

转座子 Tn917 的 DNA 序列计算机分析*

余健秀 庞义 邓日强 余榕捷 李建华

(中山大学昆虫学研究所、生物防治国家重点实验室, 广州 510275)

摘要 通过计算机分析转座子 Tn917 的全序列, 详细阐述了其物理图谱、结构功能及其转录调节机制. Tn917 的 5 个 ORFs 排列在同一条 DNA 链上, 且阅读方向都从左至右. ORF1-3 起始点的左侧翼排列有启动子序列和 Shine-Dalgarno 序列. ORF5 (编码转座酶) 和 ORF4 (编码拆分酶) 的转录方向是一致的, 翻译也紧密偶联在一起. 在 ORF3 和 ORF4 之间存在 1 个 *res* 位点, 与 Tn3 中的 *res* 位点基本同源. 翻译衰减的功能与 rRNA 甲基化酶 (由 ORF2 编码的 *erm* 基因的产物) 诱导有关, 在这个结构基因的左侧翼有 200 bp 的前导区域编码一个具调控功能的 36 个氨基酸组成的多肽 (由 ORF1 编码).

关键词 Tn917, DNA 序列, 计算机分析

分类号 Q 523

利用转座子 Tn917 及其衍生载体在枯草杆菌 (*Bacillus subtilis*, Bs) 或其它革兰氏阳性 (G^+) 菌中进行染色体突变、基因转移等已成为近年来改良菌种的一种新技术. 详细分析 Tn917 的 DNA 序列、转座作用的必要元素及其物理图谱等, 对更好地研究和利用转座子 Tn917 是十分必要的, 本文正从这方面进行了 DNA 序列计算机分析.

通过 Internet 与国际分子生物学数据库 GenBank 联网, 获得 DNA 序列. DNA 序列的分析采用 Dnasis 和 Pcgene 等软件进行, 绘图采用 Autocad 软件进行.

1 Tn917 与 Tn3 的物理图谱及转座特性比较

对 Tn917 全序列进行了计算机分析, 绘制了 Tn917 的细微结构 (图 1). 已经证明, Tn917 属 Tn3 (图 2) 家族^[1-3]. Tn917 全序列长 5 257 bp, 在同 1 条链上从 5'~3' 有 6 个 ORFs (open reading frames)^[3]. 1991 年 An 等^[4]证明先前所认为的 2 个开放读码框 ORF5 与 ORF6 实际上就是 1 个 ORF5, 且与 Tn3 类的转座酶 (Transposase) 基因 *tnpA* 有很高的同源性. 从图 1 可看出, 在 Tn917 全序列有 5 个 ORFs, Tn917 的 5 个 ORFs 排列在同 1 条 DNA 链上, 且阅读方向都是从左至右. ORF1 起始点的左侧翼有启动子序列区 “-35” (位置为 327~332), 及 “-10” (位置为 354~359) 和 1 个 Shine-Dalgarno 序列 (位置为 403~413). ORF1 的编码基因产物是 36 个氨基酸的前导肽 (leader peptide), 具衰减子功能. ORF2 是 *erm* 结构基因, 其基因产物 rRNA 甲基化酶 (rRNA methylase). ORF3

