Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3114–3118, May 1989
Biochemistry

Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase

(Northern blot hybridization/CNBr peptides/domain mapping/active sites)

CHRISTOPHER M. AMY, ANDRZEJ WITKOWSKI, JURGEN NAGGERT, BRENDA WILLIAMS, ZAFAR RANDHAWA*, AND STUART SMITH†

Children's Hospital Oakland Research Institute, 747 Fifty Second Street, Oakland, CA 94609

Communicated by P. K. Stumpf, January 30, 1989 (received for review December 8, 1988)

ABSTRACT **Overlapping cloned cDNAs representing the entire sequence of the rat fatty acid synthase mRNA have been isolated from a cDNA library and sequenced. Authenticity of the cDNA clones was supported by hybridization to fatty acid synthase mRNA and by amino-terminal sequencing of 39 fatty acid synthase CNBr fragments. The full-length fatty acid synthase mRNA is 9156 nucleotides long and includes an 84-nucleotide 5′ noncoding region, a 7515-nucleotide coding sequence, and a 1537-nucleotide 3′ noncoding region; a second mRNA species containing a shortened 3′ noncoding sequence is also transcribed in the rat. The encoded fatty acid synthase subunit contains 2505 amino acids and has a molecular weight of 272,340. Active sites and substrate binding sites were located within the sequence, thus establishing the order of domains on the multifunctional animal fatty acid synthase as condensing enzyme–transferase–dehydrase–enoyl reductase–ketoreductase–acyl carrier protein–thioesterase.**

The synthesis of fatty acids from malonyl-CoA *de novo* requires several enzymatic activities (1). In most bacteria and plants the activities exist as discrete monofunctional polypeptides, whereas in animals they are integrated into a single multifunctional polypeptide (2). Partial sequences for animal fatty acid synthases have been reported (3–6). In this paper we report the complete amino acid sequence‡ of an animal fatty acid synthase and the ordering of the seven functional domains on the multifunctional subunit.

MATERIALS AND METHODS

Isolation and Sequencing of Fatty Acid Synthase cDNA Clones. Clones characterized in this study were isolated from λgt11 cDNA libraries constructed from poly(A) RNA obtained from the mammary glands of lactating Long–Evans rats (7) or the livers of fasted-refed Long–Evans rats (8). To maximize the probability of including cDNA sequences corresponding to the 5′ end of the fatty acid synthase in the liver library, we added a specific primer [nucleotides (nt) 3183–3200 of fatty acid synthase, antisense direction, 50 ng/ml] to the reaction mixture for first-strand synthesis. Asymmetrical adaptors with dephosphorylated *Eco*RI overhangs (Pharmacia) were used in the ligation reaction, eliminating the need for *Eco*RI methylase treatment. The liver library yielded, in *Escherichia coli* Y1090r⁻, 1.2×10^7 plaque-forming units/μg of DNA and was amplified 7.6×10^4-fold (68% white plaques) before screening.

Probes for library screening were derived from established fatty acid synthase cDNAs and were labeled with [^{32}P]dCTP by random-priming (9). Inserts were subcloned (10) using pUC12 or pUC19 vectors and *E. coli* DH5α cells (Bethesda Research Laboratories). Nested deletions were constructed using BAL-31 nuclease (11). Double-stranded plasmid DNA was used directly in dideoxynucleotide sequencing reactions with purified synthetic oligonucleotide primers (3, 12).

Amino Acid Sequencing. Fatty acid synthase was purified from the livers of Long–Evans rats. The thioesterase domains were removed with trypsin, and the core polypeptides, produced by nicking near the center of the subunit, were isolated (13). Protein was carboxymethylated and digested with CNBr. The resulting peptides were fractionated by HPLC (14) and sequenced on an Applied Biosystems model 477A sequencer.

Isolation of Amino-Terminal and Carboxyl-Terminal Fragments of the Trypsin-Digested Fatty Acid Synthase Core. The two core polypeptides were separated by gel filtration in the presence of $NaDodSO_4$. The amino-terminal fragment was immobilized to CNBr-activated Sepharose and used as an affinity matrix to purify regionally specific antibodies from anti-fatty acid synthase antibodies. The carboxyl-terminal fragment was subjected directly to Edman degradation.

RESULTS AND DISCUSSION

Derivation of Fatty Acid Synthase Sequence. Our strategy for sequencing the fatty acid synthase involved the initial identification and characterization of several cDNA clones that produced immunoreactive fusion proteins. Then making use of the fact that the library contained many overlapping cDNA fragments, possibly produced by self-priming of the fatty acid synthase mRNA in the reverse transcriptase reaction, we conducted a "cDNA walk" in a 3′ → 5′ direction. As upstream sequences were established, they in turn were utilized to develop probes for the identification of clones in the library. Authenticity of the cDNA clones was confirmed by their ability to hybridize specifically to fatty acid synthase 8.3- and 9.1-kilobase (kb) mRNA species in Northern analysis (Fig. 1). The clones used to construct the complete cDNA sequence for rat mammary gland fatty acid synthase are shown in Fig. 2. In the initial immunochemical screening of the library, three unique cDNA clones were identified. Clones λFAS1 and λFAS5 produced fusion proteins recognized by antibodies specific for the thioesterase domain and together contained the complete sequence of the thioesterase and the acyl carrier protein domains as well as a 1537-base-pair noncoding region terminating in a poly(A) tail (3, 4). One additional clone, λFAS3, that produced a fusion protein recognized by antibodies specific for the amino-terminal half of the fatty acid synthase was isolated.

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.

Abbreviation: nt, nucleotide(s).

*Present address: Otsuka Pharmaceutical Company, 9900 Medical Center Drive, Rockville, MD 20850.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. X14175).

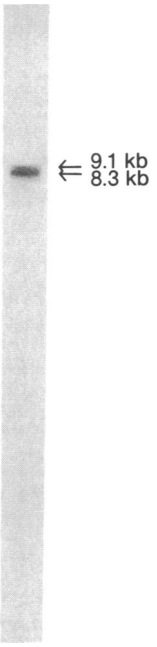

FIG. 1. Northern blot analysis of fatty acid synthase mRNA. Rat liver poly(A) RNA was electrophoresed on a 1% agarose/2.2 M formaldehyde gel (10), blotted onto nitrocellulose, and hybridized with a radiolabeled fragment of pFAS78 (see Fig. 2). The apparent lengths of the two mRNAs, estimated by comparison with the mobilities of DNA standards, were 8.7 and 8.0 kb for the 9.1- and 8.3-kb species. Identical results were obtained with rat mammary gland poly(A) RNA and cDNA probes from other regions of the fatty acid synthase sequence (see Fig. 2).

The remainder of the overlapping cDNA clones were identified by DNA hybridization and their sequences encoded a single open reading frame continuous with those of clones pFAS1, pFAS5, and pFAS3. The only major difficulty encountered in completing the sequence arose from the presence of an internal *Eco*RI site at nt 2545 that may not have been effectively protected by methylation during construction of the mammary gland library. Thus a number of cDNA clones were isolated, all of which appeared to terminate at the same 5'-end sequence as did λFAS3 (C.M.A., unpublished data). One clone, λFAS78, was identified that, when partially digested with *Eco*RI, yielded an insert extending an additional 130 nt beyond the restriction site. To ensure that this *Eco*RI site was not an artifact created during library construction, we prepared a second cDNA library from rat liver mRNA. A strategy was adopted that avoided *Eco*RI digestion and enriched the library in clones representative of this region of the fatty acid synthase. Clone λFAS54, isolated from this library, overlapped λFAS78, and its sequence confirmed that of λFAS78. Verification of the sequence permitted us to complete the cDNA walk using the mammary gland library ultimately yielding clone λFAS27. Primer-extension experiments conducted with rat liver mRNA and an oligonucleotide primer representing nt 55–87 of the antisense strand of the fatty acid synthase cDNA suggest that the mRNA may extend no more than 3 nt beyond the 5' end of clone pFAS27 (C.M.A., unpublished data). The size of the full-length mRNA, estimated using DNA standards, agreed well with that calculated from the sequence (Fig. 1). The shorter mRNA species appears to be transcribed using an alternative polyadenylylation signal $A^{8365}TTAAA$ (J.N., unpublished data). Clone λFAS27 contained a termination codon, $T^{28}GA$, in the same reading frame as the coding sequence followed, at nt 82, by a sequence (GCCATGG) typical of the vertebrate ribosome binding site (15). Thus of the possible translation start sites, the furthest upstream ATG (at nt 85) likely encodes the amino-terminal methionine residue. The nucleotide sequence of the 5' noncoding and the coding region, together with the deduced amino acid sequence, is presented in Fig. 3.

