Characterization of a Genomic and cDNA Clone Coding for the Thioesterase Domain and 3′ Noncoding Region of the Chicken Liver Fatty Acid Synthase Gene†,‡

Ranganna Kasturi, Subrahmanyam Chirala, Mehran Pazirandeh, and Salih J. Wakil*
Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030
Received March 10, 1988; Revised Manuscript Received June 16, 1988

ABSTRACT: The fatty acid synthase (FAS) of animal tissue is a dimer of two identical subunits, each with a M_r of 260 000. The subunit is a single multifunctional protein having seven catalytic activities and a site for binding of the prosthetic group 4′-phosphopantetheine. The mRNA coding for the subunit has an estimated size of 10–16 kb, which is about twice the number of nucleotides needed to code for the estimated 2300 amino acids. We have isolated a positive clone, λCFAS, containing FAS gene sequences by screening a chicken genomic library with a segment of a 3′ untranslated region of goose fatty acid synthase cDNA clone, pGFAS3, as a hybridization probe. The DNA insert in λCFAS hybridizes with synthetic oligonucleotide probes prepared according to the known amino acid sequence of the thioesterase component of the chicken liver fatty acid synthase [Yang, C.-Y., Huang, W.-Y., Chirala, S., & Wakil, S. J. (1988) *Biochemistry* (preceding paper in this issue)]. Further characterization of the DNA insert shows that the λCFAS clone contains about a 4.7-kbp segment from the 3′ end of the chicken FAS gene that codes for a portion of the thioesterase domain. Complete sequence analyses of this segment including S_1 nuclease mapping, showed that the λCFAS clone contains the entire 3′ untranslated region of the chicken FAS gene and three exons that code for 162 amino acids of the thioesterase domain from the COOH-terminal end of the fatty acid synthase. Using the exon region of the genomic clone, we were able to isolate a cDNA clone that codes for the entire thioesterase domain of chicken liver fatty acid synthase. The cDNA clone is sequenced, and the deduced amino acid sequene is consistent with amino acid sequence studies of this domain, except the deduced sequence extends the reported COOH-terminal end of the thioesterase and synthase of chicken liver by four amino acids -Val-Arg-Glu-Gly.

The biosynthesis of long-chain fatty acids from acetyl-CoA is catalyzed by two multienzyme complexes, acetyl-CoA carboxylase and fatty acid synthase. Acetyl-CoA carboxylase is involved in the carboxylation of acetyl-CoA to malonyl-CoA, while the synthase converts acetyl-CoA and malonyl-CoA into palmitate. The fatty acid synthase is possibly the most sophisticated example of the newly discovered class of multifunctional enzymes. It encompasses the entire metabolic pathway for the synthesis of palmitic acid from acetyl-CoA, malonyl-CoA, and NADPH according to the overall reaction

$$CH_3COS\text{-}CoA + 7HOOCCH_2COS\text{-}CoA + 14NADPH + 14H^+ \rightarrow CH_3(CH_2)_{14}COOH + 7CO_2 + 8CoSH + 14NADP^+ + 6H_2O$$

The synthase of animal tissues is a dimer of two identical subunits, each with a M_r of 260 000 (Mattick et al., 1983a,b; Tsukamoto et al., 1983). The subunit is a multifunctional protein containing domains for the seven partial catalytic activities and a site for the prosthetic group 4′-phosphopantetheine. Previous studies showed that the component activities are organized in three subdomains. Domain I contains the NH_2 terminus end of the subunit protein and the partial activities of β-ketoacyl synthase and the acetyl and malonyl transacylase. Domain II contains, in the order indicated, the partial activities of dehydratase, enoyl reductase, β-ketoacyl reductase, and acyl carrier protein (ACP).[1] Domain III connects to the ACP of domain II and contains the thioesterase component of the fatty acid synthase and its COOH terminus residue. In the native state the two subunits are arranged in a head-to-tail fashion so that two sites for palmitate synthesis are constructed (Wakil et al., 1983; Singh et al., 1984).

The activity and concentration of the animal fatty acid synthase vary depending upon the nutritional, hormonal, and developmental status (Wakil et al., 1983; Fischer & Goodridge, 1978; Joshi & Aranda, 1979; Kasturi & Joshi, 1982; Kasturi et al., 1984; Student et al., 1980; Weiss et al., 1980). The changes in the synthase activity and content under these various conditions are due in part to alterations in the rate of enzyme synthesis (Zehner et al., 1977; Fischer & Goodridge, 1978; Joshi & Aranda, 1979; Morris et al., 1982; Goodridge et al., 1984; Nepokroeff et al., 1984). Such changes are accompanied or preceded by corresponding changes in the levels of fatty acid synthase mRNA (Morris et al., 1982, 1984; Goodridge et al., 1984) suggesting that the regulation of synthase mRNA is predominantly at the level of gene transcription (Back et al., 1986). However, the precise molecular mechanism by which the changes in synthase protein are regulated is not yet known.

In order to gain a better understanding of the structure and function of this multifunctional protein and its regulation, it is necessary to isolate and characterize its cDNA and genomic clones. However, the large size of the animal fatty acid synthase mRNA made it difficult to isolate such clones. The

† This work was supported in part by grants from the National Institutes of Health (GM19091), The Robert A. Welch Foundation, and the Clayton Foundation.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02839.
* Author to whom correspondence should be addressed.

[1] Abbreviations: FAS, fatty acid synthase; bp, base pair(s); kb, kilobase(s); kbp, kilobase pair(s); SDS, sodium dodecyl sulfate; TE1 and TE2, oligonucleotide probes; TE, thioesterase.

0006-2960/88/0427-7778$01.50/0

synthase mRNAs of goose uropygial gland and liver (Zehner et al., 1980; Morris et al., 1982; Back et al., 1986) and lactating rat mammary gland (Mattick et al., 1981) are estimated to be about 10–16 kb, which is about twice the number of nucleotides required for the 2300 amino acids needed for the subunit protein synthesis. The excess nucleotides present in the mRNA would have to be distributed between the 5′ and 3′ untranslated regions of the molecule. In 1982 Goodridge and his colleagues isolated a cDNA clone, pGFAS3, for goose uropygial gland fatty acid synthase mRNA (Morris et al., 1982). This clone contains a DNA insert of 1.7 kbp. However, this insert coded only for the 3′ untranslated region of the synthase mRNA. In order to obtain clones that code for the synthase protein, we had to isolate the thioesterase component of the chicken liver fatty acid synthase, determine its amino acid sequence, and use those sequences for the synthesis of oligonucleotide probes to screen genomic and cDNA libraries. The present paper reports the isolation and characterization of such clones and the verification of the amino acid sequence of the thioesterase (Yang et al., 1988).

