Honig, B., Dinur, U., Nakanishi, K., Balogh-Nair, V., Gawinowicz, M. A., Arnaboldi, M., & Motto, M. G. (1979) *J. Am. Chem. Soc. 101*, 7084–7086.
Ito, M., Hiroshima, T., Tsukida, K., Schichida, Y., & Yoshizawa, T. (1985) *J. Chem. Soc., Chem. Commun.*, 1443–1444.
Kropf, A., & Hubbard, R. (1958) *Ann. N.Y. Acad. Sci. 74*, 266–280.
Lewis, B. A., Harbison, G. S., Herzfeld, J., & Griffin, R. G. (1985) *Biochemistry 24*, 4671–4679.
Liebman, P. (1973) *Biochemistry and Physiology of Visual Pigments Symposium*, pp 299–305, Springer, New York.
Lippmaa, E., Alla, M., & Tuherm, T. (1976) *Proceeding of the XIXth Ampere Congress*, pp 113–118, Groupement Ampere, Heidelberg and Geneva.
Lowe, I. J. (1959) *Phys. Rev. Lett. 2*, 285.
Lugtenburg, J. (1985) *Pure Appl. Chem. 57*, 753–762.
Maricq, M. M., & Waugh, J. S. (1979) *J. Chem. Phys. 70*, 3300–3316.
Mathies, R. A., Oseroff, A. R., & Stryer, L. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1–5.
Millet, F., Hargrave, P. A., & Raftery, M. A. (1973) *Biochemistry 12*, 3591–3593.
Oseroff, A., & Callender, R. H. (1974) *Biochemistry 13*, 4243–4248.
Ottolenghi, M. (1980) *Adv. Photochem. 12*, 97–200.
Palings, I., Mathies, R. A., Pardoen, J. A., Winkel, C., & Lugtenburg, J. (1985) *Biophys. J. 47*, 358a.
Palings, I., Pardoen, J. A., van den Berg, E., Winkel, C., Lugtenburg, J., & Mathies, R. A. (1987) *Biochemistry* (in press).
Pardoen, J. A., Winkel, C., Mulder, P. P. J., & Lugtenburg, J. (1984) *Recl.: J. R. Neth. Chem. Soc. 103*, 135–141.
Pines, A., Gibby, M. G., & Waugh, J. S. (1973) *J. Chem. Phys. 59*, 569–590.
Rothwell, W. P., & Waugh, J. S. (1981) *J. Chem. Phys. 74*, 2721–2732.
Sardet, C., Tardieu, A., & Luzzati, V. (1976) *J. Mol. Biol. 105*, 383–407.
Schaefer, J., & Stejskal, E. O. (1976) *J. Am. Chem. Soc. 98*, 1031.
Schmidt, A., Smith, S. O., Raleigh, D. P., Roberts, J. E., Griffin, R. G., & Vega, S. (1986) *J. Chem. Phys. 85*, 4248–4253.
Sefcik, M. D., Schaefer, J., Stejskal, E. O., McKay, R. A., Ellena, J. F., Dodd, S. W., & Brown, M. F. (1983) *Biochem. Biophys. Res. Commun. 114*, 1048–1055.
Shriver, J., Abrahamson, E. W., & Mateescu, G. D. (1976) *J. Am. Chem. Soc. 98*, 2407–2409.
Shriver, J., Mateescu, G., Fager, R., Torchia, D., & Abrahamson, E. W. (1977) *Nature (London) 270*, 271–273.
Shriver, J. W., Mateescu, G. D., & Abrahamson, E. W. (1979) *Biochemistry 18*, 4785–4792.
Stryer, L. (1986) *Annu. Rev. Neurosci. 9*, 87–119.
Suwelack, D., Rothwell, W. P., & Waugh, J. S. (1980) *J. Chem. Phys. 73*, 2559–2569.
van der Steen, R., Biesheuvel, P. L., Mathies, R. A., & Lugtenburg, J. (1986) *J. Am. Chem. Soc. 108*, 6410–6411.
Wald, G. (1968) *Science (Washington, D.C.) 162*, 230–242.
Yamaguchi, A., Unemoto, T., & Ikegami, A. (1981) *Photochem. Photobiol. 33*, 511–516.
Yeager, M. J. (1975) *Brookhaven Symp. Biol. No. 27*, 3–36.

cDNA Sequences of Two Apolipoproteins from Lamprey†

M. Pontes,‡ X. Xu,§ D. Graham, M. Riley, and R. F. Doolittle*
Department of Chemistry, University of California, San Diego, La Jolla, California 92093
Received August 11, 1986; Revised Manuscript Received October 30, 1986

ABSTRACT: The messages for two small but abundant apolipoproteins found in lamprey blood plasma were cloned with the aid of oligonucleotide probes based on amino-terminal sequences. In both cases, numerous clones were identified in a lamprey liver cDNA library, consistent with the great abundance of these proteins in lamprey blood. One of the cDNAs (LAL1) has a coding region of 105 amino acids that corresponds to a 21-residue signal peptide, a putative 8-residue propeptide, and the 76-residue mature protein found in blood. The other cDNA (LAL2) codes for a total of 191 residues, the first 23 of which constitute a signal peptide. The two proteins, which occur in the "high-density lipoprotein fraction" of ultracentrifuged plasma, have amino acid compositions similar to those of apolipoproteins found in mammalian blood; computer analysis indicates that the sequences are largely helix-permissive. When the sequences were searched against an amino acid sequence data base, rat apolipoprotein IV was the best matching candidate in both cases. Although a reasonable alignment can be made with that sequence and LAL1, definitive assignment of the two lamprey proteins to typical mammalian classes cannot be made at this point.

