Biochem. J. (1987) 243, 597–601 (Printed in Great Britain)

Cloning and sequencing of the medium-chain *S*-acyl fatty acid synthetase thioester hydrolase cDNA from rat mammary gland

Jürgen NAGGERT,* Brenda WILLIAMS,* DANIEL P. CASHMAN† and Stuart SMITH*‡
*Children's Hospital Oakland Research Institute, 747 52nd Street, Oakland, CA 94609, and †Department of Medicine, University of California, San Francisco, CA 94143, U.S.A.

cDNA clones coding for the medium-chain *S*-acyl fatty acid synthetase thioester hydrolase (thioesterase II) from rat mammary gland were identified in a bacteriophage λgt11 library and their nucleotide sequences were determined. The predicted coding region spans 263 amino acid residues and includes a sequence identical with that of a peptide derived from the enzyme active site. The rat thioesterase II cDNA sequence exhibits homology with that of a thioesterase found in duck uropygial glands.

INTRODUCTION

The medium-chain fatty acids characteristic of milk fat are synthesized in the lactating mammary glands of non-ruminants by the action of a tissue-specific chain-terminating enzyme, medium-chain *S*-acyl fatty acid synthetase thioesterase hydrolase (thioesterase II), on fatty acid synthetase [1–3]. Thioesterase II from rat mammary-gland epithelial cells is a monomer of approx. 30000 Da that, although not an integral part of the fatty acid synthetase, is able to hydrolyse acyl chains of intermediate chain length from the 4′-phosphopantetheine prosthetic group of the synthetase [4]. Recently the amino acid sequence of a 57-residue segment of the enzyme containing a di-isopropyl phosphorofluoridate-labelled serine residue has been elucidated, and this indicates that thioesterase II is a serine-active-site esterase [5]. In order to study the tissue-specific expression of thioesterase II, we started by synthesizing, cloning and sequencing the thioesterase II cDNA.

EXPERIMENTAL

Materials

Restriction endonucleases, S_1 nuclease, ribonuclease T_1, reverse transcriptase, DNA polymerase (Klenow fragment) and T_4 DNA ligase were purchased from Boehringer (Mannheim, Germany), λgt11-immunoscreening kit was from Clontech Laboratories (Palo Alto, CA, U.S.A.), deoxy- and dideoxy-ribonucleoside triphosphates were from Pharmacia (Piscataway, NJ, U.S.A.), [α-[^{35}S]thio]dCTP and [α-^{32}P]dCTP were from NEN (Boston, MA, U.S.A.), [γ-^{32}P]ATP was from ICN Biomedicals (Irvine, CA, U.S.A.) and oligo(dT)–cellulose type 3 and dT_{12-18} were from Collaborative Research (Waltham, MA, U.S.A.). Bacteriophage λgt11 [6], Packagene extract, *Escherichia coli* Y1090r$^-$ [7] and *E. coli* Y1089r$^-$ [7] were supplied by Promega Biotech (Madison, WI, U.S.A.), and plasmid pUC12 [8], *E. coli* JM101 [9] and RNA size standards by Bethesda Research Laboratories (Gaithersburg, MD, U.S.A.). Oligonucleotides were synthesized on a Vega Coder 300 (Vega Biotechnologies, Tucson, AZ, U.S.A.) by using cyanoethylphosphoramidite chemistry, deblocked, passed through a Sep Pac (C_{18}) cartridge (Waters, Milford, MA, U.S.A.) and used without further purification.