* 国家 863 计划 (863-401-03-01-01) 和广东省自然科学基金重点资助项目

收稿日期: 1997-04-02 余健秀, 男, 28 岁, 博士研究生

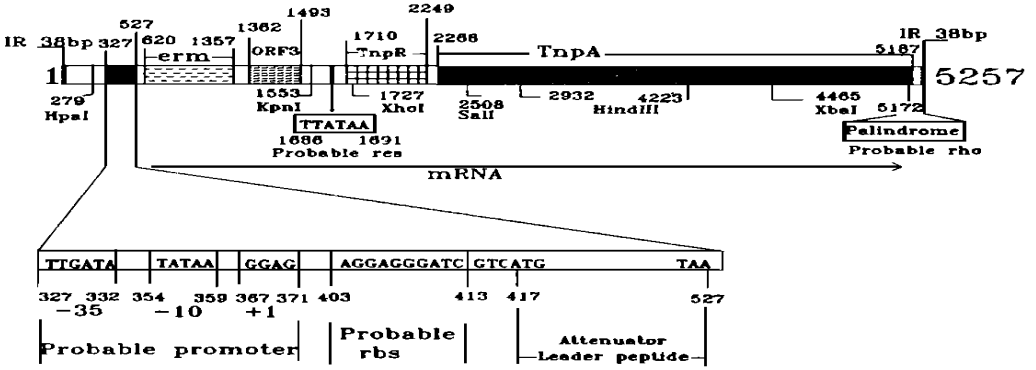


图 1 转座子 Tn917 的物理图谱及其结构

Fig. 1 Physical map and structure of the transposon Tn917

是 44 个氨基酸的多肽, 其功能未明. ORF4 的编码基因产物类似于 Tn₃ 的拆分酶 (resolvase) TnpR 的阻遏蛋白, 含 180 个氨基酸. ORF5 的编码基因产物类似于 Tn₃ 的转座酶 TnpA. Tn917 的 ORF5 和 ORF4 之间的间隔只有 17 bp, ORF5 的左侧翼没有 Shine-Dalgarno 结构, 所以 ORF4 和 ORF5 的转录和翻译是非常紧密偶联在一起的. ORF4-5 的左侧翼估计也有 1 个潜在的启动子序列. ORF1~3 的左侧翼有末端重复序列 (1~73), 另也存在 1 个几乎与之相同的内部同向重复序列 (1457~1529). Tn917 的另一端即 3 末端有 1 个倒转重复序列 (5220~5257), 与左侧 5' 末端和内部重复序列前 38 个碱基相对称.

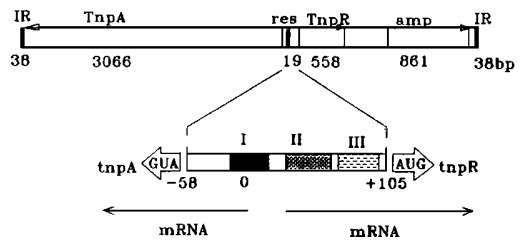


图 2 转座子 Tn₃ 的细微结构

Fig. 2 Structure of the transposon Tn₃ (转引自孙乃恩等, 1990)

一般转座过程分 2 个阶段, 第 1 阶段共合体 (cointegrate) 形成; 第 2 阶段共合体拆分 (resolution of cointegrate). 每个阶段由转座子上的 1 个基因所编码的酶起催化作用, 其中 tnpA 基因编码转座酶, tnpR 编码拆分酶. 互补试验表明, 这 2 个基因的作用是反式的 (trans-acting). Tn917 的末端倒转重复 (inverted repeat) 序列长 38 bp (图 3), 2 个末端重复序列十分接近, 其中任何一个的缺失都会顺式 (cis-acting) 地妨碍转座功能. 转座酶能够与 DNA 单链结合, 它是识别并切割转座子的两端以及在靶点序列 (“ATATA”) 上造成 5 bp 的参差不齐单链末端所必需的, 同时转座酶对末端倒转重复序列的识别是高度特异性的.



图 3 Tn917 末端倒转重复序列比较

Fig. 3 Comparison of the terminal inverted repeat sequences of Tn917

Tn₃ 的转座酶由 1021 个氨基酸组成, 而 Tn917 的转座酶由 974 个氨基酸组成. 在 Tn₃

中, tnpR 基因编码的拆分酶含 185 个氨基酸, 它除具有拆分酶活性外, 还有调控 tnpA 基因及 tnpR 基因转录的阻遏蛋白活性, tnpA 和 tnpR 基因之间隔以 163 bp 富含 A-T 的调控区域, 2 个基因的转录从这里起始, 朝相反的方向进行 (图 2). TnpR 正是通过与这段调控区域的结合而实现对 tnpA 和 tnpR 转录的阻遏. TnpR 结合在 res (resolution sequence) 位点而造成切口, DNA 的磷酸二酯键断裂所释放的能量被储藏在蛋白质-DNA 键中, 这些能量供应 1 个 res 位点和另一位点重新连接起来的需要. 拆分酶的识别位点的核苷酸顺序是 5'-TTATAA-3'; 2 条链的断裂点相隔 2 bp, 由此产生参差不齐的粘性末端, 其中 5 端各与 1 个分子 TnpR 共价结合 (图 4). 在 Tn917 中, tnpA 和 tnpR 之间间隔仅有 17 bp, 且 tnpR 和 tnpA 的转录方向一致和翻译紧密偶联在一起, 因此类似与 Tn3 上的 tnpR 和 tnpA 之间的 res 位点结构区域, 是不可能存在于 Tn917 上的 tnpR 和 tnpA 之间的. 通过计算机分析 Tn917 的 DNA 全序列, 最有可能的 res 位点结构区域在 ORF3 和 ORF4 (tnpR) 之间的 208 bp (1 493~ 1 710). 1984 年 Perkins 等^[2]证明了 208 bp 区域内的 kpn 位点 (1 553) 上如果插入含有 cat 基因外源 DNA 片段, 会影响复合体的分离. 图 1 也表明, Tn917 中也存在与 Tn3 相同的拆分酶识别序列“TTATAA” (位置为 1 686~ 1 691). 虽 Tn917 的 res 位点区域与 Tn3 所处位置不同, 但其核心识别序列基本相同 (图 5). Tn917 的 208 bp 区域内还存在 1 个对称的倒转复序列区 GTACCATAAACGACCGTTTATGGTAC (位置为 1 549~ 1 573), 此区域富含 A-T, 经计算达 75.9%.



