*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 4387–4391, June 1989
Biochemistry

# Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA

(multienzyme complex/enzyme structure/protein sequence)

KRZYSZTOF P. HOLZER, WEI LIU, AND GORDON G. HAMMES

Department of Chemistry, Cornell University, Ithaca, NY 14853

*Contributed by Gordon G. Hammes, March 3, 1989*

**ABSTRACT** **The complete amino acid sequence of chicken liver fatty acid synthase [acyl-CoA:malonyl-CoA *C*-acyltransferase (decarboxylating, oxoacyl- and enoyl-reducing, and thioester-hydrolyzing), EC 2.3.1.85] has been determined from the corresponding cDNA sequence. A 5.3-kilobase-pair (kbp) region of cDNA coding for chicken fatty acid synthase has been cloned and sequenced that is contiguous to the 2.3-kbp region previously sequenced [Yuan, Z., Liu, W. & Hammes, G. G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6328–6331]. The cDNA codes for the remaining 1677 amino acids of the previously unsequenced region of the protein. The amino acid sequence contains peptides known to be associated with the NADPH binding site of the enoylreductase active center, the acetyl/malonyltransacylase active site, the "waiting" site containing cysteine, and a pyridoxal 5′-phosphate binding site. Locations of the NADPH binding site of the β-ketoacylreductase active site and of the dehydratase active site are proposed on the basis of protein sequence homologies to catalytic sites in other enzymes. The molecular weight of the complete polypeptide chain is 267,288. A linear functional map of the chicken fatty acid synthase derived from its primary sequence is presented.**

Chicken fatty acid synthase [acyl-CoA:malonyl-CoA *C*-acyltransferase (decarboxylating, oxoacyl- and enoyl-reducing, and thioester-hydrolyzing), EC 2.3.1.85] is a multienzyme complex responsible for the synthesis of palmitic acid from acetyl-CoA and malonyl-CoA with NADPH as the reducing agent. The kinetics and stereochemistry of many of the six separate enzymatic activities of the complex have been described (1). The complex consists of two identical polypeptides of $M_r \approx 250{,}000$ (2). Limited tryptic digests of chicken liver fatty acid synthase lead to the identification of three major functional domains (2–4). Domain I, $M_r$ 127,000, carries the acetyl- and malonyltransacylases as well as the β-ketoacylsynthase activities. Domain II, $M_r$ 107,000, contains the dehydratase, β-ketoacylreductase, and enoylreductase activities as well as the acyl carrier protein. Domain III, $M_r$ 33,000, carries the thioesterase active site. These molecular weights require a cDNA of about 7 kilobase pairs (kbp) to code for the entire polypeptide chain. The cDNA sequence of a 2.3-kbp 3′-terminal fragment of the fatty acid synthase coding region has been presented (5). The translated amino acid sequence contains peptides identified to be parts of the acyl carrier 4′-phosphopantetheine and thioesterase active sites. In this work the cDNA sequence of an additional 5.3 kbp, 5′-terminal to the previously described fragments, coding for domains I and II is presented.*

## MATERIALS AND METHODS

Most of the materials and methods used in this work have been described (5–7). *Escherichia coli* strains used were JM101 and JM109 for M13 phage manipulation (8), Y1090 for λgt11 phage library screening (9), and C600 and C600 *hfl* for work with λgt10 phage (9).

cDNA libraries were prepared by Clontech from mRNA isolated from adult Leghorn rooster livers. The first-strand synthesis reaction for each library was primed with oligodeoxyribonucleotides synthesized by using sequence information from the 5′ terminus of the clone obtained from the preceding library.

## RESULTS AND DISCUSSION

Repeated screening of the first library, which was constructed in phage λgt11 by priming the reverse transcription reaction with two oligodeoxyribonucleotides found 130 and 310 bp, respectively, from the 5′ terminus of the previously determined fatty acid synthase cDNA sequence (5), yielded 75 clones, which were analyzed by *Eco*RI restriction mapping. Selected clones were analyzed by Southern blotting (10). All clones hybridizing the synthetic oligodeoxyribonucleotide used for library screening contained an *Eco*RI insert ≈800 bp long. This fragment was subcloned and sequenced by the strategy shown in Fig. 1. The sequence of the fragment (Fig. 2) was found to overlap the 5′ end of the previously published chicken fatty acid synthase partial cDNA sequence (5).

Two oligodeoxyribonucleotides with sequences from the 5′ end of the 800-bp sequence described above were used to prime first-strand cDNA synthesis in the construction of a second library. The mRNA was denatured with 2 mM methylmercuric hydroxide prior to cDNA synthesis to remove possible secondary structures that may have been responsible for consistent termination of the first-strand synthesis reaction 800 bp away from the priming site in the construction of the first library. The cDNA was subcloned into λgt10 and amplified once. Screening with a synthetic oligodeoxyribonucleotide, which hybridizes to the noncoding strand of the fatty acid synthase cDNA in the 5′ region of the fragment cloned from the first library, yielded >20 independent clones. The sizes of the inserts in λgt10 estimated by *Eco*RI restriction mapping and electrophoresis in 0.7% agarose ranged from 0.9 to 3.1 kbp. Inserts with sizes over 2.6 kbp yielded two non-λgt10 *Eco*RI fragments, of which one was 2.6 kbp. Southern analysis of the clones showed that the 2.6-kbp fragment hybridized to the oligodeoxyribonucleotide used to screen the library. This is consistent with the presence of an *Eco*RI restriction site 2.6 kbp to the 5′ side of the library priming site in the fatty acid synthase coding region. Restriction mapping of the longest insert with *Hin*dIII and *Bgl* II restriction endonucleases yielded the map shown in Fig. 1. The fragments indicated on the map were subcloned into appropriate sites of M13mp19 in the orientations indicated. The *Bgl* II fragments were subcloned into the *Bam*HI site. Sequencing was performed according to the scheme in Fig. 1, and the resulting sequence is shown in Fig. 2.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M22987).

