

## 技术与方法

## 一种有效提高纳豆激酶表达量的方法

潘秋辉<sup>1</sup> 谢秋玲<sup>2\*</sup> 洪岸<sup>2</sup> 蒲含林<sup>2</sup> 刘翠兰<sup>2</sup> 廖美德<sup>2</sup>(成都体育学院运动医学系 成都 610041)<sup>1</sup> (广州暨南大学生物工程研究所 广州 510632)<sup>2</sup>

**摘要:** 为提高纳豆激酶在大肠杆菌中的表达量,在不改变蛋白质序列的前提下,对纳豆激酶氨基端编码前20个氨基酸的序列进行改造,以降低此区域的GC含量及稀有密码子数量,其中有3个位点直接采用T,另有4个位点采取不确定突变的方式,产生16种组合,由此而得16条序列,利用生物信息软件DNASIS v2.5对这些序列的二级结构进行模拟,取得自由能绝对值最低的6条序列设计引物,克隆至表达载体pET-3c,导入大肠杆菌BL21(DE3),IPTG诱导表达,其中有3株菌形成明显高效表达。

**关键词:** 纳豆激酶, 翻译起始区, 密码子偏好性

**中图分类号:** Q784 **文献标识码:** A **文章编号:** 0253-2654 (2002) 06-0059-05

## CLONING AND HIGH EXPRESSION OF NATTOKINASE WITH AN EFFECTIVE STRATEGY

PAN Qiu-Hui<sup>1</sup> XIE Qiu-Lin<sup>2</sup> HONG An<sup>2</sup> PU Han-Lin<sup>2</sup> LIU Cui-Lan<sup>2</sup> LIAO Mei-De<sup>2</sup>(Sports Medical Department of Chengdu Physical Education Institute, Chengdu 610041)<sup>1</sup>(Bioengineering Institute of Jinan University, Guangzhou 510632)<sup>2</sup>

**Abstract:** Computer program DNASIS was used to design the site-directed mutation for optimizing the expression of nattokinase in *E. coli*. The secondary structure of the TIR (Translation Initiation Region) is a determinant factor for translation initiation rate, meanwhile, codon preference plays an important role, too. According to this two principles, 3 sites at the beginning of the open reading frame of nattokinase gene were definitely changed to T, and another 4 sites randomly changed. Then computer program DNASIS v2.5 was utilized to analyze the total 16 TIR sequences resulted from the combination of the 4 randomly mutated sites. Six sequences with the top level of free formation energy ( $\Delta G^0$ ) were chosen for subsequent cloning. By PCR using synthetic primers containing the 7 changed sites described above, 6 nattokinase gene were amplified and cloned to pET-3c respectively. BL21 (DE3) cells were transduced and induced to express recombinant nattokinase. Three high-expressional clones were obtained, indicating that computer program-aided design for optimizing expression of foreign genes in *E. coli* is useful.

**Key words:** Nattokinase, TIR, Codon preference

纳豆是日本的传统发酵食品,它由枯草杆菌发酵大豆而成,纳豆同时也作为一种传统药品用于心脑血管疾病的防治。纳豆激酶(Nattokinase, NK)是从日本传统食品纳豆中提取出来的具有溶血栓作用的一种酶,因其与链激酶、尿激酶等溶血栓药物具有同样的功能,故最初被命名为纳豆激酶。NK具有安全性好、易被人体吸收、作用直接迅速、持续时间长、可由细菌发酵生产因而造价低廉等优点,是一种新型溶血栓药物。但直接由枯草杆菌发酵产量较低,我们曾尝试用基因工程方法直接在大肠杆菌 BL21 (DE3) 中发酵纳豆激酶,未见明显的表达。

\* 联系人 Tel: 020-85220221, Fax: 020-85226616, E-mail: ojds@jun.edu.cn

收稿日期: 2001-11-19, 修回日期: 2002-02-18

翻译起始区是指基因编码框起始密码子上下游一段对基因表达具有重要影响的区域,此区域的二级结构会直接影响基因的表达。因此 TIR 区域的 GC 含量过高会造成二级结构过于稳定,使得转录单位无法顺利通过,影响基因表达<sup>[1]</sup>。此外,大肠杆菌对编码基因的密码子存在一定的偏好性,使用其偏好的密码子有利于基因的表达。综上所述,我们对纳豆激酶基因编码区 5' 端的部分序列进行了改造,但前提是改变蛋白质的氨基酸序列。