Confirmation of the predicted amino acid sequence was obtained in part by direct protein sequencing. We prepared and fractionated CNBr fragments from the fatty acid synthase core (Fig. 4) that contains all the domains of the multifunctional subunit except the thioesterase for which protein sequence data was available (4). Two CNBr peptides predicted only by the first of the possible start sites were identified (Table 1, fragments 1 and 2), establishing that $A^{85}TG$ is the codon for the amino-terminal methionine. The amino terminus appears to be blocked since we were unable to obtain any sequence data from the intact fatty acid synthase polypeptide. Thirty-seven additional CNBr fragments were identified with sequences predicted correctly by the cDNA sequence. Two of the sequenced CNBr fragments (fragments 24 and 28) were not preceded by methionine residues in the predicted sequence. Fragment 28 resulted from the tryptic cleavage at Lys-1281 incurred during preparation of the trypsin-digested fatty acid synthase core. This was confirmed when the 99-kDa carboxyl-terminal tryptic core peptide (residues 1282–2199) was isolated and found to have the same amino-terminal sequence as CNBr fragment 28. The most likely explanation for the isolation of CNBr fragment 24 is the occurrence of a polymorphism at residue 1145 in the outbred Long–Evans strain. The codon for this residue was unambiguously established in both clones pFAS3 and pFAS13 as CTG so that a C → A point mutation at nt 3517

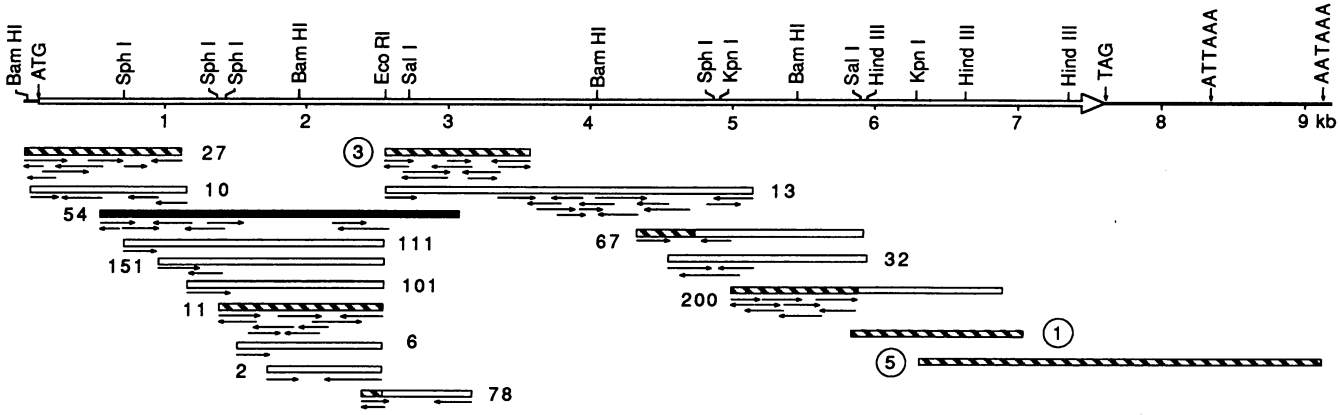

FIG. 2. Strategy for constructing the nucleotide sequence for rat fatty acid synthase from overlapping cDNA clones. The individual λ clones used in this construction are shown beneath a physical map showing restriction enzyme sites within the sequence. The open reading frame deduced from this sequence is shown as an open arrow. The two signals for polyadenylylation of the 8.3- and 9.1-kb mRNAs (ATTAAA and AATAAA, respectively) are shown. Clones are represented as numbered bars with arrows indicating direction and extent of sequencing reactions. The sequence determination for clones 1 and 5 has been published (3, 4). All clones were from a rat mammary gland cDNA library except λFAS54, which was from a rat liver library (see text). Clones with circled numbers were identified by antibody screening; hatched whole or partial bars represent parts of the sequence that were used as probes for Northern blot hybridization to verify that cloned sequences were authentic fatty acid synthase (see Fig. 1).

CGCCCTGGGATCCCTGCCTGTCAGGAGTGACCAGCCCAAGCCCTGGCTTGTCGTCTGCCTCCAGATCCCAGACAGAGAAGAGCC 84

ATG GAG GAG GTG GTG ATA GCC GGT ATG TCC GGG AAA TTG CCC GAG TCA GAG AAC CTG CAG GAG TTC TGG GCC AAC CTC ATT GGC GGT GTG GAC ATG GTC ACA GAC GAT GAC AGG AGG TGG 204
Met Glu Glu Val Val Ile Ala Gly Met Ser Gly Lys Leu Pro Glu Ser Glu Asn Leu Gln Glu Phe Trp Ala Asn Leu Ile Gly Gly Val Asp Met Val Thr Asp Asp Asp Arg Arg Trp 40

AAG GCT GGG CTC TAT GGG TTG CCT AAG CGG TCT GGA AAG CTG AAG GAT CTG TCC AAG TTC GAC GCC TCC TTT TTT GGG GTC CAC CCC AAG CAG GCA CAC ACA ATG GAC CCG CAG CTC CGG 324
Lys Ala Gly Leu Tyr Gly Leu Pro Lys Arg Ser Gly Lys Leu Lys Asp Leu Ser Lys Phe Asp Ala Ser Phe Phe Gly Val His Pro Lys Gln Ala His Thr Met Asp Pro Gln Leu Arg 80

CTG CTG CTG GAA GTC AGC TAT GAA GCT ATT GTG GAC GGA GGT ATC AAC CCG GCC TCA CTC CGA GGA ACA AAC ACT GGT GTC TGG GTG GGT GTG AGT GGT TCC GAG GCG TCG GAG GCC CTG 444
Leu Leu Leu Glu Val Ser Tyr Glu Ala Ile Val Asp Gly Gly Ile Asn Pro Ala Ser Leu Arg Gly Thr Asn Thr Gly Val Trp Val Gly Val Ser Gly Ser Glu Ala Ser Glu Ala Leu 120

AGC AGA GAT CCT GAG ACT CTT CTG GGC TAC AGC ATG GTG GGC TGC CAG AGA GCA ATG ATG GCC AAC CGG CTC TCT TTC TTC TTC GAC TTC AAA GGA CCC AGC ATT GCC CTG GAC ACA GCC 564
Ser Arg Asp Pro Glu Thr Leu Leu Gly Tyr Ser Met Val Gly Cys Gln Arg Ala Met Met Ala Asn Arg Leu Ser Phe Phe Phe Asp Phe Lys Gly Pro Ser Ile Ala Leu Asp Thr Ala 160