Apolipoproteins are the principal agents for transporting cholesterol and fatty acids in vertebrate blood. At least eight distinct types occur in mammals, including A-I, A-II, A-IV, B, C-I, C-II, C-III, and E; they are synthesized either in the liver or in the intestine, or both. The evolution of mammalian apolipoproteins has received much attention from the viewpoint of sequence comparison. Dayhoff and Hunt (1977) compared the sequences of the three small human apolipoproteins, C-I, C-II, and C-III, and concluded that all three were related to the larger protein A-I. At the same time, Fitch (1977) showed how a series of internal duplications of an 11 amino acid

† This investigation was supported by a grant from the National Institutes of Health (HL26873). Portions of this work are based on a Ph.D. dissertation submitted by M.P. to the University of California, San Diego.
‡ Present address: Department of Neurobiology, Harvard Medical School, Cambridge, MA.
§ Present address: Department of Biology, University of Science and Technology, Hefei, People's Republic of China.

0006-2960/87/0426-1611$01.50/0

residue segment could have given rise to the observed sequence of human apolipoprotein A-I. McLachlan (1977) came to a similar conclusion about the nature of the internal duplication, although he demurred on the common ancestry of A-I and C-III, contending that the observed similarities could just as well be a convergent resemblance resulting from the existence of amphipathic helices needed in these proteins. More recently, Boguski et al. (1984) showed that rat apolipoprotein A-IV contains 13 tandem repeats of a 22 amino acid segment, the fundamental sequence of which is consistent with an amphipathic helix. Subsequently, these same authors (Boguski et al., 1985) compared the sequences of rat apolipoproteins A-I, A-IV, and E and concluded they were all descended from a common ancestor. They made the point further that the diversification of duplicated genes had resulted in lipid binding proteins with distinct albeit overlapping functions.

In all of the cases cited above, some conjecture was made about early events in apolipoprotein evolution, including estimates as to when the various duplication events might have occurred. As in all such matters involving the comparison of contemporary protein sequences where extrapolation is made backward in time, there is a degree of uncertainty stemming from assumptions about the constancy of rates of change. There is an independent approach to the problem, however, that can confirm the antiquity of a particular event. Thus, comparison of comparable proteins from distant creatures can often bracket the times of gene duplication and divergences. For example, if mammals and cyclostome fish each express a protein, the two sequences of which are similar at a sufficiently high level of confidence, we can legitimately presume that their common ancestor also had that protein, some 450 million years ago. If a particular internal duplication or set of duplications is present in both, then the event logically predated the divergence of the two groups.

To this end, we have been isolating and characterizing the principal proteins of lamprey blood plasma in an effort to determine the chronology of events that has led to contemporary vertebrate blood plasma. We have previously reported cDNA sequences for the γ and β chains of lamprey fibrinogen (Strong et al., 1985; Bohonus et al., 1986) and have a study of the lamprey fibrinogen α chain under way. At the same time, we had initiated a study of lamprey plasma albumin (Kuyas et al., 1983), during the course of which we encountered two small proteins with properties reminiscent of apolipoproteins. The proteins were inadvertently exposed during an attempt to sequence an albumin preparation by direct stepwise degradation. It developed that the albumin itself has a blocked amino terminus, and as a result, it was expeditiously possible to infer amino-terminal sequences for the two small molecular weight contaminants. Oligonucleotide probes were synthesized on the basis of these sequences and the lamprey liver cDNA library screened. A suitable selection was made from the many clones that were found to hybridize with these probes, and the full cDNA sequences for the two proteins were obtained. The sequences reveal that these abundant proteins have all the characteristics of typical mammalian apolipoproteins. Although apolipoproteins from fish have been isolated and characterized before (Nelson & Shore, 1974; Chapman et al., 1978), to our knowledge no sequence data have been reported previously.

Materials and Methods

Lampreys. The lampreys used in these experiments were *Petromyzon marinus* taken in various streams in New Hampshire during their spring spawning runs. The handling of the frozen blood plasma and preparation of protein fractions

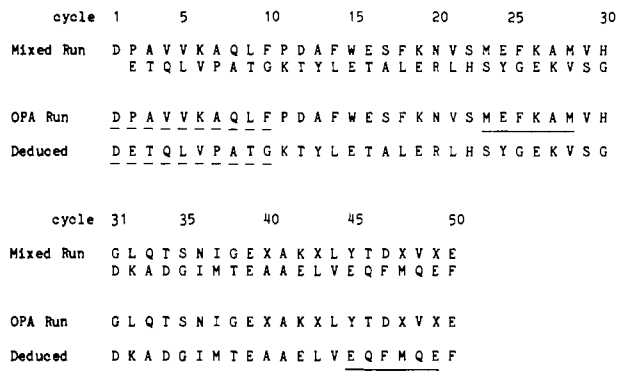

FIGURE 1: Amino-terminal sequences determined in two different automatic sequencer runs. In the first (top), two different residues were observed at each step beginning with residue 2, which showed proline and glutamic acid. Accordingly, a second run was conducted in which the preparation was treated with OPA after the first cycle in order to block the glutamic acid residue. Only a single residue was observed thereafter, allowing the second sequence to be inferred by difference. (These experiments were conducted by Dr. Kenneth Watt.) The portions of the sequence upon which oligonucleotide probes were based are underlined (solid lines). Synthetic peptides were made corresponding to the amino-terminal 9 and 10 residues, respectively (underlined, dashes).

have been described in a previous article (Doolittle et al., 1976), as has the construction of a lamprey liver cDNA library (Strong et al., 1985).

Amino-Terminal Sequences. Amino-terminal sequences were determined with a Beckman Model 850 spinning cup sequencer. An initial run on a lamprey plasma fraction containing water-soluble protein, chiefly albumin, revealed two amino acids at all steps except the first, which was aspartic acid. The two residues found at position 2 were glutamic acid and proline. Accordingly, a second sequencer run was initiated, and after the first cycle, the preparation was treated with *o*-phthalaldehyde (OPA),[1] which reacts with primary but not secondary amines. The sequencer run was then resumed and a singular surviving sequence beginning with Asp-Pro determined. Comparison of the single sequence with the results of the initial mixed run allowed the second sequence to be deduced (Figure 1). These experiments were conducted by Dr. Kenneth Watt of Cetus Corp., Oakland, CA.

