

链霉菌质粒 pSGL1 最小复制子序列测定及分析*

张 华 洪 斌 李 元

(中国医学科学院 协和医科大学 医药生物技术研究所 北京 100050)

摘 要 质粒 pSGL1(7.4kb)是从球孢链霉菌(*Streptomyces globisporus*)中分离得到的一个高拷贝质粒,已证明其最小复制子位于 *Sau*3AI 酶切的 2.0kb 片段上。对该片段进行亚克隆,测序后数据表明该片段是一个新序列。仅有一个开放阅读框架(ORF R)位于最小复制子中,推测其编码的蛋白质含有滚环复制质粒复制酶的特定序列。

关键词 质粒, 最小复制子, 滚环复制, 复制酶

分类号 Q753 **文献标识码** A **文章编号** 0001-6209(1999)04-0327-32

链霉菌是一种丝状革兰氏阳性细菌。关于链霉菌的重组 DNA 技术在过去的十几年里发展迅速,到目前为止根据不同的需要已经在链霉菌中构建了大量的质粒载体。质粒 pSGL1 是本室从球孢链霉菌(*Streptomyces globisporus*)中分离得到的一个高拷贝质粒^[1]。已经证明其最小复制子位于 *Sau*3AI 酶切的 2.0kb 片段上,并构建了含有该片段的质粒 pSGLS3^[2]。本工作对该片段进行了亚克隆,测得该片段的全序列,并对其进行了分析,为利用它构建新的功能良好的载体系统提供依据。

1 材料和方法

1.1 菌种和质粒

大肠杆菌 *E. coli* JM109, 重组变铅青链霉菌 *S. lividans* TK54 为本室保存。质粒 pUC18, pSGL1, pSGLS3 为本室保存。

1.2 培养基

1.2.1 YEME: 酵母膏 3g, 蛋白胨 5g, 麦芽提取物 3g, 葡萄糖 10g, 蔗糖 34g, 氯化镁 1g, 定容至 1L, pH7.0, 用于 *S. lividans* TK54 液体培养。

1.2.2 LB: 酵母膏 5g, 蛋白胨 10g, 氯化钠 10g, 定容至 1L, pH7.0, 用于 *E. coli* 液体培养, 加 1.2% 琼脂用于 *E. coli* 固体培养。

1.3 酶和试剂

限制性内切酶, DNA 聚合酶和 DNA 连接酶购自美国 BRL 公司。

1.4 质粒的提取及 DNA 片段回收

大肠杆菌质粒和链霉菌质粒大量提取采用碱裂解法^[3]。质粒及 DNA 片段回收采用

本文中的缩写: SSI, single strand initiation; RBS, ribosome binding site; ori, original region, bp, base pair; ORF, open reading frame; Tsr^r, thiostrepton resistant.

* 本课题获中国医学科学院基金资助(No. 952001)

收稿日期: 1997-11-05, 修回日期: 1998-01-23

ISCO 公司的 Little Blue Tank 进行。

1.5 亚克隆过程

提取并回收含有最小复制子的重组质粒 pSGLS3。将该质粒用 *Xho*I/*Cla*I, *Kpn*I/*Bam*HI, *Sma*I/*Xho*I, *Kpn*I/*Pvu*II 酶切后分离, 0.7kb, 0.7kb, 0.7kb, 0.5kb 片段, 分别与用 *Sal*I/*Kpn*I, *Kpn*I, *Bam*HI, *Sma*I/*Sal*I, *Kpn*I/*Hinc*II 酶切的 pUC18 连接。转化大肠杆菌后挑取重组子, 得到重组质粒 S-1, S-2, S-3, S-4。

