Structure of Human Milk Bile Salt Activated Lipase†,‡

Tadashi Baba, Deborah Downs, Kenneth W. Jackson, Jordan Tang, and Chi-Sun Wang*

Oklahoma Medical Research Foundation, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma 73104

Received May 10, 1990; Revised Manuscript Received September 12, 1990

ABSTRACT: The structure and some functional sites of human milk bile salt activated lipase (BAL) were studied by cDNA cloning and chemical analysis of the enzyme. Eighteen cDNA clones of human BAL were identified from lactating human breast cDNA libraries in λgt11 and λgt10 with antibody and synthetic oligonucleotides as probes. The sequence of four clones was sufficient to construct a 3018-bp BAL cDNA structure. This sequence codes for an open reading frame of 742 amino acid residues. There is a putative signal sequence of 20 residues which is followed by the amino-terminal sequence of BAL, and the mature BAL contains 722 amino acid residues. The cDNA sequence also contains a 678-base 5′-untranslated sequence, a 97-base 3′-untranslated region, and a 14-base poly(A) tail. The sequence of a 1.8-kbp insert of clone G10-4A differs from that of the other cDNA in that it contains a deletion of 198 bases (1966–2163) corresponding to 66 amino acid residues. By use of BAL cDNA as probe, it was found that the major molecular species of BAL mRNA in human mammary gland HBL-100 cells had a size of 2.9 kb and two minor species had sizes of 3.8 and 5.1 kb by Northern blot analyses. The deduced BAL protein structure contains in the carboxyl-terminal region 16 repeating units of 11 amino acids each. The repeating units have the basic structure Pro-Val-Pro-Pro-Thr-Gly-Asp-Ser-Gly-Ala-Pro with only minor substitutions. The amino acid sequence of human BAL is related to that of pancreatic lysophospholipase, cholesterol esterase, cholinesterase, acetylcholinesterase, and thyroglobulin. Ten of the 14 cyanogen bromide fragments of diisopropyl fluorophosphate inhibited human milk BAL were isolated, determined for N-terminal sequences, analyzed for amino sugars, and tested for some functional properties. These chemical studies established that the active site of human milk BAL is located at serine-194, the N-glycosylation site is present at asparagine-187, the O-glycosylation region is in the 16 repeating units near the C-terminus, and the heparin binding domain is in the N-terminal region. We have also determined the location of disulfide bridges as Cys64–Cys80 and Cys246–Cys257. The cyanogen bromide cleavage and the partial sequencing of CNBr peptides also confirmed the location of methionines in the polypeptide chain as well as the deduced cDNA sequence of BAL.

Human milk contains a bile salt activated lipase (BAL)[1] at a very high level of 0.5–1% of the milk protein (Wang, 1986; Olivecrona & Bengtsson, 1984). The physiological function of this lipase is the digestion of milk fat ingested by infants. Not only is this enzyme stable to acidity in the pH range of infant stomachs (Hernell, 1975; Fredrikzon et al., 1978), its contribution to fat digestion is further supported by the results of the nutritional experiments in human infants (Williamson et al., 1978; Alemi et al., 1981) and in an animal model (Wang et al., 1989). Human BAL has been purified to homogeneity (Wang & Johnson, 1983), and its specificity and kinetic properties are well established (Wang, 1981; Wang & Lee, 1985; Wang et al., 1983, 1988). For the specific hydrolysis of physiological substrates, the binding of bile salts to BAL is essential because it participates in the substrate binding as part of the catalytic mechanism (Hernell, 1975; Wang, 1981; Wang & Lee, 1985; Wang et al., 1983, 1988). In addition, BAL binds heparin with yet unclear physiological and enzymic reasons. These properties suggest that BAL comes from a distinct class of lipase that differs from lipoprotein lipase and other pancreatic lipase. Thus, BAL is an interesting model for the understanding of the structure and function relationships of lipases. For this reason, we embarked on a study of structure and function relationships of this enzyme with a combination of cDNA cloning and chemical studies. In this paper, we report the complete structure of human milk BAL, its functional domains, and its relationship with rat pancreatic lysophospholipase (RPLL) (Han et al., 1987), cholesterol esterase (Kissel et al., 1989), bovine pancreatic cholesterol esterase/lysophospholipase (Kyger et al., 1989), acetylcholinesterase (Schumacher et al., 1986), cholinesterase (Lockridge et al., 1987) and thyroglobulin (Mercken et al., 1985).

Experimental Procedures

Materials

Human milk BAL was purified as described previously (Wang & Johnson, 1983). Rabbit antiserum against human BAL was prepared as reported (Wang, 1981). cDNA libraries from lactating human breast tissue in λgt11 and λgt10 were purchased from Clontech. Iodogen was purchased from Pierce Chemical Co. The radioisotopes iodine-125, [γ-^{32}P]ATP, [α-^{32}P]dATP, and [α-^{35}S]dATP and nylon filters (Hybond-N) were purchased from Amersham. The enzymes used in recombinant DNA manipulations and the nick translation kit

†This work was supported by National Institutes of Health Grant HD-23472.

‡The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05301.

*Address correspondence to this author at the Oklahoma Medical Research Foundation, 825 N.E. 13th St., Oklahoma City, OK 73104.

[1] Abbreviations: BAL, bile salt activated lipase; RPLL, rat pancreatic lysophospholipase.

0006-2960/91/0430-500$02.50/0 © 1991 American Chemical Society

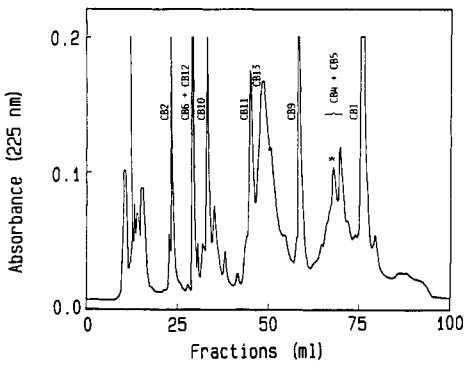


FIGURE 1: Fractionation of [^{3}H]DFP-labeled CNBr-BAL by cation exchange HPLC. Major fractions are characterized by amino acid sequencing analyses as shown in Table II. The peak containing the radioactivity with labeled with an asterisk.

were obtained from Bethesda Research Laboratory. The DNA sequencing kits and reagents were obtained from U.S. Biochemical and Boehringer. The oligo(dT)–cellulose spin-column kit was obtained from 5 Prime → 3 Prime, Inc. The NY-TRAN nylon membranes were purchased from Schleicher & Schuell. Other reagents were the highest grade commercially available and used without further purification. Human HBL-100 cells were obtained from ATCC. Total RNA from human HeLa cells and KG-1 cells were a gift from Dr. J. M. Gimble, Immunobiology and Cancer Research Program, Oklahoma Medical Research Foundation. Both of these cell lines can be obtained from ATCC.

Methods

N-Terminal Amino Acid Sequence Analysis. Automated Edman degradations were performed according to Hewick et al. (1981) in a Model 470A gas-phase protein sequencer equipped with a Model 120A on-line phenylthiohydantoin amino acid analyzer (Applied Biosystems, Inc.).

Isolation of Heparin-Binding CNBr Fragment of Human Milk BAL. Cyanogen bromide cleavage of BAL (70 mg) was carried out under the conditions described by Steers et al. (1965). The heparin-binding CNBr peptide was purified by affinity chromatography on a heparin–Sepharose column (Wang & Johnson, 1983). In this procedure, the CNBr fragments were applied onto the heparin–Sepharose column (2 × 10 cm) preequilibrated with 50 mM NH_4OH–HCl buffer, pH 8.5. The unretained fraction was eluted with 200 mL of the same buffer. The heparin-binding peptide, which was monitored by absorbance at 280 nm, was then eluted with the same buffer containing 0.3 M NaCl. The fractions containing the heparin-binding peptide were pooled, lyophilized, redissolved with 1 mL of distilled water, and desalted on a Sephadex G-50 column (2 × 25 cm). The final yield of lyophilized heparin-binding peptide was about 4 mg.

Polyacrylamide Gel Electrophoresis. Urea–sodium dodecyl sulfate–polyacrylamide gel electrophoresis was carried out in 10% acrylamide gel in the presence of 0.1% sodium dodecyl sulfate and 8 M urea, and the gel was stained with Coomassie brilliant blue as previously described (Wang, 1981).

