中 国 科 学
1988 年 4 月　　B 辑　　第 4 期

胰泌素基因的化学合成和克隆

钱苏雯　陈常庆　李载平
（中国科学院上海生物化学研究所）

摘　要

本文报道的内容为胰泌素基因的化学合成和克隆．胰泌素为一由 27 个氨基酸组成的肠胃道多肽激素，该多肽合成基因的顺序由计算机辅助设计，共 216 个核苷酸碱基；合成基因选用酵母细胞偏爱的密码子，并在其中设置了一些便于今后用于基因改造、表达调控和产物分离所需的位点，排除了可能影响酶促连接反应的各种重复顺序；用化学法（固相磷酸三酯法和固相亚磷酸三酯法）合成了 12 个基因片段，长度为 12 寡聚核苷酸至 24 寡聚核苷酸，聚丙烯酰胺凝胶电泳分离纯化合成产物；利用 T4 DNA 连接酶组建完整的胰泌素基因，克隆于 pWR 13 质粒中，构建了一个含胰泌素基因的新质粒 pWS 1．

近年来，DNA 合成技术发展迅速，它在重组 DNA 技术方面起着重要的作用．合成 DNA 技术的应用之一——构建多肽或蛋白的基因方面的研究工作已成果累累．许多多肽激素如：生长激素释放抑制因子[1]、胰岛素[2]、生长激素[3]、β-内啡肽[4]、表皮生长因子[5]、干扰素[6]、类生长因子[7]等已由合成基因的表达而获得，有些产品已投放市场应用，有的刚已处于中试阶段．国内脑啡肽和强啡肽的基因片段已化学合成[8,9]．　这里报道的是胰泌素基因的化学合成和克隆工作．

胰泌素是由 27 个氨基酸组成的活性多肽，它是由胃、十二指肠、空肠和回肠的 S 细胞分泌的消化道激素，具有重要的生理作用，它能促进胰的外分泌，增加胰液总量及所含的重碳酸盐，以中和十二指肠内的胃酸；同时还参与胃酸分泌的调节，增加非糖尿病患者的血液胰岛素水平．在临床上胰泌素可用于胰腺病患的诊断和十二指肠球部溃疡的治疗[10,11]．但天然来源限制了它的理论研究和临床应用．日本角慎一郎等人已于 1982 年化学合成了胰泌素基因并在大肠杆菌中得到了初步表达[12]．现我们也通过 DNA 化学合成构建了胰泌素基因，但在合成设计中使用的是酵母偏爱的密码子，并从多方面考虑可能影响基因表达的因素，设置了以该基因作为表达调控研究模型所需的各种酶切位点．整个基因分段合成后，克隆于 pWR 13 质粒中，得到一新质粒 pWS 1；按此设计得到的基因可能适于在酵母系统中表达．

一、实 验 材 料

1. 化学合成试剂

本文 1987 年 3 月 30 日收到修改稿．

(1) 溶剂

合成所用的溶剂在用于化学合成之前都经过除去可能存在的干扰反应的杂质和脱水的处理. 每1000ml分析纯吡啶中加入50g高锰酸钾迴流2h后蒸出; 再加60g对甲苯磺酰氯,迴流2h,蒸出;最后加100g氢氧化钾迴流2h后蒸出,保存于放有钾3 A或4 A型分子筛的棕色瓶中. 乙酸酐重蒸后,也保存于放有钾3 A或4 A型分子筛的棕色瓶中. 所有溶剂如: 吡啶、乙酸酐、二氯甲烷、二氧六环、乙酸乙酯均为上海试剂总厂分析纯级产品.

(2) 药品

四种脱氧核苷 dA, dC, dG, dT, 1,2,4-三唑, N, N, N′, N′-四甲基胍、吡啶-4-醛肟,3-氨基-1,2,4-三唑,1% 交联度的氨甲基聚苯乙烯树脂为美国 SIGMA 公司产品, 对氯苯酚,丁二酸酐为上海试剂总厂产品,对甲苯磺酰氯、三氯乙酸、二环己基碳二亚胺系余山化工厂产品,4,4′-二甲氧基三苯基氯甲烷为中国科学院上海生物化学研究所东风试剂厂产品.

4种保护核苷 ($^{DMT}dA^{B}Z$, $^{DMT}dC^{B}Z$, $^{DMT}G^{iBu}$, ^{DMT}dT) 主要按 G,S. Ti 的方法[13]制备, 对一氯苯磷酸二氯,2,4,6-三甲基苯磺酰硝基三唑 (MSNT) 以及带有四个保护核苷的聚苯乙烯树脂按文献[14—16]制备.