Construction of a rat mammary-gland cDNA library

RNA was isolated from lactating rat mammary glands by the guanidinium thiocyanate extraction procedure [10] and poly(A)-enriched RNA was obtained by oligo(dT)–cellulose chromatography [10]. Integrity of the isolated poly(A)-enriched RNA was confirmed by the observation of sharp distinguishable bands upon urea/agarose-gel electrophoresis [11]. A cDNA library was constructed by the method of Huynh *et al.* [12], with the following modifications. (*a*) Poly(A)-enriched RNA (2 μg) and dT_{12-18} primer were denatured by heating together for 3 min at 70 °C and then quenched on ice. (*b*) After the synthesis of the first strand the nucleic acids were precipitated from 2.5 M-ammonium acetate with ethanol (final concentration 65%, v/v) and dried by vacuum centrifugation. (*c*) Synthesis of the second strand was done by ribonuclease H/DNA polymerase I replacement synthesis [13]. The reaction was carried out for 1 h at 12 °C followed by 2 h at 22 °C in 100 μl of 20 mM-Tris/HCl buffer, pH 7.5, containing 4 mM-$MgCl_2$, 1.5 mM-β-NAD^+, 80 mM-KCl, 1 mM-dithiothreitol, 50 μg of bovine serum albumin/ml, 10 mM-$(NH_4)_2SO_4$, 0.1 mM-d(A,C,G,T)TP including 10 μCi of [α-^{32}P]dCTP (500 Ci/mmol), 100 units of *E. coli* DNA ligase/ml, 250 units of DNA polymerase I/ml and 150 units of ribonuclease H/ml. (*d*) The cDNA was ligated to linkers, restricted with *Eco*RI and size-fractionated by electrophoresis in a 2% agarose gel. DNA greater in length than 600 bp was electro-eluted and precipitated from 2.5 M-ammonium acetate with ethanol (final concentration 65%, v/v). (*e*) cDNA (0.1 μg) and dephosphorylated λgt11 arms (1 μg) were co-precipitated and ligated, in a volume of 10 μl, for 20 h at 14 °C by using 0.5 unit of T_4 DNA ligase. Half of the ligation mix was packaged in bacteriophage λ particles by using 100 μl of Packagene extract. The library was amplified in

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00311.
‡ To whom correspondence should be addressed.

Y1090r and the lysate was kept at 4 °C. The yield was 2.1×10^6 plaque-forming units/μg of cDNA and the titre was 2.2×10^{11} plaque-forming units/ml with 76% white plaques.

Screening of the mammary-gland cDNA library

The library was initially screened [12] with affinity-purified monospecific anti-(rat thioesterase II) antibodies [14]: 16 positive clones were identified, two of which (λTE2-3 and λTE2-6) were further investigated. The library was rescreened with a ^{32}P-5′-end-labelled oligonucleotide probe [10] complementary to the coding-strand 5′-end of the λTE2-6 insert. The plaque lifts were done by a standard procedure [10]. The filters were prehybridized for 5 min at 68 °C and then for 30 min at 22 °C in 0.9 M-NaCl/90 mM-Tris/HCl buffer, pH 7.5, containing 6 mM-EDTA, 50 μg of heparin/ml, 100 μg of yeast rRNA/ml, 0.1% (w/v) SDS and 0.1% Nonidet P-40. Hybridization was carried out under the same conditions for 20 h at 22 °C with ^{32}P-5′-end-labelled oligonucleotide (0.5 μCi/ml). The filters were washed twice for 5 min at 22 °C in 0.3 M-NaCl/20 mM-NaH_2PO_4/2 mM-EDTA/0.1% SDS adjusted to pH 7.4 with NaOH and twice for 15 min at 37 °C in 30 mM-NaCl/2 mM-NaH_2PO_4/0.2 mM-EDTA/0.1% SDS adjusted to pH 7.4 with NaOH. Radioautography was performed overnight at −70 °C with Fuji RX film and Quanta III intensifying screens (Du Pont, Wilmington, DE, U.S.A.). The rescreening identified 49 positive clones that were analysed by Southern hybridization [10] with the use of random prime-labelled λTE2-6 insert [15]. One representative clone (λTE2-5) was further investigated.

Subcloning and sequencing

Bacteriophage DNA from plaque-purified clones λTE2-5 and λTE2-6 was isolated [16] and the *Eco*RI inserts were subcloned into plasmid pUC12 by standard procedures [10]. Plasmid DNA was isolated from 50 ml of bacterial cultures [17]. After the ammonium acetate precipitation step residual RNA was digested with ribonuclease T_1 (1000 units/ml) and the reaction mixture was extracted with phenol. The pUC12 inserts were sequenced directly by the dideoxy method [18], with minor modifications: plasmid DNA (1 pmol) was denatured in alkali and co-precipitated with synthetic primer (2–8 pmol) as described previously [19]. [α-^{35}S]thio]dCTP (450/fCi/mmol) was used instead of [α-^{32}P]dCTP, the C-reaction mix was changed to 83.2 μM each in dATP, dGTP, dTTP and dCTP and the reaction temperature was 43 °C. One-fourth of the products of individual reactions was analysed on a 50 cm sequencing wedge gel (0.2–0.4 mm, 8% acrylamide in 8 M-urea/Tris/borate [10]), run at 68 °C with 2400 V applied. The gels were dried on to a glass plate and exposed to Fuji RX film at 20 °C for 36 h.

