

简报

牛凝乳酶原cDNA的序列及其 缺失机制的分析

谭思元 张渝英 刘彤瑶

刘年娟 杨开宇

(中国科学院微生物研究所, 北京)

前文报道, pCT5及pCT15均带有凝乳酶原cDNA的插入片段。根据酶谱分析, 前者5'端缺少约80bp, 后者在编码区域中部有缺失^[1]。我们从pCT5及pCT15出发, 完成了凝乳酶原cDNA的序列分析, 并讨论了pCT15的缺失机制。

实验部分

整个工作分两步进行: 第一, 完整基因的拼接; 第二, 序列分析。

(一) 拼接是与构建表达质粒同时进行的。首先, 将pDR720中 Trp 启动子部分 (Hind III-BamH I 片段) 亚克隆到pUC19中, 通过 BamH I 切点引入人工接头。人工接头含有起始密码子 ATG、凝乳酶原前五个氨基酸的编码区以及新的 BamH I 切点, 这样得到 pTrL9。其次, 将 pCT5中的凝乳酶原基因 3'端约1.1kb的 Pst I 片段亚克隆到pUC19中, 以便利用pUC19上的多酶切位点。截取此亚克隆的 EcoR I-Sma I 片段, 以及pCT15中目的基因5'端的 BamH I-EcoR I 片段, 将二者同时引入pTrL9中, 获得含有 Trp启动子、ATG及凝乳酶原cDNA的质粒pTLC23 (图1)。

(二) 对pTLC23中凝乳酶原cDNA进行亚克隆, 其战略如图2所示。克隆载体分别为M13mp18、M13mp19以及 pBluescript(+), pBluescript(-)。后二者在辅助病毒作用下, 可获取单链DNA^[2]。制备单链模板DNA, 并按双脱氧末端终止法进行序列分析^[3]。

结果与讨论

(一) 凝乳酶原的核苷酸序列见图3。已知凝乳酶原有A、B两种类型。主要区别在于286位的氨基酸残基不同。前者为 Asp, 后者为Gly。根据图3, 可以确定我们所获得的cDNA为编码凝乳酶原B的基因。

(二) 关于凝乳酶原cDNA的序列已有三个实验室用Maxam & Gilbert的化学法进行了测定^[4-6], 但结果不尽相同, 与Foltmann的氨基酸序列分析结果^[7]进行比较, 也有一些区别。现在, 我们提供了用双脱氧末端终止法所测得的结果, 这对于分析已有数据之间的矛盾将有一定的帮助。

Harris等^[5]所报道的凝乳酶原cDNA序列中, 在384、960位为A, 769位为T, 而Moir等^[6]的结果分别为G、G、C, 并认为前者用化学法只测定了一条链, 可能有误。我们的数据为Moir等的论点提供了旁证。

根据cDNA序列所导出的凝乳酶原氨基酸序列中, 在214位, Harris等^[5]测定为Asn, 而Nishimori等^[4], Mcir等^[6]和我们的结果都为Asp; 在319位, Nishimori等测定为Gly, 而Harris等、Moir等和我们的结果都为Ser。再对照Foltmann^[7]所报道的直接氨基酸序列分析结果, 结论是在214位应为Asp, 319位应为Ser。