亚磷酸三酯法化学合成试剂为美国 ABI 公司产品.

2. 酶促连接和克隆所用的酶和试剂

(1) 菌体和质粒 DNA

大肠杆菌 JM 83 细胞和质粒 pWR 13 由中国科学院上海细胞生物学研究所郭礼和提供, 感受态细胞制备和保存方法按 Morrison 方法[17].

(2) 酶

T_4 多聚核苷酸激酶为中国科学院生物物理研究所生化工厂产品, T_4 DNA 连接酶, 限制性内切酶 PstI, EcoRI, BamHI 等由本组工具酶小组提供, PvuII 为美国 BRL 产品.

(3) 酶反应液和培养基等

T_4 DNA 连接酶反应液 50mmol/L Tris·HCl (pH 7.6),10mmol/L $MgCl_2$, 10mmol/L 二巯基苏糖醇 (DTT), 10mmol/L ATP, 10mmol/L 亚精胺.

T_4 多聚核苷酸激酶反应液 50mmol/L Tris·HCl (pH 7.6), 10mmol/L $MgCl_2$, 10mmol/L DTT, 10mmol/L 亚精胺.

内切酶 (RE) 反应液 50mmol/L Tris·HCl (pH7.6), 50mmol/L KCl, 10mmol/L $MgCl_2$, 10mmol/L DTT.

TE 缓冲液 10mmol/L Tris·HCl, 1mmol/L EDTA, pH8.0.

(γ-^{32}P) ATP (10mci/ml, 3000 ci/mmol/L), (α-^{32}P)ATP (1mci/ml, 410 ci/mmol/L) 为美国 Amersham 公司产品.

LB 培养基 10g 蛋白胨,5g 酵母抽提物, 5g 氯化钠溶于 1000ml 去离子水中, 高压灭菌.

LB 琼脂培养基 10g 蛋白胨,5g 酵母抽提物,5g 氯化钠,15g 琼脂粉, 溶于 1000ml 去离子水中,高压灭菌.

杂交所用的所有溶液按文献配制[18].

二、方法和结果

1. 基因的设计

胰泌素合成基因的顺序在计算机的辅助下设计完成，这里考虑以下因素：

(1) 密码的采用　选用酵母高表达蛋白所偏爱的密码[19]构建整个合成基因，并在编码胰泌素 27 个氨基酸的框架两端加上起始码 ATG 和终止码 TAG.

(2) 酶切位点的设置　在编码框架的 5′ 端和 3′ 端分别设置了内切酶 PstI 和 EcoRI 识别位点的粘性末端，便于合成基因直接、定向地克隆于质粒中；在 ATG 位点前设置了一个 HpaI 位点，HpaI 酶作用后给出平头末端，因而利用适当的接头片段可以使合成基因与表达质粒的蛋白的阅读框架相一致. 最后我们在 ATG 之后和 TAG 之前分别设置了 KpnI 和 SfaNI 位点，这二个位点可以用于除去 ATG 或 TAG，通过适当的接头片段，可将基因组建成串联式或带有一定的表达调控区的表达体系或便于表达蛋白分离纯化形式的阅读框架.

(3) 羟胺降解位点 (—Asn ↓ Gly—) 的设置　在 ATG 之前，设置了羟胺降解位点 (—Asn ↓ Gly—) 的编码区，可用于表达杂合蛋白的分离纯化，关于这一点将于后面加以讨论.

(4) 重复顺序的排除　基因顺序中存在的直向或反向重复序列会干扰合成片段的酶促连接反应；利用计算机检索出可能存在的重复顺序，在不影响上述各种位点的前提下，通过改变出现重复顺序区域的密码子来去除这些重复顺序，最后得到的合成基因顺序中已不存在 6 个以上核苷酸组成的重复顺序.