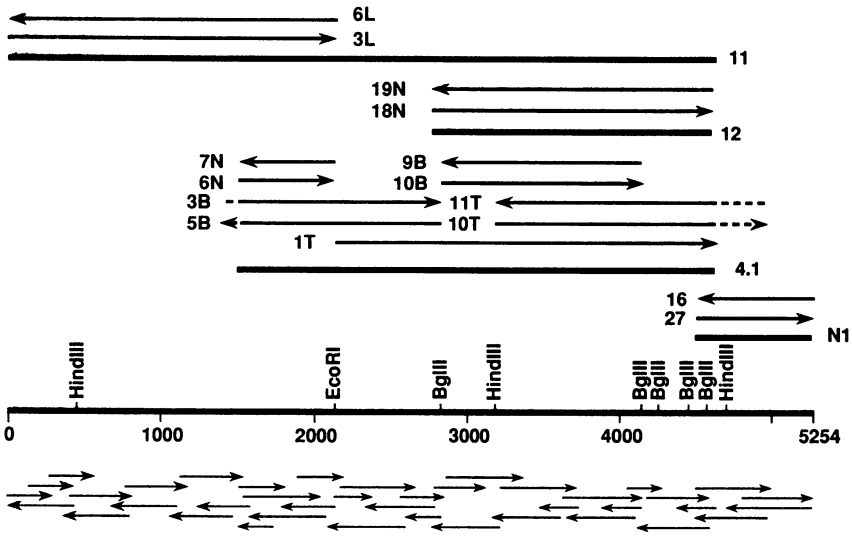


FIG. 1. Map of the chicken liver fatty acid synthase cDNA fragment. Clones used in sequencing are shown above the restriction map. Solid lines indicate clones in phage λgt vectors. Lines with arrows indicating orientation represent clones in M13mp19. Dashed lines represent λgt10 sequences. The sequencing strategy for the cDNA is shown below the restriction map.

