

马铃薯 α -淀粉酶基因的克隆及生物信息学分析

王永刚, 马建忠*, 马雪青, 周贤婧, 李昆鹏, 赵虎彪
(兰州理工大学生命科学与工程学院, 甘肃 兰州 730050)

摘要: 利用 RT-PCR 方法从马铃薯(*Solanum tuberosum*)茎段总 RNA 中扩增、克隆 *amyA1* 基因(NCBI 登录号 GQ406048.1)。采用半定量 RT-PCR 方法检测 *amyA1* 基因在马铃薯茎、叶等不同组织中的表达强度, 表明在茎组织中的表达丰度略高。利用生物信息学软件分析 *amyA1* 密码子的偏好性; 同时对 AmyA1 氨基酸的理化性质、细胞内定位、保守结构及高级结构进行预测。基于 NCBI 数据库中有物种代表性的 29 种 α -淀粉酶基因序列构建了基因进化树。*amyA1* 基因全长 1224bp, 可编码一条理论分子质量为 46.40kD、407 个氨基酸残基组成的、可能为亲水性的胞外酶。与 NCBI 已登记的马铃薯 α -淀粉酶基因(登录号 M79328.1)核苷酸及氨基酸序列同源性达 98%。第 20~348 位点范围内的氨基酸残基含有与淀粉酶 13 家族及亚家族相似的催化活性域(PF00128、SM00624), 第 349~407 位点范围内的氨基酸残基含有 α -淀粉酶 C-末端 β 折叠区域(PF07821)。蛋白质结构预测表明氨基酸残基序列有维持淀粉酶活性的(β/α)桶状结构以及其他几个功能域结构。所构建的基因进化树表明, 两个马铃薯 α -淀粉酶基因与木薯、苹果的序列同源性较高, 与菜豆的次之, 与水稻、大麦、玉米等单子叶植物的序列同源性较低。

关键词: 马铃薯; α -淀粉酶基因; 克隆; 生物信息学

Cloning and Sequence Analysis of Potato α -Amylase Gene

WANG Yong-gang, MA Jian-zhong*, MA Xue-qing, ZHOU Xian-jing, LI Kun-peng, ZHAO Hu-biao
(School of Life Science and Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: A cDNA for potato α -amylase was amplified by RT-PCR and then cloned. The sequence analysis showed that the cDNA had a 1224bp-ORF, which encodes an α -amylase with 407 amino acid residues (GenBank accession number: GQ406048.1). Semi-quantitative RT-PCR assay was used to detect the expression of the α -amylase genes in potato leaves and stems, and the results showed that the expression in stems was a little stronger than that in leaves. The amino acid sequence was bioinformatically analyzed, including its codon usage bias, physical and chemical properties, subcellular localization and conserved structures. Totally 29 α -amylase genes from the same or different species were taken from the GenBank for constructing a phylogenetic tree. The bioinformatical analyses showed that the putative protein shared 98% identity with a published potato α -amylase (M79328.1) at the amino acid level. The deduced α -amylase also contained a catalytic domain (PF00128 and SM00624) between sites 20 and 348 and a C-terminal beta-sheet domain between sites 349 and 407, which were similar to their counterparts in the amylase family 13. The postulated eight-stranded alpha/beta barrel was also found in the enzyme, which is thought as an active site of α -amylase. According to the constructed phylogenetic tree, the two genes from potato presented closer homology to those from cassava and apple, followed by kidney bean, but distant homology to those from the monocotyledonous plants including barley, rice and maize.

Key words: potato; α -amylase; cloning; bioinformatic analysis

中图分类号: Q785

文献标识码: A

文章编号: 1002-6630(2010)19-0216-05

收稿日期: 2010-01-03

基金项目: 兰州理工大学博士启动基金项目; 兰州市科技局工业科技攻关及产业化项目(07-02-44)

作者简介: 王永刚(1983—), 男, 助教, 硕士, 研究方向为生化与分子生物学。E-mail: 412316788@163.com

* 通信作者: 马建忠(1963—), 男, 研究员, 博士, 研究方向植物基因工程。E-mail: majz@lut.cn

α -淀粉酶家族(α -amylase family)是最庞大的糖苷水解酶家族,约有30余个成员,包括水解酶、转移酶和异构酶。其中大部分成员又同属于糖苷水解酶家族13(glycoside hydrolase family 13, GH13)^[1]。目前,人们利用基因工程技术相继从动物、植物、真菌、细菌中克隆出了各种淀粉酶基因。按来源不同,淀粉酶被分为细菌淀粉酶、真菌淀粉酶、植物淀粉酶和动物淀粉酶等。同时,人们还利用突变等技术克隆了许多极端抗性 α -淀粉酶的基因,扩大其在工业领域中的应用^[2]。

