

# 化学与酶促相结合合成人生长激素基因

赵向阳, 高朝辉, 金璐, 王曦, 孟庆繁, 滕利荣  
(吉林大学生命科学学院, 长春 130012)

**摘要** 根据人生长激素(hGH)的氨基酸序列和大肠杆菌密码子的偏好性, 全面优化 hGH 的密码子, 添加表达调控元件后合成序列的全长为 1040 bp, 采用化学方法拆分成 30 条单链寡核苷酸. 采用改进的两步法拼接成全基因, 得到 2 个全序列正确的基因. DA-PCR 和 OE-PCR 拼接产物经 T7 核酸内切酶 I 处理, 合成基因中的碱基错误率降低了 93.93%.

**关键词** 人生长激素; 基因合成; 两步法

**中图分类号** Q516; O629.8

**文献标识码** A

**文章编号** 0251-0790(2008)10-1995-04

人生长激素(Human Growth Hormone, hGH)是由脑垂体前叶嗜酸性细胞分泌的 191 个氨基酸组成的非糖基化的蛋白质类激素, 分子量约为 22000<sup>[1]</sup>, 可以促进人体生长, 影响蛋白质、糖和脂类的代谢<sup>[2]</sup>. 可用于治疗儿童期的生长激素缺乏症、Turner 综合症、Prader-Willi 综合征、慢性肾衰竭等多种病症<sup>[3]</sup>. 早期人类生长激素只能由尸体的脑垂体提取<sup>[4]</sup>, 现已被基因工程重组产品完全替代<sup>[5]</sup>, 编码基因采用 mRNA 逆转录得到, 密码子是人类细胞偏爱的, 由其构建的工程菌常常因为稀有密码子对应的 tRNA 数量不足而导致蛋白表达的意外终止, 表达量受到限制<sup>[6,7]</sup>.

最近, 人们尝试用酶促与化学合成相结合的方法来重新合成 hGH 全基因, 构建高效表达的工程菌. 本文对 hGH 的密码子进行了全局优化, 尝试改良合成方法<sup>[8]</sup>以降低合成基因中的碱基错误率, 合成 hGH 基因为进一步的表达奠定了基础.

## 1 实验部分

### 1.1 材料

单链寡核苷酸由上海生工合成, 经 PAGE 纯化; pfu 高保真聚合酶购自 Ferments; 限制性内切酶、T4 DNA 连接酶和 T7 核酸内切酶 I 均购自 NEB 公司.

### 1.2 hGH 基因的合成方案

从 GeneBank 中查询 hGH 的氨基酸序列信息, 采用 *E. coli* 偏爱的密码子逆翻译成 DNA 序列, 拆分成 60 bp 左右的单链寡核苷酸, 采用化学法和酶促法拼接成全长的双链基因.

### 1.3 实验方法

1.3.1 单链寡核苷酸的化学合成 利用 ABI3900 DNA 自动合成仪, 采用亚磷酰胺三酯固相合成法合成单链寡核苷酸.

1.3.2 DA-PCR 拼接及反应条件优化 外端单链寡核苷酸稀释为终浓度 100 nmol/L, 内部单链寡核苷酸稀释为终浓度 25 nmol/L. 反应体系: 10 倍 PCR 缓冲液 5  $\mu$ L, dNTPs(脱氧核糖核苷酸混合物)4  $\mu$ L, pfu DNA 聚合酶 0.5  $\mu$ L, 加水补足 50  $\mu$ L. PCR 反应条件: 94  $^{\circ}$ C, 30 s; 53  $^{\circ}$ C, 30 s; 72  $^{\circ}$ C, 1 min; 重复 20 个循环. 为了降低反应过程中的碱基错配率, 优化内部单链寡核苷酸浓度、内外部单链寡核苷酸浓度比和退火温度等 3 个条件.