Screening of the second library with an oligodeoxyribonucleotide found near the 5′ terminus of the sequence obtained from the experiment described above, yielded 12 independent clones. The longest insert in λgt10 was 4.9 kbp long and yielded 2.6-kbp and 2.3-kbp non-λgt10 fragments when digested with *Eco*RI. The 5′-terminal 2.3-kbp fragment was

```
        10         20         30         40         50         60         70         80         90
AGAACCTGCT CAATGGGGTT GATATGGTCA CAGAGGACGA TCGGAGGTGG AAGCCAGGGA TTTATGGACT GCCCAAAAGA AATGGAAAGC
       100        110        120        130        140        151        160                   175
TCAAGGACAT AAAAAAATTC GATGCCTCCT TCTTTGGGTC CACCCCAAAC AAGCTCATACA ATG GAT CCT CCA GTT CGC TTG TTG
                                                                    MET Asp Pro Pro Val Arg Leu Leu
                  190                 205                 220                 235                 250
TTG GAA GTT TCT TAT GAG GCT ATT TTG GAT GGA GGC ATT AAT CCA ACT GCC CTC CGT GGC ACA GAC ACG GGT GTA
Leu Glu Val Ser Tyr Glu Ala Ile Leu Asp Gly Gly Ile Asn Pro Thr Ala Leu Arg Gly Thr Asp Thr Gly Val
                  265                 280                 295                 310                 325
TGG GTT GGT GCA AGT GGC TCA GAA GCT GCT GAA GCC CTT AGC CAA GAT CCA GAA GAG CTT TTG GGA TAC AGT ATG
Trp Val Gly Ala Ser Gly Ser Glu Ala Ala Glu Ala Leu Ser Gln Asp Pro Glu Glu Leu Leu Gly Tyr Ser Met
                  340                 355                 370                 385                 400
ACT GGC TGC CAG CGT GCT ATG CTT GCC AAC AGG ATT TCT TAC TTC TAT GAT TTT ACA GGA CCA AGC TTA ACT ATC
Thr Gly Cys Gln Arg Ala Met Leu Ala Asn Arg Ile Ser Tyr Phe Tyr Asp Phe Thr Gly Pro Ser Leu Thr Ile
                  415                 430                 445                 460                 475
GAC ACA GCC TGC TCC TCC AGT CTC ATG GCT TTA GAA AAT GCT TAT AAA GCA ATT CGT CAC GGA CAG TGC AGT GCA
Asp Thr Ala Cys Ser Ser Ser Leu MET Ala Leu Glu Asn Ala Tyr Lys Ala Ile Arg His Gly Gln Cys Ser Ala
                  490                 505                 520                 535                 550
GCC CTG GTA GGA GGG GTC AAC ATT CTG CTG AAG CCC AAC ACT TCT GTG CAG TTC ATG AAG CTG GGC ATG CTT AGT
Ala Leu Val Gly Gly Val Asn Ile Leu Leu Lys Pro Asn Thr Ser Val Gln Phe Met Lys Leu Gly Met Leu Ser
                  565                 580                 595                 610                 625
CCT GAT GGT GCC TGC AAG GCT TTC GAT GTT TCA GGA AAT GGG TAT TGT CGC TCT GAA GCT GTT GTT GTT GTG CTC
Pro Asp Gly Ala Cys Lys Ala Phe Asp Val Ser Gly Asn Gly Tyr Cys Arg Ser Glu Ala Val Val Val Val Leu
                  640                 655                 670                 685                 700
TTG ACC AAG AAA TCC ATG GCT AAA CGC GTC TAT GCC ACT ATA GTC AAT GCT GGG AGT AAC ACT GAT GGC TTT AAG
Leu Thr Lys Lys Ser Met Ala Lys Arg Val Tyr Ala Thr Ile Val Asn Ala Gly Ser Asn Thr Asp Gly Phe Lys
                  715                 730                 745                 760                 775
GAG CAA GGT GTG ACA TTC CCA TCT GGA GAG ATG CAG CAG CAG CTG GTT GGT TCT CTG TAC AGA GAA TGT GGT ATC
Glu Gln Gly Val Thr Phe Pro Ser Gly Glu Met Gln Gln Gln Leu Val Gly Ser Leu Tyr Arg Glu Cys Gly Ile
                  790                 805                 820                 835                 850
AAG CCT GGA GAT GTG GAG TAT GTT GAA GCT CAT GGG ACA GGC ACC AAG GTT GGA GAT CCT CAA GAA GTA AAT GGC
Lys Pro Gly Asp Val Glu Tyr Val Glu Ala His Gly Thr Gly Thr Lys Val Gly Asp Pro Gln Glu Val Asn Gly
                  865                 880                 895                 910                 925
ATT GTA AAT GTC TTC TGC CAG TGT GAG AGA GAG CCT CTG TTA ATT GGA TCA ACC AAG TCA AAC ATG GGT CAT CCA
Ile Val Asn Val Phe Cys Gln Cys Glu Arg Glu Pro Leu Leu Ile Gly Ser Thr Lys Ser Asn Met Gly His Pro
                  940                 955                 970                 985                1000
GAG CCT GCT TCT GGG CTT GCT GCA TTA GCC AAG GTC ATT CTT TCT CTG GAA CAT GGA CTG TGG GCT CCA AAT CTT
Glu Pro Ala Ser Gly Leu Ala Ala Leu Ala Lys Val Ile Leu Ser Leu Glu His Gly Leu Trp Ala Pro Asn Leu
                 1015                1030                1045                1060                1075
CAT TTC AAT GAT CCA AAT CCA GAT ATT CCT GCT TTA CAC GAT GGC TCC TTG AAG GTG GTT TGC AAA CCA ACA CCG
His Phe Asn Asp Pro Asn Pro Asp Ile Pro Ala Leu His Asp Gly Ser Leu Lys Val Val Cys Lys Pro Thr Pro
                 1090                1105                1120                1135                1150
GTG AAA GGT GGC CTT GTC AGC ATC AAT TCT TTT GGC TTT GGA GGC TCT AAT GCT CAT GTT ATT CTG AGG CCA AAT
Val Lys Gly Gly Leu Val Ser Ile Asn Ser Phe Gly Phe Gly Gly Ser Asn Ala His Val Ile Leu Arg Pro Asn
                 1165                1180                1195                1210                1225
GAG AAG AAA TGT CAG CCT CAA GAG ACT TGT AAC TTG CCA AGA CTG GTT CAA GTT TGT GGC AGA ACA CAG GAA GCT
Glu Lys Lys Cys Gln Pro Gln Glu Thr Cys Asn Leu Pro Arg Leu Val Gln Val Cys Gly Arg Thr Gln Glu Ala
                 1240                1255                1270                1285                1300
GTG GAA ATA CTA ATT GAA GAA AGC AGG AAA CAT GGA GGA TGC AGT CCA TTT TTA AGC CTG CTC AGT GAT ATC TCT
Val Glu Ile Leu Ile Glu Glu Ser Arg Lys His Gly Gly Cys Ser Pro Phe Leu Ser Leu Leu Ser Asp Ile Ser
                 1315                1330                1345                1360                1375
GCA GTT CCT GTA TCT TCT ATG CCC TAC AGG GGC TAC ACA CTA GTT GGC ACT GAG AGT GAC ATA ACA GAG ATT CAG
Ala Val Pro Val Ser Ser Met Pro Tyr Arg Gly Tyr Thr Leu Val Gly Thr Glu Ser Asp Ile Thr Glu Ile Gln
                 1390                1405                1420                1435                1450
CAA GTT CAA GCA TCT GGT AGA CCA CTC TGG TAC ATC TGC TCA GGC ATG GGA ACA CAG TGG AAA GGT ATG GGC CTG
Gln Val Gln Ala Ser Gly Arg Pro Leu Trp Tyr Ile Cys Ser Gly Met Gly Thr Gln Trp Lys Gly Met Gly Leu
                 1465                1480                1495                1510                1525
AGC CTT ATG AAA TTG GAT CTG TTT CGC CAG TCT ATA TTG CGC TCA GAT GAG GCT TTG AAG AGC ACA GGA CTG AAG
Ser Leu Met Lys Leu Asp Leu Phe Arg Gln Ser Ile Leu Arg Ser Asp Glu Ala Leu Lys Ser Thr Gly Leu Lys
                 1540                1555                1570                1585                1600
GTC TCA GAC CTG CTT CTG AAT GCA GAT GAG AAC ACT TTT GAT GAC ACT GTC CAT GCT TTT GTT GGA CTA GCT GCT
Val Ser Asp Leu Leu Leu Asn Ala Asp Glu Asn Thr Phe Asp Asp Thr Val His Ala Phe Val Gly Leu Ala Ala
                 1615                1630                1645                1660                1675
ATA CAG ATT GCC CAA ATT GAT GTG CTA AAG GCT GCG GGT CTG CAA CCT GAT GGG ATT TTG GGC CAC TCA GTG GGA
Ile Gln Ile Ala Gln Ile Asp Val Leu Lys Ala Ala Gly Leu Gln Pro Asp Gly Ile Leu Gly His Ser Val Gly
                 1690                1705                1720                1735                1750
GAA CTA GCT TGT GGC TAT GCA GAT AAT TCC TTA AGT CAT GAA GAA GCT GTT CTT GCT GCT TAT TGG AGG GGC CGA
Glu Leu Ala Cys Gly Tyr Ala Asp Asn Ser Leu Ser His Glu Glu Ala Val Leu Ala Ala Tyr Trp Arg Gly Arg
                 1765                1780                1795                1810                1825
TGT GTG AAA GAG GCC AAA TTG CCC CCG GGA GGG ATG GCT GCT GTT GGT CTG ACA TGG GAG GAA TGT AAG CAG CGC
Cys Val Lys Glu Ala Lys Leu Pro Pro Gly Gly Met Ala Ala Val Gly Leu Thr Trp Glu Glu Cys Lys Gln Arg
```

