

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

[12] 发明专利申请公开说明书

[21] 申请号 98808383.3

C12N 15/54
C12N 9/12 A61K 38/45
C07K 16/40 C12Q 1/68
C12Q 1/48 C12N 15/11
A61K 31/70

[43]公开日 2000年10月18日

[11]公开号 CN 1270634A

[22]申请日 1998.7.1 [21]申请号 98808383.3

[30]优先权

[32]1997.7.1 [33]US [31]60/051,410

[32]1997.7.19 [33]US [31]60/053,018

[32]1997.7.21 [33]US [31]60/053,329

[32]1997.8.4 [33]US [31]60/054,642

[32]1997.9.9 [33]US [31]60/058,287

[86]国际申请 PCT/US98/13835 1998.7.1

[87]国际公布 WO99/01560 英 1999.1.14

[85]进入国家阶段日期 2000.2.22

[71]申请人 卡姆比亚生物系统有限责任公司

地址 美国华盛顿

[72]发明人 A·奇利安 D·鲍泰尔

[74]专利代理机构 中国国际贸易促进委员会专利商标事
务所

代理人 樊卫民

权利要求书 5 页 说明书 53 页 附图页数 68 页

[54]发明名称 脊椎动物端粒酶基因和蛋白质及其用途

[57]摘要

本发明提供了编码脊椎动物端粒酶的核酸分子。还提供了基因产物、表达载体和适于表达端粒酶的宿主细胞。公开了鉴定端粒酶活性抑制剂的方法和抑制剂组合物。

I S S N 1 0 0 8 - 4 2 7 4

权利要求书

1. 一种分离的编码脊椎动物端粒酶的核酸分子。
2. 权利要求 1 的分离的核酸分子，其中所说的脊椎动物是人。
3. 权利要求 1 的核酸分子，其中的核酸分子包含图 1 中所示的序列，或者在正常严谨条件下可与图 1 中所示序列的互补序列杂交的序列，前提是该核酸分子不是 EST AA281296。
4. 权利要求 1 的核酸分子，其中的核酸分子编码图 1 或 11 中所示的氨基酸序列或其变体。
5. 一种分离的编码图 11 中所示的任何一个氨基酸序列的核酸分子，或者在正常严谨条件下可与编码图 11 所示氨基酸序列的互补序列杂交的核酸分子，前提是该核酸分子不是 EST AA281296。
6. 一种分离的含有图 10 中所示的任何一个序列的核酸分子，或者在正常严谨条件下可与图 10 所示序列的互补序列杂交的核酸分子。
7. 一种寡核苷酸，它含有图 1 中所示序列的 10 至 100 个连续核苷酸或其互补序列。
8. 一种寡核苷酸，它含有图 10 中所示序列的 10 至 100 个连续核苷酸或其互补序列。
9. 权利要求 7 或 8 的寡核苷酸，其中的寡核苷酸被标记。
10. 权利要求 9 的寡核苷酸，其中的标记是一种放射性标记、化学发光标记或生物素标记。
11. 一种表达载体，它含有一个与权利要求 1-6 的任一核酸分子有效连接的异源启动子。
12. 权利要求 11 的表达载体，其中的载体选自细菌载体、逆转录病毒载体、腺病毒载体和酵母载体。
13. 一种含有权利要求 11 或 12 的载体的宿主细胞。
14. 权利要求 13 的宿主细胞，其中的细胞选自人细胞、猴子细胞、小鼠细胞、大鼠细胞、酵母细胞和细菌细胞。
15. 权利要求 13 的宿主细胞，其中的细胞是人细胞。

16. 一种分离的含有脊椎动物端粒酶蛋白质的蛋白质。
17. 权利要求 16 的蛋白质，其中的脊椎动物是人。
18. 权利要求 16 的蛋白质，其中的蛋白质含有图 1 或 11 中所示的氨基酸序列或其变体。
 19. 脊椎动物端粒酶蛋白质的一部分。
 20. 权利要求 19 的部分蛋白质，其中该部分蛋白质的氨基酸序列示于图 1 中。
 21. 权利要求 19 的部分蛋白质，其中该部分蛋白质的氨基酸序列示于图 11 中。
 22. 权利要求 19 的部分蛋白质，其中该部分蛋白质为 10 至 100 个氨基酸长。
 23. 一种可与权利要求 16 或 19 的蛋白质特异结合的抗体。
 24. 一种可与由选自 1 区、 α 区、 β 区、2 区和 3 区的序列编码的多肽特异结合的抗体。
 25. 权利要求 24 的抗体，其中的抗体为单克隆抗体。
 26. 一种产生权利要求 14 的抗体的杂交瘤。
 27. 一种在正常严谨条件下能够与编码脊椎动物端粒酶特异性杂交的核酸探针，前提是该探针不与图 1 中所示的 1624-2012 位核苷酸杂交。
 28. 权利要求 27 的探针，其中的探针为 12 至 200 个核苷酸长。
 29. 权利要求 27 的探针，其中的探针为 20 至 50 个核苷酸长。
 30. 权利要求 17 的探针，其中的核酸分子具有图 1 中所示的序列或其互补序列。
 31. 权利要求 17 的探针，其中的核酸分子被标记。
 32. 一对能够特异扩增编码人端粒酶的核酸分子的全部或部分的寡核苷酸引物。
 33. 权利要求 32 的引物，其中的核酸分子含有图 1 中所示的序列或其互补序列。
 34. 权利要求 32 的引物，其中的核酸分子含有图 11 中所示的任何一个序列或其互补序列。

35. 权利要求 32 的引物，其中的引物对能够特异地扩增包括 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区的全部或部分的序列。

36. 权利要求 35 的引物，其中的引物位于图 1 中所示的 222、1950、2131-2166、2287-2468、2843、或 3157 位核苷酸的侧翼。

37. 权利要求 36 的引物，其中引物对中只有一个引物位于图 1 中所示的 222、1950、2131-2166、2287-2468、2843、或 3157 位核苷酸的侧翼，而引物对中的另一个引物具有相应于图 10 中所示的序列之一的序列或其互补序列。

38. 一对能够特异扩增图 10 中所示的基因组序列的寡核苷酸引物，其中的引物扩增至少 1 至 38 位核苷酸。

39. 一种可与 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区的核酸序列特异杂交的寡核苷酸。

40. 权利要求 39 的寡核苷酸，其中的寡核苷酸为 15 至 36 个碱基。

41. 一种诊断患者癌症的方法，包括制备肿瘤 cDNA，利用特异扩增人端粒酶核酸序列的引物扩增肿瘤 cDNA，其中端粒酶核酸序列的检测是癌症诊断的指征。

42. 权利要求 41 的方法，进一步包括比较端粒酶序列与对照的扩增量，其中端粒酶核酸序列比对照量的增加是癌症诊断的指征。

43. 权利要求 41 的方法，其中的引物跨越 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区，其中扩增的模式是癌症诊断的指征。

44. 权利要求 43 的方法，其中的引物为 Htel 内含子 T 和 Htel 723B。

45. 权利要求 44 的方法，其中的引物为 Htel335T 和 Htel1022B。

46. 一种确定细胞中端粒酶 RNA 表达模式的方法，包括从分离自细胞中的 mRNA 制备 cDNA，利用权利要求 35 的引物扩增 cDNA，由此确定端粒酶 RNA 表达的模式。

47. 权利要求 46 的方法，进一步包括通过与具有 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区序列的全部或部分的寡核苷酸杂交，检测扩增产物。

48. 一种诊断患者癌症的方法，包括确定端粒酶 RNA 表达的模式，

包括从由肿瘤 RNA 合成的 cDNA 扩增端粒酶，并通过与具有 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区序列的全部或部分的寡核苷酸杂交检测扩增产物，由此确定端粒酶 RNA 表达的模式，其中表达模式是癌症诊断的指征。

49. 权利要求 48 的方法，进一步包括将表达模式与从对照癌症所得的表达模式相比较。

50. 一种非人类转基因动物，其细胞中含有脊椎动物端粒酶基因，该基因与能够使该基因表达的启动子有效地连接。

51. 权利要求 50 的动物，其中的动物是小鼠。

52. 权利要求 50 的动物，其中的启动子是组织特异性的。

53. 权利要求 50 的动物，其中的端粒酶基因是图 11 中所示的任何一个核酸序列。

54. 一种小鼠，其细胞中具有通过与非功能性端粒酶基因进行同源重组而被破坏的内源性端粒酶基因，其中的小鼠不能表达内源性端粒酶。

55. 一种脊椎动物端粒酶活性抑制剂，其中的抑制剂可与端粒酶结合但不是核苷类似物。

56. 权利要求 55 的抑制剂，其中的脊椎动物是人。

57. 权利要求 55 的抑制剂，其中的抑制剂是与人类端粒酶 mRNA 互补的反义核酸。

58. 权利要求 57 的抑制剂，其中的反义核酸与 α 区、 β 区、2 区、3 区、或 X 区互补。

59. 权利要求 55 的抑制剂，其中的抑制剂是核酶。

60. 一种治疗癌症的方法，包括给患者施用治疗有效量的权利要求 55 的抑制剂。

61. 一种核酸分子，其含有由选自图 10 中所示的 1 区、 α 区、 β 区、2 区或 3 区的序列构成的序列及其变体。

62. 一种鉴定端粒酶活性效应物的方法，包括：

(a) 向端粒酶蛋白质、RNA 组分和模板混合物中加入候选效应物，其中的端粒酶蛋白质由分离的权利要求 1 的核酸分子编码；

(b) 检测端粒酶活性; 和

(c) 比较步骤(b)的活性与没有候选效应物的对照混合物的活性,
由此鉴定效应物。

63. 权利要求 62 的方法, 其中的效应物是抑制剂。

64. 权利要求 62 的方法, 其中的核酸分子编码人端粒酶。

说明书

脊椎动物端粒酶基因和蛋白质及其用途

技术领域

本发明通常涉及端粒酶，特别涉及人的端粒酶基因和蛋白质及其在诊断和治疗中的用途。

发明背景

非环状的染色体需要一种特化机制，用于在每次细胞分裂后维持染色体的末端，因为负责染色体 DNA 复制的聚合酶不能完全复制线性 DNA 分子，因而产生“末端复制问题”。为了解决这一问题，真核细胞依赖于一种酶，即端粒酶，将短的、特别是富含 G 的、相对保守的重复序列增加到染色体的末端。这些重复结构被称为端粒。

端粒的存在是细胞生存所必需的。即使缺少了一个端粒也会导致酵母（一种真核细胞）中的细胞循环停滞（Sandell 和 Zakian, 细胞 (Cell) 75: 729, 1993）。端粒在复制中缩短；端粒酶修复端粒。因此，如预料的那样，端粒酶的活性首先在活跃分裂的细胞中被检测到。象这样，端粒酶的活性在单细胞生物中是组成型存在，而在更复杂的生物中是被调节的，在胚系、胚胎组织和细胞以及肿瘤细胞中相对丰富。与此相反，在正常的人类体细胞中很难检测到端粒酶的活性。而且，不仅复制的终止会造成端粒酶减少，最新的数据显示端粒酶抑制可能是这个过渡期的一个关键事件。端粒酶与复制活性之间似乎直接的关系促使人们猜想端粒酶抑制剂可能是一种“通用的”癌症治疗物，基本上对所有的肿瘤类型都有效，而端粒酶的刺激物则可能克服所观察到的正常细胞的自然衰老。

受这些模型的激励，在分离与克隆端粒酶的过程中已经对端粒酶进行了大量的描述。已表明端粒延长的机制主要在于端粒重复序列的富含 G 的链。这条向染色体的 3' 端延伸的富含 G 的链是由端粒酶延伸的，

其是来自 RNA 组分的一种核糖核蛋白质，在此作为模板。已经分离并克隆了这种复合体的不同组分。复合体的 RNA 组分已被从许多不同的生物中分离并克隆，包括人（Feng 等人，科学（Science）269: 1236, 1995）、小鼠和其他哺乳动物、酿酒酵母（*Saccharomyces cerevisiae*）、四膜虫（*Tetrahymena*）、*Euplotes*、以及 *Oxytricha*（参考 Singer 和 Gottschling, 科学, 266: 404, 1994; Lingner 等人, 基因与发育（Genes & Develop.）8: 1984, 1994; 以及 Romero 和 Blackburn, 细胞, 67: 343, 1994）。蛋白质组分相对难分离。最近，已经测定了几种蛋白质组分的核苷酸序列（从四膜虫中分离的一个 80kD/95kD 二聚体蛋白质，WO96/19580；和从人体分离的一个 67kD 蛋白质，WO97/08314）。

本发明公开了端粒酶的核苷酸和氨基酸序列、这些序列在诊断和治疗中的用途，并提供其它相关的优点。

发明概述

一方面，本发明通常提供分离的编码脊椎动物端粒酶（及其变体）的核酸分子。脊椎动物的代表性例子包括哺乳动物如人、古老的猴子（例如猕猴、黑猩猩和狒狒）、狗、大鼠和小鼠，以及非哺乳动物类生物如鸟类。在一个优选实施方案中，提供了编码脊椎动物端粒酶的核酸分子，该核酸分子含有图 1 中所示的序列，或者在严谨条件下可与图 1 中所示序列的互补序列杂交，前提是该核酸分子不是 EST AA281296。

在另一个优选实施方案中，核酸分子含有图 11 中所示的任何序列或编码图 11 中所示的任何氨基酸序列，或者在正常严谨条件下可与其序列的互补序列杂交，前提是该核酸分子不是 EST AA281296。在其它实施方案中，核酸分子含有图 10 中所示的任何序列，或在正常严谨条件下可与其序列的互补序列杂交。

另一方面，本发明提供了含有图 1 所示的序列的 10 至 100 个连续核苷酸的寡核苷酸或其互补序列，以及图 10 所示序列的 10 至 100 个连续核苷酸的寡核苷酸或其互补序列。寡核苷酸可用可检测的标记物进行标记。

而在另一方面，本发明提供了一种表达载体，它含有一个与人端粒酶核酸分子有效相连的异源启动子。该载体可选自细菌载体、逆转录病毒载体、腺病毒载体和酵母载体。还提供了含有这些载体的宿主细胞。

另一方面，本发明提供一种分离的含有人端粒酶蛋白质的蛋白质。该蛋白质可含有图 1 所示的氨基酸序列或其变体，或含有图 11 所示的任一氨基酸序列或其变体。在一个相关方面，该蛋白质是人端粒酶蛋白质的一部分，它可衍生自图 1 或 11 所示的序列。在优选实施方案中，该蛋白质长度为 10 至 100 个氨基酸。

在其它方面，本发明提供了可与人端粒酶蛋白质或其部分蛋白质特异结合的抗体。

在一个优选方面，本发明提供了能够在正常严谨条件下与编码人端粒酶的核酸分子特异杂交的寡核苷酸（例如核酸探针或引物）。在某些实施方案中，核酸分子具有可检测的标记。在某些实施方案中，核酸分子是根据不能与图 1 所示的 1624-2012 位核苷酸杂交而选择的。在本发明的某些实施方案中，核酸探针或引物可能与野生型端粒酶序列有一个或多个核苷酸的差别。

