

黑曲霉糖化酶高产和低产菌株糖化酶基因 调控区的克隆及其分析比较*

钟丽婵 乔殿华 唐国敏 杨开宇

(中国科学院微生物研究所 北京 100080)

摘要 以 PCR 合成的糖化酶高产菌株黑曲霉 (*Asp.niger*) T21 糖化酶基因 5' 近端非编码区 588bp (EcoRI-BamHI) 的序列为探针, 从 T21 染色体 DNA 中克隆到近 2.0kb 的糖化酶基因 5' 端非编码区序列, 并以此序列为探针从糖化酶低产菌株黑曲霉 3.795 (T21 的诱变出发株) 的染色体 DNA 中克隆到 1.5kb 的糖化酶基因 5' 端非编码区序列。该二序列的分析测定结果表明, 其结构特征与文献报道的黑曲霉糖化酶基因 5' 端非编码区的基本一致, 被称为“核心启动子” (Core promoter) 的 TATAAAT 框及 GCAAT 框, 分别在翻译起始点的 -109bp 及 -178bp 处。此外, 在曲霉 *amdS*, *amyB* 基因中已发现有调控功能的 CCAAT 序列存在于 -449bp 和 -799bp 处。高产和低产菌株糖化酶基因 5' 端非编码区序列的分析比较结果表明, 有 9 个部位的碱基发生了变化。此实验结果为进一步研究黑曲霉糖化酶基因在转录水平上的调控规律打下了基础。

关键词 黑曲霉, 糖化酶, 5' 端非编码区

黑曲霉糖化酶是诱导酶, 淀粉能诱导酶的形成。已有的研究表明, 培养基中淀粉浓度与糖化酶活力及糖化酶 mRNA 含量呈正相关。黑曲霉糖化酶高产菌株与低产菌株的酶活力与糖化酶 mRNA 含量比也有一致的对应关系。因此, 与已研究的其他丝状真菌^[1]一样, 黑曲霉糖化酶的基因表达也主要在转录水平上受调控。黑曲霉糖化酶基因启动子是丝状真菌中已知的几个强启动子之一, 因而是目前丝状真菌表达载体使用最多的启动子之一。通过黑曲霉糖化酶高、低产菌株糖化酶基因 5' 端非编码区序列的比较及进一步的研究不仅可发现重要的基因顺式调控元件, 为深入研究参与糖化酶基因转录的蛋白质因子及最终阐明其转录水平上的调控规律奠定基础, 同时还可直接指导构建外源蛋白高效表达载体和工业丝状真菌育种。因此, 获得黑曲霉糖化酶基因 5' 端非编码区序列是必要的第一步。前文^[2]已报道黑曲霉糖化酶高产菌株糖化酶结构基因的分离及序列测定, 本文报道黑曲霉糖化酶高、低产菌株糖化酶基因 5' 端非编码区序列的克隆及两序列的分析比较结果。

1 材料和方法

1.1 材料

1.1.1 菌株及质粒: 黑曲霉 (*Asp.niger*) T21 为中国科学院微生物研究所糖化酶研究小组

* 国家自然科学基金资助项目。

本文于 1995 年 4 月 17 日收到。

经多次诱变所获得的糖化酶高产菌株;黑曲霉 3.795 为本组保存的糖化酶低产菌株,是 T21 的诱变出发株;大肠杆菌 (*E.coli*) DH₅ 为本组保存,用作宿主菌;质粒 pGlap5 含有 PCR 合成的黑曲霉糖化酶基因 5' 近端非编码区序列,由本组构建^[3],用于制备 Southern 印迹分析及原位杂交所用之探针;质粒 pUC19 用作克隆及序列分析中亚克隆的载体。

1.1.2 培养基:细菌生长培养基为 LB 培养基。黑曲霉 T21 及 3.795 摇瓶培养基同前文^[2]。

1.1.3 酶及其他试剂:限制性内切酶、连接酶、碱性磷酸酯酶为 Boehringer Mannheim, Promaga, New England Biolabs, 华美生物工程公司、中国医学科学院基础医学研究所友谊生物制品公司产品。切刻平移试剂盒为 Promaga 公司产品。DNA 序列分析试剂盒 (T₇ Sequencing kit) 为 Pharmacia 产品。同位素 ³²P 及 ³⁵S 为 Dupont 公司产品。

