Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6618–6621, September 1986
Medical Sciences

Cloning and DNA sequence analysis of the cDNA for the precursor of the β chain of bovine follicle stimulating hormone

(follitropin)

F. S. ESCH*†, A. J. MASON‡, K. COOKSEY*, M. MERCADO*, AND S. SHIMASAKI*

*Laboratories for Neuroendocrinology, The Salk Institute for Biological Studies, La Jolla, CA 92037; and ‡Department of Developmental Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080

Communicated by Roger Guillemin, May 13, 1986

ABSTRACT **Follicle stimulating hormone (FSH) plays essential roles in the maintenance and development of oocytes and spermatozoa in normal reproductive physiology. FSH possesses two subunits, α and β, the latter being responsible for FSH biological specificity. We have cloned and sequenced the cDNA encoding the FSH β chain from a bovine anterior pituitary cDNA library. The mature molecule is 109 amino acids long and is preceded by a 20-amino acid putative signal peptide. RNA gel blot analysis of bovine pituitary RNA shows that the mRNA encoding β chain of FSH is ≈1.7 kilobases in length.**

Follicle stimulating hormone (FSH) is a glycoprotein hormone secreted by the anterior pituitary (1, 2). It induces growth of the seminiferous tubules, maintains spermatogenesis, and is responsible for the growth of the follicle that, in turn, produces estrogen and stimulates endometrial growth in the uterus. FSH is one member of a family of four structurally related glycoprotein hormones that includes chorionic gonadotropin (CG), luteinizing hormone (LH), and thyroid stimulating hormone (TSH). These polypeptides consist of two dissimilar subunits that are noncovalently linked together. The α subunit is common to all four hormones within a single species (3), while the β subunits are unique and confer biological specificity. In as much as the β-subunit gene sequences for human CG and LH (4, 5), rat LH and TSH (6–8), and mouse and bovine TSH (9, 10) have been characterized, it is surprising that the DNA sequence encoding the β chain of FSH has yet to be described from any species. The amino acid sequences for the ovine, human, equine, and porcine β chains of FSH are known, and we have used highly conserved regions of amino acid sequence within these structures to construct two long synthetic oligonucleotide probes. These probes were used to screen a bovine anterior pituitary cDNA library, and the resultant positive clones were characterized by DNA sequence analysis.

MATERIALS AND METHODS

Bovine anterior pituitary polyadenylylated mRNA was prepared from freshly frozen tissues (11) and used to construct an oligo(dT)-primed cDNA library of 6×10^6 clones in which double-stranded cDNA was initially ligated to an adaptor containing *Eco*RI and *Sal* I sites and subsequently ligated to the λgt10 vector (12). The two long synthetic oligonucleotide probes in Fig. 1 were synthesized with an Applied Biosystems Model 380A DNA synthesizer and purified by reverse phase liquid chromatography (15). These probes were 5′ end-labeled using polynucleotide kinase and [^{32}P]ATP (16) and were used to screen 250,000 clones. Prehybridizations (2 hr) and hybridizations (≥12 hr) of blotted nitrocellulose filters (10,000 plaque-forming units per filter) with ^{32}P-labeled probes were carried out

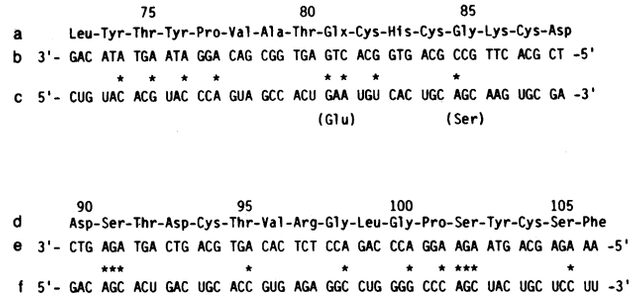

FIG. 1. Synthetic oligonucleotide probes for the isolation of bovine β chain of FSH coding sequences. The amino acid sequences of ovine, human, porcine, and equine β chain of FSH(73–88) (*a*) and β chain of FSH(90–106) (*d*) (the numbering system begins with the first amino acid of human β chain of FSH) were used to design long synthetic oligonucleotide probes (*b* and *e*, respectively) whose codon selections were based upon mammalian codon frequency usage tables (13) and the substitution of thymidine for cytosine (14) in "wobble" positions to preserve approximately 50/50, G+C/A+T ratios in the probes. Mismatches between the probe sequences and the actual bovine FSH β-chain mRNA sequence (*c* and *f*) are marked (*).