随着石油等不可再生性能源的日益消耗以及人类发展过程中对能源的日益需求,寻找新的可再生性能源作为后石油时代的替代能源,迫在眉睫。马铃薯作为世界第三大高产经济作物,有着潜在的作为乙醇能源来源的作物优势。通过克隆马铃薯自身淀粉酶基因,构建能够利用马铃薯为原料的转淀粉酶基因工程酵母菌。利用生物发酵法,生产燃料用乙醇^[3],增加马铃薯附加值,解决现有的能源危机。本研究通过克隆马铃薯不同组织的 α -淀粉酶基因,了解其萌发生长过程中的组织分布情况,并对其序列进行生物信息学分析,以期深入了解马铃薯 α -淀粉酶的结构、功能和性质,为进一步实现其基因工程改造以及在酵母等不同遗传系统中的表达积累必要的理论知识。

1 材料与方法

1.1 材料

马铃薯组培苗(*Solanum tuberosum* cv. Shepody)本实验室培养。

1.2 菌株、质粒与试剂

E. coli DH5 α 、质粒 pET-32a(amp^r) 本实验室保存; Trizol 试剂、T₄ DNA 连接酶、pfu DNA 聚合酶、dNTP、PrimeScriptTM 反转录酶、牛小肠碱性磷酸酶(CIAP)、胶回收试剂盒(ZYMO)、焦碳酸二乙酯(DEPC)、限制性内切酶 大连宝生物工程有限公司; 其余试剂均为分析纯。

1.3 方法

1.3.1 总 RNA 的提取和 cDNA 第一链的合成

按照 Sangon 公司 Trizol 试剂操作说明书提取马铃薯叶片、茎段总 RNA,方法略有改动,在苯酚氯仿处理前先用水饱和酚等体积处理。

参照 Takara 公司反转录酶操作说明以下游引物合成 cDNA 第一链,反转录产物-20℃保存备用。

1.3.2 RT-PCR 法^[4-5]扩增 α -淀粉酶基因

根据马铃薯的 α -淀粉酶基因(GeneBank 登录号 M79328.1)序列设计扩增马铃薯夏波蒂组培苗叶片、茎秆中 α -淀粉酶基因的引物。上游引物:5'-GCGGCCGC TGTATAATCTCAACTACAATCCATT-3'; 下游引物:

5'-GCGGCCGCTTTAAGTAACGGAATAGCTGTA AGA-3'; 上下游引物都设计了 Not I 酶切位点。以逆转录产物为模板,PCR 扩增 α -淀粉酶基因。

扩增条件:扩增条件为 95℃ 1min; 95℃ 45s, 65℃ 45s, 72℃ 1min, 30 个循环; 72℃ 10min。PCR 产物经过 PCR 产物纯化试剂盒纯化后,参照 T4 Polynucleotide Kinase 说明书,将 PCR 产物磷酸化处理,保存备用。

1.3.3 载体 pET-amy A 的构建

参照碱裂解法提取 pET-32a 质粒,利用 EcoR V 酶切、牛小肠碱性磷酸酶去磷酸化处理,凝胶回收,保存备用。

将 T4 Polynucleotide Kinase 处理的 PCR 产物克隆到平末端化的 pET-32a 质粒,构建重组质粒 pET-amy A,转化至 *E. coli* DH5 α ,通过氨苄抗性筛选阳性克隆。

1.3.4 α -淀粉酶基因测序和生物信息学分析

阳性克隆经酶切鉴定正确,送往上海生工测序。DNA 测序序列用 CLC Protein Workbench 5.0 软件分析拼接,得到正确 α -淀粉酶基因序列和氨基酸序列。提交到 GenBank 上,用 BLAST 软件与 NCBI 收录的基因序列和蛋白序列进行比较。同时对测序结果进行核酸、蛋白质性质及结构的分析、预测。

1.3.4.1 核苷酸性质分析

核苷酸的一级结构取决于4种碱基的引用频率以及碱基间的排列顺序,DNA 分子中不同排列顺序的 DNA 区段构成特定的功能单位,决定氨基酸的种类、排列及蛋白质高级结构的功能。运用软件 CHIPS 对 α -淀粉酶基因进行有效密码子数统计,再运用 Codon W 和 CUSP 计算有效密码子数(effective number of codons, ENc)、CDS 区的 GC 含量、密码子中第3位碱基的 GC 含量(GC_{3s})和密码子的使用频率。最后将 α -淀粉酶基因的密码子偏好性与酵母基因组的密码子偏好性“/1000”使用频率比值进行比较分析^[6],预测其酵母中的表达情况。