FIG. 2. (*Figure continues on the next page.*)

```
                1840                1855                1870                1885                1900
TGT CCT CCA AAC GTG GTA CCA GCA TGT CAC AAC TCT GAG GAT ACT GTC ACT GTT TCG GGG CCT CTG GAT TCT GTG
Cys Pro Pro Asn Val Val Pro Ala Cys His Asn Ser Glu Asp Thr Val Thr Val Ser Gly Pro Leu Asp Ser Val
                1915                1930                1945                1960                1975
TCT GAG TTT GTA ACC AAA CTG AAG AAA GAT GGG GTG TTT GCA AAG GAG GTG CGC AGC GCC GGA GTT GCA TTT CAT
Ser Glu Phe Val Thr Lys Leu Lys Lys Asp Gly Val Phe Ala Lys Glu Val Arg Ser Ala Gly Val Ala Phe His
                1990                2005                2020                2035                2050
TCC TAT TAC ATG GCA TCC ATT GCA CCA GCA CTG CTC AGT GCA CTG AAA AAG GTC ATT CCA CAC CCT AAG CCT CGT
Ser Tyr Tyr Met Ala Ser Ile Ala Pro Ala Leu Leu Ser Ala Leu Lys Lys Val Ile Pro His Pro Lys Pro Arg
                2065                2080                2095                2110                2125
TCA GCA CGG TGG ATC AGT ACA TCT ATC CCT GAA TCT CAG TGG CAG AGT GAT CTT GCT AGG AAT TCC TCT GCA GAG
Ser Ala Arg Trp Ile Ser Thr Ser Ile Pro Glu Ser Gln Trp Gln Ser Asp Leu Ala Arg Asn Ser Ser Ala Glu
                2140                2155                2170                2185                2200
TAT CAT GTG AAC AAC CTA GTG AAT CCT GTG CTG TTC CAT GAA GGC CTG AAG CAT ATT CCA GAG AAT GCT GTT GTA
Tyr His Val Asn Asn Leu Val Asn Pro Val Leu Phe His Glu Gly Leu Lys His Ile Pro Glu Asn Ala Val Val
                2215                2230                2245                2260                2275
GTG GAG ATT GCT CCA CAT GCT CTC TTA CAG GCT ATC TTG AGG AGA ACT TTG AAG CCA ACT TGC ACT ATT CTA CCT
Val Glu Ile Ala Pro His Ala Leu Leu Gln Ala Ile Leu Arg Arg Thr Leu Lys Pro Thr Cys Thr Ile Leu Pro
                2290                2305                2320                2335                2350
CTG ATG AAG AAG GAC CAC AAA AAT AAC TTG GAG TTC TTC CTA ACG CAG ACT GGA AAG ATT CAT TTA ACT GGG ATA
Leu Met Lys Lys Asp His Lys Asn Asn Leu Glu Phe Phe Leu Thr Gln Thr Gly Lys Ile His Leu Thr Gly Ile
                2365                2380                2395                2410                2425
AAT GTT CTT GGA AAT AAC TTG TTC CCA CCT GTG GAA TAC CCT GTC CCT GTG GGA ACA CCT CTC ATT TCT CCA TAT
Asn Val Leu Gly Asn Asn Leu Phe Pro Pro Val Glu Tyr Pro Val Pro Val Gly Thr Pro Leu Ile Ser Pro Tyr
                2440                2455                2470                2485                2500
ATC AAA TGG GAC CAC AGC CAA GAC TGG GAT GTT CCA AAA GCT GAA GAC TTC CCC TCA GGT TCC AAA GGC TCT GCG
Ile Lys Trp Asp His Ser Gln Asp Trp Asp Val Pro Lys Ala Glu Asp Phe Pro Ser Gly Ser Lys Gly Ser Ala
                2515                2530                2545                2560                2575
TCT GCT TCA GTC TAC AAC ATC GAT GTG AGT CCT GAC TCT CCT GAC CAT TAC TTG GTT GGC CAT TGC ATT GAT GGC
Ser Ala Ser Val Tyr Asn Ile Asp Val Ser Pro Asp Ser Pro Asp His Tyr Leu Val Gly His Cys Ile Asp Gly
                2590                2605                2620                2635                2650
AGA GTC CTG TAC CCA GCA ACT GGG TAC TTA GTG CTG GCG TGG CGA ACT CTG GCA CGA TCT CTT GGC ATG GTC ATG
Arg Val Leu Tyr Pro Ala Thr Gly Tyr Leu Val Leu Ala Trp Arg Thr Leu Ala Arg Ser Leu Gly Met Val Met
                2665                2680                2695                2710                2725
GAA CAA ACA GCT GTT ATG TTT GAA GAA GTT ACA ATC CAT CAG GCA ACT ATC CTT CCC AAA AAG GGA TCA ACA CAG
Glu Gln Thr Ala Val Met Phe Glu Glu Val Thr Ile His Gln Ala Thr Ile Leu Pro Lys Lys Gly Ser Thr Gln
                2740                2755                2770                2785                2800
CTG GAA GTA CGA ATC ATG CCT GCT TCT CAC AGC TTT GAA GTG TCA GGG AAT GGG AAT TTG GCT GTG AGT GGG AAG
Leu Glu Val Arg Ile Met Pro Ala Ser His Ser Phe Glu Val Ser Gly Asn Gly Asn Leu Ala Val Ser Gly Lys
                2815                2830                2845                2860                2875
ATC TCC CTC CTA GAA AAC GAT GCT CTG AAG AAC TTT CAT AAC CAG CTG GCT GAC TTT CAG AGT CAA GCA AAC GTG
Ile Ser Leu Leu Glu Asn Asp Ala Leu Lys Asn Phe His Asn Gln Leu Ala Asp Phe Gln Ser Gln Ala Asn Val
                2890                2905                2920                2935                2950
ACT GCG AAG TCT GGC CTC TTG ATG GAA GAT GTT TAC CAA GAG CTG CAT CTT CGT GGA TAT AAC TAT GGA CCA ACT
Thr Ala Lys Ser Gly Leu Leu Met Glu Asp Val Tyr Gln Glu Leu His Leu Arg Gly Tyr Asn Tyr Gly Pro Thr
                2965                2980                2995                3010                3025
TTT CAG GGT GTT CTG GAA TGC AAC AGT GAA GGA AGT GCA GGG AAA ATT CTG TGG AAT GGA AAC TGG GTA ACC TTC
Phe Gln Gly Val Leu Glu Cys Asn Ser Glu Gly Ser Ala Gly Lys Ile Leu Trp Asn Gly Asn Trp Val Thr Phe
                3040                3055                3070                3085                3100
CTT GAC ACC CTG CTA CAC TTG ATA GTC TTA GCA GAG ACT GGG CGC AGT CTA CGA TTG CCC ACC AGG ATT CGC TCA
Leu Asp Thr Leu Leu His Leu Ile Val Leu Ala Glu Thr Gly Arg Ser Leu Arg Leu Pro Thr Arg Ile Arg Ser
                3115                3130                3145                3160                3175
GTG TAT ATT GAC CCT GTG CTT CAT CAG GAG CAG GTG TAC CAG TAC CAG GAC AAT GTA GAA GCT TTT GAT GTT GTT
Val Tyr Ile Asp Pro Val Leu His Gln Glu Gln Val Tyr Gln Tyr Gln Asp Asn Val Glu Ala Phe Asp Val Val
                3190                3205                3220                3235                3250