1.2 方法

1.2.1 染色体 DNA 制备:参照文献[4]所述方法,用氯化苄提取。菌体培养及菌丝体收集方法同前文^[2]。

1.2.2 质粒制备、DNA 片段分离回收、DNA 片段脱磷、DNA 连接、转化、Southern 印迹分析及原位杂交均按文献[5]所述进行。

1.2.3 核酸序列分析:以双链 DNA 为模板,按 T₇ Sequencing kit 操作程序进行。

2 结果和讨论

2.1 Southern 印迹分析

已分离到黑曲霉糖化酶高、低产菌株糖化酶基因从翻译起始密码 ATG 到 5' 端上游 EcoR I 位点间 204bp 的旁侧序列^[2], 本工作拟分离 EcoR I 位点上游更长片段的 5' 旁侧序列。以 pGlap5 质粒的 EcoR I - BamH I 片段 (588bp)^[3] 为探针, 对 EcoR I 完全酶切的黑曲霉 T21 染色体 DNA 作 Southern 印迹分析, 结果表明 ~2.0kb 的 EcoR I - EcoR I 片段含有黑曲霉糖化酶基因 5' 旁侧序列 (图略)。同样, 黑曲霉 3.795 的染色体 DNA Southern 印迹分析结果表明, 1.8kb 的 xho I - EcoR I 片段含有黑曲霉糖化酶基因 5' 端上游序列 (图略)。

2.2 黑曲霉 T21 及 3.795 糖化酶基因 5' 端非编码区序列的克隆

黑曲霉 T21 染色体 DNA 用 EcoR I 完全酶切, 电泳分离, 回收 1.5kb ~ 2.3kb 片段, 与载体 pUC19 连接, 转化大肠杆菌 DH5, 获得转化子, 以 pGlap5 质粒的 EcoR I - BamH I 片段为探针, 经原位杂交从转化子中筛选出 3 个阳性克隆。用同样方法, 以已克隆到的 T21 糖化酶基因 5' 端非编码区序列的 Xho I - EcoR I 片段为探针, 从黑曲霉 3.795 的染色体 DNA 中克隆到 1.5kb 的糖化酶基因 5' 端非编码区序列。所获得的黑曲霉 T21 及 3.795 糖化酶基因 5' 端上游序列的限制性内切酶图谱与 Fowler 等^[6] 报道的 *Asp.niger* ATCC10864 的基本一致 (图 1)。

2.3 黑曲霉 T21 及 3.795 糖化酶基因 5' 端非编码区序列的测定

图 2 为测定的近 2.0kb 的黑曲霉 T21、1.5kb 的黑曲霉 3.795 及 Fowler 等^[6] 报道的近 2.0kb 的黑曲霉 ATCC10864 糖化酶基因 5' 端非编码区的序列。三者都在 -46bp 及 -70bp 处有二个可能的转录起始点 CAAC 及 GAAG^[7] (ATCC10864 为 GAAAG); 丝状真菌中可能起确定转录起始点作用的 TATAAAT 框位于 ATG 上游 109bp。类似真核

基因上游启动子元件 UPE (Upstream Promoter Elements) 的 GCAAT 序列在 ATG 上游 178bp 的位置, 紧靠二个转录起始点的上游分别有一个 CT 富含区 CTTCTTCCTTC 和 CTTCTCCCTTTT (ATCC10864 为 CTTCTCCCCCTC)。被认为可能与糖化酶基因高表达及淀粉诱导有关的序列 GGGGACGGCGAATCCCCGGG^[8] 存在-425bp 至-406bp。

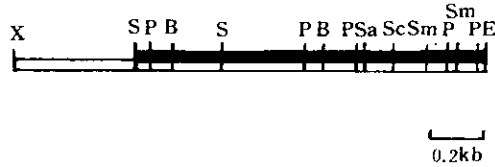


图1 黑曲霉 T21 及 3.795 糖化酶基因 5'端非编码区的限制性内切酶图谱

Fig.1 Restriction maps of the 5'-flanking regions of the glucoamylase-encoding gene

(*glaA*) from *A. niger* T21 and 3.795

□ 黑曲霉 T21 糖化酶基因 5'端非编码区

The 5' upstream region of the *A. niger* T21 *glaA* gene:

■ 黑曲霉 3.795 糖化酶基因 5'端非编码区

The 5' upstream region of the *A. niger* 3.795 *glaA* gene.

