

## 猪生长激素cDNA的全序列分析

齐顺章 王辛中 周顺伍 贾锋 王华岩 夏立 李娟

(北京农业大学动物生物化学教研室, 北京)

本文报道了两个猪生长激素cDNA的全序列分析的结果。猪生长激素cDNA 是我们由猪垂体mRNA经过反向转录、克隆和筛选后获得的。文中还将这两个序列与Seeburg等(1983)报道的序列以及 Vize等(1987)报道的猪基因组生长激素基因的序列进行了比较和讨论。

**关键词** cDNA序列; 猪生长激素

1983年Seeburg等<sup>[1]</sup>首次发表了猪生长激素cDNA的全序列。然而他们的cDNA对成熟猪生长激素来说是全长的, 但对猪前生长激素来说则不够完整, 没有起始密码子, 更没有5'端的非编码区序列。为了研制猪基因工程生长激素, 我们由猪垂体mRNA经过反向转录、克隆和筛选, 获得了多个猪生长激素cDNA的克隆<sup>[2,3]</sup>。对其中两个进行酶切分析的结果表明, 它们在5'端和3'端都比Seeburg等发表的cDNA长很多。为了探讨猪生长激素cDNA的结构, 从而探讨其mRNA的结构, 我们对这两个cDNA进行了全序列分析。

### 材料和方法

#### (一) 材料

菌种*E. coli* JM107购自BRL; 猪生长激素cDNA (pGHcDNA) 本室制备; 噬菌体M<sub>13</sub>mp18, M<sub>13</sub>mp19 RFDNA 本室制备; Hind III、Sma I购自BRL; EcoR I、Acc I 及T4DNA连接酶购自华美公司; DNA聚合酶I (大片段) 本室制备;  $\alpha$ -<sup>32</sup>P-dCTP购自Amersham公司; 四种脱氧核苷酸和四种双脱氧核苷酸购自BRL; 通用引物购自BioLab; 1号探针由吴瑞先

生提供; 2号探针为应用生物工程公司协助本室合成。

#### (二) 方法

以M<sub>13</sub>为载体, 按Sanger法<sup>[3]</sup>进行序列分析。每次可读出约350个核苷酸序列。为了进行全序列分析, 我们采用了下述方案: 首先用Hind III和EcoR I将克隆在质粒pUC19中的猪生长激素cDNA切下, 然后将切下的cDNA片段重组在M<sub>13</sub>mp18和M<sub>13</sub>mp19中, 分别用以分析cDNA的5'端和3'端的一段序列。其后将含有pGH cDNA的M<sub>13</sub>mp18RF DNA用Acc I切去pGH cDNA 5'端的一个片段, 用以分析此酶切位点下游的一段序列。将含有pGH cDNA (代号Jia) 的M<sub>13</sub>mp19RF DNA用Sma I和EcoR I切去pGH cDNA的3'端一个片段。用以分析Sma I酶切位点上游的一段序列。这样就得到了pGH cDNA的全序列, 并且在分析出的片段之间有相当长的一段序列是重叠的, 如图1所示。

此外, 对其中一个cDNA (代号Wang) 还用1号探针 (由成熟猪生长激素N-端第一个氨基酸的密码子开始) 和2号探针 (由成熟猪生长激素N-端第81号氨基酸

本文于1988年1月21日收到

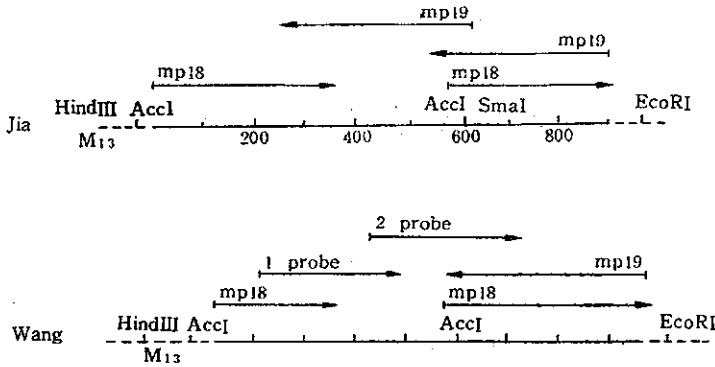


图 1 pGH cDNA 全序列分析过程图示

Fig. 1 Diagram of the pGH cDNAs sequencing

的密码子开始)为引物,分别分析了这两个探针下游的一段序列(图1),这样就增多了分析片段之间的重叠部分,保证了分析结果的可靠性。

## 结 果

图版 I 是我们一次分析的放射自显影图谱,从其中可精确读出约 350 个核苷酸序列。对代号 Jia 的 pGH cDNA 共得到 4 个这样的图谱,对代号 Wang 的则得到 5 个(参见图 1,其余图片从略)。由于片段之间的重叠部分读出的序列(其中有的是从 5' 端向 3' 端读的,有的则是从 3' 端向 5' 端读的)是相同的,因而判断分析的结果是可靠的。现将我们分析的两个 pGH cDNA 的全序列列于图 2。图 2 中还列举了 Seeburg 等发表的序列以供比较。

