



## Rapid Communication

# The Precursor Protein of the Structural Apolipoproteins of Lipophorin: cDNA and Deduced Amino Acid Sequence\*

KIM SUNDERMEYER,† JUDY K. HENDRICKS,† SARVAMANGALA V. PRASAD,† MICHAEL A. WELLS†‡

Received 24 May 1996; revised and accepted 18 June 1996

A 10 138 bp cDNA from the fat body of the tobacco hornworm, *Manduca sexta*, which encodes the precursor protein for apolipoprotein aPOLP-1 and -II, the structural apolipoproteins of the insect lipoprotein, lipophorin, has been cloned and sequenced. The cDNA has a single 9915 bp open reading frame beginning at an initiating ATG at bp 59 and extending to a stop codon at position 9974. This open reading frame encodes a 3305 amino acid protein with a molecular mass of 366 812 Da. Signal peptide cleavage is predicted to occur after residue 23, leaving a 3,282 amino acid precursor protein. The precursor protein is arranged with apoLp-II at the amino terminal end and apoLp-I at the carboxy terminal end. At present, the site of cleavage of the precursor protein to generate apoLp-I and -II is unknown. Copyright © 1996 Elsevier Science Ltd

*Manduca sexta* Apolipoprotein Lipophorin

### INTRODUCTION

Insects have a single high density hemolymph lipoprotein called lipophorin, which is composed of 60% protein and 40% lipid (Soulages and Wells, 1994). Lipophorin is a nearly spherical particle which consists of a hydrophobic core, composed of transported lipids, primarily diacylglycerol, with smaller amounts of sterol, triacylglycerol, fatty acid and hydrocarbon, surrounded by a hydrophilic surface coat of apolipoproteins and phospholipid. Lipophorin contains two surface apolipoproteins, and each particle contains one molecule of apolipoprotein-II (apoLp-I; Mr ≈250 000) and one molecule of apolipoprotein-II (apoLp-II; Mr ≈80 000). Lipophorin is similar to the vertebrate high density lipoproteins in its size and to the vertebrate low density lipoproteins in its structural organization. In all insect systems, except the developing egg, lipophorin delivers lipids to tissues without endo-

cytosis of the lipoprotein particle (Soulages and Wells, 1994).

The mechanism of intracellular assembly of lipophorin is not well understood. Lipophorin is synthesized in the fat body and secreted into the hemolymph where it serves as a reutilizable to transport lipids between insect tissues (Soulages and Wells, 1994; Prasad *et al.*, 1986. Venkatesh *et al.*, 1987; Weers *et al.*, 1992). ApoLp-I and -II arise from a common precursor (Weers *et al.*, 1993), which is similar to the situation in the major egg protein vitellogenin (Raikhel and Dhadialla, 1992). In this paper we present the cDNA and deduced amino acid sequence of the apoLp-I and -II precursor protein from *Manduca sexta*.

### MATERIALS AND METHODS

The apoLp-I-II precursor protein mRNA sequence was determined using clones isolated from four different cDNA libraries that were constructed using mRNA from SECOND day fifth instar *M. sexta* fat body. Clone 1 (bp 4703–7918) was isolated from a λgt11 random primed cDNA expression library (Cole *et al.*, 1987) by probing with polyclonal antibodies to ApoLp-I. Clone 2 (bp 4024–5572) was isolated from a second random primed

\*This sequence has been deposited in GenBank under accession number U57 651.

†Department of Biochemistry and Center for Insect Science, Biological Sciences West, The University of Arizona, Tucson, AZ 85721, U.S.A.