Abbreviations: X, XhoI; S, Sall; P, PstI; B, BamHI; Sa, SacI; Sc, ScaI; Sm, SmaI; E, EcoRI.

Littljohn 和 Nagata 分别发现 CCAAT 序列在曲霉 *amdS* 基因和 *amyB* 基因表达中具有调控功能^[8]。该序列在黑曲霉糖化酶基因中存在于-449bp 及-799bp 处。在-381bp 还有一个 TCAAT 序列, 该序列存在于许多曲霉及其他丝状真菌基因中, 对基因表达可能有调控的功能。此外, 在-472bp 有一个类似 TATAAAT 框的序列 TATAAT。要确定以上各序列在黑曲霉糖化酶基因表达调控中是否具有重要性及其功能作用, 有待进一步研究。

黑曲霉糖化酶低产菌株经诱变至高产菌株后, 其酶活力增加至原来的 15 倍。迄今为止, 黑曲霉糖化酶基因的表达调控被认为主要是在转录水平上的调控, 与高水平表达有关的序列在-800bp 之内^[6,9]。比较黑曲霉糖化酶高产菌株 T21 与低产菌株 3.795 的糖化酶基因 5'端非编码区序列可以看到, 两者自-204bp 至 ATG 的序列完全一致, 在-660bp 至-348bp 有 3 个部位显示碱基改变, 而在-874bp 至-1478bp 还有 6 个部位的碱基发生了变化。黑曲霉 3.795 诱变至 T21 后, 其糖化酶活力的增加从转录水平上的调控来分析, 或是因糖化酶基因顺式调控元件碱基突变或是因调节基因突变引起了调节蛋白与 DNA 结合位置或状态的改变。据 Fowler 等的报道, *A. niger* ATCC 10864 是糖化酶高产菌株。将上述黑曲霉 T21 和 3.795 糖化酶基因 5'端非编码区序列中碱基发生了变化的 9 个部位与 *A. niger* ATCC 10864 的相应部位比较, 可以看到高产菌株 T21 和 ATCC10864 的碱基有 5 个部位是一致的, 即相对低产菌株而言, 二者都发生了相同的碱基突变。这 5 个部位的碱基突变可能是引起酶活力增加的原因。其余 4 个部位只有 T21 发生碱基改变, ATCC10864 则与 3.795 一致。因此, 通过将黑曲霉 T21 与 3.795 糖化酶基因 5'端非编码区序列中发生碱基变化的相应序列进行互换及缺失分析试验, 就有可能鉴别出重要的调

1
2 CTCGAGGATTGTCTGAACATGACATTCCGGGCCACGGCCGACC ACTCGGACGCGAAGTGCCTGCTGTCT GGATCTTTGGCGGAGGCTTTGAAC -1875
3 CTCGAGGATTGTCTGAACATGACATTCCGGGCCACGGCAACCCAACTCGGACCGGAA TCCCGTCTGCTGTCTCGGATCTTTGGCGGAGGCTTTGAAC

1
2 TGGTTCAAAGGCCATGTATGACGGCACACCGATGGTATCATCGTCGATAGACAAGAATAGCCTATCGTGTGGTAGCGATGAACATCCGGTAGGGGGCTT -1773
3 TGGTTCAAAGGCCATGTATGACGGCACACCGATGGTATCATCGTCGATAGACAAGAATAGCCTATCGTGTGGTAGCGATGAACATCCGGTAGGGGGCTT