合成基因顺序确定后，将其分成 12 个片段，长度为 12 至 24 寡聚核苷酸，上、下链酶促连接片段之间存在至少由 6 个核苷酸组成的互补区，见图 1

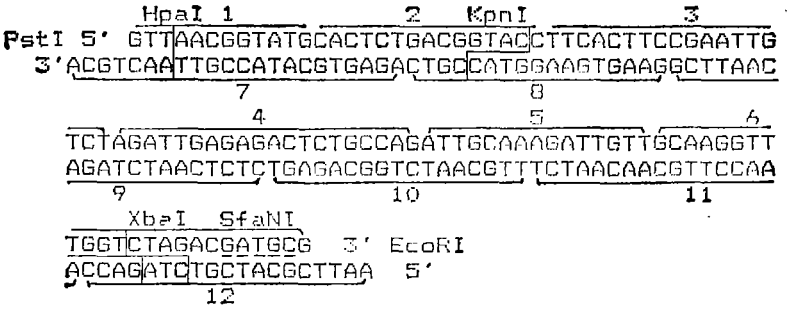

图 1　胰泌素合成基因顺序及分段情况

2. 寡聚核苷酸的化学合成

8 个寡聚核苷酸片段 (SE-1, SE-2, SE-3, SE-4, SE-5, SE-8, SE-9, SE-11) 通过固相磷酸三酯法[20]合成，固相合成的具体操作过程见表 1. 固相化学合成完成后，通过下列步骤从固相载体聚苯乙烯树脂上切下合成片段，除去磷酸基、碱基以及 5′端羟基上的保护基：将带有合成片段的固相载体先用 0.5mol/L N,N,N′,N′-四甲基胍/吡啶-4-醛肟的吡啶/水(9:1, V/V)溶液于 37℃ 下反应 24h，取上清，减压浓缩至胶状物后，再加浓氨水 (28%)，55℃下反应 4—5h，浓缩除去氨水；加入 80%冰乙酸水溶液，室温反应 30min；用乙酸乙酯抽提 2—3 次，样

表 1 固相化学合成循环

步 骤	溶剂或试剂	体积(ml)	振摇时间(min)	操作次数
1.洗涤	二氯甲烷	2	0.5	4
2.脱除三苯甲基	2%三氯醋酸/二氯甲烷	2	0.5	2
3.洗涤	A) 二氯甲烷 B) 吡啶	2	0.5	3
4.共蒸发脱水	单体/吡啶	0.5	/	2
5.缩合*	MSNT*/吡啶	0.3	60(25℃)	1
6.洗涤	吡啶	2	0.5	3
7.封闭-OH	10%乙酸酐/0.1%DMAP*/吡啶	2	5	1
8.洗涤	吡啶	2	0.5	3

* 缩合反应比例: 树酯上的核苷酸: 单体: MSNT = 1:5: 15μmol.
MSNT: 2, 4, 6-三甲基苯磺硝基三唑.
DMAP: 二甲基氨基吡啶.

品浓缩后,经葡聚糖 G-25 柱去盐,取 100pmol 粗样品进行 5′-^{32}P 标记,鉴定馏出液中指定长度的核苷酸片段的存在;最后用 20% 聚丙烯酰胺凝胶电泳在变性条件(7 mol/L 尿素)下分离出具有指定长度的寡聚核苷酸片段.

另外四个寡聚核苷酸片段 (SE-6, SE-7, SE-10, SE-12) 则通过固相亚磷酸三酯法合成[20], 整个化学反应操作在 DNA 合成仪 (ABI380A) 上进行, 得到的粗产物也是用聚丙烯

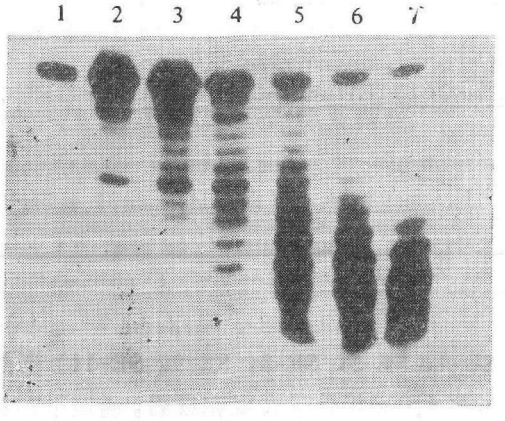

图 2 合成的寡聚核苷酸 SE-10 (17-mer) 5′-^{32}P 标记的放射自显影结果
(1 ——纯化的 17-mer; 2 至 7 ——未纯化的 17-mer 经 Sephadex G-25 柱洗脱的馏份)