RESULTS AND DISCUSSION

A cDNA library was constructed with poly(A)-enriched RNA from lactating rat mammary glands. Initially 600000 recombinant clones were screened with anti-(thioesterase II) antibodies, and 16 positive plaques were found and purified and their bacteriophage DNA was isolated. Two clones (λTE2-3 and λTE2-6) with the largest inserts (approx. 2.3 kb) showed the same restriction pattern and produced lysogens containing an immunoreactive fusion protein of about 140000 Da.

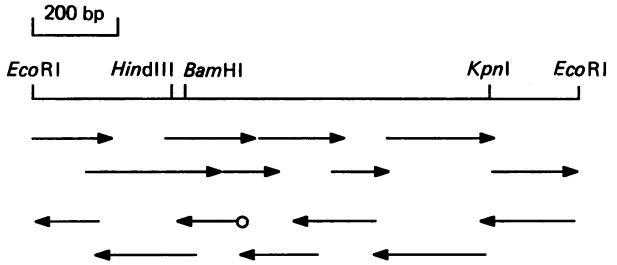

Fig. 1. Partial restriction map and sequencing strategy used for the rat thioesterase II cDNA clone pTE2-512

The arrows indicate the position of the primers and the direction of the sequence as determined by the dideoxy-chain-termination procedure. A circle marks the oligonucleotide used to screen the cDNA library.

The insert from λTE2-6 was subcloned in plasmid pUC12 and sequencing reactions were performed by using the M13 17-mer (−40) universal primer. The deduced protein sequence was identical with a region of 25 amino acid residues surrounding the active site of thioesterase II [5]. However, the coding sequence for the *N*-terminal part of the active-site peptide was missing. Therefore an oligonucleotide complementary to the RNA sequence from the *C*-terminal part of the active-site peptide was synthesized and used to rescreen the library (Fig. 1): 49 positives were identified, 30 of which carried a 1.3 kb insert; the remaining 19 carried smaller inserts. A representative of the larger inserts (λTE2-5) was subcloned into plasmid pUC12 (named pTE2-512) and sequenced (Fig. 1).

The sequence of the insert from clones λTE2-6 and λTE2-3 extends 1000 bp beyond the poly(A) tail found in λTE2-5. Northern-blot and Southern-blot analyses with labelled probes corresponding to the different 3′-non-coding sequences of clones λTE2-6 and λTE2-5 indicated that these sequences are associated with different mRNA species copied from distant regions of the genome (results not shown). It seems probable then that the extended 3′-non-coding sequence found in the long insert of λTE2-6 and λTE2-3 may represent an artifact and may have arisen by ligation of a thioesterase II cDNA to an unrelated cDNA species during construction of the library. However, the possibility that they are representative of mRNA species formed from the same gene by alternative splicing events cannot be completely excluded until the thioesterase II genomic DNA sequence is established.

The entire sequence of the *Eco*RI insert of pTE2-512 is presented in Fig. 2. The 1296 bp sequence contains an open reading frame of 789 bp, starting with a methionine codon 57 bp from the 5′-end after an in-frame termination codon at position −15. Proximal to the methionine codon, at position −4, is the sequence CA and, at position +4, a G believed to be part of a ribosome-binding site in higher eukaryotes [20,21]. The thioesterase II initiation sequence TCACAGA*AUG* G is partially homologous to the consensus sequence CCRCC*AUG* G. Similar deviations from the consensus sequences are found in the rat prepro-elastase I gene (**CCACAAC*AUG* C**) and the rat thyrotropin pre-β-subunit gene (**TCAAAGC*AUG* A**) [20].

