Sinha, N. (1983) *Tetrahedron Lett. 24*, 5348.
Sundquist, W. I., Lippard, S. J., & Stollar, B. D. (1986) *Biochemistry 25*, 1520.
Ushay, H. M., Tullius, T. D., & Lippard, S. J. (1981) *Biochemistry 20*, 3744.
van der Veer, J. L., Ligtvoet, G. J., van den Elst, H., & Reedijk, J. (1986) *J. Am. Chem. Soc. 108*, 3860.
van Hemelryck, B., Guittet, E., Chottard, G., Girault, J. P., Huynh-Dinh, T., Lallemand, J.-Y., Igolen, J., & Chottard, J. C. (1984) *J. Am. Chem. Soc. 106*, 3037.
van Hemelryck, B., Guittet, E., Chottard, G., Girault, J. P., Herman, F., Huynh-Dihn, T., Lallemand, J.-Y., Igolen, J., & Chottard, J. C. (1986) *Biochem. Biophys. Res. Commun. 130*, 758.
Villani, G., Hübscher, U., & Butour, J. L. (1988) *Nucleic Acids Res. 16*, 4407.
Vrâna, O., Brabec, V., & Kleinwächter, V. (1986) *Anticancer Drug Des. 1*, 95.
Wilson, W. D., Heyl, B. L., Reddy, R., & Marzilli, L. G. (1982) *Inorg. Chem. 21*, 2527.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley & Sons, New York; and references cited therein.
Yoon, C., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332.

Cloning and Characterization of the Human Colipase cDNA[†,‡]

Mark E. Lowe,[*,§] Jerry L. Rosenblum,[§] Patricia McEwen,[∥] and Arnold W. Strauss[§,∥]

Department of Pediatrics and Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63114

Received July 26, 1989; Revised Manuscript Received September 7, 1989

ABSTRACT: Pancreatic lipase hydrolyzes dietary triglycerides to monoglycerides and fatty acids. In the presence of bile salts, the activity of pancreatic lipase is markedly decreased. The activity can be restored by the addition of colipase, a low molecular weight protein secreted by the pancreas. The action of pancreatic lipase in the gut lumen is dependent upon its interaction with colipase. As a first step in elucidating the molecular events governing the interaction of lipase and colipase with each other and with fatty acids, a cDNA encoding human colipase was isolated from a λgt11 cDNA library with a rabbit polyclonal anti-human colipase antibody. The full-length 525 bp cDNA contained an open reading frame encoding 112 amino acids, including a 17 amino acid signal peptide. The predicted protein sequence contains 100% of the published protein sequence for human colipase determined by chemical methods, but predicts the presence of five additional NH_2-terminal amino acids and four additional COOH-terminal amino acids. Comparison of the predicted protein sequence with the known sequences of colipase from other species reveals regions of extensive identity. In vitro translation of mRNA transcribed from the cDNA gave a protein of the expected molecular size that was processed by pancreatic microsomal membranes. Sequence analysis of the in vitro translation product after processing demonstrated signal peptide cleavage and the presence of a human procolipase, as exists in the pig and horse colipases. DNA blot analysis was consistent with the presence of a single gene for colipase. RNA blot analysis demonstrated tissue-specific expression of colipase mRNA in the pancreas. Thus, we report, for the first time, a cDNA for colipase. The cDNA predicts a human procolipase and suggests that there may also be processing at the COOH-terminus. The regions of identity with colipase from other species will aid in defining the interaction with lipase and lipids through site-specific mutagenesis.

Pancreatic lipase is essential for the intraduodenal conversion of dietary triglycerides to more polar monoglycerides and fatty acids (Borgstrom & Erlanson-Albertsson, 1984). In vitro, pancreatic lipase activity is inhibited by bile salts. The activity is restored by the addition of another pancreatic protein, colipase, to the reaction. Studies on patients with steatorrhea suggest that the degree of fat malabsorption is proportional to the colipase concentration in the pancreatic juices (Gaskin et al., 1984). The combination of this observation and the marked effect of colipase on lipase activity in vitro suggests that colipase is essential for pancreatic lipase activity under physiological conditions.

The molecular events of colipase interaction with pancreatic lipase have not been elucidated. Colipase binds to pancreatic lipase in the presence or absence of lipids (Erlanson-Albertsson, 1980; Sternby & Erlansson-Albertsson, 1982) and, also, binds to interfaces such as lipid micelles (Borgstrom, 1976). The amino acids involved in the properties have not been identified. Various studies, relying on spectral methods or chemical modification, have implicated amino acids or regions that may be important in these properties of colipase (Borgstrom & Erlanson-Albertsson, 1984). Both types of studies may be monitoring the effects of a conformational change in the protein on the residues that are modified or followed spectrophotometrically. Site-specific mutagenesis affords the opportunity to obtain detailed information about the rela-

[†] This work was supported by NIH Grant DK33487A.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02883.
[*] Address correspondence to this author at the Department of Biochemistry and Molecular Biophysics, Box 8231, Washington University School of Medicine, St. Louis, MO 63110.
[§] Department of Pediatrics.
[∥] Department of Biochemistry and Molecular Biophysics.

