Rat mammary gland fatty acid synthase: localization of the constituent domains and two functional polyadenylation/termination signals in the cDNA

Michael Schweizer*, Kenji Takabayashi†, Thomas Laux‡, Karl-Friedrich Beck and Rosemarie Schreglmann

Institut für Mikrobiologie und Biochemie der Universität Erlangen-Nürnberg, Staudtstraße 5, D-8520 Erlangen, FRG

Received October 21, 1988; Revised and Accepted December 16, 1988 Accession no. X13415

ABSTRACT
The rat fatty acid synthase (FAS) is active only as a dimer, although the eight component functions are contained in a single polypeptide chain. Using mRNA from lactating rat mammary glands a cDNA expression library was established. With the overlapping immunologically positive clones we have an 8.9kb cDNA sequence for rat FAS. In the 3'-nontranslated region of the rat FAS cDNA we find a prototype polyadenylation/termination signal and 779 nucleotides upstream, a mutated one. Both of these polyadenylation/termination signals are used and give rise to two equally abundant mRNA species which are coordinately regulated. In the derived amino acid sequence we could locate six of the eight component functions; their order is NH$_2$- beta-ketoacyl synthase - acetyl/malonyl transferases - enoyl reductase - acyl carrier protein - thioesterase -COOH. Comparison of FAS from different sources shows that the primary sequence is conserved only for the active residues and the amino acids in their immediate vicinity.

Introduction
The activity of a multifunctional enzyme results from the interaction of several discrete active sites within the same polypeptide. In the case of the vertebrate fatty acid synthase (FAS) eight component functions are found on a single polypeptide, alpha [1,2]. With a molecular mass of 250 000, this is one of the largest multifunctional polypeptides known [3,4]. In the active enzyme the two polypeptides are found in a head-to-tail arrangement [5,6]. This symmetrical structure for the vertebrate FAS permits the juxtaposition of the two intrinsic thiol groups in the enzyme, viz. the active cysteine thiol group of the beta- ketoacyl synthase and the pantetheine thiol group of the acyl carrier protein. The former and the latter thiol groups bind the fatty acyl and the incoming
malonyl residues, the substrates for the condensing reaction, respectively. Furthermore, a dimeric molecule is essential for the beta-ketoacyl reductase [7].

Limited proteolysis of native chicken and rabbit FAS clearly shows that three fragments were obtained, each containing one or more of the partial FAS functions [1,8, 9,10,11]. This elegant approach permits a relative ordering of the domains within each proteolytic fragment. Since the partial amino acid sequences immediately surrounding the active substrate binding sites are known for most of the component FAS functions from many systems [1,2], it should be feasible to determine within the primary structure of FAS as derived from the cDNA sequence the exact linear positions of the substrate active sites.

As a source of mRNA for the cDNA expression library we used lactating rat mammary glands since this tissue has an elevated FAS activity [12,13,14], which could be due to an increased mRNA content. Naturally we were interested in obtaining long cDNA clones since the FAS gene must be encoded by a mRNA with a minimum length of 6kb. Using the isolated FAS cDNA clones we could identify two FAS mRNA species in both mammary gland and liver. Sequencing of the cDNA revealed that two polyadenylation/termination sites are the reason for these two FAS mRNA species. We also present evidence that the two rat FAS mRNA species, in addition to being equally abundant, are coordinately regulated. Most importantly, the sequence data enabled us to locate six of the eight active centres of FAS within the derived amino acid sequence. We comment on the significance of the homology observed in and around the active centres of the FAS component activities in several different organisms. Since the rat FAS cDNA and the two genes encoding yeast FAS have been sequenced in their entirety [15,16,17,18, 19] it is possible to compare these two organisms over longer sequence stretches.

MATERIALS AND METHODS

Biological materials - 71-18 (F' lacI97 deltaM15 pro delta (lac pro)) was the recipient for M13 phage transfection [20]. BNN97 (=BNN93 (lambda gt11) hsdR' hsdM' supE thr leu thi...
lacY1 tonA21) was used for antibody purification and Y1090 (delta lacU169 proA' delta lon araD139 strA supF (trpC22::Tn10) hsdR'hsdM' (pMC9 = pBR 322-lacI') (Promega, Heidelberg) was used as the bacterial lawn when screening the rat cDNA library [21].