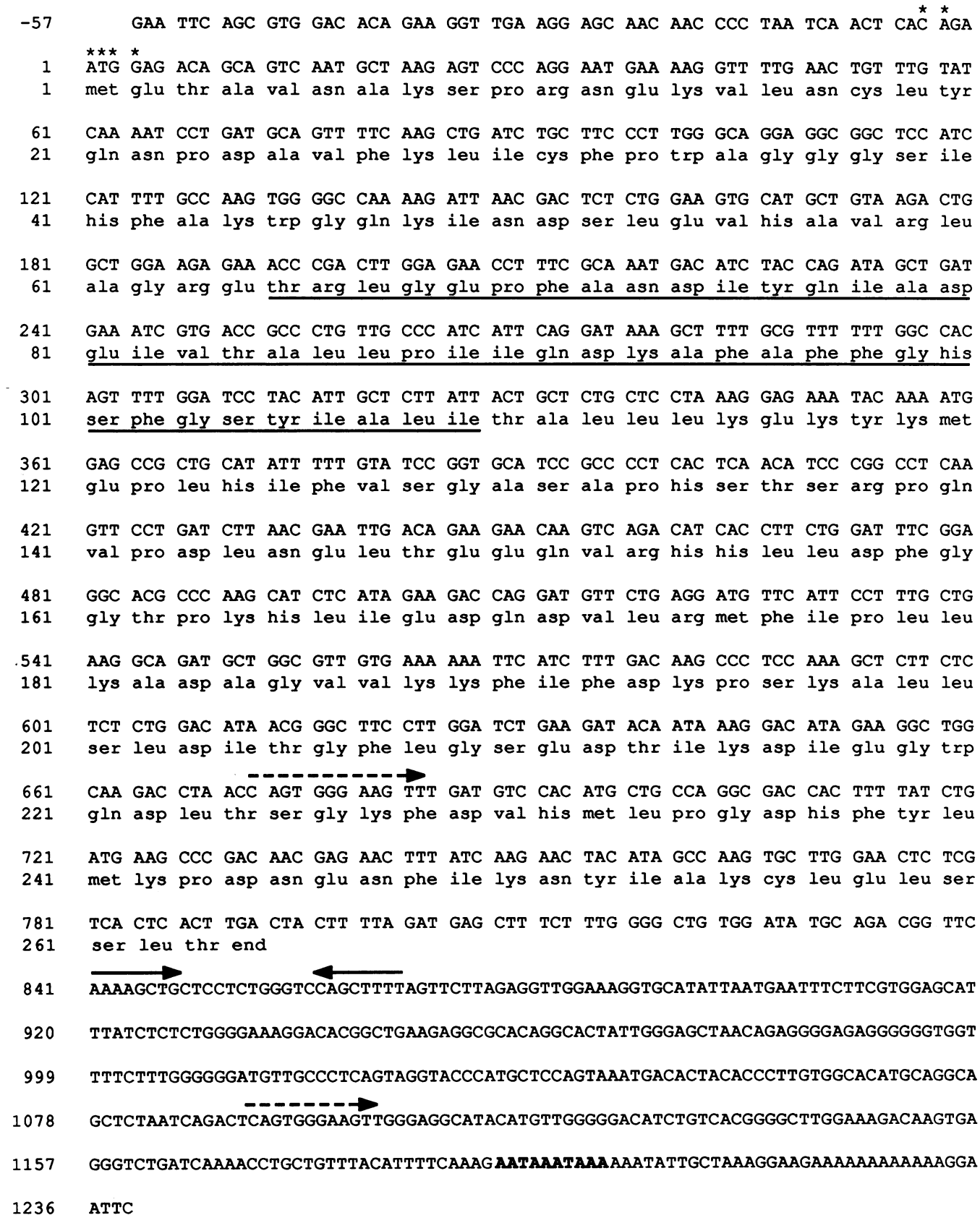

Fig. 2. Sequence of the rat thioesterase II cDNA clone pTE2-512

The deduced amino acid sequence is also shown. The polyadenylation signal is marked in **bold** type. Direct (---►) and inverted (◄——) repeats in the 3′-non-coding region are indicated by arrows. Asterisks indicate the conserved nucleotides of the translation initiation site. The sequence confirmed by Edman degradation of thioesterase II peptides [5] is underlined.

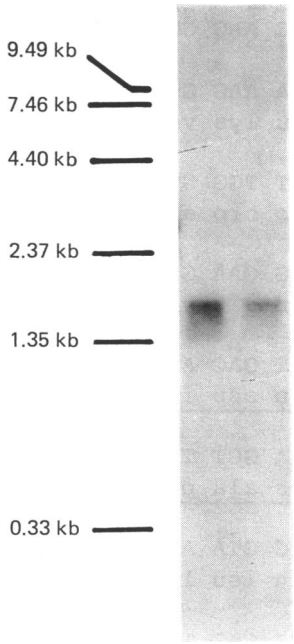

Fig. 3. Northern-blot analysis of lactating rat mammary-gland RNA

Poly(A)-enriched RNA (1 μg) from two different isolations was denatured with formamide, separated on a 1.2% agarose/2.2 mM-formaldehyde gel and transferred to nitrocellulose [22]. The RNA was hybridized [22] to ^{32}P-random-prime-labelled [15] insert from clone pTE2-512. Markers were RNA size standards obtained from Bethesda Research Laboratories.