从图 2 可以看出:

1. 我们的两个 cDNA 对猪前生长激素来说都是全长的,因为都有翻译的起始密码子和终止密码子。而且起始密码子都是相邻的两个,一个为 GTG,一个为 ATG。

2. 两个 cDNA 无论在 5' 端或 3' 端都比 Seeburg 等报道的长很多。在 3' 端除了都有一长段非编码区以外,都出现了多聚

A。其中之一为 17 个 A,另一个则至少为 70 个 A,这些是 Seeburg 等的序列中所没有的。在 5' 端除了有完整的引导序列(Seeburg 等的引导序列不完整,没有起始密码子)外,还都有很长一段非编码区。和 Vize 等(1987)<sup>[4]</sup>发表的猪基因组生长激素基因(图 3)的 5' 端非编码区序列相比,代号 Wang 的 pGHcDNA 的 5' 端距 Vize 推定的转录起始核苷酸相差 15 个核苷酸。代号 Jia 的 cDNA 则直到与 Vize 推定的转录起始核苷酸仅距 8 个核苷酸处仍与基因组基因的序列一致。不过代号 Jia 的 cDNA 在此上游还有长达接近 70 个核苷酸的序列,此段序列与基因组基因的序列不符。如果这段序列是在克隆 cDNA 时随机重组上的一个 cDNA 片段(不是生长激素的,讨论见后),而 Vize 所推定的转录起始位点也是正确的话,仍能说明我们的反向转录已接近模板 mRNA 的 5' 末端,在我们见到的报道中尚未发现有这样长的 cDNA。

3. 将我们的两个以及 Seeburg 等的 pGH cDNA 加以对比,可以看出它们的核苷酸序列虽然基本上相同,但无论在编码区或非编码区彼此都有些差异(图 2 中已用 \* 号标出)。其中在为成熟猪生长激素的编码区中,虽然在核苷酸中有所不