‡Author for correspondence. Tel.: 520 621-3847; Fax: 520 621-9288; Email: wells@mozart.biosci.arizona.edu

cDNA library (APGEM-2) that was probed with a subclone from the 5' end of the clone 2. Clone 3 (bp 6959–10 138) was obtained from a  $\lambda$ ZAP II oligo-dT primed cDNA library by probing with a subclone from the 3' end of the clone 1. Clone 4 (bp 1–4412) was obtained by making a 5' cDNA library using Stratagene's ZAP EXPRESS library kit. In order to overcome secondary structure in the mRNA, a hot start technique was used in which all the components of the first strand synthesis reaction, except the reverse transcriptase and RNase Inhibitor, were heated to 65°C. The temperature was reduced to 50°C and the RNase Inhibitor and reverse transcriptase (Stratagene's Stratascript RNase H<sup>-</sup> MMLV-RT) were added. This reaction mixture was incubated for one hour, after which the temperature was lowered to 40°C and 50 U more of reverse transcriptase were added. cDNA synthesis was allowed to proceed for another hour. The rest of the library construction was performed as described in the Stratagene library protocol. Clone 4 was identified by probing with a 380 bp PCR product corresponding to the 5' end of clone 2.

All clones were sequenced by making Exo III deletion subclones using Promega's Erase-A-Base kit (Henikoff, 1984) and then using USB's Sequenase 2.0 kit (Sanger *et al.*, 1977). Gaps between the Exo III deletion subclones were filled in by sequencing with primers made at the University of Arizona sequencing service using an Applied Biosystem's model 373 automated DNA sequencer (Ansoorge *et al.*, 1987; Smith *et al.*, 1986).

Tryptic peptides isolated from purified apoLp-I or apoLp-II were sequenced at the Harvard Microchemical Facility.

Northern blot analysis was carried out as described by Noriega and Wells (1993), using second day fifth instar fat body total RNA. The blot was first probed with a random primed probe that annealed to bases 9758–10 138 in the apoLp-I–II sequence (3' probe). The blot was then stripped and probed with a single stranded probe that annealed to bases 709–1810 in the apoLp-I–II sequence (3' probe).

Sequence Analysis was carried out using the Wisconsin Sequence Analysis Package by Genetics Computer Group, Inc. and through programs available on the World Wide Web, for example through "Pedro's Bimolecular Research Tools."

## RESULTS AND DISCUSSION

The 10 138 bp cDNA sequence (GenBank Accession Number U57 651, sequence not shown) has a single 9915 bp open reading frame beginning at an initiating ATG at bp 59 and extending to a stop codon at position 9974. This open reading frame encodes a 3305 amino acid protein (Fig. 1) with a molecular mass of 366 812 Da (excluding the initiating Met). Although the polyA tail was lost during cloning, the cDNA has a consensus polyadenylation sequence, AATAAA, starting at bp 10 122.

Signal peptide cleavage is predicted to occur after residue 23 (Von Heijne, 1986)

Three tryptic peptides from purified apoLp-II were located at positions 33–44, 677–693 and 700–709, and one tryptic peptide from purified apoLp-I was located at positions 2897–2920 (Fig. 1, underline). These results show that the precursor protein is arranged with apoLp-II at the amino terminal end and apoLp-I at the carboxy terminal end. The precursor protein contains two strong glycosylation sites (NXT/S, where X is any residue other than D, W or P), one in apoLp-II (residues 643–645) and one in apoLp-I (residues 2769–2771), and one weak glycosylation site (NXT/S, where X is D, W or P) in apoLp-I (residues 1844–1846), which is consistent with the fact that both apolipoproteins are glycosylated (Shapiro *et al.*, 1984). No other sites for post-translational modification were found.

The site of cleavage of the precursor protein to generate apoLp-I and -II is unknown. There are no recognizable protein processing sites between residues 700 and 800, where the likely site of processing occurs, and several attempts to determine the amino-terminal sequence of apoLp-I were unsuccessful. However, the carboxy terminal most tryptic peptide of apoLp-II, identified by tandem mass spectroscopy peptide sequence analysis, ends at residue 709. This peptide does not have a R or K residue at its carboxy terminal, as would be expected for a tryptic peptide, which might mean that residue 709 is the carboxy terminal of apoLp-II. If this indeed represents the cleavage site, it would produce apoLp-II with  $M_r = 74,875$  and apoLp-I with  $M_r = 289,551$ . The apparent discrepancy in the predicted molecular mass for apoLp-II, based on an SDS-PAGE value of ~80 kDa, is not large and within the error of SDS-PAGE molecular mass determinations. For example, Willott *et al.* (1989) found a similar difference for arylphorin which has about the same mass as apoLp-II. The discrepancy for apoLp-I seems larger, based on an SDS-PAGE value of ~250 kDa. However, apoLp-I is known to bind lipid with high affinity (Kawooya *et al.*, 1989), which might explain an abnormal mobility in SDS-PAGE and calibration of the gels in this molecular weight range is not highly accurate. It is also possible that during the cleavage reaction a polypeptide of significant size is removed. However, in the absence of hard data, further speculation about the molecular mass of apoLp-I is not warranted.