1.6 大肠杆菌的转化

按文献^[3]进行。

1.7 序列测定

由 ABI 公司的自动测序仪进行。

1.8 相关序列计算机分析

由 blast 库自动进行序列比较。启动子及回文结构分析由程序 GOLDKEY 进行。

2 结果

2.1 重组质粒的鉴定

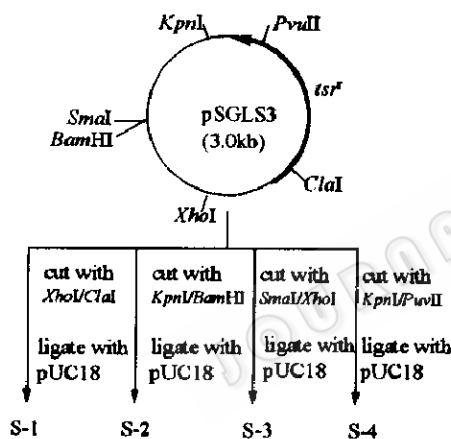


图 1 质粒 pSGLS3 亚克隆构建

Fig. 1 Construction of the subcloning of the plasmid pSGLS3

质粒 pSGLS3 是质粒 pSGL1 的最小复制子 *Sau*3AI 2.0kb 片段与硫链丝菌素抗性基因 *tsr'* 连接后形成的一个的衍生质粒。为了测定该 2.0kb 片段的序列, 将该质粒 pSGLS3 进行了亚克隆(图 1)。

含有链霉菌质粒 pSGLS3 片段的重组质粒 S-1, S-2, S-3, S-4 的酶切鉴定如图 2~5。S-1 用 *Xho*I 酶切后证实 *Xho*I 位点消失, 用 *Sst*II, *Eco*RI 单酶切得到一分子量 3.4kb 的片段。用 *Eco*RI/*Pst*I 酶切可得到克隆的 0.7kb 的小片段。S-2 用 *Kpn*I 和 *Eco*RI 单酶切得到一分子量为 3.4kb 的片段, 用 *Eco*RI/*Hind*III 酶切可得到克隆的 0.7kb 的小片段。S-3 用 *Sma*I 单酶切可得到 3.4kb 片段, 用 *Bam*HI 酶切可得到一小于 0.7kb 的小片段和大于 2.7kb 的大片段, 用 *Eco*RI/*Pst*I 酶切可得到克隆的 0.7kb 的小片段。S-4 用 *Kpn*I 单酶切得到一分子量为 3.2kb 的片段, 用 *Eco*RI/*Pst*I 双酶切可得到克隆的 0.5kb 小片段。

酶切鉴定证明重组质粒 S-1, S-2, S-3, S-4 中含有质粒 pSGL1 的最小复制子片段。

2.2 序列测定及分析

2.2.1 DNA 分析: 提取重组质粒 S-1, S-2, S-3, S-4 后在 ABI335 自动测序仪上进行序列测定, 测得一个长 1990bp 的序列(图 6)。整个序列中 G+C 含量为 69.8%, 这与链霉菌染色体 G+C 含量在 70% 左右相吻合。将该序列输入计算机后与 blast 库进行自动分析, 证明其为一新的序列。局部 DNA 相似分析表明该序列的 152~620 区域与链霉菌质粒 pSG5 的 2208~2701 区域的同源性较高, 而该区域位于质粒 pSG5 的非编码区。该序列

与链霉菌质粒 pIJ101 几乎无同源性。

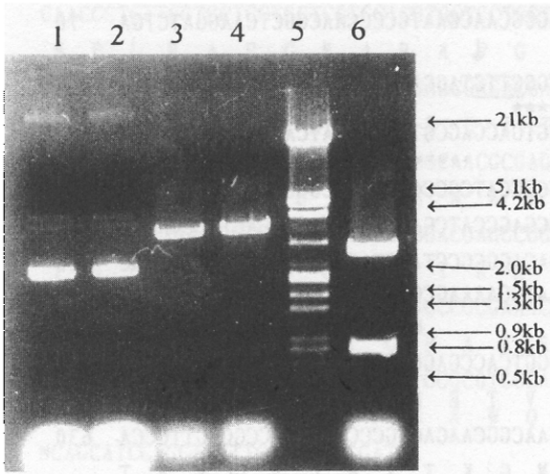


图 2 S-1 酶切鉴定

Fig. 2 Restriction endonuclease analysis of recombinant plasmid S-1
1. S-1DNA; 2. S-1 DNA/*Xho*I; 3. S-1DNA/*Sst*II; 4. S-1 DNA/*Eco*RI; 5. λ DNA/*Eco*RI + *Hind*III; 6. S-1DNA/*Eco*RI + *Pst*I.