pOH TTTTTTTTTTTTGGTGGGAAAGAAAGAACTTTTATTGGGATGTTTAAAGTGGGGGACTCCAGGAAACACA  
 Jia ACAC TAGGACCCAGCTCCCCAGACCACTCAGGGACCTGTGGACAGCTCACC GGCT GTG ATG  
 met met  
 Wang AGCTCCTCAGACCACTCAGGGACCTGTGGACAGCTCACC GGCT GTG ATG  
 met met  
 Jia GCT GCA GGC CCT CGG -20 ACC TCC GTG CTC CTG GCT TTC GCC CTG CTC -10  
 ala ala ggc pro arg thr ser val leu leu ala phe ala leu leu cys  
 Wang GCT GCA GGC CCT CGG ACC TCC GGC CTC CTG GCT TTC GCC CTG CTC TGC  
 ala ala gly pro arg thr ser ala leu leu ala phe ala leu leu cys  
 Seeburg ACC TCC GTG CTC CTG GCT TTC GCC CTG CTC TGC  
 thr ser val leu leu ala phe ala leu leu cys  
 Jia CTG CCC TGG ACT CAG GAG GTG GGC -1 GGC TTT CCA GCC ATG CCC TTG TCC  
 leu pro trp thr gln glu val gly ala phe pro ala met pro leu ser  
 Wang CTG CCC TGG ACT CGG GAG GTG GGA GGC TTT CCA GCC ATG CCC TTG TCC  
 leu pro trp thr arg glu val gly ala phe pro ala met pro leu ser  
 Seeb CTG CCC TGG ACT CAG GAG GTG GGA GGC TTT CCA GCC ATG CCC TTG TCC  
 leu pro trp thr gln glu val gly ala phe pro ala met pro leu ser  
 Jia AGC CTA TTT 10 GCC AAC GCC GTG CTC CGG GCC CAG CAC 20 CAC CAA CTG  
 ser leu phe ala asn ala val leu arg ala gln his leu his gln leu  
 Wang AGC CTA TTT GCC AAT GCC GTG CTC CGG GCC CAG CAC CTG CAC CAA CTG  
 ser leu phe ala asn ala val leu arg ala gln his leu his gln leu  
 Seeb AGC CTA TTT GCC AAC GCC GTG CTC CGG GCC CAG CAC CTG CAC CAA CTG  
 ser leu phe ala asn ala val leu arg ala gln his leu his gln leu  
 Jia GCT GCC GAC ACC TAC AAG GAG TTT GAG CGG GCC TAC ATC CCG GAG GGA  
 ala ala asp thr tyr lys glu phe glu arg ala tyr ile pro glu gly  
 Wang GCT GCC GAC ACC TAC AAG GAG TTT GAG CGG GCC TAC ATC CCG GAG GGA  
 ala ala asp thr tyr lys glu phe glu arg ala tyr ile pro glu gly  
 Seeb GCT GCC GAC ACC TAC AAG GAG TTT GAG CGG GCC TAC ATC CCG GAG GGA  
 ala ala asp thr tyr lys glu phe glu arg ala tyr ile pro glu gly  
 Jia CAG AGG TAC TCC ATC CAG AAC GCC CAG GCT 50 GGC TTT TGC TTC TCG GAG  
 gln arg tyr ser ile gln asn ala gln ala gln ala phe cys phe ser glu  
 Wang CAG AGG TAC TCC ATC CAG AAC GCC CAG GCT GGC TTT TGC TTC TCG GAG  
 gln arg tyr ser ile gln asn ala gln ala gln ala phe cys phe ser glu  
 Seeb CAG AGG TAC TCC ATC CAG AAC GCC CAG GCT GGC TTT TGC TTC TCG GAG  
 gln arg tyr ser ile gln asn ala gln ala gln ala phe cys phe ser glu  
 Jia ACC ATC CCG GCC 60 ACC GGC AAG GAC GAG GCC CAG CAG AGA 70 TCG GAC  
 thr ile pro asp pro thr gly lys asp glu ala gln gln arg ser asp  
 Wang ACC ATC CCA GCC ACC GGC AAG GAC GAG GCC CAG CAG AGA TCG GAC  
 thr ile pro asp pro thr gly lys asp glu ala gln gln arg ser asp  
 Seeb ACC ATC CCG GCC ACC GGC AAG GAC GAG GCC CAG CAG AGA TCG GAC  
 thr ile pro asp pro thr gly lys asp glu ala gln gln arg ser asp  
 Jia GTG GAG CTG CTG CGC TTC TCG CTG CTG 80 CTC ATC CAG TCG TGG CTC GGG  
 val glu leu leu arg phe ser leu leu leu leu ile gln ser trp leu gly  
 Wang GTG GAG CTG CTG CGC TTC TCG CTG CTG CTC ATC CAG TCG TGG CTC GGG  
 val glu leu leu arg phe ser leu leu leu leu ile gln ser trp leu gly  
 Seeb GTG GAG CTG CTG CGC TTC TCG CTG CTG CTC ATC CAG TCG TGG CTC GGG  
 val glu leu leu arg phe ser leu leu leu leu ile gln ser trp leu gly  