Materials and Methods

Materials. Restriction enzymes were purchased from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim and were used according to each manufacturer's recommendations. [γ-^{32}P]ATP and [α-^{32}P]dNTPs (<3000 Ci/mmol specific activity) were obtained from Amersham Corp. S_1 nuclease was purchased from Boehringer Mannheim, and DNA polymerase 1, the large-fragment type, was purchased from New England Biolabs. Bacterial alkaline phosphatase and T4 polynucleotide kinase were obtained from Bethesda Research Laboratories. All of the other chemicals used were of the highest degree of purity available.

Screening of λ Plaques for Chicken Fatty Acid Synthase Sequences. The 700-bp *Pvu*II/*Pvu*II fragment of the goose pGFAS3 cDNA clone was subcloned into the pUC18 vector and used for screening a chicken genomic library, which was constructed in λ1059 and provided by Dr. M. Tsai, Department of Cell Biology, Baylor College of Medicine. About 30000 pfu was screened for chicken FAS gene sequences with the nick-translated 700-bp *Pvu*II/*Pvu*II fragment of goose pGFAS3 cDNA as a hybridization probe. Plaques were transferred to nitrocellulose filters. The DNA was denatured and fixed onto nitrocellulose filters according to the procedure of Benton and Davis (1977). The filters were prehybridized overnight at 65 °C in 6× SSC, 50 mM Tris, pH 7.5, 1 mM EDTA, and 4× Denhardt's solution. Hybridization was carried out by incubating the filters overnight in a solution containing 0.5 × 10^6 cpm/mL ^{32}P-labeled probe, 3× SSC, 25 mM Tris, pH 7.5, 0.5 mM EDTA, and 2× Denhardt's solution. The filters were washed two times each in 3× SSC and 0.2% SDS at 65 °C, in 1.5× SSC and 0.2% SDS at 65 °C, and in 2× SSC at room temperature. Each wash was for 15 min. The positive clones were identified by autoradiography.

Miscellaneous Procedures. The large-scale preparation of the λ genomic clone, λCFAS DNA, was carried out according to the methods described by Maniatis et al. (1982). The DNA fragments of the λCFAS clone were subcloned into plasmid pUC18. The large-scale plasmid DNA preparations were made from the subclones by an alkaline lysis method (Birnboim & Doly, 1979) followed by two rounds of ethidium bromide/cesium chloride equilibrium centrifugation.

Southern Blotting and Hybridization. DNA fragments in agarose gel were transferred onto a nitrocellulose filter by the method of Southern (1975). Prehybridization, hybridization, and washing conditions were similar to those that were described for the in situ plaque hybridization when the ^{32}P-labeled 700-bp *Pvu*II/*Pvu*II fragment was used as a hybridization probe. When 5′ end-labeled oligonucleotide probes were used for Southern hybridization, the blots were prehybridized and hybridized as described by Reilly et al. (1982). Prehybridization was carried out overnight at 30 °C in 6× SSC containing 5× Denhardt's. Hybridization was carried out at 30 °C in 6× SET containing 5× Denhardt's and 0.5 × 10^6 cpm/mL 5′ end-labeled oligonucleotide probes. The blots were washed five times for 30 min each at 30 °C and three times for 10 min each at 37 °C. Autoradiography was established with a Du Pont Corenex intensifying screen at −70 °C.

RNA Isolation. White Leghorn chickens, obtained from a local supplier, were starved for 3 days and fed for 2 days with a low-fat, high-carbohydrate diet. The livers were excised and immediately processed for total RNA isolation by a guanidine hydrochloride extraction method (Adams et al., 1977; Deeley et al., 1977). Poly(A+)-containing RNA was prepared from the total RNA by chromatography on oligo(dT)–cellulose (type T3, Collaborative Research) according to the method of Aviv and Leder (1972).

Isolation of cDNA Clones. The plasmid pCFAS36 obtained from λCFAS was used as a primer to synthesize a cDNA bank according to the method of Gubler and Hoffman (1983). The detailed protocol for the plasmid-based primer extension method will be described elsewhere. About 600 plasmid-containing colonies were screened with a *Bgl*II fragment containing the $n-1$ exon of the genomic clone λCFAS as a probe according to the method of Grunstein and Hogness (1975). Among several clones that hybridized with the probe, pCFAS5 was found to be the longest clone.

DNA Sequencing. Subcloned DNA fragments were first sequenced and then joined together by sequencing overlapping fragments. The DNA fragments were dephosphorylated by bacterial alkaline phosphatase, and their 5′ ends were labeled with radioactive phosphate by use of [γ-^{32}P]ATP and T4 polynucleotide kinase (Maxam & Gilbert, 1980). The labeled DNA fragments were digested with appropriate restriction endonucleases and analyzed on 4% or 5% polyacrylamide gels. The desired fragments were isolated and subjected to the sequencing reactions of Maxam and Gilbert (1980) according to modifications described by Chirala and Wakil (1986). About 90% of the DNA sequence was determined on both strands. When the nucleotide sequences were obtained in only one direction, the sequence was confirmed by sequencing more than once. Routinely, 200–250 bp of a sequence could be read correctly. By running xylene cyanol twice to the bottom on a 6% acrylamide urea gel, we occasionally could read 300–400 bp on some fragments, especially those generated by using the most commonly used restriction enzymes, such as *Eco*RI.

S_1 Nuclease Mapping. The procedure of Berk and Sharp (1977) was used for mapping the 3′ end of the λCFAS clone. The DNA fragments were labeled at the 3′ end with the Klenow fragment of DNA polymerase I and cut with an appropriate restriction enzyme. Required fragments were isolated from polyacrylamide gels and used for S_1 nuclease mapping.

Results

Isolation and Characterization of Chicken FAS Genomic Clone. A goose FAS cDNA clone, pGFAS3, has been isolated by Morris et al. (1982) with the procedure of differential hybridization and hybrid-selected translation. We have sequenced this clone completely. Although the sequence is not confirmed (the sequences were carried out only in one direc-

