

巨大芽孢杆菌 α -淀粉酶基因核苷酸序列分析*

李 隽 蒋如璋**

(南开大学生命科学学院生化及分子生物学系 天津 300071)

摘 要 巨大芽孢杆菌(*Bacillus megaterium*)AS1.127 的淀粉酶基因的全碱基序列已被测定。结构基因由 1982bp 的单一开读框架组成。由 DNA 序列推测出的前体酶蛋白由 659 个氨基酸组成, N-端 33 个氨基酸为信号肽。成熟酶分子由 626 个氨基酸组成, 分子量为 68.676kD。该淀粉酶属糖化型 α -淀粉酶。并与枯草杆菌(*B. subtilis*)168 产生的糖化型 α -淀粉酶之间有 83.3% 的同源性。分析发现两种菌产生的酶分子的 N-端 3/4 的同源性为 90.4%, 而 C-端 1/4 的同源性只有 70%。序列排比结果说明在淀粉酶基因的趋异进化过程中, 基因突变和遗传重组都曾起过作用。

关键词 α -淀粉酶基因, 巨大芽孢杆菌

分类号 Q74 **文献标识码** A **文章编号** 1000-3061(1999)01-0068-74

α -淀粉酶(EC3.2.1.1)具有广泛的工业用途, 是 *Bacillus* 属许多菌种产生的胞外酶之一, 它以内切方式水解淀粉底物的 α -1,4 糖苷键, 依终产物的不同可将它们分为糖化型与液化型两类。液化型 α -淀粉酶能将淀粉快速液化, 其终产物为寡聚糖和糊精。而糖化型 α -淀粉酶有较强的酶切活性, 在水解可溶性淀粉时, 随水解时间的延长而产生寡聚糖、麦芽糖直至葡萄糖。前者的代表产生菌为 *B. amyloliquefaciens* (*B. a*)^[1] 和 *B. licheniformis* (*B. l*)^[2], 而后的代表产生菌为 *B. subtilis* (*B. s*)^[3]。

近年来, 已有多种 *Bacillus* 菌种的 α -淀粉酶基因已被相继克隆和做过序列分析。通过对 11 种 α -淀粉酶一级结构的排比分析^[4]发现, 其中被分析的 6 种动物源的 α -淀粉酶之间的氨基酸序列的同源性为 80%~90%。而微生物与动物和植物产生的 α -淀粉酶的氨基酸序列之间的同源性则不超过 10%。但所有上述 11 种 α -淀粉酶中都有 4 个保守序列区, 推测这些保守区与其同底物的结合和(或)催化中心有关。

已有有关 *B. megaterium* (*B. m*) 产生胞外 β -淀粉酶^[5] 和 α -淀粉酶^[6] 的报道。但 *B. m* AS1.127 菌株产生的淀粉酶与报道的都不相同。我们克隆了其淀粉酶基因, 并对它做了结构基因的全序列分析, 结果表明无论是 DNA 序列, 还是由碱基序列推测出的蛋白质的一级结构都与 *B. s* 168 产生的糖化型 α -淀粉酶之间有很高的同源性, 二者之间的差异可以以趋异进化来解释。而与 *B. m* α -淀粉酶^[8] 之间的同源性很低。

* 国家高技术研究发展计划预研项目(No. 8388021)。

** 联系人。

收稿日期:1997-09-25, 修回日期:1998-01-20。

1 材料和方法

1.1 菌株、质粒及培养基

质粒 pBX92^[7]系 *E. coli* pAT153 的衍生质粒。它携带着 2.5kb 编码 *B. m* AS 1.127 的 α -淀粉酶基因的 DNA 片段; pBX101^[7]系 *B. subtilis* 衍生质粒 pBX96^[7]的亚克隆,携带 2kb 的缺失淀粉酶 C-末端编码区及后续序列约 0.5kb 的 DNA 片段。

