Primary Structure of Apolipopophorin-III from the Migratory Locust, Locusta migratoria

 POTENTIAL AMPHIPATHIC STRUCTURES AND MOLECULAR EVOLUTION OF AN INSECT APOLIPOPROTEIN* 

(Received for publication, February 22, 1988)

Michael R. Kanost‡‡, Mark S. Boguski§, Mark Freeman¶, Jeffrey I. Gordon||**, Gerard R. Wyatt§, and Michael A. Wells‡‡

From the ¶Department of Biochemistry, University of Arizona, Tucson, Arizona 85721, the §Department of Biology, Queen’s University, Kingston, Ontario K7L 3N6, Canada, and the ||Department of Pathology and ||Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110

The amino acid sequence of an insect apolipoprotein, apolipoporphin-III from Locusta migratoria, has been deduced from the sequence of its cloned cDNA. The mature hemolymph protein consists of 161 amino acids. Optimized alignments of this protein with apolipoporphin-III from the tobacco hornworm, Manduca sexta, disclosed an overall sequence identity of only 29%, even though the two proteins are functionally equivalent. The L. migratoria sequence is composed of 12 repeating peptides that are variable in length. Six amphipathic helical segments of varying length were identified in each protein using a newly described algorithm for detecting such secondary structures. The degree of sequence identity between the two insect apoproteins is considerably less than that observed among orthologous mammalian apolipoproteins. However, calculation of the rates of synonymous and non-synonymous nucleotide substitutions indicates that the insect genes may be evolving at rates similar to the mammalian apolipoprotein genes. Further comparative analyses of insect and mammalian apolipoproteins should provide insights about the limits of sequence diversity tolerated by their predicted amphipathic helical domains.

* This work was supported by Grants HD-07951 and HL-18577 from the United States National Institutes of Health and from the Canadian National Sciences and Engineering Research Council. 

‡‡ To whom correspondence should be addressed: Dept. of Biochemistry, University of Arizona, Tucson, AZ 85721.

** Established Investigator of the American Heart Association.

Lipids are transferred between insect tissues by lipophorin, the major lipoprotein of insect hemolymph (Beenackers et al., 1985; Chino et al., 1985; Shapiro et al., 1988). Lipophorin contains two apoproteins, apolipoporphin-I (apoLp-I, \( M_r \approx 250,000 \)) and apolipoporphin-II (apoLp-II, \( M_r \approx 85,000 \)) (Ryan et al., 1984). In the African migratory locust, Locusta migratoria, and the tobacco hornworm, Manduca sexta, which use lipids as the major fuel for flight, the peptide adipokinetic hormone is released during flight and causes lipid stored in the fat body to associate with lipophorin for transport to flight muscles. This loading of lipophorin with diacylglycerol results in a larger, less dense lipoprotein particle which contains a third apoprotein, apolipoporphin-III (apoLp-III, \( M_r \approx 20,000 \)) (Shapiro and Law, 1983; Van der Horst et al., 1984; Goldsworthy et al., 1985; Chino et al., 1986). apoLp-III exists free in the hemolymph in resting animals and associates with lipophorin during lipid loading until each particle contains 16 molecules of apoLp-III in M. sexta (Wells et al., 1987) and nine molecules (Chino and Yazawa, 1986) or 14 molecules (Van der Horst et al., 1988) of apoLp-III in L. migratoria.

M. sexta apoLp-III is an asymmetric, lipid-binding protein (Kawooya et al., 1986) which contains no carbohydrate (Kawooya et al., 1984). Its role in lipid transport appears to be to increase the lipid carrying capacity of lipophorin by covering the expanding hydrophobic surface resulting from diacylglycerol uptake (Kawooya et al., 1986; Wells et al., 1987). Like mammalian apolipoproteins M. sexta apoLp-III is composed of tandemly repeated amino acid sequences with a high potential for forming amphipathic helical structures (Cole et al., 1987), which may be responsible for its lipid binding activity.

L. migratoria apoLp-III is a glycoprotein of \( M_r = 19,000 \) and exists as three species that are indistinguishable in size and amino acid composition but differ in isoelectric point (Chino and Yazawa, 1986). apoLp-III from M. sexta and L. migratoria function equally well in an in vitro system for production of low density lipophorin (Van der Horst et al., 1988), yet the two proteins differ markedly in amino acid composition, and the L. migratoria protein contains 11% carbohydrate (Chino and Yazawa, 1986). In this report we describe the primary structure of L. migratoria preapoLp-III deduced from a cloned cDNA. Computer-assisted comparative sequence analyses of the orthologous (Fitch and Margoliash, 1970) M. sexta and L. migratoria apolipoproteins were undertaken to identify structural similarities which underlie their functional equivalence. In addition the availability of these apolipoprotein sequences allowed us to examine the extent and rates of sequence divergence between the two apoLp-III genes.

MATERIALS AND METHODS

Animals—African migratory locusts (Locusta migratoria migratoria) were reared at Queen’s University as previously described (Chen et al., 1978).