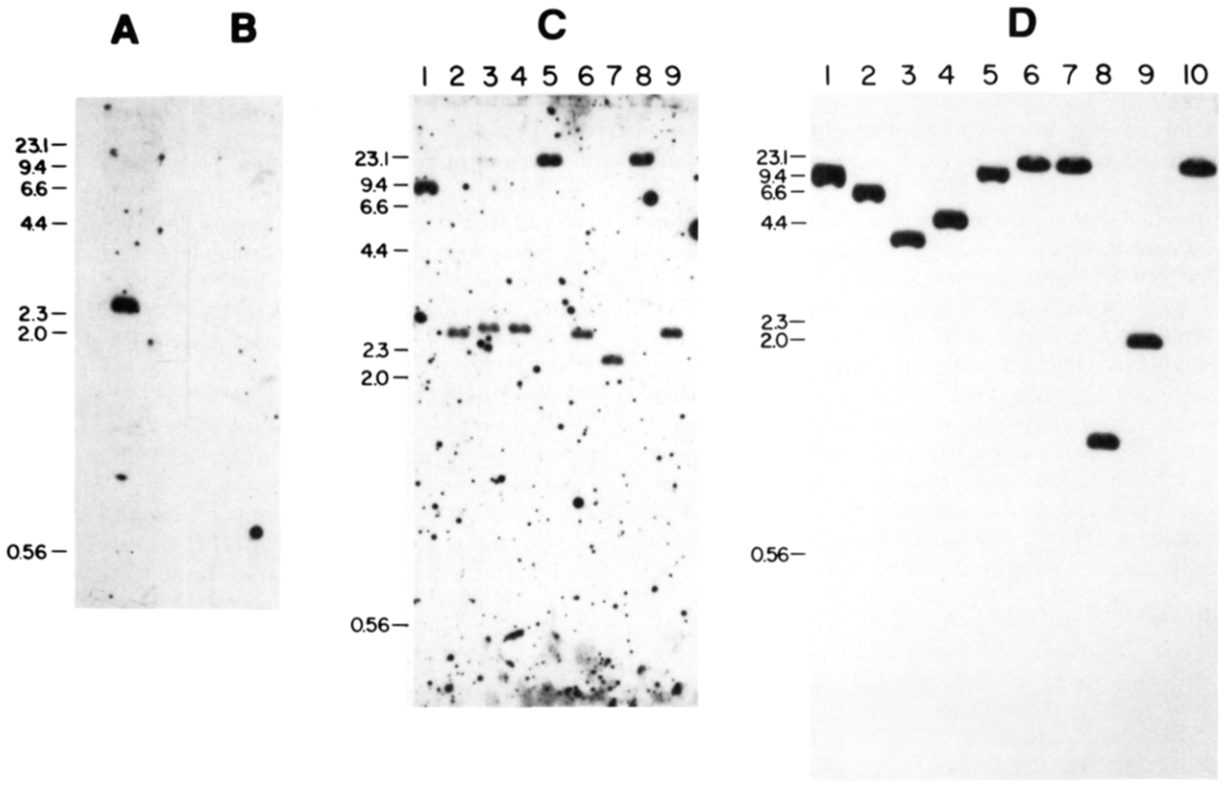

FIGURE 1: Southern blot analysis of chicken λCFAS clone. Phage DNA containing the chicken FAS gene sequences was digested with restriction enzymes. The DNA fragments were subjected to electrophoresis on a 0.8% agarose gel and transferred to a nitrocellulose filter by the standard Southern blotting technique (Southern, 1975). Prehybridization/hybridization and washing procedures were carried out as described under Materials and Methods. λ *Hin*dIII fragments were used as markers. Blots A and B contain the *Eco*RI/*Hin*dIII fragments of the λCFAS clone and were probed with oligonucleotide probes TE1 and TE2, respectively. Blot C contains the *Bam*HI, *Bam*HI/*Eco*RI, *Eco*RI, *Eco*RI/*Hin*dIII, *Hin*dIII, *Pvu*II, *Pvu*II/*Eco*RI, *Xba*I, and *Xba*I/*Eco*RI fragments of the λCFAS clone in lanes 1–9, respectively, and was probed with TE1. Blot D contains the *Bam*HI, *Bam*HI/*Eco*RI, *Bam*HI/*Xba*I, *Bam*HI/*Hin*dIII, *Eco*RI, *Xba*I, *Xba*I/*Hin*dIII, *Xba*I/*Eco*RI, *Hin*dIII/*Eco*RI, and *Hin*dIII fragments, in lanes 1–10, respectively, and was probed with the goose FAS cDNA fragment.

tion; unpublished data), it clearly suggests to us that this cDNA does not code for a protein, but more likely, it codes for an untranslated region of synthase mRNA. Near its 3′ end the cDNA has "CCTTTCTT" sequences which were found to be moderately repeated in the chicken genome (Fisch et al., 1985). Moreover, this cDNA clone was prepared according to G-C tailing protocols which resulted in the presence of stretches of G "and" C at the ends. In order to prevent nonspecific hybridization signals, a central *Pvu*II fragment (700 bp) devoid of these sequences was isolated and subcloned into pUC18. This fragment was then isolated from pUC18, nick translated, and used as a probe to screen a chicken genomic library which was constructed in the λ1059 vector. Since we were not certain about the extent of homology between untranslated regions for goose and chicken, we used moderately stringent conditions for washing the filters after hybridization (see Materials and Methods). About 30000 pfu was screened. A positive clone containing a 13.5-kbp insert was isolated and designated as λCFAS.

The DNA of λCFAS was cleaved with restriction endonucleases and subjected to Southern hybridization analyses using oligonucleotide probes synthesized according to known amino acid sequences of the thioesterase domain of the fatty acid synthase (Yang et al., 1988). Two such probes were prepared. The first, TE1, represented amino acid sequence -His-Asn-Glu-Tyr-Glu-Glu-Gly- (residues 244–250 inclusive) of a peptide region near the COOH terminus. A second probe, TE2, represented amino acid sequence -Asp-Cys-Met-Lys-Gln-Ile- (residues 84–89 inclusive) of a peptide near the active serine site of the thioesterase (Yang et al., 1988). Probe TE1 hybridized strongly to a 2.7-kbp *Eco*RI fragment as shown

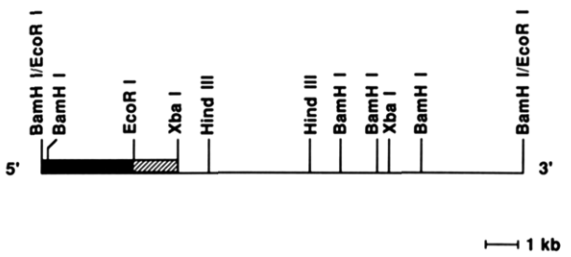

FIGURE 2: Restriction endonuclease mapping of the chicken λCFAS clone. The solid box represents the region that hybridized to the TE1 probe. The hatched box is the portion that hybridized to the goose FAS cDNA probe.