Sprague-Dawley male rats weighing 200 - 300 grammes were the source of liver tissue. Mammary gland tissue was obtained from four Sprague-Dawley lactating female rats (Ivanovas, Kissleg). Livers were obtained either from rats which had been starved for two days or starved for two days followed by refeeding a fat-free carbohydrate rich diet (Altromin, Lage) for at least 16 hours [14].

RNA preparation and characterization - RNA was isolated from lactating rat mammary gland and liver tissue using guanidinium hydrochloride in combination with successive ethanol precipitations [22]. The purity and quality of the RNA was checked by its absorption spectrum between 220 and 300 nm and its ability to stimulate in vitro translation. Poly (A)-rich RNA was obtained using either "messenger activated paper" (instruction manual, ORGENICS Ltd., Yavne, Israel) [23] or oligo (dT)-cellulose column chromatography (Paesel, Frankfurt). Northern analysis and size determination of the FAS mRNA were performed by electrophoresis of glyoxal/DMSO denatured RNA in 6.6% formaldehyde/ 0.8% agarose gels [24, 25]. Radioactively labelled fragments of lambda Hind III were used as standards [26].

Synthesis and cloning of cDNA - cDNA was synthesized according to the method of Gubler and Hoffman [27] with an Amersham cDNA kit (Braunschweig) and a cDNA library of about 800,000 plaque forming units in Y 1090 (hsdR' hsdM') [21] was constructed.

Screening for FAS cDNA clones - The unamplified library was screened with a polyclonal antiserum against rat FAS isolated from rat liver. FAS positive candidates as determined histochemically with horseradish-peroxidase-conjugate second antibody (Bio-Rad Laboratories, München, instruction manual) were subjected to rescreening and tested for lack of cross hybridization to alpha-2-inhibitor3 [28]. The isolation of human FAS and the corresponding antiserum will be described elsewhere.

DNA manipulations - DNA from the positive hybrid plaques was isolated using DEAE cellulose (DE 52) chromatography (Whatman, München)[29]. For sequence analysis the cDNA was subcloned into appropriate M13 vectors [30]. All manipulations with M13 derivatives were carried out as described previously [17]. cDNA sequencing was performed using either the universal primer for M13 (Amersham, Braunschweig) or sequence derived 17-mer oligonucleotide primers prepared with the Applied Biosystem synthesizer type 381A. Generally, the primers were used without further purification. The sequence reactions were carried out initially with the Klenow fragment of DNA polymerase I (Pharmacia, Freiburg) [31] and later with the modified bacteriophage T7 DNA polymerase [32] sequencing kit (USB-C) obtained through Renner, Heidelberg. Sequence reaction products were separated on wedge-shaped 5% polyacrylamide/7.5M urea gels [33]. The DNA sequence data generated were

569
analysed on a VAX 11-780 terminal using the UWGCG sequence analysis software package, version 5 (1987) [34]. For homology searches we used the GenBank, EMBL and NBRF Nucleic, NBRF Protein Sequence Data Banks (Release 50, 11, 30, 12; 1987). Radioactive probes for Southern hybridization (nitrocellulose BA 85, Schleicher and Schüll, Dassel) were labelled by nick translation [35] or 5'-endlabelling [36]. Restriction endonucleases and DNA modifying enzymes were purchased from Boehringer, Mannheim; Pharmacia, Freiburg or Gibco BRL, Eggenstein and used according to the manufacturers' instructions.

RESULTS

Stepwise isolation of FAS cDNA with a length of 8.9kb

The tissue specific increase of FAS activity in rat mammary glands during lactation [14] may be due at least in part to an increased transcription of the FAS mRNA. Therefore, for the express purpose of isolating FAS cDNA clones, we constructed an oligo d(T)-primed cDNA library, using poly (A)-rich RNA prepared from lactating rat mammary glands. The cDNA synthesized was cloned into the expression vector lambda gt11. The presence of a beta-galactosidase/FAS fusion protein should give rise to a positive immune response with our polyclonal rat FAS antiserum.

Of the 300 000 recombinant plaque forming units 21 reacted with our rat FAS antiserum. Three successive rounds of rescreening confirmed the first result. About 70% of the clones showed a significant reaction with human FAS antiserum [37]. Hybridization characteristics of six of these clones, together with those of three others which reacted only with rat FAS antiserum, are presented in Table 1.