Purification and Analysis of apoLp-III Protein—Hemolymph from mature adult male locusts was diluted 1:2 with cold 0.1 M sodium phosphate, pH 7.0, 0.15 M NaCl, 5 mM l-phenyl-2-thiourea, 5 mM EDTA, and hemocytes were removed by centrifugation. Diluted cell-
free hemolymph (21 ml) was heated in a boiling water bath for 10 min. Precipitated protein was removed by centrifugation (5009 x g, 10 min), and the supernatant was applied to a column (90 x 2.5 cm) of Sephacryl S-200 and eluted with 0.1 M sodium phosphate, pH 7.6. Fractions which contained apolp-III (detected by sodium dodecyl sulfate-polyacrylamide gel electrophoresis) were combined, dialyzed against 0.1 M acetic acid, and lyophilized, yielding 40 mg of protein. Part of this preparation (3.5 mg) was further purified by reverse phase high performance liquid chromatography (µBondapak C-18, Waters, 25 x 0.5 cm). Samples were dissolved in 0.1% trifluoroacetic acid and eluted with a linear gradient of 0-64% acetonitrile in 0.1% trifluoroacetic acid (45 min, 1 ml/min). Fractions containing apolp-III were pooled and lyophilized, yielding 2.4 mg of protein. apoLp-III prepared by this method contained a single band when analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Kanost et al., 1987). This preparation of apoLp-III was used as an antigen for production of antibodies in rabbits and for analysis of amino acid composition and NH2-terminal sequence. To determine amino acid composition, apoLp-III was hydrolyzed in vacuo in 6 M HCl at 110 °C for 24 h and analyzed on a Beckman model 120C analyzer modified for single column analysis. NH2-terminal amino acid sequence was determined by automated Edman degradation as described by Flynn et al. (1983).

cDNA Cloning and Sequencing—RNA from fat body of adult female locusts was isolated as previously described (Chinzei et al., 1982), and poly(A)+ RNA was prepared by binding to cligo(DT)-cellulose (Aviv and Leder, 1972). Synthesis of cDNA, construction of a cDNA library in λ-gp11, and screening of the library with anti-Locusta apolp-III antisera were performed as described by Cole et al. (1987). Five-thousand recombinant plaques yielded eight positive clones. Four of these were purified to homogeneity, and the EcRI insert of the largest clone (X-LmApoIII-8) and its restriction fragments were subcloned into pTZ18R and pTZ19R. Single-stranded DNA generated from the fl origin of replication of these plasmids (Dente et al., 1983) was sequenced by the dideoxy chain termination method (Sanger et al., 1980), using modified T7 DNA polymerase (United States Biochemical) (Tabor and Richardson, 1987).

Computer-assisted Analysis of Sequencing Data—All calculations were carried out on a MicroVAX II (Digital Equipment Corp.) running the VMS operating system (version 4.4). Program AMPHI for the prediction of amphipathic helical segments was adapted from the published FORTRAN source code (see Appendix to Margalit et al., 1983). Program LWL85 for estimating synonymous and non-synonymous rates of nucleotide substitution (Li et al., 1985) was kindly provided by Dr. Wen-Hsiung Li (University of Texas Health Science Center at Houston).

RESULTS AND DISCUSSION

Nucleotide Sequence of apolp-III cDNA—The strategy for sequencing the apolp-III cDNA is shown in Fig. 1 and the nucleotide and deduced amino acid sequence in Fig. 2. The 679-base pair sequence of the apolp-III cDNA clone contains an open reading frame beginning with an ATG codon at nucleotide 35 and extending to nucleotide 571, followed by a 98-base pair 3’-untranslated sequence ending with a poly(A) tail (Fig. 2). The five nucleotides upstream from the ATG are CGCGC which match the consensus sequence for eukaryotic initiation sites (CCA/GCC) (Kozak, 1984).

Apolp-III Amino Acid Sequence—The cDNA codes for a 198-base pair 3” untranslated sequence ending with a poly(A) tail (Fig. 2). The five nucleotides upstream from the ATG are CGCGC which match the consensus sequence for eukaryotic initiation sites (CCA/GCC) (Kozak, 1984).

Fig. 1. Sequencing strategy for locust apolp-III cDNA. The boxes show the sense coding for mature apolp-III, and the shaded area indicates the sequence coding for the signal peptide. Arrows show the direction and extent of each sequence determination; bp, base pairs.

Fig. 2. cDNA and amino acid sequence of apolp-III from L. migratoria. Amino acids in the mature hemolymph protein are assigned positive numbers, and those in the putative signal peptide are assigned negative numbers. Amino acids which are underlined were confirmed by protein sequencing. (*) At Asn (16) and Asn (83) mark possible NH2-linked glycosylation sites.