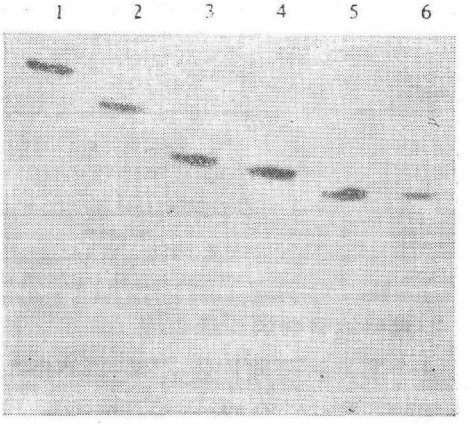

图 3 纯化的寡聚核苷酸 5′-^{32}P 标记的放射自显影结果
(1 —— SE-6, 24-mer 4 —— SE-3, 18-mer
2 —— SE-7, 22-mer 5 —— SE-10, 17-mer
3 —— SE-12, 19-mer 6 —— SE-11, 17-mer)

酰胺凝胶电泳分离纯化的，图 2 和图 3 示出了寡聚核苷酸的合成粗产物和纯化产物的 5′-^{32}P 标记的分析结果.

3. 胰泌素基因的组建

(1) 5′ 端磷酸化

除两侧 5′ 端片段 SE-1，SE-12 外，其余 10 个片段各取 1000pmol，真空抽干，在以下条件下进行磷酸化反应： 加 2μl 10 × 多聚核苷酸激酶反应液，2μl(γ-^{32}P) ATP (3.3pmol/ml，10μci/μl)，4 单位 T_4 激酶，10μl 水，37℃ 下反应 15min，再加入 1μl 10mmol/L ATP，

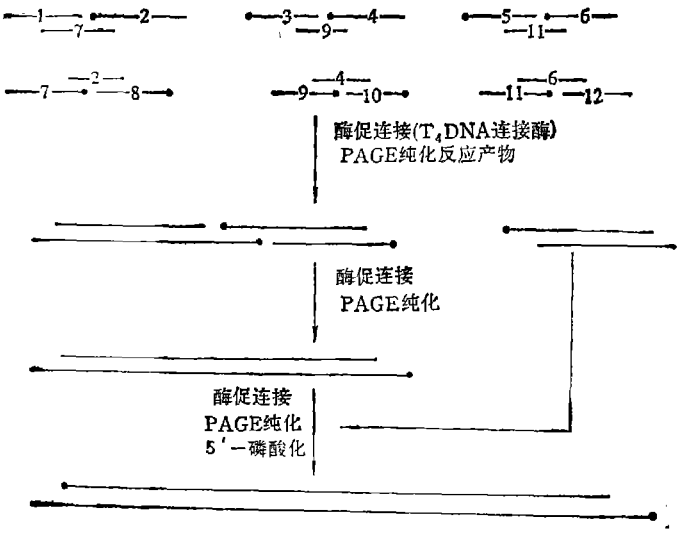

图 4　胰泌素合成基因的构建

(●—— 5′-磷酸基)

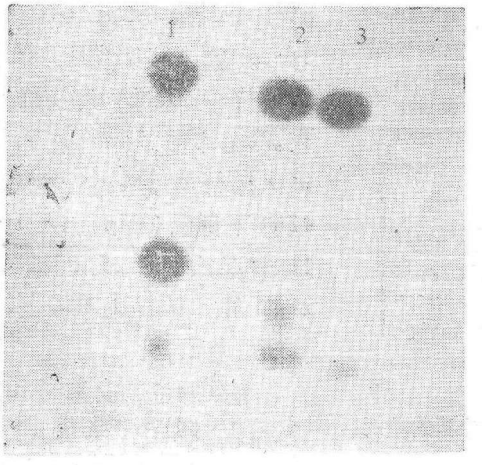

图 5　寡聚核苷酸酶促连接反应的放射自显影结果

(1——SE-7 + SE-8 (SE-2 为模板)　22-mer + 17-mer,
2——SE-9 + SE-10 (SE-4 为模板)　20-mer + 17-mer,
3——SE-11 + SE-12(SE-6 为模板)　17-mer + 19-mer)