噬菌体 M13mp18 用于 DNA 序列分析的载体。*E. coli* JM109 用于 M13 噬菌体的受体。

LB 培养基用于细菌培养。H 培养基^[8]用于 M13 噬菌体复制型和单链 DNA 的转化。Terrific Broth(ABI 公司推荐)用于 M13 单链 DNA 的制备。

1.2 DNA 序列分析

质粒提取采用碱法^[9]并经 CsCl-EB 平衡超速离心纯化。pBX101 DNA 以 *Pvu* II 酶切,并以外切核酸酶 *Bal*31 作不同时间酶解,用 DNA 聚合酶 I(Klenow)将 DNA 末端补平后,再用内切酶 *Eco*RI 切取,以得到不同长度的 DNA 目的片段,并插入噬菌体 M13mp18 的 *Sma*I/*Eco*RI 位点上。从而得到用于测序的重组体 M13 噬菌体。相继的亚克隆之间约相差 200~300bp。

由质粒 pBX101 切取 *Cla*I/*Eco*RI(约 300bp)片段,以及由 pBX92 的 0.9kb *d1Pvu* II-*Eco*RI 片段直接插入 M13mp18 的 *Eco*RI/*Sma*I 位点,用来测定 α -淀粉酶基因 3' 一端编码区及 α -淀粉酶基因的后续部分的 DNA 序列。

为核对 3'-端的部分 DNA 序列,采用 PCR 扩增测序法对编码 α -淀粉酶基因 C-末端的近 500bp(1633~2116)做了重复测序。所用引物分别为 5'CAATGATCAGCTGACGGT(1635~1650),和 5'CATTC TACAGATAATCAA(2116~2098)。

核苷酸序列测定采用 ABI 公司 370A 型 DNA 自动序列分析仪。按该公司推荐的方法操作。每次反应可读出约 500bp 核苷酸。因为每个梯度酶切片段之间相隔仅 200~300bp,因而可以在可读序列的重叠区内相互验证,保证了被测序列的精确性。

采用 PC/Gene 软件(InterlignGenetics Inc.)处理测序结果及蛋白质结构的推测和比较。

1.3 酶及试剂

限制酶、核酸外切酶、*Bal*31 和 T4DNA 连接酶购自中国医学科学院基础所或美国 New England Biolabs; DNA 聚合酶(Klenow)和 Taq-DNA 聚合酶购自华美公司; X-gal、IPTG、尿素、丙烯酰胺、甲叉双丙烯酰胺和过硫酸胺购自 BRL; 4 种荧光标记引物购自 ABI 公司; dNTP 购自 Pharmacia 公司。

2 结 果

2.1 *B. m* α -淀粉酶基因的 DNA 序列及酶的一级结构

图 1 示包含 α -淀粉酶基因的 DNA 碱基序列,全长为 2499bp。 α -淀粉酶基因的编码区为自 172~2154bp 的开读框架,编码区全长为 1982bp。起始密码子为 ATG。核糖体 SD 序列为 AATAAGGAG。终止密码子为 ATTAAT。随后相隔约 60bp 有一反向重复序列