179-amino acid polypeptide with a molecular weight of 19,142. This agrees well with an estimated molecular weight of 19,000 for apolp-III translated in vitro and immunoprecipitated. The NH2-terminal amino acid sequence of apolp-III isolated from hemolymph begins with the 19th amino acid encoded by the cDNA sequence. The 92 residues determined by Edman degradation are in good agreement with the deduced sequence with the exceptions of residue 16 (discussed below) and residue 18, which was Thr in the cDNA and Ser in the protein.

When the site of cotranslational cleavage of the signal peptide was predicted by the method of Von Heijne (1986) using the computer program SIEQSEQ2 (Folz and Gordon, 1987) the highest probability for cleavage was at Ala-16 ( residue −3 in Fig. 2), and the second highest was at Pro-18 ( residue −1 in Fig. 2). The accuracy of this predictive scheme is ~80% when the sites with the highest or second highest probability for cleavage are tested using in vitro cotranslational cleavage assays (Folz and Gordon, 1987). Thus, it is probable that the −1 position of preapoLp-III’s signal peptide is occupied by a proline residue. During the course of Von Heijne’s initial assessment of the sites of cotranslational proteolytic processing of eukaryotic preproteins, it appeared that signal peptide cleavage after a proline was prohibited (von Heijne, 1983). However, a few examples of processing after a proline have been recently noted. These include pheasant prelysozyme (Weisman et al., 1986) as well as mutant human apolipoprotein AII produced using oligodeoxynucleotide-directed site-specific mutagenesis where the pentapeptide prosegment had been removed and a proline introduced at the −1 position (Folz et al., 1988).

The deduced amino acid sequence of the mature protein beginning at Asp (1) is a 161-residue polypeptide with a molecular weight of 17,220. The molecular weight of apolp-III isolated from hemolymph is approximately 19,000 (determined by sedimentation equilibrium) (Chino and Yazawa, 1986). The difference may be accounted for by the 11% carbohydrate content of the fully processed apolp-III (Chino and Yazawa, 1986). The amino acid sequence contains two putative NH2-linked glycosylation sites (Asn-X-Ser/Thr)
TABLE I
Amino acid composition of L. migratoria apoLp-III

| Amino acid | cDNA sequence | Amino acid analysis |
|------------|---------------|---------------------|
| Asp and Asn| 18            | 20                  |
| Thr        | 11            | 11                  |
| Ser        | 11            | 9                   |
| Glu and Gln| 31            | 32                  |
| Pro        | 7             | 5                   |
| Gly        | 4             | 5                   |
| Ala        | 29            | 29                  |
| Val        | 9             | 8                   |
| Met        | 0             | 0                   |
| Ile        | 5             | 6                   |
| Leu        | 15            | 17                  |
| Tyr        | 0             | 0                   |
| Phe        | 2             | 2                   |
| His        | 7             | 6                   |
| Lys        | 7             | 8                   |
| Arg        | 3             | 3                   |
| Trp        | 2             | 1                   |
| Cys        | 0             | 0                   |

*Calculated from Chino and Yazawa (1986). Very similar results were obtained upon amino acid analysis of our preparation of apoLp-III.

(Bause, 1983) at residues 16 and 83. Residue 16 of the protein could not be determined by Edman degradation, which indicates that it is probably glycosylated.

The amino acid composition calculated from the deduced sequence is in good agreement with that determined by analysis of the protein (Table I). The sequence confirms that the mature protein lacks methionine, as noted by Chino and Yazawa (1986). This raises questions about the labeling of secreted apoLp-III with [35S]methionine reported by Izumi et al. (1987). We have purified a different locust hemolymph protein of \( M_r = 19,000 \) which does contain methionine, which may explain this result.

Repeating Units and Potential Amphipathic Regions—We have previously shown (Cole et al., 1987) that \( M. sexta \) apoLp-III is composed almost entirely of repeated tetradecapeptides with amphipathic helical potential and postulated that these peptides constitute the structural basis for lipid binding activity as in the mammalian apolipoproteins (reviewed in Boguski et al., 1986a). It was of interest, therefore, to determine if the locust protein has a comparable periodic structure and to analyze sequence relationships between the two proteins.

An optimized alignment between the locust and hornworm sequences was generated using the computer program ALIGN (Dayhoff et al., 1978), and the results are displayed in Fig. 3. The overall sequence identity is 29%, allowing for gaps representing hypothetical insertion and/or deletion mutations. This relatively low degree of sequence similarity was somewhat surprising, and its functional and evolutionary significance is discussed below.