本文于1988年10月20日收到。

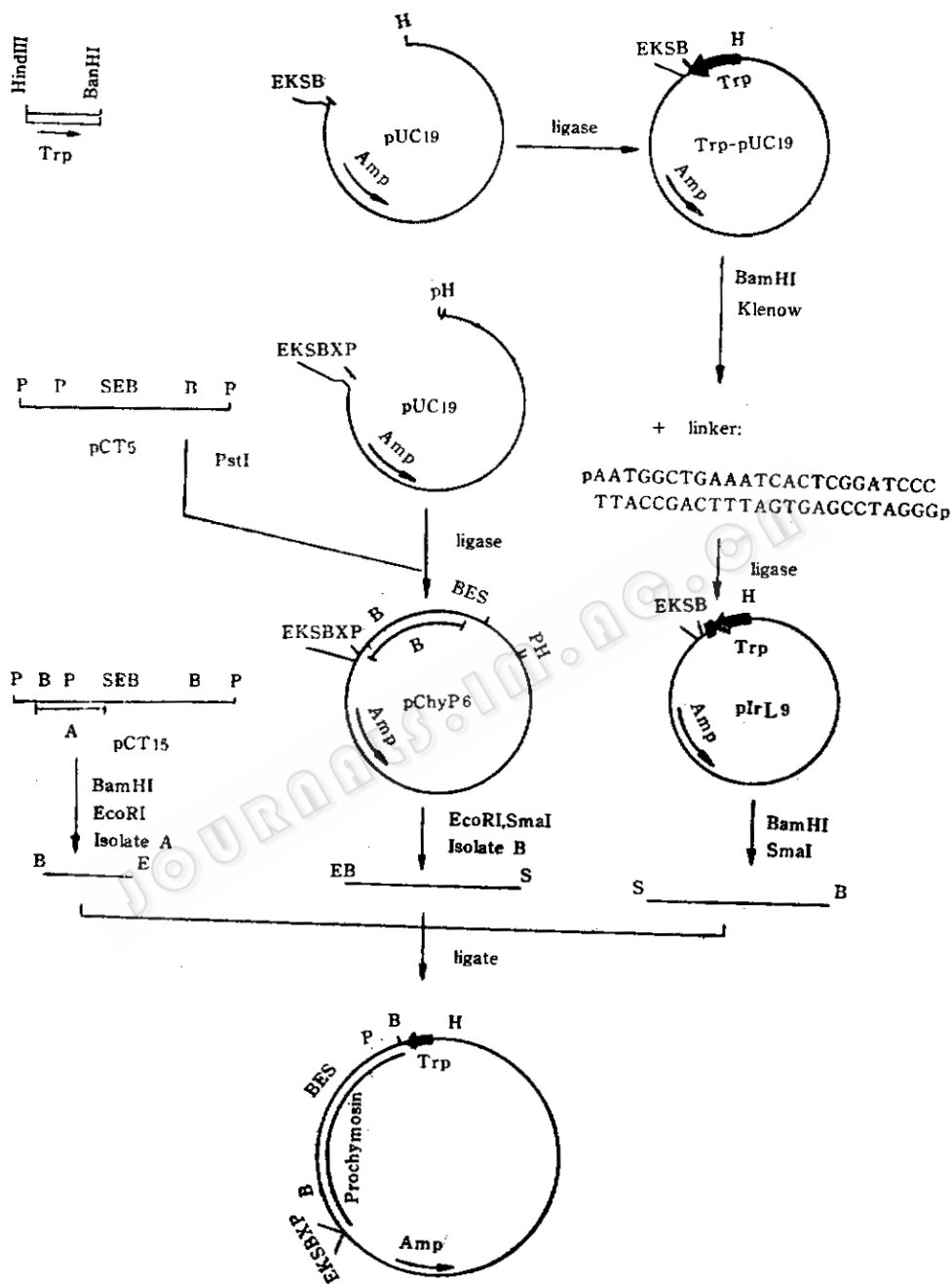


图 1 质粒pTLC23的构建

Fig. 1 Construction of plasmid pTLC23

Abbreviations: B, BamHI; E, EcoRI; H, HindIII
K, KpnI; P, PstI; S, SmaI; X, XbaI

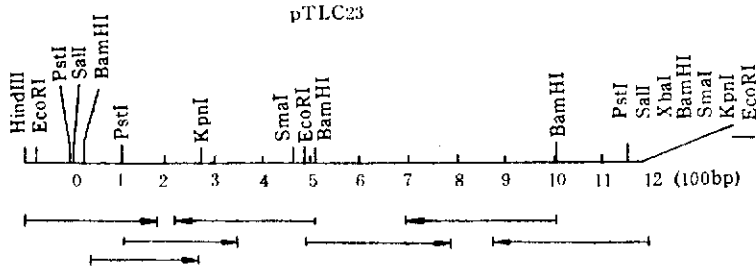


图 2 凝乳酶原 cDNA 的限制性内切酶谱及其克隆战略
Fig. 2 Restriction map and sequencing strategy for prochymosin cDNA