GTT GAC CGC TGT CTT GAT AGC CTC AAA GCA GGA GGT GTT CAG ATC AAT GGA CTT CAT GCC TCG GTG GCA CCA CGG
Val Asp Arg Cys Leu Asp Ser Leu Lys Ala Gly Gly Val Gln Ile Asn Gly Leu His Ala Ser Val Ala Pro Arg
                3265                3280                3295                3310                3325
CGA CAA CAG GAG CGG ATC TCT CCC ACT CTG GAA AAA TTC TCC TTT GTT CCC TAT ATT GAG AGT GAC TGT TTG TCT
Arg Gln Gln Glu Arg Ile Ser Pro Thr Leu Glu Lys Phe Ser Phe Val Pro Tyr Ile Glu Ser Asp Cys Leu Ser
                3340                3355                3370                3385                3400
TCC AGT ACC CAG CTT CAT GCC TAC CTG GAG CAC TGC AAA GGC CTG ATC CAG AAA TTA CAA GCT AAG ATG GCA TTG
Ser Ser Thr Gln Leu His Ala Tyr Leu Glu His Cys Lys Gly Leu Ile Gln Lys Leu Gln Ala Lys Met Ala Leu
                3415                3430                3445                3460                3475
CAC GGA GTC AAA CTA GTT ATC CAT GGC CTA GAA ACC AAC GGG GCT GCT GCA GGA TCC CCA CCC ACA CAG AAG GGC
His Gly Val Lys Leu Val Ile His Gly Leu Glu Thr Asn Gly Ala Ala Ala Gly Ser Pro Pro Thr Gln Lys Gly
                3490                3505                3520                3535                3550
CTT CAG CAT ATC CTT ACT GAA ATC TGC CAT CTG GAA CTG AAT GGA AAC CTA CAT TCT GAG CTG GAA CAG ATT GTG
Leu Gln His Ile Leu Thr Glu Ile Cys His Leu Glu Leu Asn Gly Asn Leu His Ser Glu Leu Glu Gln Ile Val
                3565                3580                3595                3610                3625
ACT CAG GAG AAG ATG CAC CTC CAG GAC GAT CCC CTT CTC AAT GGC TTG CTG GAT TCT TCA GAG TTG AAG ACT TGC
Thr Gln Glu Lys Met His Leu Gln Asp Asp Pro Leu Leu Asn Gly Leu Leu Asp Ser Ser Glu Leu Lys Thr Cys
                3640                3655                3670                3685                3700
CTG GAT GTG GCA AAG GAG AAC ACG ACC AGT CAC AGG ATG AAG ATA GTG GAG GCT CTG GCA GGA AGT GGA CGT CTG
Leu Asp Val Ala Lys Glu Asn Thr Thr Ser His Arg Met Lys Ile Val Glu Ala Leu Ala Gly Ser Gly Arg Leu
                3715                3730                3745                3760                3775
TTC TCT CGT GTC CAA AGT ATT CTG AAT ACT CAG CCC CTG TTG CAG CTG GAC TAC ATT GCC ACT GAC TGC ACC CCT
Phe Ser Arg Val Gln Ser Ile Leu Asn Thr Gln Pro Leu Leu Gln Leu Asp Tyr Ile Ala Thr Asp Cys Thr Pro
                3790                3805                3820                3835                3850
GAA ACT CTT TCA AAT GAT GAA ACA GAG CTG CAC GAT GCT GGA ATC TCC TTT AGC CAG TGG GAT CCC TCT AGC CTT
Glu Thr Leu Ser Asn Asp Glu Thr Glu Leu His Asp Ala Gly Ile Ser Phe Ser Gln Trp Asp Pro Ser Ser Leu
                3865                3880                3895                3910                3925
CCC TCT GGA AAT CTG ACC AAT GCT GAC CTG GCA GTA TGC AAC TGT TCA ACA AGT GTT CTG GGG AAC ACA GCT GAA
Pro Ser Gly Asn Leu Thr Asn Ala Asp Leu Ala Val Cys Asn Cys Ser Thr Ser Val Leu Gly Asn Thr Ala Glu
                3940                3955                3970                3985                4000
ATT ATC TCT AAC TTA GCA GCT GCA GTG AAA GAA GGA GGG TTT GTT TTG CTG CAC ACC CTT CTT AAA GAG GAA ACT
Ile Ile Ser Asn Leu Ala Ala Ala Val Lys Glu Gly Gly Phe Val Leu Leu His Thr Leu Leu Lys Glu Glu Thr
                4015                4030                4045                4060                4075
CTT GGA GAA ATT GTC AGC TTT CTT ACA AGT CCA GAC CTA CAG CAA GAG CAC AGC TTC CTG TCT CAG GCA CAG TGG
Leu Gly Glu Ile Val Ser Phe Leu Thr Ser Pro Asp Leu Gln Gln Glu His Ser Phe Leu Ser Gln Ala Gln Trp
                4090                4105                4120                4135                4150
GAG GAG TTA TTC AGC AAG GCC TCA TTG AAT CTG GTT GCA ATG AAG AGA TCT TTC TTT GGC TCA GTT ATT TTC CTG
Glu Glu Leu Phe Ser Lys Ala Ser Leu Asn Leu Val Ala Met Lys Arg Ser Phe Phe Gly Ser Val Ile Phe Leu
                4165                4180                4195                4210                4225
TGT CGA CGG CAG TCC CCT GCC AAA GCA CCC ATT CTT CTG CCA GTA GAT GAC ACT CAT TAT AAG TGG GTT GAC TCC
Cys Arg Arg Gln Ser Pro Ala Lys Ala Pro Ile Leu Leu Pro Val Asp Asp Thr His Tyr Lys Trp Val Asp Ser
                4240                4255                4270                4285                4300
TTA AAG GAG ATC TTG GCT GAC TCA TCA GAG CAG CCT CTG TGG TTG ACT GCC ACC AAT TGT GGG AAC TCT GGA ATT
Leu Lys Glu Ile Leu Ala Asp Ser Ser Glu Gln Pro Leu Trp Leu Thr Ala Thr Asn Cys Gly Asn Ser Gly Ile
                4315                4330                4345                4360                4375
TTG GGT ATG GTG AAC TGC CTC CGC CTG GAA GCA GAG GGC CAC AGA ATC AGG TGT GTG TTT GTT TCC AAC CTG AGC
Leu Gly Met Val Asn Cys Leu Arg Leu Glu Ala Glu Gly His Arg Ile Arg Cys Val Phe Val Ser Asn Leu Ser
                4390                4405                4420                4435                4450
CCT TCA TCA ACT GTC CCA GCC ACT AGT CTT TCT TCC CTG GAG ATG CAG AAG ATT ATT GAG AGA GAT CTG GTG ATG
Pro Ser Ser Thr Val Pro Ala Thr Ser Leu Ser Ser Leu Glu Met Gln Lys Ile Ile Glu Arg Asp Leu Val Met
                4465                4480                4495                4510                4525
AAT GTG TAT CGT GAT GGA AAG TGG GGT TCC TTC AGG CAT CTC CCA TTG CAG CAA GCT CAG CCT CAG GAG CTG ACA
Asn Val Tyr Arg Asp Gly Lys Trp Gly Ser Phe Arg His Leu Pro Leu Gln Gln Ala Gln Pro Gln Glu Leu Thr
                4540                4555                4570                4585                4600
GAA TAT GCC TAC GTA AAT GTG TTG ACT CGT GGA GAT CTC TCT TCC CTT CGT TGG ATT GTT TCC CCA CTT CGA CAC
Glu Tyr Ala Tyr Val Asn Val Leu Thr Arg Gly Asp Leu Ser Ser Leu Arg Trp Ile Val Ser Pro Leu Arg His
```