in duplicate in 2× hybridization buffer [10× SSC (1.5 M NaCl/0.15 M sodium citrate, pH 7.0)/10 mM $NaPO_4$, pH 7.2/10× Denhardt's solution (0.2% Ficoll/0.2% polyvinylpyrrolidone/0.2% bovine serum albumin)/herring sperm DNA at 0.2 mg/ml/*Escherichia coli* RNA at 20 μg/ml/0.2% $NaDodSO_4$] diluted to 1× concentration with distilled H_2O and formamide to give a final 40% (vol/vol) formamide concentration (17). These incubations were performed at 37°C using 1–3 μCi of ^{32}P-labeled probe (1 Ci = 37 GBq) and 5–10 ml of hybridization buffer per filter in 15-cm diameter glass crystallizing dishes. The filters were then washed in 0.2× SSC/0.1% $NaDodSO_4$ for 30 min at 40°C and exposed to x-ray film with intensifying screens at −80°C. Positive clones were screened again to isolate pure plaques. The cDNA inserts were excised with *Sal* I and inserted into the M13mp18 phage vector for further manipulations including DNA sequence analyses using the dideoxy chain termination method (18). Both strands of positive clones were completely sequenced to ensure sequence accuracy. RNA gel blot (19, 20) analyses of bovine pituitary RNA were done using nick-translated cDNA inserts (specific activity, 10^8 cpm/μg) from the positive clones (16).

RESULTS

The long synthetic oligonucleotide probes in Fig. 1 were designed around the most highly conserved regions of the FSH β-chain amino acid sequence derived from four different

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.

Abbreviations: FSH, follicle stimulating hormone; CG, chorionic gonadotropin; LH, luteinizing hormone; TSH, thyroid stimulating hormone; bp, base pair(s); kb, kilobase(s).

†To whom correspondence should be addressed.