文献报道根据 cDNA 序列所导出的凝乳酶原一级结构都表明在 202 位为 Asn, 而氨基酸序列分析结果为 Asp, 并认为这很可能是在氨基酸序列分析时发生了脱氨, 以致将 Asn 误定为 Asp, 我们的结果支持这一论点。

(三) Moir 等^[6]和我们在构建凝乳酶原 cDNA 库时都发现, 在某些克隆的编码区域中部发生缺失, 对于这种缺失的原因一直没有确切的解释。经过 DNA 序列分析, 确证 pCT15 缺失部分为 862—960 位。对照 Hidaka 等^[8]报道的包括内含子在内的凝乳酶原基因全序列发现, 这一段正好在内含子 G 和内含子 H 之间。已知, 内含子剪切总是以 GT 起始, AG 告终, 而这段序列的末尾正好是 AG, 因此, 有理由认为, 这段序列是在 mRNA 前体后加工时被删除的。一种可能性是在剪切内含子 G 时错误地利用这个信号将此片段同时切下, 然后再剪切内含子 H, 这样拼接成

功的成熟 mRNA 便缺失了这一段。另一种可能是在剪切时利用内含子 G 的起始信号和内含子 H 的终止信号, 把两个内含子及它们中间的外显子序列同时删除。按照上述机理完全可以解释 Moir 等所获得的克隆 R207。在 R207 中, 缺失部位为内含子 E 和 F 之间的外显子序列, 其结尾也正好是 AG, 这些特点为缺失提供了条件。至于这种缺失机理是否适用于凝乳酶原以外的 cDNA 库中, 尚不得而知。

在已报道的 5 个凝乳酶原 cDNA 库中, 有 2 个 cDNA 库中出现外显子缺失的克隆, 说明这种缺失不是偶然现象。如上分析, 这种缺失正是体内 mRNA 前体错误后加工引起的, 而不是人为的结果。考虑到生物体内各种机能都是在严格调控的基础上进行的, 我们怀疑这类外显子序列的缺失是否与调控机制有关, 以适应某种生理需要, 而不是一种简单的“错误”后加工。