in Figure 1, blot A, while probe TE2 did not hybridize to any of the fragments (Figure 1, blot B), suggesting that the clone λCFAS has DNA segments that code for only a portion of the thioesterase domain. In order to map further the thioesterase coding segment of the λCFAS clone, digestions with seveal restriction endonucleases were carried out, and the fragments were probed with TE1 oligonucleotides. The results of these analyses are shown in Figure 1C. Similar blots were also probed with the ^{32}P-labeled 700-bp *Pvu*II goose pGFAS3 cDNA probe (Figure 1D). On the basis of these results, a restriction map for the chicken λCFAS DNA was constructed, and regions hybridizing with pGFAS3 and TE1 were identified. As shown in Figure 2, the λCFAS clone contains an insert of about 13.5 kbp. Chicken FAS gene sequences were located within 4.9 kbp from one end of the DNA fragment. The region hybridizing to the goose cDNA probe was located within 2.1 kbp of the *Eco*RI/*Hin*dIII fragment. The region spanning the 2.7-kbp *Eco*RI/*Eco*RI fragment contained the

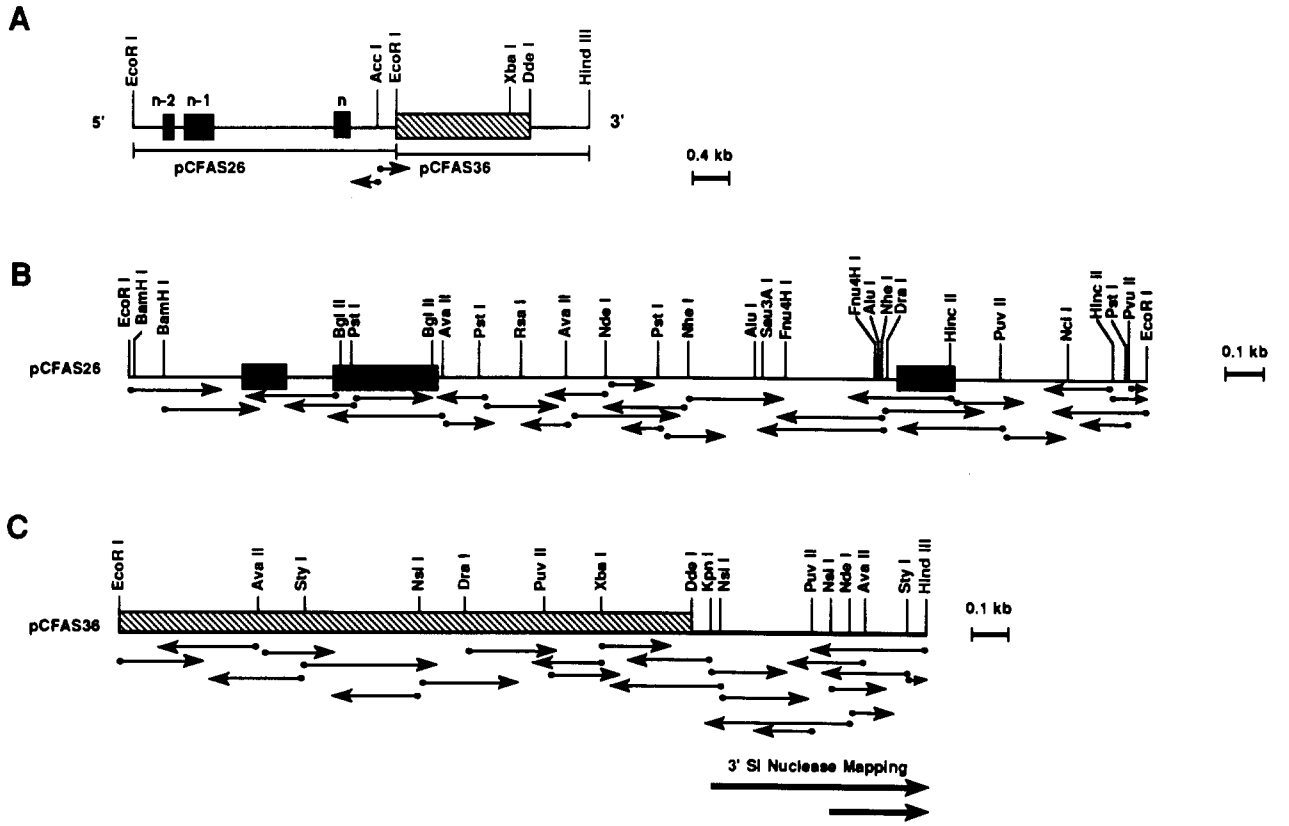

FIGURE 3: Restriction mapping and the sequencing strategy for the chicken λCFAS clone. The top line (A) shows the FAS gene segment of the chicken λCFAS clone. It is divided into two fragments and subcloned to obtain pCFAS26 and pCFAS36, lines B and C, respectively. The restriction sites shown were those used for the DNA sequence determination. Overlapping sequences were obtained by sequencing relevant fragments. The arrows indicate the direction and extent of sequencing. The heavy arrows in the pCFAS36 fragment show the fragments used in the 3′ S_1 nuclease mapping. The solid boxes represent the exons. The hatched box represents the region of homology to the goose cDNA clone, pGFAS3.

sequences that hybridized to the TE1 probe (Figure 2). These fragments (2.1 and 2.7 kbp) were isolated and subcloned into pUC18. Subclones pCFAS26 and pCFAS36 containing the 2.7-kbp *Eco*RI/*Eco*RI fragment and the 2.1-kbp *Eco*RI/*Hin*dIII fragment, respectively, were used for sequence analysis.

Sequence Analysis of λCFAS. In order to determine the region and extent of coding sequences of thioesterase that the λCFAS clone contains, subclones pCFAS26 and pCFAS36 were completely sequenced. The restriction map and sequencing strategy are shown in Figure 3. The region of the λCFAS clone connecting these two subcloned fragments was also sequenced. All sequences were confirmed as described under Materials and Methods. The thioesterase coding sequences were distributed in three exons interrupted by the introns as shown in Figures 3 and 4. The presence of the exons was deduced from the known amino acid sequence of thioesterase (Yang et al., 1988). The exon sequences in the genomic clone pCFAS26 are indicated by the protein sequences below the nucleotide sequences of Figure 5. The conserved consensus splice junction sequences are indicated by boldfaced letters as shown in Figure 5. Since the thioesterase domain constitutes the COOH-terminal region of the chicken liver fatty acid synthase (Mattick et al., 1983a,b), these exons are numbered $n-2$, $n-1$, and n, where the exon n was considered to be the last exon in the genome of the chicken fatty acid synthase. The translational reading frame of the exon n ends in an ochre codon, and the sequence beyond this until the putative poly(A+) addition sequence (see below) is considered to be the untranslated region present on the 3′ side of chicken poly(A+) mRNA.

