a number of components were subsequently isolated by high resolution preparative gel isoelectric focusing. The series II fractionation was based on dialysis at low ionic strength, as in *A. polyphemus*. Three B proteins were isolated in sufficient purity (Fig. 7b) and quantity for further study. Table 1 includes their amino acid compositions. Their amino termini were blocked, as were the termini of B proteins in *A. polyphemus*. Two of the *A. pernyi* proteins contained a single methionine residue each, and were sequenced in duplicate, after cleavage with cyanogen bromide. Fig. 3b identifies these proteins on a 2-dimen-

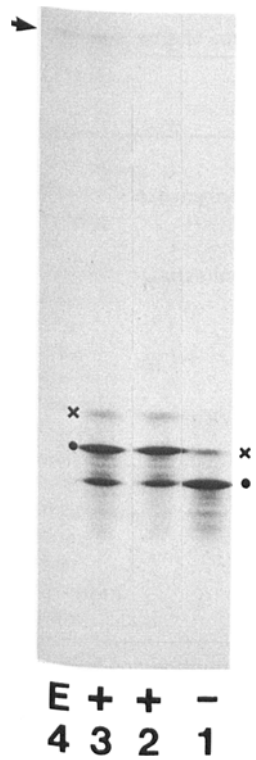

Fig. 5. Isoelectric focusing profiles of the B6–f1 component (from isolation series IV; repurified by a second cycle of isoelectric focusing), before (–) and after (+) digestion with pyroglutamate aminopeptidase. Minor bands may be contaminants or artifacts of tight ampholyte binding during refocusing (Rodakis 1978). Enzymatic digestion moves the major B6–f1 band (dot) and the contaminants toward the basic end, as expected for removal of pyrrolidonecarboxylate and unmasking of a free amino terminus. The residual material at the original position of B6–f1 is of unknown nature; it does not disappear after digestion with excess enzyme (lane 3). Arrow indicates a protein band due to the enzyme preparation (cf. lane 4, E). The upper half of the gel was blank and is not shown. Conditions were: 100 μg of B6–f1 digested with 1.5×10^{-4} units (lane 2) or 1.5×10^{-3} units (lane 3) of enzyme for 16 h at 30°C in 50 mM Tris-HCl pH 7.2, 1 mM Na_2 EDTA, 15 mM mercaptoethanol

sional gel, and Fig. 4 shows the corresponding sequences, comparing them to those of *A. polyphemus*.

Table 1 shows that the amino acid compositions of the *A. pernyi* B proteins are highly similar to each other, and resemble those of *A. polyphemus*. Indeed, Fig. 4 reveals that the similarities extend to the level of primary sequence: the *A. pernyi* sequences following the internal methionine are approximately as homologous to various *A. polyphemus* sequences as the latter are to each other. For example, the 37 residues determined from the *A. pernyi* protein, B1–c12, differ from the corresponding 35 residues of the *A. polyphemus* sequences, pc401 and B6–f1, by only 4 or 5 replacements and 2 single-residue deletions (approximately 82% homology).

The *A. pernyi* and *A. polyphemus* B proteins also appear to have similarly localized methionines and basic residues, suggesting that the regions for which sequence comparisons are not yet available may also be homologous. The location of the single methionine of the *A. pernyi* protein, B1–c1, measured from the blocked amino terminus, was estimated as approximately 43 ±5% of the total sequence length by cyanogen bromide digestion followed by electrophoretic analysis of the resulting two peptide fragments on SDS-urea polyacrylamide gels according to Swank and Munkres (1971): the amino-blocked peptide had an apparent molecular weight of 4,900 ±300, while the peptide with a free amino terminus had an apparent molecular weight of 6,500 ±500 and yielded the sequence Tyr-Cys–Gly-Ala-Val, as expected (Fig. 4). By comparison, the complete sequences of two disparate B proteins of *A. polyphemus* (401a and 10a, Jones and Kafatos, 1982) show the methionine location at 45% and 43% of the mature protein length from the blocked amino terminus, respectively. Similarly, the complete sequences of 401a and 10a show the basic residues to be located near the amino and carboxy termini, predicting the existence of an internal tryptic peptide equal to 82% or 62% of the total length, respectively. Limit tryptic digestion of the *A. pernyi* protein, B1–c1, followed by analysis of the products on a Swank and Munkres (1971)-type SDS-urea polyacrylamide gel, revealed the presence of a tryptic peptide with an apparent molecular weight of 9,000 ±500, or approximately 75 ±5% of the total protein length.