FIG. 2. (*Figure continues on the next page.*)

```
                4615                4630                4645                4660                4675
TTC CAA ACA ACC AAT CCA AAT GTT CAG CTC TGC AAA GTC TAC TAT GCA TCT CTC AAT TTC CGG GAC ATT ATG CTG
Phe Gln Thr Thr Asn Pro Asn Val Gln Leu Cys Lys Val Tyr Tyr Ala Ser Leu Asn Phe Arg Asp Ile Met Leu
                4690                4705                4720                4735                4750
GCA ACA GGA AAG CTT TCT CCA GAT GCT ATC CCT GGT AAC TGG ACG TTG CAG CAG TGC ATG CTG GGC ATG GAG TTC
Ala Thr Gly Lys Leu Ser Pro Asp Ala Ile Pro Gly Asn Trp Thr Leu Gln Gln Cys Met Leu Gly Met Glu Phe
                4765                4780                4795                4810                4825
TCA GGA CGG GAC CTG GCT GGA AGG AGA GTG ATG GGA TTG CTG CCA GCA AAA GGG CTG GCG ACA GTG GTG GAC TGT
Ser Gly Arg Asp Leu Ala Gly Arg Arg Val Met Gly Leu Leu Pro Ala Lys Gly Leu Ala Thr Val Val Asp Cys
                4840                4855                4870                4885                4900
GAC AAG AGG TTT CTA TGG GAA GTG CCT GAA AAC TGG ACT CTG GAA GAA GCA GCT TCG GTG CCT GTG GTT TAT GCC
Asp Lys Arg Phe Leu Trp Glu Val Pro Glu Asn Trp Thr Leu Glu Glu Ala Ala Ser Val Pro Val Val Tyr Ala
                4915                4930                4945                4960                4975
ACT GCT TAT TAT GCT TTG GTG GTT CGA GGT GGT ATG AAG AAG GGG GAG AGT GTC CTC ATT CAC TCT GGC TCA GGA
Thr Ala Tyr Tyr Ala Leu Val Val Arg Gly Gly Met Lys Lys Gly Glu Ser Val Leu Ile His Ser Gly Ser Gly
                4990                5005                5020                5035                5050
GGT GTG GGC CAA GCA GCC ATT GCC ATC GCC TTG AGC ATG GGC TGC CGT GTT TTT GCT ACT GTA GGC TCT GCT GAG
Gly Val Gly Gln Ala Ala Ile Ala Ile Ala Leu Ser Met Gly Cys Arg Val Phe Ala Thr Val Gly Ser Ala Glu
                5065                5080                5095                5110                5125
AAA CGT GAG TAT CTC CAA GCA AGG TTC CCA CAG CTG GAT GCT AAT AGC TTT GCC AGC TCC CGA AAT ACA ACC TTT
Lys Arg Glu Tyr Leu Gln Ala Arg Phe Pro Gln Leu Asp Ala Asn Ser Phe Ala Ser Ser Arg Asn Thr Thr Phe
                5140                5155                5170                5185                5200
GAG CAA CAC ATA CTG CGA GTT ACC AAT GGG AAA GGT GTC AAC CTT GTG TTA AAT TCC TTG GCA GAA GAG AAG CTC
Glu Gln His Ile Leu Arg Val Thr Asn Gly Lys Gly Val Asn Leu Val Leu Asn Ser Leu Ala Glu Glu Lys Leu
                5215                5230                5245        5254
CAA GCC AGT TTG CGT TGT CTT GCT CAA CAT GGG CGC TTC TTG GAA ATA GGC AAA
Gln Ala Ser Leu Arg Cys Leu Ala Gln His Gly Arg Phe Leu Glu Ile Gly Lys
```