Screening of cDNA Libraries. About 5 × 10^5 plaques from a λgt11 cDNA library of lactating human breast tissue were screened at 22 °C with rabbit antibodies against human BAL. Rabbit antiserum against BAL was purified on an affinity column of immobilized human milk BAL on Sepharose 4B (Wang et al., 1989). The recovered antibodies were iodinated with iodine-125 and Iodogen (Maxwell & Fox, 1978). For the screening of λgt10 library, about 10^5 plaques were transferred to Hybond-N membranes and probed with syn-

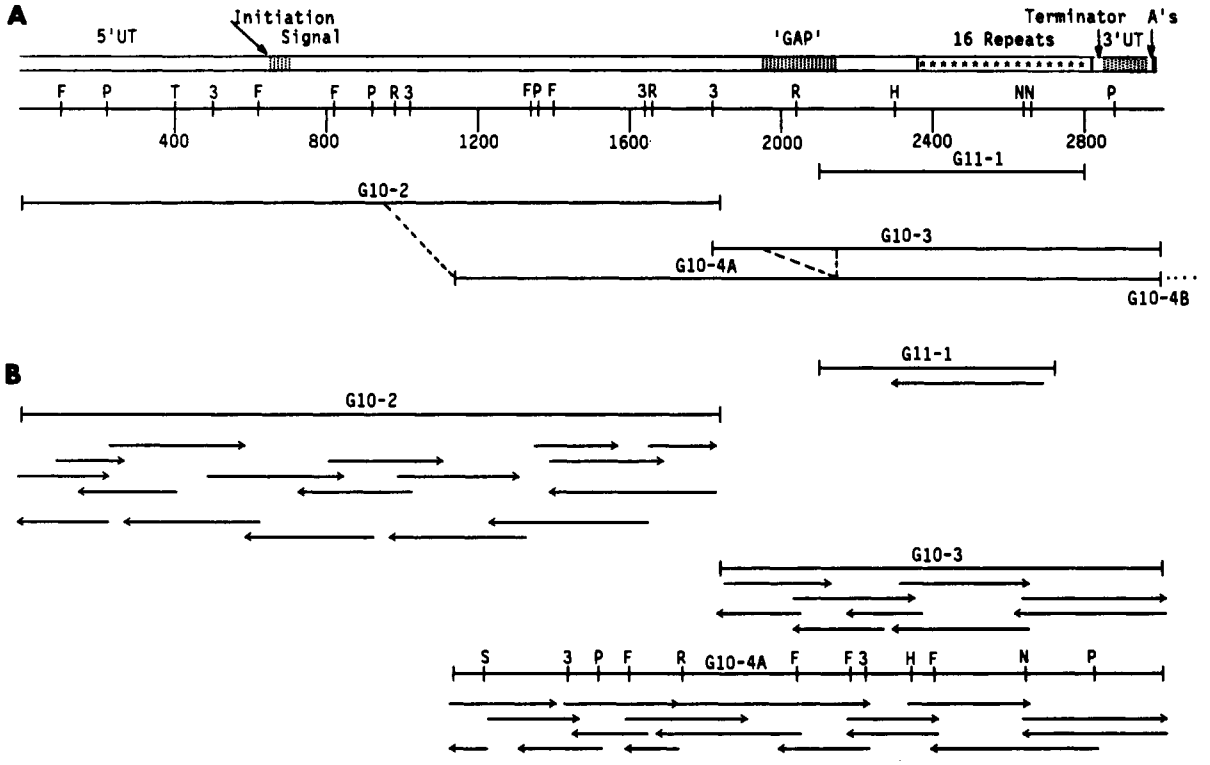


FIGURE 2: Schematic presentation of the cDNA structure of human milk bile salt activated lipase clones from lactating mammary gland. (Panel A) BAL cDNA structure is established from four clones: G11-1, G10-2, G10-3, and G10-4A, as shown by horizontal lines. The bar on the top represents the entire cDNA with different regions marked: 5′-untranslated region (5′UT), initiation codon, leader sequence, the repeating region (16 repeats), termination codon, 3′-untranslated region (3′UT), poly(A) tail (A's), and a GAP region which is a deletion in clone G10-4A. The relationships of G10-4A to other clones are indicated by dashed lines. The nucleotide numbers and the restriction sites are shown on the second line. Restriction sites: T, *Taq*I; 3, *Sau*3AI; F, *Hin*fI; N, *Nar*I; R, *Rsa*I; H, *Hha*I; P, *Pst*I; S, *Sma*I. (Panel B) Schematic presentation of sequence data. Horizontal arrows represent the directions and the covering regions of the sequencing data.

```
CTCAATTGGAGGATCAAAGTTGAGAAAAGTAATATTCGACATTTTTCGATTCAACGGAGT      60

GGCCACCAAGACGATGTCATAGAAGTCTGAACGAGTCTCAGTTCCAATTTGGTAGACCAC     120

TTCATACATCTTTGTTGGATTTCCTGTGTACTTGGTCTTTGTTTTCTCCTCGATGTACAT     180

TACTGAGCCAGATATAAGATTGCTTTTGGATGCCTGCAGAAGCCCTGAGCAAACAAGTTT     240

ATTGCCACCTTCTACTGCCCAAAGGCCAGAATCAGAACAGGACAGTGACACCGCCCCCAC     300

AAAGGCATTGATGTCCGTGCTTTGGCCATAATTGACCCTCATAACAGGAGCAATCATTTC     360

ATTGAGGAACTTCTCAGAAAAGCCGGCCTTTTGCAAGGTTTCAAGAAGTGTTCGATTAAG     420

CATTCCAAGGAAGTCATCTCCTCCTAGAGCATGAAGTAATTTTTCGACACTACTGAAGGC     480

ATAGTCATGAGACTGGTAGCGGTAGATCCTCATGAACTTGTCTAACACGTCCTCTACCCA     540

CATGTGCATACGGAGGGATTGAAATCCATAGCGCCAAACTAATTTAATCACGTTAATTAT     600

GAACCAGTTGCTCTCCTCAAATACCAGAGTCTCTCCATTATATATCCCCAGTAGGCCACC     660

                  MetGlyArgLeuGlnLeuValValLeuGlyLeuThrCysCys
CAGAGGCTGATGCTCACCATGGGGCGCCTGCAACTGGTTGTGTTGGGCCTCACCTGCTGC     720
                    1                                   10
TrpAlaValAlaSerAlaAlaLysLeuGlyAlaValTyrThrGluGlyGlyPheValGlu
TGGGCAGTGGCGAGTGCCGCGAAGCTGGGCGCCGTGTACACAGAAGGTGGGTTCGTGGAA     780
            20                                  30
GlyValAsnLysLysLeuGlyLeuLeuGlyAspSerValAspIlePheLysGlyIlePro
GGCGTCAATAAGAAGCTCGGCCTCCTGGGTGACTCTGTGGACATCTTCAAGGGCATCCCC     840
            40                                  50
PheAlaAlaProThrLysAlaLeuGluAsnProGlnProHisProGlyTrpGlnGlyThr
TTCGCAGCTCCCACCAAGGCCCTGGAAAATCCTCAGCCACATCCTGGCTGGCAAGGGACC     900
            60                                  70
LeuLysAlaLysAsnPheLysLysArgCysLeuGlnAlaThrIleThrGlnAspSerThr
CTGAAGGCCAAGAACTTCAAGAAGAGATGCCTGCAGGCCACCATCACCCAGGACAGCACC     960
            80                                  90
TyrGlyAspGluAspCysLeuTyrLeuAsnIleTrpValProGlnGlyArgLysGlnVal
TACGGGGATGAAGACTGCCTGTACCTCAACATTTGGGTGCCCCAGGGCAGGAAGCAAGTC    1020
           100                                 110
SerArgAspLeuProValMetIleTrpIleTyrGlyGlyAlaPheLeuMetGlySerGly
TCCCGGGACCTGCCCGTTATGATCTGGATCTATGGAGGCGCCTTCCTCATGGGGTCCGGC    1080
           120                                 130
HisGlyAlaAsnPheLeuAsnAsnTyrLeuTyrAspGlyGluGluIleAlaThrArgGly
CATGGGGCCAACTTCCTCAACAACTACCTGTATGACGGCGAGGAGATCGCCACACGCGGA    1140
           140                                 150
AsnValIleValValThrPheAsnTyrArgValGlyProLeuGlyPheLeuSerThrGly
AACGTCATCGTGGTCACCTTCAACTACCGTGTCGGCCCCCTTGGGTTCCTCAGCACTGGG    1200
           160                                 170
AspAlaAsnLeuProGlyAsnTyrGlyLeuArgAspGlnHisMetAlaIleAlaTrpVal
GACGCCAATCTGCCAGGTAACTATGGTCTTCGGGATCAGCACATGGCCATTGCTTGGGTG    1260
           180                 *               190                 ♦
LysArgAsnIleAlaAlaPheGlyGlyAspProAsnAsnIleThrLeuPheGlyGluSer
AAGAGGAATATCGCGGCCTTCGGGGGGGACCCCAACAACATCACGCTCTTCGGGGAGTCT    1320
           200                                 210
AlaGlyGlyAlaSerValSerLeuGlnThrLeuSerProTyrAsnLysGlyLeuIleArg
GCTGGAGGTGCCAGCGTCTCTCTGCAGACCCTCTCCCCCTACAACAAGGGCCTCATCCGG    1380
           220                                 230
ArgAlaIleSerGlnSerGlyValAlaLeuSerProTrpValIleGlnLysAsnProLeu
CGAGCCATCAGCCAGAGCGGCGTGGCCCTGAGTCCCTGGGTCATCCAGAAAAACCCACTC    1440
           240                                 250
PheTrpAlaLysLysValAlaGluLysValGlyCysProValGlyAspAlaAlaArgMet
TTCTGGGCCAAAAAGGTGGCTGAGAAGGTGGGTTGCCCTGTGGGTGATGCCGCCAGGATG    1500
           260                                 270
AlaGlnCysLeuLysValThrAspProArgAlaLeuThrLeuAlaTyrLysValProLeu
GCCCAGTGTCTGAAGGTTACTGATCCCCGAGCCCTGACGCTGGCCTATAAGGTGCCGCTG    1560
           280                                 290
AlaGlyLeuGluTyrProMetLeuHisTyrValGlyPheValProValIleAspGlyAsp
GCAGGCCTGGAGTACCCCATGCTGCACTATGTGGGCTTCGTCCCTGTCATTGATGGAGAC    1620
           300                                 310
PheIleProAlaAspProIleAsnLeuTyrAlaAsnAlaAlaAspIleAspTyrIleAla
TTCATCCCCGCTGACCCGATCAACCTGTACGCCAACGCCGCCGACATCGACTATATAGCA    1680
           320                                 330
GlyThrAsnAsnMetAspGlyHisIlePheAlaSerIleAspMetProAlaIleAsnLys
GGCACCAACAACATGGACGGCCACATCTTCGCCAGCATCGACATGCCTGCCATCAACAAG    1740
```

```
                340                                 350
GlyAsnLysLysValThrGluGluAspPheTyrLysLeuValSerGluPheThrIleThr
GGCAACAAGAAAGTCACGGAGGAGGACTTCTACAAGCTGGTCAGTGAGTTCACAATCACC  1800
                360                                 370
LysGlyLeuArgGlyAlaLysThrThrPheAspValTyrThrGluSerTrpAlaGlnAsp
AAGGGGCTCAGAGGCGCCAAGACGACCTTTGATGTCTACACCGAGTCCTGGGCCCAGGAC  1860
                380                                 390
ProSerGlnGluAsnLysLysLysThrValValAspPheGluThrAspValLeuPheLeu
CCATCCCAGGAGAATAAGAAGAAGACTGTGGTGGACTTTGAGACCGATGTCCTCTTCCTG  1920
                400                                 410