在一个相关方面，本发明提供了能够特异扩增编码人端粒酶的核酸分子的全部或部分的一对寡核苷酸引物。在特定实施方案中，核酸分子含有图 1、图 11 中所示的序列或其互补序列。在优选实施方案中，这对引物能够特异扩增含有 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区的全部或部分的序列。在一个相关方面，本发明提供可与 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区中的核酸序列特异杂交的寡核苷酸。

本发明还提供了在患者身上诊断癌症的方法。这些方法包括制备肿瘤 cDNA，并利用特异扩增人端粒酶核酸序列的引物扩增肿瘤 cDNA，其中端粒酶核酸序列的检测是癌症诊断的一个指征。可以比较检测序列的数量同正常对照细胞中扩增的端粒酶序列的数量，其中与对照相比端粒酶核酸序列的增加是癌症诊断的指征。

而在另一方面，本发明提供了一种测定细胞中端粒酶 RNA 表达模式的方法，包括从由细胞中分离的 mRNA 中制备 cDNA，利用权利要求 35

的引物扩增 cDNA，由此测定端粒酶 RNA 表达的模式。在优选实施方案中，本方法还包括通过与具有 1 区、 α 区、 β 区、2 区、3 区、X 区或 Y 区的所有或部分序列的寡核苷酸杂交而检测扩增产物。这些方法可用于诊断患者的癌症，其中 RNA 的表达模式是癌症诊断的指示。

本发明还提供非人类转基因动物，其细胞中含有人端粒酶基因，该基因与能够使其表达的启动子有效地相连。在优选实施方案中，该动物是小鼠而启动子是组织特异性的。在一个相关方面，本发明提供一种小鼠，其细胞中的内源性端粒酶基因通过与无功能的端粒酶基因进行同源重组而被破坏，其中该小鼠不能表达内源性端粒酶。

本发明还提供人端粒酶活性的抑制剂，以及鉴定端粒酶活性抑制剂的方法，其中该抑制剂与端粒酶结合而且不是核苷类似物。抑制剂可能是一种与人端粒酶 mRNA 互补的反义核酸、核酶等等。该抑制剂可用于治疗癌症。

本发明还提供了鉴定端粒酶活性效应物的方法，包括一般步骤 (a) 向端粒酶蛋白质、RNA 组分和模板的混合物中加入候选效应物，其中端粒酶蛋白质是由按上述分离的核酸分子编码的；(b) 检测端粒酶的活性，以及 (c) 比较步骤 (b) 的活性与没有候选效应物的对照混合物的活性，由此鉴定效应物。在其它实施方案中，效应物是一种抑制剂。而在其它实施方案中核酸分子编码人端粒酶。

参考下列详述和附图，本发明的这些方面和其它方面将显而易见。另外，下面列出了各种参考，它们详细描述了某些方法或组合物（例如质粒，等等），并且因此被完整地引入本文作为参考。

附图简述

图 1A-E 表示人端粒酶的 DNA 序列 (SEQ ID NO:) 及预测的氨基酸序列 (SEQ ID NO:)。

图 2 表示 *Euplotes aediculatus* p123 (SEQ ID NO:)、酵母 (EST2) (SEQ ID NO:) 和人 (HT1) 端粒酶蛋白质 (氨基酸 29-1132) 序列的对比。图中指出了逆转录酶基元。三种蛋白质的高度同源区定义

为端粒酶区。这些序列是用 ClustalW 对比的。

图 3 是 Northern 分析的扫描图象，表示端粒酶催化亚单位在 LIM 1215 结肠癌细胞中表达，而在 CCD 原代成纤维细胞中不表达。一个大约 3.8kb 的 mRNA 与 hT1 探针杂交。另外一个高分子量的交叉杂交的 mRNA 在上面加箭头表示。指出了与 polyA+RNA 制备物中核糖体 (RNA) 的交叉杂交。相同的印迹还与来自 GAPDH 作为上样对照 (下面一组) 的探针杂交。分子量标记的大小以 kb 表示。

图 4 是 Southern 分析的扫描图象，表示端粒酶催化亚单位是由单基因编码的，并且不能在 LIM 1215 细胞中扩增。用 hT1 基因探针探测从人周围血细胞和 LIM 1215 细胞系中分离的基因组 DNA。印迹还含有稀释的探针质粒作为检测敏感性的对照。质粒大约被稀释为 10、5 和 1 个基因组当量。H, HindIII; E, EcoRI; P, Pst I; X, XbaI; B, BamHI。

图 5 表示从不同组织中合成的 cDNA 的扩增结果。扩增是利用来自 hT1 cDNA 序列的引物进行的，hT1 cDNA 序列横跨 hT1 基因中的内含子，将扩增产物印迹并用来自 hT1 序列的放射性标记寡核苷酸探测。还在同样的样品上用来自作为上样对照的 β -肌动蛋白基因的一对引物进行扩增。a: hT1 cDNA 对照; b: 人基因组 DNA 对照; c: 无模板对照; d: 正常结肠 RNA; e: 正常睾丸 RNA; f: 正常淋巴细胞 RNA; g: 黑素瘤 RNA (脑转移); h: 黑素瘤 RNA (皮下踝转移); i: 黑素瘤 RNA (肝转移); j: 黑素瘤 RNA (肺转移); k: 黑素瘤 RNA (腋淋巴结转移); l: 黑素瘤 RNA (皮肤转移); m: 乳腺癌 RNA; n: 乳腺癌 RNA; o: 乳腺癌 RNA; p: 乳腺癌 RNA。

图 6 表示的结果说明 hT1 在转接期前 (pre-crisis) 细胞和转接期后 (post-crisis) 细胞系中的表达。上图: 利用初始 EST 内的引物进行的嵌套扩增。下图: 利用 β -肌动蛋白引物进行的对照 RT-PCR。a: BET-3K 第 7 代细胞(p7)(转接期前); b: BET-3K 第 32 代细胞(转接期后); c: BFT-3K 第 14 代细胞(转接期前); d: BFT-3K 第 22 代细胞(转接期后); e: BFT-3B 第 15 代细胞(转接期前); f: BFT-3B 第 29 代细胞(转接期后); g: GM897(ALT); h: IICF/c(ALT); i: IICF-T/B1(ALT); j: 无模板对照。

图 7A-C 表示 hT1 转录物的几种不同的剪接模式。A, 六种剪接变体的示意图。B, 几个鉴定的 RNA 变体的组合。C, RNA 变体的推测的外显子/内含子接点的序列。在 A 部分标出了变体。给出了变体 3 的完整 DNA 序列 (及所翻译的蛋白质) (SEQ ID NO:)。标出了相应于一个可能的 c-Abl/SH3 结合位点的氨基酸。推测的外显子/内含子用|标出, 序列坐标 (coordinate) 如图 1。推测的剪接外显子为小写字母而推测的未剪接的内含子为大写字母。

图 8 表示在不同的肿瘤样品中 hT1 转录物的各种剪接模式。利用 HT2026F 和 HT2482R 引物, 在由 HT1875F 和 HT2781R 引物产生的原代 RT-PCR 产物上进行嵌套扩增 (14 个循环)。a: 肺癌; b: 淋巴瘤; c: 肺癌; d: 成神经管细胞瘤; e: 淋巴瘤; f: 淋巴瘤; g: T47D; h: 嗜铬细胞瘤; i: 淋巴瘤; j: 神经胶质瘤; k: 淋巴瘤; l: 无模板对照。

图 9 表示从 LIM 1215cDNA 合成的 cDNA 的扩增结果。如图中所示, 逆转录基元 A 被从含有序列 α 的剪接变体中删除掉。引物组合有: a, HTM2028F+HT2356R; b, HT2026F+HT2482R; c, HTM2028F+HT2482R; d, HT2026F+HT2482R。

图 10A-B 表示端粒酶变体区域的 DNA 序列。

图 11A-W 表示作为示范的变体端粒酶蛋白质的 DNA 和氨基酸序列。

图 12 是端粒酶活性分析的扫描图象。

图 13A-D 表示质粒 pAK128.4 的示意图及该质粒的 DNA 序列。

图 14A-E 表示质粒 pAK128.7 的示意图及该质粒的 DNA 序列。

图 15A-D 表示质粒 pAK128.14 的示意图及该质粒的 DNA 序列。

发明详述

在开始阐述本发明之前, 对此处用到的某些术语加以定义会有助于理解本发明。

如此处所用的, “野生型端粒酶” 通常指一种多肽, 它向染色体的末端酶促合成含有简单重复序列 (例如 CCCTAA, 参考 Zakian, 科学 270: 1601, 1995) 的核酸序列。已推测出一个代表性的人野生型端粒酶的氨

氨基酸序列，并示于图 1 (SEQ ID NO.)。在本发明中，应当理解本发明的端粒酶不仅包括野生型蛋白质，还包括野生型蛋白质序列的变体（包括等位基因）。这些变体不必表现出酶促功能。简单地说，这些变体可以由天然多态性造成的，包括 RNA 剪接变体，通过遗传重组形成，或是通过重组方法学合成的，而且，这种变体与野生型蛋白质的区别可以是一个或多个氨基酸的替换、插入、缺失、重排等等。典型地，当变体是合成的时，氨基酸替换是保守的，就是说，氨基酸替换在极性、非极性、芳香的、带电荷的等等相同组中的氨基酸。在逆转录酶基元区与野生型序列同源的区域内，变体与野生型具有至少 90% 的氨基酸序列同一性，在某些实施方案中，具有大于 92%、95%，或 97% 的同一性。在逆转录酶基元区以外，变体优选地具有 75% 的氨基酸同一性，在某些实施方案中，具有至少 80%、85%、90%、92%、95% 或 97% 的同一性。

本领域的技术人员将知晓，一个编码端粒酶的核苷酸序列可以不同于附图中的野生型序列；这是由于密码子的简并性、核苷酸多态性，或氨基酸的差异造成的。在其它实施方案中，变体应在正常严谨的条件下优先与野生型核苷酸序列杂交，杂交条件大约是比较天然双链的 T_m 值低 25-30°C（例如 1M Na⁺，65°C；5X SSPE，0.5% SDS，5X Denhardt 氏溶液，65°C 或同等条件；通常参考 Sambrook 等人，《分子克隆实验手册》(Molecular Cloning: A laboratory Manual)，第二版，冷泉港出版，1987；Ausubel 等人，《现代分子生物学技术》(Current Protocols in Molecular Biology)，Greene 出版公司，1987)。其它短寡核苷酸的 T_m 值可通过下面的公式计算出： $T_m = 81.5 + 0.41\%(G+C) - \log(Na^+)$ 。低严谨杂交是在大约低于 T_m 值 40°C 的条件下进行的，而高严谨杂交是在大约低于 T_m 值 10°C 的条件下进行的。在逆转录酶基元区变体优选地与野生型序列具有至少 75% 的同一性，优选地至少 80%、85% 同一性，而最优选地具有至少 90% 的核苷酸同一性。

如此处所用的，“启动子”指含有指导与其相连基因转录的元件的一种核苷酸序列。最低条件是，启动子含有 RNA 聚合酶结合位点。更为典型的是，在真核生物中，启动子序列含有控制基因表达的速度和时间的

其它转录因子的结合位点。这种位点包括 TATA 盒、CAAT 盒、POU 盒、AP1 结合位点等等。启动子区还可以含有增强子元件。当一个启动子与一基因相连以使该基因得以转录时，就称为“有效连接”。

“分离的核酸分子”指一种多核苷酸分子，其形式为独立的片段或一个较大的核酸构件的一个成分，是从其来源细胞中分离的（包括它正常所在的染色体）至少曾经为基本上纯的形式。核酸分子可以包括各种核苷酸，包括 DNA、RNA、核苷酸类似物，或其组合。

I. 端粒酶、端粒酶基因和基因产物

如上所述，本发明提供了涉及脊椎动物端粒酶基因和基因产物的组合物，以及利用这些基因和基因产物的方法。根据此处提供的公开内容，端粒酶基因可从表达端粒酶的各种细胞类型中分离到，包括无限增殖化细胞或转化细胞。如此处所列举的，来自人细胞的编码端粒酶的 cDNA 和变体被鉴定、分离，并表征。通过用编码端粒酶的表达载体转化的宿主细胞可以很容易地生产端粒酶蛋白质。

A. 端粒酶基因的分离

如此处所述，本发明提供编码端粒酶的基因。在本发明的一个实施方案中，一个编码人端粒酶的基因可以利用从 EST 序列中设计的引物对通过 cDNA 文库扩增而鉴定。EST 序列 GenBank Accession No. AA281296，是通过与 *Euplotes aediculatus* 端粒酶基因 (GenBank Accession No. U95964; Lingner 等人，科学 276: 561, 1997) 的序列同一性和相似性鉴定的。*Euplotes* 端粒酶基因和 EST 的序列对比显示大约 38% 的氨基酸同一性和 59% 的氨基酸相似性。

端粒酶基因可从基因组 DNA 或 cDNA 中分离到。当需要启动子区或其它侧翼区时，优选基因组 DNA。在染色体载体如 YAC (酵母人工染色体)、细菌载体如 λ EMBL3、 λ gt10、粘粒或质粒中构建的基因组 DNA 文库，和在细菌载体 (例如 λ ZAPII)、质粒或其它载体中构建的 cDNA 文库都是适合筛选的。这种文库可以利用本领域的方法和技术进行构建

(参考 Sambrook 等人,《分子克隆实验手册》,冷泉港出版,1989)和从商业渠道购买(例如,Clontech 公司, Palo Alto, CA)。DNA 可以从脊椎动物细胞中分离,如人细胞、小鼠细胞,其它啮齿类动物或灵长类动物细胞、鸟类细胞等等。

在一个实施方案中,端粒酶基因是利用 cDNA 文库的 DNA 为模板通过扩增分离的。利用报道的 EST 序列,可以分离人端粒酶。简单地说,就是根据 EST 核苷酸序列设计几套扩增引物。这种引物的例子列于表 2 (还可参考实施例 1)。从具有高端粒酶活性的细胞中制备的 cDNA 文库的扩增是优选的。此处描述的引物扩增了一个片段,该片段具有从来自 LIM1215 cDNA 文库的 EST 序列预测的长度。LIM1215 是一人结肠癌细胞系。通过 DNA 序列分析证实该片的性质。

利用于一个可与 EST 引物之一相连的载体序列杂交的引物,在反应中扩增了含有附加序列的 DNA 片段。通过利用克隆位点任意一侧的载体引物,与 EST 引物结合使用,分离了一个衍生自 h-TEL(人端粒酶)的 3' 区的 1.6kb 片段和一个衍生自 5' 区的 0.7kb 片段。利用一对 EST 序列内部引物,通过扩增证实了这些片段含有端粒酶编码序列。将这两个片段克隆到 pBluescript 并进行 DNA 序列分析。附加的 DNA 序列是通过 C-RACE 和扩增技术得到 cDNA 的 5' 端,以及通过从 cDNA 文库中杂交并分离克隆获得的。