FIG. 2. Sequence of the cDNA fragment coding for domains I and II of the chicken liver fatty acid synthase. The underlined DNA region overlaps the previously published sequence (5). Underlined amino acid sequences indicate protein sequences that have been obtained independently and are discussed in the text.

subcloned into M13mp19, and the unknown part was sequenced according to the scheme presented in Fig. 1. The resulting sequence is shown in Fig. 2.

The amino acid sequence encoded by nucleotides 5183–5254 of the sequence reported here matches exactly the 5′-terminal sequence of the fatty acid synthase cDNA fragment reported previously (5). The open reading frame of both sequences is continuous (Fig. 2). The cDNA fragment sequenced by Yuan *et al.* (5) codes for domain III (thioesterase) and the acyl carrier protein region of domain II of fatty acid synthase. The sequence reported here codes for domain I and the remaining part of domain II.

The amino acid sequence coded by nucleotides 5024–5056 corresponds exactly to the sequence of the enoylreductase active site tryptic fragment isolated by Chang and Hammes (11). The amino acid sequence coded by nucleotides 1928–1954 matches exactly the sequence of the other pyridoxal 5′-phosphate-labeled tryptic fragment of fatty acid synthase (11). The latter fragment is located in domain I of the enzyme in agreement with peptide mapping (4).

The amino acid sequence coded by nucleotides 1652–1678 corresponds exactly to the essential serine site of acetyl/malonyltransacylase of chicken fatty acid synthase (12). This serine "loading" site is located in domain I of fatty acid synthase. The cysteine-containing "waiting" site peptide identified by iodoacetamide labeling (12) and located in domain I (13) is encoded by nucleotides 359–427.

Nucleotides 3185–3202 and 245–265 code for the amino acid sequences Cys-Leu-Asp-Ser-Leu-Lys and Val-Trp-Val-Gly-Ala-Ser-Gly, which match those of the fluorescent fragments isolated from the trypsin digest of chicken fatty acid synthase labeled with the nucleotide analog 2-[(4-bromo-2,3-dioxobutyl)thio]-1,$N^6$-ethenoadenosine 2,5-diphosphate (S.-I. Chang and G.G.H., unpublished results). The match of the known sequences with the protein sequence deduced from the cDNA indicates that the reading frame of the sequence is correct and that no omissions in nucleotides are present.

The chicken liver fatty acid synthase polypeptide contains two NADPH binding sites. The site at the enoylreductase active center was found by comparison to a known peptide sequence, as described above. There is no available protein sequence information for the NADPH binding site of the β-ketoacylreductase active center. A homology search between the amino acid sequence surrounding the enoylreductase NADPH binding site and the rest of the translated amino acid sequence of the enzyme was performed. The comparison located a region of extensive homology (Table 1) 70 residues to the N terminus side of the enoylreductase site. The region is encoded by nucleotides 4781–4834 (Fig. 2). This sequence probably represents the β-ketoacylreductase NADPH binding site. Both reductase active centers are located within a polypeptide of $M_r$ 10,000, which is consistent with trypsin-digest data (4). However, the studies of trypsin digestion suggested that enoylreductase is located to the N-terminal side of the β-ketoacylreductase, whereas our results suggest the reverse order (Fig. 3).

Table 1. Homology between the amino acid sequences of the regions containing the enoylreductase and the proposed β-ketoacylreductase sites*

| Active site | Amino acid sequence |
|---|---|
| Enoylreductase | M G – – – C R V F A T V G S A E K R |
| | X X x x x x X X X x x x X X |
| β-Ketoacylreductase | M G L L P A K G L A T V V D C D K R |

*The sequences are aligned to give the best homology. No amino acids are omitted. The symbol – is used to denote a gap in a sequence introduced to improve alignment. Homologous residues are marked with X, and conservative substitutions are marked with x.

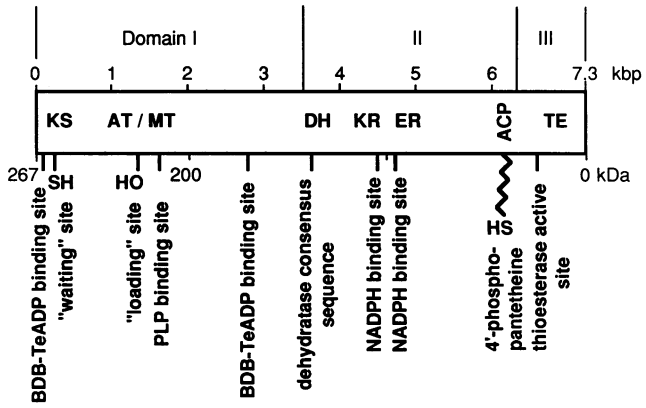


FIG. 3. Linear functional map of chicken fatty acid synthase constructed on the basis of primary sequence comparison with known peptide sequences. TE, thioesterase; ACP, acyl carrier protein; ER, enoylreductase; KR, β-ketoacylreductase; DH, dehydratase; AT/MT, acetyl/malonyltransacylase; and KS, ketoacylsynthase. The locations of the SH "waiting" site, the OH "loading" site, a pyridoxal 5′-phosphate (PLP) binding site, the 2-[(4-bromo-2,3-dioxobutyl)thio]-1,$N^6$-ethenoadenosine 2,5-diphosphate (BDB-TeADP) binding sites, the dehydratase active site, the NADPH binding sites of enoyl reductase and β-ketoacylreductase, the 4′-phosphopantetheine group, and the thioesterase active site are indicated.