TGC TCC TCT AGC CTA CTG GCA CTA CAG AAT GCC TAT CAG GCT ATC CGC AGT GGG GAG TGC CCT GCT GCC ATT GTG GGC GGG ATC AAC CTG CTG CTA AAG CCT AAC ACC TCT GTG CAG TTC 684
CYS Ser Ser Ser Leu Leu Ala Leu Gln Asn Ala Tyr Gln Ala Ile Arg Ser Gly Glu Cys Pro Ala Ala Ile Val Gly Gly Ile Asn Leu Leu Leu Lys Pro Asn Thr Ser Val Gln Phe 200

ATG AAG CTA GGC ATG CTC AGC CCC GAT GGC ACC TGC AGA TCC TTT GAT GAT TCA GGG AAC GGG TAT TGC CGT GCT GAG GCT GTC GTG GCA GTT CTG CTG ACT AAG AAG TCC TTG GCT CGG 804
Met Lys Leu Gly Met Leu Ser Pro Asp Gly Thr Cys Arg Ser Phe Asp Asp Ser Gly Asn Gly Tyr Cys Arg Ala Glu Ala Val Val Ala Val Leu Leu Thr Lys Lys Ser Leu Ala Arg 240

CGA GTC TAT GCC ACT ATT CTG AAT GCC GGG ACG AAC ACA GAT GGC TGC AAG GAG CAA GGC GTG ACA TTC CCC TCT GGA GAA GCC CAG GAA CAA CTC ATC CGT TCT CTG TAT CAG CCG GGC 924
Arg Val Tyr Ala Thr Ile Leu Asn Ala Gly Thr Asn Thr Asp Gly Cys Lys Glu Gln Gly Val Thr Phe Pro Ser Gly Glu Ala Gln Glu Gln Leu Ile Arg Ser Leu Tyr Gln Pro Gly 280

GGT GTG GCC CCC GAG TCT CTT GAA TAT ATT GAA GCC CAT GGC ACG GGC ACC AAG GTG GGG GAC CCC CAG GAA CTG AAC GGC ATT ACT CGG TCC CTG TGT GCT TTC CGC CAG AGC CCT TTG 1044
Gly Val Ala Pro Glu Ser Leu Glu Tyr Ile Glu Ala His Gly Thr Gly Thr Lys Val Gly Asp Pro Gln Glu Leu Asn Gly Ile Thr Arg Ser Leu Cys Ala Phe Arg Gln Ser Pro Leu 320

TTA ATT GGC TCC ACC AAA TCC AAC ATG GGA CAC CCT GAG CCT GCC TCG GGG CTT GCA GCC CTG ACC AAG GTG CTG TTA TCC CTA GAA AAT GGG GTT TGG GCC CCC AAC CTG CAT TTC CAC 1164
Leu Ile Gly Ser Thr Lys Ser Asn Met Gly His Pro Glu Pro Ala Ser Gly Leu Ala Ala Leu Thr Lys Val Leu Leu Ser Leu Glu Asn Gly Val Trp Ala Pro Asn Leu His Phe His 360

AAC CCC AAC CCT GAA ATC CCA GCA CTT CTT GAT GGG CGG CTG CAG GTG GTC GAT AGG CCC CTG CCT GTT CGT GGT GGC ATC GTG GGC ATC AAC TCG TTT GGC TTC GGA GGT GCC AAT GTT 1284
Asn Pro Asn Pro Glu Ile Pro Ala Leu Leu Asp Gly Arg Leu Gln Val Val Asp Arg Pro Leu Pro Val Arg Gly Gly Ile Val Gly Ile Asn Ser Phe Gly Phe Gly Gly Ala Asn Val 400

CAC GTC ATC CTC CAG CCC AAC ACA CAG CAG GCC CCA GCA CCT GCC CCA CAT GCT GCC CTA CCG CAT TTG CTG CAT GCC AGT GGA CGG ACC ATG GAG GCA GTG CAG GGC CTG CTG GAA CAG 1404
His Val Ile Leu Gln Pro Asn Thr Gln Gln Ala Pro Ala Pro Ala Pro His Ala Ala Leu Pro His Leu Leu His Ala Ser Gly Arg Thr Met Glu Ala Val Gln Gly Leu Leu Glu Gln 440

GGC CGC CAG CAC AGT CAG GAC TTG GCC TTT GTG AGC ATG CTC AAT GAC ATT GCA GCA ACC CCT ACA GCA GCC ATG CCC TTC AGA GGT TAC ACT GTG TTA GGT GTT GAG GGC CAT GTC CAG 1524
Gly Arg Gln His Ser Gln Asp Leu Ala Phe Val Ser Met Leu Asn Asp Ile Ala Ala Thr Pro Thr Ala Ala Met Pro Phe Arg Gly Tyr Thr Val Leu Gly Val Glu Gly His Val Gln 480

GAA GTG CAG CAA GTG CCT GCC AGC CAG CGC CCA CTC TGG TTC ATC TGC TCA GGG ATG GGC ACA CAG TGG CGT GGA ATG GGG CTG AGC CTT ATG CGC CTG GAC AGT TTC CGT GAG TCC ATC 1644
Glu Val Gln Gln Val Pro Ala Ser Gln Arg Pro Leu Trp Phe Ile Cys Ser Gly Met Gly Thr Gln Trp Arg Gly Met Gly Leu Ser Leu Met Arg Leu Asp Ser Phe Arg Glu Ser Ile 520

CTG CGC TCT GAT GAG GCT CTG AAG CCC TTG GGA GTC AAA GTG TCA GAC CTG CTG CTG AGC ACT GAT GAG CAC ACC TTT GAT GAC ATC GTG CAT TCC TTT GTG AGC CTC ACC GCC ATC CAG 1764
Leu Arg Ser Asp Glu Ala Leu Lys Pro Leu Gly Val Lys Val Ser Asp Leu Leu Leu Ser Thr Asp Glu His Thr Phe Asp Asp Ile Val His Ser Phe Val Ser Leu Thr Ala Ile Gln 560

ATT GCC CTC ATC GAC CTG CTG ACG TCT ATG GGG CTG AAA CCT GAT GGC ATC ATT GGG CAC TCC TTG GGA GAG GTT GCC TGT GGC TAT GCA GAT GGC TGT CTC TCC CAG AGA GAG GCT GTG 1884
Ile Ala Leu Ile Asp Leu Leu Thr Ser Met Gly Leu Lys Pro Asp Gly Ile Ile Gly His SER Leu Gly Glu Val Ala Cys Gly Tyr Ala Asp Gly Cys Leu Ser Gln Arg Glu Ala Val 600

CTT GCA GCC TAC TGG AGA GGC CAG TGC ATT AAG GAT GCC AAC CTT CCG GCT GGA TCC ATG GCA GCT GTT GGT TTG TCC TGG GAA GAA TGT AAA CAA CGC TGC CCT CCT GGT GTG GTG CCT 2004
Leu Ala Ala Tyr Trp Arg Gly Gln Cys Ile Lys Asp Ala Asn Leu Pro Ala Gly Ser Met Ala Ala Val Gly Leu Ser Trp Glu Glu Cys Lys Gln Arg Cys Pro Pro Gly Val Val Pro 640

GCC TGC CAC AAC TCT GAG GAC ACT GTG ACC ATC TCT GGA CCT CAG GCT GCA GTG AAT GAA TTT GTG GAG CAG CTA AAG CAA GAG GGC GTG TTT GCC AAG GAG GTG CGA ACA GGT GGC CTG 2124
Ala Cys His Asn Ser Glu Asp Thr Val Thr Ile Ser Gly Pro Gln Ala Ala Val Asn Glu Phe Val Glu Gln Leu Lys Gln Glu Gly Val Phe Ala Lys Glu Val Arg Thr Gly Gly Leu 680