一个对照人端粒酶的编译 DNA 序列和预测的氨基酸序列示于图 1。如图中所示,对照端粒酶的编码区为 3396 个碱基长并有一段大约 620 个碱基长的 3' 非翻译区。预测的氨基酸序列为 1132 个氨基酸长并被描绘为四个主要结构域: N-端、基本、逆转录酶(RT)和 C-端。而且,人端粒酶含有与其它端粒酶(例如来自 *Euplotes* 和 *S. pombe* 的端粒酶)和逆转录酶同源的区域。这些基元在本文和在 Kilian 等人(人类分子遗传学 (Human Molecular Genetics), 12: 2011-2019, 1997)中被鉴定为结构域 1、2、A、B、C 和 D, 在 Nakamura 等人(科学, 277: 955-959)中为结构域 1、2、A、B'、C、D 和 E, 而在 Meyerson 等人(细胞, 90: 785-795, 1997)中为基元 1-6。不管采用的名称如何,这些基元都含有氨基酸 621-

626 (基元 1) 和 631-634 (基元 2)、708-720 (基元 A)、827-839 (基元 B)、863-871 (基元 C), 以及 895-902 (基元 D)。因为这些基元的边界是根据与其它端粒酶的相似性和同一性确定的, 所以各个基元的功能性边界可能不同。

另外, 对照端粒酶序列的变体是通过扩增获得的, 这在本文中有描述。其 DNA 和预测的氨基酸序列示于图 11 中, 并在下面有更详细的讨论。简单地说, 这些变体中的一些变体编码被截短的蛋白质, 而其它变体具有不同的 C-端序列。这些变体似乎是由于不同的 RNA 剪接造成的, 因为端粒酶似乎是人类中的一个单拷贝基因 (见实施例 2)。

或者, 其它方法也可用于获得编码端粒酶的核酸分子。例如, 通过利用抗体或可与端粒酶反应的抗体进行筛选, 从表达文库中获得编码端粒酶的核酸分子 (参考 Sambrook 等人, 《分子克隆实验手册》, 第二版, 冷泉港出版, NY, 1987; Ausubel 等人, 《现代分子生物学技术》, Greene 出版公司和 Wiley-Interscience 公司, NY, 1995)。在另一个实施方案中, 编码端粒酶的核酸分子可以通过杂交筛选 cDNA 文库或基因组文库获得。用于杂交筛选的寡核苷酸可以根据此处提供的人端粒酶的 DNA 序列进行设计。用于筛选的寡核苷酸通常至少 11 个碱基长, 更常用的是至少 20 或 25 个碱基长。在一个实施方案中, 寡核苷酸为 20-30 个碱基长。这样的寡核苷酸可以通过自动化方式合成。为了便于检测, 寡核苷酸可以方便地加以标记, 通常是在 5' 端用一个报告分子, 如一种放射性核素 (如 ^{32}P)、酶标记、蛋白质标记、荧光标记或生物素标记。根据载体的不同, 文库通常为菌落或噬菌体, 并将重组 DNA 转移到尼龙膜或硝酸纤维素膜上。杂交条件根据寡核苷酸的长度和 GC 含量加以调整。经过变性、中和, 和将 DNA 固定到膜上, 使膜与标记探针杂交。合适的杂交条件可以在 Sambrook 等人, 同上, Ausubel 等人, 同上中找到, 而且杂交溶液中可以含有添加剂如四甲基氯化铵或其它离液剂或亲水剂以增加杂交的特异性 (参考 PCT/US97/17413)。杂交后, 利用合适的检测方法显示出杂交菌落或噬菌体, 并且随后将其分离并增殖。候选菌落或扩增片段可以通过任何一种方法来证实其含有端粒酶 DNA。例如, 候选菌落可以与第

二个非重叠探针杂交或进行 DNA 序列分析。通过这些方法，适用于本发明的含有端粒酶基因或基因片段的菌落就被分离出来。

端粒酶 DNA 还可以通过 cDNA 或基因组 DNA 扩增获得。用于扩增全长 cDNA 的寡核苷酸引物优选地衍生自位于编码区的 5' 或 3' 端的序列。基因组序列的扩增将利用跨越内含子序列的引物，并且采用适于长扩增产物的条件（参考 Promega 目录）。简单地说，用作扩增引物的寡核苷酸优选地不能具有自我互补序列，其 3' 端也不能有互补序列（以防止引物二聚体的形成）。优选地，引物具有约 50% 的 GC 含量并且具有限制性位点以便于克隆。通常，引物为 15 至 50 个核苷酸长，更优选地 20 至 35 个核苷酸长。引物与 cDNA 或基因组 DNA 退火并进行足够的扩增循环以产生可检测的产物，优选地是一种可以容易地通过凝胶电泳和染色观察到的产物。扩增片段被纯化并插入到载体（例如，病毒、噬菌粒或质粒载体，例如 λ gt10 或 pBS(m13+)）中并增殖。

来自许多品种的端粒酶基因可以利用此处提供的组合物进行分离。对于密切相关的品种，人的序列或其部分序列可用作探针筛选基因组或 cDNA 文库。例如，含有催化位点（大约相应于图 1 的 605-915 位氨基酸）的端粒酶基因的一个片段可被标记，并用作探针筛选从小鼠、灵长类动物、大鼠、狗、或其它脊椎动物、温血动物、或哺乳动物类构建的文库。在正常严谨条件下进行一个初始杂交可以得到编码端粒酶的克隆或片段。如果没有观察到杂交，可以进行松（低）严谨条件杂交。改变杂交严谨度的指导方法可以从 Sambrook 等人，同上，以及其它众所周知的途径中获得。这样的探针还可以用于进化上具有多样性的品种如果蝇，尽管杂交条件通常更为松弛。

其它方法也可以用于从非人品种中分离端粒酶基因。这些方法包括，但不限于，利用来自保守区（如逆转录酶基元）的引物进行扩增，利用来自端粒酶的不同区域（包括逆转录酶区）的简并引物进行扩增，利用抗体探测表达文库，利用端粒酶 RNA 探测表达文库等等。通过氨基酸相似性和/或核酸同一性鉴定一基因序列为端粒酶基因。通常，氨基酸相似性，是鉴定端粒酶所优选的，它允许有保守性的差别。对于不同的品种，

氨基酸相似性通常至少为 30%，优选地至少 40%或至少 50%。核酸同一性可能较低所以难以评价。几种方便易得的计算机分析程序，如 BLASTN 和 BLASTP，有助于确定基因和基因产物的相关性。候选的端粒酶基因的检测是利用此处描述的一种功能分析方法或其它相当的方法检测其酶活性。

B. 变体端粒酶基因

此处提供的端粒酶核酸或氨基酸序列的变体（包括等位基因），可以很容易地从天然变体（如多态性，剪接变体，突变体）中分离、合成、或构建。根据用途不同，突变体可以构建成表现不同的或有缺陷的端粒酶功能。特别有用的端粒酶基因编码一种缺少酶活性但具有显性失活表型的蛋白质。而且端粒酶变体缺失一种或多种已知的端粒酶活性，包括逆转录酶活性、溶核活性、端粒结合活性、dNTP 结合活性，以及端粒酶 RNA (hTR)结合活性。

本领域的技术人员知道现在已经建立了许多形成突变体的方法（通常参考，Sambrook 等人，同上；Ausubel 等人，同上）。简单地说，形成少量核苷酸替换突变体的优选方法是利用跨越待突变的碱基并含有突变的碱基的寡核苷酸。寡核苷酸可与互补单链核酸杂交，而第二条链从寡核苷酸处延伸合成。相似地，缺失和/或插入突变体可以通过各种任何已知的一种方法进行构建。例如，将基因用限制性酶消化并重新连接，这样一些序列被删除，或者与分离的具有粘性末端的片段相连，这样就产生了具有插入或大的替换的突变体。在另一个实施方案中，变体是通过“外显子重排 (exon shuffling)”形成的（参考美国专利 No.5,605,793）。变体序列还可以通过“分子进化(Molecular Evolution)”技术形成（参考美国专利 No.5,723,323）。其它形成变体序列的方法可以在如 Sambrook 等人（同上）和 Ausubel 等人（同上）中找到。变体序列的验证通常是通过限制性酶作图、序列分析、或探针杂交进行的，尽管其它方法也可以采用。将双链核酸转化宿主细胞，通常是大肠杆菌，但是其它原核生物、酵母、或大型真核生物也可以采用。标准的筛选技术，例如核酸杂

交、扩增、以及 DNA 序列分析，将鉴定突变体序列。

在优选实施方案中，变体端粒酶的酶活性失活并赋予宿主细胞以显性失活的表型。不管实际机制如何，当显性失活的端粒酶在宿主细胞中表达时，天然有活性的端粒酶就被失活。在催化结构域中，逆转录酶基元具有保守的天冬氨酸残基。人端粒酶还含有这些重要的残基：Asp712, Asp718, Asp868, 和 Asp869。这些天冬氨酸残基的一个或多个突变为非保守性的氨基酸（例如丙氨酸）将可能破坏酶促活性或影响端粒的缩短。对于每个这样的突变体，检测其显性失活现象。在某些实施方案中，优选的突变体是显性失活的并引起老化的表型。其它显性失活变体可以通过删除一个或多个逆转录酶基元或改变某些区域而形成，这些区域与 DNA 引发(priming)（如基元 E）、RNA 组分结合位点、模板结合位点、金属离子结合位点（如基元 C）等等有关。

在其它实施方案中，编码端粒酶的核酸分子可以与另一个核酸分子融合在一起。我们将会看到，融合伴侣基因在某些实施方案中会贡献出一个编码区。于是，可能需要仅仅利用端粒酶的催化位点（例如氨基酸 609-915）、单个逆转录酶基元（上述的）、此处描述的任何剪接变体端粒酶、端粒酶 RNA 结合位点等等。融合伴侣选择部分取决于应用的目的。融合伴侣可被用于改变端粒酶的特异性、提供一个报告基因功能、提供一个用于鉴定或纯化技术的标签序列，等等。报告序列或标签序列可以是任何一种蛋白质，它使得可以方便地并敏感地测量或分离基因产物，而不干扰端粒酶的功能。对于报告基因功能， β -葡糖醛酸糖苷酶（美国专利 No:5,268,463）、绿色荧光蛋白质和 β -半乳糖苷酶都是很容易得到的 DNA 序列。肽标签是一个短的序列，通常衍生自天然蛋白质，可以被抗体或其它分子所识别。肽标签包括 FLAG[®]、Glu-Glu 标签（Chiron 公司，Emeryville, CA）KT3 标签（Chiron 公司）、T7 基因 10 标签（Invitrogen, La Jolla, CA）、T7 主要衣壳蛋白质标签（Novagen, Madison, WI）、His6（六组氨酸）、以及 HSV 标签（Novagen）。除了标签以外，其他类的蛋白质或肽，例如谷胱甘肽-S-转移酶也可以采用。

C. 由端粒酶基因衍生的片段和寡核苷酸

另外，可以分离或构建部分端粒酶基因或基因片段以用于本发明。例如，通过众所周知的方法可以从模板 DNA，如质粒 DNA 中分离到限制性片段，而 DNA 片段，包括限制性片段可以通过扩增产生。而且，寡核苷酸可以经合成或从重组 DNA 分子中分离到。本领域的技术人员知道，还有其它方法也可用于获得具有至少部分端粒酶序列的 DNA 或 RNA 分子。而且，对于特殊的应用，这些核酸可以通过本领域已知的技术，利用放射性标记（例如， ^{32}P 、 ^{33}P 、 ^{35}S 、 ^{125}I 、 ^{131}I 、 ^3H 、 ^{14}C ）、荧光标记（例如，FITC、Cy5、RITC、Texas 红）、化学发光标记、酶、生物素等等进行标记。

获得这些片段的方法在本领域是众所周知的。特别适用于本发明的端粒酶部分含有催化位点、单个逆转录酶基元、推测的内含子序列（参考图 10）等等。寡核苷酸通常是通过自动化方式合成的；合成的方法和仪器是很容易得到的（例如应用生物系统公司，CA）。寡核苷酸可以含有非天然形成的核苷酸，例如核苷酸类似物、被修饰的骨架（如肽骨架）、核苷酸衍生物（如生物素化的核苷酸），等等。如此处所用的，寡核苷酸是指至少约 7 个核苷酸并通常不超过约 100 个核苷酸的核酸序列。通常，寡核苷酸大约为 10 至 50 个碱基之间，更常用的是大约 18 至 35 个碱基之间。寡核苷酸可以是单链的，或在某些情况下是双链的。如此处所用的，部分核酸是指含有比完整的母核酸序列少的多核苷酸。例如，端粒酶编码序列的一部分含有少于全长的端粒酶序列。一个“部分”通常至少是约 7 个核苷酸，也可以多至 10、20、25 或更多的核苷酸。一个片段是指具有任意长度的多核苷酸分子，并可以包括寡核苷酸，尽管常用但并不限于，术语寡核苷酸用于指短的多核苷酸，而术语片段用于指长的多核苷酸。

被用作扩增引物和杂交筛选探针的寡核苷酸可以根据此处给出的人端粒酶的 DNA 序列进行设计。用于扩增全长 cDNA 的寡核苷酸引物优选地衍生自 5' 和 3' 端的序列。选择扩增特定区域的引物用于形成大小容易检测的产物。在优选实施方案中，选择位于要进行其它 RNA 剪接的序

列侧翼的引物。在优选实施方案中，选择了一套引物，使得跨越被剪接进（spliced-in）的序列的产物和跨越被剪接出（spliced-out）的序列的产物都具有合适的大小，在相同的条件下可被检出。在其它实施方案中，用两套引物检测其它剪接的 RNA。例如一套引物位于剪接结合部侧翼以检测被剪接出的产物。第二套引物非常靠近结合部（这样被剪接出的扩增产物与引物二聚体具有相同的大小或比引物二聚体略长）或一套或多套引物来自被剪接进序列（这样被剪接出的 RNA 不会产生任何产物）。

扩增引物优选地没有自我互补序列，其 3' 端也没有互补序列（以防止形成引物二聚体）。优选地，引物具有大约 50% 的 GC 含量并含有限制性位点以便于克隆。扩增引物通常至少 15 个碱基并且通常不长于 50 个碱基，尽管在一些情况和条件下可使用更短的或更长的引物。更为常用的引物为 17 至 40 个碱基长、17 至 35 个碱基长、或 20 至 30 个碱基长。引物与 cDNA 或基因组 DNA 退火并进行足够的扩增循环，通常是 20 至 40 个循环，以产生通过凝胶电泳和染色或通过杂交容易检测的产物。扩增片段被纯化并插入载体如 λ gt10 或 pBS(M13+) 中，增殖、分离并进行 DNA 序列分析，进行杂交等等。