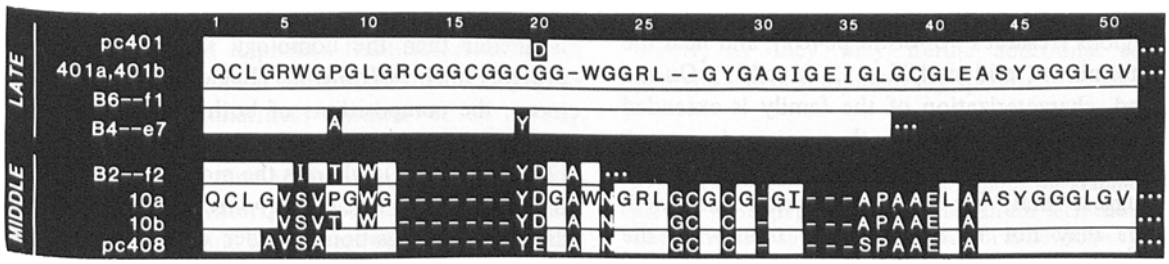

Fig. 6. Summary of sequences from the amino terminal region ("arm") of B proteins, beginning with the cyclized glutamine (pyrrolidonecarboxylate) which is removed by pyroglutamate aminopeptidase (see Fig. 5). Conventions as in Fig. 4. Sequences of the two B subfamilies (developmentally "late" and "middle") are shown separately

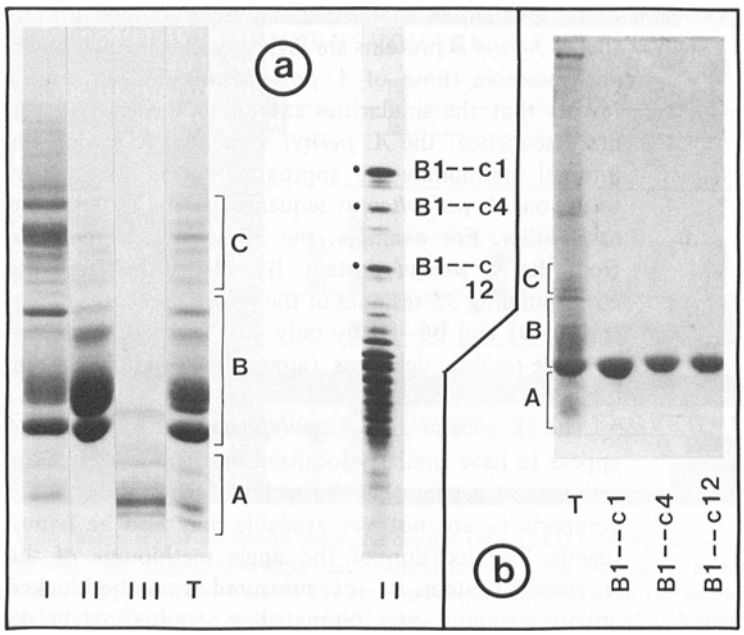

Fig. 7. Characterization of coarse fractions and isolated B family components of *A. pernyi*. The fractions were obtained by BioGel-P150 column chromatography and were characterized by isoelectric focusing, as shown in Fig. 2a and 2b of Rodakis et al. 1982. a (left) shows the SDS-electropherograms of the column fractions, of which II is essentially pure B proteins; the isoelectric focusing profile of II is also shown, with dots indicating the components excised. b (right) shows characterization of these components by SDS-electrophoresis