GCC TTC CAC TCC TAC TTC ATG GAA GGA ATT GCC CCC ACG CTG CTG CAG GCT CTC AAG AAG GTG ATC CGG GAG CCA CGG CCA CGC TCA GCA CGC TGG CTC AGC ACC TCT ATC CCT GAG GCC 2244
Ala Phe His Ser Tyr Phe Met Glu Gly Ile Ala Pro Thr Leu Leu Gln Ala Leu Lys Lys Val Ile Arg Glu Pro Arg Pro Arg Ser Ala Arg Trp Leu Ser Thr Ser Ile Pro Glu Ala 720

CAG TGG CAG AGC AGC CTG GCC CGC ACA TCT TCT GCT GAG TAC AAC GTC AAC AAC CTG GTG AGC CCT GTG CTC TTC CAG GAA GCA CTG TGG CAC GTC CCC GAG CAC GCC GTG GTG CTG GAG 2364
Gln Trp Gln Ser Ser Leu Ala Arg Thr Ser Ser Ala Glu Tyr Asn Val Asn Asn Leu Val Ser Pro Val Leu Phe Gln Glu Ala Leu Trp His Val Pro Glu His Ala Val Val Leu Glu 760

ATT GCA CCC CAT GCA CTG TTG CAG GCT GTC CTG AAG CGA GGC GTG AAG CCT AGC TGC ACC ATC ATC CCC TTG ATG AAG AGG GAC CAT AAA GAT AAC TTG GAG TTC TTC CTC ACC AAC CTC 2484
Ile Ala Pro His Ala Leu Leu Gln Ala Val Leu Lys Arg Gly Val Lys Pro Ser Cys Thr Ile Ile Pro Leu Met Lys Arg Asp His Lys Asp Asn Leu Glu Phe Phe Leu Thr Asn Leu 800

GGC AAG GTG CAC CTC ACA GGC ATC GAC ATC AAC CCT AAT GCC TTG TTC CCA CCT GTG GAA TTC CCG GTT CCC CGA GGG ACT CCT CTC ATC TCC CCT CAC ATC AAG TGG GAC CAC AGT CAG 2604
Gly Lys Val His Leu Thr Gly Ile Asp Ile Asn Pro Asn Ala Leu Phe Pro Pro Val Glu Phe Pro Val Pro Arg Gly Thr Pro Leu Ile Ser Pro His Ile Lys Trp Asp His Ser Gln 840

ACT TGG GAT ATC CCA GTT GCT GAA GAC TTC CCC AAC GGT TCC AGC TCC TCC TCA GCT ACA GTC TAC AAC ATT GAC GCC AGT TCC GAG TCA TCT GAC CAC TAC CTG GTC GAC CAC TGC ATT 2724
Thr Trp Asp Ile Pro Val Ala Glu Asp Phe Pro Asn Gly Ser Ser Ser Ser Ser Ala Thr Val Tyr Asn Ile Asp Ala Ser Ser Glu Ser Ser Asp His Tyr Leu Val Asp His Cys Ile 880

GAC GGC CGT GTC CTC TTC CCT GGC ACT GGC TAC CTG TAC CTG GTG TGG AAG ACA CTG GCT CGA AGC CTG AGC TTG TCC CTA GAA GAG ACC CCT GTG GTG TTT GAG AAC GTG ACA TTT CAT 2844
Asp Gly Arg Val Leu Phe Pro Gly Thr Gly Tyr Leu Tyr Leu Val Trp Lys Thr Leu Ala Arg Ser Leu Ser Leu Ser Leu Glu Glu Thr Pro Val Val Phe Glu Asn Val Thr Phe His 920

CAG GCC ACC ATC CTG CCC AGG ACA GGA ACC GTG CCT CTG GAG GTG CGG CTG CTA GAG GCC TCA CAT GCA TTT GAG GTG TCT GAC AGT GGC AAC CTG ATA GTG AGC GGG AAA GTG TAC CAG 2964
Gln Ala Thr Ile Leu Pro Arg Thr Gly Thr Val Pro Leu Glu Val Arg Leu Leu Glu Ala Ser His Ala Phe Glu Val Ser Asp Ser Gly Asn Leu Ile Val Ser Gly Lys Val Tyr Gln 960

TGG GAA GAC CCT GAC TCC AAG TTA TTC GAC CAC CCA GAA GTC CCG ATC CCC GCC GAG TCC GAG TCT GTC TCC CGC TTG ACG CAG GGA GAA GTA TAC AAG GAG CTG CGG CTA CGT GGC TAT 3084
Trp Glu Asp Pro Asp Ser Lys Leu Phe Asp His Pro Glu Val Pro Ile Pro Ala Glu Ser Glu Ser Val Ser Arg Leu Thr Gln Gly Glu Val Tyr Lys Glu Leu Arg Leu Arg Gly Tyr 1000

GAC TAT GGC CCT CAT TTC CAG GGC GTC TAT GAG GCC ACC CTC GAA GGT GAG CAA GGC AAG CTG CTC TGG AAA GAC AAC TGG GTG ACC TTC ATG GAC ACA ATG CTG CAG ATA TCC ATC CTG 3204
Asp Tyr Gly Pro His Phe Gln Gly Val Tyr Glu Ala Thr Leu Glu Gly Glu Gln Gly Lys Leu Leu Trp Lys Asp Asn Trp Val Thr Phe Met Asp Thr Met Leu Gln Ile Ser Ile Leu 1040

GGC TTC AGC AAG CAG AGT CTG CAG CTA CCC ACC CGT GTG ACT GCC ATC TAT ATT GAC CCT GCA ACC CAC CTG CAG AAG GTG TAC ATG CTG GAG GGA GAC ACT CAA GTG GCT GAC GTG ACC 3324
Gly Phe Ser Lys Gln Ser Leu Gln Leu Pro Thr Arg Val Thr Ala Ile Tyr Ile Asp Pro Ala Thr His Leu Gln Lys Val Tyr Met Leu Glu Gly Asp Thr Gln Val Ala Asp Val Thr 1080

ACG AGC CGC TGT CTG GGC GTG ACC GTC TCT GGT GGT GTC TAC ATT TCG AGA CTA CAG ACA ACA GCA ACC TCA CGG CGG CAG CAG GAA CAG CTG GTC CCC ACC CTG GAG AAG TTT GTC TTC 3444
Thr Ser Arg Cys Leu Gly Val Thr Val Ser Gly Gly Val Tyr Ile Ser Arg Leu Gln Thr Thr Ala Thr Ser Arg Arg Gln Gln Glu Gln Leu Val Pro Thr Leu Glu Lys Phe Val Phe 1120

ACA CCC CAT GTG GAG CCT GAG TGC CTG TCT GAG AGT GCT ATC CTG CAG AAA GAG CTG CAG CTG TGC AAG GGT CTG GCA AAG GCT CTG CAG ACC AAG GCC ACC CAG CAA GGG CTG AAG ATG 3564
Thr Pro His Val Glu Pro Glu Cys Leu Ser Glu Ser Ala Ile Leu Gln Lys Glu Leu Gln Leu Cys Lys Gly Leu Ala Lys Ala Leu Gln Thr Lys Ala Thr Gln Gln Gly Leu Lys Met 1160

ACA GTG CCT GGG CTA GAG GAC CTT CCC CAG CAT GGA CTG CCT CGA CTC TTG GCT GCT GCC TGC CAG CTG CAG CTC AAC GGG AAC CTG CAA CTG GAG TTA GGT GAG GTA CTG GCT CGA GAG 3684
Thr Val Pro Gly Leu Glu Asp Leu Pro Gln His Gly Leu Pro Arg Leu Leu Ala Ala Ala Cys Gln Leu Gln Leu Asn Gly Asn Leu Gln Leu Glu Leu Gly Glu Val Leu Ala Arg Glu 1200