GAATCACTCTGCCGGTGTGCTTTGATAGAGAGTGTGTGGTAAGTTC ⁻³⁵ AAATGTAAGCGTAAACATAATTA ⁻¹⁰ TCCAGTCTTCACAACAATTG	90
AAAGGAGGAAGCTGAAGAAAAGATAGAGGAATTTTTGACTCCGCAGTCAGTCTTCAAAAA ^{amyO} TCAAATAAAGGAGTCTCAAAAA ^{SD} TGTTAAA	180
AAACGATTCAAAACCTCTTTACTGCCGTTATTCGCCGGATTTTTATTGCTGTTCATTGGTTTTGTGACGGCCAGCGGGTGCAAAACGCT	270
K R F K T S L L P L F A G F L L L P H L V L S G P A G A N A	
GAAACTGCAAAACAAATCGAATGAGGTGACCGATTCAATCGGTCAAAAACGGGACCATCTTCATGCAATGGTCAATCAATACGTTA	360
E T A N K S N E V T D S S V K N G T I L H A W N W S F N T L	
ACACAAAATAGAAAGAGATTCGGTATGCGGGTTATGCAAGCAATTCAGACGCTCCCGATTAACCAAGTAAAGGAAGGCAACCAAGGAGAT	450
T Q N M K E I R D A G Y A A I Q T S P I N Q V K E G N Q G D	
AAAAGCATGCGAACTGGTACTGGCTCTACAGCGACATCGTACCAATTCGGCAACCGTACTTAGGCACTGAACAAGAAATTAAGGAC	540
K S M S N W Y W L Y Q P T S Y Q I G N R Y L G T E Q E F K D	
ATGTGTGACGGCCGAAAAGTATGCGGTAAAGTCATGTCGATCGCGTTGTCATCATACCACAGCGAATTAAGCCGCAATTTCTGAT	630
M C A A A A E K Y G V K V I V D A V V N H T T S D Y G A I S D	
GAGATTAAGCGTATTCAAACTGGACCCATGGAAACACACAAATTAATAATGGTCCGACCGATGGGACATCACTCAAAATGCAATGGCTT	720
E I K R I P N W T H G N T Q I K N W S D R W D I T Q N A L L	
GGCGTATGATGGAATACTCAGAACTACTGAGGTGCAAGCCTACCTGAAAGGTTTCTGGAAAGAGCATTGAATGACGGAGCAGACGGG	810
G L Y D W N T Q N T E V Q A Y L K G F L E R A L N D G A D G	
TTCCGCTATGATGCCCAAGCATATAGAGCTCCCGATGATGGGAATACGGCAACCAATTTGGCCGAATATCAACAAATACATCGGGG	900
F R Y D A A K H I E L P D D G N Y G S Q F W P N I T N T S A	
GAGTTCAAATACGGAGAAATCTGCAAGACAGCGCTCCAGAGATACGGCTTATGCGAATTAATGAATGTCAGCGCTTCAACTATGGG	990
E F Q Y G E I L Q D S A S R D T A Y A N Y M N V T A S N Y G	
CATTCATCAGATCCGCTTTAAAGAAATCGTAATCTGAGTGTGCGAATATCTCCCATTAATGCAATGATGTCGTCGGGACAAGTATGTC	1080
H S I R S A L K N R N L S V S N I S Y A S N I S V A D K L V	
ACATGGTGGAAATCAGATGATAGTATGCGAATGATGTAAGAGTCCACATGGATGAGTGAAGCATATTCGTTTAGGCTGGCCAGT	1170
T W V E S H D T Y A N D D E E S T W M S D D D I R L G W A V	
ATTTGGTCCCGCTCAGGAAGCAAGCCTCTTTTCTTTCCAGACCTGAGGGCCGAGGAAATGGTGAAGATTTCCCGGAAAAAGTCAAATA	1260