Discussion

The partial sequences of nine B proteins presented in this report, in combination with the known sequences of three types of B genes (Jones and Kafatos 1982), give a reasonably representative picture of the B family. The seven partially sequenced B proteins of *A. polyphemus* can be considered representative, since they are mostly major components, are synthesized at various developmental stages of choriogenesis (see below), and are diverse in size. A limitation is the absence of sequences from many minor components, including those that belong to the early developmental class and those which are relatively acidic (isoelectric focusing bands gl through il; see Fig. 3a); such components may well deviate from the norm more than the components examined to date. A second limitation is the absence of protein sequences for certain regions known only from DNA clones: between the amino-terminal and central regions (residues 52–68 in pc401), and near the carboxy terminus (residues 105–153 in pc401). On the other hand, characterization of the family is extended by sequences from another moth species, *A. pernyi*, and by analyses of composition and peptide structure which suggest that, for the components examined, differences may not be much greater than what the currently available sequences indicate.

On the basis of DNA sequences from three types of genes, we had previously suggested the existence of a conservative central region or polypeptide domain in the B proteins of *A. polyphemus*, and more variable amino-terminal and carboxy-terminal polypeptide "arms" (Hamodrakas et al. 1982). This suggestion is borne out by the diverse sequences presented in this report. The central region is seen not only to be conservative in sequence, but also rarely subject to deletions or insertions: only a single one-residue deletion was detected among the *A. polyphemus* components examined, and only two deletions between *A. polyphemus* and *A. pernyi*. By contrast, the amino-terminal arm defines two protein subfamilies, which differ from each other substantially, by both replacements and deletions/insertions, although the members of each subfamily considered separately appear conservative.

The limited sequence data from *A. pernyi*, together with compositional analyses and comparisons of cyanogen bromide and tryptic peptides (see Results), clearly indicate that the B family exists in *A. pernyi*, and has properties similar to those in *A. polyphemus*. The extensive interspecies homology in the central domain (Fig. 4) is greater than the homology seen between B and C proteins in *A. polyphemus* (Regier et al. 1983). Furthermore, the compositions of both individual and total B proteins in *A. polyphemus* and *A. pernyi* are very similar (Tables 1 and 2), whereas the most numerous subgroups of C proteins in *A. polyphemus* have significantly different compositions (Regier et al. 1983). However, we do not know directly whether the two B subfamilies of *A. polyphemus*, or different subfamilies, are represented in *A. pernyi*.

Comparisons with the A protein family (Rodakis et al. 1982; Jones and Kafatos 1982) are instructive. The A

Table 2. Amino acid composition of two chorion protein families (mol%)

	A Family		*B Family*	
	A. polyphemus	*A. pernyi*	*A. polyphemus*	*A. pernyi*
CAM-Cysteine	8.4	9.6	> 5.6	6.3
Aspartate (+ Asparagine)	2.5	2.6	< 4.0	4.1
Threonine	3.4	2.4	3.0	3.3
Serine	3.0	2.4	< 3.8	3.7
Glutamate (+ Glutamine)	4.2	4.2	4.7	4.3
Proline	4.0	3.1	4.5	3.9
Glycine	32.7	31.7	32.2	28.8
Alanine	13.7	12.9	11.4	14.2
Valine	7.2	7.2	> 6.1	6.6
Methionine	0.1	0.2	< 0.5	1.1
Isoleucine	3.8	4.5	3.6	4.0
Leucine	6.4	6.6	< 7.9	8.4
Tyrosine	6.3	9.0	6.9	5.9
Phenylalanine	1.0	1.1	< 1.6	1.7
Histidine	0.0	0.0	0.0	0.0
Lysine	1.0	0.6	0.4	0.7
Tryptophan	0.5	0.3	< 1.2	0.7
Arginine	1.8	1.6	< 2.4	2.4