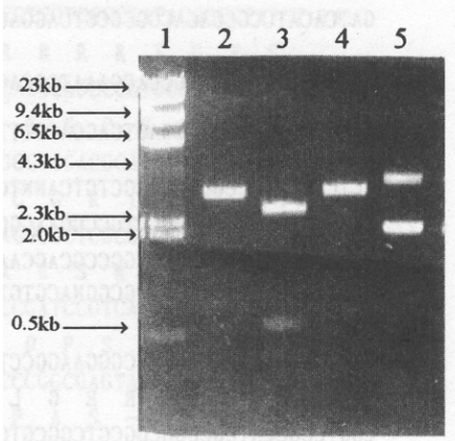


图 3 S-2 酶切鉴定

Fig. 3 Restriction endonuclease analysis of recombinant plasmid S-2
1. λ DNA/*Hind*III; 2. S-2DNA/*Kpn*I; 3. S-2DNA/*Eco*RI + *Pst*I; 4. S-2DNA/*Eco*RI; 5. S-2 DNA.

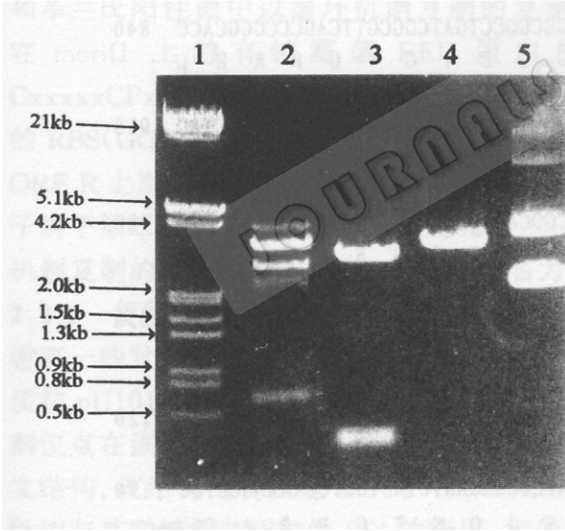


图 4 S-3 酶切鉴定

Fig. 4 Restriction endonuclease analysis of recombinant plasmid S-3
1. λ DNA/*Eco*RI + *Hind*III; 2. S-3DNA/*Eco*RI + *Pst*I; 3. S-3DNA/*Bam*HI; 4. S-3DNA/*Sma*I; 5. S-3 DNA.

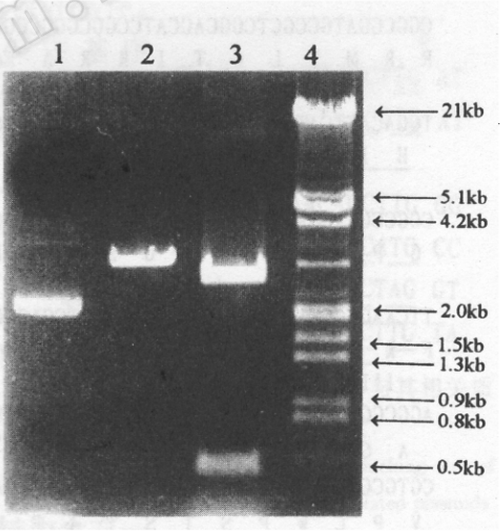


图 5 S-4 酶切鉴定

Fig. 5 Restriction endonuclease analysis of recombinant plasmid S-4
1. S-4DNA; 2. S-4DNA/*Kpn*I; 3. S-4DNA/*Eco*RI + *Pst*I; 4. λ DNA/*Eco*RI + *Hind*III.