ValProThrGluIleAlaLeuAlaGlnHisArgAlaAsnAlaLysSerAlaLysThrTyr
GTGCCCACCGAGATTGCCCTAGCCCAGCACAGAGCCAATGCCAAGAGTGCCAAGACCTAC  1980
                420                                 430
AlaTyrLeuPheSerHisProSerArgMetProValTyrProLysTrpValGlyAlaAsp
GCCTACCTGTTTTCCCATCCCTCTCGGATGCCCGTCTACCCCAAATGGGTGGGGGCCGAC  2040
                440                                 450
HisAlaAspAspIleGlnTyrValPheGlyLysProPheAlaThrProThrGlyTyrArg
CATGCAGATGACATTCAGTACGTTTTCGGGAAGCCCTTCGCCACCCCCACGGGCTACCGG  2100
                460                                 470
ProGlnAspArgThrValSerLysAlaMetIleAlaTyrTrpThrAsnPheAlaLysThr
CCCCAAGACAGGACAGTCTCTAAGGCCATGATCGCCTACTGGACCAACTTTGCCAAAACA  2160
                480                                 490
GlyAspProAsnMetGlyAspSerAlaValProThrHisTrpGluProTyrThrThrGlu
GGGGACCCCAACATGGGCGACTCGGCTGTGCCCACACACTGGGAACCCTACACTACGGAA  2220
                500                                 510
AsnSerGlyTyrLeuGluIleThrLysLysMetGlySerSerSerMetLysArgSerLeu
AACAGCGGCTACCTGGAGATCACCAAGAAGATGGGCAGCAGCTCCATGAAGCGGAGCCTG  2280
                520                                 530
ArgThrAsnPheLeuArgTyrTrpThrLeuThrTyrLeuAlaLeuProThrValThrAsp
AGAACCAACTTCCTGCGCTACTGGACCCTCACCTATCTGGCGCTGCCCACAGTGACCGAC  2340
                540                                 550
GlnGluAlaThrProValProProThrGlyAspSerGluAlaThrProValProProThr
CAGGAGGCCACCCCTGTGCCCCCCACAGGGGACTCCGAGGCCACTCCCGTGCCCCCCACG  2400
                560                                 570
GlyAspSerGluThrAlaProValProProThrGlyAspSerGlyAlaProProValPro
GGTGACTCCGAGACCGCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCG  2460
                580                                 590
ProThrGlyAspSerGlyAlaProProValProProThrGlyAspSerGlyAlaProPro
CCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCC  2520
                600                                 610
ValProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSerGlyAla
GTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCC  2580
                620                                 630
ProProValProProThrGlyAspSerGlyAlaProProValProProThrGlyAspSer
CCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGCCGCCCACGGGTGACTCC  2640
                640                                 650
GlyAlaProProValProProThrGlyAspAlaGlyProProProValProProThrGly
GGCGCCCCCCCCCGTGCCGCCCACGGGTGACGCCGGGCCCCCCCCCCGTGCCGCCCACGGGT  2700
                660                                 670
AspSerGlyAlaProProValProProThrGlyAspSerGlyAlaProProValThrPro
GACTCCGGCGCCCCCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCCGTGACCCCC  2760
                680                                 690
ThrGlyAspSerGluThrAlaProValProProThrGlyAspSerGlyAlaProProVal
ACGGGTGACTCCGAGACCGCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCCTGTG  2820
                700                                 710
ProProThrGlyAspSerGluAlaAlaProValProProThrAspAspSerLysGluAla
CCCCCCACGGGTGACTCTGAGGCTGCCCCTGTGCCCCCCACAGATGACTCCAAGGAAGCT  2880
                720
GlnMetProAlaValIleArgPheEnd
CAGATGCCTGCAGTCATTAGGTTTTAGCGTCCCATGAGCCTTGGTATCAAGAGGCCACAA  2940

GAGTGGGACCCCAGGGGCTCCCCTCCCATCTTGAGCTCTTCCTGAATAAAGCCTCATACC  3000

CCTGAAAAAAAAAAAAAA                                            3018
```

FIGURE 3: cDNA and amino acid sequences of human milk bile salt activated lipase. Nucleotide sequence is derived from clones G10-2 and G10-3 (Figure 2). The nucleotide numbers are from the 5′ end and shown on the right margin of each line. The predicted amino acid sequence is numbered from the known amino terminus position of matured enzyme. A single potential N-linked glycosylation site is marked by an asterisk. The active site serine is marked by a diamond. The region of 198 nucleotides (1966–2163) deleted in clone G10-4A (GAP region in Figure 2) and the polyadenylation signal are underlined.

thetic oligonucleotides by use of plaque hybridization (Huynh et al., 1985). Probe RP had the repeating unit sequence from clone G11-1 of the λgt11 library. Several other probes were designed and synthesized on the basis of the 61-residue amino acid sequence of the amino-terminal CNBr fragment (see previous section). The oligonucleotides which produced positive results were probe RP, 5′-CCCCGGGCCTCAG-TGGCACCCGCGT-3′, and probe NT1, 5′-CTGCAGCAAATGGGATGCCCTTG/AAAG/AATG/ATCC/GAC-3′ (based on amino-terminal sequence of resi-

dues 30–37). A second screening of 6×10^4 plaques of the λgt10 library was carried out with a *Sau*3AI fragment of clone G10-4A.

Subcloning and DNA Sequence Determination. Phage DNA from positive clones obtained from screening was prepared by the plate lysate method (Maniatis et al., 1982a) followed by the procedure of Bensen and Taylor (1984). cDNA inserts and their restriction fragments from the positive clones were subcloned (Maniatis et al., 1982b) into pUC18, pUC19, M13mp18, and M13mp19 vectors. DNA sequencing using single- or double-strand templates was carried out with the dideoxynucleotide chain-termination method (Sanger et al., 1977).

RNA Purification and Northern (RNA) Blot Analysis. The total RNA was purified by a modification of the method of Chomczynski and Sacchi (1987). The guanidium isothiocyanate cell suspension or tissue homogenates were acidified with sodium acetate, extracted with phenol–chloroform, and twice precipitated with 2-propanol. The further isolation of poly(A+) RNA was performed with an oligo(dT)–cellulose spin-column kit. Total RNA or poly(A+) RNA were suspended in 100 μL of diethyl pyrocarbonate treated water. Twenty-microliter volumes containing approximately 5 μg of poly(A+) RNA or 20 μg of total RNA were mixed with 10 μL of 5.5% formaldehyde–50% formamide, heated at 65 °C for 5 min, and loaded onto a 1% agarose–0.55% formaldehyde gel (Thomas, 1980). The gel was electrophoresed at 40 V for a period of 6–7 h and transferred by capillary action to a NYTRAN nylon membrane, and the RNA was cross-linked by UV irradiation (Church & Gilbert, 1984). Prior to hybridization, the blots were washed for 1 h at 65 °C in 0.1% sodium dodecyl sulfate (SDS)–1× sodium chloride–sodium citrate buffer (SSC; 150 mM NaCl plus 15 mM sodium citrate, pH 7.0).