A. pernyi compositions were obtained from fractions II and III of Fig. 7a. *A. polyphemus* compositions were from fraction p/p and the weighted average of fractions s/s and p/s in Regier et al. 1978b. Note that nine types of residues show small but consistent differences between the A and B families in both species

proteins also show a conserved central domain, which is invariant in length, and variable amino-terminal arms that have extensive deletions/insertions and permit classification of the known protein sequences into subfamilies. The A protein family is shared between saturniid (*A. polyphemus* and *A. pernyi*) and bombycid (*Bombyx mori*) moths, i.e. is thought to be as old as these moth families, probably at least 50,000,000 years old (Rodakis et al. 1982). The A subfamilies are shared between *A. polyphemus* and *A. pernyi*, and at the DNA level their sequence diversification is so extensive as to suggest that they may also have been seperate since saturniids and bombycids diverged (Rodakis et al. 1982); a single known A protein sequence from *B. mori* belongs to a different subfamily. By analogy, it is likely that the B family and subfamilies defined in this report are comparably ancient, although additional studies are necessary to test this possibility rigorously.

The presence of comparable A and B proteins in *A. polyphemus* and *A. pernyi* is further supported by compositional comparisons of individual (Table 1) and total (Table 2) A and B protein fractions. Although A and B proteins are similar in composition (and show limited sequence similarities; Tsitilou et al. 1980), significant compositional differences between these two families do exist, and tend to be maintained in the two species (Table 2). Interspecies homologies are also evident by cross-hybridization, between *A. polyphemus* cDNA clones and *A. pernyi* mRNA (Moschonas 1980).

Of special interest is the observation that the *A. polyphemus* B subfamilies defined by the amino-terminal arm sequences may be developmentally specific (Sim et al. 1979; Regier and Kafatos, unpublished). One subfamily includes B4–e7 and B6–f1 plus the related genes of the 401 type, all of which are expressed chiefly in the late period of development. The second subfamily includes B2–f2 and gene types 10 and 408, all of which are expressed chiefly during the middle period of development. Accordingly, the two subfamilies are provisionally designated "late" and "middle", respectively. *A. pernyi* mRNA sequences which cross-hybridize with middle or late *A. polyphemus* cDNA clones are also developmentally middle or late, respectively (Moschonas 1980). If future data reinforce this observation, it would appear that the B subfamilies were formed a long time ago and have maintained their respective developmental controls ever since. It would be interesting to determine amino-terminal sequences of B proteins with different developmental specificities, i.e. early proteins, or late components such as the centrally-sequenced B1–d7, (which is an aeropyle-specific very late protein; Regier et al. 1980). Such components may also prove to belong to the same two subfamilies characterized in the present report, or they may define additional subfamilies, perhaps also developmentally specific.

Acknowledgements. We thank C. Phillips for graphics, B. Klumpar for photography and E. Fenerjian for secretarial assistance. The work was supported by grants or fellowships from the National Institutes of Health, the National Science Foundation and the University of Athens (to F.C.K.), from the European Molecular Biology Organization (to G.C.R.), from the Greek Ministry of Coordination (to N.K.M.) and from the National Institutes of Health (to J.C.R.).

References

Doolittle RF (1977) Sequencing peptides and proteins lacking free amino groups. In: Needleman SB (ed) Advanced methods in protein sequence determination. Springer, New York, pp 38–54

Efstratiadis A, Kafatos FC (1976) The chorion of insects: Techniques and perspectives. In: Last JA (ed) Eukaryotes at the subcellular level. Methods in molecular biology, vol. 8. Marcel Dekker, New York, pp 1–124

Gross E (1967) The Cyanogen Bromide Reaction. In: Hirs CHW (ed) Methods in enzymology, Vol. XI. Academic Press, New York, pp 238–255

Hamodrakas SJ, Jones CW, Kafatos FC (1982) Secondary structure predictions for silkmoth chorion proteins. Biochim Biophys Acta 700:42–51

Jones CW, Kafatos FC (1980a) Coordinately expressed members of two chorion multigene families are clustered, alternating, and divergently oriented. Nature 284:635–638

Jones CW, Kafatos FC (1980b) Structure, organization and evolution of developmentally-regulated chorion genes in a silkmoth. Cell 22:855–867