Interestingly, the clones pcRFAS 44, pcRFAS 34 and pcRFAS 64 with the strongest immune reactions have the smallest inserts and according to their hybridization patterns are contained within the 1.8kb Eco RI/Xho I fragment of pcRFAS 40 (cf.Figure 1). pcRFAS 40, 2.6kb long, hybridized with more than half of the 21 clones (data not shown), whereas pcRFAS 22, roughly 2.0kb, hybridized with itself, pcRFAS 15, pcRFAS 43 and albeit weakly, with pcRFAS 66. The lengths and positions of these clones relative to each other are shown in Figure 1. The most gratifying result of the crosshybridization
was the fact that clone pcRFAS 43 hybridized with both pcRFAS 40 and pcRFAS 22. pcRFAS 43 has an exceptionally large insert of at least 6.0kb but showed a weak immune response in the immunological screening. Since the insert of pcRFAS 40 was so abundant in the library, it is reasonable to assume that the 3'-end of the mRNA - the startpoint of the cDNA synthesis - is located therein. Taken together, the three cDNA clones, pcRFAS 43, pcRFAS 40 and pcRFAS 22, represent about 6.4kb of cDNA and in theory are capable of encoding a polypeptide of 250 000, the size of the octafunctional subunit of vertebrate FAS.

However in Northern analyses with total RNA from mammary gland and liver, FAS mRNA apparently has a length of at least 8.8kb (see Figure 2a) [38]. It was therefore necessary to extend the cDNA in the presumptive 5'-direction. First of all pcRFAS 15, pcRFAS 66 and pcRFAS 30 were further characterized. Clone pcRFAS 15 turned out not to extend beyond the end of the 5'-end of pcRFAS 22. On the other hand, pcRFAS 66 provided an 1.0kb extension of the cDNA in the postulated 5'-direction. The position of clone pcRFAS 30 within pcRFAS 66 is such that it does not cross-hybridize with pcRFAS 22, although it does extend further in the 5'-direction. The intrinsic Eco RI restriction site in the cDNA is the 3'-boundary of pcRFAS 30 and provides the link with pcRFAS 22 (see Figure 1 and below).

To improve the chance of having a full length cDNA, the
library was rechecked by hybridization with pcRFAS 66. The one positive clone, pcRFAS 6, with an insert of 1.6kb brought the cDNA up to a total length of 8.9kb.

The two FAS mRNA are regulated coordinately

The cDNA clone pcRFAS 40 was used to estimate the size of the mRNA species corresponding to the cDNA. Much to our surprise, in lactating mammary gland and total rodent liver RNA pcRFAS 40 hybridized to two mRNA species, both of which are larger than the postulated length of the FAS mRNA. The upper band was estimated to be 9500 nucleotides and the lower is apparently 8800 nucleotides long (Figure 2a, lanes 2-5). All of the cDNA clones in Figure 1 hybridized to the two mRNA species (data not shown).

The two signals in the Northern analyses with total
Fig. 2 Northern blot analysis of FAS mRNA. (a) 10μg of total RNA (lanes 1 and 3) and 2.5μg poly (A)-rich RNA (lane 2) prepared from lactating rat mammary glands, 10μg of total RNA from the livers of rats which had been starved and then refed (lane 4) or only starved (lane 5). In lanes 2 to 5 the probe was pcRFAS 40 and in lane 1 the Xba I/Sst I restriction fragment (X/S) of pcRFAS 43. (b) The relative positions of the hybridization probes used to obtain the Northern blot in (a) are indicated as black bars. The Xba I restriction site arose from the last five nucleotides of the rat FAS cDNA and the first A of the poly (A) tail.

Liver RNA are observed easily in the liver of an animal fed a special diet. The increase in the amount of FAS mRNA is brought about by refeeding starved animals a fat-free carbohydrate rich diet for 16 hours before sacrificing. In the livers taken from animals which have been starved for 16 hours and not refed, only a very faint hybridization signal with pcRFAS 40 is seen after prolonged exposure (Figure 2a, lanes 4 and 5). On the basis of these results we conclude that the two species of FAS mRNA are equally abundant and apparently coordinately regulated in the liver.