Table 2. Comparison of the primary structure of the proposed dehydratase site of chicken liver fatty acid synthase with homologous regions from other dehydratases

| Dehydratase | Amino acid sequence | Ref. |
|---|---|---|
| Chicken liver fatty acid synthase | 1247-C S T S V L G N T A E I I S N L A A A | This paper |
| Yeast fatty acid synthase $\beta$ subunit | 1612-F V D M V L P N T A L K T S I Q H V G | 14, 15 |
| Yeast threonine dehydratase *ILV1* | 132-C – – S – A G N H A N G V A F – A A K | 16 |
| Rat liver L-serine dehydratase *SDH2* | 62-C S – S – A G N – A G M A T A Y A A R | 17 |
| | 96-P S P L T A G N – A G M A T A Y A A R | 18 |
| *E. coli* biodegradative threonine dehydratase *tdc* | 81-C – – S – A G N H A Q G V S L S C A M | 19 |
| *E. coli* threonine dehydratase *ilvA* | 83-I T A S – A G N H A Q G V A F S S A R | * |
| *E. coli* D-serine dehydratase | 171-A V G S T – G N – – L G L S I G I M – | 20 |

The sequences are aligned to give the best homology. The symbol – denotes a gap in a sequence introduced to improve alignment. Residues homologous to those in the chicken liver sequence are underlined.
*EMBL/GenBank Genetic Sequence Database (1986) GenBank (Bolt, Beranek, and Newman Labs., Cambridge, MA), Release 58.0, accession no. KO3503.

The chicken liver fatty acid synthase dehydratase site was located by comparison with known dehydratase sequences (Table 2). Nucleotides 3893–3949 code for an amino acid sequence with significant homologies to five other dehydratases. The differences between the metabolic functions and origins of the different enzymes account for the sequence differences. The presence of the common motif Ser-$Xaa_{0-3}$-(Ala or Leu)-Gly-Asn-(Thr or His)$_{0-1}$-Ala-$Xaa_{6-7}$-$Ala_{1-2}$ points to the importance of these conserved residues in the dehydratase activity of these enzymes. The dehydratase activity is located in domain II of the chicken liver fatty acid synthase in agreement with trypsin digest data (21).

Analysis of the cDNA sequence revealed only one in-frame ATG start codon (nucleotides 149–151) between the sequence (nucleotides 245–265) encoding one of the 2-[(4-bromo-2,3-dioxobutyl)thio]-1,$N^6$-ethenoadenosine 2,5-diphosphate binding sites and an upstream inframe nonsense codon. Thus, the methionine encoded by this codon must represent the N terminus of the protein, unless it is removed by posttranslational modification. The calculated molecular weight of the polypeptide encoded by the cDNA is 267,288, in agreement with the approximate molecular weight of chicken liver fatty acid synthase determined from the electrophoretic mobilities of the enzyme and its fragments (2, 21).

The cDNA sequence of chicken liver fatty acid synthase, combined with the amino acid sequences of labeled tryptic fragments and computer analysis of the translated sequence, permits construction of the linear functional map of this multienzyme complex that is presented in Fig. 3.

The sequencing of the full cDNA coding for the rat fatty acid synthase has been reported recently in an abstract (22). Analysis of homologies between the chicken and rat fatty acid synthase sequences will yield information about conserved, functionally important domains and about the evolutionary relationship of the two enzymes.

We thank Drs. B.-K. Tye and J. Calvo for the use of their sequencing facilities. This work was supported by a grant from the National Institutes of Health (GM 13292).

1. Hammes, G. G. (1985) *Curr. Top. Cell. Regul.* **26,** 311–324.
2. Wakil, S. J., Stoops, J. K. & Joshi, V. C. (1983) *Annu. Rev. Biochem.* **52,** 536–579.
3. Mattick, J. S., Tsukamoto, Y., Nickless, J. & Wakil, S. J. (1983) *J. Biol. Chem.* **258,** 15291–15299.
4. Tsukamoto, Y., Wong, H., Mattick, J. S. & Wakil, S. J. (1983) *J. Biol. Chem.* **258,** 15312–15322.
5. Yuan, Z., Liu, W. & Hammes, G. G. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6328–6331.
6. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
7. Fristensky, B., Lis, J. & Wu, R. (1982) *Nucleic Acids Res.* **10,** 6451–6463.
8. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
9. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 56, 110.
10. Southern, E. (1985) *J. Mol. Biol.* **98,** 503–517.
11. Chang, S.-I. & Hammes, G. G. (1989) *Biochemistry* **28,** in press.
12. Chang, S.-I. & Hammes, G. G. (1988) *Biochemistry* **27,** 4753–4750.
13. Yuan, Z. & Hammes, G. G. (1985) *J. Biol. Chem.* **260,** 13532–13538.
14. Schweizer, M., Roberts, L. M., Höltke, H.-J., Takabayashi, K., Höllerer, E., Hoffmann, B., Müller, G., Köttig, H. & Schweizer, E. (1986) *Mol. Gen. Genet.* **203,** 479–486.
15. Chirala, S. S., Kuziora, M. A., Spector, D. M. & Wakil, S. J. (1987) *J. Biol. Chem.* **262,** 4231–4240.
16. Kielland-Brandt, M. C., Holmberg, S. & Litske Petersen, L. G. (1984) *Carlsberg Res. Commun.* **49,** 567–575.
17. Noda, C., Ito, K., Nakamura, T. & Ichihara, A. (1988) *FEBS Lett.* **234,** 331–335.
18. Ogawa, H., Miller, D. A., Dunn, T., Su, Y., Burcham, J. M., Peraino, C., Fujioka, M., Babcock, K. & Pitot, H. C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5809–5813.
19. Datta, P., Gross, T. J., Omnaas, J. R. & Patil, R. V. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 393–397.
20. Schlitz, E. & Schmitt, W. (1981) *FEBS Lett.* **134,** 57–62.
21. Tsukamoto, Y. & Wakil, S. J. (1988) *J. Biol. Chem.* **263,** 16225–16229.
22. Amy, C., Naggert, J., Witkowski, A., Williams, B., Randhawa, Z. & Smith, S. (1988) *J. Cell Biol.* **107,** 412a, abstr. 2346.