图 4 Tn917 的 TnpR 的作用方式

Fig. 4 Action of the Tn917 TnpR

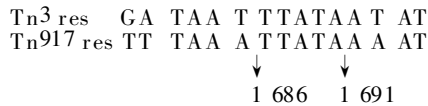


图 5 内部拆分子列 (res) 比较

Fig. 5 Comparison of internal resolution sequences

Tn91 的 res 位点区域位于偶联的 ORF4-的操纵子调控转录区域内. Tn3 上的 3 个结合点长 30~ 40 bp, 且彼此有一定的同源性, 但分析 Tn91 的 208 bp 的同源性时, 未发现类似现象, 因此认为 Tn91 的 208 bp 不存在类似的 Tn3 中有 3 个结合位点, 但肯定存在 TnpR 的结合位点的关键区域, 而其它区域对拆分酶和转录阻遏也有功能上的影响. 由于拆分酶识别序列非常靠近 ORF4, 仅隔 18 bp, 估计结合位点的关键区域也正是 ORF 转录的调控区域, 基本上重叠在一起. TnpR 与结合位点的结合, 形成了以 TnpR 为媒介的拆分, 同时, 这结合妨碍了 tnpR (ORF4) 和 tnpA (ORF5) 基因的转录. 所以, 虽然 Tn91 与 Tn3 的 res 位点所处位置不一样, 但因 tnpR 和 tnpA 转录方向的一致性和偶联性, TnpR 蛋白同样实现了对 tnpR 和 tnpA 转录的阻遏调控.

2 Tn91 的转录终止子

根据转录终止子 (transcriptional terminator) 在基因上所处位置的不同, 可以把终止信号分为 2 种类型: ① 存在于基因的末端, 转录到此形成 1 个完整的 mRNA, 这类终止信号不具适应环境的调节功能. Tn91 中, 在 tnpA 基因末端序列连同远 erm 端倒转重复序列构成转录终止子. ② 存在于基因的内部, 它的终止功能受环境因素的影响, 在不需要该基因产物的环境中, 终止子发挥作用, 形成不完整的 mRNA, 降低基因的表达. 这类终止信号特称为弱化子 (attenuator), 又称衰减子. Tn91 存在上述 2 类转录终止子.

Tn91 存在前导肽衰减子. Tn91 的 *erm* 基因的操纵子包括 1 个 36 个氨基酸组成的前导肽, 定位于 *erm* 结构基因的 5' 端, 这个多肽是由 ORF1 编码的, 而 *erm* 基因由 ORF2 编码的, 两者相隔 91 bp. 这个前导肽被认为具有翻译衰减 (translational attenuation) 的功能. Horinouchi 等^[5]认为 Tn91 的前导区包括前导肽的一半 C 端部分 DNA 序列可以形成潜在的非依赖 rho 蛋白的转录终止子的回文结构 (palindrome), 前导肽的翻译在一定程度上影响了回文结构不同形式的组合, 从而控制结构基因即下游基因的转录. 我们用计算机分析了前导肽序列区 (位置为 417~ 527), 结果也证实了这个设想. 表 1 是 Tn91 前导肽序列区等 (417~ 527) 形成的回文结构.

表 1 Tn91 前导肽序列区等 (417~ 527) 形成的回文结构

Tab. 1 Palindrome site reference (normal 417~ 527)

长度	AGE	自由能 ¹⁾	位置	序列
23	64	- 12.56	438	5' A AT GTAGAT AA 3' T GT CATCTA CA ^A
21	60	- 2.51	445	5' AT AAAAC ATC 3' AG TTTTG TCA ^A
37	61	-	451	5' C A TCT ACTGTTTT GAAA C ^A 3' A T TAG TGACAAAA ATCA G ^A
21	60	-	470	5' G AC TAA AA AC ^A 3' C GC ATT AG TG ^A
20	60	-	485	5' T T ACG C A GA T) 3' A T TGC A T AA A)
51	60	- 23.03	492	5' G A T AAA TA C G T T AGATT AA T TC CT A _C 3' A T T TTT CA G T A T TCTAA TC A GT GA C