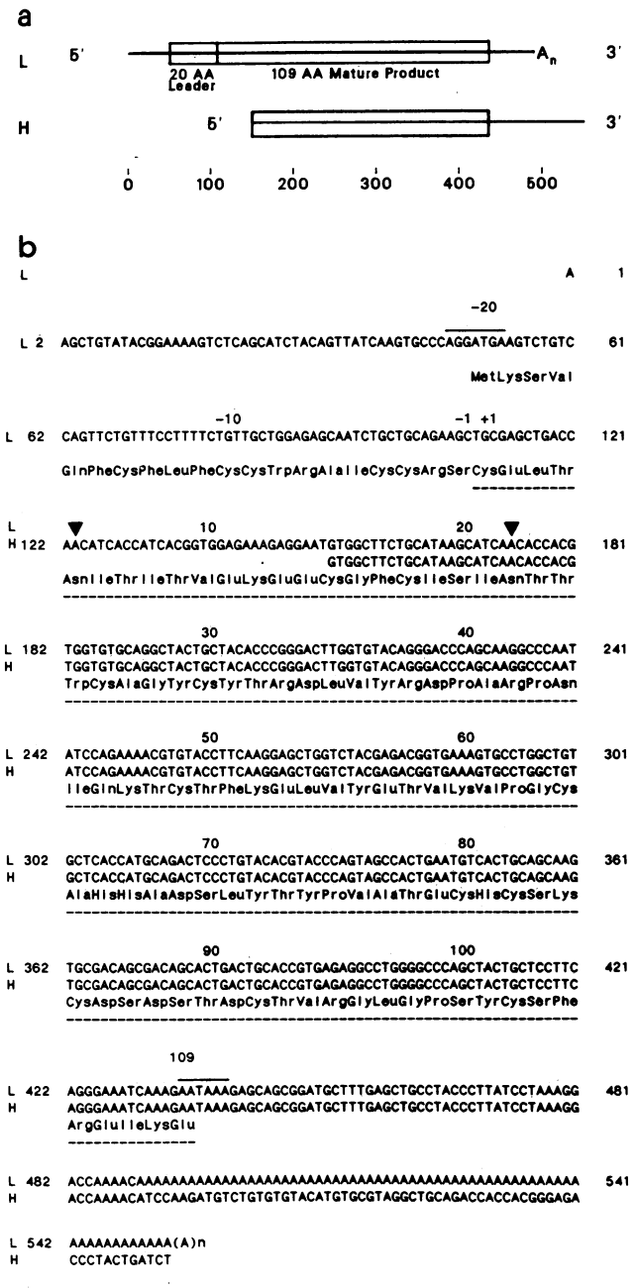

FIG. 2. Amino acid and nucleotide sequence of bovine FSH β-chain cDNA clones H and L. (*a*) Schematic representation of bovine FSH β-chain clones. Coding sequences are boxed while untranslated sequences are represented with a line. The scale is in nucleotides from the 5′ end of the L clone. (*b*) Nucleotide and predicted amino acid sequence of the bovine FSH β-chain clones. Nucleotides are numbered at the ends of lines, and amino acids are numbered throughout. The coding sequence for the putative signal peptide (−20 to −1) is unmarked, the sequence for the mature hormone is underlined, and the potential N-glycosylation sites are indicated by (▼). The PuXXATGPu translation initiation consensus sequence (21) and the AATAAA polyadenylylation signal (22) are overlined. Sequence differences between the L and H clones were found only in the 3′ terminus (bases 491–553).

species. The chosen codon selections (13, 17) resulted in only 8 base mismatches in the 47-mer (probe 73–88) and 11 base mismatches in the 50-mer (probe 90–106) when compared to the bovine mRNA sequence. Only two positive clones were found when 250,000 clones were screened with these probes. DNA sequence analyses ultimately confirmed that both positives did encode the FSH β-chain sequence. A subsequent screening of 820,000 clones with a nick-translated

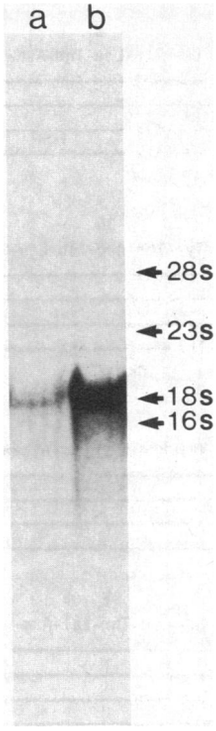

FIG. 3. RNA gel blot analysis of bovine pituitary RNA for β chain of FSH. Total RNA (10 μg, lane a) and polyadenylylated RNA (5 μg, lane b) were electrophoresed in a 1.4% agarose/formaldehyde (19) gel and transferred to nitrocellulose (20). The H clone cDNA insert was nick-translated and used as a hybridization probe. The filter was washed in 0.1× SSC/0.1% $NaDodSO_4$ for 30 min at 50°C and radioautographed with an intensifying screen at −80°C. The positions of the 16s, 23s, 18s, and 28s rRNA standards are indicated by arrows. Identical results were obtained using a nick-translated 257-bp *Ava* II (nucleotides 224–481) L clone fragment as a hybridization probe (data not shown).

FSH β-chain cDNA insert yielded only six positive clones.

The full cDNA and deduced protein sequences for the two positive clones are shown in Fig. 2. The 490-base-pair (bp) clone L contains a 49-bp 5′-untranslated region and a probable ATG initiation codon at position 50 whose surrounding sequence is consistent with that of most eukaryotic initiation codons (21). Following this is a putative 20-amino acid signal peptide (23, 24) and the 109-residue mature hormone; a TAA termination codon immediately follows the carboxyl-terminal glutamic acid of the mature sequence. The 401-bp clone H is identical to clone L except that (*i*) the 5′ sequence of clone H extends only to amino acid residue 16 (cysteine) of the mature hormone, (*ii*) the 3′-terminal sequences (bases 491–553) of the two clones are different, and (*iii*) the H clone lacks a poly(A) tail and apparently utilizes a different (and unseen) polyadenylylation signal (22) than the L clone.

RNA gel blot analysis of bovine pituitary RNA (Fig. 3) indicates that the full length mRNA encoding the β chain of FSH is ≈1.7 kilobases (kb) long. Hence, our longest clone (L) with its poly(A) tail is ≈1.1 kb short of a full length transcript, but does contain the entire coding sequence of the polypeptide.

DISCUSSION

Fig. 4 compares the structures of β chain of FSH from five different species. These highly conserved sequences show the most variability in their carboxyl termini. The absolute assignment of the amino-terminal amino acid of mature bovine β chain of FSH is not possible without NH_2-terminal protein sequence analysis data from the intact protein. However, signal peptides generally contain a central region