0006-2960/90/0429-823$02.50/0

tionship of colipase structure to function.

To understand the action of lipase at a molecular level, we have isolated a cDNA clone for human colipase. The cDNA was characterized by dideoxynucleotide sequencing, primer extension, RNA and DNA blot hybridization, and in vitro translation of RNA made from the cDNA. A protein of 112 amino acids with a 17 amino acid signal peptide is encoded by the cDNA and the mRNA expression is specific for pancreatic colipase.

Materials and Methods

Isolation of Human Pancreatic RNA. Total RNA (2.8 mg) was isolated from 1.0 g of human pancreas by the guanidinium isothiocyanate method (Stratowa & Rutter, 1987). Twenty-five micrograms of poly(A)-containing RNA was isolated from 1.0 mg of total RNA on an oligo(dT)–cellulose column (Aviv & Leder, 1972).

Construction of a λgt11 cDNA Library. A cDNA library was generated from 10 μg of mRNA using the Librarian XI system (Invitrogen) according to the manufacturer's directions. The first-strand cDNA synthesis was primed with oligo(dT) primers. *Not*I/*Eco*RI linkers were ligated to the double-stranded cDNA. Size selection was accomplished on a Sephadex G-100 column primarily to remove the excess linkers. Alkaline agarose gel electrophoresis of the first-strand cDNA demonstrated that the majority of the cDNA was between 200 and 5000 bases in size.

Isolation of Human Colipase cDNA Clones. About 200000 plaques from the unamplified library were screened with polyclonal antibody to human colipase as previously described (Mierendorf et al., 1987). The antibody was produced in rabbits using human colipase that was purified to homogeneity from human pancreas following extensive delipidation by published methods (Sternby & Borgstrom, 1979). ^{125}I Protein A was used to identify positive clones.

DNA Sequence Analysis. The sequence at both ends of each clone was determined by the dideoxy-nucleotide chain termination method (Sanger et al., 1977). The largest clone was sequenced on both strands using synthetic oligonucleotides complementary to internal sequences as primers.

RNA Blot Analysis. Twenty micrograms of total RNA from various human tissues and AR42J rat pancreatic acinar cells (ATCC CRL 1492) was separated on denaturing agarose gels and transferred to nylon filters as described (Maniatis et al., 1982). A probe derived from the entire colipase insert was labeled by the random primer method and incubated with the immobilized RNA in 50% formamide and 1.0 M NaCl at 50 °C. The filters were washed in 1.0 × SSC and 1.0% SDS at room temperature for 15 min (low stringency) and in 0.1 × SSC and 1.0 SDS at 65 °C for 15 min (high stringency). RNA from the AR42J cells was prepared by the SDS–proteinase K method (Dillmann et al., 1983). All other RNA species were prepared by the guanidine isothiocyanate method (Stratowa & Rutter, 1987).

DNA Blot Analysis. Fifteen micrograms of human genomic DNA isolated from liver was digested with *Eco*RI or *Hin*dIII, and the products were separated by agarose gel electrophoresis. The DNA was denatured with alkali and transferred to a nylon membrane. The entire cDNA for colipase was labeled by the random primer method and incubated with the immobilized DNA under the same conditions used for the RNA blot.

Primer Extension. A 30 bp synthetic oligonucleotide complementary to the 5′ coding region of the mRNA was labeled with T4 polynucleotide kinase (Maniatis et al., 1982) and employed as a primer to copy the 5′ end of the mRNA present in 20 μg of total human pancreatic RNA (Calzone et al., 1987). The products were analyzed on a 6% acrylamide/urea gel. The molecular size was determined by comparison to a sequencing ladder.

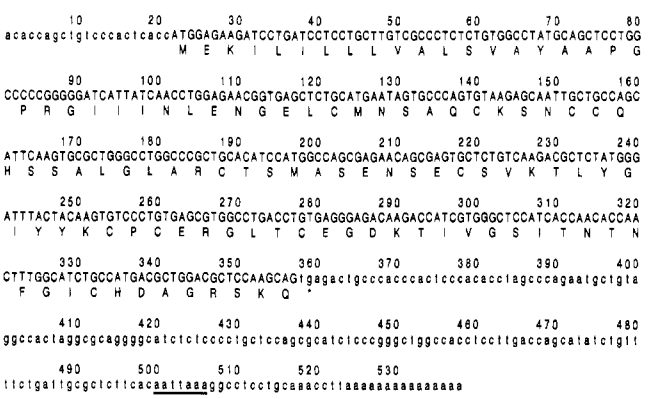

FIGURE 1: Nucleotide sequence and the predicted amino acid sequence for human colipase. The complete nucleotide sequence for the cDNA of human colipase is shown. The 5′- and 3′-untranslated regions are given in lower case letters. The open reading frame encoding human colipase is given in upper case letters. The predicted protein sequence is given below each codon in the single-letter amino acid code. The polyadenylation addition signal sites are underlined.