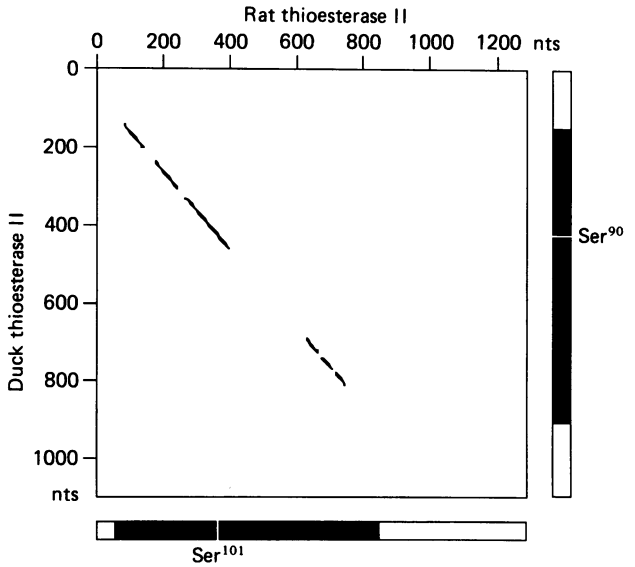

Fig. 4. Homology matrix comparison between rat mammary-gland thioesterase II and mallard-duck uropygial-gland thioesterase II

By using the computer program DNA Inspector II [28] a homology matrix was calculated with a window size of 70 nucleotide residues allowing for 31 mismatches. cDNA sequences are indicated by open bars, the coding sequences by solid bars. The position of the active-site serine residue is given.

The *Eco*RI site at the 5′-end of pTE2-512, GAATTCA, differs from the linker sequence GGAATTCC used during construction of the cDNA library. It seems probable, then, that it is an internal *Eco*RI site within the thioesterase II transcript, and pTE2-512 thus lacks the complete 5′-non-coding region of the thioesterase II mRNA. This view is supported by a Northern-blot analysis, which shows thioesterase mRNA to be approx. 1590 nucleotide residues long (Fig. 3).

Most eukaryotic mRNAs contain a conserved hexanucleotide AATAAA in the 3′ untranslated region, 15–25 nucleotide residues upstream from the polyadenylation site, required for correct processing and polyadenylation [23,24]. In rat thioesterase II there is a polyadenylation signal AATAAATAAA, composed of two overlapping consensus sequences, 22 nucleotide residues upstream of the poly(A) tail. The same signal was recently found in the cDNA for human antileukoproteinase from cervix uterus [25]. Whether dual overlapping polyadenylation signals are more efficient than single signals in promoting mRMA processing is currently not known. At 48 nucleotide residues downstream from the termination codon an eight-nucleotide-residue inverted repeat is found, which could form a hairpin structure.

The open reading frame of 263 amino acid residues, translating to a calculated molecular mass of 29471 Da, is in good agreement with the value obtained by electrophoresis of the protein in SDS/polyacrylamide gels [4]. The serine residue at position 101 in the protein sequence is part of a motif Gly-Xaa-Ser-Xaa-Gly frequently found in serine-active-site esterases [26]. This serine residue was recently shown by labelling experiments to be indeed the reactive serine residue of thioesterase II [5].

Rat thioesterase II is the second medium-chain hydrolase to be sequenced. A similar enzyme has been found in the uropygial glands of waterfowl, which synthesize and secrete lipids containing medium-chain fatty acids [27]. The cDNA sequences of mallard-duck and rat mammary-gland thioesterases show significant homology (Fig. 4). Three areas of homology are seen in the sequence coding for the region between the *N*-terminus of the proteins and the region around the active-site serine residue. In a stretch of 50 nucleotide residues around the serine codon 78% are conserved. The highest values in the other areas reach 60%. Another, smaller, cluster of homologies is in sequences coding for the region near the *C*-termini, reaching values between 48 and 55%. A large stretch of over 200 nucleotide residues in the centre of the coding region is void of homologies. Overall homology of the amino acid sequences encoded by the rat and mallard cDNAs is 40%. The rat and duck thioesterase II enzymes form complexes with their respective fatty acid synthetases that differ markedly in stability. Whereas the avian enzymes associate to form a stable complex even in the absence of substrates, the complex formed by the mammalian enzymes dissociates rapidly following completion of the catalytic process [29]. A structural basis for this functional difference has not been established, and it is tempting to speculate that the region of divergent sequence may play a role in complex-formation. Although the duck thioesterase II cDNA, as well as that of the rat, shows an inverted repeat (seven bases) 48 nucleotide residues downstream from