Synthetic Oligonucleotides. Oligonucleotide probes based on the deduced amino acid sequences were synthesized at the U.C.S.D. Peptide–Oligonucleotide Synthesis Facility. In the case of LAL1, the 20-nucleotide probe corresponded to residues 23–29 (Figure 1):

Met- Glu- Phe- Lys- Ala- Met- Va(l)

ATG GAR TTY AAR GCR ATG GT

where R = a mixture of both purines and Y = a mixture of both pyrimidines. Actually, the anticomplement of the sequence was made:

5′ACCATYGCYTTRAAYTCCAT3′

In the case of LAL2, a 20-nucleotide probe was based on residues 45–51 (Figure 1):

Glu- Gln- Phe- Met- Gln- Glu- Ph(e)

GAR CAR TTY ATG CAR GAR TT

In this case, also, the anticomplement of the sequence was what was actually made:

5′AAYTCYTGCATRAAYTGYTC3′

The probes were labeled with [γ-^{32}P]ATP through the action

[1] Abbreviations: OPA, *o*-phthalaldehyde; $NaDodSO_4$, sodium dodecyl sulfate; LAL, lamprey apolipoprotein.

```
GGA CGT GCA GGT GCC TTG CTC CAG CGA CCA AAG CCA AAG AAG AAG ATG AAG CTC CAC GTG GCT GCC CTG GCG ACT CTC GCC GTC GTC TGC
                                                           Met Lys Leu His Val Ala Ala Leu Ala Thr Leu Ala Val Val Cys
                                                           -29                                   -20

ATC CTG GCT GCA GGG TCC GAG GCC GCG CCC AAG GCG ATG TCC GAC CCG GCC GTG GTC AAG GCC CAG CTC TTC CCC GAC GCC TTC TGG GAG
Ile Leu Ala Ala Gly Ser Glu Ala Ala Pro Lys Ala Met Ser Asp Pro Ala Val Val Lys Ala Gln Leu Phe Pro Asp Ala Phe Trp Glu
               -10                                    -1  +1                                 +10

AGC TTC AAG AAT GTG TCC ATG GAG TTC AAG AAG ATG GTG CAC GGC CTG CAG ACC TCC AAC ATC GGG GAG CAC GCC AAG AGC CTG TAC ACC
Ser Phe Lys Asn Val Ser Met Glu Phe Lys Lys Met Val His Gly Leu Gln Thr Ser Asn Ile Gly Glu His Ala Lys Ser Leu Tyr Thr
           +20                                    +30                                 +40

GAC ACG GTG GCC GTG CTG ACC CCG TAC CTC CAG AAG ATC CGC GAG AAC GTC ACC AAG ATG TAC CAG GTG TAC GTG GAG TCC AAG CAG CAC
Asp Thr Val Ala Val Leu Thr Pro Tyr Leu Gln Lys Ile Arg Glu Asn Val Thr Lys Met Tyr Gln Val Tyr Val Glu Ser Lys Gln His
           +50                                    +60                                 +70

TAG ACT CCT CAC TCG CGC ACA CAC ACA GGG GGG CAC ACC ACA CGC ACA GCG GCA CCA CAC AGA CGG GGT AAA CTG GGC ACC GAA CAT AGA
Stop

GGC CTC GCG GCT GAT CGT GAA GAC AGG GAC ATG TGA CGT GCC CTG CGC AGC TCT CCC TAC GCG GCC CCC ACC CAC CTG TGG ATG TGG CAC

AGA ACC AGC AGG GAT GAT GAT AAG ATG ATA AAG ACA CAC AGA CAC ACA CAC GCC CCG CAG ACT CGC CAC ACG CGG AGA GAC CCA CAG ACT

CCC AGA CAC CCA CAG ACA CTC CCA GAC ACC ACC CAC AGA CTC CCA CAG ACA CCA CCC ACA GAC ACC ACC CAC AGA CTC CCA CAG ACA CCC

AGA GAC TCC CAC AGA CAC CAC CCA CAG ACT CCC ACA AAC ACC ACC CAC AGG CTG CCA CAC GCG GCG GAG AGA CCC ACA GGT TCG GGC CGC

AAG ACG CCG GGC CGT ATC CGT GGC GCT TTG CTT CCA TAC AAG TCT CAC GGC GAT CTC AAT CGA CCA CCA CCC ACG CTG CAT CTC CAC TTT

CCA CCG TCC CGA AGC ACC CCA GCG GCA GCT GTC CCT ACG CAG CCG CCG GCT GCG GTC GGG GGC CTA TCT CTC TTC TCT ACC GCA CTC GGA

GCA GCG GCG GCG GCG GCG ACG ATC GCC GTC CCA CGC TGC ACG CCC ACG AAT TGT TGG GGG GGG GGG AGG AGG TGA CTC GTT TGG TTT TAT

TCA GAA GGT GGG AAT GTC GTG AAT GTT AAT TGA AAA TGA TAC AGA CCC AAA TAA ACG CGT TGT GTA ACA ATG AAA AAA
```

FIGURE 2: Nucleotide sequence of lamprey lipoprotein LAL1 cDNA and corresponding amino acid sequence.

of T4 polynucleotide kinase. Hybridizations were performed at 45–50 °C in Dazey Seal-A-Meal bags with 100 ng of probe per filter (specific activity = 10^8 cpm/μg) as described by Wallace et al. (1981). Filters were washed at 40–50 °C at the same ionic strength as used in hybridization (0.9 M NaCl) prior to autoradiography. Minipreps of plasmids from hybridization-positive clones were prepared, and the plasmid DNA was treated with appropriate restriction enzymes before electrophoresis on agarose gels (Maniatis et al., 1982). Larger preparations of selected clones were grown (Norgard et al., 1979) and the plasmids purified by RPC-5 Analog chromatography (Thompson et al., 1983). In some cases, fragments were obtained upon isolation of the insert after digestion with *Pst*I and preparative electrophoresis on agarose gels. DNA restriction fragments were labeled either by the T4 kinase procedure (5′ label) or through the action of terminal transferase (3′ label); some fragments were labeled at the 3′ end by the "fill-in" method with the Klenow fragment (Maniatis et al., 1982). DNA sequencing was accomplished by the Maxam and Gilbert (1980) procedure with some additional modification (Ruben & Schmid, 1980). A limited amount of sequence was also obtained by the dideoxy method (Sanger et al., 1977; Biggin et al., 1983) after cloning fragments into M13 (Messing, 1983).