In Vitro Translation of Human Pancreatic Lipase. Colipase mRNA was generated from the cDNA in pGEM 4z (Promega) using the Riboprobe method according to the manufacturer's instructions. The mRNA was translated in a reticulocyte lysate system in the presence or absence of dog pancreatic microsomes (Mueckler & Lodish, 1986). The [^{35}S]methionine-labeled products were analyzed by SDS–polyacrylamide gel electrophoresis and fluorography (Laemmli, 1970). The product processed by the microsomes was transferred to Immobilon membranes, the radioactive colipase was detected by radioautography, and the bands were cut from the membrane (Matsudaira, 1987). The amino acid sequence was determined directly from the immobilized protein on an Applied Biosystems gas-phase sequencer. Endoglycosidase H digestion was performed as described previously (Mueckler & Lodish, 1986). Isolation of the products from the translation mixture with polyclonal antibodies was done as previously described (Gordon et al., 1983).

Results and Discussion

Screening for Colipase cDNA. Seven clones were initially isolated with antibody screening, and the longest clones were subcloned into pGEM vectors and sequenced. None of these original clones contained a translation initiation site. A probe was prepared from an *Eco*RI/*Sac*I 80 bp 5′-end fragment, and the library was rescreened. Of the isolated clones, one insert was larger than the others and was sequenced completely on both strands. The insert contained 525 bp plus a poly(A) tail. The cDNA had a 336 bp open reading frame preceded by 21 bp of 5′-untranslated region and followed by 165 bp of 3′-untranslated region (Figure 1). A single polyadenylation additional signal is present at 500 bp. The site is followed by a poly(A) tail.

Amino Acid Sequence Predicted from the cDNA. The open reading frame encodes a 112 amino acid protein with a calculated molecular mass of 11954 Da. The known protein sequence of human colipase determined by chemical methods (Sternby et al., 1984) begins at amino acid 23 of the sequence predicted from the cDNA, and the two sequences are identical until amino acid 91. At this point, the previously reported chemical sequence ends, but our predicted sequence contains four additional amino acids (Figure 2). This homology

```
      -10      -1 1   5    10         20         30
MEKILILLLVALSVAYA APGPRGIIINLENGELCMNSAQCKSNCCQHSSALGLARC
 signal peptide        ::::::::::::::::::::::::::::::::::
                       GIIINLENGELCMNSAQCKSNCCQHSSALGLARC
                              10         20

40         50         60         70         80         90   95
TSMASENSECSVKTLYGIYYKCPCERGLTCEGDKTIVGSITNTNFGICHDAGRSKQ*
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TSMASENSECSVKTLYGIYYKCPCERGLTCEGDKTIVGSITNTNFGICHDAG*
     40         50         60         70              86
```

FIGURE 2: Comparison of the predicted protein sequence with the protein sequence determined by chemical methods. The protein sequence predicted from the cDNA for human colipase is aligned with the protein sequence of human colipase previously determined on the isolation protein by chemical methods (Sternby et al., 1984). The predicted sequence is given in the top line. The amino acids are numbered starting with the NH_2-terminus after the signal peptide is removed. The amino acids of the signal peptide are assigned negative numbers.

```
            -10       -1      5    10         20         30
HUMAN  MEKILILLLVALSVAYA APGPRGIIINLENGELCMNSAQCKSNCCQHSSALG
PIG                      V D          DE    L            DTI S
HORSE                    V D  V       A  I L   E   E  HQEES S
                         pro
           40         50         60         70         80
HUMAN  LARCTSMASENSECSVKTLYGIYYKCPCERGLTCEGDKTIVGSITNTNFG
PIG     L  ALK R       AF    V                    SL
HORSE      AAK R       AW    V                 QV  L      M
                             lipid binding
           90   95
HUMAN  ICHDAGRSKQ*
PIG       NV   DSEDKTARLS*
HORSE     F  A  DS*
```

FIGURE 3: Comparison of human, pig, and horse colipase. The predicted protein sequence for human colipase is aligned with the protein sequences of pig and horse colipase that were previously determined from the isolated proteins by chemical methods (Borgstrom & Erlanson-Albertsson, 1984). The single-letter amino acid code is used. Only the differing amino acids are given for the pig and horse sequences, and only one of the two horse isoforms is given. The predicted signal peptide for the human sequence is assigned negative numbers. Boxes are around the NH_2-terminal propeptide (Pro) and the region of central homology (lipid binding) that includes the neighboring tyrosine residues (shown by underlining). Conserved lysine residues are circled.

confirms that our cDNA encodes human colipase.