适用于筛选基因组、cDNA 或其它类型（如突变体端粒酶序列）文库，探测 Southern、Northern、或 Northwestern 印迹、扩增产物等等的寡核苷酸杂交探针可以根据此处提供的序列进行设计。用于杂交的寡核苷酸通常至少 11 个碱基长，一般长度少于 100 个碱基，优选至少 15 个碱基长，至少 20 个碱基长，至少 25 个碱基长，而优选地为 20-70、25-50、或 30-40 个碱基长。为了便于检测，寡核苷酸可以被方便地予以标记，通常在 5' 端用报告分子，如放射性核素（如 ^{32}P ）、酶标记、蛋白质标记、荧光标记、或生物素进行标记（参考 Ausubel 等人，和 Sambrook 等人，同上）。根据载体的不同，文库通常为菌落或噬菌体，而重组 DNA 被转移到尼龙膜或硝酸纤维素膜上。经过变性、中和、并将 DNA 固定到膜上，使膜与标记探针杂交，并洗膜。利用合适的检测方法会显示出杂交菌落或噬菌体，然后将其分离并增殖。将核酸转移至膜上进行杂交的方法是众所周知的。在某些实施方案中，加入杂交溶液的添加剂，例如离液

剂（如四甲基氯化铵）或亲水剂（如三氯乙酸铵；参考 PCT/US97/17413）以增加杂交的敏感性和特异性。如果在经过了与杂交条件相当的条件下（此处表示为低于 T_m 值的度数）的洗膜后，探针仍然保持在可检测的退火状态下，该探针就特异地与核酸杂交。

D. 人端粒酶的剪接变体

除了图 1 的对照端粒酶 DNA 和蛋白质序列以外，还观察到几种 RNA 剪接变体。尽管一些变体反映了不完全加工的 mRNA，值得注意的是这种变体在预先选择的多腺嘌呤化 mRNA 的 RNA 样品（LIM1215）中非常丰富。这些发现，以及它们在 RT 结构域中的成簇情况，说明插入变体似乎更反映了 hT1 蛋白质表达的调节情况。例如，外显子被删除的变体（参考 α , β , 图 7）似乎是编码变体蛋白质的不同成熟类型。支持不同蛋白质的其它证据来自 LIM1215 cDNA 文库中的 cDNA 克隆的序列分析，与对照序列相比该文库含有缺失和插入。

至少有七种推测的内含子保留在 mRNA 中（参考图 7，它列出了 7 个内含子中的 6 个）。内含子是独立保留的，因此，一个特定的 mRNA 可以没有、有任何 1 个、两个、等等直至 7 个内含子。由 7 个独立剪接形成的不同 mRNA 的最大数目是 2^7 ，或 128 个不同的 mRNA。这些内含子的 DNA 序列示于图 10 中。最靠近 5' 端的内含子，称为序列“X”，其长度未知，并且只列出了部分序列。

对照端粒酶序列（图 1）包括内含子 α 和内含子 β 。在下面的讨论中，各个内含子的存在与否及其位置所起的作用均基于这是唯一的改变。我们将会看到，一个特定的内含子可以改变翻译产物的序列，而不管其它内含子是被剪接进还是剪接出。例如，内含子 1 的存在产生一个移码和截短的蛋白质，而不管内含子 α 、 β 、2 或 3 是被剪接进还是被剪接出。

内含子“X”的存在产生一个截短的蛋白质，它含有大约 600 个 N-端氨基酸，缺少所有的逆转录酶基元。在 222 位碱基上内含子“Y”的存在产生一个终止在内含子外三个密码子内的移码的蛋白质。因为 Y 内含子富含 GC（大约为 78%），测序很困难，所以内含子 Y 引起大约 35 个

氨基酸的插入而不是移码。

位于 1950 位核苷酸的内含子 1 为 38bp 长，它在 mRNA 上的存在引起移码，并最终翻译出一个截短的蛋白质（终止密码子在 1973 位核苷酸）。这个截短的蛋白质仅含逆转录酶结构域 1 和 2。

内含子 α 位于 2131-2166 碱基之间，常常观察到它从端粒酶 mRNA 中被剪接出去。从这样一个 RNA 翻译的蛋白质缺失了 12 个氨基酸，即失去了逆转录酶基元 A。这个基元似乎是逆转录酶功能所必须的；在酵母 EST2 蛋白质中这个结构域内的单个氨基酸的突变产生具有显性失活功能的蛋白质，并且导致细胞老化和端粒缩短。

在 2286-2468 位删除 β -外显子的另一个变体序列编码一个截短的蛋白质，这是由于在与 2269 位碱基相连的 2287 位碱基处发生了阅读移码，因此在 2605 位碱基处形成终止密码子。这个变体蛋白质具有逆转录酶的结构域 1、2、A、B 和部分 C，但缺少另一个基元；除了逆转录酶结构域基元，在 hT1 的 β 插入序列中鉴定的另一个序列基元（AVRIRGKS）与 P-环基元共有序列 AXXXXGK（S）相匹配（Saraste 等人，生化科学进展（Trends Biochem. Sci.）15: 430-434, 1990）。该基元见于大量的蛋白质家族中，包括许多激酶、细菌 dnaA、recA、recF、mutS 和结合 ATP 的解旋酶（Devereaux 等人，核酸研究(Nucl. Acids Res.），12: 387-395, 1984）。因此 P-环仅仅存在于大多数所分析的 RNA 样品中 h-TEL 的一个亚群中，而在几种肿瘤样品中完全不存在（图 8）。

2843 碱基处的内含子 2 含有一个框内终止密码，产生一个具有完整的逆转录酶结构域区但失去了 C-末端的截短的蛋白质。因为 C-末端可能起调控作用，所以该蛋白质的活性可能会受到影响。当保留内含子 3 时，也产生一个较小的蛋白质，因为此内含子含有一个框内终止密码子。所以，此蛋白质的 C-末端序列被改变了。此种蛋白质可能具有的活性现在还不知道。HIV-1 逆转录酶的晶体结构显示，缺少 RNAase 结构域的较短的蛋白质（p51）受到了 C-末端向催化裂隙的“结合”折叠的抑制。如果假设 hT1 采取与 HIV-RT 相似的结构，那么 C-末端 hT1 蛋白质变体将反映出相似的调控机制。

除了缺少对照 C-末端结构域的变体外，在 2157 碱基处具有内含子 3 的变体表达一个不同的 C-末端结构域。而且，由内含子 3 提供的编码区具有潜在的 SH3 结合位点，SGQPEMEPPRRPSGCVG，这个序列与在如共济失调毛细血管扩张突变的（ATM）蛋白质中发现的共有序列 c-Abl SH3 结合肽（PXXXXPXXP）相匹配。另一个这种基元的例子见于 hT1 蛋白质的 N-末端内，该肽序列为 HAGPPSTSRPPRPWDTP。在端粒酶 cDNA 中发现了其它不同的 C-末端结构域；EST12462（GenBank Accession No.AA299878）直至 2157 位碱基具有大约 50 个碱基的相同序列，然后开始与对照端粒酶序列以及内含子不同。这个新序列在 50 个碱基内具有一个内部终止密码子，这将产生一个截短的 C-末端。

在一个 ALT 细胞系中检测到的变体（图 6，泳道 i）说明了 hT1 的碱性结构域可能至少在某些 ALT 细胞系中对于 ALT 机制有所贡献。有趣的是，这个 ALT 细胞系表达 hTR 基因。ALT 的一个可能的机制似乎涉及在 TRAP 分析中失活的调节失常的端粒酶组分。

下表概括了剪接变体和所形成的蛋白质。为了简便起见，对于每个形成的蛋白质仅列出了一种变体。而且如上所述，Y 内含子的存在似乎引起移码，从而形成截短的蛋白质，但是不引起插入。因此，Y 内含子的每个阅读框都存在，而此表的构建似乎插入不造成截短的蛋白质。这些内含子的一个独立的组合会产生 128 种不同的 mRNA 序列。表 1 中的变体的 DNA 和氨基酸序列示于图 11。

表 1

蛋白质	插入序列					
	Y	1	α	β	2	3
截短的#1	0	+	0	0	0	0
截短的#2	0	0	+	0	0	0
对照蛋白质	0	0	+	+	0	0
截短的#3	0	0	+	+	+	0
不同 C-端	0	0	+	+	0	+
缺少基元 A	0	0	0	+	0	0
截短的#3, 缺少基元 A	0	0	0	+	+	0
缺少基元 A;不同的 C-端	0	0	0	+	0	+
截短的#1(版本 2)	+	+	0	0	0	0
截短的#2(版本 2)	+	0	+	0	0	0
对照蛋白(版本 2)	+	0	+	+	0	0
截短的#3(版本 2)	+	0	+	+	+	0
不同 C-端(版本 2)	+	0	+	+	0	+
缺少基元 A(版本 2)	+	0	0	+	0	0
截短的#3(版本 2)	+	0	0	+	+	0
缺少基元 A;不同 C-端(版本 2)	+	0	0	+	0	+

E. 载体、宿主细胞和表达与产生蛋白质的方法

端粒酶蛋白质可在多种宿主生物体中表达。在一个实施方案中，端粒酶是在细菌，如大肠杆菌中表达的，已经构建了许多大肠杆菌表达载体，这些载体都是容易获得的。其它合适的宿主生物体包括其它细菌种类，以及真核生物，如酵母（例如酿酒酵母）、哺乳动物细胞（如 CHO 和 COS-7）、和昆虫细胞（例如 Sf9）。

将编码端粒酶、端粒酶的一部分、一个变体、其融合蛋白质等等的 DNA 序列引入到适合于宿主的表达载体中。在某些实施方案中，将端粒酶插入到一种载体中以便产生一种融合蛋白质。端粒酶序列衍生自一个

现存的片段、cDNA 克隆、或是经合成的序列。合成的一个优选方法是利用一套位于该编码区或其所需部分侧翼的引物，从 cDNA 中扩增该基因。如上面所讨论，端粒酶序列可以含有具有多密码子的各氨基酸的不同的密码子。选择最适合宿主种类的不同的密码子。通常还在引物序列中插入限制位点，这些位点是根据载体上的克隆位点加以选择的。如果需要的话，可以将翻译起始密码子和终止密码子设计到引物序列中。

最低要求是，载体必须含有启动子序列。载体中可以含有其它调控序列。这些序列包括转录终止信号序列、分泌信号序列、复制起点、选择性标记等等。这些调控序列彼此有效地结合在一起以进行转录或翻译。

此处用于端粒酶表达的质粒含有为了在宿主细胞（如细菌）中表达该蛋白质而设计的启动子。合适的启动子来源广泛而且在本领域众所周知。诱导型或组成型启动子是优选的。用于在细菌中表达的这些启动子包括 T7 噬菌体和其它噬菌体的启动子，例如 T3、T5 和 SP6，以及 *trp*、*lpp* 和 *lac* 操纵子。还可以采用杂合启动子（参考美国专利 No. 4,551,433），例如 *tac* 和 *trc*。用于在真核细胞中表达的启动子包括 P10 或杆状病毒/昆虫细胞表达系统的多角体基因启动子（参考美国专利 Nos. 5,243,041, 5,266,317, 4,745,051, 及 5,169,784）、MMTV LTR、CMV IE 启动子、RSV LTR、SV40、金属硫蛋白质启动子（参考美国专利 No. 4,870,009）以及其它诱导型启动子。为了表达该蛋白质，将启动子与端粒酶蛋白质的编码区有效地连接起来。

控制端粒酶转录的启动子本身可以受到阻遏子的控制。在一些系统中，通过改变细胞的生理条件可以使启动子去阻遏，例如，通过加入一个能够竞争性结合阻遏子的分子、或者通过改变生长培养基的温度。优选的阻遏子蛋白质包括，但不限于，负责 IPTG 诱导的大肠杆菌 *lacI* 阻遏子、温度敏感型 λ cI857 阻遏子等等。大肠杆菌阻遏子是优选的。

在其它优选实施方案中，载体还含有转录终止序列，终止序列中有一段序列，它提供一种信号，通过聚合酶识别所选启动子而终止转录，以及/或者提供聚腺苷酸化的信号序列。

优选，载体能够在宿主细胞中复制。因此，当宿主细胞是一种细菌

时，载体优选地含有细菌的复制起始区。优选的细菌复制起始区包括 fl-ori 和 col E1 复制起始区，特别是衍生自 pUC 质粒的复制起始区。在酵母中，ARS 或 GEN 序列可用于保证复制的进行。哺乳动物细胞中最常用的系统是 SV40 的复制起始区。

质粒也优选地含有至少一个在宿主中起作用的选择性标记。选择性标记基因包括任何赋予宿主一种表型的基因，该表型使得转化细胞能够被鉴定出来并选择性地生长。细菌宿主的合适的选择性标记基因包括氨苄青霉素抗性基因 (Amp^r)、四环素抗性基因 (Tc^r) 和卡那霉素抗性基因 (Kan^r)。在本发明中卡那霉素抗性基因是优选的。真核生物的合适的标记基因通常需要宿主中的互补缺陷型 (例如，tk-宿主中的胸苷激酶)。但是，也可用药物标记 (例如，G418 抗性和潮霉素抗性)。

编码端粒酶的核苷酸序列还可以含有分泌信号，通过这个信号使蛋白质被合成为前体蛋白质，随后被加工并分泌出去。形成的经过加工的蛋白质可以从壁膜间隙或发酵培养基中回收。合适的分泌信号来源广泛并且在本领域是众所周知的 (von Heijne, 分子生物学杂志(J. Mol. Biol.)184: 99-105, 1985)。在大肠杆菌 (或其它宿主) 中起作用的原核的和真核的分泌信号都可以采用。在本发明中优选的分泌信号包括，但不限于，这些由下列大肠杆菌基因编码的分泌信号: pelB(Lei 等人, 细菌学杂志(J. Bacteriol.)169: 4379, 1987)、phoA、ompA、ompT、ompF、ompC、 β -内酰胺酶和碱性磷酸酶。

本领域的技术人员知晓，有大量的适于在细菌细胞中表达的载体并且这些载体都是很容易得到的。例如 pET 系列 (Novagen, Madison, MI)、tac 和 trc 系列 (Pharmacia, Uppsala, 瑞典)、pTTQ18(Amersham International plc, 英国)、pACYC177、pGEX 系列等载体都适合于端粒酶的表达。杆状病毒载体，如 pBlueBac (参考，美国专利 Nos. 5,278,050、5,244,805、5,243,041、5,242,687、5,266,317、4,745,051、以及 5,169,784; 可从 Invitrogen 公司得到, San Diego) 可用于在昆虫细胞中表达端粒酶，例如在草地夜蛾(Spodoptera frugiperda) sf9 细胞中 (参考美国专利 No. 4,745,051)。用于端粒酶表达的宿主的选择部分是由载体所决定的。商业

上可得的载体是与适当的宿主成对提供的。

大量用于在真核细胞中表达的合适的载体是可以得到的。这些载体包括 pCMVLacI、pXT1(Stratagene Cloning Systems, La Jolla, CA); pCDNA 系列、pREP 系列、pEBVHis (Invitrogen, Carlsbad, CA)。在某些实施方案中, 端粒酶基因被克隆到基因导向载体中, 例如 pMCIneo、pOG 系列载体 (Stragene)。

端粒酶蛋白质通过标准方法分离, 例如亲和层析、分子排阻层析、金属离子层析、离子交换层析、HPLC, 以及其它已知的蛋白质分离方法。(通常参考 Ausubel 等人, 同上; Sambrook 等人, 同上)。一个被分离纯化的蛋白质用考马斯兰染色后在 SDS-PAGE 上为一条带。