Antibodies to Synthetic Peptides. Synthetic peptides corresponding to the amino-terminal segments of the two sequences (Figure 1) were obtained from the U.C.S.D. Peptide–Oligonucleotide Synthesis Facility. In the case of LAL1, the first 13 residues from the amino terminal were made; for LAL2, the amino-terminal decapeptide was made, a tyrosine replacing the phenylalanine at position 10 for attachment purposes. The peptides were attached to bovine serum albumin by the bis(diazobenzidine) procedure (Bassiri et al., 1979). Antibodies were raised in rabbits and appropriately diluted antisera used for Western blotting (Burnette, 1981). ^{125}I-Labeled sheep anti-rabbit antibodies were used for detection.

Computer Searching and Alignment. Amino acid sequences were searched against an updated version of the NEWAT data base (Doolittle, 1981), as well as release 6.0 of the National Biomedical Research Foundation Atlas (George et al., 1986). Alignments were made and tested for significance as described by Feng et al. (1985).

Secondary Structure Predictions. Computer programs based on the methods of Chou and Fasman (1978) and Garnier et al. (1978) were used to predict the secondary structure of the inferred proteins.

Results

All told, 19 clones were identified that hybridized with the oligonucleotide to LAL1; 14 others gave a strong signal with the probe for LAL2. The largest of these were chosen for characterization.

LAL1. Two clones with overlapping inserts were used to obtain the full sequence, including an apparent 21-residue signal peptide, an 8-residue pro-piece, and a 76-residue mature protein (Figure 2). The 3′-noncoding region extended 792 nucleotides before ending in poly(A). A consensus sequence for poly(A) attachment begins 25 nucleotides upstream of the poly(A) itself. There were two differences between the translated DNA sequence and the interpreted sequencer results shown in Figure 1. The two residues 27 were reversed, Ala occurring in LAL2 and Lys in LAL1, and residue-40 turned out to be His instead of Gly. The first change noted resulted in the probe actually having two mismatches among the 20 nucleotides, but this did not seem to affect the quality of the screening.

Table I: Amino Acid Compositions[a] of LAL1, LAL2, and Lipoprotein Layer

amino acid	LAL1[b]		LAL2[b]		LAL1 + LAL2[b]		high-density lipoprotein[c]
Asn	3.95	7.90	0.60	4.17	1.64	5.32	
Asp	3.95		3.57		3.68		6.2
Thr	6.58		6.55		6.54		7.2
Ser	6.58		4.17		4.91		5.6
Gln	6.58	13.16	2.98	14.29	4.09	13.91	
Glu	6.58		11.31		9.82		13.7
Pro	3.95		4.17		4.09		3.8
Gly	2.63		6.55		5.32		6.3
Ala	6.58		16.07		13.09		14.6
Cys	0.00		0.00		0.00		
Val	11.84		5.95		7.77		8.1
Met	3.95		2.98		3.27		2.8
Ile	2.63		4.76		4.09		3.2
Leu	6.58		11.31		9.82		10.3
Tyr	5.26		1.79		2.86		2.9
Phe	5.26		1.79		2.86		3.0
His	3.95		2.38		2.86		3.1
Lys	10.53		8.33		9.02		8.3
Arg	1.32		4.17		3.27		3.3
Trp	1.32		0.60		0.80		

[a] Values as mole percents. [b] Percents calculated from cDNA sequence. [c] 24-h hydrolysis; duplicate runs.

The amino acid composition of the mature LAL1 protein (Table I) resembles that of many mammalian apolipoproteins. It has no cysteine and only a single tryptophan residue. Its theoretical molecular weight is 8793. A computer search of amino acid sequence collections revealed that the closest known sequence is rat apolipoprotein A-IV. Secondary structure prediction revealed that the sequence of the mature protein is largely helix-permissive, especially over the course of its amino-terminal half.

LAL2. Two clones containing inserts that hybridized with the probe to LAL2 were fully characterized. All told, 1126 nucleotides were sequenced, including 166 of 5′-noncoding region, 576 coding nucleotides, and 379 nucleotides of 3′-noncoding region. Neither a poly(A) tail nor an AATAAA consensus was identified, although two strings of seven and nine A's, respectively, were observed in the 3′-noncoding region (Figure 3).

The coding region corresponds to a 23-residue signal and a 168-residue mature protein, the inferred amino acid composition of which is similar to that of known apolipoproteins (Table I). Its theoretical molecular weight is 18 123. In this case, also, the cDNA sequence was at variance with the sequencer run at two steps: one was the already mentioned switch at residue 27; the other was a Glu/Gln change at residue 49. As judged by the computer programs used, the sequence is helix-permissive throughout its length. A computer search of the amino acid sequence collection again listed rat apolipoprotein A-IV as the most similar sequence in the collection.

Codon Usage. As observed in the other lamprey cDNA sequences we have determined (Strong et al., 1985; Bohonus et al., 1986), codon usage in lampreys is distinctly biased in that there is a very strong preference for G or C in the third position of all codons (Tables II and III).