As in \( M. sexta \) apoLp-III, repeating units are present in the amino acid sequence of \( L. migratoria \) apoLp-III. This can be seen by comparison with the \( M. sexta \) sequence in which 12 repeated sequences, approximately 14 residues in length, have previously been defined (labeled \( I-XIII \) in Fig. 3). The presence of repeated sequences in the locust protein was independently verified using the computer program RELATE as well as comparison matrix methods (Boguski et al., 1986a, data not shown). The 12 repeating peptides that comprise most of the locust sequence (Fig. 3) are more variable in length than the hornworm peptides, so that no average repeat length can reasonably be defined. Patterns of residues corresponding to the first eight (most highly conserved) positions in the \( M. sexta \) tetradecapeptide repeat unit are identifiable in the locust sequence and recapitulate the following motif:

```
hydrophobic-acid/amide-acid/amide-hydrophobic-
1 2 3 4
hydrophilic-basic-basic-hydrophobic
5 6 7 8
```

(Fig. 3, Table II). However, the sequences of these repeating units are highly degenerate (see below).

The repeating sequences of \( M. sexta \) apoLp-III have the potential to assume amphipathic, helical conformations (Cole et al., 1987). To investigate the structural potentials of the \( L. migratoria \) sequence, we have employed a new optimized algorithm for detecting amphipathic helical structures (Margalit et al., 1987; Cornette et al., 1987). This method, which has been used successfully to predict immunodominant sites for helper T-cell stimulation (Margalit et al., 1987), has a number of advantages over previous methods for associating amphipathic secondary structure in a protein with the sequence of hydrophobicity values of its residues (see Cornette et al., 1987). The method applies a power spectrum procedure (least squares fit of a sinusoid) on a sequence of overlapping blocks of hydrophobicity values and identifies those blocks for which the maximum intensity occurs at a frequency near 100 Hz, reflecting a periodicity of 3.6 residues/turn (the structure of an \( \alpha \)-helix). The method then determines whether those sequence blocks, if arranged in a helical conformation, would be amphipathic. An amphipathic segment is considered stable if it contains at least three consecutive blocks of the right periodicity. Finally, the amphipathic character of the segment is quantified by combining the magnitudes of the intensity peaks in the power spectrum (amphipathic index) with the length of the segment to yield an amphipathic score (see Margalit et al., 1987 for a complete exposition of this method).

The amphipathic segments predicted by this approach in both \( M. sexta \) and \( L. migratoria \) apoLp-III are shown in Table II. Six amphipathic helical segments of varying lengths are present in each protein. All of these segments incorporate at least some of the characteristics of the repeating units previously defined in the hornworm protein. However, there is not an exact correspondence between these repeat units and the predicted amphipathic segments (Fig. 3, Table II). Furthermore, there is not an exact correspondence between the locust and hornworm sequences in terms of their predicted amphipathic regions (although considerable overlap does exist). In \( M. sexta \) apoLp-III long sequences with high amphipathic scores occur at the termini of the molecule, while in \( L. migratoria \) the segments with the highest amphipathic scores are located more centrally. This lack of correspondence is undoubtedly due to the low degree of sequence identity between the two proteins as well as the presence of multiple insertion and/or deletion mutations. One notable alignment of residues that are predicted to be within amphipathic regions involves repeat IX of \( M. sexta \) apoLp-III and repeat XI of the \( L. migratoria \) sequence (Fig. 3). These repeats are each exactly 14 residues long and are identical in six positions. All but one of the remaining positions are conservative substitutions.

There are several other interesting features of the repeating elements present in the amphipathic segments. Although positions 1 and 4 maintain predominantly hydrophobic resi-

\( ^{a} \) M. R. Kanost, unpublished data.
by this basic amino acid (although 2 of 8 residues include the uncharged amide derivatives (which nevertheless lack significance because in a large number of naturally occurring proteins lysine but none of the locust position 6 residues are occupied). Helical segments, it occurs as the ultimate or penultimate COOH-terminal residue much more frequently than would be expected by chance alone (Margalit et al., 1987).

Finally, the average amphipathic score for segments of *M. sexta* apoLp-III (31.1) is greater than the average score for the *L. migratoria* protein (26.8). Because the amphipathic score is related to the intensity of the periodicity, this observation probably reflects the fact that the periodic structure of the locust protein is degenerate. Even though there is little sequence identity between apoLp-III of *L. migratoria* and *M. sexta*, they are functionally equivalent, even in hybrid low density lipophorins which contain *L. migratoria* apoLp-1 and apoLp-II with *M. sexta* apoLp-III (Van der Horst et al., 1988). These hybrid lipoproteins are functional in *vitro* in loading with diglyceride from locust fat body, and they can both be used as substrates.