I G S R S G S T P L F F S R P E G G G N G V R F P G K S Q I	
GGAGATCCGGGAGCGCTTATTTAAAGCAGCGCATCACTGCGGTCAATCAATTTCAACAATGAATGGCCGGGACCGCTCAGGAATC	1350
G D R G S A L F K D Q A I T A V N Q F H N E M A G Q P E E L	
TCAATCCGAATGGCAACAATCAAAATTTATGAATACCGCGCTCAAAGCGTGTGCTGGCAAATGCAGGATCGTCTCTGTCACC	1440
S N P N G N N Q I F M N Q R G S K G V V L A N A G S S S V T	
ATCAATCTTCAACGAAATTAACCTGACGGCAGGTATGATAATAGGGCCGGCCGGTTCATTTCAAGTAGGCAACGGCAACTGACAGGT	1530
I N T S T K L P D G R Y D N R A G A G S F A G C K K L T G	
ACGATCAATGCCAGATCCCGCGTGTCTTTGCTCTGATGATATGGAAATGCCCTCATGTCTTCTTGAGAATACCAAAACAGAGGCA	1620
T I N A R S A A V L C P D D I G N A P H V F L E N Y Q T E A	
GTCCATCTTTCAATGATCAGCTGACGGTCAACCTCGGTGCAAAATGCGAAAACAAACAAAGCGGTTTACCAAAATCAATAATGGGCAGCAG	1710
V H S F N D Q L T V T L R A N A K T T K A V Y Q I N N G Q	
ACAGCATTTAAGGATGGAGCCGATTAAGCATCGGGAAGAAGATCCAATCGGCACGACATACAACGTCAAGTTAACCGGAACGACCGC	1800
T A F K D G D R L T I G K E D P I G T T Y N V P K L T G T N G	
GAGGGTGCATCGAAGCAAGAAATACAGCTTTGTCAAAAAAGACCGCTCCAAACCAACATCATTTGGTATCAAAAATCCGGATCATTGG	1890
E G A S R T Q E Y T F V K K D P S Q T N I I G Y Q N P D H W	
GGCAATGTAATGCTTATATTAACAACATGATGGAGCGGGCCATAGAATTAACCGGATCGTGGCCGGGAAAGCCATGACTAAGAAT	1980
G N Y N A I Y I Y K H D G G A I E L T G S V P G K A M T K N	
GCAGATGGAATTTACACTGACGCTGCTGCGAATGCGGATACGGCCGACGCCAAAGTATTTTAAACAATGGCAGCGCCCAAGTCCCC	2070
A D G I Y T L T L P A N A D T A D A K V I F N N G S A Q V P	
GCACAGAACATCCCGCTTTGATATGTCGCAATGCGTTTATATAACAACCTCTGGTTTGAATGGTTATCTTCCGCATTAATGAATAGAG	2160
G Q N H P G F D Y V Q N G L Y N N S G L N G Y L P H	
ACATCATCACATCAAAACCTTGTCTTGTAGGAAATCAATCAGCTGAAAGATGACAACATCTAAAACCGGTTCTGTTTTAGGATGTT	2250
GTCAACAATCTGAAACCATCACCGGTATTTTTTGTCCAAAAATCAAACCTTTCTCTTGGCAAAGTTTGTGAAGTGTGCACAATATAA	2340
ATGTGAAATCTTCACAAGGTAAAAACAGACTCCACGAAACACTGGAGGATGAGCATACATGTAACGAACAGGATGAACAAAGTAGCAT	2430
TAATCGGAGCAGGTTTTGTTGGAAGCAGTTATGCATTTGCGTTAATTAACCAAGGGATCCGGGAATTC	2499