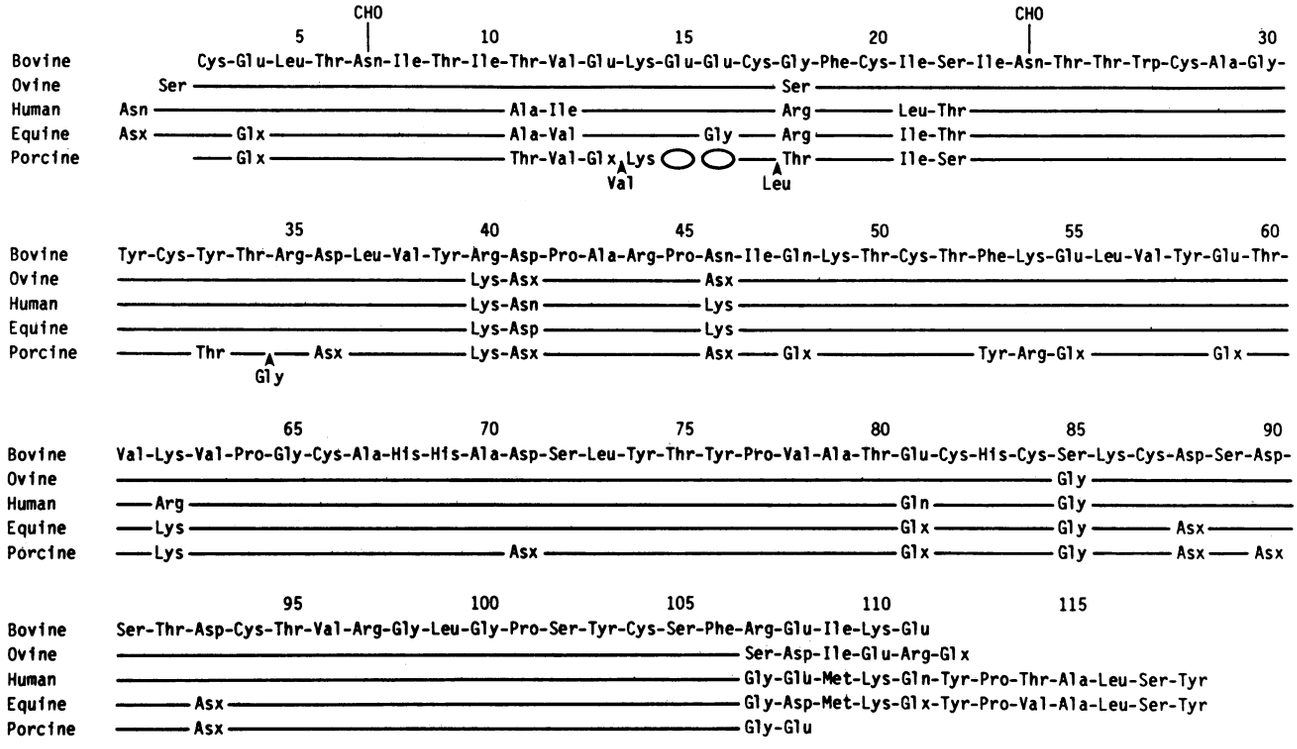

FIG. 4. A comparison of the mature bovine β chain of FSH sequence with previously described β chain of FSH sequences from four different species. The ovine (25), human (26), equine (27), and porcine (28) sequences were all obtained from protein sequence analyses data. The porcine sequence is aligned with gaps (○) and inserts (▲) to show maximum homology. Putative sites of N-glycosylation are indicated by (-CHO). The numbering system begins at asparagine-1 for human FSH.

rich in hydrophobic amino acids with large side chains and terminate in a residue with a small neutral side chain (e.g., alanine, glycine, or serine) (23, 24); the utilization of serine (−1) as the last amino acid in the leader sequence fulfills these requirements and is consistent with the identification of 12 homologous cysteines in the mature β subunits of virtually all the characterized members of this glycoprotein hormone family (i.e., the β chain of CG, TSH, LH, and FSH). While signal peptides typically contain zero to two cysteines, it is interesting to note that the β chain of FSH signal sequence is unusually cysteine-rich, containing five cysteines.