1.3.4.2 蛋白质性质分析^[7]

蛋白质理化性质分析: ProtParam tool(<http://www.expasy.org/tools/protparam.html>)分析序列的分子质量、理论等电点、氨基酸组成等理化性质;疏水性预测和分析:瑞士生物信息学研究所的 ExPASy 服务器上的 ProtScale 程序(<http://www.expasy.ch/tools/protscale.html>)对氨基酸序列作疏水性分析;二级结构预测分析:利用 3D-PSSM 分析蛋白质序列的二级结构(<http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html>)。保守结构域的预测和分析:通过 <http://www.ncbi.nlm.nih.gov/Structure/CDD> 保守结构域预测程序,预测相关的 α -淀粉酶基因序列的保守结构域。蛋白序列跨膜区分析:丹麦科技大学(DTU)的 CBS 服务器上的 TMHMM Server v.2.0

程序进行蛋白序列跨膜区分析(<http://www.cbs.dtu.dk/services/TMHMM/>)。蛋白质在细胞内的定位分析: 采用 PSORT 软件进行分析(<http://www.psort.org>)。蛋白质的立体结构预测和观看采用 SISS-MODEL(<http://swissmodel.expasy.org/SISS-MODEL.html>)。

1.3.4.3 基于 α -淀粉酶基因的物种进化分析

采用近邻结合法(neighbor-joining)将 NCBI 中不同物种的 29 种 α -淀粉酶序列数据集合在一起, 利用 CLC Protein Workbench 5.0 软件多重序列比对并绘制基因进化树, 进一步了解功能结构域的进化^[8]。

2 结果与分析

2.1 总 RNA 的提取及 RT-PCR 扩增目的基因

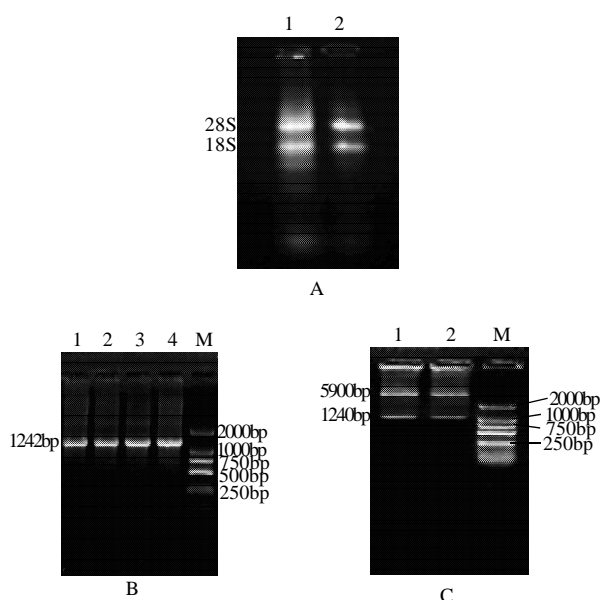


图1 马铃薯不同组织 α -淀粉酶基因的克隆及阳性质粒的酶切鉴定
Fig.1 Cloning of α -amylase genes from different tissues of potato and restriction enzyme digestion for identification of positive plasmid

按照 Trizol 试剂提取 RNA 操作说明书提取总 RNA, 用 Cary 50 紫外分光光度计分别检测来源叶片和茎段的 RNA 溶液在 280nm 和 260nm 波长处的吸光度。 A_{260nm}/A_{280nm} (用 R 表示) 分别为 1.92、1.97。 R 值在 1.82~2.20 之间, RNA 质量合乎要求。同时用琼脂糖凝胶电泳检测 RNA 的完整性。28S 和 18S 条带明亮、清晰, 且 28S 的亮度约为 18S 条带的两倍, RNA 的质量很好(图 1A)。总 RNA 经反转录合成 cDNA 第一链, 以此为模板, PCR 扩增

α -淀粉酶基因。PCR 产物经 1.0% 琼脂糖凝胶电泳检测。源于叶片和茎段的 cDNA 都能扩增出一条约 1242bp 的目的条带, 与预期大小一致, *amyA1* 基因的实际全长为 1224bp(图 1B)。