1
2 CGGGTTTCTCCCGGAAAGGAGATTCTGGAGGACGGGTCCGCCAACTTAGTCTACT ACCAACCGCTT CCCTCAGTGGGTGCCGACAACATCGAGGGC -1673
3 CGGGTTTCTCCCGGAAAGGAGATTCTGGAGGACGGGTCCGCCAACTTAGTCT TTAGACCA GCCTTGCCT AGTGGGTGGCGACAACATCGAGGGC

1
2 TTTGGTGAGACCCAGACAAGGTGACAATCTGGGAGAATCAGCAGGGGCTATTTCTGTCTTTGATCAGATGATCTTGTACGACGGAAACATCGCTTACAAG -1571
3 TTTGGTGAGACCC AGACAAGGTGACAATCTGGGAGAATCAGCAGGGGCTATTTCTGTCTT GATCAGATGATCT GTACGACGGAAACATCGCTTACAAG

1
2 GA AAGCCCTTGTTCGGGGAGCCATCATGGACTCCGGTAGTGTGTTCCCGCAGACCCCTGTCCGACGGGGTCAAGGGA CAGCAAGTATATGATG CGGTTG -1472
3 GACAAGCCCTTGTTCGGG AGCCATCATGGACTCCGGTA TG GTTCCCGCAGACCCCTGTCCGACGGGGTCAAGGGATCAGCAAGTATATGATG CGGTTG

1
2 TGGACTCTGCAGGCTGTTCCCTTCCAACGACACCCCTGGCTGTCTGCGTGACGTAGACTACACCGACTATCTCAATGGCGGAAACTCCGTGCCGGGATCC -1370
3 TGGACTCTGCAGGCTGTTCCCTTCCAACGACACCCCTGGCTGTCTGCGTGACGTAGACTACACCGACTATCTCAATGGCGGAAACT CGTGCCGGGATCC

1
2 TAGGTTATCACTCCGTGGCGCTATCATATGTGCCTCGACAGACGGGACGGCATT GTCGGCGTCGCAG TTTTGGTAAAGCAGGGAAGTATGCGCGG -1270
3 TAGGTTATCA CCGTGGCGCTATCATATGTGCCTCGACAGACGGGACGGCATTGTGCGCGTCGCAG TTTTGGTAAAGCAGGGAAGTATGCGCGG

1
2 GTCCCATTCATCGTGGGGACCAAGAGGATGAGGGGACCTTGTTCGCCTGTTTCAGTCCCTTACGACGATCGACGAGGTAGTGCAGTATCTGGGCACCTACT -1168
3 GTCCCATTCATCGTGGGGACCAAGAGGATGAGGGGACCTTGTTCGCCTGTTTCAGTCCCTTACGACGATCGACGAGGTAGTGCAGTATCTGGGCACCTACT

1
2 TCTTCTATGACGCTAGCCAGGAGCAGCTTGAAGAATTAGTGGCCCTGTACCCAGACACCACACATATGGGTCTCCCTCAGGACGGGCACGGCCAAACCT -1066
3 TCTTCTATGACGCTAGCCAGGAGCAGCTTGAAGAATTAGTGGCCCTGTACCCAGACACCACACATATGGGTCTCCCTCAGGACGGGCACGGCCAAACCT

1
2 GGTATCCGCAATTTAAGCGATTGGCCGCATTCTCGGGCACTTGGTCTTACCATTACCCGGCGG CATTCTGTCTATGACAGAGGAGCTCTCCCTGACC -965
3 GGTATCCGCAATTTAAGCGATTGGCCGCATTCTCGGGCACTTGGTCTTACCATTACCCGGCGG CATTCTGTCTATGACAGAGGAGCTCTCCCTGACC

1
2 TCCCGAAATGGTCGTAACCTGGCGACCTATGACTATGGCAGCCAAATCTGGGGACCTTCCATGGAAGTGACCTGCTGCAGGTGTTCTATGAGATCAAGCCGA -863
3 TCCCGAAATGGTCGTAACCTGGCGACCTATGACTATGGCAGCCAAATCTGGGGACCTTCCATGGAAGTGACCTGCTGCAGGTGTTCTATGGATCAAGCCGA