图 1 *B. m* α-淀粉酶基因的碱基序列及其由核苷酸序列推出的酶蛋白的一级结构

Fig. 1 It shows both nucleotide sequence of *B. m* α-amylase gene and the amino acid sequence of the enzyme deduced from the nucleotide sequence

The proenzyme is composed of 659 amino acids and the first 33 amino acids at N-terminal are signal peptide

及多T区,推测为转录终止信号。*B. m* α -淀粉酶结构基因上游区与 *B. s* α -淀粉酶上游区^[10]碱基序列排比有88%同源性。启动子位于-150—121bp的区域,并与葡萄糖阻遏有关的操纵子(*amyO*)相重叠^[12],SD序列AATAAGGAG与16S rRNA 3'端有很强的互补性。

质粒pBX101所携带的为包含 *B. m* α -淀粉酶基因的1964bp的DNA片段,它是质粒pBX92^[7]携带的野生型片段的3'端缺失535bp(1965~2499)的自发突变型。

由DNA推测的 α -淀粉酶前体蛋白由659氨基酸组成。前体 α -淀粉酶的N-端为信号肽序列,长度为33个氨基酸。成熟酶由626氨基酸组成,分子量为68.676kD。等电点pI=5.1。

2.2 *B. m* 与 *B. s* α -淀粉酶比较

B. m 和 *B. s* α -淀粉酶的氨基酸序列排比见图2。分析比较二者的氨基酸组成发现,它们之间有83.8%的氨基酸是等同的,4.6%的氨基酸虽不等同,但却是被相似氨基酸置换。如果将多肽链作分段比较,二者肽链的前3/4肽段(1—459)的同源性为90.4%,如将相似氨基酸计算在内则为94.3%。而肽链的近C-末端1/4(460—659)的同源性只有70%,包括相似氨基酸在内也只有76%。

B. m α -淀粉酶的信号肽的氨基酸组成与 *B. s* 的相比,33个氨基酸中有4个发生过突变。即 *B. s* 淀粉酶的第-31,-8,-4和-2分别由A..A..A..S突变为 *B. m* α -淀粉酶的K.S.G.N.。预期-31的A→K使 *B. m* α -淀粉酶N-端亲水性增强。而-4和-2的突变则增加了信号肽酶识别切割的准确性^[11,12]。

2.3 *B. m* α -淀粉酶的活性中心

T. Mutsunra *et al.*^[13]通过对米曲霉(*Aspergillus oryza*)产生的Taka淀粉酶的X-衍射晶体结构分析推测出其可能的淀粉酶活性中心和底物结合的四个位点的序列。Nakajima *et al.*^[4]通过比较11种不同来源的 α -淀粉酶的一级结构,确定了4个同源区,其结果亦与Mutsunra *et al.*的一致。*B. m* α -淀粉酶的和Taka α -淀粉酶的4个同源区吻合得很好。与 *B. s* α -淀粉酶相比,所不同的只是在第2个同源区中有一个相似氨基酸(Phe/Tyr)置换(资料未列出)。

3 讨 论

B. m 与 *B. s* α -淀粉酶同属糖化型 α -淀粉酶,二者的氨基酸序列同源性约90%。而和 *B. a* 和 *B. l* 的液化型 α -淀粉酶之间的同源型只有约37%(资料未列出),说明了两种类型的 α -淀粉酶基因的非同源性。

与Metz^[6]报道的 *B. megaterium* 的 α -淀粉酶相比,二者的同源性无论在DNA水平上或者蛋白质水平上都都很低(约30%)。而且所报道的酶还具有4- α -D-葡萄糖苷转移酶(EC 2.4.1.24)活性。因此,可以判断这两者所产生的淀粉酶是性质不同的两种淀粉酶。在分类上两菌株是否为同一菌种自然也在质疑之列。