Fig. 2 shows a Northern blot using second day fifth instar fat body total RNA that was first probed with a random primed probe that annealed to bases 9758–10 138 in the apoLp-I–II sequence (5' probe), then the blot was stripped and probed with a single stranded probe that annealed to bases 709–1810 in the apoLp-I–II sequence (3' probe). It can be seen that both probes detected the same ~10 kb mRNA.

A sequence homology search using the BLAST network service at the NCBI showed that the apoLp-I–II precursor protein was 19.8% identical and 42.9% similar to human apolipoprotein B. However, the alignment of

1 MGKSNRLLLSV *LFVISVLWKA* AYGNKQCQIA CKGSSSPSFA AGQKYNYGVE GTVSVYLTGA  
61 DNQETSLKML GQASVSAISN CELELSVHNM VLSGPDGKKY PCPQIEKFPV RFSYQDGRVG  
121 PEICAAEDDS RRSLNIKRAI ISLLQAEQKP SVQVDVFGVC PTEVSSSQEG GAVLLHRSRD  
181 LSRCAHREQG RNDVFNISAN PDAGIKDLQV LQSMNLVESK VNNGVPEKVS AIEEYLYKPF  
241 SVGENGARAK VHTKLTLSGK GGAGGGNAHC TESRSIIFDV PHGTSSASGN LNSVISAVKE  
301 TARTVANDAS SKSAGQFAQL VRIMRTSSKD DLMRIYSQVK AHQLEKRVYL DALLRAGTGE  
361 SIEASIQILK SKDLSQLEQH LVFLSLGNAR HVNPNALKAA AGLLDMPNLP KEVYLGAGAL  
421 GGAYCREHDC HNVKPEGIVA LSNKLGSKLQ NCRPKNKPDE DVVVAIKGI RNIRHLEDSL  
481 IDKLVHCAVD NNVKARVRV ALEAFHADPC SAKIHKAMD IMKNRQLDSE IRIKAYLAVI  
541 ECPCSHSASE IKNLLDSEPV HQVGNFITSS LRHIRSSSNP DKQLAKKHYG QIRTPNKFKV  
601 DERKYSFYRE MSYKLDALGA GGSVDQTVIY SOTSFLPRSV NFNLTVDLFG QSYNVMELGG  
661 RQGNLDRVVE HFLGPKSFLR TEDPQALYDN LVKRFQESKK KVEDSLSRGR RSIKSEIDVF  
721 DKNLKAESAP YNNELDLDIY VKLFGTDAVF LSFQDDKGF DFNKMLDQILG GCNSGINKAK  
781 HFQQEIRSHL LFMDAELAYP TSVGLPLRLN LIGAATARLD VATNIDIRQI FQSPONAKAD  
841 IKFVPSTDFE ISGAFIIDAD AFSTGIKVI TNLHSSTGTVH NAKVLENGRG IDLQIGLPVD  
901 KQELIAASSD LVFVTAEKGQ KEKQKVIKME KGENEYSACF DQLSGPLGLT MCYDMVLPFP  
961 IVNRNDKLD SIAKAMGKWP LSGSAKFKLFL EKNDLRGYHI KAVVKEDKA GRRSFELLLD  
1021 TEGAKTRRSQ LTGEAVYNEN EVGVKLGLEA VGKVIYGHIW AHKKNELVA SVKGLDDIE  
1081 YSGKLGFSVQ GNEHRAVYKP IFEYSLPDGS SPGSKKYEK IDGQVIRECD GRVTKYTFDG  
1141 VHVNLQNAEK PLEICGSVST VAQPREVEFD VEVKHYASLK GSWKGSVDVL AFNNQLNPKI  
1201 NFDLKGKFEN TDSMHNELDI HYGPNRGDNN ARITFSQILK YHVENSKNFN VITKNNLEIR  
1261 AVPFKLVANA DVDPKKIDID IEGQLQDKSA GFNL DARTHI KKEGDYSIKV KANLNNANLE  
1321 AFSRRDIVNA EKS NVENYID MKGVGRYELS GFVLHKT KPN DVNVGFIGHL KINGGGKNE D  
1381 FKINIGHIET PAVFSSHATI SGRGDIIDY LLKIMRTANP NGNFKLVIK SIAANGQYKV  