### TABLE II

Predicted amphipathic segments in insect apolipophorins-III

| Midpoints of | Sequence | Range of angles | Amphipathic score |
| blocks       |          |                |                  |
|--------------|----------|----------------|------------------|
| **Manduca sexta** |          |                |                  |
| 7-30         | GGNAfEEMKFAKEFKQTF | 85-120         | 60.5             |
| 37-46        | KNTDQFKNLKQDSDS  | 90-125         | 22.4             |
| 50-57        | SDSVLQSLAFSSS   | 80-120         | 16.7             |
| 82-89        | ARQNVEKTAELRK   | 95-120         | 19.9             |
| 115-124      | VQTVFQSLKLAKEVA | 80-120         | 20.4             |
| 140-159      | LAKPIKVQIAVQ    | 80-120         | 46.4             |
| **Locusta migratoria** |      |                |                  |
| 9-16         | VNAIEAVQQLQNTT | 100-135        | 17.7             |
| 21-29        | TVNNAEHELHETGL | 95-120         | 18.4             |
| 70-87        | HQGSVAELNBFNRN | 80-120         | 44.5             |
| 93-101       | ATSAQPADQLSILQS | 95-135         | 19.9             |
| 124-149      | SVAEAAAPVPSA    | 80-125         | 50.5             |
| 151-154      | LQNSGQSAV      | 90-105         | 10.2             |

dues, there is considerable variability in other regions, particularly in the *L. migratoria* sequence. For example, 69% of the residues occupying positions 2 and 3 in the *M. sexta* segments are acidic whereas only 28% of the corresponding residues in the locust segments are acidic with most of the remainder being uncharged amide derivatives (which nevertheless have similar relative hydrophobicity values). Seventy-five percent of the residues in position 6 of the hornworm segments are lysine but none of the locust position 6 residues are occupied by this basic amino acid (although 2 of 8 residues include arginine and histidine). Lysine is thought to have special significance because in a large number of naturally occurring helical segments, it occurs as the ultimate or penultimate COOH-terminal residue much more frequently than would be expected by chance alone (Margalit et al., 1987).

Finally, the average amphipathic score for segments of *M. sexta* apoLp-III (31.1) is greater than the average score for the *L. migratoria* protein (26.8). Because the amphipathic score is related to the intensity of the periodicity, this observation probably reflects the fact that the periodic structure of the locust protein is degenerate. Even though there is little sequence identity between apoLp-III of *L. migratoria* and *M. sexta*, they are functionally equivalent, even in hybrid low density lipophorins which contain *L. migratoria* apoLp-1 and apoLp-II with *M. sexta* apoLp-III (Van der Horst et al., 1988). These hybrid lipoproteins are functional in *vitro* in loading with diglyceride from locust fat body, and they can both be used as substrates.
by lipase from locust flight muscle. Thus, the two proteins must have some conserved features, perhaps intrasequence cooperative effects or a pattern of folding of amphipathic helices, which contribute to these functions. ApoLp-III from *L. migratoria* has recently been crystallized (Holden et al., 1988), and determination of its structure by x-ray crystallography will show which parts of the sequence actually form amphipathic helices. Site-directed mutagenesis of the *L. migratoria* and *M. sexta* cDNA clones should then provide a novel model system for examining experimentally the limits of sequence variation in predicted amphipathic helical structures.

**Molecular Evolution**—As stated previously the degree of amino acid sequence identity between the locust and horn-worm proteins is only 29% even with the introduction of gaps in the alignment to maximize the number of matching residues. Despite the fact that the majority of nonidentical residues represent conservative substitutions, this degree of sequence identity appears rather low compared with that of other orthologous proteins, including even the mammalian apolipoproteins, which are known to have evolved rapidly (Boguski et al., 1986a). Rat and human apoA-I, for instance, are 64% identical in their amino acid sequences. The biochemical properties of amphipathic sequences may not depend upon a precise sequence of amino acids but rather a particular spatial distribution of sets of interchangeable amino acids with similar chemical properties (Boguski et al., 1986b). Nevertheless, there must be some limits on sequence variation that preserve functional equivalence.

These issues prompted us to examine more closely the molecular evolution of these insect apolipoproteins. The common ancestor of humans and rodents existed prior to the great mammalian radiation during the Cretaceous period approximately 75 million years ago, and the exopterygote and endopterygote insects are believed to have diverged during the Carboniferous period approximately 320 million years ago (Hennig, 1981), although the fossil record and divergence times for insects are less reliable than those for mammals. Thus, the insect and mammalian proteins may have evolved at similar rates but with the former having had a longer time to accumulate sequence changes. To test this hypothesis, we have used the method of Li et al. (1985) to estimate the rates of nucleotide substitution in the apoLp-III genes. This method has been used to study the rates of evolution in the mammalian apolipoprotein gene family (Luo et al., 1986).

The method estimates the rates of both synonymous and nonsynonymous nucleotide substitution. Nucleotide substitutions are considered synonymous (phenotypically silent) if they result in no change in the amino acid coded for and nonsynonymous if they do. It is important to distinguish the two because it is known that the rate of synonymous substitution is generally much greater than that of nonsynonymous changes. The method yields two values, $K_s$ and $K_a$, which represent the number of (synonymous) substitutions/synonymous site and the number of (nonsynonymous) substitutions/nonsynonymous site, respectively. Nucleotide substitution rates are obtained by dividing $K_s$ or $K_a$ by the time elapsed since divergence of the species. $K_s$, $K_a$ values, and respective rates for insect apoLp-III and selected mammalian apolipoproteins are presented in Table III. The $K_s$ values for the four mammalian apolipoprotein mRNA sequences are similar, with an average $K_s$ of 0.69 substitutions/synonymous site. The average rate of synonymous substitution is $4.6 \times 10^{-9}$ substitutions/synonymous site/year. As previously pointed out by Luo et al. (1986), this rate is very similar to the average synonymous rate, $4.65 \times 10^{-9}$, for 35 mammalian genes. The $K_a$ value for insect apoLp-III, 1.06, is considerably higher than the average $K_a$ for mammalian apolipoproteins. However, when this value is converted to a rate ($1.66 \times 10^{-9}$) the result is that the insect genes are actually evolving nearly three times more slowly than the mammalian genes at their respective synonymous