AGG CTC CTG CTG CCA GAA GAC CCT CTG ATC AGT GGC CTC CTT AAC TCC CAG GCC CTC AAG GCC TGC ATA GAC ACA GCC CTG GAG AAC CTG TCT ACT CTC AAG ATG AAG GTG GTG GAG GTG 3804
Arg Leu Leu Leu Pro Glu Asp Pro Leu Ile Ser Gly Leu Leu Asn Ser Gln Ala Leu Lys Ala Cys Ile Asp Thr Ala Leu Glu Asn Leu Ser Thr Leu Lys Met Lys Val Val Glu Val 1240

CTG GCT GGA GAA GGC CAC TTG TAT TCC CAC ATC TCA GCA CTG CTC AAC ACC CAG CCT ATG CTG CAA CTG GAG TAT ACA GCC ACC GAC CGG CAC CCC CAG GCC CTG AAG GAT GTT CAG ACC 3924
Leu Ala Gly Glu Gly His Leu Tyr Ser His Ile Ser Ala Leu Leu Asn Thr Gln Pro Met Leu Gln Leu Glu Tyr Thr Ala Thr Asp Arg His Pro Gln Ala Leu Lys Asp Val Gln Thr 1280

AAG CTG CAG CAG CAT GAT GTA GCA CAG GGC CAG TGG GAC CCT TCT GGT CCT GCT CCT ACC AAC CTG GGT GCT CTT GAC CTT GTG GTG TGC AAC TGT GCG TTA GCC ACC CTG GGG GAT CCA 4044
Lys Leu Gln Gln His Asp Val Ala Gln Gly Gln Trp Asp Pro Ser Gly Pro Ala Pro Thr Asn Leu Gly Ala Leu Asp Leu Val Val Cys Asn Cys Ala Leu Ala Thr Leu Gly Asp Pro 1320

GCC CTG GCC CTG GAC AAC ATG GTA GCT GCC CTC AAG GAT GGT GGT TTC CTG CTA ATG CAC ACA GTG CTC AAA GGA CAT GCC CTT GGG GAA ACC CTG GCC TGC CTC CCT TCT GAG GTG CAG 4164
Ala Leu Ala Leu Asp Asn Met Val Ala Ala Leu Lys Asp Gly Gly Phe Leu Leu Met His Thr Val Leu Lys Gly His Ala Leu Gly Glu Thr Leu Ala Cys Leu Pro Ser Glu Val Gln 1360

CCT GGG CCC AGC TTC TTA AGC CAG GAA GAG TGG GAG AGC CTG TTC TCA AGG AAG GCA CTG CAC CTG GTG GGC CTT AAA AAG TCA TTC TAC GGT ACT GCG CTG TTC CTG TGC CGC CGT CTC 4284
Pro Gly Pro Ser Phe Leu Ser Gln Glu Glu Trp Glu Ser Leu Phe Ser Arg Lys Ala Leu His Leu Val Gly Leu Lys Lys Ser Phe Tyr Gly Thr Ala Leu Phe Leu Cys Arg Arg Leu 1400

AGC CCA CAG GAC AAG CCC ATC TTC CTG CCT GTG GAG GAT ACT AGT TTC CAG TGG GTG GAC TCT CTG AAG AGC ATT CTG GCC ACA TCC TCC TCC CAG CCT GTG TGG CTA ACA GCC ATG AAC 4404
Ser Pro Gln Asp Lys Pro Ile Phe Leu Pro Val Glu Asp Thr Ser Phe Gln Trp Val Asp Ser Leu Lys Ser Ile Leu Ala Thr Ser Ser Ser Gln Pro Val Trp Leu Thr Ala Met Asn 1440

TGC CCC ACC TCA GGT GTG GTA GGC TTG GTG AAC TGT CTC CGA AAA GAG CCG GGT GGA CAC CGG ATT CGG TGT ATC CTG CTG TCC AAC CTC AGC AGC ACA TCT CAC GTC CCC AAG CTG GAC 4524
Cys Pro Thr Ser Gly Val Val Gly Leu Val Asn Cys Leu Arg Lys Glu Pro Gly Gly His Arg Ile Arg Cys Ile Leu Leu Ser Asn Leu Ser Ser Thr Ser His Val Pro Lys Leu Asp 1480

CCT GGC TCT TCA GAG CTA CAG AAG GTG CTA GAG AGT GAT CTG GTG ATG AAC GTG TAC AGG GAC GGT GCC TGG GGT GCC TTC CGT CAC TTC CAG TTA GAG CAG GAC AAG CCC GAG GAG CAG 4644
Pro Gly Ser Ser Glu Leu Gln Lys Val Leu Glu Ser Asp Leu Val Met Asn Val Tyr Arg Asp Gly Ala Trp Gly Ala Phe Arg His Phe Gln Leu Glu Gln Asp Lys Pro Glu Glu Gln 1520

ACA GCA CAT GCC TTT GTA AAC GTC CTT ACC CGA GGG GAC CTT GCC TCC ATC CGC TGG GTC TCT TCT CCC CTG AAA CAC ATG CAG CCG CCC TCG AGC TCA GGA GCA CAG CTC TGC ACT GTC 4764
Thr Ala His Ala Phe Val Asn Val Leu Thr Arg Gly Asp Leu Ala Ser Ile Arg Trp Val Ser Ser Pro Leu Lys His Met Gln Pro Pro Ser Ser Ser Gly Ala Gln Leu Cys Thr Val 1560

TAC TAT GCC TCA CTG AAC TTC CGA GAT ATC ATG CTG GCC ACG GGC AAG CTG TCC CCT GAT GCC ATT CCA GGT AAA TGG GCC AGC CGG GAC TGC ATG CTT GGC ATG GAG TTC TCA GGC CGT 4884
Tyr Tyr Ala Ser Leu Asn Phe Arg Asp Ile Met Leu Ala Thr Gly Lys Leu Ser Pro Asp Ala Ile Pro Gly Lys Trp Ala Ser Arg Asp Cys Met Leu Gly Met Glu Phe Ser Gly Arg 1600

GAT AAG TGC GGC CGG CGT GTG ATG GGG CTG GTA CCC GCA GAA GGC CTG GCC ACC TCA GTC CTG TTA TCA CCC GAC TTC CTC TGG GAT GTA CCC TCT AGC TGG ACC CTG GAG GAG GCG GCT 5004
Asp Lys Cys Gly Arg Arg Val Met Gly Leu Val Pro Ala Glu Gly Leu Ala Thr Ser Val Leu Leu Ser Pro Asp Phe Leu Trp Asp Val Pro Ser Ser Trp Thr Leu Glu Glu Ala Ala 1640

TCT GTG CCT GTT GTC TAC ACC ACC GCC TAC TAC TCC TTA GTA GTG CGT GGT CGT ATT CAG CAC GGG GAA ACT GTG CTC ATT CAC TCG GGC TCC GGT GGT GTG GGC CAA GCG GCC ATT TCC 5124
Ser Val Pro Val Val Tyr Thr Thr Ala Tyr Tyr Ser Leu Val Val Arg Gly Arg Ile Gln His Gly Glu Thr Val Leu Ile His Ser Gly Ser Gly Gly Val Gly Gln Ala Ala Ile Ser 1680