1441 TDADGKGNGL I I IDFKINR KIKGDVRF TA KEPVFNADID LFLNFEKDNS DKVHFSTY NK  
1501 KTDKVM DTKN KLEYAGKRTE VNIHQD GILA VTGKAHTVAE LVLPTERCLS LKIDHDGAFK  
1561 DGLYNGHMDM TISDAPKRG S GASTISYK GK VSNSNLDQEI IDYEQINFK LKDGNLQST  
1621 FSLKNNPDGD KFKYEFKSDV NGNLI PKPAN LVATGTYSNS ENEIDETYRL KGSYGS DIGF  
1681 ELAGVGTIKF LDAGDKKYLD DYTTLTVRLPF EKAHDIKWVS TVLFLQPQGG EMTEYTLVES  
1741 VQINADVYKI DANGKVGPKN GYGAVKVLVP HVEPFVLDYN YKSSHEGEKN NNYVELKTKY  
1801 GKGKSASMV DSSYAPHYST LKVKANTPNN DKFKKLDVTV HSKNPSPDAY SNSVVVDADG  
1861 RVYKIDSSIV LSKAHPVLDI QYHSPSSDKI RRLYLQGS SL SSTQ GKLEVK VDNINDICLD  
1921 AVSEANVQKD NVAFKVVANA KELGWKNYGI DISSKDSGSG KRLEFHATND NKNVLSGSTS  
2981 FISKQEGQKT IIEGSGSVKV KEEQKSANFK YIRTVFTDSN EKGVETFFNV ALGERSYVAE  
2041 SRVTNYEYKN SYVYCEEKQ CAHAEIQSKI DMSTPGMIVN VINAGLDLRK LGVAPELGLQ  
2101 MRDEVSDRRP PRFTLDLHIN KEDRKYHLHA YNTPENGHYA SGVTVRLPSR VMALEYTLTH  
2161 PTSQDLFFPI KGEACLDL DK NRPGHKTSAR FLVDYSNSGS EDKAVAEIGF FHPKIEKEAV  
2221 IRLNAFMKRP ENGCFKIESS ASLCHSALGT DRVAKVMFET TPNSVKFLAD TPFVKAIDVE  
2281 GSFNVNQOQR TQOCLFRICL LEGKPVQMSA LVKDYQYEF TTEESNRKLS YVGHLIPEKR  
2341 VDISTDIILS GDKKNIAHGA LFLQDNLVKS DYGLSKENFN YFLNALKKDL DTLEDRIKNV  
2401 GEKASKDVEA VTQRAAPYFK KVEDNFRAEW NRFYQEIADD KVFKEISHVF NEIVQYIAKF  
2461 IDEILQGTKR SWTPSCRPTL SHPRNREMYK KQIEPQVKQL YDTLGALMKE YLDGVIDVVA  
2521 HFAAIVTDFE EKHKAELOEL TNVFTEIFKD LTRLVVAQLK ELPPKIAQIY NDIVSQITNM  
2581 PFVVVLQEKW KEFNFAERAV QLVSQAYEAF SKILPTDELK EFAKALNAYL LKKIKEEKME  
2641 ESKELPRAVR EAGQRVLLIT SIPALAVRRP RLRRWTWHHL KLAVGAGASA PSLGAASWSA  
2701 LRQLAAGDGP PALAPRGLPT AQLDPLDEVP NKLRAVVVNG QHIFTFDGRH LTFPGTCRYV  
2761 LIHDHVD RNF TVLMQLANGQ PKALVLEDKS GTIIE LKDNG QVILNCQSHG FVVEQDVFA  
2821 FRQTSGRIGL CSKYGLMAFC TSKFEVCYFE VNGFYLGKLP GLLGDGNNEP YDDFRMPNGK  
2881 ICSSSESEFGN SYRLSRSCPA ANAPAHDDHQ MHAPLPKPC RVFSGTSPLR PLSIMLDIAP  
2941 FRQACIHAVT GADADKDLQO ACDLARGYRR SRSRGCCPPR CPTPACAART ATGPGSWATP  
3001 TSNCPTDSL ISSSPLRPLR TTPAHYKNMV VPLVSQLVDM LKGKHCTDIK VFLVGHSTKH  
3061 PYPILYDTDL KLKNAKVSFD DKSRYDRIPF VKTGHEK FDS YSKTVVDFLN YIKIELGITN  
3121 IEASQGQIFD LPLRPGAVKH VIFVTGGPTI SQFFLLETVR ALRNKVIIDE MAMSASLVTS  
3181 TPGLKIGGGK NAAQIVGYEK HGVLLLGEKK QSKDSEAVRA TLEVEDDPFS DAVEFANGVV  
3241 FSASNYAALP AGQOKQFIQT AAHNI IQRMW REQIVQOCTC VFVDPFRVRS VCFNKARTEV  
3301 ARRRK