Identification of Corresponding Plasma Proteins. Initially, the two lipoproteins were serendipitously identified in a crude lamprey albumin preparation. A water-soluble fraction denoted "LMPP", for "lamprey major plasma protein", was found to contain two smaller molecular weight components that corresponded to two prominent bands observed upon $NaDodSO_4$ electrophoresis of lamprey plasma (Figure 4).

Table II: Codon Usage in Lamprey Apolipoprotein 1[a]

TTT (F)	0	TCT (S)	0	TAT (Y)	0	TGT (C)	0
TTC (F)	4	TCC (S)	5	TAC (Y)	4	TGC (C)	1
TTA (L)	0	TCA (S)	0	TAA (*)	0	TGA (*)	0
TTG (L)	0	TCG (S)	0	TAG (*)	1	TGG (W)	1
CTT (L)	0	CCT (P)	0	CAT (H)	0	CGT (R)	0
CTC (L)	4	CCC (P)	2	CAC (H)	4	CGC (R)	1
CTA (L)	0	CCA (P)	0	CAA (Q)	0	CGA (R)	0
CTG (L)	5	CCG (P)	2	CAG (Q)	5	CGG (R)	0
ATT (I)	0	ACT (T)	1	AAT (N)	1	AGT (S)	0
ATC (I)	3	ACC (T)	4	AAC (N)	2	AGC (S)	2
ATA (I)	0	ACA (T)	0	AAA (K)	0	AGA (R)	0
ATG (M)	5	ACG (T)	1	AAG (K)	10	AGG (R)	0
GTT (V)	0	GCT (A)	2	GAT (D)	0	GGT (G)	0
GTC (V)	4	GCC (A)	8	GAC (D)	3	GGC (G)	1
GTA (V)	0	GCA (A)	1	GAA (E)	0	GGA (G)	0
GTG (V)	8	GCG (A)	3	GAG (E)	6	GGG (G)	2

[a] Total codon count (including terminator) = 106.

Table III: Codon Usage in Lamprey Apolipoprotein 2[a]

TTT (F)	0	TCT (S)	0	TAT (Y)	0	TGT (C)	0
TTC (F)	3	TCC (S)	2	TAC (Y)	4	TGC (C)	0
TTA (L)	0	TCA (S)	0	TAA (*)	0	TGA (*)	1
TTG (L)	3	TCG (S)	2	TAG (*)	0	TGG (W)	1
CTT (L)	0	CCT (P)	0	CAT (H)	1	CGT (R)	0
CTC (L)	8	CCC (P)	3	CAC (H)	4	CGC (R)	4
CTA (L)	0	CCA (P)	1	CAA (Q)	4	CGA (R)	1
CTG (L)	14	CCG (P)	3	CAG (Q)	2	CGG (R)	1
ATT (I)	1	ACT (T)	0	AAT (N)	0	AGT (S)	0
ATC (I)	7	ACC (T)	4	AAC (N)	1	AGC (S)	4
ATA (I)	0	ACA (T)	0	AAA (K)	1	AGA (R)	0
ATG (M)	7	ACG (T)	10	AAG (K)	14	AGG (R)	1
GTT (V)	0	GCT (A)	3	GAT (D)	1	GGT (G)	1
GTC (V)	3	GCC (A)	14	GAC (D)	5	GGC (G)	9
GTA (V)	0	GCA (A)	1	GAA (E)	1	GGA (G)	1
GTG (V)	9	GCG (A)	13	GAG (E)	18	GGG (G)	1

[a] Total codon count (including terminator) = 192.

```
GGG GGG GGG GGG GGG GCT TTT TGA CAG CGG GCA AAC TGT CTC CGG CAG CCA ACG AGA CCC CCC CCT CCT CTC TCT CTT CTG CAC GCA GAC

AGT TTG CCC GCT GTC AAA AAG AAA GGA GGG ACC GGA ACG CCG CCA ACC ACC ATC ACC ACC ACC AGA AGC AGC CCA AGA CCT ACG AGC GTG

GCC ACC ATG CTG TCG ACG AAG ATG ACC CAC GCA GCG GGG CTC CTC CTG CTC GTG CTG ACG GCC TAC GTG CAA GCC GAC GAG ACG CAG CTG
       Met Leu Ser Thr Lys Met Thr His Ala Ala Gly Leu Leu Leu Leu Val Leu Thr Ala Tyr Val Gln Ala Asp Glu Thr Gln Leu
                   -20                                        -10                                        -1   +1

GTG CCG GCC ACG GGC AAG ACG TAC TTG GAG ACG GCC CTG GAG AGG CTG CAT AGC TAC GGC GAG GCG GTC AGC GGC GAC AAA GCT GAC GGC
Val Pro Ala Thr Gly Lys Thr Tyr Leu Glu Thr Ala Leu Glu Arg Leu His Ser Tyr Gly Glu Ala Val Ser Gly Asp Lys Ala Asp Gly
                +10                                        +20                                        +30

ATC ATG ACC GAG GCC CGC GAG CTC GTG GAG CAA TTC ATG GAG GAG TTC CAA GCG AAG GCG CTG CCC GAG GGT GTC ACG ACG CAC AAG CTC
Ile Met Thr Glu Ala Arg Glu Leu Val Glu Gln Phe Met Glu Glu Phe Gln Ala Lys Ala Leu Pro Glu Gly Val Thr Thr His Lys Leu
                +40                                        +50                                        +60

GCC GAG GAA ATG GCC GAG GCG GCC AAC GCC AAG CTG GTG CCC ATC CTG CAG GCG GCC AAG GCC GGC ATC GAG CGC GTC ACG GCG CAC CTG
Ala Glu Glu Met Ala Glu Ala Ala Asn Ala Lys Leu Val Pro Ile Leu Gln Ala Ala Lys Ala Gly Ile Glu Arg Val Thr Ala His Leu
                +70                                        +80                                        +90

CAC GAG TCG GCG CCG CTC ATC ATC AAG GTG CGC GGC TTC ATC GAG TCC AAG CGC GGC GTG ATG TGG GCC TAC CTG GCG GCG CTG GCT GAG
His Glu Ser Ala Pro Leu Ile Ile Lys Val Arg Gly Phe Ile Glu Ser Lys Arg Gly Val Met Trp Ala Tyr Leu Ala Ala Leu Ala Glu
              +100                                       +110                                       +120

CGG GCG AAG AAG GCC AAG CTG GAC GAC ACC CTC AAG GGC GTG CGA GAT GGC TTG ACG GCC CTC ACC TTG TCC AGC ATT GGA CTG ATG AAG
Arg Ala Lys Lys Ala Lys Leu Asp Asp Thr Leu Lys Gly Val Arg Asp Gly Leu Thr Ala Leu Thr Leu Ser Ser Ile Gly Leu Met Lys
              +130                                       +140                                       +150

GAG GTG AAG GAG ATC GCG AGC GCG CCG CCA CCC GCT CAA TGA GAA GTC CCC GCC GAG GAG AGT CGG GGT GGT GGT GGT GGT GGT GGG GTC
Glu Val Lys Glu Ile Ala Ser Ala Pro Pro Pro Ala Gln Stop
              +160

ATG GCA AGC CGT GCG TTG CTG CCA ACT CCC GGG GTG GAG AAG GTT TAG TTG GTA GTG GAG CCC TTT CCC TGT TTC TTA CGT GCT GCT CTT

TAA GAC GAC AAA AAA ATA TAT ATA CGT GAG GTC ATT TAC TGG AGT GTG GGT TGG GGG GCA TTT TTT TTT TGT TCT TGC CGT AAA TGC TTC

TTC GGG TTG TTT TGT GGG ACA GCT TGA AGT GCA GGC GGG GGG ATT GCC CTC CAG AGA TGT TAT CGA GGT CAC CTA GGG AAA AAA AAA TGT

TTG AGC GGA CTG CCG GGA AAG TTT CGA CTG CTT TGA AAT ATT CTT TAA AAG TCT GAA TCT TTC CCC CCC CCC CCC CCC CCC
```