1
2 ACTATGACGCGAGTTCAGCCACACGTAATATCTGAGTTTTGTATACACCGTGGATCCGAACTCCAATCGGGGGAGTACATGGAATGGCCCAAGTGGCAGG -761
3 ACTATGACGCGAGTTCAGCCACACGTAATATCTGAGTTTTGTATACACCGTGGATCCGAACTCCAATCGGGGGAGTACATGGAATGGCCCAAGTGGCAGG

1
2 AATCAGCAGTGTGATGAATTTCCGAGGCAACAGCGCCAGTCTCCTTACCGATGATTTCCGCAACGGGACATAGAGTTATCCTGCAAGAAACCCGGCGCT -659
3 C CGACAGTGTGATGAATTTCCGAGGCAACAG CAGTCTCCTTACCGATGATTTCCGCAACGGGACATAGAGTTATCCTGCAAGAAACCCGGCGCT

1
2 TCCACATCTGATGCCATTGGCGG AGGGGTCCGGAGGTCAGGGAC TAGCCTTATGAACGTAATGATGGAAGTGTCTGGCCTCGGCAAAGGATATATAGGGT -558
3 TCCACATCTGATGCCATT CGGGAGGGTCCGGAGGTCAGGGAC TAGCCTTATGAACGTAATGATGGAAGTGTCTGGCCTCGGCAAAGGATATATAGGGT

1
2 CATAATAAGTAGTACTAGTTATATTAATGGAAGGGTATATACCACCGCTTGGACCTTGGGACCTGCATTATAGCTTCCCGTTAGGTATAAATACCCTGTGTTA -456
3 CATAATAAGTAGTACTAGTTATATTAATGGAAGGGTATATACCACCGCTTGGACCTTGGGACCTGCATTATAGCTTCCCGTTAGGTATAAATACCCTGTGTTA

1
2 TAGCAGCCAATCAAGCCACCACCGCTCGACCGGG GACGGCGAATCCCCGGGAATTGAAATAAATGCAATTCAGGTCAATCGCGCCAGCGATGGACACATC -355
3 TAGCAGCCAATCAAGCCACCACCGCTCGACCGGG GACGGCGAATCCCCGGGAATTGAAATAAATGCAATTCAGGTCAATCGCGCCAGCGATGGACACATC

1
2 TCCAAG CACAGGGCCATTCTCGAGTGGCGGTGATTGAGTCAATTTCCCGGGGGCGGGCCGACACCGCGATAGGCTGGTCTTCCACACCACCGGAGAT -253
3 TCCAAGGCACAGGGCCATTCTCGAGTGGCGG GGATACAGTGCA TTCCCGGGGGCGGGCCGACAC CCGATAGGCTGGTCTTCCACACCACCGGAGAT

```

1 TCGTCGTTCTGAAGAGCTGAAGTGGCGAGATGGTCTCTGCAGGAATTC AAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTG ATGCAT
2 TCGTCGTTCTGAAGAGCTGAAGTGGCGAGATGGTCTCTGCAGGAATTC AAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTG ATGCAT - 159
3 TCGTCGTTCTGAAGAGCTGAAGTGGCGAGATGGTCTCTGCAGGAATTC AAGCTAGATGCTAAGCGATATTGCATGGCAATATGTGTTGTTGATGCAT

1 ATGCATGTGCTTCTCTCCCTCAGCTTCCCC TCGTGCAGATG AGGTTTGGCTATAAAATTGAAGTGGTTGGTC GGGGTTCCGTGAGGGGCTGAA GT GCT
2 ATGCATGTGCTTCTCTCCCTCAGCTTCCCC TCGTGCAGATG AGGTTTGGCTATAAAATTGAAGTGGTTGGTC GGGGTTCCGTGAGGGGCTGAA GT GCT - 63
3 ATGCATGTGCTTCTCTCCCTCAGCTTCCCC TCGTGCAGATGTTGGTTTGGCTATAAAATTGACGTGGTCAGGGGT CAGTGAGGGGTGAAAGTAGCT

1 TCCTCCCTTT AGACG CAACTGAGAGCCTGAGCTTCATCCCAGCATCATTACACCTC A GCAATG
2 TCCTCCCTTT AGACG CAACTGAGAGCCTGAGCTTCATCCCAGCATCATTACACCTC A GCAATG +3
3 TCCTCCCTTCAAGACGCGCAACTGAGTGACTGAGCTTCATCCTCAACATCGGTACTCCCCACCGCAATG