FIG. 3. (*Figure continues on the opposite page.*)

```
ATT GCC CTT AGC CTG GGC TGC CGA GTC TTC ACC ACT GTG GGC TCC GCT GAG AAG CGA GCT TAC CTC CAG GCC AGA TTC CCT CAG CTG GAT GAC ACC AGC TTT GCT AAC TCT CGA GAC ACA  5244
Ile Ala Leu Ser Leu Gly Cys Arg Val Phe Thr Thr Val Gly Ser Ala Glu LYS Arg Ala Tyr Leu Gln Ala Arg Phe Pro Gln Leu Asp Asp Thr Ser Phe Ala Asn Ser Arg Asp Thr  1720

TCG TTT GAG CAG CAT GTG TTA CTG CAC ACA GGT GGC AAA GGG GTG GAC CTG GTC CTC AAC TCC CTG GCA GAA GAG AAG CTG CAG GCC AGT GTG CGG TGC TTG GCT CAG CAT GGC CGC TTC  5364
Ser Phe Glu Gln His Val Leu Leu His Thr Gly Gly Lys Gly Val Asp Leu Val Leu Asn Ser Leu Ala Glu Glu Lys Leu Gln Ala Ser Val Arg Cys Leu Ala Gln His Gly Arg Phe  1760

CTA GAG ATC GGC AAA TTT GAT CTT TCT AAC AAC CAC CCT CTG GGC ATG GCC ATC TTC TTG AAG AAC GTC ACT TTC CAT GGG ATC CTG CTG GAT GCA CTT TTT GAG GGG GCC AAC GAC AGC  5484
Leu Glu Ile Gly Lys Phe Asp Leu Ser Asn Asn His Pro Leu Gly Met Ala Ile Phe Leu Lys Asn Val Thr Phe His Gly Ile Leu Leu Asp Ala Leu Phe Glu Gly Ala Asn Asp Ser  1800

TGG CGG GAG GTG GCA GAG CTG CTG AAG GCC GGC ATC CGT GAT GGG GTT GTG AAG CCT CTC AAG TGT ACA GTG TTT CCC AAG GCC CAG GTG GAG GAC GCC TTC CGA TAC ATG GCT CAA GGA  5604
Trp Arg Glu Val Ala Glu Leu Leu Lys Ala Gly Ile Arg Asp Gly Val Val Lys Pro Leu Lys Cys Thr Val Phe Pro Lys Ala Gln Val Glu Asp Ala Phe Arg Tyr Met Ala Gln Gly  1840

AAA CAT ATT GGC AAA GTC CTT GTC CAG GTA CGG GAG GAG GAG CCC GAG GCT ATG CTG CCA GGG GCT CAG CCC ACC CTG ATT TCC GCC ATC TCC AAG ACC TTC TGC CCA GAG CAT AAG AGT  5724
Lys His Ile Gly Lys Val Leu Val Gln Val Arg Glu Glu Glu Pro Glu Ala Met Leu Pro Gly Ala Gln Pro Thr Leu Ile Ser Ala Ile Ser Lys Thr Phe Cys Pro Glu His Lys Ser  1880

TAC ATC ATC ACT GGT GGC CTA GGT GGC TTT GGC CTG GAA CTG GCC CGG TGG CTT GTG CTT CGT GGG GCC CAA AGG CTT GTA CTA ACT TCC CGA TCT GGA ATC CGC ACA GGC TAC CAA GCC  5844
Tyr Ile Ile Thr Gly Gly Leu Gly Gly Phe Gly Leu Glu Leu Ala Arg Trp Leu Val Leu Arg Gly Ala Gln Arg Leu Val Leu Thr Ser Arg Ser Gly Ile Arg Thr Gly Tyr Gln Ala  1920

AAG CAC GTT CGG GAG TGG AGG CGC CAG GGC ATC CAT GTG CTA GTG TCG ACA AGC AAT GTC AGT TCA CTG GAG GGG GCC CGT GCT CTC ATC GCT GAA GCC ACA AAG CTT GGG CCC GTT GGA  5964
Lys His Val Arg Glu Trp Arg Arg Gln Gly Ile His Val Leu Val Ser Thr Ser Asn Val Ser Ser Leu Glu Gly Ala Arg Ala Leu Ile Ala Glu Ala Thr Lys Leu Gly Pro Val Gly  1960

GGT GTC TTC AAC CTG GCC ATG GTT TTA AGG GAT GCC ATG CTG GAG AAC CAG ACT CCA GAA CTC TTC CAG GAT GTC AAC AAG CCC AAG TAC AAT GGC ACC CTG AAC CTT GAC AGG GCG ACC  6084
Gly Val Phe Asn Leu Ala Met Val Leu Arg Asp Ala Met Leu Glu Asn Gln Thr Pro Glu Leu Phe Gln Asp Val Asn Lys Pro Lys Tyr Asn Gly Thr Leu Asn Leu Asp Arg Ala Thr  2000

CGG GAA GCC TGT CCT GAG CTG GAC TAC TTT GTG GCC TTC TCC TCT GTA AGC TGC GGG CGT GGT AAT GCT GGC CAA TCC AAC TAT GGC TTC GCC AAC TCT ACC ATG GAG CGT ATT TGC GAA  6204
Arg Glu Ala Cys Pro Glu Leu Asp Tyr Phe Val Ala Phe Ser Ser Val Ser Cys Gly Arg Gly Asn Ala Gly Gln Ser Asn Tyr Gly Phe Ala Asn Ser Thr Met Glu Arg Ile Cys Glu  2040

CAG CGC CGG CAC GAT GGC CTC CCA GGT CTT GCC GTG CAA TGG GGT GCC ATT GGT GAC GTG GGC ATT ATC TTG GAA GCG ATG GGT ACC AAT GAC ACA GTC GTT GGC GGC ACA CTG CCA CAG  6324
Gln Arg Arg His Asp Gly Leu Pro Gly Leu Ala Val Gln Trp Gly Ala Ile Gly Asp Val Gly Ile Ile Leu Glu Ala Met Gly Thr Asn Asp Thr Val Val Gly Gly Thr Leu Pro Gln  2080

CGC ATC TCC TCC TGC ATG GAG GTG CTG GAC CTC TTC CTG AAT CAG CCC CAC GCA GTC CTG AGC AGT TTT GTG CTG GTT GAG AAG AAA GCT GTG GCC CAT GGT GAT GGT GAA GCC CAG AGG  6444
Arg Ile Ser Ser Cys Met Glu Val Leu Asp Leu Phe Leu Asn Gln Pro His Ala Val Leu Ser Ser Phe Val Leu Val Glu Lys Lys Ala Val Ala His Gly Asp Gly Glu Ala Gln Arg  2120

GAT CTG GTG AAA GCA GTG GCA CAC ATC CTA GGC ATC CGC GAC CTC GCA GGG ATT AAC CTG GAC AGC TCG CTG GCA GAC CTC GGC CTG GAC TCG CTC ATG GGT GTG GAA GTG CGC CAG ATC  6564
Asp Leu Val Lys Ala Val Ala His Ile Leu Gly Ile Arg Asp Leu Ala Gly Ile Asn Leu Asp Ser Ser Leu Ala Asp Leu Gly Leu Asp SER Leu Met Gly Val Glu Val Arg Gln Ile  2160

CTG GAA CGT GAA CAT GAT CTG GTG CTA CCC ATT CGT GAA GTA CGG CAA CTC ACA CTG CGG AAG CTT CAG GAA ATG TCC TCC AAG GCT GGC TCA GAC ACT GAG TTG GCA GCC CCC AAG TCC  6684
Leu Glu Arg Glu His Asp Leu Val Leu Pro Ile Arg Glu Val Arg Gln Leu Thr Leu Arg Lys Leu Gln Glu Met Ser Ser Lys Ala Gly Ser Asp Thr Glu Leu Ala Ala Pro Lys Ser  2200

AAG AAT GAT ACA TCC CTG AAG CAG GCC CAG CTG AAT CTG AGT ATC CTG CTG GTG AAC CCT GAG GGC CCT ACC TTA ACA CGA CTC AAC TCA GTG CAG AGC TCT GAG CGG CCT CTG TTC CTG  6804
Lys Asn Asp Thr Ser Leu Lys Gln Ala Gln Leu Asn Leu Ser Ile Leu Leu Val Asn Pro Glu Gly Pro Thr Leu Thr Arg Leu Asn Ser Val Gln Ser Ser Glu Arg Pro Leu Phe Leu  2240

GTG CAC CCC ATT GAA GGT TCC ATC ACT GTG TTC CAC AGC CTG GCT GCC AAG CTC AGT GTG CCC ACC TAC GGT CTG CAG TGC ACC CAA GCG GCC CCC CTG GAC AGC ATT CCA AAC CTG GCT  6924
Val His Pro Ile Glu Gly Ser Ile Thr Val Phe His Ser Leu Ala Ala Lys Leu Ser Val Pro Thr Tyr Gly Leu Gln Cys Thr Gln Ala Ala Pro Leu Asp Ser Ile Pro Asn Leu Ala  2280

GCC TAC TAC ATT GAT TGC ATC AAG CAG GTG CAG CCT GAG GGG CCC CAC CGA GTG GCT GGG TAT TCT TTT GGA GCT TGT GTA GCC TTC GAG ATG TGC TCC CAG CTG CAG GCC CAG CAG GGC  7044
Ala Tyr Tyr Ile Asp Cys Ile Lys Gln Val Gln Pro Glu Gly Pro His Arg Val Ala Gly Tyr SER Phe Gly Ala Cys Val Ala Phe Glu Met Cys Ser Gln Leu Gln Ala Gln Gln Gly  2320

CCA GCC CCC GCC CAC AAC AAC CTC TTC TTG TTT GAT GGC TCA CAC ACC TAC GTA TTG GCG TAC ACC CAG AGC TAC CGG GCA AAG CTG ACC CCA GGC TGT GAG GCT GAG GCT GAA GCT GAA  7164
Pro Ala Pro Ala His Asn Asn Leu Phe Leu Phe Asp Gly Ser His Thr Tyr Val Leu Ala Tyr Thr Gln Ser Tyr Arg Ala Lys Leu Thr Pro Gly Cys Glu Ala Glu Ala Glu Ala Glu  2360

GCC ATA TGC TTC TTC ATT AAG CAG TTT GTT GAT GCA GAG CAT AGC AAG GTG CTA GAG GCC CTG CTA CCA CTG AAG AGC CTG GAG GAC CGG GTT GCT GCT GCT GTG GAC CTC ATC ACT AGA  7284
Ala Ile Cys Phe Phe Ile Lys Gln Phe Val Asp Ala Glu His Ser Lys Val Leu Glu Ala Leu Leu Pro Leu Lys Ser Leu Glu Asp Arg Val Ala Ala Ala Val Asp Leu Ile Thr Arg  2400

AGC CAC CAG AGC CTG GAC CGC CGT GAC CTG AGC TTT GCT GCC GTG TCC TTC TAC TAC AAG CTT CGA GCC GCC GAC CAG TAT AAA CCC AAG GCC AAG TAC CAC GGC AAT GTG ATC CTG CTG  7404
Ser His Gln Ser Leu Asp Arg Arg Asp Leu Ser Phe Ala Ala Val Ser Phe Tyr Tyr Lys Leu Arg Ala Ala Asp Gln Tyr Lys Pro Lys Ala Lys Tyr His Gly Asn Val Ile Leu Leu  2440

CGG GCC AAG ACA GGT GGC ACC TAC GGC GAG GAC TTG GGT GCC GAT TAC AAC CTG TCC CAG GTG TGT GAT GGG AAG GTG TCT GTG CAC ATC ATT GAG GGT GAC CAT CGT ACG CTG CTG GAG  7524
Arg Ala Lys Thr Gly Gly Thr Tyr Gly Glu Asp Leu Gly Ala Asp Tyr Asn Leu Ser Gln Val Cys Asp Gly Lys Val Ser Val His Ile Ile Glu Gly Asp His Arg Thr Leu Leu Glu  2480

GGC AGG GGC CTG GAG TCT ATC ATC AAC ATC ATC CAC AGC TCC CTG GCT GAG CCT CGA GTG AGT GTA CGG GAG GGC TAG
Gly Arg Gly Leu Glu Ser Ile Ile Asn Ile Ile His Ser Ser Leu Ala Glu Pro Arg Val Ser Val Arg Glu Gly  *
```