FIGURE 3: Nucleotide sequence of lamprey LAL2 cDNA and corresponding amino acid sequence.

In spite of their abundance, the two proteins (LAL1 and LAL2) proved difficult to purify. That they were genuine lipoproteins was demonstrated by the fact that the two proteins appear in the void volume when plasma is gel filtered on Sephadex G-150, in spite of their small molecular weights. Beyond that, they are the major components of the high-density lipoprotein layer separated by flotation ultracentrifugation at density 1.21 (Figure 4, left panel). They are absent from the "lipoprotein cake" that is floated off at densities greater than 1.10; the most abundant protein observed in the "cake" is too large to enter acrylamide gels and may be the equivalent of apolipoprotein B. The amino acid composition of the high-density lipoprotein fraction is similar to the combined inferred composition of the two proteins, implying that they exist in about equivalent molar concentrations (Table I). Delipidation of the lipoprotein (3:1 ethanol:ether ratio) invariably resulted in a material that could not be dissolved in anything short of 4 M urea–1% sodium dodecyl sulfate. Accordingly, the high-density lipoprotein fraction was incubated in that solvent and then passed over a G-100 column equilibrated with 8 M urea–0.1% $NaDodSO_4$, under which conditions the two proteins were only partially separated.

Western Blots. Lamprey plasma (5× diluted with phosphate-buffered saline) was subjected to $NaDodSO_4$ electrophoresis, and the proteins were transferred to nitrocellulose by electroblotting. Blots were treated with (a) nonimmune serum, (b) antiserum to a synthetic peptide corresponding to the amino terminus of LAL1, or (c) antiserum to a synthetic peptide based on the amino terminus of LAL2. The antibodies to the synthetic peptides from LAL1 reacted with a 10-kilodalton component and those to LAL2 with a 20-kilodalton component (Figure 4, right panel).

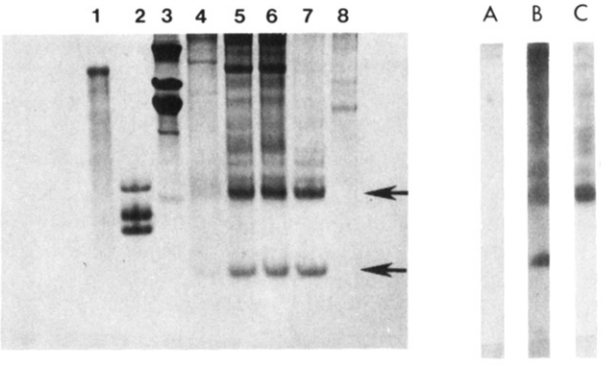

FIGURE 4: Dodecyl sulfate–polyacrylamide (7.5%) electrophoresis of lamprey plasma (left) and corresponding Western blot (right). Arrows denote lipoprotein 1 (M_r 10000) and lipoprotein 2 (M_r 20000). Left gel tracks: (1) lamprey transferrin; (2) reference substances—chymotrypsinogen, myoglobin, and lysozyme; (3) lamprey fibrinogen; (4) lamprey albumin (crude); (5 and 6) lamprey plasma; (7) high-density lipoprotein fraction; (8) lipoprotein "cake" from ultracentrifuged lamprey plasma. In the gel tracks shown at the right, a companion gel of lamprey plasma was probed with (A) nonimmune serum control, (B) anti-LAL1 peptide, and (C) anti-LAL2 peptide.