CTGCAGGTCGACGGATCAATG																				
Ala	Glu	Ile	Thr	Arg	Ile	Pro	Leu	Tyr	Lys	Gly	Lys	Ser	Leu	Arg	Lys	Ala	Leu	Lys	Glu	20
GCT	GAA	ATC	ACT	CGG	ATC	CCT	CTG	TAC	AAA	GGC	AAG	TCT	CTG	AGG	AAG	GCS	CTG	AAG	GAG	
His	Gly	Leu	Leu	Glu	Asp	Phe	Leu	Gln	Lys	Gln	Gln	Tyr	Gly	Ile	Ser	Ser	Lys	Tyr	Ser	40
CAT	GGG	CTT	CTG	GAG	GAC	TTC	CTG	CAG	AAA	CAG	CAG	TAT	GGC	ATC	AGC	AGC	AMG	TAC	TCC	
Gly	Phe	Gly	Glu	Val	Ala	Ser	Val	Pro	Leu	Thr	Asn	Tyr	Leu	Asp	Ser	Gln	Tyr	Phe	Gly	60
GGC	TTC	GGG	GAG	GTG	GCC	AGC	GTG	CCC	CTG	ACC	AAC	TAC	CTG	GAT	AGT	CAG	TAC	TTT	GGG	
Lys	Ile	Tyr	Leu	Gly	Thr	Pro	Pro	Gln	Glu	Phe	Thr	Val	Leu	Phe	Asp	Thr	Gly	Ser	Ser	80
AAG	ATC	TAC	CTC	GGG	ACC	CCG	CCC	CAG	GAG	TTC	ACC	GTG	CTG	TTT	GAC	ACT	GGC	TCC	TCT	
Asp	Phe	Trp	Val	Pro	Ser	Ile	Tyr	Cys	Lys	Ser	Asn	Ala	Cys	Lys	Asn	His	Gln	Arg	Phe	100
GAC	TTC	TGG	GTA	CCC	TCT	ATC	TAC	TGC	AAG	AGC	AAT	GCC	TGC	AAA	AAC	CAC	CAG	CCG	TTC	
Asp	Pro	Arg	Lys	Ser	Ser	Thr	Phe	Gln	Asn	Leu	Gly	Lys	Pro	Leu	Ser	Ile	His	Tyr	Gly	120
GAC	CCG	AGA	AAG	TGG	TCC	ACC	TTC	CAG	AAC	CTG	GGC	AAG	CCG	CTG	TCT	ATC	CAC	TAC	GGG	
Thr	Gly	Ser	Met	Gln	Gly	Ile	Leu	Gly	Tyr	Asp	Thr	Val	Thr	Val	Ser	Asn	Ile	Val	Asp	140
ACA	GSC	AGC	ATG	CAG	GGC	ATC	CTG	GGC	TAT	GAC	ACC	GTC	ACT	GTC	TCC	AAC	ATT	GTG	GAC	
Ile	Gln	Gln	Thr	Val	Gly	Leu	Ser	Thr	Gln	Glu	Pro	Gly	Asp	Val	Phe	Thr	Tyr	Ala	Glu	160
ATC	CAG	CAG	ACA	GTA	GGC	CTG	AGC	ACC	CAG	GAG	CCC	GGG	GAC	GTC	TTC	ACC	TAT	GCC	GAA	
Phe	Asp	Gly	Ile	Leu	Gly	Met	Ala	Tyr	Pro	Ser	Leu	Ala	Ser	Glu	Tyr	Ser	Ile	Pro	Val	180
TTC	GAC	GGG	ATC	CTG	GGG	ATG	GCC	TAC	CCC	TGC	CTC	GCC	TCA	GAG	TAC	TCS	ATA	CCC	GTG	
Phe	Asp	Asn	Met	Met	Asn	Arg	His	Leu	Val	Ala	Gln	Asp	Leu	Phe	Ser	Val	Tyr	Met	Asp	200
TTT	GAC	AAC	ATG	ATG	AAC	AGG	CAC	CTG	GGT	GCC	CAA	GAC	CTG	TTC	TGG	GTT	TAC	ATG	GAC	
Arg	Asn	Gly	Gln	Glu	Ser	Met	Leu	Thr	Leu	Gly	Ala	Ile	Asp	Pro	Ser	Tyr	Tyr	Thr	Gly	220
AGG	AAT	GGC	CAG	GAG	AGC	ATG	CTC	ACG	CTG	GGG	GCC	ATC	GAC	CCG	TCC	TAC	TAC	ACA	GGG	
Ser	Leu	His	Trp	Val	Pro	Val	Thr	Val	Gln	Gln	Tyr	Trp	Gln	Phe	Thr	Val	Asp	Ser	Val	240
TCC	CTG	CAC	TGG	GTG	CCC	GTG	ACA	GTG	CAG	CAG	TAC	TGG	CAG	TTC	ACT	GTG	GAC	AGT	GTG	
Thr	Ile	Ser	Gly	Val	Val	Val	Ala	Cys	Glu	Gly	Gly	Cys	Gln	Ala	Ile	Leu	Asp	Thr	Gly	260
ACC	ATC	AGC	GGT	GTG	GTT	GTG	GCC	TGT	GAG	GGT	GGC	TGT	CAG	GCC	ATC	CTG	GAC	ACG	GGC	
Thr	Ser	Lys	Leu	Val	Gly	Pro	Ser	Ser	Asp	Ile	Leu	Asn	Ile	Gln	Gln	Ala	Ile	Gly	Ala	280
ACC	TCC	AAG	CTG	GTG	GGG	CCC	AGC	AGC	GAC	ATC	CTC	AAC	ATC	CAG	CAG	GCC	ATT	GGA	GCC	
Thr	Gln	Asn	Gln	Tyr	Gly	Glu	Phe	Asp	Ile	Asp	Cys	Asp	Asn	Leu	Ser	Tyr	Met	Pro	Thr	300
ACA	CAG	AAC	CAG	TAC	GGT	GAG	TTT	GAC	ATC	GAC	TGC	GAC	AAC	CTG	AGC	TAC	ATG	CCC	ACT	
Val	Val	Phe	Glu	Ile	Asn	Gly	Lys	Met	Tyr	Pro	Leu	Thr	Pro	Ser	Ala	Tyr	Thr	Ser	Gln	320
GTG	GTC	TTT	GAG	AIC	AAT	GGC	AAA	ATG	TAC	CCA	CTG	ACC	CCC	TCC	GCC	TAT	ACC	AGC	CAG	
Asp	Gln	Gly	Phe	Cys	Thr	Ser	Gly	Phe	Gln	Ser	Glu	Asn	His	Ser	Gln	Lys	Trp	Ile	Leu	340
GAC	CAG	GGC	TTC	TGT	ACC	AGT	GGC	TTC	CAG	AGT	GAA	AAT	CAT	TCC	CAG	AAA	TGG	ATC	CTG	
Gly	Asp	Val	Phe	Ile	Arg	Glu	Tyr	Tyr	Ser	Val	Phe	Asp	Arg	Ala	Asn	Asn	Leu	Val	Gly	360
GGG	GAT	GTT	TTC	ATC	CGA	GAG	TAT	TAC	AGC	GTC	TTT	GAC	AGG	GCC	AAC	AAC	CTC	GTG	GGG	
Leu	Ala	Lys	Ala	Ile	***														370	
CTG	GCC	AAA	GCC	ATC	TGATC	ACATCGCTGACCAAGAACCTCACTGTCCCCACACAC														380