## 1 材料与方法

### 1.1 材料

**1.1.1 菌株与质粒:**质粒 pET-3c 与大肠杆菌 BL21 (DE3) 购自 Novagen; 枯草杆菌 (*Bacillus subtilis*) 由华南理工大学食品与生物工程学院生物工程系郭勇教授惠赠, 大肠杆菌 DH5 $\alpha$  由本室保存。

**1.1.2 工具酶与试剂:**限制性内切酶、T4 连接酶、Taq 酶均购自 NEB; 蛋白质分子量标准购自 Pharmacia; 质粒纯化、PCR 纯化、切胶纯化试剂盒均购自 QIAGEN。

### 1.2 方法

**1.2.1 枯草杆菌基因组 DNA 的提取:**按照文献 [2] 的方法抽提。

**1.2.2 纳豆激酶基因的扩增与克隆:**根据 genebank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) 提供的序列并参照文献 [3] 设计引物扩增纳豆激酶的编码序列。上游引物: 5' > ATCT CATATG GCG CAA TCT GTT CCT TAT GGC < 3', 划线部分为 Nde I 酶切位点; 下游引物: 5' > ATCG GGATCC TTA TTA TTG TGC AGC TGC TTG < 3', 其中划线部分为 BamH I 位点。以枯草杆菌基因组 DNA 为模板扩增, 条件: 94℃ 变性 4 min, 94℃ 30 s, 60℃ 30 s, 72℃ 30 s, 33 个循环, 72℃ 延伸 10 min, 切胶回收 PCR 产物, Nde I 与 BamH I 分别双酶切 PCR 产物与 pET-3c, 并按等摩尔比连接, 16℃ 30 min, 转化感受态细胞 DH5 $\alpha$ , 酶切鉴定。

**1.2.3 编码序列突变引物的设计:**我们选取纳豆激酶 5' 端编码前 20 个氨基酸 (除去起始密码子) 的区域进行改造, 密码子第 3 位摇摆碱基尽量选用 AT 以降低 GC 含量, 此外尽量选用大肠杆菌的偏好密码子 (根据网站 [www.kazusa.or.jp/codon](http://www.kazusa.or.jp/codon) 的统计), 这样一共选取了 7 个位点, 其中 3 个直接使用 T, 但另有 4 个位点选用 A 或 T 后增加了密码子的稀有性, 如编码丙氨酸的密码子有 4 种, 它们在所统计的大肠杆菌基因编码框中的使用频率分别为 16.4% (GCU), 20.6% (GCA), 24.9% (GCC), 32.0% (GCG), 显然, 若将原序列中的 GCG 改成 GCA, 其偏好性必然有所下降, 我们将此位点称为不确定突变位点, 并在此位点同时引入两种核苷酸, 一共有 4 个不定位点, 就得到 16 种不同的序列组合。并据此设计上游引物如下: 5' > ATCG CATATG GC (AG)<sup>1</sup> CA (AG)<sup>2</sup> TCT GTT CCT TAT GG (C→T) ATT TCT CAA ATT AAA GCG CCG GC (AG)<sup>3</sup> CTT CA (G→T) TCT CA (AG)<sup>4</sup> GG (C→T) TAC ACA GGC TCT AAC GTA AAA < 3'。其中, 划线部分为 Nde I 酶切位点, 上标 1、2、3、4 处分别同时引入两种突变, 一共得到 16 条引物。选取 NK 基因起始密码子上下游共 90 个碱基作为 TIR 的范围, 其中上游 30 个碱基为载体 pET-3c 上的固有序列, 下游 60 个碱基为改造后的基因编码序列, 利用二级结构预测软件 DNASIS 对 TIR 进行 RNA 二级结构的模拟, 取前 6 个结构最不稳定 (自由能绝对值最小) 的引物, 进行表达。共同的下游引物: 5' > ATCG GGATCC TTA TTA TTG TGC AGC TGC TTG < 3', 其中划线部分为 BamH I 位点。