| mRNA sequence (mature protein) | $K_s$ | $K_a$ |
|-------------------------------|------|------|
| apoA-I                        | 0.72 | 1.67 |
| apoA-II                       | 0.73 | 2.33 |
| apoE                          | 0.64 | 1.27 |
| apoLp-III                     | 1.08 | 1.61 |

**TABLE III**

**Comparative rates of nucleotide substitution**

The values in parentheses are standard errors. Values of $K_s$ and $K_a$ for apolipoproteins A-I, A-II, and E are taken from Table VI in Luo et al. (1986) and are based on comparisons of human and rat mRNA sequences. Values of $K_s$ and $K_a$ for apolipoprotein A-IV and apolipoprotein-III were computed according to Li et al. (1985) based on comparisons of the human and rat mRNA sequences for apoA-IV and on the locust and hornworm sequences for apoLp-III. Only mRNA sequences coding for the mature proteins, i.e. excluding signal peptides, were used. Rates of nucleotide substitution, expressed as number of nucleotide substitutions/site/year, are based upon divergence times of 75 million years for the mammalian sequences and 320 million years for the insect sequences.

**Codon usage in apolipopophorin-III**

| Manduca sexta | TTT Phe 2 | TCT Ser 1 | TAT Tyr 1 | TGG Cys 0 |
|---------------|----------|-----------|-----------|-----------|
| TTC Phe 6 | TCC Ser 8 | TAC Tyr 0 | TGC Cys 0 |
| TTA Leu 0 | TCA Ser 0 | TAA 0 | TGA |
| TGG Leu 10 | TCT Pro 2 | CAT His 0 | CCG Arg 0 |
| CTC Leu 1 | CCC Pro 0 | CAC His 4 | CCG Arg 2 |
| CTA Leu 0 | CCA Pro 0 | CAA Gln 3 | CGA Arg 0 |
| CGT Leu 7 | CGF Pro 0 | CGG Gln 12 | CGG Arg 0 |
| ATT Ile 0 | ACT Thr 0 | AAT Asn 0 | AGT Ser 1 |
| TTC Ile 1 | ACC Thr 6 | ACG Asn 10 | AGC Ser 2 |
| ATA Ile 1 | ACA Thr 1 | AAA Lys 5 | AGA Arg 0 |
| ATG Met 2 | TAC Thr 0 | AAT Asn 0 | AGG Arg 0 |
| GTT Val 0 | GCT Ala 0 | CAT Asp 1 | GGT Gly 0 |
| GTT Val 2 | GCC Ala 10 | GAC Asp 8 | GGC Gly 4 |
| GTA Val 1 | GCA Ala 1 | GAG Glu 3 | GGA Gly 1 |
| GTG Val 6 | GCG Ala 16 | GAG Glu 16 | GGG Gly 0 |

| Locusta migratoria | TTT Phe 0 | TCT Ser 2 | TAT Tyr 1 | TGG Cys 0 |
|--------------------|----------|-----------|-----------|-----------|
| TTC Phe 2 | TCC Ser 1 | TAC Tyr 0 | TGC Cys 0 |
| TTA Leu 0 | TCA Ser 1 | TAA 0 | TGA 1 |
| TGG Leu 10 | TCT Pro 2 | CAT His 0 | TGG Trp 2 |
| CTC Leu 1 | CCG Pro 4 | CAC His 6 | CGC Arg 1 |
| CTA Leu 0 | CGA Pro 1 | CAG Gln 2 | CGA Arg 0 |
| CCG Leu 10 | CGG Pro 1 | CAG Gln 16 | CGG Arg 1 |
| ATT Ile 0 | ACT Thr 1 | AAT Asn 1 | AGT Ser 0 |
| ATA Ile 1 | ACC Thr 8 | ACG Asn 13 | AGC Ser 4 |
| ATG Met 2 | TAC Thr 0 | AAT Asn 0 | AGG Arg 1 |
| GTG Val 0 | GCT Ala 1 | GAG Glu 3 | GGG Gly 0 |
| GTT Val 2 | GCC Ala 10 | GAC Asp 8 | GGC Gly 4 |
| GTA Val 1 | GCA Ala 1 | GAG Glu 3 | GGA Gly 1 |
| GTG Val 4 | GCG Ala 9 | GAG Glu 13 | GGG Gly 0 |

**TABLE IV**

For mammalian genes, the result is that the insect and mammalian proteins may have evolved at similar rates but with the former having had a longer time to accumulate sequence changes. To test this hypothesis, we have used the method of Li et al. (1985) to estimate the rates of nucleotide substitution in the apoLp-III genes. This method has been used to study the rates of evolution in the mammalian apolipoprotein gene family (Luo et al., 1986).