**Sequencing of the FAS cDNA**

The cDNA clones were sequenced over their combined length according to the strategy shown in Figure 1. In summary, it may be said that of the 8936 nucleotides
sequenced, the coding region contains 7293 nucleotides and represents a polypeptide of 2431 amino acids. From the first methionine at position 102 (Figure 3) to the stop codon TAG (position 7395-7397) the cDNA encodes a protein with a molecular mass of 264 124. This value is in agreement with 250 000 as estimated by SDS gel electrophoresis [3]. Since upstream ATG codons are rare in vertebrate mRNA's [39] we postulate that the ATG at nucleotide 102 of the cDNA sequence is the authentic ATG codon for FAS. Supporting evidence for the choice of this methionine as the first amino acid is provided by the sequence immediately surrounding this ATG codon. With the A at position -3 and G at +4 and the lack of T residues in the preceding 12 nucleotides this region may well function as an eucaryotic ribosome binding site and translation initiator.

Clone pcRFAS 43 was sequenced from its 3'-end over a length of 1.6kb. In the sequenced overlapping regions of pcRFAS 40 and pcRFAS 43 the sequences of both clones were found to be identical. Approximately 1.8kb of pcRFAS 43 which overlap with the 5'-end of pcRFAS 40 and the 3'-part of pcRFAS 22 were also sequenced. In order to confirm the intrinsic Eco RI restriction site in the cDNA 450bp of pcRFAS 66 within which the Eco RI site is located were sequenced. From the clones pcRFAS 30 and pcRFAS 66 a length of 560bp overlaps with pcRFAS 6. In this region we observed two nucleotide differences; one silent at position 1232 (Figure 3) changing the valine codon from GTT in pcRFAS 6 to GTG in pcRFAS 30. The second change at position 1108 changes the codon CAG for glutamine in pcRFAS 6 to CCG for arginine in pcRFAS 30.

Mapping of the activities for FAS

Within the coding region we were able to locate the substrate binding sites for six component functions of FAS. The positions of the active centres are indicated by underlining and an asterisk in the derived amino acid sequence (Figure 3).
The locations of the substrate binding sites were deduced by comparison with the known amino acid sequences of the peptides immediately surrounding the active amino acid residues in each of these partial FAS activities as known for a number of eucaryotes and procaryotes [1,2]. The first substrate binding site is the beta-ketoacyl synthase (KS), its active residue is the 87th amino acid of the FAS polypeptide (nts 360-362). 420 amino acids further on we located the active serine residue (nts 1620-1622) of the doubly active site for the acetyl/malonyl transferases (AMT). At amino acid 1621 we find the consensus sequence - GSA - corresponding to the active site for the enoyl reductase (ER, nts 4962-4964). The acyl carrier protein (ACP) and thioesterase I (TE) substrate binding sites are located at positions 2077 (nts 6330-6332) and 2228 (nts 6783-6785), respectively.

Two functional polyadenylation/termination signals

By dissecting the 3′-nontranslated end of the FAS mRNA into two parts; viz. probe 1, a restriction fragment common to pcRFAS 43 and pcRFAS 40 and probe 2, the region of pcRFAS 43 which extends beyond clone pcRFAS 40 we obtained hybridization probes specific for each of the RNA species (Figure 2b). Probe 1, like pcRFAS 40, hybridizes with both mRNA species (Figure 2a, lanes 2-5) whereas probe 2 recognizes only the larger of the two (Figure 2a, lane 1).

From the cDNA sequence ( Figure 3) we know that the 3′-nontranslated region contains 1542 nucleotides. Within this region there are two polyadenylation/termination signals, the
prototype polyadenylation signal "AATAAA" and upstream of this, a mutated form "AATTAA", both are underlined in Figure 3. The two signals are separated by 779 nucleotides which reflects the difference in the sizes of the two FAS mRNA species (Figure 2a and Figure 3). The poly (A)-tail of pcRFAS 40 is at nucleotide 8152 which is exactly 17 nucleotides after the mutated polyadenylation signal "AATTAA". The prototype "AATAAA" polyadenylation signal lies 15 nucleotides before the poly (A)-stretch of pcRFAS 43. The mutated polyadenylation signal was found in the course of sequencing clone pcRFAS 40; sequence data from the corresponding region of pcRFAS 43 revealed the presence of the mutated signal, apparently unused, between nucleotides 8130 and 8135 (Figure 3).