Nick Translation and Hybridization. For Northern blot analyses, G10-3 was used as probe. The probes were radiolabeled with [α-^{32}-P]dATP, 3000 Ci/mmol, to a specific activity of $>10^8$ cpm/μg by Nick translation. Northern blots were prehybridized for 4 h at 42 °C in 10 mL containing 50% formamide, 5× SSC, 5× Denhardt solution, 0.1% SDS, 5 mM EDTA, 50 mM sodium phosphate, pH 6.8, 10 μg of poly(adenylic acid)/mL, 100 μg of yeast RNA/mL, and 100 μg of sheared salmon sperm DNA/mL. The radiolabeled probes were heated at 100 °C for 5 min and suspended in 5 mL of 40% formamide, 10% dextran sulfate, 5× SSC, 1× Denhardt solution, 20 mM sodium phosphate, pH 6.8, 0.2% SDS, 5 mM EDTA, 10 μg of poly(adenylic acid)/mL, 100 μg of yeast RNA/mL, and 100 μg of salmon sperm DNA/mL. The prehybridization solution was drained from the blot and replaced with the radiolabeled probe in hybridization solution, and the blot was incubated at 42 °C for 14 h. The blot was then washed at low stringency (2× SSC–0.1% SDS, 50 °C, 30 min) and high stringency (0.1× SSC–0.1% SDS, 50 °C, 30 min), wrapped in cellophane while wet, and exposed with an intensifier screen for 3–6 days at −70 °C.

Amino Acid and Hexosamine Analysis. Samples for amino acid analysis were hydrolyzed with constant-boiling HCl in sealed evacuated tubes at 108 °C for 20–24 h. Samples for hexosamine analysis were hydrolyzed similarly, except the time and temperature were 10 h and 100 °C, respectively. Amino acid and hexosamine data were correlated by the addition of β-thienylalanine to the original samples as an internal standard. After hydrolysis, HCl was removed by Speed-Vac concentration. All analyses were performed on a fully automated System Gold amino acid analyzer (Beckman Instruments, Inc., San Ramon, CA) by cation exchange chromatography and postcolumn detection by reaction with ninhydrin. Amino acids were separated by utilization of a two-buffer elution system (sodium citrate, pH 3.28 and pH 7.4) as described by the manufacturer. Under these conditions, the hexosamines coelute with Met, Ile, Leu, β-thienylalanine, Tyr, and Phe and thus interfere with their quantitation. To resolve the hexosamines from these amino acids, the first buffer was changed to 0.2 N sodium citrate, pH 4.25. Under these conditions, some other amino acids were not resolved, but Met, Ile, Leu, β-thienylalanine, Tyr, and Phe eluted well ahead of glucosamine and galactosamine. All samples hydrolyzed for amino acid analysis were run with both procedures, and the data were correlated by normalizing the arginine values. Samples hydrolyzed for hexosamine analysis were run with the pH 4.25 starting buffer.

Labeling of BAL with Diisopropyl Fluorophosphate (DFP). The labeling of BAL with DFP was performed in a total volume of 5 mL with 4 mg/mL BAL, 1 mM [^{3}H]DFP with a specific activity of 10 μCi/μmol, and 2 mM taurocholate in Tris buffer, pH 7.5, and incubated at room temperature for 1 h. At the end of the incubation period there was an almost quantitative (98%) inactivation of BAL by assay of the esterase activity of BAL with *p*-nitrophenyl acetate as substrate (Wang & Johnson, 1983).

Carboxymethylation. Reduction and alkylation of disulfide-linked peptides were performed according to the method of Anfinsen and Harber (1961).

Separation and Sequencing of CNBr Peptides. [^{3}H]DFP-labeled human milk BAL were subjected to cyanogen bromide cleavage (Steers et al., 1965). The resulted peptides were dissolved in the initial equilibration solution and separated by cation exchange HPLC on a 9.4 mm × 200 mm (sulfoethyl)aspartamide column (PolyLC, Columbia, MD). The column was equilibrated with 5 mM KH_2PO_4 in 35/65 CH_3CN/H_2O, pH 3.0 (solvent A). Elution was performed with a linear gradient of 0.5 M KCl in the equilibration solution (solvent B) at a flow rate of 2 mL/min. When necessary, the CNBr peptides were further purified by reverse-phase HPLC on a Synchropak RP-P C18 column (4.6 mm × 250 mm, Synchrom, Inc., Lafayette, IN). The column was equilibrated with 0.1% trifluoroacetic acid/H_2O (solvent A). Elution was effected by a linear gradient of 0.08% trifluoroacetic acid/CH_3CN (solvent B) at a flow rate of 1.0 mL/min. The eluents from cation exchange HPLC and reverse phase were both monitored for absorbance at 215 nm and also for ^{3}H radioactivity. The numbering of the CNBr peptides, CB1 to CB14, is sequential from the N-terminal end. The elution position of CNBr peptides with the cation exchange HPLC of the DFP-labeled BAL is shown in Figure 1. The materials from these peaks were subjected to characterization by automated Edman degradation. The peaks containing fragments CB2, CB10, and CB11 were judged to be heterogeneous and were further purified to purity by reverse-phase HPLC and eluted at 29%, 22% and 19% of solvent B, respectively. Disulfide-linked CB4 and CB5 was also further purified by reverse-phase HPLC and eluted at 37% of solvent B. After reduction of disulfide bond and carboxymethylation (Anfinsen & Haber, 1961), CB4 was eluted at 39% while CB5 was eluted at 33% of solvent B. Purified peptides were subjected to analyses of N-terminal sequences, amino acid compositions, and amino sugars.

RESULTS

Isolation and N-Terminal Sequence Determination of Heparin-Binding CNBr Fragment of Human Milk BAL. In

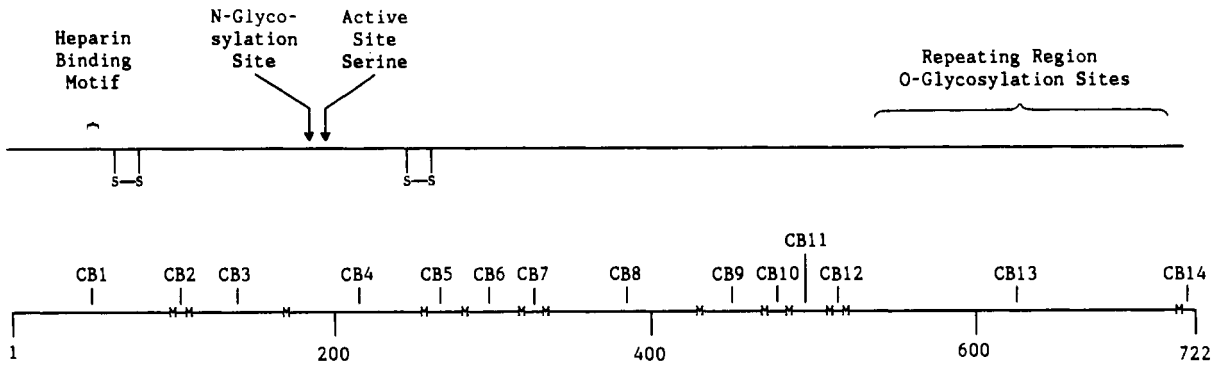


FIGURE 4: Structure organization and some functional sites of human milk BAL. The upper panel shows the N- and O-glycosylation sites, heparin-binding motif, active site serine, and the location of repeating units. The lower panel shows the locations of methionine residues in the BAL sequence and the corresponding cyanogen bromide (CB) peptides.