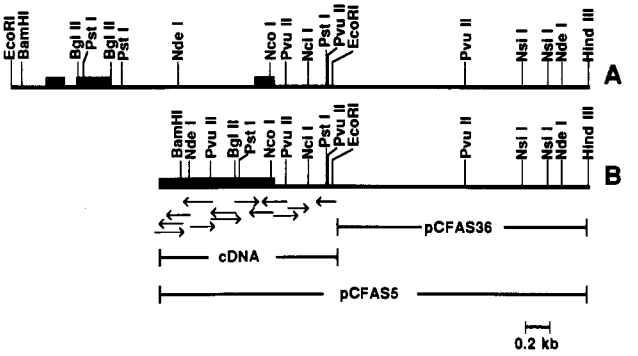

FIGURE 4: Restriction mapping and the sequence strategy of the cDNA clone. The top line (A) shows the region of genomic clone coding for fatty acid synthase mRNA, showing relevant restriction sites at the 3′ end and exons n and $n-1$ for comparison with the cDNA clone. The DNA insert of pCFAS36 (see Figure 3C) was cut at the *Eco*RI site and used as a primer for cDNA extension and cloning. Line B shows the cDNA extension in the pCFAS5 clone and the sequencing strategy used.

Isolation and Characterization of a cDNA Clone Coding for the Thioesterase Domain of Chicken Fatty Acid Synthase. We have isolated a cDNA clone using the $n-1$ exon as a probe (see Materials and Methods). Among the various cDNA clones isolated, pCFAS5 was found to contain the longest DNA insert. Restriction analyses and sequence determination of pCFAS5 revealed that the region between the *Nco*I site in the nth exon and the downstream *Eco*RI site of pCFAS36 was totally retained (Figure 4). This segment (*Nco*I–*Eco*RI) of the DNA was sequenced in order to ensure the identity of this untranslated region of pCFAS5 with that

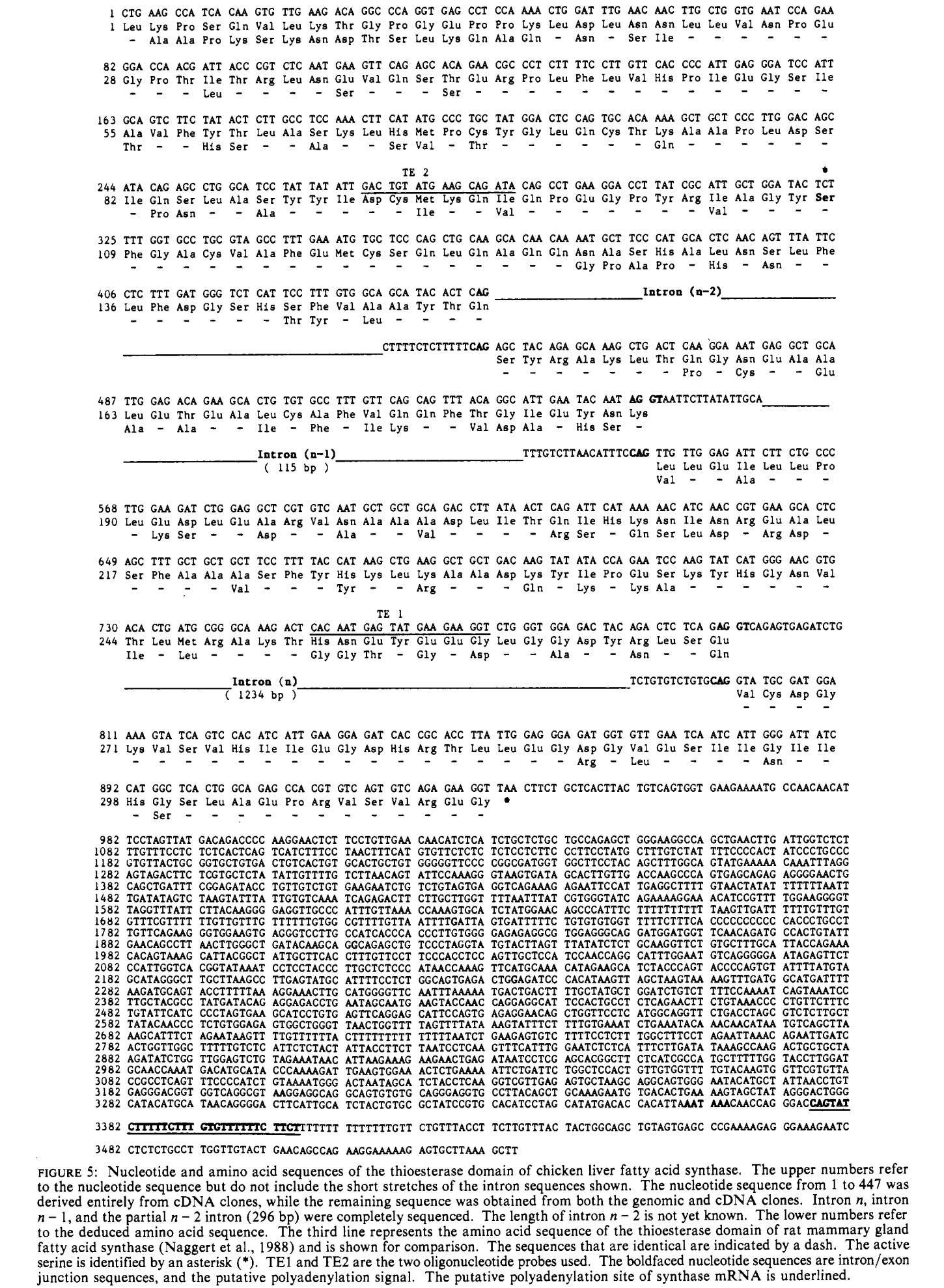

FIGURE 5: Nucleotide and amino acid sequences of the thioesterase domain of chicken liver fatty acid synthase. The upper numbers refer to the nucleotide sequence but do not include the short stretches of the intron sequences shown. The nucleotide sequence from 1 to 447 was derived entirely from cDNA clones, while the remaining sequence was obtained from both the genomic and cDNA clones. Intron n, intron $n-1$, and the partial $n-2$ intron (296 bp) were completely sequenced. The length of intron $n-2$ is not yet known. The lower numbers refer to the deduced amino acid sequence. The third line represents the amino acid sequence of the thioesterase domain of rat mammary gland fatty acid synthase (Naggert et al., 1988) and is shown for comparison. The sequences that are identical are indicated by a dash. The active serine is identified by an asterisk (*). TE1 and TE2 are the two oligonucleotide probes used. The boldfaced nucleotide sequences are intron/exon junction sequences, and the putative polyadenylation signal. The putative polyadenylation site of synthase mRNA is underlined.

found in the genomic DNA (λCFAS). Once the identity was established, the cDNA was completely sequenced. The combined sequence of the cDNA and the genomic DNA clones is presented in Figure 5 along with the deduced amino acid sequence. The nucleotide and the deduced amino acid sequences agreed completely with the sequences of the exon regions of the λCFAS genomic clone. Exons $n-2$, $n-1$, and n provided the sequences for amino acid residues 150–182, 183–286, and 267–311, respectively, while the cDNA clone provided the remaining sequences (Figure 5). The deduced amino acid sequence starting from Lys at residue number eight, as shown in Figure 5, matched perfectly with the amino acid sequence of the thioesterase domain (Yang et al., 1988).