**1.2.4 纳豆激酶基因的突变和克隆：**以 NK 基因的 PCR 产物（1.2.2 的产物）为模板，分别利用 6 条突变引物配合共同的下游引物进行扩增，条件同 1.2.3，*Nde* I 与 *Bam*H I 分别双酶切 PCR 产物与 pET-3c，并按等摩尔比连接，16℃ 30 min，转化感受态细胞 DH5 $\alpha$ ，酶切鉴定后，将重组子转化表达菌株 BL21（DE3）。

**1.2.5 重组纳豆激酶的表达与鉴定：**挑取重组子培养过夜，次日按 1% 转接至 20mL LB 培养基中，培养至  $OD_{600}$  = 0.2，加 IPTG 至终浓度 1mmol/L，继续培养 4 h，离心收集菌体，超声破碎，取上清，SDS-PAGE 分析鉴定表达产物。

**1.2.6 活性鉴定：**采用纤维蛋白平板法测定纳豆激酶活性。将纤维蛋白原，凝血酶和琼脂糖制成琼脂糖纤维蛋白平板，各取 10 $\mu$ L 裂解液加到平板上，37℃ 温箱中放置 18h，根据溶圈直径，估算待检品的活性。

2 结果

2.1 TIR 分析结果

我们利用软件 DNASIS 对 4 个不确定突变位点上的 16 种不同组合所形成的序列以及原始的序列进行二级结构的预测，并按照自由能大小排序，结果如表 1 所示（原始序列的  $\Delta G^0$  为 -18.3kcal/mol）。所有序列的 5' 端起始密码子以前都是 pET-3c 载体上的一段共

表 1 NK 基因 5'端 TIR 4 个位点不同突变组合所形成的 16 种序列的  $\Delta G^0$  (kcal/mol)

次序	位点	$\Delta G^0$	次序	位点	$\Delta G^0$	次序	位点	$\Delta G^0$
No. 1	A A A A	-18.0	No. 7	G A G A	-18.3	No. 13	A A G A	-22.0
No. 2	A A A G	-18.0	No. 8	G G A A	-18.3	No. 14	A G G A	-22.0
No. 3	A G A A	-18.0	No. 9	G G A G	-18.3	No. 15	A G G G	-22.0
No. 4	A G A G	-18.0	No. 10	G G A A	-18.3	No. 16	G A G G	-22.3
No. 5	G A A G	-18.0	No. 11	A A G G	-21.6			
No. 6	G A A A	-18.3	No. 12	G G G G	-21.9			

同序列：ATAA TTT TGT TTA ACT TTA AGA AGG AGA TAT ACA T CAT。我们取前 5 种自由能绝对值小于原始基因的序列以及一条等于原始基因的序列设计引物并分别克隆入大肠杆菌 BL21（DE3）进行表达（表 1）。

2.2 目的基因的扩增与克隆

以枯草杆菌基因组为模板，PCR 扩增后，1.0% 的琼脂糖凝胶电泳，约 820bp 处可见一明显条带，与预计产物大小相吻合；克隆入 pET-3c 载体后，酶切鉴定，插入片段大小正确（图 1）。