FIG. 3. The nucleotide sequence of a composite fatty acid synthase cDNA and its predicted amino acid sequence. The sequence of the 1537-nt 3′ noncoding region (4) has been omitted. Sequences near the active sites of the condensing enzyme (around Cys-161), transferase (around Ser-581), and thioesterase (around Ser-2302), the point of attachment of the 4′-phosphopantetheine (Ser-2151), the two nucleotide binding sites (near Gly-1670 and Gly-1886), and the pyridoxal phosphate binding site of the enoyl reductase (Lys-1698) are underlined.

could result in the conservative replacement of leucine with methionine. The predicted amino acid composition of the whole fatty acid synthase agrees well with that determined experimentally (16), but the calculated molecular weight (272,340, not including the amino-terminal blocking group) is slightly higher than that determined (240,000) by equilibrium sedimentation (16) and $NaDodSO_4$/PAGE (17). The partial sequence of the liver enzyme, derived by cDNA sequencing of clone pFAS54 and direct amino acid sequencing of CNBr fragments is in agreement with the sequence of the mammary gland enzyme predicted by cDNA sequencing (with the exception at residue 1145, mentioned above). These results lend strong support to our suggestion that the fatty acid synthases of these tissues are identical (18).

Ordering of the Functional Domains. The active site cysteine of the condensing domain was localized within the sequence Asp-Thr-Ala-Cys[161]-Ser-Ser-Ser-Leu, which is identical to that of a region of the avian fatty acid synthase shown to be protected from labeling with iodoacetamide by the substrate acetyl-CoA (19). The active site of the trans-

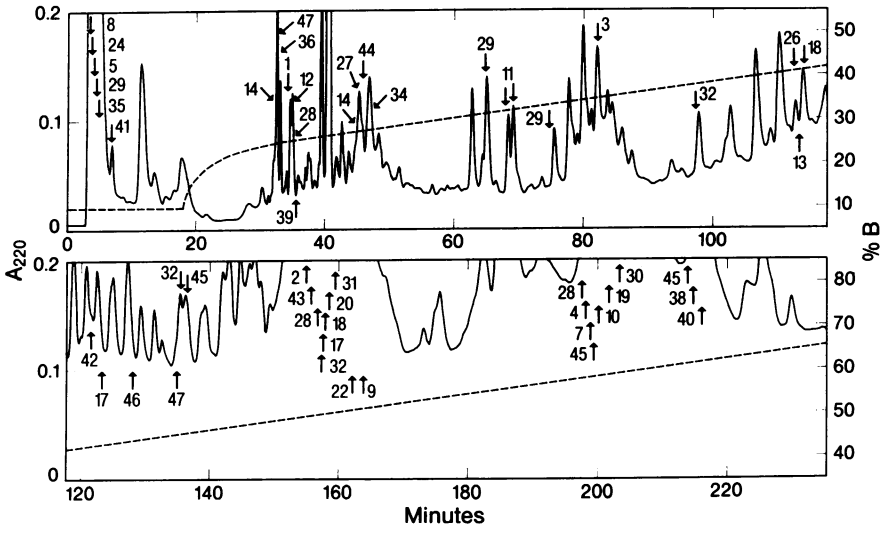

FIG. 4. Fractionation of CNBr fragments derived from the fatty acid synthase. Core polypeptides (0.6 mg) were digested with CNBr and chromatographed on a reversed-phase C_4 column (0.45 × 15 cm, 300-Å pore, 5-μm particle size; Vydac, Hesperia, CA). Solvent A is 0.1% trifluoroacetic acid in H_2O. Solvent B is 0.1% trifluoroacetic acid in acetonitrile/2-propanol/H_2O, 3:1:1 (vol/vol). The gradient is indicated by the dashed line. Arrows indicate the location of specific CNBr fragments identified according to the numbering system in Table 1.