在一个实施方案中, 端粒酶蛋白质是以六组氨酸融合蛋白质的形式表达的, 并通过含金属离子的层析法, 如镍偶联珠层析进行分离。简单地说, 编码 His₆ 的序列与编码端粒酶的 DNA 序列连在一起。尽管 His₆ 序列可以定位在分子中的任何地方, 它优选地连接在 3' 端紧靠终止密码子之前的位置。His-hT1 融合蛋白质可以通过这多种方法的任何一种进行构建。一个方便的方法是利用含有 His₆ 密码子的下游引物扩增 TEL 基因。

F. 端粒酶的肽和蛋白质

在本发明的一个方面中, 提供了具有端粒酶序列的肽。在此处和其它地方描述的分析中, 这些肽可被用作免疫原制备抗体、用作端粒酶功能的抑制剂或增强剂。这些肽一般为 5 至 100 个氨基酸长, 更常采用 10 至 50 个氨基酸。这些肽很容易以自动化方式经化学合成 (PerkinElmer ABI 肽合成仪) 或由商业途径获得。这些肽可以进一步通过各种方法加以纯化, 包括高效液相层析。而且, 这些肽和蛋白质可以含有这 20 种天然形成的氨基酸以外的氨基酸, 或含有这些氨基酸的衍生物和修饰物。

本发明中特别感兴趣的肽具有内含子序列 (图 10)、逆转录酶基元等等。在某些实施方案中, 端粒酶蛋白质具有图 1 或 11 中的氨基酸序列, 或其至少 8 个氨基酸长的一部分 (可以为 10、15、20 个氨基酸或更长)。在其它实施方案中, 该蛋白质有一个或多个氨基酸被替换、插入、缺失。

而在其它实施方案中，该蛋白质具有由一个核酸序列决定的氨基酸序列，该核酸序列在正常严谨条件下可与图 11 中任何一个序列杂交。如上面所指出，端粒酶的变体包括等位基因变体。

II. 端粒酶分析

有多种方法可用于测定端粒酶的活性及表达。这些方法包括测量端粒酶延伸端粒 DNA 底物能力、溶核活性、引物（端粒）结合活性、dNTP 结合活性、端粒酶 RNA（hTR）结合活性的体外分析，体内功能获得（gain-of-function）分析，体内功能失去(loss-of function)分析，原位杂交，RNA 酶探针保护、Northern 分析，cDNA 扩增，抗体染色等等。

A. 催化活性分析

多种催化活性的分析方法在美国专利 Nos 5,629,154; 5,639,613; 5,645,986 等等中有描述。在一个传统的端粒酶活性分析方法中，使用具有宿主端粒序列的单链 DNA 引物（例如[TTAGGG]_n）和端粒酶（参考 Shay 等人，分子遗传学方法（Methods in Molecular Genetics）5: 263, 1994; Greider 和 Blackburn, 细胞 43: 405, 1985; Morin, 细胞 59: 521, 1989; 美国专利 No. 5,629,154）。一个优选分析方法将基于去污剂的提取法与基于扩增的分析法结合起来。这种被称为 TRAP（端粒重复扩增技术）的方法具有更高的敏感性（Kim 等人，科学 266: 2011, 1994）。简单地说，在 TRAP 方法中，端粒酶合成延伸引物，延伸引物再作为扩增的模板。端粒酶产物是用衍生自寡核苷酸的非端粒区的引物和衍生自端粒区的引物扩增的。当例如通过凝胶电泳分析扩增产物，如果有端粒酶活性时可以观察到产物梯（ladder of products）。这种方法的排列已被描述（Krupp 等人，核酸研究，25: 919, 1997; Savovsky 等人，核酸研究 24: 1175, 1996）。而且，还有其它的端粒酶分析方法可以采用（Faraoni 等人，化学疗法杂志(J. Chemother)8: 394, 1996, 描述了一种体外化学敏感性分析方法; Tatematsu 等人，癌基因（Oncogene）13: 2265, 1996, 描述一种“延伸 PCR 分析法”；Lin 和 Zakian, 细胞 81: 1127, 1995,

描述一种酵母的体外分析方法)。

另外，催化活性或其它活性可以通过体外重建系统进行测定(参考实施例)。简单地说，分析方法，例如此处描述的那些方法，是利用由重组产生的纯端粒酶蛋白质和其它必须组分，例如端粒酶 RNA 组分、如 WO/98/14593 中描述的其它蛋白质进行的。

B. 其它活性分析

溶核活性可以通过如 Collins 和 Grieder, 基因与发育 (Genes and Development) 7:1364,1993)描述的技术进行评价。溶核活性是从位于 DNA 模板 5' 边界的一段核苷酸序列的 3' 端切除一个核苷酸(从端粒重复序列 TTAGG 切除 G)。简单地说，此活性可以通过一个反应进行测定，该反应利用一个具有被封闭的 3' 核苷酸的核酸模板，即它不能作为聚合酶的引物，除非它经过溶核活性被除去。

端粒结合活性及分析在如 Harrington 等人, 生物化学杂志(J. Biol. Chem)270: 8893, 1995 中有描述。一般地，可以采用任何检测蛋白质与核酸相互作用的方法，如凝胶移位分析。DNTP 和 RNA 结合活性分析方法在如 Morin, 欧洲癌症杂志(Eur. J. Cancer)33: 750 中有描述。

C. 功能获得与丢失

体内功能获得分析可以通过将编码端粒酶的表达载体转染没有或几乎不可检测的内源活性的细胞而进行。然后通过如此处描述的那些体外分析方法测定活性。另一种功能获得分析方法可以在肿瘤细胞或其它表达端粒酶或逆转录酶的细胞中进行。端粒酶基因被转染到细胞中，并以高水平表达，将这些细胞用逆转录酶的抑制剂处理。然后观察到端粒酶活性对这些抑制剂的敏感性降低。而且，在酵母端粒酶突变体 EST2 中可以测定功能恢复情况。

功能丢失可以在表达高水平的端粒酶活性的细胞中测定，例如 LIM 1215 细胞或其它肿瘤细胞。在这种分析中，将通常构建在表达载体中的反义寡核苷酸分子导入细胞中。端粒酶基因通过消除端粒酶活性得到证

实。在另一种分析中，抑制端粒酶功能的抗体被用于展示功能性分子。

D. 端粒酶的表达

端粒酶在各种细胞中的表达可以通过标准方法利用此处提供的序列进行分析。例如，用放射性或荧光标记探针（片段或寡核苷酸）的原位杂交可用于组织切片或固定化细胞。或者，可以将 RNA 从细胞中分离出来并用于 Northern 分析、Rnase 探针保护分析等等。特定区域的探针和变体特异的探针将形成各种端粒酶转录物的表达分布曲线。

在一个优选实施方案中，通过扩增分析端粒酶的表达。用于扩增端粒酶的引物对（包括扩增特殊变体的引物对），被用于扩增从细胞 RNA 合成的 cDNA。cDNA 可以由总 RNA 或 poly(A)+RNA 合成。用于 RNA 分离的方法和技术是众所周知的。cDNA 的合成可以由寡聚（dT）引物、随机引物（如 dN₆）、端粒酶特异引物等等起始。引物的选择至少部分取决于 RNA 的量和分析的目的。扩增引物是设计用于扩增脊椎动物细胞中存在的任何一种变体、几种特定变体的组合或所有变体的。选择与引物的长度、碱基含量、扩增产物的长度等等相称的扩增条件。有各种扩增系统可以采用（参考 Lee 等人，《核酸扩增技术》，生物技术丛书，Eaton 出版，Natick, MA, 1997; Larrick, 《PCR 技术：定量 PCR》，生物技术丛书，Eaton 出版，Natick, MA, 1997）。

其它用于定性和定量测定基因表达的分析方法都是众所周知的。当端粒酶 mRNA 的量很充足时，Rnase 探针保护和 Northern 分析法是可行的。当只有极少的细胞时，需要进行单细胞分析，或者当样品中端粒酶 RNA 所占的比例极低时，采用扩增技术则是优选的。Rnase 探针保护分析特别适用于检测剪接变体、突变体、以及定量这些 RNA。

如上所讨论的，在优选实施方案中，各种 RNA 种类的表达是受到监测的。不同种类的分析可以采用任何能够将某一类与其它类区分开来的方法。于是，通过 Northern 确定长度、Rnase 探针保护、克隆和扩增是一些可行的方法。在优选实施方案中，采用 Rnase 探针保护和扩增的方法。对于 Rnase 探针保护，该探针一般是衍生自对照序列和内含子

连接处的一个片段，或者是衍生自内含子插入序列位点周围的序列的一个片段。例如，对照端粒酶上的一个跨越 1950-1951 位核苷酸的片段（如 1910-1980 位核苷酸）将保护对照序列上 71 个碱基的片段，它将保护具有内含子 1 的端粒酶上的 41 个碱基和 30 个碱基长的两个片段。相反，含有 1910-1950 位核苷酸和内含子 1 的 30 个碱基的一个片段将保护内含子 1 变体的 71 个碱基的片段和对照端粒酶的 41 个碱基的片段。用于 Rnase 探针保护的片段通常是选择 30 至 400 个碱基范围的长度，并且将片段进行定位以产生容易区分的保护产物。

另一种可用于区分变体的方法是扩增。扩增引物的设计和扩增的策略在上面有所描述。简单地说，优选那些能够将每个被剪接进或剪接出的变体单独扩增的引物。可以进行多重反应以鉴别那些发生过不止一次剪接进或剪接出事件的变体。

测定端粒酶蛋白质的方法在本发明中同样有用。例如，端粒酶抗体可用于使组织切片或透化细胞染色。抗体还可用于通过免疫沉淀、Western blot 等方法检测蛋白质。而且，使用此处描述的抗体可以确定端粒酶或端粒酶变体的亚细胞定位情况。

E. 端粒酶抗体

此处所讨论的端粒酶蛋白质、片段或肽的抗体的制备很容易。这种抗体可以特异地识别野生型端粒酶蛋白质而不能识别突变体（或变体）蛋白质，或者能识别突变体（或变体）端粒酶蛋白质而不能识别野生型蛋白质，或者同时识别突变体（或变体）及野生型蛋白质。抗体可用于分离蛋白质、抑制（拮抗物）蛋白质的活性，或增强（激动剂）蛋白质的活性。同时，抗体的发展对与端粒酶作用的小分子的分析将有所帮助。

在本发明中，抗体被理解为包括单克隆抗体、多克隆抗体、抗特异抗体、抗体片段（如 Fab，和 $F(ab')_2$ ）， F_V 可变区，或互补决定区）。如果抗体与端粒酶蛋白质的结合 K_d 值大于或等于 $10^{-7}M$ ，优选地大于或等于 $10^{-8}M$ ，这种抗体一般被认为是特异性的。单克隆抗体或其结合对方的亲和性通过本领域普通技术人员可以很容易地测定（参考 Scatchard，

纽约科学院年报(Ann. N.Y. Acad. Sci.)51: 660-672, 1949)。

简单地说，多克隆抗体的制品可以容易地从大量温血动物如兔子、小鼠或大鼠中产生。将动物用端粒酶蛋白质或其肽进行免疫，优选地将它们连接到一个载体蛋白质如匙孔血蓝蛋白质上。给药途径包括腹膜内、肌内、眼内、或皮下注射，通常是在佐剂中（如弗氏完全或不完全佐剂）进行。特别优选的多克隆抗血清在分析中表现出至少比背景高三倍的结合活性。

单克隆抗体还可以容易地利用传统技术从杂交瘤细胞系中产生（参考美国专利 Nos. RE32,011, 4,902,614, 4,543,439 和 4,411,993；还参考《抗体实验手册》(Antibodies: A laboratory Manual), Harlow 和 Lane 主编，冷泉港实验室出版，1988)。简单地说，在一个实施方案中，受试动物如大鼠或小鼠被用端粒酶或其一部分注射。该蛋白质以在一种佐剂如弗氏完全或不完全佐剂中的乳剂形式给药，以增加免疫反应。在最初免疫后的一至三周，动物一般被加强免疫，并用熟知的方法测试其对蛋白质的反应性。收获脾和/或淋巴结并使其无限增殖化。可以采用各种无限增殖化技术，如由 Epstein-Barr 病毒介导或融合以产生杂交瘤的技术。在一个优选实施方案中，无限增殖化的发生是通过与合适的骨髓瘤细胞系融合形成分泌单克隆抗体的杂交瘤。合适的骨髓瘤细胞系包括，例如，NS-1 (ATCC NO. TIB 18)，和 P3X63-Ag8.653 (ATCC NO. CRL 1580)。优选的融合伴侣不表达内源抗体基因。融合后，将细胞在培养基中培养，该培养基中含有使融合的脾和骨髓瘤细胞选择性生长的试剂，如 HAT (次黄嘌呤、氨基蝶呤、和胸苷)。大约七天后，筛选存在能够与端粒酶蛋白质作用的抗体的杂交瘤。大量分析方法可以采用，包括如对流免疫电泳、放射免疫分析、放射免疫沉淀、酶联免疫吸附分析 (ELISA)、点杂交分析、Western 杂交、免疫沉淀、抑制或竞争分析，以及夹心法技术（参考美国专利 Nos. 4,376,110 和 4,486,530；还参考《抗体实验手册》，Harlow 和 Lane 主编，冷泉港实验室出版，1988）。

也可以利用其他技术来构建单克隆抗体（参考 Huse 等人，科学 246: 1275-1281, 1989; Sastry 等人，美国国家科学院院报(Proc. Natl. Acad. Sci.