B. m 与 *B. s* α -淀粉酶在氨基酸组成上,某些肽段有广泛的同源性,在第459个氨基酸之前与 *B. s* α -淀粉酶之间有90%的同源性,而其余的肽段的氨基酸序列的同源性较低。这一方面说明 *B. m* 和 *B. s* α -淀粉酶来自共同的祖先,另一方面说明在进化过程中,

```
MFKKRFKTSLPLFAGFLLLFHLLVLSG PAGANAETANKSNEVTDS SVKNG -50
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
MFAKRFKTSLPLFAGFLLLFHLLVLAGPAAASAETANKSNELTAPSIKSG -50

TILHAWNWSFNLTQNMKEIRDAGYAAIQ TSPINQVKEGNQGD KSMSNWy -100
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
TILHAWNWSFNLT KHNMKDIHDAGYTAIQ TSPINQVKEGNQGD KSMSNWy -100

WLYQPTSYQIGNRYLGT EQEFKDMCAAAEKYGVKVIVD AVVNHTTSDYGA -150
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
WLYQPTSYQIGNRYLGT EQEFKEMCAAAEEYGIKVIVDAVINHTTSDYAA -150

ISDEIKRIPN WTHGNTQIKNWSDRWDITQ NALLGLYDWNTQNT EVQAYLK -200
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
ISNEVKSIPN WTHGNTQIKNWSDRWDV TQNSLLGLYDWNTQNTQV QSYLEK -200

GFLERALNDGADGFR YDAAKHIELPDDGN YGSQF WPNITNTS AEFQYGEI -250
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
RFLDRALNDGADGFR YDAAKHIELPDDG S YGSQF WPNITNTS AEFQYGEI -250

LQDSASRD TAYANYMNV TASNYGHSIR SALKNRNL SVSNISHYASDV SAD -300
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
LQDSASRDAAYAN YMDV TASN YGHSIR SALKNRNL GVSNISHYASDV SAD -300

KLV TWVESHDTY ANDDEESTWMSDDD IRLGWAV IGSRS GSTPLFFSRPEG -350
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
KLV TWVESHDTY ANDDEESTWMSDDD IRLGWAV IASRS GSTPLFFSRPEG -350

GGNGVRFP GKSSQIGDRG S ALFKDAQAITAVNQ FHNEMAGQPEEL SNPNNGN -400
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
GGNGVRFP GKSSQIGDRG S ALFE DQAITAVNR FHNVMAGQPEEL SNPNNGN -400

QIFMNQRGS KGVVLANAGSSSVT INTSTKLPDGRYDN RAGAGSFQVANGK -450
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
QIFMNQRGS HGVVLANAGSSSVS INTATKLPDGRYDN KAGAGSFQVNDGK -450

LTGTINARSA AVLPDDIGNAPH VFLENYQTEAV HSFNDQLTVTLRANAK -500
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
LTGTINARSA VAVLPDDIAKAPH VFLENYKTGV THSFNDQLTTILRADAN -500

TTKAVYQINNG-QQTAFKGDRLTIGKEDPI GTTYNVKLTGTNGEGASRT -549
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
TTKAVYQINNGP DDRRLRMEINSQSEKEIQFGKTYT IMLKGTNSDGVTRT -550

QEYTFVKKDP SQTNIIGYQNP DHGWNV NAYIYKHDGGGAI ELTGSWP GK A -599
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
EKYSFVVKRDP ASAKTIGYQNPN HWSQV NAYIYKHDGSRVIELTGSCPGKP -600

MTKNADGIYT LTLPANADTADAKVIFNNGSAQVPGQNHPGFDYVQNGLYN -649
|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
MTKNADGIYT LTLPADTDTTNAKVIFNNGSAQVPGQNP GFDYV LNALYN -650

NSGLNGYLPH -659
|||.|||.|||.|||.
DSGLSGSLPH -660
```

图2 *B. m* α -淀粉酶的一级结构与 *B. s* α -淀粉酶的一级结构的序列排比

Fig.2 The alignment of *B. megaterium* and *B. subtilis* α -amylase protein sequences

The *B. s* protein sequence was adopted from Yang^[3]

曾因基因突变和遗传重组而导致了趋异性分化。因为 *B. m* 和 *B. s* α -淀粉酶的活性功能

区都位于酶分子的前半部分,这也暗示酶分子 C-末端部分有更多潜在的可变性。

缺失突变研究表明即使 *B. m* α -淀粉酶 C-端缺失 195 个氨基酸长的肽段(由粒 pBX952 编码),仍具 α -淀粉酶活性^[7]。有趣的是缺失编码 C-端 62 个氨基酸肽段(由粒 pBX96 编码)的 *B. m* 突变型 α -淀粉酶的酶学特性的变化。该突变型酶表现如下表型多效性效应:米氏常数(K_m 值)比野生型小 1 倍,因而活性提高 1 倍;最适反应温度由 50℃ 增加至 60℃;最适 pH 由 6.0 增高到 6.5;热稳定性也明显增高^[14,15]。由于 *B. m* α -淀粉酶的 4 个活性功能区都位于酶分子的 N-端 315 个氨基酸之前,所以 C-端的氨基酸序列的功能可能是参与酶分子活性构象的维系,从而影响酶分子活性构型及稳定性,以及酶与底物的结合及反应速率等理化特性上。