1.3.3 T7 核酸内切酶 I 处理 将合成的中间片段进行凝胶电泳, 回收后, 溶于 T7 核酸内切酶 I 缓冲液中. 于 94  $^{\circ}$ C 变性 3 min, 缓慢降温至室温, 加入 0.5  $\mu$ L T7 核酸内切酶 I, 于 37  $^{\circ}$ C 温育 1 h, 于 55  $^{\circ}$ C

再温育 1 h 后, 用凝胶电泳回收。

1.3.4 OE-PCR 法拼接全基因 反应体系: 10 倍 PCR 缓冲液 5  $\mu\text{L}$ , dNTP 4  $\mu\text{L}$ , 回收产物 1  $\mu\text{L}$ , 引物 P1 为 1  $\mu\text{L}$ , 引物 P2 为 1  $\mu\text{L}$ , pfu DNA 聚合酶 0.5  $\mu\text{L}$ , 加水补足 50  $\mu\text{L}$ . PCR 反应条件: 94  $^{\circ}\text{C}$ , 30 s; 53  $^{\circ}\text{C}$ , 30 s; 72  $^{\circ}\text{C}$ , 2 min; 重复 20 个循环. 产物用凝胶电泳回收后, 再次使用 T7 核酸内切酶 I 处理。

1.3.5 测序载体 pUC-hGH 的构建 将合成的双链基因用 Sac I 和 Xho I 双酶消化. pUC18 质粒经双酶切后用碱性磷酸酶处理, 凝胶回收, 与目的基因于 16  $^{\circ}\text{C}$  连接, 过夜. 重组载体电击导入 *E. coli* JM109 感受态细胞中, 于 37  $^{\circ}\text{C}$  温育 1 h, 涂布于氨苄青霉素、X-gal 和 IPTG 的平板上过夜培养, 挑白色菌落, 提取质粒, 经酶切和 PCR 法鉴定后, 将阳性克隆送交生工测序。

## 2 结果与讨论

### 2.1 密码子优化

为了实现人生长激素在大肠杆菌中的高水平表达, 根据大肠杆菌密码子的偏爱性, 避免使用 *E. coli* 稀有的密码子, 将某些氨基酸残基的密码子进行替换, 保持编码氨基酸序列不变, 防止由于 tRNA 的衰竭对表达产生影响<sup>[9]</sup>. 人源基因中编码 Leu 的密码子 cta、编码 Arg 的密码子 agg 和 cgg、编码 Pro 的密码子 ccc 的数量分别为 4 个、5 个、1 个、5 个, 分别用密码子 ctg, cgt, cgt, ccg 代替. 这样在优化后的基因中不再含有 *E. coli* 中的稀有密码子, 使优化后的基因中 G + C 含量为 49%. 将 Frances 等<sup>[10]</sup>报道的人源 hGH 基因序列与全新设计的基因序列进行对比, 其相应的氨基酸与 DNA 序列如图 1 所示。

```

Amino acid sequence:  F P T I P L S R L F D N A M L R A H R L
Original DNA sequence: ttc cca acc att ccc tta tcc agg ctt ttt gac aac get atg ctc cgc gcc cat cgt ctg
Synthetic DNA sequence: ttc ccg acc atc ccg ctg tet cgt ctg ttc gac aac get atg ctg cgt gct cac cgt ctg

21 H Q L A F D T Y Q E F E E A Y I P K E Q
61 cac cag ctg gcc ttt gac acc tac cag gag ttt gaa gaa gcc tat atc cca aag gaa cag
61 cac caa ttg gct ttc gac acc tac cag gaa ttc gaa gag gcc tac atc ccg aaa gaa cag

41 K Y S F L Q N P Q T S L C F S E S I P T
121 aag tat tca ttc ctg cag aac ccc cag acc tcc ctc tgt ttc tca gag tet att ccg aca
121 aaa tac tet ttc ctg cag aac ccg cag acc tet ctg tgc ttc tet gaa tet atc ccg acc

61 P S N R E E T Q Q K S N L E L L R I S L
181 ccc tcc aac agg gag gaa aca caa cag aaa tcc aac cta gag ctg ctc cgc atc tcc ctg
181 ccg tct aac cgt gaa gaa acc cag cag aaa tct aac ctg gaa ctg ctg cgt atc tet ctg

81 L L I Q S W L E P V Q F L R S V F A N S
241 ctg ctc atc cag tgc tgg ctg gag ccc gtg cag ttc ctc agg agt gtc ttc gcc aac agc
241 ctg ctg atc cag tet tgg ctg gaa ccg gtt cag ttc ctg cgt tet gtt ttc gct aac tet

101 L V Y G A S D S N V Y D L L K D L E E G
301 ctg gtg tac ggc gcc tet gac agc aac gtc tat gac ctc cta aag gac cta gag gaa ggc
301 ctg gtt tac ggt get tet gac tct aac gtt tac gac ctg ctg aaa gac ctg gaa gaa ggt

121 I Q T L M G R L E D G S P R T G Q I F K
361 atc caa acg ctg atg ggg agg ctg gaa gat ggc agc ccc cgg act ggg cag atc ttc aag
361 atc cag acc ctg atg ggt cgc ctg gaa gac ggt tct ccg cgt acc ggt cag atc ttc aaa

141 Q T Y S K F D T N S H N D D A L L K N Y
421 cag acc tac agc aag ttc gac aca aac tca cac aac gat gac gca cta ctc aag aac tac
421 cag acc tac tet aaa ttc gac acc aac tet cac aac gac gac gct ctg ctg aaa aac tac

161 G L L Y C F R K D M D K V E T F L R I V
481 ggg ctg ctc tac tgc ttc agg aag gac atg gac aag gtc gag aca ttc ctg cgc atc gtc
481 ggt ctg ctg tac tgc ttc cgt aaa gac atg gac aaa gtt gaa acc ttc tta aga atc gtt

181 Q C R S V E G S C G F * *
541 cag tgc cgc tct gtg gag ggc agc tgt ggc ttc tag
541 cag tgc cgt tet gtt gaa ggt tet tgc ggt ttc tag tga
  