2.3 NK 基因的突变与克隆

以 NK 的 PCR 产物为模板，No.1 至 No.6 的突变序列为引物行 PCR 扩增，装入 pET-3c 载体中，酶切产物约 820bp（图 2）。

2.4 重组子在大肠杆菌中的表达

经 IPTG 诱导表达 4 h 后，SDS-PAGE 电泳，发现 No.3、No.5、No.6 在 27kD 处均有一明显条带，与纳豆

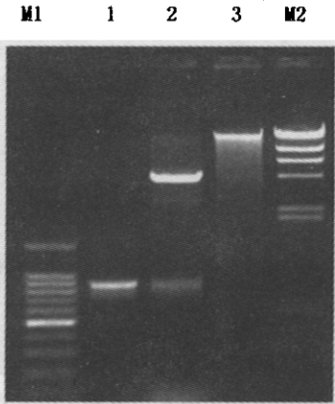


图 1 NK 基因的扩增及克隆  
M1 100bp ladder (依次为 1.5, 1.0, 0.9, 0.8, 0.7, 0.6, 0.5kb 等), 1 NK 基因 PCR 产物, 2 pET-3c-NK/*Bam*H I + *Nde* I, 3 枯草杆菌基因组 DNA, M2  $\lambda$ DNA/*Hind* III Marker (23, 9.4, 6.5, 4.3, 2.3, 2.0kb 等)

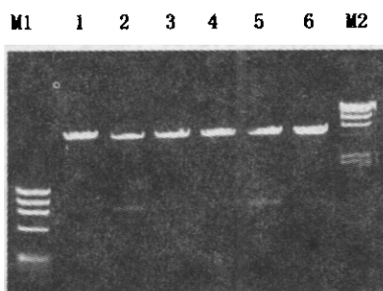


图2 NK基因的突变与克隆

M1  $\Phi$ x174/Hae III marker (1.35, 1.08, 0.82, 0.6kb 等), 1-6 No.1 至 No.6 重组质粒/*Bam*H I + *Nde* I, M2  $\lambda$ DNA/*Hind* III Marker (23, 9.4, 6.5, 4.3, 2.3, 2.0kb)

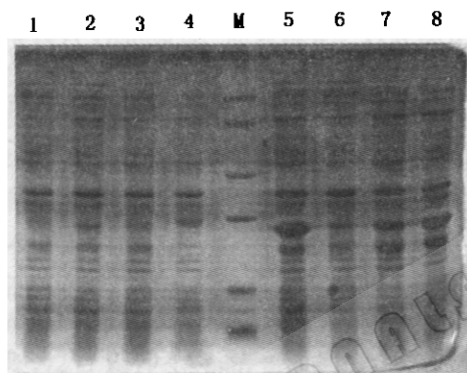


图3 表达产物 SDS-PAGE 电泳结果

1 重组子/未诱导, 2 包含原始序列的重组子/诱导, 3-8.No.1-No.6 重组子/诱导

M 蛋白质标准 (97, 66, 45, 30, 20, 14.4kD)

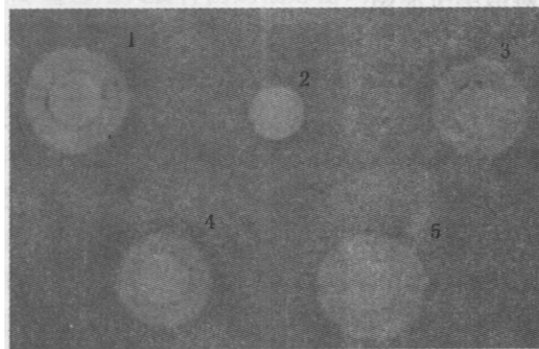


图4 重组子的活性检测

1 No.3 纤溶圈, 2 阴性对照, 3 No.5 纤溶圈, 4 No.6 纤溶圈, 5 阳性对照

激酶分子量大小吻合。薄层扫描结果表明, 表达量分别占菌体蛋白的 22%, 9% 和 18%, 其余重组子均有少量表达 (图 3)。

## 2.5 活性测定

选用转导 pET-3c 空载体的 BL21 (DE3) 裂解液为阴性对照, 未见明显纤溶圈, 枯草杆菌裂解液为阳性对照, 纤溶圈较明显, No.3、No. 5、No.6 产物均有纤溶活性 (图 4)。

## 2.6 序列分析

3 个表达克隆上游由 PCR 引入的突变序列 (前 20 个氨基酸的编码序列) 如下, 划线部分为突变碱基, NK 基因的其余部分可在 GeneBank 中获得:

No.3: ATG GCA CAG TCT GTT CCT TAT GGT ATT TCT CAA  
ATT AAA GCG CCG GCA CTT CAT TCT CAA GGT TAC  
ACA GGC TCT AAC GTA AAA;

No.5: ATG GCG CAA TCT GTT CCT TAT GGT ATT TCT CAA  
ATT AAA GCG CCG GCA CTT CAT TCT CAG GGT TAC  
ACA GGC TCT AAC GTA AAA;

No.6: ATG GCG CAA TCT GTT CCT TAT GGT ATT TCT CAA  
ATT AAA GCG CCG GCA CTT CAT TCT CAA GGT TAC  
ACA GGC TCT AAC GTA AAA.