Comparison of our predicted sequence with the chemically derived protein sequences of pig and horse colipase reveals 76% and 70% identity, respectively (Figure 3). In particular, there is a central region (amino acids 53–73) that is highly conserved, 95% between human and pig, and 86% between human and horse. This region contains the three neighboring tyrosines that have been implicated in the binding of colipase to lipid micelles by NMR (Weiloch & Falk, 1978), fluorescence spectra (McIntyre et al., 1987), UV spectra (Rathelot et al., 1981; Sari et al., 1978), and chemical modification with tetranitromethane or *N*-acetylimidazole (Erlanson et al., 1977; Erlanson-Albertsson, 1980). Colipase binding to lipids is also abolished by chemical modification of lysines with ethyl thiotrifluoroacetate. There are conserved lysines at positions 24, 60, and 73. Lysine-60 flanks the region of the neighboring tyrosines, and lysine-73 is just downstream. The central region of homology may contain the lipid binding site of colipase. Site-specific mutagenesis in this region should enable the identification of the amino acids that are important in this necessary property of colipase.

Colipase has been isolated as a proform containing an NH_2-terminal pentapeptide in both pig and horse [see Borgstrom and Erlanson-Albertsson (1984) for review and references]. No proform has been reported for human colipase. Our predicted sequence is consistent with a proform of the human colipase because a pentapeptide homologous to the propeptides of pig and horse colipase is present (Figure 3). Removal of this pentapeptide would result in the NH_2-terminal glycine found in the protein sequence for human colipase by chemical methods (Figure 2). The presence of the proform predicts a 17 amino acid signal peptide terminating in Ala-Tyr-Ala, a suitable recognition site for signal peptidase (von Heijne, 1983).

Pig procolipase is less active than colipase in restoring pancreatic lipase activity with mixed micelles or in the presence of bile salts (Erlanson-Albertsson & Larsson, 1981). Removal of the pentapeptide from procolipase by limited trypsin digestion restores colipase activity (Weiloch et al., 1981). Because trypsin digestion removes the pentapeptide, procolipase is thought to be secreted by the pancreas and processed by trypsin in the gut lumen. Although trypsin can remove the pentapeptide, the sequence arginine-glycine is the preferred recognition and cleavage site for thrombin (Lundblad et al., 1976). The presence of this sequence raises the possibility that another protease, with a substrate specificity similar to thrombin, is responsible for the activation of procolipase. In either case, the secretion of a procolipase that is poorly active in stimulating lipase activity against mixed micelles is a potential mechanism for protecting the pancreas from autodigestion by lipase in a fashion analogous to the secretion of inactive protease zymogens by the pancreas.

The mechanism by which removal of the NH_2-terminal pentapeptide increases colipase activity is not known. Spectral studies suggest that the activation of procolipase leads to conformational changes in the region of the neighboring tyrosines (Rathelot et al., 1981). Deletions from the NH_2-terminus by Edman degradation have demonstrated the importance of exposing the hydrophobic NH_2-terminal sequence isoleucine-7–9 for colipase activity against mixed micelles containing phospholipids (Erlanson-Albertsson & Larsson, 1981). Additionally, the binding of trypsin-activated colipase to pancreatic lipase in the presence of mixed micelles was increased by an order of magnitude compared to the binding of procolipase in the absence of lipids (Rathelot et al., 1981).

The residues in colipase that are necessary for the binding of colipase to lipase have not been identified. Chemical modification of free carboxylic groups in pig colipase with glycine methyl ester modified 4 out of 13 free carboxylic groups (Erlanson et al., 1977). The modified colipase could bind to lipid micelles but could not anchor lipase on the micelles. The four modified amino acids were located by isolation of tryptic peptides. Of the modified residues, only Glu-15 and Asp-72 are conserved in all species (Figure 3). The role of these residues in colipase interaction with pancreatic lipase could be defined by site-specific mutagenesis.

Trypsin also cleaves a COOH-terminal peptide from colipase (Borgstrom & Erlanson-Albertsson, 1984). Pig colipase, with 10–12 amino acids removed from the COOH-terminus, restores pancreatic lipase activity against phospholipid-containing micelles better than the full-length protein. Comparison of our derived protein sequence from human colipase cDNA with the reported protein sequence of colipase isolated from pancreas reveals that four additional COOH-terminal amino acids are predicted (Figure 2). This discrepency suggests that human colipase is similarly susceptible to protease cleavage at the COOH-terminus. Whether or not the processing at the COOH-terminus affects human colipase activity is not known.