FIGURE 1. Deduced amino, acid sequence of the precursor protein for apolipoprotein-I and -II. The sequence begins with the initiating MET. There is a putative 23 amino acid signal sequence shown in italics. Single underlined sequences correspond to the amino acid sequence of peptides isolated from a tryptic digestion of pure apolipoprotein-II, and the double underlined sequence corresponds to the amino acid sequence of a peptide isolated from a tryptic digestion of pure apolipoprotein-I.

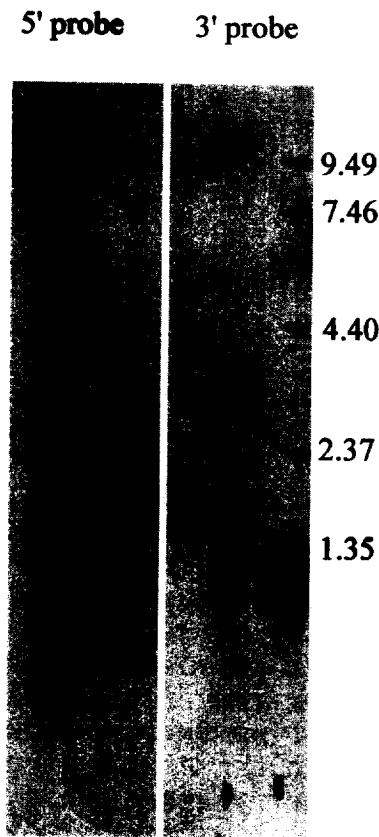


FIGURE 2. Northern blot of *Manduca sexta* second day fifth instar fat body total RNA. The blot was first probed with a random primed probe that annealed to bases 9758–10 138 in the apoLp-I-II sequence (5' probe). The blot was then stripped and probed with a single stranded probe that annealed to bases 709–1810 in the apoLp-I-II sequence (5' probe).

the apoLp-I-II precursor with apolipoprotein B did not reveal any regions of high identity or similarity, rather there was a scattering of small regions of similarity or identity through the entire sequence. Alignment of the apoLp-I-II precursor to other vertebrate and invertebrate apolipoproteins did not yield any significant similarities. The precursor protein showed some similarity to lower vertebrate vitellogenins (sturgeon, frog and chicken), but no significant similarity to insect vitellogenins (mosquito, bollweevil, or silkworm). Because vitellogenins contain two peptide chains produced by cleavage of a precursor protein, we looked carefully to determine whether there was any homology around the putative cleavage sites, but found none.