The method estimates the rates of both synonymous and nonsynonymous nucleotide substitution. Nucleotide substitutions are considered synonymous (phenotypically silent) if they result in no change in the amino acid coded for and nonsynonymous if they do. It is important to distinguish the two because it is known that the rate of synonymous substitution is generally much greater than that of nonsynonymous changes. The method yields two values, $K_s$ and $K_a$, which represent the number of (synonymous) substitutions/synonymous site and the number of (nonsynonymous) substitutions/nonsynonymous site, respectively. Nucleotide substitution rates are obtained by dividing $K_s$ or $K_a$ by the time elapsed since divergence of the species. $K_s$, $K_a$ values, and respective rates for insect apoLp-III and selected mammalian apolipoproteins are presented in Table III. The $K_s$ values for the four mammalian apolipoprotein mRNA sequences are similar, with an average $K_s$ of 0.69 substitutions/synonymous site. The average rate of synonymous substitution is $4.6 \times 10^{-9}$ substitutions/synonymous site/year. As previously pointed out by Luo et al. (1986), this rate is very similar to the average synonymous rate, $4.65 \times 10^{-9}$, for 35 mammalian genes. The $K_a$ value for insect apoLp-III, 1.06, is considerably higher than the average $K_a$ for mammalian apolipoproteins. However, when this value is converted to a rate ($1.66 \times 10^{-9}$) the result is that the insect genes are actually evolving nearly three times more slowly than the mammalian genes at their respective synonymous
sites. Li et al. (1985) point out that their method tends to give underestimates when the degree of sequence divergence is large, which is the case for the two apoLp-I11 sequences, and that these rates should be interpreted with caution. One observation which may explain this result is the codon bias which is present in the apoLp-I11 sequences (Table IV). For most amino acids there is a bias toward the use of codons with C or G in the third position in both M. sexta and L. migratoria apoLp-III. This may reflect some selection pressure for these codons in synonymous mutations in apoLp-III, perhaps due to differences in tRNA availability (Ikemura 1985). It is known that in prokaryotes, highly expressed genes have very biased codon usage and low rates of synonymous substitutions (Sharp and Li, 1987). Perhaps selection at the level of translation has also occurred in the apoLp-III genes.

The average $K_r$ for the mammalian apolipoprotein mRNA sequences is 0.26, which corresponds to an average rate of $1.73 \times 10^{-8}$ substitutions/non-synonymous site/year. The average non-synonymous rate for 35 mammalian proteins is $0.88 \times 10^{-9}$ (Li et al., 1985), indicating that the apolipoproteins are evolving nearly twice as rapidly as the average mammalian protein at non-synonymous sites. The $K_r$ for apoLp-III (0.74) is about three times higher than the average $K_r$ for the mammalian apolipoproteins, but when the differences in divergence times are factored in, the rate for apoLp-III is about 35% slower than that of the mammalian apolipoproteins at non-synonymous sites. Again, the rate of substitution in the insect genes may be underestimated due to their degree of sequence divergence. Perhaps, then, the insect and mammalian apolipoproteins are actually evolving at similar rates at non-synonymous sites. Still, apoLp-III may be evolving somewhat more rapidly than the average mammalian gene.

ApoLp-III is the first protein to have been sequenced in both L. migratoria and M. sexta. A large database of other orthologous sequences from these and other species will be essential for placing the rate of apolipoprotein sequence change in the larger context of insect protein evolution.

Acknowledgments—We thank Dr. T. G. Flynn for amino acid composition and sequence analysis, Dr. Wen-Hsiung Li for providing computer programs, and Dr. Paul Allen for first bringing to our attention the power spectrum method of predicting amphipathic helices.

REFERENCES

Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 264-268

Bause, E. (1985) Biochem. J. 209, 331-336

Beenakkere, A. T. M., Van der Horst, D. J., and Van Marrewijk, W. J. A. (1986) Prog. Lipid Res. 24, 19-67

Boguski, M. S., Freeman, M., Elhourbay, N. A., Taylor, J. M., and Gordon, J. I. (1986a) J. Lipid Res. 27, 1011-1034

Boguski, M. S., Birkenmeier, E. H., Elhourbage, N. A., Taylor, J. M., and Gordon, J. I. (1986b) J. Biol. Chem. 261, 6398-6407

Chen, T. T., Strahlerendorf, P. W., and Wyatt, G. R. (1978) J. Biol. Chem. 253, 5325-5331