2.2 α -淀粉酶基因的组织表达分析

马铃薯块茎在萌发期间, 块茎中 α -淀粉酶的活力与其生长有密切的关系。它能促进块茎萌发和早期幼苗生长, 同时在生长过程中其他组织和细胞内都有 α -淀粉酶基因的表达。如图 1B 所示, 利用半定量 RT-PCR 分析表明, 在叶片及茎段中都有 α -淀粉酶基因的表达, 在茎段组织中表达丰度略高。通过进一步克隆测序证明马铃薯在生长过程中在不同组织中都有淀粉酶的表达。

2.3 重组质粒的鉴定

通过筛选获得阳性克隆, 摇菌并且提取质粒, 获得重组质粒 pET-amyA, 利用 PCR 引物设计中引入的 *Not* I 切点检测重组质粒的正确性, 如图 1C 所示, 重组质粒经 *Not* I, 产生 1230bp 目的条带, 根据大小判断与预期的一致, 送交上海生工测序。

2.4 α -淀粉酶基因的生物信息学分析

2.4.1 核苷酸序列分析

从测序结果可知, 获得的 α -淀粉酶基因全长 1224bp, 编码 407 个氨基酸, 经过分析整合比对, 命名为 amyA1。提交到 NCBI, 得到登录号 GQ406048.1。利用 CLC Protein Workbench 5.0 软件分析, 该序列含有 1 个开放阅读框架(open reading frame, ORF)。NCBI/Blast 的核苷酸序列同源性分析表明, 此序列与 GenBank 中已登记的 M79328.1 源于马铃薯的 α -淀粉酶基因的核苷酸序列同源性高达 98%, 氨基酸序列的同源性高达 98%, 与 AK323104.1 番茄 α -淀粉酶基因的核苷酸序列同源性高达 97%。这可能是由于种属及品种间差异所引起的, 对该基因序列多样性的研究有一定意义^[9]。

2.4.2 密码子分析

根据所得到的 α -淀粉酶基因序列, 利用 CHIPS 和 Codon W 软件对 *amyA1* 基因统计得到的有效密码子, 显示除了编码 Pro 的 CCC 及编码 Arg(R) 的 CGA 没有出现外, 其他氨基酸的各种密码子在该基因序列中都有出现, ENc 值为 53.57, 与 ENc 值为 60 接近, 表明马铃薯在合成 α -淀粉酶的过程中, 密码子的兼并性及偏好性并不明显。此外, 整个序列中碱基对 GC 含量为 44.1%, GC_{3s} 含量为 39.95%, 说明该基因中的 G 和 C 碱基绝大部分分布于密码子的第 3 位。此外通过比较马铃薯 α -淀粉酶基因序列中有效密码子及酵母偏好性密码子使用频率的比值, 可以发现, 马铃薯和酵母的偏好性有较小差异。马铃薯 α -淀粉酶基因序列和酵母的密码子“/1000”比

值中有 17 种密码子的值大于等于 2.0 或小于等于 0.5。由此可以初步预测该基因在酵母中能够顺利表达。同时为该基因表达选择合适的表达系统,改造密码子提高基因的表达式提供了依据。

2.4.3 蛋白质序列分析

2.4.3.1 理化性质的预测

amyA1 氨基酸 407, 相对分子质量 46395.1, 分子式为 $C_{2087}H_{3141}N_{577}O_{606}S_{12}$ 。理论等电点为 6.95。其中负电荷氨基酸残基(Asp+Glu)46 个, 占总氨基酸含量的 11.3%, 正电荷氨基酸残基(Arg+Lys)45 个, 占总氨基酸含量的 11.06%。不稳定参数为 45.7, 属于不稳定蛋白。半寿期 30h, Kyte&Doolittle 算法预测蛋白质的疏水性, 其疏水性平均值为 -0.56, 脂溶指数为 70.71, 预测该蛋白为疏水性脂溶蛋白。用 Prot Scale 预测 α -淀粉酶序列的疏水性/亲水性的结果表明, 多肽链第 230 位的丝氨酸(Ser)具有最低的分值 -3.144, 第 102 位的丙氨酸(Ala)具有最高的分值 1.656。参考 ProtScale 软件, 依据氨基酸分值越低亲水性越强和分值越高疏水性越强的规律, 可以看出, 在第 230 位的 Ser 亲水性最强, 第 102 位的 Ala 疏水性最强。就整体来看, 亲水性氨基酸均匀分布在整条肽链中, 且多于疏水性氨基酸。因此, 整个多肽链表现为亲水性, 可推测该蛋白质为亲水性蛋白。同时利用 TMHMM 2.0 进行蛋白质序列的跨膜区分析表明, 该蛋白由里到外在 50~70bp 和 309~327bp 处存在两个跨膜螺旋, 中心位置分别为 59bp Ala 和 318bp Thr。经过 PSORT 软件对马铃薯 α -淀粉酶基因蛋白亚细胞定位的预测, 表明该基因蛋白在细胞外的可能性为 9.97。与 P23314 基因蛋白性质匹配, 属于胞外蛋白酶类。