2.2.2 蛋白质分析:整段序列中从 ATG 开始的 ORF 有 14 处,根据 Blast 库检索其中 13 个都无编码蛋白质的可能性,只有一个 1548bp 的 ORF R 为有功能的 ORF。推测其编码的蛋白质的 N 端与链霉菌质粒 pSG5 的复制酶 REP 有很高的同源性。它有三个保守区

域:futltxxx(motif1), xpHuHuuux(motif2), uxxYuxKxxx(motif3), 这些基序都是在噬菌体

GATCACATCCGCCACACGGCGCGTCAGGAGGACCGGGGGCAACGAATGCCCAACGGCTCAAGGATCTGA 70
↓
ACGGGCCTCATGGGGCCAGGAAATGGCAGCACCCCGGCTTGTAGGAGGCGGGGGCCGCGCCCCCGA 140

ATCGGCGTCGCATGCCGAGTGAGGACGTTTGTCTGTGACGAGCGTAACGAGATCAGCGTAACCCGGTA 210

CCTCGCCCCCGCGGACCCCGCTGTCAANTGCGGGACCCCATCGCCCGGACGCGGGCAAGCGGCCCGCT 280
NGTGTACTGCTCGGACGCTGCCGGAAGGGGCAAAGCGAGCCATCGAGCGGGGCACTGCACTGCCACCA 350
CGACCGCTGAGGCACGACCGCGCGCAGCAAATCCGGACAGGCGCTACCTACTACCGGAGAAAGTGCAAC 420
CCGTGCAGGTACAGGCGCTGCGGNACGTGTCCGGAATGCAAAAGCGGACAGCCCCGGGATCGCGACCCG 490
■■■ M R T R
CGATGCCAAGCTGGACCTCCGGGAAGGCCTCCAGCGCGTCACCGAGGACAAGGGGCTGAAGGGCTCGCGC 560
D A K V D L R E G L Q R V T E D K G L K G C G
CGGTCCGCCATCGGGGCGCGCTCGGCGTCTGCTCTGAACGGCAAGACGGCCACCTGTCCGGCGTTGCCA 630
R S A I A G G V G V V L N G K T A H L S G V A T
CGTGGGAAAGATCCACCTCTGCCCGTCTGCTCAGCCAAGATCCGGGCGCGCGGTCCGAGGAGATCGA 700
C G K I H L C P V C S A K I R A A R S E E I D
IV
CACGGTCACGGCGCGCTGGCAGGCAGCAGCATGGGCTGGCGATGATGACGCTACCCCTCCGGCACTGC 770
T V T G A W Q A A E H G L A M M T L T L R H Y
I
CGGCGGATGCCGCTCGGCACCATCCGGCGGGCGGAGAGGGCGGCGCTGATCGCGTTACGCCCGGGCACC 840
R R M P L G T I R R A E R G G L I G V Q P R H L
II
TGCACATTCATACCATCGTCTGTGGATCCTGGCCCGTTGGGTCCGCTCCGTGGAGAGTCAGGGGGCTAC 910
H I H T I V L W I L A R W V R S V E S Q G A T
II
CGGCCCCAACGACGCTGGACTCCGGATCGACGTCCCCCGGCGGAGACCCCAACAGTTACGCCGCTACCTG 980
G P T T V D S G S T S P G R R P Q Q F S R Y L
III
TTCAAGTCCCAGGACGGCAAGGCGCGTTTCGAGCGCTGGGCGCGGGCGCGGAGCTGACCACTGCCGACA 1050
F K S Q D G K A R F E R W A P G A E L T S A D K
III
AGGCCGGGCGTAAGGCGTCCCGGATGCCGTTCCGAGGTGGCGGACGGGGCGGCGCGCGGCTCGCCGAGGA 1120
A G R K A S R M P F E V A A G A A A G L A E D
CGTGGCGCTGTGGCAGGTACGAGCACGGCGCCCATGGGATCCGGGAATCTACTGGTCCAAANGCCTGC 1190
V P L W P S T S T A P M G S G Q S T G P ? A C
GGGCACTGCTCGGCAAGCTCGTGCAAGCTCGACGACCGACCGAGGAGATTGCCTCCGAGGACGTTG 1260
G H C S A S S C K L D E R T D E E I A S E D V G
CGGCACTGGCGTGCCTGATCCCGGAGAGACCTCGTACCGCGTGATCCTCCGGCACCGTGGCCGCTC 1330
G T G V A L I P A E T W Y G V I L R H R C R S
GCTGACCTGATTGGGGGCGCGGAGATCGGGCGGCTGCGCGGCTCCAGTCCGCTCATCAGCGATGGGG 1400
L D L I R G R R D R R R R R G P V A H H R M G
GCTGACTGGGGCGGGACGTTCTCCCGTCCCGGACCCGACCNACTGACCCTGTTCTGGCCCGGG 1470
A ? V G P G R S P G P G D P D ? L T L F W P G G