Table I: Sequencing Analyses of CNBr Peptides Isolated from Cation HPLC[a]

fractions	sequence
CB2	IWIYG-
CB6 + CB12[b]	LHYVG- (CB6)
	GSS- (CB12)
CB10	IAYXTN-
CB11	GDSAVPTH-
CB13	KRSLRTNF-
CB9	PVYPK-
CB4 + CB5[c]	AIAWV- (CB4)
	AQCLK- (CB5)
CB1	AKLGAVYTEG-

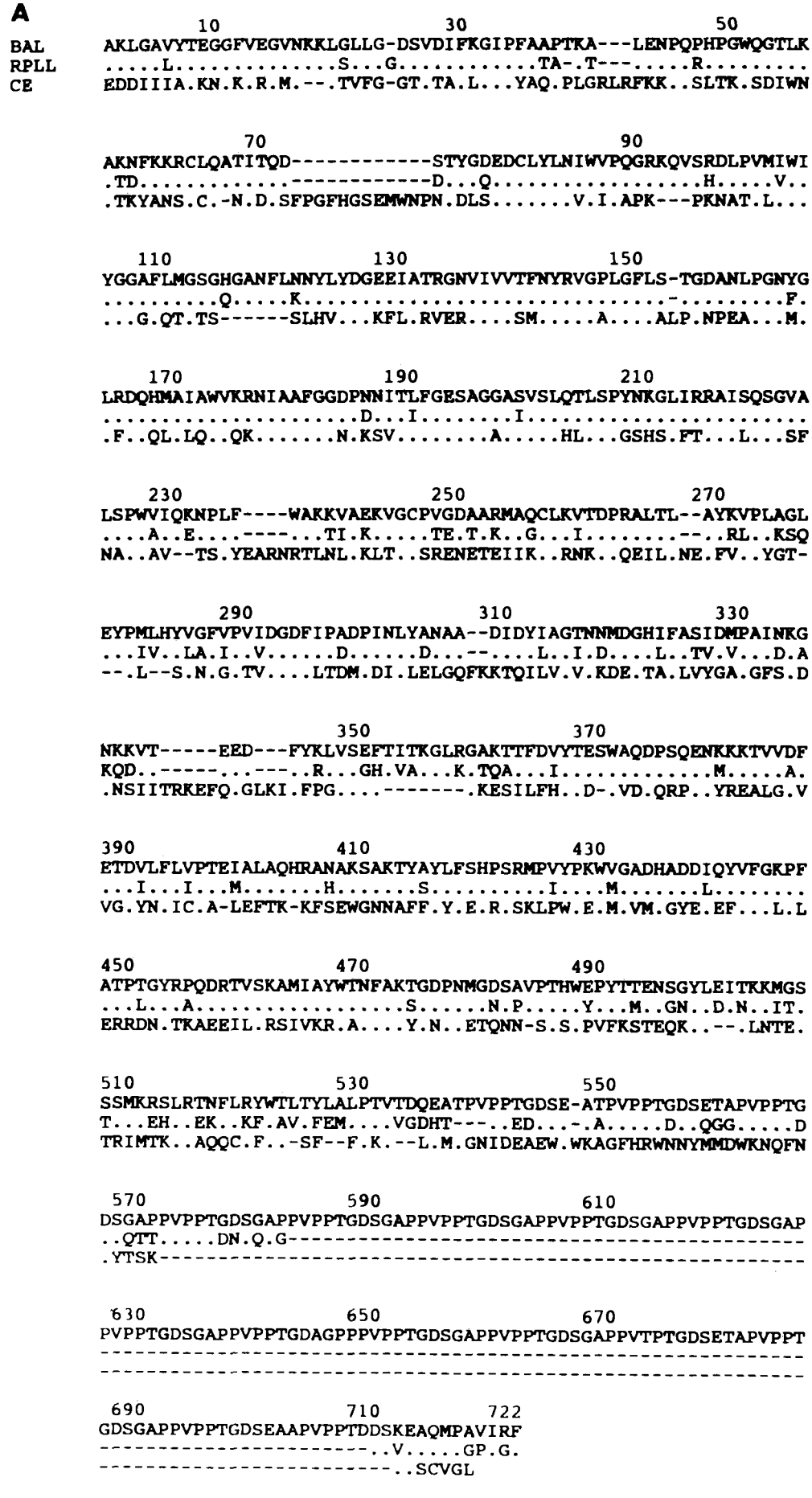
[a]The separation pattern of CNBr-BAL is shown in Figure 1. [b]CB6 represents the major and CB12 represents the minor component of this fraction. [c]The N-terminal sequences of the two peaks for CB4 + CB5 in Figure 1 are the same except that the early peak contained the DFP radioactivity. The DFP-labeled fraction was further carboxymethylated and separated by reverse-phase HPLC, and CB4 and CB5 sequences were determined. Only CB4 contained the DFP radioactivity.

order to construct the cDNA probe for the screening of the cDNA library from the human mammary gland cDNA library, we decided in the early phase of the study to isolate the heparin-binding peptide from the CNBr-degraded BAL because of the ease of purifying the peptide for the deduction of amino acid sequence and for the subsequent probe design. The heparin-binding peptide as isolated from the heparin–Sepharose column was judged to be pure by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, with an apparent molecular weight of 12000. From the automated amino acid sequence analyses we obtained information on the 61 N-terminal amino acid residues of the fragment. The sequence is identical with that derived from the cDNA sequence analyses (see below). The first 23 residues of this sequence (Wang & Johnson, 1983) are identical with the reported N-terminal sequence for BAL except that the 11th residue is glycine instead of the lysine found in the current study. The above results also indicated that the heparin-binding peptide sequence is located in the amino-terminal region of BAL.

BAL cDNA Structure. A total of 18 positive clones were identified from three different screenings. Five positive clones were obtained from human mammary gland cDNA library in λgt11 with BAL antibody as probe. From the λgt10 library, probes RP and NT1 produced three and two positive clones, respectively. Eight positive clones were obtained from the same λgt10 library with as probe a 402 base pair *Sau*3AI fragment (nucleotides 1638–2237, without the "GAP" region; see below) of a partial BAL cDNA obtained above (clone G10-4A). Restriction mapping combined with Southern blots of these clones suggested that all five λgt11 clones were related. The longest clone among these five, Gll-1, was about 0.8 kbp. From the two screenings of the λgt10 library, the longest clones from each probe were G10-2 (1.9 kbp, positive with probe NT1), G10-4 (4 kbp, positive with probe RP), and G10-3 (1.1 kbp, positive with *Sau*3AI fragment of G10-4A). All four clones overlapped with each other as judged by the mapping and Southern blot results. These clones and various fragments from them were subcloned and determined for nucleotide sequences, as summarized in Figure 2. Clones G10-2 and G10-3 and the 5′ side of an *Eco*RI fragment of G10-4 (G10-4A, 1.8 kbp) were completely sequenced (Figure 2). The 3′ *Eco*RI fragment of G10-4 (G10-4B, 2.2 kbp) was downstream from the poly(A) sequence of BAL and is connected to G10-4A by an *Eco*RI linker sequence. Since the G10-4B sequence is totally unrelated to that of BAL, it was obvious that G10-4A and G10-4B were inserted together into clone G10-4 during the ligation step of the library construction. For this reason, we did not study further the sequence of G10-4B. The partial sequence of G11-1 was identical with the 3′-region sequence of G10-3. Since the latter was completely sequenced, the sequencing of clone G11-1 did not continue.

Sequence data indicated that the human BAL cDNA sequence is contained in the combined sequence of clones G10-2 and G10-3 (Figures 2 and 3). This sequence (Figure 3) contains an open reading frame which codes for 742 amino acid residues between the initiation codon (nucleotides 679–681) and the stop codon (nucleotides 2905–2907). The reason that this particular Met codon is chosen as the initiation site, over another potential site upstream, is based on the presence of an optimal initiation flanking sequence, ACCATGG (nucleotides 676–682) (Kozak, 1986). Also, there are 20 predominantly hydrophobic residues between this Met site and the amino-terminal position of the matured BAL (residue 1 in Figure 3). The length of this 20-residue region is appropriate for the signal sequence of BAL. The amino-terminal 61 residues determined by Edman degradation are in complete agreement with the deduced amino acid sequence (Figure 3, residues 1–61). There are 97 bases in the 3′-untranslated region between the termination codon and the 14-base poly(A) tail. The BAL cDNA sequence contains a region of 16 highly similar, internally repeating sequences near the carboxyl terminus (between nucleotides 2343 and 2880). The deduced protein structure of this region forms 16 highly similar repeating units of 11 residues each (Figures 3 and 4). About one-third of the amino acids in this region are prolines. This accounts for the high proline content of BAL (Wang, 1981). The high aspartic acid and glutamic acid content of this region (12.5%) could contribute to the low p*I* (3.7) of BAL. Human