USA)86: 5728-5732, 1989; Altling-Mees 等人, 分子生物学策略(Strategies in Molecular Biology)3:1-9,1990; 描述重组技术)。简单地说, 从 B 细胞群体中分离 mRNA, 并用于在合适的载体如 λ ImmunoZap(H)和 λ ImmunoZap(L)中构建重链和轻链免疫球蛋白 cDNA 表达文库。这些载体可以被单独筛选或共表达以形成 Fab 片段或抗体(参考 Huse 等人, 同上; Sastry 等人, 同上)。阳性噬菌斑可以随后被转变为能够高水平表达大肠杆菌单克隆抗体片段的非裂解质粒。

相似地, 抗体的部分或片段, 例如 Fab 和 Fv 片段, 可以利用传统的酶解或重组 DNA 技术进行构建, 以形成分离的抗体可变区。在一个实施方案中, 来自产生目的单克隆抗体的杂交瘤编码可变区的基因是利用可变区的核苷酸引物扩增的。这些引物可以通过本领域普通技术人员进行合成, 或者从商业渠道购买(如 Stratacyte 公司, La Jolla, CA)。扩增产物插入到载体如 ImmunoZAPTMH 或 ImmunoZAPTML (Stratacyte) 中, 然后再导入大肠杆菌、酵母、或基于哺乳动物的系统中进行表达。利用这些技术, 可以产生大量的含有 V_H 和 V_L 结构域融合的单链蛋白质(参考 Bird 等人, 科学 242: 423-426, 1988)。另外, 可利用技术将“鼠源”抗体变为“人源”抗体, 而不改变该抗体的结合特异性。

一旦获得了合适的抗体, 可以通过许多本领域普通技术人员熟知的那些方法进行分离或纯化(参考《抗体实验手册》, Harlow 和 Lane 主编, 冷泉港实验室出版, 1988)。合适的技术包括肽或蛋白质亲和柱层析、HPLC 或 RP-HPLC、在蛋白质 A 或蛋白质 G 柱上纯化, 或这些技术的任何组合。

F. 与端粒酶作用的蛋白质

可以通过一种方法如酵母双杂交结合系统检测直接与端粒酶作用的蛋白质。简单地说, 在双杂交系统中, 构建 DNA-结合结构域-端粒酶蛋白质融合体(如 GAL4-端粒酶融合体)并转染到含有与选择性标记基因连接的 GAL4 结合位点的细胞中。可以使用完整的端粒酶蛋白质或端粒酶的亚区。还构建了融合到 GAL4 激活结构域的 cDNA 文库并共转染。

当 cDNA-GAL4 激活结构域融合体中的 cDNA 编码与端粒酶作用的蛋白质时，选择性标记就会表达。培养含有 cDNA 的细胞，分离并鉴定该构件。还可以采用其它分析方法鉴定与端粒酶作用的蛋白质。这些方法包括 ELISA、Western 杂交、共免疫沉淀等等。

III. 端粒酶活性的抑制剂和增强剂

候选的抑制剂和增强剂（统称为“效应物”）可从多种渠道分离到或获得，例如细菌、真菌、植物、寄生虫、化合物文库（如组合化学文库）、随机肽等等。效应物还可以是端粒酶的肽或变体肽、端粒酶变体、反义核酸、端粒酶抗体、端粒酶启动子活性的抑制剂等等。抑制剂和增强剂还可以根据从 X-射线晶体学确定的蛋白质结构进行合理设计（参考，Livnah 等人，科学 273: 464, 1996）。在某些优选实施方案中，抑制剂作用于特异的端粒酶，例如一种变体。

抑制剂的作用是通过阻止端粒酶与核蛋白质复合体的其他组分或端粒的结合，通过引起结合蛋白质的解离，或者通过其它机制。抑制剂的作用可以是直接的或者是间接的。在优选实施方案中，抑制剂干扰端粒酶蛋白质与端粒酶 RNA 或端粒的结合。在其它优选实施方案中，抑制剂是小分子。在一个最优选实施方案中，抑制剂使得细胞停止复制。抑制剂应当具有最低的副作用并且优选地是无毒的。能够透过细胞的抑制剂是优选的。

在其它优选实施方案中，效应物是以显性失活方式起作用的端粒酶蛋白质或肽（参考，Ball 等人，现代生物学(Current Biology)7: 71, 1997; 现代生物学 6: 84, 1996）。例如，能够竞争性抑制端粒酶与端粒结合的端粒酶的肽将破坏端粒的延长。一般地，这些肽具有天然的序列，但是变体可以具有更强的活性（参考，Ball 等人，同上）。变体可通过此处描述的方法进行构建。其它的肽可能结合端粒酶并抑制其一种或多种活性，但是不具有端粒酶的氨基酸序列。这种肽可以通过此处描述的方法加以鉴别。这些蛋白质或肽还可以增强端粒酶的活性。为了有效的抑制，肽类抑制剂优选地是由被转染或感染到宿主细胞中的载体表达的，但是载

体也可以是通过其它方法导入的，例如脂质体介导的融合等等。真核生物载体为大家所熟知并且很容易得到。载体包括质粒、基于病毒的载体等等。

在另一个优选实施方案中，抑制剂是核酶。“核酶”是指一种能够切割端粒酶核酸序列的核酸分子。核酶由 DNA、RNA、核酸类似物、或这些分子的任何组合形式（例如 DNA/RNA 杂合体）构成。“核酶基因”指一种核酸分子，当它被转录为 RNA 时能够产生核酶，而“核酶载体”指一种能够转录目的核酶基因的部件，由 DNA 或 RNA 构成。在本发明的某些实施方案中，载体可以含有一个或多个限制性位点和选择性标记。此外，根据所选择的载体和宿主细胞，此处描述的载体中可以含有另外的元件如复制起点、聚腺苷酸化位点、和增强子。

如上所述，本发明还提供能够抑制端粒酶基因表达的核酶。简单地说，可以产生大量的核酶用于本发明中，例如包括发夹核酶（参考如，Hampel 等人，核酸研究 18: 299-304, 1990, EPO360, 257, 和美国专利 No.5,254,678），锤头核酶（参考如，Rossi, J.J.等人，药物治疗(Pharmac. Ther.) 50: 245-254, 1991; Forster 和 Symons, 细胞 48: 211-220, 1987; Haseloff 和 Gerlach, 自然(Nature)328: 596-600, 1988; Walbot 和 Bruening, 自然 334: 196, 1988; Haseloff 和 Gerlach, 自然 334: 585, 1988; Haseloff 等人, U.S. Patent No.5,254,678），丁型肝炎病毒核酶（参考如，Perrotta 和 Been, 生物化学 (Biochem.) 31: 16, 1992），I 组内含子核酶例如那些基于四膜虫核糖体 RNA 的核酶（参考如，Cech 等人，美国专利 No. 4,987,071），RNase P 核酶（参考如，Takada 等人，细胞 35: 849, 1983）；以及大量的能够切割目的或所选靶序列的其它核酸结构（参考如，WO95/29241, 和 WO95/31551）。在本发明的某些实施方案中，核酶的天然结构可以被改变以包含能够增加稳定性的四环或其它结构（参考如，Anderson 等人，核酸研究 22: 1096-1100, 1994; Cheong 等人，自然 346: 680-682, 1990），或使核酶抵抗 RNase 或内切酶活性的结构（参考如，Rossi 等人，药物治疗 50: 245-254, 1991）。

在本发明的一个实施方案中，提供了能够切割端粒酶核酸序列的发

夹核酶和锤头核酶。简单地说，产生发夹核酶以使它们能够识别靶序列 $N_3XN^*GUC(N_{>6})$ ，其中 N 是 G、U、C、或 A，X 是 G、C、或 U，而 * 是切割位点。相似地，产生锤头核酶以使它们能够识别序列 NUX，其中 N 是 G、U、C、或 A。锤头或发夹核酶的其他核苷酸是根据靶序列周围的核苷酸和锤头共有序列确定的（参考 Ruffner 等人，生物化学 (Biochemistry.)29: 10695-10702, 1990）。特定核酶的制备和应用在 Cech 等人（美国专利 No. 4,987,071）中被描述。核酶优选地是由导入宿主细胞中的载体表达的。

本发明的核酶，以及编码这些核酶的 DNA 可以利用已发表的技术很容易地产生（例如，Promega, Madison Wis., Heidenreich 等人，J. FASEB 70: 90-6, 1993; Sproat, 生物技术现代观点 (Curr. Opin. Biotechnol.)4: 20-28, 1993）。或者，核酶可以由编码核酶并与 RNA 聚合酶启动子（例如，SP6 或 T7）有效连接的 DNA 或 cDNA 分子产生。RNA 核酶是由 DNA 或 cDNA 分子转录产生的。

在其它优选实施方案中，抑制剂消除了端粒酶的启动子活性。真核启动子含有被 RNA 聚合酶结合的序列和其它参与转录单位调控的蛋白质。端粒酶的转录似乎是被高度调控的；该蛋白质主要在干细胞、胚胎细胞和癌细胞中表达，而在大多数体细胞中以低水平表达或根本就不表达。于是，启动子是抑制剂潜在的靶点。抑制剂破坏或阻止一个或多个控制端粒酶转录的因子的结合，引起转录减少或停止。转录水平仅仅需要低到使得至少一种端粒酶消失的足够低的水平。

本发明的另一个抑制剂是端粒酶编码或非编码序列的反义 RNA 或 DNA。针对特定 mRNA 分子的反义核酸已经表现出抑制所编码蛋白质表达的活性。根据此处的端粒酶序列，设计反义序列并优选地插入适于转染宿主细胞的载体中，并表达反义序列。反义序列可以结合 hT1 RNA 的任何部分。在某些实施方案中，设计了与一个或多个变体特异结合的反义序列。特异结合的意思是在生理条件下，反义序列与具有互补序列的 RNA 结合，而不与其它 RNA 结合。因为含有任何特定内含子序列的端粒酶 RNA 可以是一组由于剪接变体独立排列组合造成的异源变体，所以

不只一种 RNA 被结合并被钝化。此处的反义多聚核苷酸至少 7 个核苷酸长，一般不长于 100 至 200 个碱基，更典型的至少 10 至 50 个碱基长。关于反义分子的设计和导入细胞的方法方面的考虑可在美国专利 Nos. 5,681,747； 5,734,033； 5,767,102； 5,756,476； 5,749,847； 5,747,470； 5,744,362； 5,716,846）中找到。

另外，在某些情况下需要端粒酶活性或表达的增强剂。有时，提高细胞的繁殖潜力将具有治疗作用。例如，在受伤或生病之后器官的再生或分化，受伤后神经细胞或脑细胞的生长，在骨髓移植中用到的造血干细胞或其它器官干细胞的繁殖等十分有限，并因此得益于端粒酶的增强剂。增强剂稳定内源蛋白质、增加转录或翻译、或通过其它机制起作用。正如本领域的技术人员所清楚的，上面提供的许多指导同样适用于增强剂的设计。

抑制剂和增强剂的筛选分析方法将随抑制剂的类型和受抑制的活性的性质不同而改变。分析方法包括 TRAP 分析或方差分析、不基于扩增的聚合酶分析、酵母双杂交法、用脊椎动物端粒酶转染的酵母中阻遏的释放等等。为了筛选与端粒酶启动子相互作用的化合物，通过报告基因进行的分析是很方便的。

IV. 端粒酶的应用

端粒酶的核苷酸序列和端粒酶蛋白质被用于本发明的许多地方。在优选实施方案中，本发明的组合物被用作诊断剂或治疗剂。

A. 诊断剂

编码端粒酶和/或蛋白质的 mRNA 的表达可用于正在分裂的细胞的检测，特别是肿瘤细胞和干细胞。检测方法包括抗体染色或用于检测蛋白质的标记端粒酶结合化合物，mRNA 的核酸原位杂交，在 DNA “芯片”上的杂交，Northern 分析，Rnase 探针保护，通过 PCR 或其它方法的扩增，连接酶介导的扩增等等。而且，RNA 剪接变体的表达可以方便地通过扩增、RNase 探针保护、其它公开的方法等分析。特别地，围绕频繁

剪接变体位点的寡核苷酸引物例如此处描述的引物（例如，Htel 内含子 T 和 HT2482R）可用于检测不同细胞类型中的剪接变体。如实施方案中所示，不同的肿瘤细胞类型表现出不同的 RNA 剪接变化。可以确定剪接变体的类型与肿瘤所处的时期、转移潜力等的关系。这样，对特定变体的分析可用作一种诊断剂。具有增强的端粒酶活性的细胞，例如肿瘤细胞或过量增生性细胞可通过此处描述的任何一种方法进行定性或定量分析来鉴别。典型地，比较目的细胞和其来自相同个体或不同个体的正常对等细胞之间的端粒酶活性或表达。肿瘤或过量增殖造成活性增强这一指征是通过直接比较或通过检测其它细胞中的活性建立的，在这些细胞中已知缺失了端粒酶活性或表达。另外，可利用此处描述的分析方法监视肿瘤的发展或对治疗的反应，并比较一段时间内的活性或表达。

在一个 ALT 细胞系中检测到的表达端粒酶的变体提示，hT1 这一基本结构域至少在一些 ALT 细胞系中促进了 ALT 机制。ALT 的一个可能的机制涉及在 TRAP 分析中被钝化的调节异常的端粒酶组分。于是，变体的鉴别对随后的肿瘤发生有所帮助。

不同的 mRNA 剪接是在高等真核生物中调节基因表达的一种常见机制，并且有许多组织特异的、发育特异的和性别特异的剪接发生改变的例子。重要的是，15%与哺乳动物疾病状态有关的突变影响剪接类型（Horowitz 和 Krainer, 遗传学进展(Trends Genet.)10: 100-106, 1994）。细胞生理状态的改变也能引起剪接类型改变。实际上，肿瘤发生本身已经提示可以通过适应不同的剪接机制而提高 mRNA 剪接变体的表达。尽管其它新的、细微不同的剪接 hT1 变体在肿瘤发展中可能起作用，但是在与正常细胞相比的各种肿瘤细胞中，以及在与生命周期有限的转捩前细胞相比的转捩后细胞系中发现的主要转录物不同的相对表达水平，似乎是在肿瘤的发生和发展中起主要作用。另外，在睾丸和结肠隐窝以及肿瘤细胞系中发现的 hT1 的不同剪接变体的存在，显示出该基因在正常发育中受到复杂的调节。

在大多数肿瘤和所有端粒酶呈阳性的无限增殖化细胞系中都观察到了主要 hT1 产物的表达。所以 hT1 的转录调控是端粒酶活性及其它功能

调节的一个主要方面。例如，除了维持种系中端粒的长度以外，端粒酶还在愈合染色体断裂中起作用。端粒酶的组成可以根据这些功能角色的不同而改变。

因此，内含子序列在诊断实践中特别有用。例如，疾病，例如癌、衰老、伤口的愈合、神经元再生、再生性细胞（如干细胞）的检测与鉴别是确定有效疗法的重要前奏。鉴于这一考虑，对伤口愈合的检测有助于开发并鉴别改良的化合物。当前，伤口愈合分析昂贵并耗时，而扩增或基于杂交的分析快速而费用低。在这些实践中，检测可以是定量的或是定性的。在定性分析中，一种变体序列（例如以不同方式剪接的内含子）的特定扩增引物对或杂交探针可被用于检测变体序列的存在与否。

在本发明中有用的探针包括可与图 10 中列出的序列或其互补序列杂交的核酸分子。杂交探针一般至少 24 个碱基，但其变化范围可以从 12 个碱基至全长序列。探针可以包含不与 hT1 DNA 或 RNA 杂交的其它序列。探针一般是 DNA，但也可以是 RNA、PNA、或其衍生物。选择对于探针长度和杂交方法合适的杂交条件（例如，在尼龙支持物上，在硅片上）。杂交条件在本领域是众所周知的。图 10 中的一个序列是基因组序列，在端粒酶 mRNA 中找不到。由这个序列衍生的探针可被用于在 RNA 样品和扩增反应中检测基因组 DNA。杂交探针可用放射性标记物、化学发光标记物、或任何大量其它已知的标记物加以标记。

杂交可以在 mRNA 样品、cDNA 样品上进行，可固定到固体支持物上、在溶液中、或在组织中原位进行等等。一种杂交分析是与固定在固体底物如经功能化处理的玻璃片或硅片上的寡核苷酸进行退火。这种芯片可从商业途径获得，或根据下面文献中提供的方法和步骤进行制备：如 PCT/US94/12282；美国专利 No. 5,405,783；美国专利 No.5,412,087；美国专利 No. 5,424,186；美国专利 No. 5,436,327；美国专利 No. 5,429,807；美国专利 No. 5,510,270；WO 95/35505；美国专利 No.5,474,796。寡核苷酸一般是以阵列的方式排列的，这样就能够确定每个寡核苷酸序列的位置。