the termination codon and an 11-base direct repeat 39 and 441 nucleotide residues upstream from the polyadenylation site, no significant homology could be found in the 3′-non-coding region of the cDNAs. In particular, unlike that of the rat thioesterase II, the duck sequence has a conventional polyadenylation signal. The presence of significant homology in the coding regions of the mammary-gland and uropygial-gland cDNAs and the similar function of the encoded thioesterases in regulating the product-specificity of the lipogenic pathway indicates that these enzymes may share a common evolutionary ancestry.

This study was supported by Grants HD 12588 and AM 16073 from the National Institutes of Health.

REFERENCES

1. Knudsen, J., Clark, S. & Dils, R. (1976) Biochem. J. **160**, 683–691
2. Libertini, L. J., Linn, C. Y. & Smith, S. (1976) Fed. Proc. Fed. Am. Soc. Exp. Biol. **35**, 1671
3. Libertini, L. J. & Smith, S. (1978) J. Biol. Chem. **253**, 1393–1401
4. Libertini, L. J. & Smith, S. (1979) Arch. Biochem. Biophys. **192**, 47–60
5. Randhawa, Z. I., Naggert, J., Blacher, R. W. & Smith, S. (1987) Eur. J. Biochem. **162**, 577–581
6. Young, R. A. & Davis, R. W. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 1194–1198
7. Young, R. A. & Davis, R. W. (1983) Science **222**, 778–782
8. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Gene **33**, 103–119
9. Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. **9**, 309–321
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
11. Locker, J. (1979) Anal. Biochem. **98**, 358–367
12. Huynh, T. V., Young, R. A. & Davis, R. W. (1984) in DNA Cloning Techniques: A Practical Approach (Glover, D., ed.), pp. 49–78, IRL Press, Oxford
13. Okayama, H. & Berg, P. (1982) Mol. Cell. Biol. **2**, 161–170
14. Pawlak, J. & Smith, S. (1986) Cancer Res. **46**, 4712–4719
15. Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. **137**, 266–267
16. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984) Experiments with Gene Fusion, pp. 96–97, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
17. Agellon, L. B. & Chen, T. T. (1986) Gene Anal. Tech. **3**, 86–89
18. Hattori, M. & Sakaki, Y. (1986) Anal. Biochem. **152**, 232–238
19. Zagurssky, R., Banmeister, K., Lomax, N. & Berman, R. (1986) Gene Anal. Tech. **2**, 89–94
20. Kozak, M. (1984) Nucleic Acids Res. **12**, 857–872
21. Kozak, M. (1986) Cell (Cambridge, Mass.) **44**, 283–292
22. Seed, B. (1982) in Genetic Engineering: Principles and Methods (Setlow, J. K. & Hollaender, A., eds.), vol. 4, pp. 91–102, Plenum Press, New York
23. Proudfoot, N. J. (1984) Nature (London) **307**, 412–413
24. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) Cell (Cambridge, Mass.) **41**, 349–359
25. Heinzel, R., Appelhans, H., Gassen, G., Seemuller, U., Machleidt, W., Fritz, H. & Steffens, G. (1986) Eur. J. Biochem. **160**, 61–67
26. Dayhoff, M. O., Barker, W. C. & Hardman, J. K. (1976) Atlas of Protein Sequencing and Structure, vol. 5, pp. 53–66, National Biomedical Research Foundation, Washington
27. Poulose, A. J., Rogers, L., Cheesbrough, T. M. & Kolattakudy, P. E. (1985) J. Biol. Chem. **260**, 15953–15958
28. Gross, R. H. (1986) Nucleic Acids Res. **14**, 591–596
29. Mikkelsen, J., Witkowski, A. & Smith, S. (1987) J. Biol. Chem. **262**, 1570–1574

Received 23 December 1986/6 February 1987; accepted 11 February 1987