A


10 30 50
BAL AKLGAVYTEGGFVEGVNKKLGLLG-DSVDIFKGIPFAAPTKA---LENPQPHPGWQGTLK
RPLL .....L.....S...G.....TA-.T---.....R.....
CE EDDIIIA.KN.K.R.M.--.TVFG-GT.TA.L...YAQ.PLGRLRFKK..SLTK.SDIWN
70 90
AKNFKKRCLQATITQD----------STYGDEDCLYLNIWVPQGRKQVSRDLPVMIWI
.TD.....----------D...Q.....H.....V..
.TKYANS.C.-N.D.SFPGFHGSEMWNPN.DLS.....V.I.APK---PKNAT.L...
110 130 150
YGGAFLMGSGHGANFLNNYLYDGEEIATRGNVIVVTFNYRVGPLGFLS-TGDANLPGNYG
.....Q.....K.....-.....F.
...G.QT.TS------SLHV...KFL.RVER....SM.....A....ALP.NPEA...M.
170 190 210
LRDQHMAIAWVKRNIAAFGGDPNNITLFGESAGGASVSLQTLSPYNKGLIRRAISQSGVA
.....D...I.....I.....
.F..QL.LQ..QK.....N.KSV.....A.....HL...GSHS.FT...L...SF
230 250 270
LSPWVIQKNPLF----WAKKVAEKVGCPVGDAARMAQCLKVTDPRALTL--AYKVPLAGL
....A..E.....TI.K.....TE.T.K..G...I.....--..RL..KSQ
NA..AV--TS.YEARNRTLNL.KLT..SRENETEIIK..RNK..QEIL.NE.FV..YGT-
290 310 330
EYPMLHYVGFVPVIDGDFIPADPINLYANAA--DIDYIAGTNNMDGHIFASIDMPAINKG
...IV..LA.I..V.....D.....D...--....L..I.D....L..TV.V...D.A
--.L--S.N.G.TV....LTDM.DI.LELGQFKKTQILV.V.KDE.TA.LVYGA.GFS.D
350 370
NKKVT-----EED---FYKLVSEFTITKGLRGAKTTFDVYTESWAQDPSQENKKKTVVDF
KQD..-----.....R...GH.VA...K.TQA...I.....M.....A.
.NSIITRKEFQ.GLKI.FPG....-------.KESILFH..D-.VD.QRP..YREALG.V
390 410 430
ETDVLFLVPTEIALAQHRANAKSAKTYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPF
...I...I...M.....H.....S.....I....M.....L.....
VG.YN.IC.A-LEFTK-KFSEWGNNAFF.Y.E.R.SKLPW.E.M.VM.GYE.EF...L.L
450 470 490
ATPTGYRPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGS
...L...A.....S.....N.P.....Y...M..GN..D.N..IT.
ERRDN.TKAEEIL.RSIVKR.A....Y.N..ETQNN-S.S.PVFKSTEQK..--.LNTE.
510 530 550
SSMKRSLRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSE-ATPVPPTGDSETAPVPPTG
T...EH..EK..KF.AV.FEM....VGDHT----..ED...-.A.....D..QGG.....D
TRIMTK..AQQC.F..-SF--F.K.--L.M.GNIDEAEW.WKAGFHRWNNYMMDWKNQFN
570 590 610
DSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAP
..QTT.....DN.Q.G----------
.YTSK----------
630 650 670
PVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAPPVPPTGDSGAPPVTPTGDSETAPVPPT

---
690 710 722
GDSGAPPVPPTGDSEAAPVPPTDDSKEAQMPAVIRF
----------..V.....GP.G.
----------..SCVGL

B

HUMAN MILK BILE SALT-ACTIVATED LIPASE (BAL)

Repeat #	
1	PVPPTGDSEAT
2	PVPPTGDSETA
3	PVPPTGDSGAP
4	PVPPTGDSGAP
5	PVPPTGDSGAP
6	PVPPTGDSGAP
7	PVPPTGDSGAP
8	PVPPTGDSGAP
9	PVPPTGDSGAP
10	PVPPTGDAGPP
11	PVPPTGDSGAP
12	PVPPTGDSGAP
13	PVTPTGDSETA
14	PVPPTGDSGAP
15	PVPPTGDSEAA
16	PVPPTDDSKEA

RAT PANCREATIC LYSOPHOSPHOLIPASE (RPLL)

Repeat #	
1	HTPPEDDSEAA
2	PVPPTDDSQGG
3	PVPPTDDSQTT
4	PVPPTDNSQAG

C

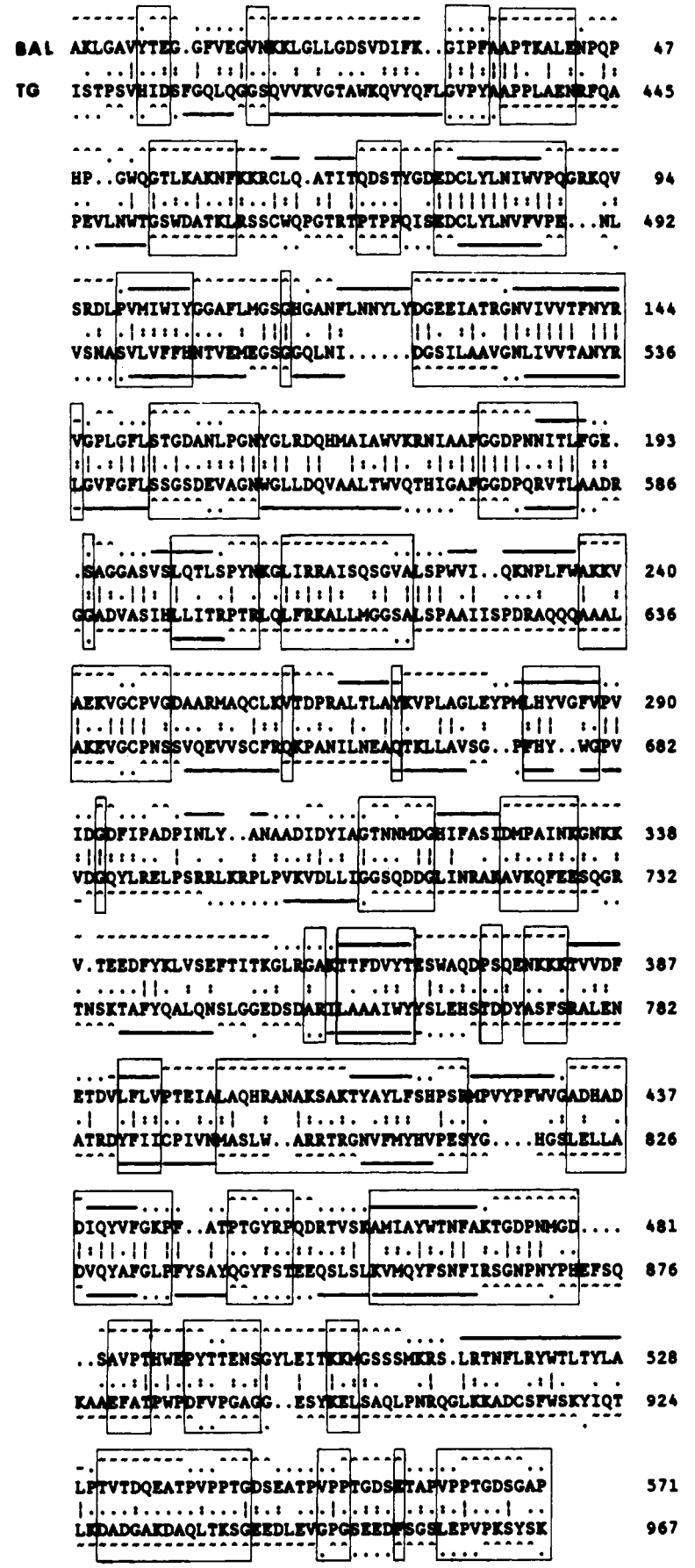


FIGURE 5: (Panel A) Alignment of amino acid sequences of human milk bile salt activated lipase (BAL), rat pancreatic lysophospholipase (RPLL), and cholinesterase (CE). Residues identical with those of BAL are marked with dots on RPLL and CE sequences, and aligned spaces are represented by dahses. (Panel B) Alignment of 14 internal repeating sequences of human milk BAL with the corresponding four internal repeats in RPLL. (Panel C) Comparison of primary and secondary structures of human bile salt activated lipase (BAL, residues 1–571) and a region of rat thyroglobulin (TG, residues 396–967). The alignment of the two sequences was done with a computer program (Devereux et al., 1984) based on creating maximum relationships. Four levels of relatedness between the corresponding residues in two sequences are shown between the sequences: identical residues, vertical lines; strongly similar, two dots; weakly similar, one dot; not related, unmarked. The computer-generated secondary prediction based on Chou and Fasman (1978) is shown above the BAL sequence and below the TG sequence. Secondary structures: α-helix, squiggle; β-structure, solid line; turn, inverted v; random coil, dot. Boxed areas represent regions of the two sequences with strong secondary structure similarities.

BAL is known to be a glycoprotein (Wang, 1981); a single potential N-glycosylation site is observed at residue 187. The molecular weight of mature BAL calculated from the deduced amino acid sequence is 76 282.