图 3 凝乳酶原cDNA的核苷酸序列

Fig. 3 Nucleotide sequence of prochymosin cDNA

核苷酸及氨基酸序列的编号均从凝乳酶原开始。斜线表示凝乳酶原和凝乳酶的连接处

The nucleotide sequence and amino acid sequence is numbered from the beginning of prochymosin.

The oblique line indicates the junction between prochymosin and chymosin

参 考 文 献

- [1] 谭恩元等: 生物工程学报, 4(2):91—97, 1988.
 [2] Russel, M. et al.: *Gene*, 45:333—338, 1986.
 [3] Sanger, F. et al.: *J. Mol. Biol.*, 143:161—178, 1980.
 [4] Nishimori, K. et al.: *J. Biochem.*, 91:1085—88, 1982.
 [5] Harris, T. J. R. et al.: *Nucleic Acids Res.*, 10(7):2177—78, 1982.
 [6] Moir, D. et al.: *Gene*, 19:127—138, 1982.
 [7] Foltmann, R. I. et al.: *Proc. Natl. Acad. Sci., USA*, 74:2321—24, 1977.
 [8] Hidaka, M. et al.: *Gene*, 43:197—203, 1986.

NUCLEOTIDE SEQUENCE OF BOVINE PROCHYMO SIN cDNA AND THE MECHANISM OF cDNA DELETION

Tan Siyuan Zhang Yuying Liu Tongyao

Liu Nianjuan Yang Kaiyu

(*Institute of Microbiology, Academia Sinica, Beijing*)

A plasmid pTLC23 harbouring entire cDNA sequence coding for prochymosin was constructed from clones pCT5 and pCT15. The nucleotide sequence of the complete cDNA insert determined by using dideoxy sequencing method indicates that the cloned DNA apparently represents the coding sequence for prochymosin B. Based on the data presented here, the differences between the DNA sequences reported by Nishimori et al., Harris et al. and Moir et al. were discussed as well.

DNA sequence analysis reveals that the deletion in pCT15 covers positions 862—960. The missing sequence is characterized as an entire exon between intron G and intron H of prochymosin gene and by ending with splicing signal AG, suggesting the deletion in middle region of coding sequence results from the incorrect splicing of pre-mRNA.

Key words

Prochymosin; molecular cloning; recombinant DNA; deletion; pre-mRNA splicing