Determination of 3′ End of Chicken FAS mRNA. In order to determine the length of the 3′ untranslated region of chicken FAS poly(A+) mRNA, S_1 nuclease mapping was performed with various fragments from both λCFAS and the subclones pCFAS26 and pCFAS36. Preliminary S_1 mapping indicated that the region between the exon *n* and the second *Eco*RI site (see Figures 2 and 3) and most of the *Eco*RI/*Hin*dIII 2.1-kbp fragment is protected by the chicken mRNA from S_1 nuclease digestion (data now shown). Thus the 3′ end of FAS mRNA appeared to be close to the *Hin*dIII site within a *Eco*RI/*Hin*dIII (2.1 kbp) fragment.

Finer mapping of the 3′ end of FAS mRNA was achieved by utilizing the 3′ end labeled 540-bp *Kpn*I/*Hin*dIII and 156-bp *Nsi*I/*Pvu*II fragments (Figure 6). Gel electrophoresis of S_1 nuclease resistant products revealed the presence of protected fragments of approximately 400 and 86–115 bases long as shown in panels A and B of Figure 6, respectively. Approximate locations of the putative 3′ end of chicken FAS poly(A+) mRNA are indicated in the sequence shown in Figure 5. This prediction was based on the size estimates of the truncation products with φX174 *Hae*III markers, as well as identification of the region by running sequencing reactions on the shorter template (*Nsi*I–*Pvu*II) as shown in Figure 6B. The mRNA and DNA hybridizations were carried out at varying temperatures ranging from 37 to 55 °C, and there was no change in the pattern.

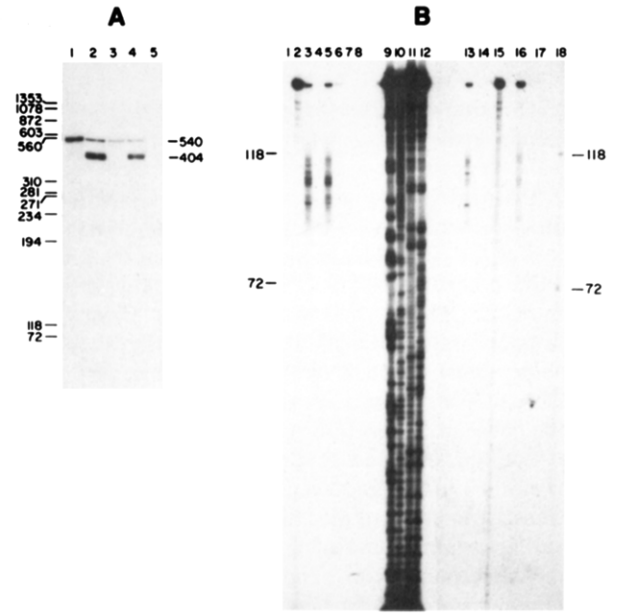

FIGURE 6: S_1 endonuclease mapping of the 3′ end of chicken FAS mRNA. In panel A, the *Kpn*I/*Hin*dIII fragment (540 bp) of pCFAS36, labeled at the *Kpn*I site, was hybridized with chicken poly(A+) RNA at 45 and 55 °C and then incubated with 1000 units of S_1 nuclease at room temperature for 30 min. The amount of poly(A+) RNA or tRNA used was 25 μg for each reaction. The protected fragments were analyzed on 4% polyacrylamide–urea gels. Lane 1, the 540-bp *Kpn*I/*Hin*dIII fragment; lanes 2 and 4, the S_1-resistant fragments after hybridizing with chicken poly(A+) RNA at 45 and 55 °C, respectively; lanes 3 and 5, the S_1-resistant fragments after hybridizing with tRNA at 45 and 55 °C, respectively. In panel B, the 156-bp *Nsi*I/*Pvu*II fragment labeled at the *Nsi*I site was used in S_1 nuclease analysis. The protected fragments were subjected to 8% polyacrylamide–urea gel electrophoresis. Lanes 1 and 18, the *Hae*III fragments of φX174 marker DNA; lanes 2 and 15, the *Nsi*I/*Pvu*II fragment, without hybridization or S_1 treatment; lanes 3, 5, and 7, the fragments hybridized with poly(A+) RNA at 37, 45, and 55 °C, respectively, and treated with S_1 nuclease at room temperature; lanes 4, 6, and 8, the fragments hybridized to tRNA at 37, 45, and 55 °C, respectively, and treated with S_1 nuclease at room temperature; lanes 9–12, G, A, C, and T sequencing reactions of the 3′ end-labeled *Nsi*I/*Pvu*II fragment; lanes 13 and 14, the fragments hybridized with poly(A+) RNA and tRNA, respectively, at 45 °C and treated with S_1 nuclease at 0 °C; lanes 16 and 17, the fragments hybridized with poly(A+) RNA and tRNA, respectively, at 45 °C and treated with S_1 nuclease at room temperature.

Discussion

We have isolated and characterized a genomic clone, λCFAS, that contains the sequences coding for a portion of the thioesterase domain and the entire 3′ untranslated region of the chicken fatty acid synthase gene. The λCFAS clone has an insert of about 13.5 kbp. Restriction mapping and Southern hybridization analyses revealed that the insert contained both coding and noncoding regions of the chicken FAS gene. This DNA fragment contains 4736 bp and is located between the *Eco*RI and *Hin*dIII restriction sites of the map shown in Figure 2. The complete nucleotide sequence of this region was determined (Figure 5). The results showed that 486 bp of this sequence code for 162 amino acid residues from the COOH terminus of the thioesterase domain of the fatty acid synthase as determined by amino acid sequence analyses (Yang et al., 1988). This sequence is distributed in three exons, $n-2$, $n-1$, and n, containing 99, 252, and 135 bp, respectively.

The synthetic oligonucleotide TE1, which was prepared according to the amino acid sequences near the COOH terminus of the thioesterase (Yang et al., 1988), hybridized to the $n-1$ exon (Figure 5). However, the oligonucleotide TE2, which was prepared according to the amino acid sequences near the active serine of the thioesterase, did not hybridize to the λCFAS alone, indicating that this region of the thioesterase molecule was upstream of the $n-2$ intron. Moreover, the size of intron $n-2$ is not known, and the genomic clone λCFAS ends somewhere within this intron. The existence of intron $n-2$ was confirmed when the cDNA was sequenced. As shown in Figure 5, the intron–exon junction sequences at the 3′ end of this intron as well as the 3′ end of the $n-2$ exon, conforming to the consensus splice junction sequences as predicted by Sharp (1981).