## 3 讨论

成熟的纳豆激酶有 275 个氨基酸, 它是由前肽原经过加工而成, 前肽原则包括 29 个氨基酸的信号肽和 77 个氨基酸的多肽再加 275 个氨基酸的成熟肽。国内刘北域等人<sup>[3]</sup>报道在大肠杆菌中成功表达了纳豆激酶原, 张淑梅等人<sup>[4]</sup>则表达了纳豆激酶的融合蛋白, 直接表达 NK 的 275 氨基酸的成熟肽尚无报道。我们采用了高度商品化的大肠杆菌表达系统 BL21 (DE3) -pET-3c 直接表达了纳豆激酶的成熟形式。但起初, 没有改变纳豆激酶的编码基因, 未见明显产物, 而同时在 DH5 $\alpha$ -PBV220 温度诱导表达系统中则得到高表达 (另文发表), 由此认为原始的纳豆激酶并不适合在 BL21 (DE3) -

pET-3c 系统中表达。

一般认为外源基因在大肠杆菌中的表达很大程度上决定于 TIR 区域, 这是由于此

区域有一些重要的顺式结构, 如: SD 序列、起始密码子、以及 -3 位碱基<sup>[5]</sup>。甚至有人认为 TIR 区的自由能绝对值每增加 1.4kcal/mol, 基因的翻译起始率即会下降 10%<sup>[6]</sup>。但是 TIR 的范围尚无明确定义, Ganoza 认为应包括起始密码子上下游 70 个核苷酸的范围<sup>[7]</sup>, 考虑到本研究中起始密码子上游的序列为载体序列, 已经过优化, 我们将重点放在 ATG 下游的序列, 并将调整范围延长到 60 (即 20 个氨基酸) 个核苷酸。同时我们也考虑到密码子的偏好性, 大肠杆菌稀有密码子的存在会造成很多严重后果: mRNA 的半衰期降低, 翻译提前终止, 和移框突变等<sup>[8]</sup>。综上所述, 我们在降低 GC 含量的同时, 尽量采用大肠杆菌偏好的密码子。

本研究中菌株 No.6 的自由能与原始序列相同且较前 5 种序列为低, 但表达量在所有序列中位居第二, 与计算机的预测结果并不完全一致, 可能是由于 TIR 内部的碱基有可能与外部的碱基发生配对, 形成更大范围的二级结构, 超出了计算机的分析范围, 因此人为地适当延长 TIR 范围可以增加预测的可靠性。

## 参 考 文 献

- [1] Rex G, Surin B, Besse G, *et al.* J Biol Chem, 1994, **269** (27): 8118~8127.
- [2] Fujita M S, Nomura K C, Hong K G, *et al.* Biochem Biophys Res Commun, 1993, **197** (3): 1340~1347.
- [3] 刘北城, 官孝群, 宋后燕. 上海医科大学学报, 1999, **26** (6): 401~404.
- [4] 张淑梅, 张云湖, 赵晓祥, 等. 中国生物化学与分子生物学报, 1999, **15** (6): 912~915.
- [5] Barrick D, Villanueva K, Childs J, *et al.* Nucleic Acids Research, 1994, **22** (7): 1287~1295.
- [6] de Smit M H, van Duin J. Proc Natl Acad Sci, 1990, **87** (19): 7668~7672.
- [7] Ganoza M C, Kofoed E C, Marliere P, *et al.* Nucleic Acids Res, 1987, **15** (1): 345~360.
- [8] Andersson S G, Kurland C G. Microbiol Rev, 1990, **54** (2): 198~210.