Jones CW, Kafatos FC (1982) Accepted mutations in a gene family: Evolutionary diversification of duplicated DNA. Mol Evol 19:87–103

Jones CW, Rosenthal N, Rodakis GC, Kafatos FC (1979) Evolution of two major chorion multigene families as inferred from cloned cDNA and protein sequences. Cell 18:1317–1332

Kafatos FC, Regier JC, Mazur GD, Nadel MR, Blau HM, Petri WH, Wyman AR, Gelinas RE, Moore PB, Paul M, Efstratiadis A, Vournakis JN, Goldsmith MR, Hunsley JR, Baker B, Nardi J, Koehler M (1977) The eggshell of insects: Differentiation-specific proteins and the control of their synthesis and accumulation during development. In: Beermann W (ed) Results and problems in cell differentiation, Vol. 8. Springer, Berlin Heidelberg New York, pp 45–145

Moschonas N (1980) Evolutionary comparisons of chorion structural and regulatory genes in two wild silkworm species. PhD Thesis, University of Athens, Athens, Greece

Paul M, Goldsmith MR, Hunsley JR, Kafatos FC (1972) Specific protein synthesis in cellular differentiation: Production of eggshell proteins by silkmoth follicular cells. J Cell Biol 55:653–680

Podell DN, Abraham GN (1978) A technique for the removal of pyroglutamic acid from the amino terminus of proteins using calf liver pyroglutamate amino peptidase. Biochem Biophys Res Commun 81:176–185

Regier JC, Kafatos FC (1981) In vivo kinetics of pyrrolidonecarboxylic acid formation in selected silkmoth chorion proteins. J Biol Chem 254:6444–6451

Regier JC, Kafatos FC, Goodfliesh R, Hood L (1978a) Silkmoth chorion proteins: Sequence analysis of the products of a multigene family. Proc Natl Acad Sci USA 75:390–394

Regier JC, Kafatos FC, Kramer KJ, Heinrikson RL, Keim PS (1978b) Silkmoth chorion proteins: Their diversity, amino acid composition, and the amino terminal sequence of one component. J Biol Chem 253:1305–1314

Regier JC, Mazur GD, Kafatos FC (1980) The silkmoth chorion: Morphological and biochemical characterization of four surface regions. Dev Biol 76:286-304

Regier JC, Kafatos FC, Hamodrakas SJ (1983) Silkmoth chorion multigene families constitute a superfamily: Comparison of C and B family sequences. Proc Natl Acad Sci USA 80: 1043–1047

Rodakis GC (1978) The chorion of the lepidopteran *Antheraea polyphemus*: A model system for the study of molecular evolution. PhD Thesis, University of Athens, Athens, Greece

Rodakis GC, Kafatos FC (1982) The origin of evolutionary novelty in proteins: How a high-cysteine chorion protein has evolved. Proc Natl Acad Sci USA 79:3551–3555

Rodakis GC, Moschonas NK, Kafatos FC (1982) Evolution of a multigene family of chorion proteins in silkmoths. Mol Cell Biol 2:554–563

Sim GK, Kafatos FC, Jones CW, Koehler MD, Efstratiadis A, Maniatis T (1979) Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families. Cell 18:1303–1316

Spande TF, Witkof B, Decani Y, Patcornic A (1970) Selective cleavage and modifications of peptides and proteins. Adv Prot Chem 24:97–260

Swank RT, Munkres KD (1971) Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. Anal Biochem 39:462–477

Terhorst C, Parham P, Mann DL, Strominger JL (1976) Structure of HLA antigens: Amino-acid and carbohydrate compositions and NH_2-terminal sequences of four antigen preparations. Proc Natl Acad Sci USA 73:910–914

Tsitilou SG, Regier JC, Kafatos FC (1980) Selection and sequence analysis of a cDNA clone encoding a known chorion protein of the A family. Nucl Acids Res 8:1987–1997

Witkop B (1968) Chemical cleavage of proteins. Science 162: 318–326

Received February 24, 1983