```

Fig. 1 Nucleotide and amino acid sequence of human growth hormone

### 2.2 序列全长及单链寡核苷酸的拆分

密码子全面优化的基因全长为 573 bp, 加上表达必须的调控元件, 拟合成序列全长为 1040 bp, 拆

分为 30 条平均长度为 56 bp, 相互重叠 20 bp 的单链寡核苷酸, 分别命名为 H1 ~ H30(图 2), 重叠部分  $T_m$  值平均为 53 °C.

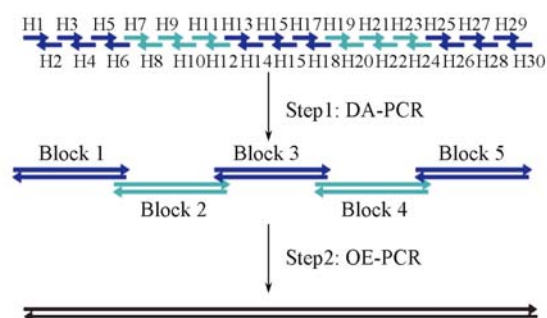


Fig. 2 Schematic diagram of the two-step gene synthesis method

### 2.3 DA-PCR 结果

为了降低重叠片段的错配引起的碱基错误, 必须优化反应体系及反应条件. 随着内部单链寡核苷酸浓度的增加, 非特异性扩增也随之增加, 内部单链寡核苷酸最佳浓度为 25 nmol/L 时. 外部与内部寡核苷酸最佳浓度比为 8:1. 退火温度对片段拼接没有明显的影响. 在最佳的反应体系和条件下, 将 30 条单链寡核苷酸每 6 条作为一组, 拼接出 5 条中间片段结果如图 3 所示.

### 2.4 OE-PCR 结果

合成的 5 条中间产物, 经 T7 核酸内切酶 I 处理后, 电泳并回收, 再加上两端的扩增引物 P1 和 P2, 通过 OE-PCR 进一步拼接成全长的基因. 将拼接产物进行电泳检测, 产物条带位于 1000 bp 处, 条带明亮、单一, 无非特异性条带.

### 2.5 测序载体 pUC-hGH 的构建

合成的基因经过 T7 核酸内切酶 I 处理后, 电泳并回收, 与 pUC18 载体经过 Sac I 和 Xho I 酶切消化, 形成两个不互补的黏性末端, 实现定向克隆. 为了进一步减少背景的干扰, 对双酶切消化的载体片段进行碱性磷酸酶处理, 去除其 3' 磷酸基团, 防止载体自连. 双酶消化的载体与基因于 16 °C 连接, 过夜. 转化到 *E. coli* JM109 感受态细胞中, 经蓝白筛选挑取白色菌落. 提取质粒用 Nde I 进行酶切及 M13 引物进行 PCR 鉴定.

### 2.6 测序结果

经过鉴定, 在 T7 核酸内切酶 I 处理前后各挑选 4 个阳性克隆, 送交上海生工进行测序. 测序结果(表 1)表明, 未经过 T7 核酸内切酶 I 处理的 4 个克隆碱基错误率为 0.79%, 无完全正确的基因. 而经过 2 次 T7 核酸内切酶 I 处理的 4 个克隆, 错误率只有 0.048%, 含有 2 个完全正确的基因, 碱基错误降低了 93.93%.