The nucleotide sequence of clone G10-4A is identical with the corresponding region of the combined sequence from clones G10-2 and G10-3, as shown in Figure 2, except that a section of 198 bases (nucleotides 1966–2163) is absent (Figures 2 and 3). This represents a deletion of 66 amino acids (residues 410–475). Several possible origins of this shorter cDNA have been considered. Since the nucleotide sequences of long and short cDNAs are otherwise identical, it suggests that they are the products of the same gene. A search of mRNA secondary structure near the "GAP" junctions did not provide any reason to suspect an erroneous copying by reverse transcriptase during the construction of cDNA (results not shown). Also, the introduction of the GAP in the shorter cDNA did not change the reading phase. These facts seem to argue against the possibility that the shorter cDNA is a cloning artifact. The two versions of cDNAs are probably derived from a difference in the splicing of BAL mRNA precursor. Although the gene structure of human BAL is not known, the sequence AAG (nucleotides 1963–1965) just before the GAP and the G's at both sides of the junction at the end of the GAP (nucleotides 2163 and 2164) are the most favorable nucleotides for occurrence at intron/exon junctions (Padgett et al., 1986). These structures are supportive for the alternative splicing explanation. However, definitive conclusion must wait for future verification.

Chemical Studies of BAL Structure. Chemical studies of human milk BAL structure were carried out with the purpose to confirm the amino acid sequence deduced from the cDNA structure, to generate additional structural information such as active site position and disulfide bridges, and to locate other functional regions of the enzyme. Since there are 13 methionines reasonably well situated in the BAL sequence (Figure 3), the strategy for the chemical studies was to isolate most of the cyanogen bromide fragments from the DFP-reacted enzyme and to use these fragments to assess structure and locations of the functional sites. The isolation of CNBr fragments is described under Methods. The N-terminal sequences of purified CNBr fragments (Table I) established their origin in the BAL structure (Figure 4) as deduced from the cDNA sequence. Additional information on BAL structure has also been obtained from analyses of the CNBr fragments.

(*1*) *Active Site Location.* Amino acid sequence deduced from cDNA structure of BAL contains a putative active site motif, GXSXG at residues 192–196, which is conserved among several lipases and esterases (Yang et al., 1989). This suggests that serine-194 may be the active site residue which is the site of reaction with DFP. In the CNBr fragments generated from BAL reacted with radiolabeled DFP, only fragment CB4 contained radioactivity. This result is consistent with the assignment of serine-194 as the active site residue.

(*2*) *Identification of Disulfide Bridges.* Fragment CB4 and CB5 were linked by a disulfide bridge (see Methods) which could be separated only after reduction and carboxymethylation. The single half-cystine in each fragment (Cys-246 and Cys-257) must be disulfide linked in native enzyme. There are four Cys residues (Figure 3) and no free -SH group in the BAL molecule (Abouakil et al., 1989); the remaining two half-cystines, Cys64 and Cys80, must also be disulfide linked to each other.

(*3*) *Glycosylation Sites.* Two CNBr fragments were found to contain amino sugars. Fragment CB13 contains 8 mol of galactosamine and 19 mol of glucosamine per mole of the peptide. Fragment CB4 contains 2 mol of glucosamine. These results indicate that these two fragments contain glycosylation sites.

(*4*) *Presence of GAP Peptide in CNBr Fragments of BAL.* Fragments CB9 and CB10 are derived from a BAL structure without a GAP region (see cDNA structure above). A comparison of PTH-amino acid yields in the N-terminal amino acid analyses of CB1 and CB9 showed that the two fragments were in approximate equimolar yield. This result suggests that if the shorter version of BAL exists, it must be presented in human milk at a very low concentration.

Northern Blots. In order to assess the possible structural relationships of human milk BAL and various liver lipases (see Discussion), Northern blot analyses were performed for RNA of rat and mouse liver with human BAL cDNA as probe. Poly(A+) RNA for human mammary gland epithelial cells HBL-100, which was used as positive control, produced a single predominant band of about 2.9 kb and two minor bands of 3.8 and 5.1 kb. mRNAs from HeLa and KB-1 cells, as expected, produced negative blots and served as negative controls. Mouse and rat liver RNAs produced negative results in Northern blots (results not shown).

Discussion

A schematic presentation of structure domains of human milk BAL based on the deduced cDNA structure and the results from the studies of the CNBr peptides is shown in Figure 4. The structure of BAL cDNA was established from the sequence determination of nearly whole lengths of both strands. Ten of the predicted 14 CNBr fragments were isolated and identified by N-terminal sequence analyses. The structural results from these two studies are in complete agreement. This information confirms that the chemical structure of human milk BAL is a single polypeptide chain of 722 amino acid residues. The enzyme has two intrachain disulfide bridges and two separate glycosylation regions. Direct evidence is provided in this study to show the linkage between Cys-246 and Cys-257 while the other disulfide bond between Cys-64 and Cys-80 is inferred because there are a total of four half-cystines and no free sulfhydryl on the enzyme. BAL most probably contains both N- and O-glycosylation sites. Fragment CB4, which has 2 mol of glucosamine (but no galactosamine), contains the only potential N-glycosylation site of the enzyme molecule in a sequence of Asn-Ile-Thr (residues 187–189). Therefore, we suggest that Asn-187 is N-glycosylated. The other glycosylation region is located in fragment CB13, which is derived primarily from the region containing the 16 repeating sequences. There are many potential O-glycosylation sites in this region since it contains 26 threonines and 16 serines but is without an N-glycosylation signal. There are 8 galactosamines and 19 glucosamines in this fragment, indicating that only part of the potential sites are glycosylated.

The structure of the repeating region is in fact most interesting. It consists of repeating 16 units of 11 residues each of proline-rich sequences near the carboxyl terminus of the enzymes (Figure 5B). The repeat numbers 3–9, 11, 12, and 14 contained the basic sequence PVPPTGDSGAP. Others were found to contain minor substitutions. The presence of aspartic acid in every repeating unit, and glutamic acid in some, renders this structure region highly acidic and contributes to the low isoelectric point (p*I* = 3.7) for BAL. The secondary structure prediction of this region revealed a strong tendency to open random coils (results not shown). Although the function of this structural region is not known, the asso-

ciation of this unique structure with the unique function of bile salt activation of BAL seems to be a reasonable hypothesis.

The conclusion that serine-194 represents the active site serine of human milk BAL is supported by the presence in fragment CB4 of a single G-X-S-X-G motif which is a common sequence for serine-type esterases (Yang et al., 1989). The fact that CB4 contains radiolabeled DFP (Figure 1) from DFP-inactivated enzyme further supports the conclusion.

Evidence has also been provided in this work that the amino-terminal CNBr fragment (CB-1, residues 1–101) retains the heparin-binding property of BAL (see Methods). Some of the heparin-binding sequences of apolipoproteins and lipases are known to be BBBXXB and BBXB (B = basic residues; X = uncharged residues) (Martin et al., 1988). These motifs are not found in fragment CB1. However, a highly basic region with the pattern of BXBXXBBB is located between residues 56 and 63 which may be the heparin-binding site of BAL.

In this work, we have provided structural evidence for the presence of two variants of human BAL cDNA. The only differences between the two cDNAs is that the shorter version contains a deletion, GAP, of 198 bp. Most probably these cDNAs are derived from alternative splicings of human milk BAL mRNA precursor which produce two different lengths of mRNA. We have so far established that the 3′ side of the GAP represents an intron/exon junction of human BAL gene (unpublished observation). This result further supports the alternate splicing explanation. The existence of two different mRNAs would predict the presence of two sizes of BAL proteins in human milk. This possibility was examined in the composition of CNBr fragments because BALs with and without GAP would produce different CNBr fragments. The fact that fragments derived from the GAP region would found in high yields indicated that the longer version of mRNA is predominantly expressed. In the present work, we did not examined all peptide fragments of minor yields. Thus, the presence of small amounts of fragments derived from the shorter version of BAL would not have been detected. The Northern blot analyses with poly(A+) RNA of a human mammary gland derived HLB-100 cell line indicated that the cells did contain BAL mRNA. Since there is only one major 2.9-kb species of BAL mRNA in this cell line, it probably represents the long version without the deletion.

The amino acid sequence of human milk BAL is related to several proteins of known sequences. Among them, the strongest relation is between BAL and rat pancreatic lysophospholipase (Han et al., 1987), RPLL (or rat pancreatic cholesterol esterase; Kissel et al., 1989), which has 65% identical residues in the alignment of the two sequences (Figure 5A). Recently, Kyger et al. (1989) have demonstrated that bovine pancreatic cholesterol esterase/lysophospholipase is also similar in structure to RPLL. The major difference of human milk BAL and RPLL is in the length of the repeating region. The closeness in structural homology of the two enzymes would suggest that RPLL is the bile salt activated lipase because such an activity is known to be present in the pancreas (Wang, 1986, 1988; Abouakil et al., 1988). The difference in the numbers of the repeating units may explain the differences in molecular weights of bile salt activated lipases from different sources (Abouakil et al., 1988). This apparent molecular weight difference may also be amplified by the presence of O-glycosylation sites in the repeating units. The human milk BAL sequence appears unrelated to several other lipases including lipoprotein lipase (Wion et al., 1987; Yang et al., 1989), hepatic triglyceride lipase (Datta et al., 1988; Martin et al., 1988), and pancreatic lipase (DeCaro et al., 1981). The negative Northern blot results of rat and mouse liver RNA, using human milk BAL cDNA as probe, tend to confirm the lack of structural relationships between the milk BAL and liver lipases.