Based on a number of studies, it has been concluded that in lipophorin apoLp-I is found predominately on the surface of the particle, whereas apoLp-II is somehow "sequestered" away from the aqueous environment (Soulages and Wells, 1994). The amino acid sequence of the precursor protein does not show any unusual sequences in apoLp-II that could explain this observation. Analysis of the predicted secondary structure of the precursor protein using the PredictProtein mail (Rost, 1996) and the PHDsec programs (Rost and Sander, 1993, 1994) did not reveal any striking secondary structure pre-

dictions for apoLp-II which might explain the sequestration of apoLp-II. Additionally, there were no long runs of hydrophobic amino acids in the sequence which might be putative lipid-binding domains, nor were there any apparent loops near the putative processing site.

## REFERENCES

- Ansorge W., Sproat B., Stegemann J., Schwager C. and Zenke M. (1987) Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* **15**, 4593–4602.
- 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) Primary structure and comparative sequence analysis of insect apolipoproteins. *J. Biol. Chem.* **262**, 11794–11800.
- Henikoff S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- Kawooya J.K., Wells M.A. and Law J.H. (1989) A strategy for solubilizing delipidated apolipoprotein with lyophosphatidylcholine and reconstitution with phosphatidylcholine. *Biochemistry* **28**, 6658–6667.
- Noriega F.G. and Wells M.A. (1993) A comparison of three methods for isolating RNA from mosquitoes. *Insect. Molec. Biol.* **2**, 21–24.
- Prasad S.V., Fernando-Warnakulasuriya G.J.P., Sumida M., Law J.H. and Wells M.A. (1986) Lipophorin biosynthesis in the larvae of the tobacco hornworm *Manduca sexta*. *J. Biol. Chem.* **261**, 17174–17176.
- Raikhel A.S. and Dhadialla T.S. (1992) Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37**, 217–251.
- Rost B. (1996) PHD: predicting one-dimensional protein structure by pro-file based neural networks. *Methods Enzymol.* **266**, 525–539.
- Rost B. and Sander C. (1993) Prediction of protein structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599.
- Rost B. and Sander C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55–72.
- Sanger F., Nicklen S. and Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Smith L.M., Saners J., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Kent S.B.H. and Hood L.E. (1986) Fluorescent detection in automated DNA sequencing analysis. *Nature* **321**, 674–679.
- Shapiro J.P., Keim P. and Law J.H. (1984) Structural studies on lipophorin: an insect lipoprotein. *J. Biol. Chem.* **259**, 3680–3685.
- Soulages J.L. and Wells M.A. (1994) Lipophorin: The structure of an insect lipoprotein and its role in lipid transport in insects. *Adv. Protein Chem.* **45**, 371–415.
- Venkatesh K., Lenz C., Bergman D.K. and Chippendale G.M. (1987) Synthesis and release of lipophorin in larvae of the southwestern corn borer *Diatraea grandiosella*. *Insect Biochem.* **17**, 1173–1180.
- Von Heijne G. (1986) A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* **14**, 4683–4690.
- Weers P.M.M., Van Der Horst D.J., Van Marrewijk W.J.A., Van Den Eijnden M., Van Doorn J.M.V. and Beenackers A.M.T. (1992) Biosynthesis and secretion of insect lipoprotein. *J. Lipid Res.* **33**, 485–491.
- Weers P.M.M., Wil J.A., Van Marrewijk W.J.A., Beenackers A.M.T. and Van Der Horst D. (1993) Biosynthesis of locust lipophorin. *J. Biol. Chem.* **268**, 4300–4303.
- Willott E., Wang X.-Y. and Wells M.A. (1989) cDNA and gene sequence of *Manduca sexta* arylphorin, an aromatic acid-rich larval serum protein. *J. Biol. Chem.* **264**, 19052–19059.

*Acknowledgements*—We thank Mary Hernandez for animal care. This work was supported by NIH grant GM 50 008.