The bovine β chain of FSH mRNA has several interesting features. (*i*) It is considerably larger (1.7 kb) than mRNAs encoding rat TSH β chain (8), LH β chain (7) (0.7 kb), or human CG β-chain (4) (0.9 kb). (*ii*) The H and L clones may reflect the presence of two RNA species for β chain of FSH. The H clone lacks a poly(A) tail and probably resulted from random priming during the reverse transcriptase step of cDNA synthesis. If the mRNA represented by the H clone possesses a long 3′-untranslated region, then the relatively high abundance of the 1.7-kb β chain of FSH mRNA seen on RNA gel blot analysis (Fig. 3) and the low frequency of β chain of FSH clones in the library could be accounted for by the difficulty of obtaining full-length reverse transcripts of large mRNAs and by the relative rarity of random priming events during first-strand cDNA synthesis. Following this line of reasoning, if the presumably shorter L clone represents only a small fraction of the total β chain of FSH mRNA population, then this low molecular weight mRNA would be difficult to detect on the RNA gel blot analysis of Fig. 3 and only the major β chain of FSH mRNA represented by the H clone would be seen. (*iii*) The AATAAA sequence in the 3′-untranslated region, like that for human CG β chain (4), serves the following multiple roles: it encodes the last amino acid (GAA for glutamic acid) and the termination codon (TAA) as well as serving as the polyadenylylation signal (22) for the L clone. Polyadenylylation signals are typically approximately 20 nucleotides away from their respective poly(A) tracts (22); it is interesting to note that this spacing for the L clone is unusually long, involving 49 bases. (*iv*) The AATAAA sequence in the H clone does not signal the initiation of a nearby poly(A) tract in that clone; these sequence variations in the 3′-untranslated regions of the H and L clones may reflect the presence of multiple genes encoding this protein or alternatively multiple polyadenylylation signals (29, 30) in a single β chain of FSH gene. Southern analyses of bovine genomic DNA should delineate whether or not multiple genes exist for β chain of FSH.

We thank F. Castillo and Luc Denoroy for synthesis and purification of synthetic oligonucleotides, R. Klepper for tissue collections, C. Saris and G. Weinmaster for helpful discussion, and W. Collins and B. Gayer for secretarial assistance. This research was supported by program grants from the National Institutes of Health (HD-09690 and AM-18811) and the Robert J. and Helen C. Kleberg Foundation.

1. Fiddes, J. C. & Talmadge, K. (1984) *Recent Prog. Horm. Res.* **40**, 43–78.
2. Sairam, M. R. (1983) *Horm. Proteins Pept.* **11**, 1–79.
3. Fiddes, J. C. & Goodman, H. M. (1981) *J. Mol. Appl. Genet.* **1**, 3–18.
4. Fiddes, J. C. & Goodman, H. M. (1980) *Nature (London)* **286**, 684–687.
5. Talmadge, K., Boorstein, W. R. & Fiddes, J. C. (1983) *DNA* **2**, 281–289.
6. Tepper, M. A. & Roberts, J. L. (1984) *Endocrinology* **115**, 385–391.
7. Chin, W. W., Godine, J. E., Klein, D. R., Chang, A. S., Tan, L. K. & Habener, J. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4649–4653.
8. Chin, W. W., Muccini, J. A. & Shin, L. (1985) *Biochem. Biophys. Res. Commun.* **128**, 1152–1158.
9. Gurr, J. A., Catterall, J. F. & Kourides, I. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2122–2126.

10. Maurer, R. A., Croyle, M. L. & Donelson, J. E. (1984) *J. Biol. Chem.* **250,** 5024–5027.
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
12. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL Press, Oxford), pp. 49–78.
13. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9,** r43–r74.
14. Agarwal, K. L., Brunstedt, J. & Noyes, B. E. (1981) *J. Biol. Chem.* **256,** 1023–1028.
15. Applied Biosystems (1984) *User Bulletin* (Applied Biosystems, Foster City, CA), No. 13.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 122.
17. Ullrich, A., Berman, C. H., Dull, T. J., Gray, A. & Lee, J. M. (1984) *EMBO J.* **3,** 361–364.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. Dobner, P. R., Kawasaki, E. S., Yu, L. Y. & Bancroft, F. C. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2230–2234.
20. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
21. Kozak, M. (1981) *Nucleic Acids Res.* **9,** 5233–5252.
22. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
23. Habener, J. F., Rosenblatt, M., Kemper, B., Kronenberg, H. M., Rich, A. & Potts, J. T. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2616–2620.
24. Kriel, G. (1981) *Annu. Rev. Biochem.* **50,** 317–348.
25. Sairam, M. R., Seidah, N. G. & Chretien, M. (1981) *Biochem. J.* **197,** 541–552.
26. Saxena, B. B. & Rathnam, P. (1976) *J. Biol. Chem.* **251,** 993–1005.
27. Fujiki, Y., Rathnam, P. & Saxena, B. B. (1978) *J. Biol. Chem.* **253,** 5363–5368.
28. Closset, J., Maghuin-Rogister, G., Hennen, G. & Strosberg, A. D. (1978) *Eur. J. Biochem.* **86,** 115–120.
29. Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. (1980) *Cell* **22,** 361–370.
30. Tosi, M., Young, R. A., Hagenbuchle, O. & Schibler, U. (1981) *Nucleic Acids Res.* **9,** 2313–2323.