```

图2 黑曲霉 T21(2)、3.795(1)及 ATCC10864(3)糖化酶基因 5'端非编码区的序列

Fig.2 Sequences of the 5' upstream regions of the *A.niger*
T21(2), 3.795(1) and ATCC10864 *gliaA* gene

翻译起始编码子定为+1;黑线表明文中所述的特异性序列;黑曲霉 T21 与 3.795 两者糖化酶基因 5'端非编码区序列之间的不同碱基用↓表示。

The translation start codon is numbered +1; Special sequences as described in the present paper are underlined; The different nts between the 5' upstream regions of the *gliaA* gene of *A.niger* T21 and 3.795 are indicated by ↓.

控元件,为进一步研究参与转录的蛋白质因子及阐明转录机制打下基础,同时还可指导实践,构建丝状真菌高效表达载体,用于工业丝状真菌育种。

参 考 文 献

- [1] Unkles S E. Gene Organization in Industrial Filamentous Fungi. In: Kinghorn J R, Turner G ed. Applied Molecular Genetics of Filamentous Fungi. Glasgow: Blackie Academic & Professional, 1992. 47.
- [2] 钟丽婵,唐国敏,杨开宇.微生物学报,1994,34(3):184~190.
- [3] 唐国敏,汤文剑,钟丽婵,等.生物工程学报,1996,12(2):178.
- [4] 朱衡,翟峰,朱立煌,等.真菌学报,1994,13(1):34~40.
- [5] Sambrook J, Fritsch E F, Maniatis T. Molecular Cloning. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989.
- [6] Fowler T, Berka R M, Ward M. *Curr Genet*, 1990, 18: 537~545.
- [7] Unkles S E. Gene Organization in Industrial Filamentous Fungi. In: Kinghorn J R, Turner G ed. Applied Molecular Genetics of Filamentous Fungi. Glasgow: Blackie Academic & Professional, 1992. 32.
- [8] Mackenzie D A, Jeenes D J, Belshaw N J et al. *Journal of General Microbiology*, 1993, 139: 2295~2307.
- [9] Verdoes J C, Punt P J, Stouthamer A H et al. *Gene*, 1994, 145: 179~187.

**CLONING, SEQUENCING AND COMPARISON OF THE 5'
FLANKING REGIONS OF glaA GENE FROM HIGH AND LOW
GLUCOAMYLASE-PRODUCING STRAINS OF
ASPERGILLUS NIGER**

Zhong Lichan Qiao Dianhua Tang Guomin Yang Kaiyu

(Institute of Microbiology, Academia Sinica, Beijing 100080)

Abstract A 2kb DNA fragment of the 5' flanking region of glaA gene was cloned from the chromosomal DNA of *Aspergillus niger* T21, a glucoamylase over producing strain, using a 588bp fragment (EcoR I - BamH I) of the 5' proximal flanking region as a probe, which had been synthesized previously by PCR from the chromosomal DNA of T21 covering EcoR I - BamH I stretch. The cloned 2kb DNA fragment was then used as the probe for cloning 1.5kb of the 5' flanking region of glaA gene from the chromosomal DNA of *Asp. niger* 3.795, a low glucoamylase-producing strain from which *A.niger* T21 was selected through mutation. Sequence analysis showed that both the 5' flanking regions of glaA gene contained the "core promoter" of TATAAAT box and GCAAT sequence at nt -109 and -178 from the start codon respectively. In addition, two CCAAT sequences existed at nt -799 and -449, which had been proven to play a functional role for expression and regulation of *Aspergillus* amdS and amyB gene. From comparison of the 5' non-coding regions of glaA gene from high and low glucoamylase-producing strains, nine sites with different bases were found. These results would prove to be valuable for further study on the regulation of transcription of the *Aspergillus niger* glaA gene.

Key words *Aspergillus niger*, Glucoamylase, 5' non-coding region