DISCUSSION

CDNA clones coding for rat FAS were isolated from a lambda gt11 library based on RNA from lactating mammary gland tissue. This tissue is actively engaged in the synthesis of most components of milk, including lactose and fatty acids and the latter may account for up to 10% of the soluble protein [13,14,40]. The special care exercised during mRNA isolation and construction of the cDNA library was rewarded by our success in obtaining a cDNA clone of 6.0kb (pcRFAS 43). The length of the cDNA was extended by about 3000 nucleotides. According to SDS gel electrophoresis each of the two component polypeptides of rat FAS has a relative molecular mass of 250 000 [3]. Therefore our cDNA is long enough to encode this polypeptide. Indeed, 102 nucleotides into the cDNA sequence we find an ATG codon in a sequence environment typical for the start codon of a eucaryotic gene [39].

Clones pcRFAS 22, pcRFAS 66, pcRFAS 30 and pcRFAS 6 have apparently arisen by fortuitous priming of the reverse transcriptase. The priming positions could be due to folding of the FAS mRNA thus bringing the poly (A) region in close proximity to other regions of the molecule [41]. The Eco RI site at position 2340 is unlikely to be the startpoint of clone pcRFAS 30, probably pcRFAS 30 is the truncated part of a longer cDNA clone which has arisen following Eco RI
restriction of the Eco RI linkered cDNA immediately prior to in vitro packaging. Further evidence that the unique Eco RI restriction site at least in some cases escaped methylation is provided by the relatively large number of cDNA clones in the library which have identical restriction patterns to clone pcRFAS 22. Obviously, pcRFAS 66 is the result of a successfully methylated cDNA molecule with yet another primer-independent initiation (Figure 1). The startpoint of pcRFAS 40 coincides with the poly (A) tail attached following termination of transcription at the mutated polyadenylation/termination signal (Figure 1 and 3).

The longest clone pcRFAS 43 arose from cDNA synthesis starting at the poly (A) tail resulting from the prototype polyadenylation/termination signal between nucleotides 8916 and 8921. Evidence that this mutated signal in FAS mRNA is not redundant is the existence of two corresponding FAS mRNA species. The measured sizes of the two FAS mRNA species may well be an overestimate since the nucleotide sequence suggests that the FAS mRNA species are in fact shorter. This is not the only example of FAS being encoded by two mRNA species [42,43].

The two FAS mRNA species could in fact arise from different FAS genes. The longer gene would have a 3'-nontranslated region with the mutated and prototype polyadenylation/termination signals and would give rise to the 9.5kb transcript; i.e. the mutated polyadenylation/termination signal is not used. The shorter gene would have a 3'-nontranslated region containing only the mutated signal since in the absence of the prototype polyadenylation/termination signal the otherwise cryptic signal is unmasked. Suggestive for the presence of two FAS genes are the nucleotide exchanges at 1108 and 1232 (Figure 3), although these could be explained by somatic mutation or inaccuracy of the reverse transcriptase. However, it would appear from our genomic Southern blotting that both transcripts arise from a single FAS gene (T. Laux, unpublished data).

Previous results from partial proteolysis of the native chicken enzyme [9] placed the beta-ketoacyl synthase in the same proteolytic fragment as the acetyl/malonyl transferases.
Our results have unequivocally established the order NH₂-KS-AMT- in rat FAS. Further along the cDNA, we found the enoyl reductase active site at nucleotides 4962-4964 (aa 1620, Figure 3). For comparative purposes we used the peptide sequence for the enoyl reductase from goose and S.cerevisiae [44,17].