Table 1. Amino acid sequences of CNBr fragments derived from the trypsin-digested fatty acid synthetase core

CNBr fragment	Position of first residue	Sequence	*n*
1	2	EEVVIAGM	2
2	10	SXKL	2
3	33	VTXXDRRWKA	1
4	76	XPQLRLLLEV	1
5	133	VGCQRAM	4
7	141	ANRLSXFFDF	1
8	202	KLGM	2
9	206	LSPDGTCR	2
10	330	GHPEPASGL	3
11	432	EAVQG	1
12	454	LNDIAATPTA	2
13	466	PFXGYTVLGV	3
14	500	GTQWR	4
17	571	GLKPDGIIGH	6
18	621	AAVGLSXEEC	2
19	688	EGIAPTLLQAL	1
20	786	KRDHKDNLEFFLT	2
22	1035	LQISILGF	2
24	1146	AKALQTK	3
26	1236	KVVEVLAGEG	3
27	1261	LQLEYTATDRHPQALKDVQT	1
28	1282	LQQHDVAQGQXDPSGP	4
29	1328	VAALKDGGFLLM	3
30	1340	XTVLKGHALGETLAC	2
31	1440	NCPTSGVVGL	2
32	1497	NVYRDGAWGA	3
34	1572	LATGKLSPDA	3
35	1593	LGM	2
36	1596	EFSGR	1
38	1777	AIFLKNVTFH	3
39	1838	AQGKHIGKVL	1
40	1859	LPGAQPTLIS	2
41	1968	VLRDAM	2
42	1974	LXNQTPELFQ	2
43	2036	EXIXEQRRHD	1
44	2068	GTNDTVVGGT	2
45	2087	EVLDLFLNQP	3
46	2154	GVEVRQILE	3
47	2186	SSKAGSDTXL	2

Some peptides were isolated from more than one digest, others were located in more than one chromatographic zone from a single digest; *n* indicates the number of times a given peptide was sequenced. In some cases the amino acid sequences of more than one peptide could be identified unambiguously from analysis of a mixture; in other cases, prior to sequencing, the peptides were purified to near homogeneity by rechromatography on C_4, C_{18}, or diphenyl columns. The single-letter amino acid code is used.

ferase domain was located by identifying the sequence Asp-Gly-Ile-Ile-Gly-His-**Ser**^{581}-Leu-Gly-Glu-Val-Ala, which is identical to that of a region of the goat enzyme (20) that is labeled by both acetyl and malonyl moieties. The finding of only one substrate loading site sequence in the entire polypeptide confirms conclusively earlier suggestions that both acetyl and malonyl moieties gain access to the enzyme by the same transferase activity (20–22). The consensus sequence Gly-Xaa-Gly-Xaa-Xaa-Gly, found in most nucleotide binding proteins (23), was found at two locations within the fatty acid synthase sequence, Gly^{1670}-Ser-Gly-Gly-Val-Gly and Gly^{1886}-Leu-Gly-Gly-Phe-Gly. The first of these motifs is near a sequence, Thr-Thr-Val-Gly-Ser-Ala-Glu-**Lys**^{1698}-Arg, containing a lysine residue that when reacted with pyridoxal phosphate resulted in specific inactivation of the enoyl reductase activity of the avian enzyme (24). The enoyl reductase of the rat enzyme was also inhibited by pyridoxal phosphate, and NADPH was protected against inhibition, whereas the ketoreductase activity was unaffected (S.S., unpublished data). Thus the first nucleotide binding site motif can be assigned to the enoyl reductase domain and the second to the ketoreductase domain. The discovery of two nucleotide binding site motifs is consistent with the earlier finding that two pyridine nucleotide binding sites can be stoichiometrically titrated per subunit of the animal fatty acid synthase (25). The sequence Gly-Leu-Asp-**Ser**^{2151}-Leu is identical to that of a pentapeptide that carries the 4′-phosphopantetheine prosthetic group of the rat acyl carrier protein domain (26). The active site region of the thioesterase domain was located at the sequence Gly-Tyr-**Ser**^{2302}-Phe-Gly (4). The only functional domain that cannot be precisely located at present is the dehydrase. However, there is only one substantial polypeptide region remaining unassigned, the region between the transferase and enoyl reductase domains, so it seems likely that the dehydrase activity resides within this region. Thus the ordering of functional domains on the animal fatty acid synthase is established: condensing enzyme–transferase–dehydrase–enoyl reductase–ketoreductase–acyl carrier protein–thioesterase. With seven functional domains located within a single polypeptide chain, this represents one of the most complex protein subunits yet sequenced. The availability of this complete sequence will permit critical study of structure–function relationships and the evolutionary origins of the multifunctional fatty acid synthase.

This work was supported by Grants AM16073 and HD12588 from the National Institutes of Health.

1. Lynen, F. (1961) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **20**, 941–951.
2. Hardie, D. G. & McCarthy, A. D. (1986) in *Mutidomain Proteins—Structure and Evolution*, eds. Hardie, D. G. & Coggins, J. R. (Elsevier, Amsterdam), pp. 229–258.
3. Witkowski, A., Naggert, J., Mikkelsen, J. & Smith, S. (1987) *Eur. J. Biochem.* **165**, 601–606.
4. Naggert, J., Witkowski, A., Mikkelsen, J. & Smith, S. (1988) *J. Biol. Chem.* **263**, 1146–1150.
5. Yuan, Z., Lie, W. & Hammes, G. G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6328–6331.
6. Kasturi, R., Chirala, S., Pazirandeh, M. & Wakil, S. J. (1988) *Biochemistry* **27**, 7778–7785.
7. Naggert, J., Williams, B., Cashman, D. P. & Smith, S. (1987) *Biochem. J.* **243**, 597–601.
8. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
9. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
11. Ausubel, F. M., Brent, R., Kingston, R. G., Moore, R. E., Smith, J. A., Seidman, J. G. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
12. Hattori, M. & Sasaki, Y. (1986) *Anal. Biochem.* **152**, 232–238.
13. Smith, S. (1981) *Methods Enzymol.* **71**, 188–200.
14. Randhawa, Z. I. & Smith, S. (1987) *Biochemistry* **26**, 1365–1373.
15. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
16. Smith, S. & Abraham, S. (1970) *J. Biol. Chem.* **245**, 3209–3217.
17. Smith, S. & Stern, A. (1979) *Arch. Biochem. Biophys.* **197**, 379–387.
18. Smith, S. (1973) *Arch. Biochem. Biophys.* **156**, 751–758.
19. Poulose, A. J., Bonsall, R. F. & Kolattukudy, P. E. (1984) *Arch. Biochem. Biophys.* **230**, 117–128.
20. Mikkelsen, J., Hojrup, P., Rasmussen, M. M., Roepstorff, P. & Knudsen, J. (1985) *Biochem. J.* **227**, 981–985.
21. Stern, A., Sedgwick, B. & Smith, S. (1982) *J. Biol. Chem.* **257**, 799–803.
22. McCarthy, A. D., Aitken, A., Hardie, D. G., Santikarn, S. & Williams, D. H. (1983) *FEBS Lett.* **160**, 296–300.
23. Wierenga, R. K. & Hol, W. G. J. (1983) *Nature (London)* **302**, 842–844.
24. Poulose, A. J. & Kolattukudy, P. E. (1983) *Arch. Biochem. Biophys.* **220**, 652–656.
25. Hsu, R. Y. & Wagner, B. J. (1970) *Biochemistry* **9**, 245–251.
26. Smith, S. & Libertini, L. J. (1979) *Arch. Biochem. Biophys.* **196**, 88–92.