2.4.3.2 功能结构域的分析^[10]

利用 3D-PSSM 在线分析得出, 组成马铃薯 α -淀粉酶基因的蛋白质序列中, 存在 14 个 α -螺旋, 11 个 β -折叠, 其他主要由无规则卷曲构成, 预测表明该蛋白的二级结构主要为 α - β 型。此外通过 Interpro Scan 对该序列进行保守序列的预测以及功能位点的注释, 发现马铃薯 α -淀粉酶在 20~348 位点范围内氨基酸与 IPR006047、IPR006589 存在相似的结构域(F00128、SM00624), 属于淀粉酶 13 家族和亚家族催化活性区域; 在 19~379 位点内氨基酸与 SSF51445 相似存在转糖苷酶活性。在 349~407 位点内氨基酸与 IPR12850 存在相似 domain(PF07821), 属于 α -淀粉酶 C-末端 β 折叠区域; 同时得出位于 68、103、108、187、212、294 位点处氨基酸有活性功能, 与钙、锌离子的结合有关。

2.4.3.3 高级结构的预测

根据 Swiss-Model 的预测结果显示, 该序列空间结

构的预测基于 PDB: 1rp8 ChainA 为模板, 序列一致性为 44.853%。空间结构见图 2, 可以明显看出该结构在活性催化区域具有典型的 $(\beta/\alpha)_8$ 桶状结构, 结合二级结构的预测表明 C-末端具有典型的 β 折叠结构, 与已知的淀粉酶的空间结构大致符合。大部分淀粉酶有 8 个 $(\beta/\alpha)_8$ 桶状结构的催化中心, 还包括 domain B、C、D。其中 domain B 具有 3 个 β 折叠和 3 个 α 螺旋, 长度和结构随来源的不同而变化; domain C 区是催化区域后面的区域, 主要由 β 折叠组成, 该区被认为是起保护酶催化活性中心疏水氨基酸稳定性的作用。



图 2 Swiss-Model 预测的 AmyA1 蛋白空间结构

Fig.2 Predicted AmyA1 protein spatial structure using Swiss-Model

2.5 基于 α -淀粉酶基因的基因树分析

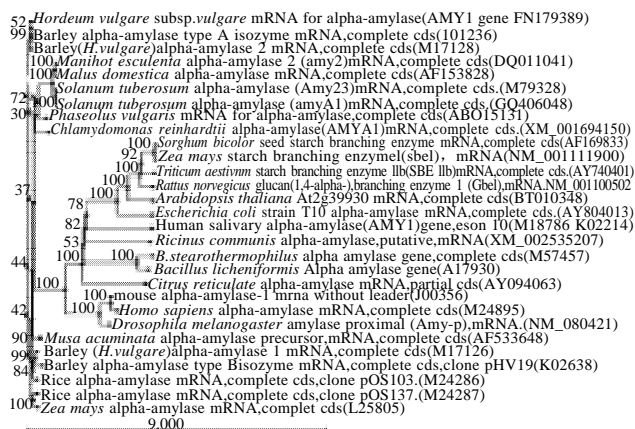


图 3 基于 α -淀粉酶基因的不同物种间的基因进化树

Fig.3 Phylogenetic tree of different species based on α -amylase gene

如图 3 所示, 已克隆的两个马铃薯 α -淀粉酶基因与木薯、苹果的序列同源性较高, 与菜豆的次之, 与水稻、大麦、玉米等单子叶植物的序列同源性较低。马铃薯与木薯均发源于南美洲, 其与苹果较高的同源性则难以理解。微生物、动物与植物来源的 α -淀粉酶的氨基酸残基序列之间的同源性不超过 10%。尽管整体的序列同源性很低, 但 α -淀粉酶共有的 A、B、C、D