From the amino acid sequence studies the COOH-terminal residue was reported to be serine (Yang et al., 1988). However, the amino acid sequences of the COOH-terminal end of the thioesterase deduced from nucleotide sequences of the coding region of the λCFAS clone gave glycine as the COOH terminal or four amino acid residues longer than the reported serine. The cDNA sequence shown in Figure 5 confirms that the *n*th exon is the last exon. The additional four amino acids which are also present in the cDNA sequence at the COOH-terminal end of the thioesterase are also the COOH-terminal end of the chicken fatty acid synthase. The loss of these four amino acids in the purified thioesterase fragment prepared from chicken fatty acid synthase by chymotryptic digest (Mattick et al., 1983a,b) may be due to proteolysis of this tetrapeptide during the preparation of the synthase itself or its cleavage to yield the thioesterase domain. These results

are consistent with the earlier observations that serine and/or glycine are probable residues at the COOH terminus of fatty acid synthase (Mattick et al., 1983a,b).

The nucleotide sequences following the translation termination codon to the polyadenylation site account for 2422 bp. This DNA segment codes for the untranslated region of chicken FAS mRNA (Figure 5). Goose pGFAS3 cDNA hybridizes to this DNA segment, and when the nucleotide sequences of the two DNA segments are compared, there is about 90% homology. This homologous region is located between the *Eco*RI and *Dde*I restriction sites (Figure 3) and includes nucleotides 1444–2911 (Figure 5). This high level of homology between these two DNA sequences is not surprising because of the relatedness of the two avian species. Since the goose pGFAS3 cDNA codes for FAS mRNA and does not contain sequences that code for fatty acid synthase (Morris et al., 1982), it is obvious that this region belongs to the untranslated portion of the chicken FAS gene. Nucleotide sequences upstream and downstream to the goose FAS homology region account for 510 and 448 bp, respectively. Both of these regions are part of the untranslated DNA segment of the chicken FAS gene as evidenced by cDNA nucleotide sequence analyses and S_1 nuclease mapping.

The complete sequence of goose pGFAS3 cDNA up to the polyadenylation signal sequence is found in the chicken genomic clone with the exception of the CCTTTCTT repeat. In the chicken DNA the corresponding region (2700–2759 bp) is short and rich in "T's" (Figure 4). Such pyrimidine-rich sequences were found earlier in both avian and mammalian genomic DNA (Birnbaum et al., 1979; Fisch et al., 1985), and so far, their importance is not apparent.

In the λCFAS clone the AAGAAA sequence is found at the site corresponding to polyadenylation signal sequence AATAAA of the goose pGFAS3 cDNA clone. This AAGAAA sequence does not appear to be involved in the 3′ end cleavage/polyadenylation of chicken FAS mRNA. It is reported that AAGAAA signals are deleterious and result in the generation of reduced levels of 3′ end formation (Birnstiel et al., 1985). However, the consensus polyadenylation signal sequence AATAAA is found starting at 3360 bp, which is within 12 bases from the putative polyadenylation site(s) determined by S_1 nuclease mapping (Figure 6). Downstream of the consensus polyadenylation signal sequence AATAAA, the 3′ end formation signal sequences "CA and G/T cluster" starting at 3377 bp (Lai et al., 1979; McLauchlan et al., 1985) are also found (Figure 5).

Back et al. (1986) have found that chicken liver contains two mRNAs for fatty acid synthase, using goose cDNA as a probe. Our S_1 mapping studies have indicated that chicken FAS mRNA is a single species, a conclusion that is further supported by the presence of only one potential poly(A+) addition signal sequence. However, the reason for the differences is not apparent at this time.

Recently, Naggert et al. (1988) reported the nucleotide sequence of a cDNA clone containing the coding sequence of the thioesterase as well as the 3′ noncoding region of the mRNA of rat mammary gland fatty acid synthase. Comparison of the amino acid sequences of the thioesterases of the two species is shown in Figure 5. In the protein coding region of these cDNAs, there are 326-bp differences out of 933 bp of the nucleotide sequence. This represents a homology of 65%. The amino acid sequences of both chicken and rat thioesterase domains differ at 94 positions corresponding to a linear homology of about 70% (Figure 5). Closer examination of the sequences of the two thioesterases shows that 15 of the first 16 amino acids at the NH_2 terminus are different and may represent the peptide linking the thioesterases to their corresponding acyl carrier protein (ACP) domains (unpublished data). The amino acid sequences in this region may not be critical as long as the spatial separation of the two domains is achieved and their interactions are unhindered. However, it is important to note that out of the 35 amino acids around the active serine sites of the two thioesterases only 4 amino acids are different, suggesting the functional importance of this region and hence its conservation. Similarly, at the COOH-terminal end of the molecule there is about 90% homology in the sequence, including the last 12 amino acids which are totally conserved. These observations suggest that, apart from the sequences around the active site, the amino acid sequences at the COOH terminus may be important in the function of the molecule.

In contrast, the nucleotide sequences of the untranslated region of the mRNA of chicken and rat synthases do not show any significant homology. The sequence homology in this region falls apart immediately after the TAA stop codon. This is unlike the nucleotide sequence homology seen between the untranslated regions in chicken and goose fatty acid synthase mRNAs. Thus, only the protein coding regions have maintained high levels of homology, suggesting that the complex multifunctional fatty acid synthases are conserved through evolution.

ACKNOWLEDGMENTS

We thank Dr. Alan G. Goodridge for providing goose fatty acid synthase cDNA clone pGFAS3 and Dr. Ming-Jer Tsai for the gift of the chicken λ genomic library.

Registry No. FAS, 9045-77-6.

REFERENCES

Adams, S. L., Sobel, M. E., Howard, B. H., Olden, K., Yamada, K. M., DeCrombrugghe, B., & Paston, I. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3399–3403.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Back, D. W., Goldman, M. J., Fisch, J. E., Ochs, R. S., & Goodridge, A. G. (1986) *J. Biol. Chem. 261*, 4190–4197.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–183.

Berk, A. J., & Sharp, P. A. (1977) *Cell (Cambridge, Mass.) 12*, 721–732.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Birnboim, H. C., Sederoff, R. R., & Paterson, M. C. (1979) *Eur. J. Biochem. 98*, 301–307.

Birnstiel, M. L., Busslinger, M., & Strub, K. (1985) *Cell (Cambridge, Mass.) 41*, 349–359.

Chirala, S. S., & Wakil, S. J. (1986) *Gene 47*, 297–300.