GTTTCATCACGGCCGCGCCGGATCGGCCTGCACCGCATTTGTGCCGGCCGAACCGGATCCGCCTCGAG 1540
S S R P P G R I G L H R I C A G R T G S A S R
GAACCTCTTTCGTGCTCCTGGTCTGCTGCTCCTGGTCTGCTGCTGCTGCGCGATACGGCGTATCT 1610
N P L R A P G R A R A P G A R R R R R Y G V S
CATGGCACGGCCGCTCCGCTCCGGCGGATGAGCGAGACGACACGGCAGGGAGCCCGTGCGCCTGGTGCG 1680
H G T A G R V R R M S E T T T A G S P C A W C G
GCGAGCCGGTCCGACAGTCCGGGATCGGGCGCAACCGGAGTACTGCGGGCGGACGCATCGGGAGCTGGCC 1750
E P V R Q S G I G R N A S T A G G R I G S W P
TACCGGGCACGGCGGACGACGGCGGTTGACGAGCGGTTGCCAGCGCTCGCGGCGGACGCGTCCG 1820
T G H G G S S G R W T R R L P R R S R P Q R R
AATCGTCACTTGTCACTTGTGAAACGGGGCCCGGAAACGGGGTCCCGATCCGTCACTTGTGAAACGT 1890
N R Q L S V V E T G P A K R G P D P S V V E T S
CTCCGGGGCAGGTACGCCCCGTGATTCCGTCCGCGTGGCAGGCGGCCCCGGCGAGTAGCNGCGGCTTCT 1960
P G Q V S P V I P S A W Q A A P A S -
NCAGCATCCNGGTTTTTCGAAGCGGGATCA 1990

图 6 质粒 pSGL1 最小复制子序列

Fig. 6 Sequence of the minimal replicon of the pSGL1

★. Inverted repeat sequence; ↓. Putative nicking site; ◆. Consensus sequence of SSI;
■. Putative RBS; I、II、III、IV, Consensus motif of REP protein.

和革兰氏阳性菌中以滚环机制复制的复制酶的保守基序。在 motif1 上游有链霉菌 REP 蛋白质的保守基序 CxxxxxCPxC(motif4)。此外 ORF R 上游 10bp 处有一潜在的 RBS(GGA)。链霉菌基因的启动子可分为两种类型^[4], 在 ORF R 上游没有检索到典型的 -10, -35 区, 因此它的启动子属于第二种类型。根据上述讨论推定 ORF R 编码以滚环机制复制的复制酶 REP。将 ORF R 命名为 *rep*。