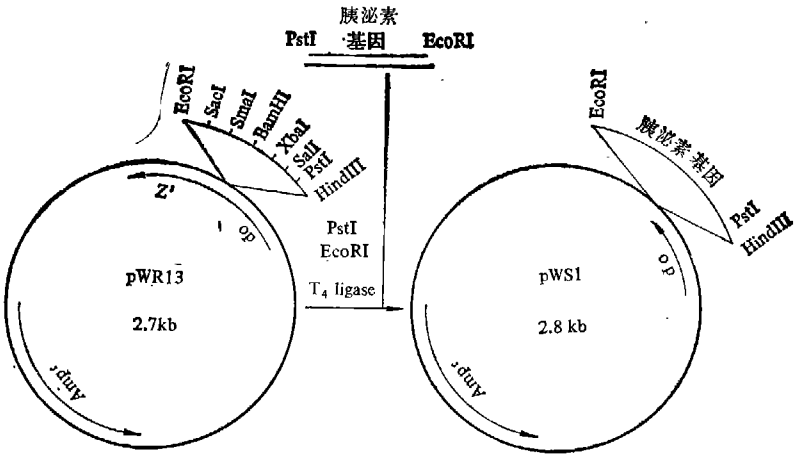

图 6 胰泌素合成基因的克隆

6 单位 T_4 激酶，37℃ 下反应 30min，最后在 95℃ 下保温 5min 使 T_4 激酶失活。

(2) 酶促连接反应

按图 4 混合等当量的寡聚核苷酸片段，混合物 95℃ 下保温 5min，缓慢冷却至室温使之相互配对，加入 2—4μl 10 × T_4 DNA 连接酶反应液，10 单位 T_4 连接酶，25℃ 下反应过夜，用聚丙烯酰胺变性胶分离连接产物（见图 5），整个胰泌素基因按图 4 所示连接而成。

4. 克隆和顺序分析

酶促连接得到的胰泌素基因按图 6 所示克隆于 pWR 13 中，1.2μg pWR 13 PstI-EcoRI 酶切片段和 1pmol 的胰泌素基因片段在 40μl 含有连接反应液和 2 单位 T_4 DNA 连接酶的溶液中于 22℃ 下反应过夜，70℃ 保温 5min 使 T_4 DNA 连接酶失活，加入 100μl 大肠杆菌 JM 83 感受态细胞，0℃ 保温 1h，然后于 42℃ 下保温 2min，加入 1ml LB 培养基，37℃ 下保温 1h，样品涂于含 15μg/ml 青霉素和 40μg/ml IPTG 及 x-gal 的 LB 琼脂平板上，置 37℃ 培养箱中培养夜过，挑出白色菌落。

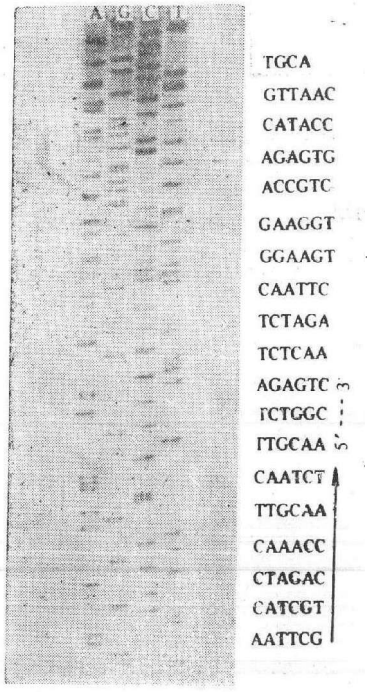

图 7 胰泌素合成基因序列分析的放射自显影结果

菌落原位杂交和 Southern 杂交按文献[21, 22]进行，以 SE-6 (24-mer) 作探针，杂交温度 50℃，其中有一菌落和从该菌落抽提出的 DNA 给出了阳性杂交结果。

利用双脱氧链终止法[23]在双链质粒上直接对克隆于 pWR 13 中的胰泌素基因进行了序列分析，证明合成基因的序列正确，见图 7。

三、讨　论

合成胰泌素基因的设计考虑了以下因素：(1) 选用酵母细胞偏爱的密码子，(2) 设置基因克隆以及增加合成基因在基因表达调控时的应用灵活性所需的酶切位点，(3) 排除可能影响酶促连接反应的重复顺序.

在低等的真核生物中酵母是一很好的表达体系，首先，当外源基因在酵母中表达时，酵母体内的酶类可以识别和加工外源蛋白的信号肽顺序，并使表达的外源蛋白分泌于细胞体外[24]；第二，α-因子引导下的外源基因的表达也可以给出分泌型的表达产物[25]，这两点对于表达产物的分离纯化是很有利的；另外，有的实验室发现在酵母细胞中表达的外源蛋白部分可被C端酰胺化，因此利用酵母表达体系可能是一条很有希望的解决C端酰胺化的途径.