Discussion

The principal reason for studying lamprey plasma proteins is to find when particular significant events occurred during the evolution of these proteins (Doolittle, 1984). In the case of the apolipoproteins, there are two general questions to ad-

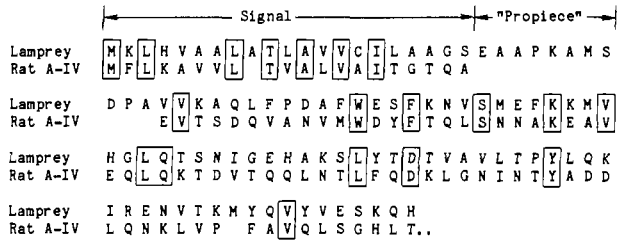

FIGURE 5: Sequence alignment of lamprey preproapolipoprotein LAL1 with amino-terminal portion of rat preapolipoprotein A-IV.

dress. First, when did the discrete duplications occur that led to the diverse types of protein observed among mammals? Barker and Dayhoff (1977) concluded that the small apolipoproteins C-I, -II, and -III share common ancestry with apolipoprotein A-I, which is longer and internally duplicated, and conjectured that the duplications occurred well in advance of the evolution of vertebrates. More recently, Boguski et al. (1985) have contended that mammalian apolipoproteins A-I, A-IV, and E also share common ancestry.

The second question has to do with the timing of the internal duplications. Human apolipoprotein A-I (245 residues) contains 13 11-residue repeats, and rat A-IV (371 residues) contains approximately twice that number. In both cases, the major elongation is actually due to serial duplication of a 22-residue segment. The virtues of an 11-residue or a 22-residue repeat for a lipid binding α-helix have been described by McLachlan (1977), who noted that three turns of an α-helix amount to 10.8 residues, and a repetitive helix would yield a continuously suitable hydrophobic face. Moreover, if the helical segments each began with a proline residue, small directional changes would result that could allow the construction of an inclusive topical covering. The question is: When did these tandem duplications take place?

The two principal lamprey apolipoproteins are small and, although highly helix-permissive, do not have obvious 11-residue repeats. Nonetheless, lamprey apolipoprotein 1 (LAL1) does have certain progenitorial features. First, the mature protein has 76 residues, which is very close to the sizes of human apolipoproteins A-II (77 residues), C-II (78 residues), and C-III (79 residues). More significant, however, LAL1 can be aligned with the amino-terminal segment of rat apolipoprotein A-IV (Figure 5). The relationship is significant in that the authentic alignment scores are 7–9 standard deviations above those of jumbled versions, whether or not the test is conducted with the entire LAL1 sequence or only that of the mature protein. That the signal peptides have 7 of 20 residues identical is partly a reflection of the restricted amino acid compositions of those entities, of course.

The alignment of LAL1 with rat A-IV points to another interesting feature: LAL1 is made as a precursor and has either a six- or eight-residue propiece. This is reminiscent of mammalian apolipoprotein A-I, which in humans and rats contain a six-residue segment between the signal peptide and the amino terminus of the mature circulating protein. In the case of the lamprey LAL1, we are not able to distinguish between a 21-residue signal and an 8-residue propiece, a situation suggested by the alignment with rat A-IV (Figure 5), and a 23-residue signal and a 6-residue propiece, which would be more in keeping with the size of the segment in mammalian A-I proteins. The cleavage of either junction by a signal peptidase fits the empirical rules for signal peptide removal devised by von Heijne (1986).

If size alone is used as a criterion, the 168-residue lamprey apolipoprotein LAL2 does not have a known mammalian counterpart. Still, our search of the data base revealed that

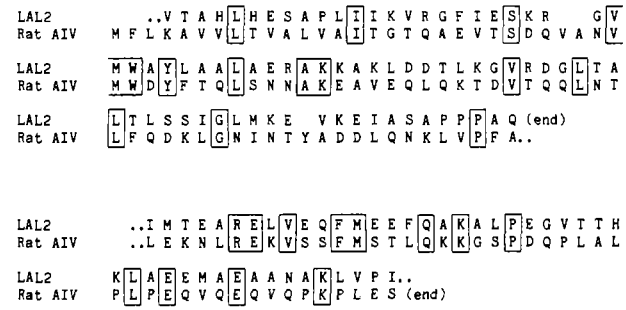

FIGURE 6: Alignment of two different segments of lamprey apolipoprotein LAL2 with sections of rat apolipoprotein A-IV. (Top) LAL2 residues 109–186 vs. rat A-IV residues 1–84. (Bottom) LAL2 residues 55–98 vs. rat A-IV residues 343–391.

rat apolipoprotein A-IV was the closest known match. In this case, however, the alignment involves different portions of the two sequences. Thus, LAL2 residues 41–80 can be aligned with positions 310–389 of the rat A-IV, in one case, and LAL2 residues 133–180 with A-IV positions 10–55 in another (Figure 6). Even allowing for the constraints of α-helix periodicity, we find that the alignments are tantalizing. They are not significant enough to warrant conjecture about potential segmental rearrangements, however.

It should be noted that the rate of evolutionary change among the apolipoproteins, as judged by the comparison of rodent (rat and mouse) and primate (humans and monkeys) sequences, is sufficiently great (Boguski et al., 1985; Barker & Dayhoff, 1977) that convincing sequence relationships to ancestors as distant as lampreys would be difficult to establish. Rodents and humans diverged about 80 million years ago, whereas lampreys split off from other vertebrate lineages 400–450 million years ago. Still, as in most proteins, different parts of the sequences change more slowly than others. In the case of apolipoprotein A-I, it was noted by Boguski et al. (1985) that the 64-residue amino-terminal domain of the mature protein is conserved to the greatest degree between rats and humans; this corresponds to the segment in A-IV that we are able to align with lamprey LAL1 (Figure 5).

Abundance of Lamprey Apolipoproteins. LAL1 and LAL2 are among the most abundant proteins in lamprey blood plasma (Figure 4), and the message abundance in liver is such that numerous clones were readily identified. It should be noted that the lampreys were collected during their annual spawning run, a fact that may bear on the abundance of these proteins, which are likely under endocrine control. It is well-known that fish plasmas contain substantial amounts of cholesterol (Larsson & Fange, 1977) and that many fish have correspondingly large amounts of lipoproteins (Chapman, 1980). Indeed, Chapman et al. (1978) have reported that, on the basis of molecular weight and amino acid composition, the two most prominent high-density lipoproteins in trout plasma resemble human apolipoproteins AI and AII. By these criteria, the two lamprey proteins we have described best fit those assignments also.