Information concerning the amino acid sequence around the serine to which the 4'-phosphopantetheine arm is attached to the acyl carrier protein (ACP), has been determined in a number of eucaryotes and procaryotes. We have assembled the data in Figure 4. Our rat ACP sequence was identified by its high homology to the 64 known residues of the rabbit ACP peptide [45]. Unexpectedly of the 13 residues purported to come from the rat ACP [46] only five agree with our sequence and that of rabbit ACP. The degree of homology for the individual substrate binding sites for all organisms compared is maximal immediately surrounding the active serine residue. Only when sequences of vertebrate FAS are compared does the homology extend over a long region as particularly well exemplified by the comparison of ACP from several different sources. In the comparison of ACP from rat and rabbit FAS there is 100% homology extending over 36 amino acids (Figure 4). The homology, however, drops to roughly 20% when the ACP of Saccharomyces cerevisiae derived from the cloned FAS2 gene [16,19] is compared to the corresponding rat cDNA sequence. Two procaryote ACP (E.coli [47] and Rhodobacter sphaeroides [48]) have a strong amino acid homology with each other and, in the region around the active serine, with the rodent ACP. Likewise the homology of spinach ACP peptide with the vertebrate, fungal and bacterial peptides is restricted to the region immediately surrounding the serine binding site. Interestingly, all ACP peptides shown here have a leucine 20 residues amino-terminal to the active serine.

Since no corresponding data is available, the beta-ketoacyl reductase and the dehydratase active centres could not be located using the homology approach. There is, however, sufficient space in the derived amino acid sequence for the domains at the positions suggested by limited proteolysis of
Fig. 4  Sequence homology within the ACP peptides of procaryotes and eucaryotes. The amino acid sequences common to the rat FAS ACP and the other sequences are boxed. The serine (S) marked with an arrowhead is the attachment site for the 4'-phosphopanthetheine. The consensus sequence is given in the lowest line. a) [45]; b) [50]; c) [51]; d) [52]; e) [53]; f) [54]; g) [19]; h) [47]; i) [48].
Fig. 5  Schematic representation of the reaction sequence of the alpha dimer of rat FAS and the linear arrangement of its active centres. The movements of the enzyme bound substrates between the two subunits (I, II) in the upper half of the dimer are indicated by arrows. The zig zag line represents the prosthetic 4'-phosphopantetheine SH group of the ACP. The relative positions of the component functions within the alpha polypeptide indicated by upper case letters are definitive; lower case letters (deh and kr) indicate the tentative positions for these activities.

the native enzyme [1,2], as indicated in Figure 5, which is a schematic representation of the linear arrangement of the domains in the primary structure of the rat FAS polypeptide as derived from the cDNA sequence and the reaction sequence within the alpha dimer. We suggest a location for the dehydratase between the acetyl/malonyl transferases and the
enoyl reductase domains and for the beta-ketoacyl reductase domain between the enoyl reductase and the acyl carrier protein. This agrees with the finding that a 107 000 dalton proteolytic fragment of chicken FAS has been shown to carry the beta-ketoacyl reductase, enoyl reductase, dehydratase and ACP functions [9]. The locations of the ACP and thioesterase I domains have been reported previously [49].

We have clearly shown that the homology within analogous domains in distantly related FAS systems is restricted to the regions immediately surrounding the substrate binding sites. This strengthens the argument that it is the architecture of a multifunctional enzyme which determines its specificity. We now know that the linear arrangement of the rat domains in the FAS enzyme does not correspond to the FAS reaction sequence. Chemically it has been shown that the smallest functional unit for rat FAS is a dimer with a head-to-tail arrangement of the component polypeptides [5]. In Figure 5 we have sketched out a possible pathway allowing fatty acid synthesis within the alpha₂ dimer. Obviously, the acyl substrates are transferred inter- and intramolecularly between the active centres. Only when the three-dimensional structure of rat and other FAS have been elucidated, will it be clear how the differences in primary structure are compensated.

Acknowledgments
Part of the results have been orally presented at the 18th FEBS Meeting, Ljubljana, Yugoslavia, June 1987. We thank G. Müller and J. Beck for synthesizing the oligonucleotides and B. Rautenstraub for helping with computer searches. The work was supported by a grant from the Deutsche Forschungsgemeinschaft to M.S.. The rat fatty acid synthase cDNA sequence has been assigned the accession number X13415 in the EMBL Data Library.

Present addresses: †Department of Biotechnology, Ciba-Geigy AG, Basel, Switzerland and §Department of Biology, University of California, Los Angeles, CA 90024, USA

*To whom correspondence should be addressed.

Abbreviations: aa, amino acids (they are given in the single letter code); bp, base pairs; kb, 1000bp; nts, nucleotides.
References