Comparison of our nucleotide sequence for colipase with the sequences in GenBank and of the predicted protein sequence with the sequences in Protein Sequence Database revealed no

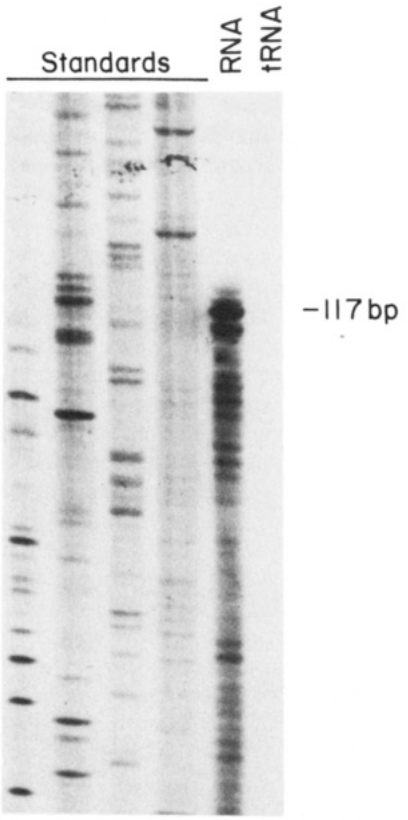

FIGURE 4: Primer extension analysis of human colipase RNA. Twenty micrograms of total pancreatic RNA isolated as described under Materials and Methods was annealed with an oligonucleotide complementary to the 5′-coding region of colipase. The labeled oligonucleotide was extended with a reverse transcriptase. The standards were from a dideoxynucleotide sequence reaction using colipase cDNA as template. As a negative control, yeast tRNA was also a template for the primer extension reaction.

homology with reported sequences. Specifically, there is no homology with ApoCII, a protein cofactor for lipoprotein lipase. This finding is of interest because lipoprotein lipase has extensive homology with pancreatic lipase, and the two, along with hepatic lipase, probably form a gene family (Kirschgessner et al., 1987). The absence of homology between colipase and ApoCII suggests that they interact with their respective lipases at domains that are not homologous.

Location of the Transcription Start Sites. Primer extension analysis of the human colipase mRNA was performed to determine the length of the 5′-untranslated region. A 30 bp oligonucleotide complementary to bases 50–80 of the mRNA was annealed to total pancreatic RNA or yeast tRNA and extended with AMV reverse transcriptase. A major band corresponding to 117 bp was present only with the pancreatic RNA as template (Figure 4). This result demonstrates that the transcription start site lies 37 bases upstream from the proposed translation initiation site.

In Vitro Translation of Human Colipase mRNA. To confirm that the cDNA encoded a protein of the proper size that could be processed by signal peptidase, mRNA was transcribed from the cDNA and translated in a reticulocyte lysate system. A primary translation product of 11 200 Da was synthesized (Figure 5). This size is in good agreement with that predicted (11 954 Da). In the presence of dog pancreatic microsomes, the primary translation product was converted to a protein of 10 700 Da, a value close to the predicted size of 10 111 Da after signal peptide cleavage. Unexpectedly, a second species of about 12 200 Da was also present. Both species were bound by polyclonal antibody

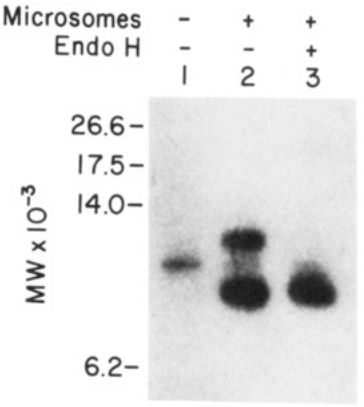

FIGURE 5: In vitro translation of human colipase mRNA. mRNA was transcribed from the human colipase cDNA and translated in a reticulocyte lysate system as described under Materials and Methods. Dog pancreatic microsomes were included in some incubations. Colipase was isolated by binding to polyclonal rabbit antibodies against human colipase and analyzed by SDS–PAGE. The samples are as follows: lane 1, the product of the primary translation; lane 2, the products present in the microsomal pellet; lane 3, the proteins in lane 2 after treatment with endoglycosidase H.

against human colipase. Furthermore, endoglycosidase H digestion of the microsomal contents converted the larger species into the smaller protein.

Confirmation that the two species were both colipase was obtained by amino acid sequence analysis of each after radiolabeling and processing in vitro (data not shown). Human colipase mRNA was translated in the presence of [^{35}S]-methionine and dog pancreatic microsomes. The products remaining with the microsomes were separated by SDS–polyacrylamide gel electrophoresis and transferred to Immobilon membranes. The bands were localized by radioautography and cut from the membrane for analysis by Edman degradation. The fractions from each cycle of the sequencer were counted, and the position of the [^{35}S]methionine was determined. The peak of radioactivity at cycle 18 for both proteins is in agreement with the sequence of procolipase (see Figure 2). This result demonstrates that both species are cleaved by signal peptidase at the same position, that both are species of colipase, and that our cDNA contains the appropriate translation initiation site for colipase.