这4个区域仍然高度保守,推测这些高度保守的区域可能组成了酶结合/催化底物的活性中心。

3 讨论

马铃薯的主要储藏物质为淀粉,淀粉的降解是其块茎萌发进而生长成植株的关键步骤,因而淀粉酶的表达与调控在马铃薯块茎萌发过程中起重要的作用。考虑到酶的专一性特性,马铃薯自身淀粉酶对马铃薯淀粉的水解效率较高、利用率高等特性,本实验主要克隆了马铃薯自身 α -淀粉酶,通过生物信息学手段,分析了 α -淀粉酶核苷酸序列、氨基酸序列的理化特性、序列结构等,与酵母偏好密码子进行对比分析,预测了在酵母中的表达情况。分析表明:马铃薯 α -淀粉酶系糖苷水解酶家族13,有完整的功能活性域^[11],同时在活性部位拥有淀粉酶的共同催化氨基酸Asp206、Glu230、Asp297^[12],为 α -淀粉酶基因在工程菌株中的转化表达提供参考。譬如在酵母表达系统中利用马铃薯生淀粉,表达淀粉酶,生产燃料乙醇。同时研究该酶的相关动力学参数,进行酶的活性、热稳定性及耐酸碱性的测定,根据已有的生物信息学分析结果,对其进行改造^[13-15],一方面提高其对马铃薯生淀粉的利用率,另一方面增加其稳定抗性等,提升其工业中的应用。

参考文献:

- [1] 王阳,李樊,张应玖. α -淀粉酶家族的新成员:多功能淀粉酶[J]. 中国生物化学与分子生物学报, 2009, 25(2): 104-109.
- [2] FITTER J. Structural and dynamical features contributing to thermostability in α -amylases[J]. Cell Mol Life Sci, 2005, 62(17): 1925-1937.
- [3] GHANG D M, YU L, LIM M H, et al. Efficient one-step starch utilization by industrial strains of *Saccharomyces cerevisiae* expressing the glucoamylase and α -amylase genes from *Debaryomyces occidentalis* [J]. Biotechnol Lett, 2007, 29(8): 1203-1208.
- [4] 迪芬巴赫 C W, 德威克斯勒 G S. PCR 实验技术指南[M]. 黄培堂, 俞炜源, 陈添弥, 等, 译. 北京: 科学出版社, 2000: 380-393.
- [5] 萨姆布鲁克 J, 拉塞尔 D W. 分子克隆实验指南[M]. 3 版. 黄培堂, 王恒禄, 周晓巍, 等, 译. 北京: 科学出版社, 2006: 636.
- [6] 赵耀, 刘汉梅, 顾勇, 等. 玉米 waxy 基因密码子偏好性分析[J]. 玉米科学, 2008, 16(2): 16-21.
- [7] TAO Fang, ZHU Suwen, FAN Jun, et al. Cloning and sequence analysis of maize *FAD2* gene[J]. Journal of Plant Physiology and Molecular Biology, 2006, 32(6): 649-656.
- [8] JANECEK S, SVENSSON B, HENRISSAT B. Domain evolution in the α -amylase family[J]. J Mol Evol, 1997, 45(3): 322-331.
- [9] PARK K H, KIM T J, CHEONG T K, et al. Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the α -amylase family[J]. Biochim Biophys Acta, 2000, 1478(2): 165-185.
- [10] MacGREGOR E A, JANECEK S, SVENSSON B. Relationship of sequence and structure to specificity in the α -amylase family of enzymes [J]. Biochim Biophys Acta, 2001, 1546(1): 1-20.
- [11] 缪可嘉, 王正祥. 淀粉酶系中淀粉结合域的结构和功能的研究进展[J]. 食品工业科技, 2007, 28(3): 242-245.
- [12] 孙连海. α -淀粉酶结构与功能关系的研究[D]. 河南: 河南农业大学, 2006.
- [13] JUGE N, N ØHR J, LE GAL-COËFFET M F, et al. The activity of barley α -amylase on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus niger* glucoamylase[J]. Biochim Biophys Acta, 2006, 1764(2): 275-284.
- [14] DAMIÁN-ALMAZO J Y, MORENO A, LÓPEZ-MUNGUÍA A, et al. Enhancement of the alcoholic activity of α -amylase AmyA from *Thermotoga maritima* MSB8 (DSM 3109) by site-directed mutagenesis [J]. Appl Environ Microbiol, 2008, 74(16): 5168-5177.
- [15] SIDDIQUI K S, FELLER G, D' AMICO S, et al. The active site is the least stable structure in the unfolding pathway of a multidomain cold-adapted α -amylase[J]. J Bacteriol, 2005, 187(17): 6197-6205.