Chino, H. (1985) Comprehensive Insect Physiology Biochemistry and Pharmacology (Kerkut, G. A., and Gilbert, L. L., eds) Vol. 10, 115-136, Pergamon Press, Oxford

Chino, H., Downer, R. G. H., and Takahashi, K. (1986) J. Lipid Res. 27, 21-29

Chino, H., and Yazawa, M. (1986) J. Lipid Res. 27, 377-385

Chinzei, Y., White, B. N., and Wyatt, G. R. (1982) Can. J. Biochem. 60, 243-251

Cole, K. D., Fernando-Warnakulasuriya, G. J. P., Boguski, M. S., Freeman, M., Gordon, J. I., Clark, W. A., Law, J. H., and Wells, M. A. (1987) J. Biol. Chem. 262, 11793-11800

Corbett, J. L., Cease, K. B., Margalith, H., Sprouge, J. L., Berzofsky, J. A., and DeLisi, C. (1987) J. Mol. Biol. 195, 659-688

Dayhoff, M. O. (1978) Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, pp. 345-375, National Biomedical Research Foundation, Washington, D. C.

Dente, L., Cesareni, G., and Cortese, R. (1983) Nucleic Acids Res. 11, 1645-1655

Faucchere, J., and Pilska, V. (1983) Eur. J. Med. Chem. 18, 365-375

Fitch, W. M., and Margoliash, E. (1970) J. Mol. Biol. 4, 67-169

Flyn, T. G., de Bold, M. L., and de Bold, A. J. (1983) Biochem. Biophys. Res. Commun. 117, 859-865

Folz, R. J., Notzwehr, S. F., and Gordon, J. I. (1988) J. Biol. Chem. 263, 2070-2078

Folz, R. J. and Gordon, J. I. (1987) Biochem. Biophys. Res. Commun. 146, 870-877

Goldsworthy, G. J., Miles, C., and Wheeler, C. H. (1985) Physiol. Entomol. 10, 151-164

Henning, W. (1981) Insect Phylogeny, pp. 283, 441, John Wiley and Sons, New York

Holden, H. M., Kanost, M. R., Law, J. H., Raymond, I., and Wells, M. A. (1988) J. Biol. Chem. 263, 3960-3962

Ikemura, T. (1985) Mol. Biol. Evol. 2, 13-34

Izumi, S., Yamazaki, K., Tomino, S., and Chino, H. (1987) J. Lipid Res. 28, 667-672

Kanst, R. M., McDonald, H. L., Bradfield, J. Y., Locke, J., and Wyatt, G. R. (1987) in Molecular Entomology, UCLA Symposium on Molecular and Cellular Biology, New Series, Vol. 49 (J. H. Law, ed) pp. 275-284, Alan R. Liss, Inc., New York

Kawooya, J. K., Keim, P. S., Ryan, R. O., Shapiro, J. P., Samarewaara, P., and Law, J. H. (1984) J. Biol. Chem. 259, 10733-10737

Kawooya, J. K., Meredith, S. C., Wells, M. A., Kezdy, F. J., and Law, J. H. (1986) J. Biol. Chem. 261, 13588-13591

Kozak, M. (1984) Nucleic Acids Res. 12, 857-872

Liu, W. H., Wu, C. L., and Law, C-C. (1985) Mol. Biol. Evol. 2, 150-174

Luo, C-C., Li, W-H., Moore, M. N., and Chan, L. (1996) J. Biol. Chem. 187, 325-340

Marghit, H., Spouge, J. L., Cornette, J. L., Cease, K. B., DeLisi, C., and Berzofsky, J. A. (1987) J. Immunol. 138, 2213-2222

Ryan, R. O., Schmidt, J. O., and Law, J. H. (1984) Arch. Insect Biochem. Physiol. 1, 376-383

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467

Shapiro, J. F., and Law, J. H. (1983) Biochem. Biophys. Res. Commun. 115, 924-931

Shapiro, J. P., Law, J. H., and Wells, M. A. (1988) Ann. Rev. Entomol. 33, 297-318

Sharp, P. M., and Li, W-H. (1987) Mol. Biol. Evol. 4, 222-230

Tabor, S., and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 4787-4771

Van der Horst, D. J., Ryan, R. O., Van Heusden, M. C., Schulz, T. K. F., Van Doorn, J. M., Law, J. H., and Beenakkere, A. M. T. (1988) J. Biol. Chem. 263, 2027-2033

Van der Horst, D. J., Van Doorn, J. M., and Beenakkere, A. M. T. (1984) Insect Biochem. 14, 493-504

von Hejnie, G. (1983) Eur. J. Biochem. 133, 17-21

von Hejnie, G. (1986) Nucleic Acids Res. 14, 4693-4690

Weiman, L. S., Drummel, B. M., and Wilson, A. C. (1986) J. Biol. Chem. 261, 2309-2313

Wells, M. A., Ryan, R. O., Kawooya, J. K., and Law, J. H. (1987) J. Biol. Chem. 262, 4172-4176