对于扩增分析，需要用于扩增的引物对或者是位于内含子附近，或者是需要内含子存在。此处公开了许多这样的引物对。其它的引物对可

可以根据此处给出的序列进行设计。一般地，引物对的设计是使其仅仅能够扩增单个内含子，但是，在一些情况下在同一个 RNA 样品中检测多个内含子则是优选的。

其它诊断方法，例如原位杂交、RNase 保护等方法可以单独使用或与上面讨论的方法联合使用。本发明提供了指导这些分析方法的原则，尽管这些技术是众所周知的。

可以构建转基因小鼠和无效突变体小鼠（如“敲除小鼠”）以促进候选抑制剂的测试。端粒酶基因优选地处于一个转基因小鼠载体构件的组织特异性启动子的控制下。过量表达端粒酶的小鼠可被用作测试抑制剂的模式系统。在这些小鼠中，过量表达端粒酶的细胞可望连续地繁殖。根据对细胞生长的观察和测量决定候选抑制剂的给药方式。能够减缓细胞生长或使细胞生长停止的抑制剂是候选的治疗剂。

端粒酶还可被转染到细胞中使各种细胞类型无限增殖化。暂时的无限增殖化可通过含有端粒酶的表达载体的不稳定转染获得。相反，通过添加或不加诱导物，可以使处于可诱导启动子控制下的端粒酶基因稳定转化子进行或不进行繁殖。相似地，端粒酶活性的抑制剂的存在与否可用于选择性地无限增殖化细胞。酵母中全部蛋白质的一部分的表达可以作为一种显性失活，因为许多的人类蛋白质与酵母中蛋白质复合体的一些组分相互作用，但是这种作用非常不完美因此是没有价值的。所以，这些基因作为显性失活。于是，酵母将最终老化。这种细胞可被用于筛选抑制剂药物，这种药物将使得酵母的生长度过老化期。

纯化的端粒酶蛋白质、对照变体蛋白质、或片段可被用于筛选抑制剂药物的分析中。这些分析将典型地在体外进行，并利用任何一种上述的方法或本领域已知的方法。该蛋白质还可以制成晶体并进行 X-射线分析以确定其三维结构。

B. 治疗剂

此处公开的组合物和方法还可在疾病和紊乱的治疗中用作治疗剂以影响细胞中的任何一种端粒酶活性。治疗是指任何对疾病或紊乱的改善，

例如减轻疾病或紊乱的症状、减小肿瘤细胞的体积等等。例如，酶活性的抑制剂可用于限制细胞的繁殖。

许多疾病和紊乱与细胞的繁殖和繁殖潜力密切相关。与不希望的繁殖相关的一种最明显的疾病是癌症。此处描述的方法和组合物可用于治疗癌症，例如黑素瘤、其它皮肤癌、成神经细胞瘤、乳腺癌、结肠癌、白血病、淋巴瘤、骨肉瘤、等等。在本发明中适于治疗的其它疾病和紊乱包括细胞过度繁殖（与相同或不同个体的正常对等细胞相比其繁殖速率增加）如平滑肌细胞增生、皮肤生长等等。而其它的疾病和紊乱将受益于端粒酶活性的提高。端粒酶的增强剂可用于刺激干细胞繁殖和可能的分化。所以，造血干细胞的扩增可用于骨髓移植中。同时许多组织具有干细胞。这些细胞的繁殖有助于伤口愈合、毛发生长、疾病如 Wilm's 肿瘤的治疗等等。

某些抑制剂或增强剂可以通过表达载体的方式给药。许多将核酸导入细胞的技术是已知的。这种方法包括逆转录病毒载体和随后的逆转录病毒感染、腺病毒或腺结合病毒载体和随后的感染、核酸与凝聚剂（如多聚赖氨酸）的复合体，这些复合体或病毒载体通过结合配体的方式导入特定的细胞类型。许多肿瘤细胞和其它细胞特异的配体在本领域是众所周知的。

如上所述，在本发明的某些方面，编码核酶、反义序列、显性失活端粒酶、端粒酶的部分等等的核酸可以通过向目的细胞引入一种功能性的基因，而被用于抑制端粒酶的活性。这可以通过向细胞传递一种合成的基因或者通过传递能够在体内转录基因产物的 DNA 或 cDNA 来实现。更特别的是，为了在体内产生该产物，编码该产物的核酸序列被置于一个真核启动子的控制之下（如 pol III 启动子、CMV 或 SV40 启动子）。当需要更特异地控制转录时，该基因可被置于一个组织或细胞特异的启动子（如将细胞导入肝中），或一个可诱导启动子的控制之下。

大量的载体可用于本发明中，包括如质粒、病毒、逆转录转座子和粘粒。代表性例子包括腺病毒载体（如 WO94/26914, WO93/9191; Yei 等人，基因疗法(Gene Therapy)1: 192: 200, 1994; Koll 等人，美国国

家科学院院报 91 (1): 215-219, 1994; Kass-Eisler 等人, 美国国家科学院院报 90 (24): 11498-502, 1993; Guzman 等人, 循环 88 (6): 2838-48, 1993; Guzman 等人, 循环研究 (Cir. Res.) 73 (6): 1202-1207, 1993; Zabner 等人, 细胞 75 (2): 207-216, 1993; Li 等人, 人基因疗法 (Hum Gene Ther.) 4 (4): 403-409, 1993; Cailaud 等人, 欧洲神经科学杂志 (Eur. J. Neurosci.) 5 (10): 1287-1291, 1993), 腺伴随 1 型病毒 (“AAV-1”) 或腺伴随 1 型病毒 (“AAV-2”) 载体 (参考 WO95/13365; Flotte 等人, 美国国家科学院院报 90 (22): 10613-10617, 1993), 活体、或弱化的丁型肝炎病毒载体和疱疹病毒载体 (如美国专利 No. 5,228,641), 以及在美国专利 No. 5, 166, 320 中公开的载体。其它代表性的载体包括逆转录病毒载体 (如 EP 0 415 731; WO90/07936; WO91/02805; WO94/03622; WO 93/25698; WO 93/25234; 美国专利 No. 5, 219, 740; WO93/11230; WO93/11230; WO93/10218。方法和其它组合物参考美国专利 No. 5, 756, 264; 5, 741, 486; 5, 733, 761; 5, 707, 618; 5, 702, 384; 5, 656, 465; 5, 547, 932; 5, 529, 774; 5, 672, 510; 5, 399, 346, 和 5, 712, 378)。

在本发明的某些方面中, 利用载体或通过各种物理方法可将核酸分子导入宿主细胞中。这些方法的代表性例子包括用磷酸钙沉淀进行转化 (Dubensky 等人, 美国国家科学院院报 81: 7529-7533, 1984), 直接显微注射核酸分子到完整的靶细胞中 (Acsadi 等人, 自然 352: 815-818, 1991), 和电穿孔法。在电穿孔法中, 悬浮在导电溶液中的细胞被置于强电场中使细胞膜被瞬时极化, 从而使核酸分子进入细胞。其它方法包括利用连接在失活腺病毒上的核酸分子 (Cotton 等人, 美国国家科学院院报 89: 6094, 1990), 脂质转染法 (Felgner 等人, 美国国家科学院院报 84: 7412-7417, 1989), 微粒轰击法 (Williamms 等人, 美国国家科学院院报 88: 2726-2730, 1991), 聚阳离子化合物如聚赖氨酸、受体特异性配体、包裹核酸分子的脂质体, 含有核酸分子的大肠杆菌被除去外层细胞壁并通过聚乙二醇融合到动物细胞中形成的球形融合体, 病毒转导 (Cline 等人, 药物治疗 29: 69, 1985; 和 Friedmann 等人, 科学 244:

1275, 1989), 和 DNA 配体 (Wu 等人, 生物化学杂志 264: 16985-16987, 1989), 以及被补骨脂素失活的病毒如仙台病毒或腺病毒。在一个实施方案中, 利用脂质体法将核酸分子导入宿主细胞中。

效应物的给药方法一般按照现有方案进行。本发明的化合物可单独给药, 或是作为药物组合物给药。简单地说, 本发明的药物组合物可含有此处描述的一种或多种抑制剂或增强剂, 并与一种或多种药学或生理学上可接受的载体、稀释剂或赋形剂结合使用。这些组合物可以含有缓冲液如中性缓冲盐水、磷酸缓冲盐水等等, 碳水化合物如葡萄糖、甘露糖、蔗糖或葡聚糖、甘露醇, 蛋白质、多肽或氨基酸如甘氨酸, 抗氧化剂、螯合剂如 EDTA 或谷胱甘肽, 佐剂 (例如氢氧化铝) 和防腐剂。另外, 本发明的药物组合物还可以含有一种或多种另外的活性成分。效应物可以进一步结合一个导向基因, 它可与繁殖细胞特异的细胞表面受体结合。

本发明的组合物可以根据指定的给药方式进行配制, 包括如经口、经鼻、经静脉、颅内、腹膜内、皮下、或肌内给药。在本发明的其它实施方案中, 此处描述的组合物可以作为缓释植入物的一部分。而在其它实施方案中, 本发明的组合物可制成冻干粉, 利用合适的赋形剂提高冻干粉和随后的再水化的稳定性。

如上所述, 本发明还提供了药物组合物。这些组合物含有任何一种上述核酶、DNA 分子、蛋白质、化学药品、载体、或宿主细胞, 以及药学或生理学上可接受的载体、赋形剂或稀释剂。一般地, 这些载体在所用的剂量和浓度内对受体应是无毒的。通常, 这些组合物的制备包括使治疗剂与缓冲液、抗氧化剂如抗坏血酸、低分子量 (低于约 10 个残基) 多肽、蛋白质、氨基酸、碳水化合物包括葡萄糖、蔗糖或糊精、螯合剂如 EDTA、谷胱甘肽和其它稳定剂和赋形剂混合。中性缓冲盐水或与非特异性血清白蛋白质混合的盐水是合适的稀释剂的范例。

另外, 本发明的药物组合物可制成不同药剂以便通过多种不同的途径给药, 包括如关节内、颅内、皮内、肝内、肌内、眼内、腹膜内、鞘内、静脉内、皮下注射或者甚至直接注射到肿瘤中。另外, 本发明的药

物组合物可置于容器内，并有提供该药物组合物使用指导的包装材料。一般地，这些指导将包括描述药剂浓度的明确说明，以及在某些实施方案中还包括赋形剂成分或稀释剂（如水、盐水或 PBS）的相对含量，这些成分是重构药物组合物所必需的。药物组合物对于诊断或治疗都是有用的。

本发明的药物组合物可以针对所治疗（或预防）的疾病采取适当的给药方式。给药的量和频率将根据这些因素确定，例如患者的条件、以及患者病情的类型和严重程度等。在临床实验中最精确地确定给药的剂量。通过适当的技术检测患者的治疗效果，这些技术包括临床恶化的观察、成像等等。

下面的实施方案是通过说明的方式提供的，而不是意在限制。

实施例

实施例 1

人端粒酶基因的鉴定与分离

一个人端粒酶基因是在从癌细胞系构建的 cDNA 文库中鉴定的。将该 cDNA 进行 DNA 序列分析（Kilian 等人，同上）。

通过 BLAST 与 GenBank 登录号 U95964 ($p=3.2 \times 10^{-6}$) 的 Euplotes 端粒酶序列比较，GenBank 登录号 AA281296 的 EST 序列被鉴定为一个部分端粒酶基因序列。两个序列之间的氨基酸序列的同一性大约为 38% 而氨基酸序列的相似性大约 60%。

为了获得更长的 hT1 克隆，利用 EST 序列内部引物，通过扩增的方法筛选从肿瘤细胞制备的 cDNA 文库。基于 EST 序列的引物 HT1553F 和 HT1920R 用于在不同 cDNA 文库中扩增大约 350bp 的片段。扩增反应是在“热启动”条件下进行的。扩增循环为 95℃ 4 分钟；80℃ 1 分钟；94℃ 30 秒、55℃ 30 秒、72℃ 1 分钟 30 次循环；以及 72℃ 5 分钟。在 12 个所筛选的文库中，仅从 3 个文库中检测到了预期大小（~350bp）的扩增产物。在睾丸 cDNA 文库、体细胞文库、和不同癌细胞 cDNA 文库中未检测到片段。然而，从结肠癌细胞系 LIM1215 细胞中检测到了高丰度的

350bp 片段。在这个文库以及其它几个文库中，扩增了大约 170bp 的另外一个片段。

从 LIM1215 文库中获得更长克隆的方法有两个：用 ^{32}P 标记的 EST 探针筛选噬菌斑和在文库 DNA 上扩增。通过用 EST 探针进行文库杂交获得了一个具有 1.9kb 插入片段的阳性噬菌斑，命名为 53.2。此克隆的 DNA 序列分析表明它从 EST 序列的 5' 和 3' 延伸，但是并不含有一个开放阅读框（ORF）。从文库的扩增分析获得的一个片段在序列上与 53.2 片段相似，但是还含有 36bp 和 >300bp 的另外两个序列。这两个插入片段表示出了在相对于 53.2 序列边界处剪接受体和供体序列的特征，可能代表了未剪接的内含子。用 T7 引物和 HT1553F 引物扩增，产生一个大约 1.6kb 的片段；而用 T3 引物和 HT1893R 引物扩增，产生了一个大约 0.7bp 的片段。每个片段都支持用 HTEL1553F 引物和 HT1893R 引物扩增，得到 320bp 的片段。

更长的克隆还可以通过 mRNA 样品的扩增获得。在 LIM1215 mRNA 上进行逆转录 PCR（RT-PCR）鉴定出另外一个 PCR 产物，包括一个相对于 53.2 片段的具有 182bp 插入片段的产物，它形成一个开放阅读框（ORF）。cDNA 是从分离自正常组织和肿瘤组织的 RNA 合成的。在进行嵌套扩增前用 Titan RT-PCR 系统（Boehringer-Mannheim）进行 RT-PCR。扩增条件如下：95℃ 2 分钟，94℃ 30 秒两个循环，65℃ 30 秒和 68℃ 3 分钟，94℃ 30 秒，63℃ 30 秒，68℃ 3 分钟两个循环，94℃ 30 秒，60℃ 30 秒和 68℃ 3 分钟 34 个循环。RT-PCR 产物稀释 100 倍，将其中 1 微升用 Taq 聚合酶和缓冲液 Q（Qiagen）进行嵌套扩增。扩增条件如上，除了最后一步为 14 个循环。对于正常组织和肿瘤组织，将扩增产物在 1.5% 琼脂糖凝胶上电泳分离，转移至 Zetaprobe 膜上并用放射性标记的寡核苷酸 HT1691F 探测。

分别用 cRACE 和 3' RACE 的组合，在 LIM1215mRNA 上将该 DNA 序列的 5' 和 3' 延伸，得到了命名为 hT1（图 1）的 3871bp 的片段。进行两轮 cRACE 以延伸 hT1 序列并确定转录的起始位点。用 500 纳克 LIM1215 polyA+RNA 作为模板。cDNA 第一条链的合成是用 HT1576R