天然的胰泌素C端缬氨酸为酰胺形式，而在大肠杆菌中表达的蛋白产物中C端一般并非酰胺化. 虽然已有文献报道在大肠杆菌中表达的非酰胺化胰泌素具有和天然蛋白相似的生物活性，但它的活性要比后者低，C端酰胺化可能是影响因素之一[26]. 1982 年美国 Braudbury 等人[27]报道在脑垂体中发现了引起C端酰胺化的酶，该酶可将蛋白前体中紧接着成熟蛋白的C端氨基酸的甘氨酸中的氨基转到前一氨基酸的羧基上造成C端酰胺化. 如果能得到这种酶或具有这种酶后加工活性的表达细胞体系，则多肽C端酰胺形式便可通过体外成体内后加工而获得.

合成基因中设置的羟胺 (NH_2OH) 降解位点 (—Asn $\overset{\downarrow}{-}$ Gly—) 可用于表达杂合蛋白的分离纯化. 从统计观点看，在一条长链蛋白质中出现 (—Asn—Gly—) 这样的由 2 个氨基酸组成的位点的几率要比单个—Met—位点的小得多，因此如果我们用羟胺先作用于表达的杂合蛋白，得到的多肽片段组分可能会比用溴化氰 (CNBr) 处理的简单得多，这样分离出在N—端带有—Gly—Met—的胰泌素片段就会方便得多；之后再用溴化氰除去 N—端—Gly—Met—，即可得到纯化的胰泌素多肽产物，但这种方法还未见文献报道过，是否可行有待证实.

参　考　文　献

[1] Itakura, K. et al., *Science*, **198**(1979), 1056—1063.
[2] Goeddel, D. V. et al., *Proc. Natl. Acad. Sci. USA.*, **76**(1979), 106—110.
[3] Martial, J. A. et al., *Science*, **205**(1979), 602—607.
[4] Shine, J. et al., *Nature*, **285**(1980), 456—461.
[5] Smith, J. et al., *Nucl. Acids Res.*, **10**(1982), 4467—4482.
[6] Nagata, S. et al., *Nature*, **284**(1980), 316—320.
[7] Sproat, B. S. et al., *Nucl. Acids Res.*, **13**(1985), 2959—2977.
[8] 汤锦炎等，中国科学 B 辑，1983，12：1105.
[9] 彭渤等，生物化学与生物物理学报，**18**(1986)，80—89.
[10] Zimmerman, M. J. et al., *Secretin, Cholecystokinin, Pancreozymin and Gastrin.*, Springer-Verlag, Berlin, 1973, 219—246.
[11] Zollinger, R. M. and Ellison, E. H., *Ann. Surg.*, **142**(1955), 709—728.
[12] 角慎一郎、铃木正则，化学の领域，**37**(1983)，6：19—28.
[13] G. S. Ti et al., *J. Am. Chem. Soc.*, 104(1982), 1316—1319.
[14] 池原森男、大塚荣子，蛋白质·核酸·酵素，**26**(1981)，4：259—269.
[15] Jones, S. S. et al., *Tetrahedron Letter*, **36**(1980), 3075—3085.
[16] Miyoshi, K. et al., *Nucl. acids Res.*, 8(1980), 5507—5517.
[17] Morrison, D. A., *Methods in Enzymol.*, 68(1978), 326—331.

[18] Maniatis, T. et al., *Molecular Cloning*, 1982, 446—448.
[19] Bennetzen, J. et al., *The Journal of Biological Chemistry*, **257**(1982), 3026—3031.
[20] Keiichi Itakura, et al., *Anu. Rev. Biochem.*, 53(1984), 323—356.
[21] Grunstein, M. et al., *Proc. Natl. Acad. USA.*, 72(1975), 3961—3965.
[22] Maniatis, T. et al., *Molecular Cloning*, 1982, 382—389.
[23] Guo, L. H. et al., *Nucl. Acids Res.*, 11(1983), 5521—5540.
[24] Mutt, V. et al., *Eur. J. Biochem.*, 15(1970), 513—519.
[25] Brake, A. J. et al., *Proc. Natl. Acad. USA.*, 81(1984), 4642—4646.
[26] Sumi, S. et al., *Gene*, 29(1984), 125—134.
[27] Bradbury, A. F. et al., *Nature*, 298(1982), 686—688.