Because of the endoglycosidase H sensitivity, the simplest explanation for these findings is that a portion of colipase is glycosylated. Review of the amino acid sequence reveals no suitable N-glycosylation acceptor sites. This addition of oligosaccharides to colipase may reflect broadened specificity of the glycosyl transferases during in vitro processing that may not be present in vivo or oligosaccharide addition may not be specific for Asn-X-Ser/Thr in the pancreas with pancreatic proteins. Bovine protein C is N-glycosylated at an Asn-Ala-Cys sequence (Stenflo & Fernlund, 1982). A similar sequence, Asn-Cys-Cys, is present in human colipase at Asn-26, and glycosylation may occur there (see Figure 2). The original report of the isolation of human colipase found carbohydrate present in the final protein at less than molar amounts. Possibly, this result reflected the glycosylation of a portion of colipase rather than contamination with a glycoprotein (Sternby & Borgstrom, 1979).

Tissue-Specific Expression of Human Colipase mRNA. Random primer labeled probes from the entire cDNA detected a single hybridizing band in human pancreatic RNA, but not in RNA from other tissues (Figure 6). In addition to the tissues shown, RNAs from human colon and kidney were examined, and no hybridization was detected (data not shown).

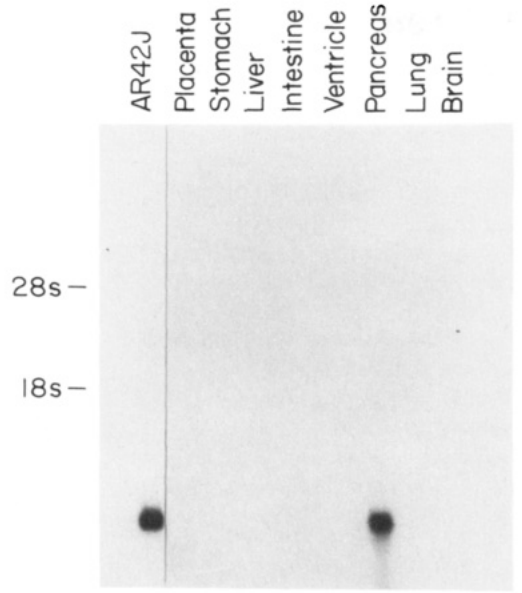

FIGURE 6: Tissue specificity of human colipase mRNA expression. RNA from various tissues and rat pancreatic acinar cell line AR42J was isolated by the guanidinium isothiocyanate method. Twenty micrograms of RNA was separated on an agarose gel and blotted to a nylon membrane for all tissues except for intestine, for which 2 μg of poly(A)-containing RNA way analyzed. The blot was probed with the random primer labeled cDNA and washed as described under Materials and Methods. The lane for AR42J was exposed 10 times longer than the other lanes.

These results demonstrate that pancreatic colipase mRNA expression is specific for pancreas and that a closely related species is not made by other tissues.

A rat pancreatic acinar cell line, AR42J, synthesizes and secretes a number of pancreatic proteins, including pancreatic lipase (Boulet et al., 1986; Scheele et al., 1988). To date, there is no evidence that these cells also synthesize colipase. The colipase cDNA probe detected a single hybridizing band in the AR42J cell RNA (Figure 6). This result demonstrates that these cells do make colipase and that the mRNA is both the same size and homologous to human mRNA for colipase.

The secretion of both pancreatic lipase and colipase is enhanced by secretin and by dietary fat (Erlanson-Albertsson et al., 1987; Kern et al., 1987; Gaskin et al., 1982). Other pancreatic enzymes do not respond to these stimuli, but have hormonal regulators of their own. For instance, caerulein causes an increase in trypsin synthesis and a decrease in amylase synthesis, but does not affect lipase synthesis (Steinhilber et al., 1988). The elements that regulate the discoordinate gene expression of pancreatic enzymes have not been described. Nucleic acid sequences necessary for trypsin and chymotrypsin stimulation by calcium are located in the 5′-flanking regions of the respective genes, but these sequences have not been identified (Stratowa & Rutter, 1986). Because lipase activity depends on the presence of colipase and the secretion of the two proteins is responsive to the same stimuli, speculation that the two genes are modulated coordinately by the same or similar regulatory elements is attractive. Isolating and characterizing the genes for pancreatic lipase and colipase is the first step in locating and describing the DNA sequences that control the tissue-specific and hormonally regulated expression of these proteins. Our cDNA for colipase provides a necessary reagent for isolating the gene for colipase.