采用 T7 核酸内切酶 I 对拼接产物进行处理, 消除各种因素产生的碱基错误. 双链 DNA 于 94 °C 温育 3 min 后完全变性, 缓慢降温, 两条单链 DNA 分子之间严格地按照碱基互补配对原则进行复性. 含有突变碱基的单链 DNA 分子复性成异源二聚体, T7 核酸内切酶 I 能够识别并进行切割. 经过琼脂糖凝胶电泳可以将序列完全正确的核酸分子分离出来, 大大降低了基因合成过程中的错误率.

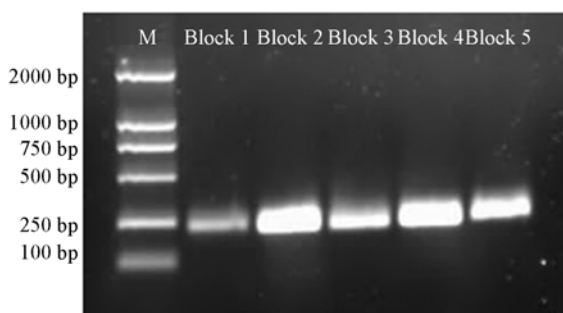


Fig. 3 Production of DA-PCR

Table 1 Summary of errors in hGH gene syntheses

Error type		Untreated	Treated
Base deletion	- G/C	4	0
	- A/T	10	1
Base insertion	+ G/C	9	0
	+ A/T	6	0
Nucleotide transition	G/C to A/T	2	1
	A/T to G/C	0	0
Nucleotide transversion	G/C to C/G	1	0
	G/C to T/A	0	0
	A/T to C/G	1	0
	A/T to T/A	0	0
Base error total		33	2
Base total		4160	4160
Error rate		0.0079	0.00048

## 参 考 文 献

- [ 1 ] Saugy M. , Robinson N. , Saudan C. . Br. J. Sports Med. [J], 2006, **40**(1) : 35—39
- [ 2 ] Nørrelund Helene. Growth Horm. Igf. Res. [J], 2005, **15**(2) : 95—122
- [ 3 ] Hintz R. L. . Brit. Med. J. [J], 2004, **328**(7445) : 907—908
- [ 4 ] Raben M. S. . J. Clin. Endocrinol Metab. [J], 1958, **18** : 901—903
- [ 5 ] David B. , Allen M. D. . Pediatrics [J], 2006, **118**(1) : 343—348
- [ 6 ] Klaus G. , Andreas P. . Biotechnol. J. [J], 2006, **1**(2) : 164—186
- [ 7 ] Kathleen M. , Julie C. , Nishihara Ilana S. , *et al.* . Proteomics [J], 2003, **3**(7) : 1365—1373
- [ 8 ] Lei Y. , Qihan D. . Nucl. Acid. Res. [J], 2004, **32**(7) : e59
- [ 9 ] LI Zhuo-Yu(李卓玉), SHI Ya-Wei(石亚伟), YUAN Jing-Ming(袁静明), *et al.* . Chem. J. Chinese Universities(高等学校化学学报) [J], 2002, **23**(3) : 394—398
- [10] Frances M. D. , David D. M. , Howard M. G. . Nucl. Acid. Res. [J], 1981, **9**(15) : 3719—3730

## Synthesis of Human Growth Hormone Gene by the Combination of Chemical and Enzymatic Method

ZHAO Xiang-Yang, GAO Chao-Hui, JIN Lu, WANG Xi, MENG Qing-Fan, TENG Li-Rong\*  
(College of Life Science, Jilin University, Changchun 130012, China)

**Abstract** According to the amino acid sequence and codon preference of *E. coli*, the human growth hormone (hGH) gene was optimized to avoid rare codons. The total length of the synthesized gene was 1040 bp because a promoter and a signal peptide sequence were added into it. Thirty oligonucleotides were designed and synthesized. The whole DNA sequence was synthesized by the improved two-step method. The sequencing result show that two complete correct gene were synthesised. The error rate of synthesized gene treated by T7 endonuclease I reduced 93.93% in comparison with the standard two-step synthesis method.

**Keywords** Human growth hormone; Gene synthesis; Two-step method

(Ed. : J, H, Z)