In summary, lamprey blood plasma contains two small but abundant apolipoproteins which are similar in many ways to mammalian apolipoproteins. Although both have sequences that resemble parts of mammalian apolipoprotein A-IV, the long series of tandem repeats that typifies the latter must have occurred since the divergence of lampreys and other vertebrates.

ACKNOWLEDGMENTS

We are grateful to Cemal Kuyas, whose experiments on lamprey albumin in our laboratory led to the eventual isolation

of these two lipoproteins, and to Kenneth Watt for the extended sequencer runs that yielded the amino acid sequence data upon which our oligonucleotide probes were patterned.

References

Barker, W. C., & Dayhoff, M. O. (1977) *J. Comp. Biochem. Physiol. 57B*, 309–315.
Bassiri, R. M., Dvorak, J., & Utiger, R. D. (1979) in *Methods of Hormone Radioimmunoassay* (Jaffey, B. M., & Behrman, H. R., Eds.) pp 46–47, Academic Press, New York.
Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.
Boguski, M. S., Elshourbagy, N., Taylor, J. M., & Gordon, J. I. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5021–5025.
Boguski, M. S., Elshourbagy, N., Taylor, J. M., & Gordon, J. I. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 992–996.
Bohonus, V. L., Doolittle, R. F., Pontes, M., & Strong, D. D. (1986) *Biochemistry 25*, 6512–6516.
Burnette, W. H. (1981) *Anal. Biochem. 112*, 195–203.
Chapman, M. J. (1979) *J. Lipid Res. 21*, 789–853.
Chapman, M. J., Goldstein, S., Mills, G. L., & Leger, C. (1978) *Biochemistry 17*, 4455–4464.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
Doolittle, R. F. (1981) *Science (Washington, D.C.) 214*, 149–159.
Doolittle, R. F. (1984) in *The Plasma Proteins* (Putnam, F. W., Ed.) Vol. IV, pp 317–360, Academic Press, New York.
Doolittle, R. F., Cottrell, B. A., & Riley, M. (1976) *Biochim. Biophys. Acta 453*, 439–452.
Feng, D. F., Johnson, M. S., & Doolittle, R. F. (1985) *J. Mol. Evol. 21*, 112–125.
Fitch, W. M. (1977) *Genetics 86*, 623–644.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
George, D. G., Barker, W. C., & Hunt, L. T. (1986) *Nucleic Acids Res. 14*, 11–15.
Kuyas, C., Riley, M., Bubis, J., & Doolittle, R. F. (1983) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 42*, 2085.
Larsson, A., & Fänge, R. (1977) *J. Comp. Biochem. Physiol. 57B*, 191–196.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, A Laboratory Manual*, pp 1–545, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
McLachlan, A. D. (1977) *Nature (London) 267*, 465–466.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Nelson, G., & Shore, V. G. (1974) *J. Biol. Chem. 249*, 536–542.
Norgard, M. V., Emigholz, K., & Monahan, J. J. (1979) *J. Bacteriol. 138*, 270–272.
Rubin, C. M., & Schmid, C. W. (1980) *Nucleic Acids Res. 8*, 4613–4619.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Strong, D. D., Moore, M., Cottrell, B. A., Bohonus, V. L., Pontes, M., Evans, B., Riley, M., & Doolittle, R. F. (1985) *Biochemistry 24*, 92–101.
Thompson, J. A., Blakesley, R. W., Doran, K., Hough, C. J., & Wells, R. D. (1983) *Methods Enzymol. 100*, 368–399.
von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683–4690.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyabe, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.

Isolation of Full-Length Putative Rat Lysophospholipase cDNA Using Improved Methods for mRNA Isolation and cDNA Cloning†

Jang H. Han,‡ Christian Stratowa,§ and William J. Rutter*
Hormone Research Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Received August 14, 1986; Revised Manuscript Received October 30, 1986

ABSTRACT: We have cloned a full-length putative rat pancreatic lysophospholipase cDNA by an improved mRNA isolation method and cDNA cloning strategy. These new methods allow the construction of a cDNA library from the adult rat pancreas in which the majority of recombinant clones contained complete sequences for the corresponding mRNAs. A previously recognized but unidentified long and relatively rare cDNA clone containing the entire sequence from the cap site at the 5′ end to the poly(A) tail at the 3′ end of the mRNA was isolated by single-step screening of the library. The size, amino acid composition, and the activity of the protein expressed in heterologous cells strongly suggest this mRNA codes for lysophospholipase [Van den Bosch, H., Aarsman, A. J., DeJong, G. N., & Van Deenen, L. M. (1973) *Biochim. Biophys. Acta 296*, 94–104].

The molecular cloning of cDNAs has played a key role in elucidating the sequence, organization, and expression of eucaryotic genes. In addition, it has provided the coding sequences for synthesizing specific proteins in heterologous systems. The methodology usually requires screening a library for the cDNA of interest and splicing partial-length cDNAs to provide the full-length cDNA. The availability of an efficient method for obtaining full-length cDNAs would greatly improve the efficacy of various screening methods of cDNA libraries and of course eliminate the need for splicing partial-length cDNAs to obtain full-length coding regions.

† This research was supported by NIH Grants AM21344, GM28520, and AM32994 to W.J.R.
* Address correspondence to this author at the Hormone Research Institute, University of California, San Francisco.
‡ Present address: Department of Virus and Cell Biology, Merck Sharp & Dohme Research Labs, West Point, PA 19486.
§ Present address: Ernst-Boehringer-Institut fur Arzneimitterforschung, Dr. Boehringer-Gasse 5-11, A-1120 Vienna, Austria.

0006-2960/87/0426-1617$01.50/0