DNA Blot Analysis. The entire cDNA was used to probe genomic DNA. Human liver genomic DNA was digested with *Eco*RI, *Hin*dIII, and *Bam*HI, none of which cut the cDNA, separated on an agarose gel, and transferred to a nylon membrane. Hybridization of the random primer labeled probe revealed a single band in the *Bam*HI digest, two major bands and a minor band in the *Eco*RI digest, and a single major band and two minor bands in the *Hin*dIII digest. The results are consistent with the presence of a single gene for colipase.

The isolation and characterization of the first cDNA for colipase is reported in this paper. Colipase activates pancreatic lipase in the presence of bile salts and mixed micelles containing phospholipids and, as such, is essential for the digestion of dietary triglycerides in the gut lumen. Domains for binding both lipase and lipids have been identified in colipase, but the precise amino acids involved are not known. The amino acids that are important to the function of colipase can be identified by the combination of site-specific mutagenesis and structural analysis of X-ray crystallography. The full-length cDNA for colipase reported here and a full-length cDNA for human pancreatic lipase isolated recently by us (Lowe et al., 1989) provide the ability to test the functions of both colipase and pancreatic lipase by changing the primary structure of both proteins.

Added in Proof

After submission of this paper, we became aware of a previous report describing a procolipase from human pancreas (Sternby & Borgstrom, 1984).

Acknowledgments

We thank Dr. Robert Haas and Harold Sims for their invaluable advice and encouragement. Drs. David Perlmutter and Jeffery Gordon were kind enough to review the manuscript.

References

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.
Borgstrom, B. (1976) *FEBS Lett. 71*, 201–204.
Borgstrom, B., & Erlanson-Albertsson, C. (1984) in *Lipases* (Borgstrom, B., & Brockman, H. L., Eds.) pp 152–183, Elsevier, New York.
Boulet, A. M., Erwin, C. R., & Rutter, W. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3599–3603.
Calzone, F. J., Britten, R. J., & Davidson, E. H. (1987) *Methods Enzymol. 152*, 611–632.
Dillmann, W. H., Barrieux, A., Neeley, W. E., & Contreras, P. (1983) *J. Biol. Chem. 258*, 7738–7745.
Erlanson, C., Barrowman, J. A., & Borgstrom, B. (1977) *Biochim. Biophys. Acta 489*, 150–162.
Erlanson-Albertsson, C. (1980) *FEBS Lett. 117*, 295–298.
Erlanson-Albertsson, C., & Larsson, A. (1981) *Biochim. Biophys. Acta 665*, 250–255.
Erlanson-Albertson, C., Larsson, A., & Duan, R. (1987) *Pancreas 2*, 531–535.
Gaskin, K. J., Durie, P. R., Hill, R. E., Lee, L. M., & Forstner, G. G. (1982) *J. Clin. Invest. 69*, 427–434.
Gordon, J. I., Sims, H. F., Lentz, S. R., Edelstein, C., Scanu, A. M., & Strauss, A. W. (1983) *J. Biol. Chem. 258*, 4037–4044.
Kern, H. F., Rausch, U., & Scheele, G. A. (1987) *Gut 28*, 89–94.
Kirchgessner, T. G., Svenson, K. L., Lusis, A. J., & Schotz, M. C. (1987) *J. Biol. Chem. 262*, 8463–8466.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Lowe, M. E., Rosenblum, J. R., & Strauss, A. W. (1989) *J. Biol. Chem. 264*, 20042–20048.
Lundblad, R. L., Kingdon, H. S., & Mann, K. G. (1976) *Methods Enzymol. 35*, 156–176.
Maniatis, T., Fritsch, E. F., & Sambrook, S. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor

Laboratory, Cold Spring Harbor, NY.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035-10038.
McIntyre, J. C., Hundley, P., & Behnke, W. D. (1987) *Biochem. J. 245*, 821-829.
Mierendorf, R. C., Percy, C., & Young, R. A. (1987) *Methods Enzymol. 152*, 458-469.
Mueckler, M., & Lodish, H. (1986) *Cell 44*, 629-637.
Rathelot, J., Canioni, P., Bosc-Bierne, I., Sarda, L., Kamoun, A., Kaptein, R., & Cozzone, P. J. (1981) *Biochim. Biophys. Acta 671*, 155-163.
Sanger, F., Miklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Sari, H., Granon, S., & Semeriva, M. (1978) *FEBS Lett. 95*, 229-234.
Scheele, G. A., Steinhilber, W., Swarovsky, B., & Kern, H. F. (1988) *J. Cell Biol. 107*, 763a.
Steinhilber, W., Poensgen, J., Rausch, U., Kern, H. F., & Scheele, G. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6597-6601.
Stenflo, J., & Fernlund, P. (1982) *J. Biol. Chem. 257*, 12180-12190.
Sternby, B., & Borgstrom, B. (1979) *Biochim. Biophys. Acta 572*, 235-243.
Sternby, B., & Erlanson-Albertsson, C. (1982) *Biochim. Biophys. Acta 711*, 193-195.
Sternby, B., & Borgstrom, B. (1984) *Biochim. Biophys. Acta 786*, 109-112.
Sternby, B., Engstrom, A., Hellman, U., Vihert, A. M., Sternby, N., & Borgstrom, B. (1984) *Biochim. Biophys. Acta 784*, 75-80.
Stratowa, C., & Rutter, W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4292-4296.
Stratowa, C., & Rutter, W. J. (1987) *Biochemistry 26*, 1617-1625.
von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17-21.
Weiloch, T., & Falk, K. (1978) *FEBS Lett. 85*, 271-274.
Weiloch, T., Borgstrom, B., Pieroni, G., Pattus, F., & Verger, R. (1981) *FEBS Lett. 128*, 217-220.