1) 单位是 kJ/mol

在 Tn91 中, 红霉素抗性的产生和转座作用能同步诱导^[1-3], 是由于存在翻译衰减多肽编码序列, 当存在低浓度 Em 且高温诱导时, *tnpA*, *tnpR*, *erm* 基因可能在同一方向转录而表达. rRNA 甲基化酶是 *erm* 基因编码的, 其表达可以被亚选择水平浓度的 Em (或 MLS) 诱导. 当不存在 rRNA 甲基化酶时, 则导致核糖体 RNA 与大环内酯相结合, 使得核糖体停留在前多肽的某一序列区, 前导肽不能顺利翻译, 这时核糖体停留在 mRNA 上不再前进, 这个事件可能便是导致下游包括 *erm*, *tnpR*, *tnpA* 基因的转录. 由表可知, 前导区转录物 11 个核苷酸 (位置为 417~ 527) 的顺序, 可以形成种种不同的茎-环结构. 由于 1 个核糖体一般可覆盖 20 个核苷酸, 当 *erm* 基因得到表达之后, 即有 rRNA 甲基化酶时, 在 rRNA 的某些腺嘌呤上甲基化, 使得大环内酯不能与核糖体相结合, 前多肽能顺利合成, 但下游基因的转录不能顺利进行而不表达.

从图 3 可看出, Tn91 远红霉素基因端 (*erm*-distal) 向内的 DNA 序列, 与近红霉素基因端 (*erm*-proximal) 向内的 DNA 序列, 这 2 个末端倒转重复序列 38 bp 中有 4 个长度不一样, 这细微差别具有生物学的意义. 远 *erm* 端可形成回文结构 (palindromic structure), 有 43 个碱基形成茎-环 (stem-loop) 结构, 其中有 18 对碱基自身配对成“茎”, 余下 7 个碱基形成不配对的“环”. 在“茎”之外的 3' 端, 有短的富含 T 的序列区. 但不是紧接着就是 T 碱基. 整个结构有点类似大肠杆菌的非依赖 rho 蛋白的转录终止子. 按照 Tinoco 等 (1977) 自

由能的计算方法, 转录形成 RNA 回文序列在形成茎-环结构时, 其 $\Delta E = -125.60 \text{ kJ/mol}$. 如果远 *erm* 基因端的倒转重复序列与近 *erm* 基因端没有 4 个碱基差异的话, 在形成茎-环结构时, ΔE 只有 -66.99 kJ/mol .

参 考 文 献

- 1 Youngman P J, Perkins J B, Losiek R. Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. *Proc Natl Acad Sci USA*, 1983, 80: 2305 ~ 2309
- 2 Perkins J B, Youngman P J. A physical and functional analysis of Tn917, a *Streptococcus* transposon in the Tn3 family that functions in *Bacillus*. *Plasmids*, 1984, 12: 119~ 138
- 3 Shaw J H, Clewell D B. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in *Streptococcus faecalis*. *J Bacteriol*, 1985, 164: 782~ 796
- 4 An F Y, Clewell D B. Tn917 transposase sequence correction reveals a single open reading frame corresponding to the *tnpA* determinant of Tn3-family elements. *Plasmid*, 1991, 25: 121~ 124
- 5 Horinouchi S, Byeon W H, Weisblum B. A complex attenuator regulates inducible resistance to macrolides, lincosamides and streptogramin type B antibiotics in *Streptococcus sanguis*. *J Bacteriol*, 1983, 154: 1252~ 1262

Analysis on DNA Sequence of the Transposon Tn917 with Computer

Yu Jianxiu* Pang Yi Deng Riqiang Yu Rongjie Li Jianhua

Abstract A physical map, structure and function as well as mechanism of transcriptional regulation of the transposon Tn917 were described in detail, employing computer analysis on its sequences. The transposon appears to have five significant ORFs, all of which occur on the same strand and are read from left to right. A consensus promoter site precedes a Shine-Dalgarno sequence that is close to the start site of ORF1-3. The translation of ORF5 (coding Transposase) and ORF4 (coding Resolvase) is coupled or closely coordinated. The Tn917 has a site between ORF3 and ORF4 homologous with *res* site of Tn3. The translational attenuation associated with the control of rRNA methylase (product of *erm* gene, corresponding ORF2) induction, within the 200 bp leader region preceding the rRNA methylase structural gene is a control peptide 36 amino acid residues (corresponding ORF1) in length.

Keywords Tn917, DNA sequence, computer analysis

* Institute of Entomology and State Key Laboratory for Biocontrol, Zhongshan University, Guangzhou 510275