2.2.3 复制起始点 ori 与单链起始 SSI 功能区: *rep* 基因上游有一些发卡结构(图 6)。比较 100~130 处的回文结构和质粒 pIJ101, pJV1, pSG5 的切割位点, 推测 pSGL1 的正链切割位点在该结构中(图 7)。在 *rep* 上游 140~220 区有一回文结构, 该结构的自由能 ΔG 为 -35.23kcal mol⁻¹。比较该结构与其它链霉菌质粒的 SSI 功能区, 它的对应链中含有一个保守的序列 TCGCAT, 推测此结构为 SSI 功能区(图 8)。

		↓
pC194	CTTATCTTG	AT
φx174	CCCAACTTG	AT
pIJ101	AACACCTTG	CG
pSG5	CGCAGCATG	CC
pJV1	GACGCCTAG	CT
pSGL1	CCCGGCTTG	TA

图 7 质粒 pSGL1 与其相关质粒的切割位点比较

Fig. 7 Analysis of nicking site of the pSGL1 and the related plasmids
↓ Nicking site; _ conserved nucleotides.

GGCGGGGGCTTAGCCGACGCTACGGCTCACTC CT GCAAAACGAGACA
::: :: : : : : : : : : : : : : : :
CCGG CGG GCGGAGGTA CCGGGTTACGCTGATCTCGTTACGCTCGTC

图 8 质粒 pSGL1 的 SSI 功能区分析

Fig. 8 Analysis of SSI function of the pSGL1

::Pairing nucleotides; _ Conserved nucleotides.

3 讨论

链霉菌质粒 pSGL1 是从 *Streptomyces globisporus* 中分离的一新质粒。pSGLS3 含有它的基本复制区。链霉菌质粒如 pIJ101, pJV1, pSG5 等都以滚环机制进行复制^[5,6]。通过这些质粒的序列测定,根据复制必需区序列的相似性,复制酶的同源性和起始点的定位可将其归为四个家庭^[7]。本文对 pSGL1 基本的复制区进行序列测定,与已知的复制起点和复制酶进行序列分析比较后可以确定它的复制机制是滚环复制。其 REP 蛋白质与 pC194 家庭中 REP 蛋白质相似。在 REP 蛋白质读框中,由于测序造成三联密码在某些位置不能确定,因此用问号代替该处的氨基酸。在该复制必需区内推测有一个 REP 蛋白质编码区,有 SSI 功能区,ori 区。这些表明了含有复制必需区的质粒 pSGLS3 能够作为构建载体的基本骨架。

参 考 文 献

- [1] 李晓平,李元.中国抗生素杂志,1992,17(5):326~332.
- [2] 洪斌,李元.微生物学报,1998,38(4),256~260.
- [3] Sambrook J, Fritsch E F, Maniatis T. Molecular cloning: A Laboratory Manual, 2nd edition, New York: Cold Spring Harbor Laboratory, 1989, 1.38.
- [4] Wukkuan R S. *Nucleic Acids Research*, 1992, 20(5):961~974.
- [5] Luis Servin-Gonzalez. *Plasmid*, 1993, 30:131~140.
- [6] Gunther M, Martin F, Vera H *et al.* *Plasmid*, 1995, 33:113~126.
- [7] Tatyana V I, Eugene V K. *Nucleic Acid Research*, 1992, 20(13):3279~3285.

NUCLEOTIDE SEQUENCE ANALYSIS OF THE MINIMAL REPLICON OF THE *STREPTOMYCES* PLASMID pSGL1

Zhang Hua Hong Bin Li Yuan

(Institute of Medicinal Biotechnology, Peking Union Medical College,
Chinese Academy of Medical Sciences, Beijing 100050)

Abstract The high-copy-number plasmid pSGL1 (7.4kb) was isolated from *Streptomyces globisporus*. Deletion experiments showed its minimal replicon is located on the 2.0kb *Sau*3 AI fragment. This fragment was subcloned. DNA sequence data analysis showed this fragment is a new sequence. Only an open reading frame with high coding probability is located on the minimal replicon. The deduced protein contains motifs characteristic of replicase for rolling-circle replication.

Key words Plasmid, Minimal replicon, Rolling-circle replication, Replicase