A·T and C·C$^+$ Base Pairs Can Form Simultaneously in a Novel Multistranded DNA Complex†

Eric L. Edwards,‡ Michael H. Patrick,‡,∥ Robert L. Ratliff,§ and Donald M. Gray*,‡
Program in Molecular and Cell Biology, Mail Stop FO 31, The University of Texas at Dallas, Box 830688, Richardson, Texas 75083-0688, and Genetics Group, Life Sciences Division, Mail Stop 886, Los Alamos National Laboratory, Box 1663, Los Alamos, New Mexico 87545

Received April 26, 1989; Revised Manuscript Received September 1, 1989

ABSTRACT: Previous experiments have established that in certain synthetic oligomeric DNA sequences, including mixtures of d(AACC)$_5$ with d(CCTT)$_5$, adenine–thymine (A·T) base pairs form to the exclusion of neighboring protonated cytosine–cytosine (C·C$^+$) base pairs [Edwards, E., Ratliff, R., & Gray, D. (1988) *Biochemistry 27*, 5166-5174]. In the present work, circular dichroism and other measurements were used to study DNA oligomers that represented two additional classes with respect to the formation of A·T and/or C·C$^+$ base pairs. (1) One class included two sets of repeating pentameric DNA sequences, d(CCAAT)$_{3-6}$ and d(AATCC)$_{4,5}$. For both of these sets of oligomers, an increase in the magnitude of the long-wavelength positive CD band centered at about 280 nm occurred as the pH was lowered from 7 to 5 at 0.1 and 0.5 M Na$^+$, indicating that C·C$^+$ base pairs formed. Even though it may have been possible for these oligomers to form duplexes with two antiparallel A·T base pairs per pentamer, no A·T base pairing was detected by monitoring the CD changes at 250 nm. Thus, spectral data showed that as few as 40% C·C$^+$ base pairs were stable in two sets of oligomers in which A·T base pairs did not form adjacent to, or in place of, C·C$^+$ base pairs. (2) Another class of oligomer was represented by d(C$_4$A$_4$T$_4$C$_4$), which was studied by CD, HPLC, and centrifugation experiments. We confirmed previous work that this sequence was able to form both types of base pairs as the pH and temperature were lowered [Gray, D., Cui, T., & Ratliff, R. (1984) *Nucleic Acids Res. 12*, 7565-7580]. We further established that at low pH, where both A·T and C·C$^+$ base pairs formed concurrently in d(C$_4$A$_4$T$_4$C$_4$), there was an increase in molecular weight above that expected for a duplex. Our interpretation of all these data is that those sequences that formed only A·T base pairs had an antiparallel-strand orientation while those that formed only C·C$^+$ base pairs probably had a parallel-strand orientation. We propose that d(C$_4$A$_4$T$_4$C$_4$) at low pH adopted a multistranded structure that contained both parallel-stranded C·C$^+$ base pairs and antiparallel-stranded A·T base pairs.

It is well-known that DNA can adopt a wide variety of conformations. The intensity of recent research on DNA structural polymorphism attests to the awareness that structures other than canonical B-form DNA may play important roles in transcriptional regulation (Nordheim & Rich, 1983), homologous recombination during meiosis or at immunoglobin switch regions (Sen & Gilbert, 1988), relief of torsional stress (Cantor & Efstratiadis, 1984), and telomere formation (Oka & Thomas, 1987; Henderson et al., 1987). Polypurine–polypyrimidine tracts can be located in regions of regulatory importance (Kilpatrick et al., 1986), and it is common for such sequences to have domains with high cytosine content (Evans

† This work was performed by E.L.E. in partial fulfillment of the requirements for the Ph.D. degree in the Program in Molecular and Cell Biology, The University of Texas at Dallas. Support was provided by NIH Research Grant GM 19060 and by Grant AT-503 from the Robert A. Welch Foundation. Work at the Los Alamos National Laboratory was supported in part by the U.S. Department of Energy.
‡ The University of Texas at Dallas.
∥ Current address: Box 41A, Saxeville Rd., Pine River, WI 54965.
§ Los Alamos National Laboratory.

0006-2960/90/0429-828$02.50/0