 Jia CCC GTG 90 CAG TTC CTC AGC AGG GTC TTC ACC AAC AGC 100 CTG GTG TTT GGC  
 pro val gln phe leu ser arg val phe thr asn ser leu val phe gly  
 Wang CCC GTG CAG TTC CTC AGC AGG GTC TTC ACC AAC AGC CTG GTG TTT GGC  
 pro val gln phe leu ser arg val phe thr asn ser leu val phe gly  
 Seeb CCC GTG CAG TTC CTC AGC AGG GTC TTC ACC AAC AGC CTG GTG TTT GGC  
 pro val gln phe leu ser arg val phe thr asn ser leu val phe gly  
 Jia ACC TCA GAC CGC GTC TAC GAG AAG CTG AAG GAC CTG GAG GAG GGC ATC  
 thr ser asp arg val tyr glu lys leu lys asp leu glu glu gly ile  
 Wang ACC TCA GAC CGC GTC TAC GAG AAG CTG AAG GAC CTG GAG GAG GGC ATC  
 thr ser asp arg val tyr glu lys leu lys asp leu glu glu gly ile  
 Seeb ACC TCA GAC CGC GTC TAC GAG AAG CTG AAG GAC CTG GAG GAG GGC ATC  
 thr ser asp arg val tyr glu lys leu lys asp leu glu glu gly ile  
 Jia CAG GCC CTG ATG CGG GAG CTG GAG 120 GAT 130 GGC CCC CGG GCA GGA CAG  
 gln ala leu met arg glu leu glu asp gly ser pro arg ala gly gln  
 Wang CAG GCC CTG ATG CGG GAG CTG GAG GAT GGC AGC CCC CGG GCA GGA CAG  
 gln ala leu met arg glu leu glu asp gly ser pro arg ala gly gln  
 Seeb CAG GCC CTG ATG CGG GAG CTG GAA GAT GGC AGC CCC CGG GCA GGA CAG  
 gln ala leu met arg glu leu glu asp gly ser pro arg ala gly gln  
 Jia ATC CTC AAG CAA 140 ACC TAC GAC AAA TTT GAC ACA AAN TTO CGC 150 GAT  
 ile leu lys gln thr tyr asp lys phe asp thr asn leu arg ser asp  
 Wang ATC CTC AAG CAA ACC TAC GAC AAA TTT GAC ACA AAN TTO CGC AGT GAT  
 ile leu lys gln thr tyr asp lys phe asp thr asn leu arg ser asp  
 Seeb ATC CTC AAG CAA ACC TAC GAC AAA TTT GAC ACA AAN TTO CGC AGT GAT  
 ile leu lys gln thr tyr asp lys phe asp thr asn leu arg ser asp

```

Jia      GAC GCG CTG CTT AAG AAC TAC GGG CTG CTC TCC TGC TTC AAG AAG GAC
asp ala leu leu lys asn tyr gly leu leu ser cys phe lys lys asp
Wang     GAC GCG CTG CTT AAG AAC TAC GGG CTG CTC TCC TGC TTC AAG AAG GAC
asp ala leu leu lys asn tyr gly leu leu ser cys phe lys lys asp
Seeb     GAC GCG CTG CTT AAG AAC TAC GGG CTG CTC TCC TGC TTC AAG AAG GAC
asp ala leu leu lys asn tyr gly leu leu ser cys phe lys lys asp

Jia      CTG CAC AAG GCT GAG ACA TAC CTG CGG GTC ATG AAG TGT CGC CGC TTC
leu his lys ala glu thr tyr leu arg val met lys cys arg phe
Wang     CTG CAC AAG GCT GAG ACA TAC CTG CGG GTC ATG AAG TGT CGC CGC TTC
leu his lys ala glu thr tyr leu arg val met lys cys arg phe
Seeb     CTG CAC AAG GCT GAG ACA TAC CTG CGG GTC ATG AAG TGT CGC CGC TTC
leu his lys ala glu thr tyr leu arg val met lys cys arg phe

Jia      GTG GAG AGC AGC TGT GCC TTC TAGTTGCTGGGCATCTCTGTTGCCCTCCCCAGTA
val glu ser ser cys ala phe
Wang     GTG GAG AGC AGC TGT GCC TTC TAGTTGCTGGGCATCTCTGTTGCCCTCCCCAGTA
val glu ser ser cys ala phe
Seeb     GTG GAG AGC AGC TGT GCC TTC TAGTTGCTGGGCATCTCTGTTGCCCTCCCCAGTA
val glu ser ser cys ala phe

Jia      CCTCCCCCTGAACCCCTGGAAAGTGCCACCCCAATGCCGTGCTGTGCTTTCTTAATAAAACCAGGT
Wang     CCTCCCCCTGAACCCCTGGAAAGTGCCACCCCAATGCCGTGCTGTGCTTTCTTAATAAAACCAGGT
Seeb     CCTCCCCCTGGTGGCCCTGGAAAGTGCCACCCCAATGCCGTGCTGTGCTTTCTTAATAAAACCAGGT

Jia      TGCATCGTAAAAAAAAAAAAAAAAAAAA
Wang     TGCATCGTA(70)
Seeb     TTGCATCGTA