The amino acid sequence of BAL is also related to that of acetylcholinesterase (Schumacher et al., 1986) and cholinesterase (Lockridge et al., 1987). The alignment of choline esterase against BAL produces 26% identical residues (Figure 5A). Alignment of BAL with acetylcholinesterase produced similar results (data not shown) since these esterases are highly related to each other. The strongest homology in these sequence comparisons occurs in the region near the active site of the esterases (Figure 5A). In fact, the sequence around the active site serine (MacPhee-Quigley et al., 1985), FGESAG, in acetylcholinesterase is completely conserved in BAL and RPLL. These observations suggest that serine-194 in the two lipases is the active site and the hydrolytic mechanism of the lipases may be similar to that of acetylcholinesterase and cholinesterase. Despite the sequence homology between BAL and acetylcholinesterase, we could not detect any acetylcholinesterase activity with BAL. The alignment of BAL against a region of rat thyroglobulin between residues 396 and 967 (DiLauro et al., 1985) produced clear homology in several short stretches (Figure 5C). However, comparison of the predicted secondary structures of the two proteins produced a close relationship throughout the entire length of BAL (Figure 5C). These observations imply that this region of thyroglobulin is related to BAL in tertiary structure and it is also an independently folded domain. The five proteins discussed above are likely diverged from a common ancestral protein in evolution.

ACKNOWLEDGMENTS

We thank Dr. Bruce Roe for advice and assistance in the DNA sequencing methods, Dr. J. M. Gimble for help in Northern blot analyses, Dr. Jean A. Hartsuck for suggestions in manuscript preparation, Mr. Gerald Koelsch for assistance in secondary structure analysis, and Ms. Carmeta Stanley and Ms. Joan Pilcher for typing the manuscript.

Registry No. L-Ser, 56-45-1; L-Asn, 70-47-3; heparin, 9005-49-6; lysophospholipase, 9001-85-8; cholesterol esterase, 9026-00-0; acetylcholinesterase, 9000-81-1; cholinesterase, 9001-08-5; DNA (human mammary gland bile salt activated lipase mRNA complementary), 130726-86-2; lipase (human mammary gland bile salt activated protein moiety reduced), 130726-84-0; lipase (human mammary gland bile salt activated protein moiety), 130726-83-9; lipase, 9001-62-1; prelipase (human mammary gland bile salt activated protein moiety reduced), 130726-80-6; pretriacylglycerol lipase (human clone G10-2/G10-3 protein moiety) with disulfides Cys64–Cys80, Cys246–Cys257, 130726-79-3; DNA (human clone G10-4A triacylglycerol lipase mRNA complementary), delete 1966 to 2163, 130726-85-1; triacylglycerol lipase (human clone G10-4A protein moiety reduced), delete 410 to 475, 130726-82-8; triacylglycerol lipase (human clone G10-4A protein moiety), delete 410 to 475, plus with disulfide bonds, 130726-81-7; pretriacylglycerol lipase (human clone G10-4A protein moiety reduced), delete 410 to 475, 130726-78-2; pretriacylglycerol lipase (human clone G10-4A protein moiety), delete 410 to 475, plus with disulfide bonds, 130726-77-1.

REFERENCES

Abouakil, N., Rogalska, E., Bonicel, J., & Lombardo, D. (1988) *Biochim. Biophys. Acta 961*, 299–308.

Abouakil, N., Rogalska, E., & Lombardo, D. (1989) *Biochim. Biophys. Acta 1002*, 225–230.

Alemi, B., Hamosh, M., Scanlon, J. W., Salzman-Mann, C.,

& Hamosh, P. (1981) *Pediatrics 68*, 484–489.
Anfinsen, C. B., & Haber, E. (1961) *J. Biol. Chem. 236*, 1361–1363.
Benson, S. A., & Taylor, R. K. (1984) *BioTechniques May/June*, 126–127.
Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.
Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.
Church, G. M., & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1991–1995.
Datta, S., Luo, C.-C., Li, W.-H., Van Tuinen, P., Ledbetter, D. H., Brown, M. A., Chen, S.-H., Liu, S.-W., & Chan, L. (1988) *J. Biol. Chem. 263*, 1107–1110.
DeCaro, J., Boudonard, M., Bonicel, J., Guidoni, A., Desnuelle, P., & Rovery, M. (1981) *Biochim. Biophys. Acta 671*, 129–138.
Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.
DiLauro, R., Obici, S., Conliffe, D., Ursini, V. M., Musti, A., Moscatelli, C., & Avvedimento, V. E. (1985) *Eur. J. Biochem. 148*, 7–11.
Fredrikzon, B., Hernell, O., Blackberg, L., & Olivecrona, T. (1978) *Pediatr. Res. 12*, 1048–1052.
Han, J. H., Stratowa, C., & Rutter, W. J. (1987) *Biochemistry 26*, 1617–1625.
Hernell, O. (1975) *Eur. J. Clin. Invest. 5*, 267–272.
Hewick, R. M., Hunkapiller, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.
Huynh, T. V., Young, R. A., & Davis, R. W. (1985) *DNA Cloning* (Glover, D. M., Ed.) Vol. I, pp 49–75, IRL Press, Oxford.
Kissel, J. A., Fontaine, R. N., Tuck, C. W., Brockman, H. L., & Hui, D. Y. (1989) *Biochim. Biophys. Acta 1006*, 227–236.
Kozak, M. (1986) *Cell 44*, 283–292.
Kyger, E. M., Weigand, R. C., & Lange, L. G. (1989) *Biochem. Biophys. Res. Commun. 164*, 1302–1309.
Lockridge, D., Bartels, C. F., Vaughan, T. A., Wong, C. K., Norton, S. E., & Johnson, L. L. (1987) *J. Biol. Chem. 262*, 549–557.
MacPhee-Quigley, K., Taylor, P., & Taylor, S. (1985) *J. Biol. Chem. 260*, 12185–12189.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982a) *Molecular Cloning*, pp 63–66, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982b) *Molecular Cloning*, pp 150–178, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Martin, G. A., Busch, S. J., Meredith, G. D., Cardin, A. D., Blankenship, D. T., Mao, S. J. T., Rechtin, A. E., Woods, C. W., Racke, M. M., Schafer, M. P., Fitzgerald, M. C., Burke, D. M., Flanagan, M. A., & Jackson, R. L. (1988) *J. Biol. Chem. 263*, 10907–10914.
Maxwell, M. A. K., & Fox, C. F. (1978) *Biochemistry 17*, 4807–4817.
Mercken, L., Simons, M. J., Swillens, S., Massaer, M., & Vassart, G. (1985) *Nature 316*, 647–651.
Olivecrona, T., & Bengtsson, G. (1984) Lipases in milk in *Lipases* (Borgstrom, B., & Brockman, H. L., Eds.) pp 205–261, Elsevier, Amsterdam, The Netherlands.
Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.
Rudd, E. A., Mizuno, N. K., & Brockman, H. L. (1987) *Biochim. Biophys. Acta 918*, 106–114.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5468.
Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T., & Taylor, P. (1986) *Nature 319*, 407–409.
Steers, E., Jr., Craven, G. R., & Anfinsen, C. B. (1965) *J. Biol. Chem. 240*, 2478–2484.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Wang, C.-S. (1981) *J. Biol. Chem. 256*, 10198–10202.
Wang, C.-S. (1986) in *Fat Absorption* (Kuksis, A., Ed.) Vol. I, pp 83–117, CRC Press, Boca Raton, FL.
Wang, C.-S. (1988) *Biochem. Biophys. Res. Commun. 155*, 950–955.
Wang, C.-S., & Johnson, K. (1983) *Anal. Biochem. 133*, 457–461.
Wang, C.-S., & Lee, D. M. (1985) *J. Lipid Res. 26*, 824–830.
Wang, C.-S., Kuksis, A., Manganaro, F., Myher, J. J., Downs, D., & Bass, H. B. (1983) *J. Biol. Chem. 258*, 9197–9202.
Wang, C.-S., Hartsuck, J. A., & Downs, D. (1988) *Biochemistry 27*, 4834–4840.
Wang, C.-S., Martindale, M. E., King, M. M., & Tang, J. (1989) *Am. J. Clin. Nutr. 49*, 457–463.
Williamson, S., Finucane, H., Ellis, H., & Gamsu, H. R. (1978) *Arch. Dis. Child. 53*, 555–563.
Wion, K. L., Kirchgessner, T. G., Lusis, A. J., Schotz, M. C., & Lawn, R. M. (1987) *Science 235*, 1638–1641.
Yang, C.-Y., Gu, Z.-W., Yang, H.-X., Rohde, M. F., Gotto, A. M., Jr., & Pownall, H. J. (1989) *J. Biol. Chem. 264*, 16822–16827.