B. m α -淀粉酶编码区与 *B. s* 相比,在 N-端 3/4(1—459 个氨基酸)肽段编码区内发生过 180 次转换和 98 次颠换突变,其中 234 次为静默突变,只有 44 次为氨基酸置换突变。同源率为 90.4%。若将被相似氨基酸置换计算在内,同源率则为 94.3%。而近 C-端的 1/4 肽段(由 460—659)曾出现过 60 次置换突变,同源率只有 70%。若包括相似氨基酸置换在内为 76%。这暗示在进化过程中,淀粉酶基因的不同部位受到自然选择压力不同,酶分子的前 3/4 与后 1/4 的进化速率的明显区别是因为后 1/4 所受到的选择压力较低^[16]。

此外,在淀粉酶的前 3/4 肽段编码区,除了单一碱基被置换或颠换外,另有 8 个位点为相邻双碱基置换突变。而后 1/4 则有 15 次相邻双碱基置换,1 次相邻 3 碱基置换和 1 次 5 个相邻碱基被置换,这表明基因重组曾参与了其趋异进化的过程^[17]。

致 谢:宗建超教授曾参与指导 α -淀粉酶基因的部分 DNA 序列分析工作,特此致谢。

参 考 文 献

- [1] K. Takkinen, R. F. Pettersson, N. Kalkkinen *et al.*, *J. Biol. Chem.*, 1983, **258**:1007.
- [2] T. Yuuki, T. Nomura, H. Tazuka, *et al.* *J. Biochem.*, **98**:1147~1156, 1985.
- [3] N. Yang, A. Gulizzi *et al.* *Nucl. Acid Res.*, 1983, **11**:237.
- [4] R. Nakajima, T. Imanaka, S. Aiba. *Appl. Microb. Biotech.*, 1986, **23**:355.
- [5] M. Thomas, F. G. Priest, J. R. Stark. *J. Gen. Microb.*, 1987, **118**:167.
- [6] R. Z. Metz. *Nucl. Acids Res.*, 1988, **16**:5203.
- [7] 吕向阳, 蒋如璋, 王桂芬. 遗传学报, 1990, **18**:185.
- [8] J. H. Miller. *Experiments in Molecular Genetics.*, Published by Cold Spring Harbor Laboratory, N. Y. 1972.
- [9] T. Maniatis, *Molecular Cloning: A Laboratory Manual.*, Ed, Published by Cold Spring Harbor, N. Y. 1982.
- [10] Michael, J. Weickert, Glenn H. Chambliss. *J. Bacter.*, 1989, **171**:3656.
- [11] G. Von Heijne. *J. Mol. Biol.*, 1984, **173**:245.
- [12] 周 毅, 蒋如璋:中国科学 B 辑, 1990(9):942.
- [13] T. Matsunra, W. Kusunoki, W. Harada *et al.* *J. Biochem.*, 1984, **95**:697.
- [14] 马 明, 杨丽珠, 王诚一等. 微生物学报, 1992, **32**:400~404.
- [15] 杨丽珠, 马 明, 王诚一等. 生物化学杂志, 1993, **9**:141.
- [16] R. R. Hudson. *Proc. Natl. Acad. Sci. USA*, 1993, **90**:7425.
- [17] D. E. Dykhuizen. L. Green. *J. Bact.*, 1991, **173**:7257.

Analysis of Nucleotide Sequence of α -amylase Gene from *Bacillus megaterium*

Li Juan Jiang Ruzhang

(Department of Biochemistry and Molecular Biology,

Nankai University, Tianjin 300071)

Abstract The nucleotide sequence of α -amylase gene from *Bacillus megaterium* AS1.127 has been determined. The structural gene is composed of 1982 bp. Proamylase deduced from nucleotide sequence is composed of 659 amino acids. 33 amino acids at N-terminal was deduced as signal peptide. The mature α -amylase is composed of 626 amino acids. The molecular weight is 68.676kD. It belongs to saccharifying α -amylase with 83.8% homogeneity with *Bacillus subtilis* 168. We find that the different parts of these two amylases possess different homogeneity ratio; the first 3/4 from N-terminal end the homogeneity is 90.4%; the 1/4 from C-terminal end only 70%. Comparison of two DNA sequences suggested that both gene mutation and genetic recombination were involved in the divergent evolution process.

Key words α -amylase gene, *Bacillus megaterium*