```

图 2 猪生长激素cDNA顺序

Fig. 2 The cDNA sequences of porcine growth hormone

```

Vize      AGGCCAGGG TATAAAA AGGCCCAAAAGGACCAA
Jia      TTTTITTTTTTGGTGGGAAGAGCACTTTTATGGCATGTTACTGGGGCACTCCAGG

Vize      TTCC A GAATCCC AGGACCACTGCCCCAGCACTCAGGCACTGTGGACAGCT
Jia      GAAC A CAACACT AGGACCACTGCCCCAGCACTCAGGCACTGTGGACAGCT
Wang     AGCTCTTCAAGCACTCAGGCACTGTGGACAGCT

Vize      CACCCGCT GTG ATG
Jia      CACCCGCT GTG ATG
Wang     CACCCGCT GTG ATG

```

图 3 两个cDNA (Jia和Wang) 及基因组DNA (Vize)5'端的序列

Fig. 3 List of the two cDNAs (Jia and Wang) and the genomic DNA(Vize) sequences on 5' end

同,但所代表的氨基酸是相同的,说明由它们表达出来的成熟生长激素是相同的。然而引导序列中却出现了两个氨基酸的差异,即代号Wang的cDNA中为丙氨酸和精氨酸的密码子处,代号Jia的和Seeburg等的cDNA中却分别为缬氨酸谷氨酰胺。

## 讨 论

1. 在我们分析的两个pGH cDNA中都有两个相邻的起始密码子(和Vize报道的基因组基因相符,Seeburg等的缺少这段序列)。Seeburg等<sup>[1]</sup>报道的牛生长激

素cDNA也有两个相邻的起始密码子(相邻两个ATG)。是否许多真核基因都有两个相邻的起始密码子?由于我们掌握的资料太少,目前当然得不出结论。但这是一个很值得注意的问题。仅就我们已知的资料而言,至少猪和牛生长激素的mRNA都有两个相邻的起始密码子。它对这些基因的表达有没有特殊意义,是很值得探讨的问题。

2. 在结果中已经提到,在代号Jia的pGH cDNA的5'端有很长一段是Vize报道的猪基因组生长激素基因的5'端所没有的,而且在整个基因(包括内含子在内)中也没有这段序列。如果在猪生长激素mRNA中确实有这段序列(我们还不能完全排除这种可能性),则是一个非常有意思的问题。不过我们认为很可能是在制得总cDNA后,在向质粒pUC19中重组时,由于用的是钝端联结,因而随机地联上了混合cDNA中的一个非生长激素的cDNA片段。如果是这样,则我们注意到我们的两个cDNA以及Seeburg等的cDNA尽管它

们的长短彼此不同(即反向转录达到的位点不同),但它们的5'端都是腺苷酸。这是纯属偶合,还是有什么意义?是值得探讨的。

3. 由于两个cDNA的5'端都已很接近转录的起始点,因而认为核糖体结合区必定在这一段非编码区序列之内。由于真核mRNA的核糖体结合区迄今尚未查明是否也有象原核mRNA那样的有特征的S.D.序列,因而不能判断出哪一段是核糖体结合区。然而我们注意到在这一段中也有AGGA以及AGGGA和TGGA样的序列。

4. 由于两个cDNA的3'端都有多聚

A尾巴,表明它们都已到达多聚A聚合酶作用(加多聚A尾巴)的位点。和Vize等的基因组基因对比,看来在猪生长激素基因的初始转录物的加工中是在尿苷酸和腺苷酸之间用核酸酶切断后,再加上多聚A尾巴的。这与Vize的推断相符。此位点与其上游的AATAAA序列相距15个核苷酸,与一般认为的相距10—15个核苷酸的规律相符。

5. 在我们的两个cDNA之间以及它们与Seeburg等的cDNA之间在核苷酸序列上都存在一些差异。这种差异意味着什么是值得探讨的。

### 参 考 文 献

- [1] Seeburg, P.H. et al.: *DNA*, 2:37—45, 1983.
- [2] 王华岩等: *北京农业大学学报*, 13(4):379—386, 1987.
- [3] Sanger, F. et al.: *Proc. Natl. Acad. Sci. USA*, 74:5463—5467, 1977.
- [4] Peter, D.V. and Wells, J.R.E.: *Gene*, 55:339—344, 1987.

## cDNA SEQUENCING OF THE PORCINE GROWTH HORMONE

Qi Shunzhang Wang Xinzhang Zhou Shunwu Jia Feng

Wang Huayan Xia Li Li Juan

(Faculty of Animal Biochemistry, Beijing Agricultural University)

Sequences of two pGH cDNAs are reported in this paper. The cDNAs were obtained by reverse transcription, cloning and screening from porcine pituitary gland mRNA in the same lab. Discussion and comparison about the two cDNA sequences, the one published by Seeburg et al. and the pGH genomic DNA sequence published by Vize et al are also provided.

### Key words

cDNA sequence; porcine growth hormone