Deeley, R. G., Gordon, J. I., Burns, A. T. H., Mullinix, K. P., Binastein, M., & Goldberger, R. F. (1977) *J. Biol. Chem. 252*, 8310–8319.

Fisch, J. E., Goldman, M. J., Wilson, R. K., Roe, B. A., & Goodridge, A. G. (1985) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 44*, 1084.

Fischer, P. W. F., & Goodridge, A. G. (1978) *Arch. Biochem. Biophys. 190*, 332–344.

Goodridge, A. G., Jenik, R. A., McDevitt, M. A., Morris, S. M., Jr., & Winberry, L. K. (1984) *Arch. Biochem. Biophys. 230*, 82–92.

Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.

Gubler, V., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Joshi, V. C., & Aranda, L. P. (1979) *J. Biol. Chem. 254*, 11783–11786.
Kasturi, R., & Joshi, V. C. (1982) *J. Biol. Chem. 257*, 12224–12230.
Kasturi, R., Joshi, V. C., & Wakil, S. J. (1984) *Arch. Biochem. Biophys. 233*, 530–529.
Lai, E. C., Stein, J. P., Catterall, J. F., Woo, S. L. C., Mace, M. L., Means, A. R., & O'Malley, B. W. (1979) *Cell (Cambridge, Mass.) 18*, 829–842.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Mattick, J. S., Zehner, Z. E., Calabro, M. A., & Wakil, S. J. (1981) *Eur. J. Biochem. 114*, 643–651.
Mattick, J. S., Nickless, J., Mizugaki, M., Yang, C.-Y., Uchiyama, S., & Wakil, S. J. (1983a) *J. Biol. Chem. 258*, 15300–15304.
Mattick, J. S., Tsukamoto, Y., Nickless, J., & Wakil, S. J. (1983b) *J. Biol. Chem. 258*, 15291–15299.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
McLauchlan, J., Gaffney, D., Whitton, J. L., & Clements, J. B. (1985) *Nucleic Acids Res. 13*, 1347–1368.
Morris, S. M., Jr., Nilson, J. H., Jenik, R. A., Winberry, L. K., McDevitt, M. A., & Goodridge, A. G. (1982) *J. Biol. Chem. 257*, 3225–3229.
Morris, S. M., Jr., Winberry, L. K., Fisch, J. E., Back, D. W., & Goodridge, A. G. (1984) *Mol. Cell. Biochem. 64*, 63–68.
Naggert, J., Witkowski, A., Mikkelsen, J., & Smith S. (1988) *J. Biol. Chem. 263*, 1146–1150.
Nepokroeff, C. M., Adachi, K., Yan, C. Y., & Porter, J. W. (1984) *Eur. J. Biochem. 140*, 441–445.
Reilly, J. G., Ogden, R., & Rossi, J. J. (1982) *Nature (London) 300*, 287–289.
Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643–646.
Singh, N., Wakil, S. J., & Stoops, J. K. (1984) *J. Biol. Chem. 259*, 3605–3611.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Student, A. K., Hsu, R. Y., & Lane, M. D. (1980) *J. Biol. Chem. 255*, 4745–4750.
Tsukamoto, Y., Wong, H., Mattick, J. S., & Wakil, S. J. (1983) *J. Biol. Chem. 258*, 15312–15322.
Wakil, S. J., Joshi, V. C., & Stoops, J. K. (1983) *Annu. Rev. Biochem. 52*, 537–579.
Weiss, G. H., Rosen, O. M., & Rubin, C. S. (1980) *J. Biol. Chem. 255*, 4751–4757.
Yang, C.-Y., Huang, W.-Y., Chirala, S. S., & Wakil, S. J. (1988) *Biochemistry* (preceding paper in this issue).
Zehner, Z. E., Joshi, V. C., & Wakil, S. J. (1977) *J. Biol. Chem. 252*, 7015–7022.
Zehner, Z. E., Mattick, J. S., Stuart, R., & Wakil, S. J. (1980) *J. Biol. Chem. 255*, 9519–9522.

Amino Acid Sequence and Posttranslational Modifications of Human Factor VII_a from Plasma and Transfected Baby Hamster Kidney Cells

Lars Thim,* Soeren Bjoern, Mogens Christensen, Else Marie Nicolaisen, Torben Lund-Hansen, Anders H. Pedersen, and Ulla Hedner

Novo Research Institute, Novo Allé, DK-2880 Bagsvaerd, Denmark

Received March 31, 1988; Revised Manuscript Received June 7, 1988

ABSTRACT: Blood coagulation factor VII is a vitamin K dependent glycoprotein which in its activated form, factor VII_a, participates in the coagulation process by activating factor X and/or factor IX in the presence of Ca^{2+} and tissue factor. Three types of potential posttranslational modifications exist in the human factor VII_a molecule, namely, 10 γ-carboxylated, N-terminally located glutamic acid residues, 1 β-hydroxylated aspartic acid residue, and 2 N-glycosylated asparagine residues. In the present study, the amino acid sequence and posttranslational modifications of recombinant factor VII_a as purified from the culture medium of a transfected baby hamster kidney cell line have been compared to human plasma factor VII_a. By use of HPLC, amino acid analysis, peptide mapping, and automated Edman degradations, the protein backbone of recombinant factor VII_a was found to be identical with human factor VII_a. Neither recombinant factor VII_a nor human plasma factor VII_a was found to contain β-hydroxyaspartic acid. In human plasma factor VII_a, the 10 N-terminally located glutamic acid residues were found to be fully γ-carboxylated whereas 9 full and 1 partial γ-carboxylated residues were found in the corresponding positions of the recombinant factor VII_a molecule. Asparagine residues 145 and 322 were found to be fully N-glycosylated in human plasma factor VII_a. In the recombinant factor VII_a, asparagine residue 322 was fully glycosylated whereas asparagine residue 145 was only partially (approximately 66%) glycosylated. Besides minor differences in the sialic acid and fucose contents, the overall carbohydrate compositions were nearly identical in recombinant factor VII_a and human plasma factor VII_a. These results show that factor VII_a as produced in the transfected baby hamster kidney cells is very similar to human plasma factor VII_a and that this cell line thus might represent an alternative source for human factor VII_a.

Factor VII is a vitamin K dependent glycoprotein which is synthesized by liver cells (Wion et al., 1985) and secreted into the blood. In its activated form (factor VII_a), the protein acts as a serine protease that participates in the extrinsic pathway of the blood coagulation leading to the formation of a fibrin clot (Davie et al., 1979). Factor VII is synthesized and secreted as a single-chain molecule (Kisiel & Davie, 1975; Radcliffe & Nemerson, 1975) consisting of 406 amino acid

0006-2960/88/0427-7785$01.50/0