引物引发的。在连接产物（利用 XL-PCR 系统）上进行的第一轮扩增采用 HT1157R 和 HT1262F 引物。扩增产物用 Qiagen 柱纯化，并进一步用 HT1114R 和 HT1553F 引物扩增。产生的 1.4kb 谱带进行 DNA 序列分析，并根据该序列设计了一套新的引物。在第二轮 cRACE 中，cDNA 第一条链用 HT220R 引物引发。第一轮扩增采用 HT0142R 和 HT0141F 引物。产物按上述纯化并用 HT0093 和 HT0163F 扩增。观察到一个 100bp 的产物，并在两个独立的实验中将其进行序列分析以确定 hT1 转录物的 5' 末端。转录物的 5' 末端还可以通过用引物 HtelFulcodT 5'-AGGAGATCTCGCGATGCCGCGCGCTC-3' 和 HtelFulcodB 5'-TCCACGCGTCCTGCCCGGGTG-3' 对 LIM1215 RNA 扩增得到。产生的扩增产物用 MluI 和 BglIII 消化并连接到剩余的端粒酶 cDNA 序列上。

转录物的 3' 远端序列是通过进行两轮扩增（XL-PCR 系统）得到的，在两轮扩增中都采用 EBHT18 作为反向引物，而在第一和第二轮扩增中分别采用 HT2761F 和 HT3114F 作为正向引物。

hT1 的大小与从 LIM1215 RNA 中丰度最高的 RNA 的 Northern 杂交（见下面）中估计的大小非常吻合。在图 1 中列出了大约 3.9kb 的 DNA 序列。在 EST 中发现的序列位于 1624-2012 位核苷酸。最大开放阅读框的预测的氨基酸序列也在图 1 中列出。如图中所示，该蛋白质为 1132 个氨基酸。

表 2

名称	寡核苷酸序列
HT0028F	5' - GCTGGTGCAGCGCGGGGACC
HT 5'Met	5' - CACAAGCTTGAATTCACATCTCACCATGAAGGAGCTGGTGGCCCCGAGT
HT0093R	5' - GGCACGCACACCAGGCACTG
HT0141F	5' - CCTGCCTGAAGGAGCTGGTG
HT0142R	5' - GGACACCTGGCGGAAGGAG
HT0163F	5' - CCGAGTGCTGCAGAGGCTGT
HT0220R	5' - GAAGCCGAAGGCCAGCACGTTCTT

HT1262F 5' - GTGCAGCTGCTCCGCCAGCACA
 HT1114R 5' - GTTCCCAAGCAGCTCCAGAAACAG
 HT1157R 5' - GGCAGTGCGTCTTGAGGAGCA
 HT1553F 5' - CACTGGCTGATGAGTGTGTAC
 HT1576R 5' - GACGTACACACTCATCAGCCAG
 HT1590F 5' - GGTCTTTCTTTTATGTCACGGAG
 HT1691F 5' - CACTTGAAGAGGGTGCAGCT
 HT1875F 5' - GTCTCACCTCGAGGGTGAAG
 HT1893R 5' - TTCACCCCTCGAGGTGAGACGCT
 HT1920R 5' - TCGTAGTTGAGCACGCTGAAC
 HT2026F 5' - GCCTGAGCTGTACTTTGTCAA
 HTM2028F 5' - CTGAGCTGTACTTTGTCAAGGACA
 HT2230F 5' - GTACATGCGACAGTTCGTGGCTCA
 HT2356R 5' - CATGAAGCGTAGGAAGACGTCGAAGA
 HT2482R 5' - CGCAAACAGCTTGTCTCCATGTC

 HT2761F 5' - CTATGCCCGGACCTCCATCAGA
 HT2781R 5' - CTGATGGAGGTCCGGGCATAG
 HT3114F 5' - CCTCCGAGGCCGTGCAGT
 HT3292B 5' - CACCTCAAGCTTTCTAGATCAGTCCAGGATGGTCTTGAAGTCA
 HT3689R 5' - GGAAGGCAAAGGAGGGCAGGGCGA
 EBHT18 5' - CACGAATTCGGATCCAAGCTTTTTTTTTTTTTTTTTT
 HT-RNA-F 5' - GGGTTGCGGAGGGTGGGC
 HT-RNA451R 5' - GCAGTGGTGAGCCGAGTCCTG
 HT-RNA598F 5' - CGACTTTGGAGGTGCCTTCA

 HTel 5'T 5' - GCTGGTGCAGCGGGGACC
 HTel979T 5' - GAGGTGCAGAGCGACTACTCCA
 HTel1335T 5' - GTCTCACCTCGAGGGTGAAG
 HTel71T 5' - GGCTGCTCCTGCGTTTGGTGGA
 HTel21B (Top) 5' - GCCAGAGATGGAGCCACC

HTel21TBct) 5' - GGGTGGCTCCATCTCTGGC
 HTel-7B 5' - CCGCACGCTCATCTCCACGT
 HTel+256B 5' - GCTTGGGGATGAAGCGGTC
 HtelIntronT 5' - CGCCTGAGCTGTACTTTGTCA
 Htel 3'COdB 5' - CACCTCAAGCTTTCTAGATCAGCTAGCGGCCAGCCCAACTCCCCT
 Htel 1210B 5' - GCAGCACACATGCGTGAAACCTGT
 Htel 1274B 5' - GTGTCAGAGATGACGCGCAGGAA
 Htel 1624b 5' - ACCCACACTTGCCTGTCTGAGT
 hTR TAC 5' - ACTGGATCCTTGACAATTAATGCATCGGCTCGTATAATGTGTGGAGGGTTGCGGAGGG
 TGGGC
 hTR 5'T7 5' - CTGTAATACGACTCACTATAGGGTTGCGGAGGGTGGGC
 hTR 3'PstI 5' - CACCTGCAGACATGCGTTTCGTCTCACGGACTCATCAGGCCAGCTGGCGACGCATGTGT
 GAGCCGAGTCCTG

 BT-177 5' - GGATCCGCCGAGAGCACCGTCTG
 BT-178 5' - CGAAGCTTTCAGTGGGCCGGCATCTGAAC
 BT-179 5' - CGAAGCTTTCACAGGCCAGCCCAACTCC
 BT-182 5' - GCGGATCCAGAGCCACGTCTTACGTC
 BT-183 5' - GCGGATCCGTTTACAGATGCCGGCCAC

实施例 2

HT1 序列及与其它端粒酶的序列比较

多重序列比较显示推测的 hT1 蛋白质与 *Euplotes* 和酿酒酵母端粒酶催化亚单位在全长范围内共线性 (co-linear) (图 2)。尽管三种蛋白质之间的整体同源性相对较低 (在所有的成对组合中大约为 40% 的相似性), 但是蛋白质的整体结构似乎非常保守。四个主要的结构域: N-末端、碱性的、逆转录酶 (RT) 和 C-末端结构域存在于所有三种蛋白质中。序列相似性最高的区域是在 RT 结构域中。显然, *Euplotes* 逆转录酶结构域的所有基元的特征都存在于 hT1 序列中, 并且与逆转录酶催化有关的所

有氨基酸残基也都保留在 hT1 序列中 (Lingner 等人, 科学 276: 561-567, 1997)。

最近, 用蛋白质磷酸酶 2A 处理人乳腺癌细胞提取物, 显示能够抑制端粒酶的活性 (Li 等人, 生物化学杂志 272: 16729-16732, 1997)。尽管还不知道这种作用是不是直接的, 但是它提高了蛋白质磷酸化对端粒酶活性调节的可能性。推测的 hT1 蛋白质不含有大量潜在的磷酸化位点, 包括 11SP 或 TP 二肽, 这些是细胞周期依赖性激酶的潜在位点。

实施例 3

端粒酶基因的特征

进行 Northern 分析和 Southern 分析以确定端粒酶转录物的大小以及端粒酶基因是否是在肿瘤细胞中扩增的。

对于 Northern 分析, polyA mRNA 是从 LIM1215 细胞和 CCD 成纤维细胞中分离的。CCD 是一个原代人成纤维细胞系。在含有去污剂 (0.1%SDS) 和 200 μ g/ml 蛋白质酶 K 的缓冲液 (0.1M NaCl, 10mM Tris, pH 7.4, 1mM EDTA) 中通过匀浆简单地将细胞裂解。将 SDS 加到裂解液中至终浓度为 0.5%, 并将裂解液在 60 $^{\circ}$ C 保温 1 小时和 37 $^{\circ}$ C 下保温 20 分钟。将裂解液与 Oligo dT-纤维素再保温 1 小时, Oligo dT-纤维素事先经在 0.1M NaOH 中预循环并经过在 0.5M NaCl, 10mM Tris pH 7.4, 1mM EDTA, 和 0.1%SDS 中平衡。离心收集树脂, 在平衡缓冲液中分批洗涤, 并装柱。mRNA 用温缓冲液 (38 $^{\circ}$ C) (10mM Tris pH 7.4, 0.1mM EDTA) 洗脱下来并用乙醇沉淀。

大约 3 微克聚腺苷酸化的 RNA 在 0.85% 的甲醛-琼脂糖凝胶上电泳 (参考 Sambrook 等人, 同上), 过夜转移至 Genescreen plus 膜 (Bio-Rad, CA) 上。将膜用 32 P 标记的端粒酶特异探针 (相应于 EST 序列的 390bp 插入片段) 杂交。经高严谨度洗膜后, 在来自 LIM1215 的 mRNA 中观察到一条很明显的~3.8kb 的谱带, 而在来自 CCD 成纤维细胞的 mRNA 中没有 (图 3)。随后, 同一张膜用甘油醛 6-磷酸脱氢酶探针进行的杂交显示两种 mRNA 的谱带强度相同, 说明各泳道中含有相似数量的高质量

的 RNA。仅在 LIM1215 RNA 中还观察到更大的转录物（特别是~8kb 非均一的谱带）的存在（图 10，上道）。这些结果证明了其它 hT1-特异的 mRNA 的存在，以及 hT1 可以优选地在肿瘤细胞而非正常细胞中表达。

为了 Southern 分析，DNA 从人的外周血液单核细胞和 LIM1215 细胞中分离。大约 10 微克 DNA 用 HindIII、XbaI、EcoRI、BamHI 和 PstI 消化，在 1% 琼脂糖凝胶中电泳，并转移至尼龙膜上。作为对照，将含有人端粒酶的质粒 DNA 滴定至每 10 微克基因组 DNA 大约 10 个拷贝、5 个拷贝和 1 个拷贝，并在相同凝胶中电泳。一个 390bp 的端粒酶基因片段（含有 EST 序列）被用 ^{32}P 标记并在正常严谨度条件下杂交。在 $2 \times \text{SSC}$ ，0.1% SDS 中 55°C 下洗膜。图 4 中给出了一个扫描磷图象。如图所示，端粒酶基因似乎没有在 LIM1215 细胞中被扩增或重排，因为 LIM1215 与 PBMC DNA 相比其杂交带型与强度没有明显的不同。而且，端粒酶似乎是一个单拷贝基因，因为除了 PstI 以外，所有的酶消化都产生单一谱带。

实施例 4

HT1 的表达模式

尽管已经将端粒酶活性与肿瘤细胞和种系广泛地联系在一起，但是直到最近才认识到某些正常的哺乳动物组织表达低水平的端粒酶活性。在原代成纤维细胞 RNA 中没有检测到 hT1 的表达，并且用 EST 区的引物扩增几种从商业途径得到的肺、心脏、肝、胰、海马区、胎儿脑以及睾丸的 cDNA 文库也没有得到任何产物。

然而，在前面已表明具有端粒酶活性的正常组织（结肠，睾丸和外周血液淋巴细胞）以及许多黑素瘤和乳腺癌样品中检测了 hT1 的表达。从正常的人结肠、睾丸和循环淋巴细胞和肿瘤样品的组织切片中分离 RNA，并进行 RT-PCR 分析。cDNA 的扩增产物很容易与因污染了基因组 DNA 而形成的产物区别开，因为利用 cDNA 为模板得到~300bp 的产物，而利用基因组 DNA 为模板得到 2.7kb 的产物。在结肠、睾丸和大多数肿瘤样品中都检测出 hT1 转录物，在淋巴 RNA 中转录物非常弱（图 5，

上道)。有趣的是，两个乳腺癌样品的 hT1 表达呈阴性，尽管以 β -肌动蛋白的扩增作为阳性对照判断它含有与其它样品数量相当的 RNA (图 5, 下道)。

端粒酶活性的获得似乎是无限增殖化过程的一个重要的方面。在许多区配的成对转接前细胞培养物和转接后细胞系中 hT1 的表达利用 RT-PCR 进行测定，随后进行嵌套引物扩增 (图 6, 上道)。根据 TRAP 分析 (Bryan 等人, EMBO J.14: 4240-4248, 1995), 这些细胞系分别是端粒酶阴性 (转接前细胞系) 和阳性 (转接后细胞系) 的。在两个匹配的成对细胞系 BFT-3B 和 BET-3K 中, hT1 仅在转接后细胞系中检测到 (比较 a 泳道和 b 泳道、e 泳道和 f 泳道)。尽管 BFT-3K 的转接后细胞系 (d, f 泳道) 显示出丰富的 hT1 谱带, 但是相同大小的片段在转接前培养样品 (c, e 泳道) 中仍然很弱。另外, 三个转接后细胞系中的两个显示存在另外一个 320bp 的意料之外的片段, 并且当在高分辨率胶上分析结肠和睾丸 mRNA 时也观察到了这个产物。

对三个无限增殖化的端粒酶阴性 (ALT) 细胞系也进行 hT1 表达分析 (图 6, g, h, i 泳道)。两个细胞系显示 hT1 表达阴性, 但是在另一个细胞系 (IICF-T/B1) 中, 又扩增到了一个大约 320bp 的产物, 这与转接后的结肠和睾丸样品相似。来自 IICF-T/B1 (ALT) 细胞系的 320bp 产物的 DNA 序列分析显示, 与期望的产物相比存在一个 38bp 的插入片段。利用相同的引物但是用基因组 DNA 为模板进行的扩增排除了这是一个基因组 DNA 扩增而不是 mRNA 扩增的可能性。在这些条件下, 扩增到一个 2.7kb 的片段, 通过部分序列分析得到确证。

实施例 5

端粒酶 mRNA 其它剪接类型的鉴定

LIM1215cDNA 文库中克隆的 DNA 序列分析和上述转接前和转接后培养物的 RT-PCR 数据表明, 在 hT1 转录物中有大量不同的序列变体。为了系统地测定变体, 利用包含整个序列的引物对进行 RT-PCR。在 N-末端和碱性结构域中没有发现变体, 而在逆转录酶结构域中发现了几个

变体，以及在 C-末端结构域发现了更少的变体。最明显的是，在逆转录酶基元 A 和逆转录酶基元 B 之间有几个 RNA 变体（图 7A）。

mRNA 样品是利用传统方法从几种不同的肿瘤中制备的。这些肿瘤是：（1）SLL 肺癌，（2）淋巴瘤 C，（3）肺癌，（4）成神经管细胞瘤，（5）淋巴瘤 B，（6）淋巴瘤 E，（7）肿瘤样品 47D，（8）嗜铬细胞瘤，（9）淋巴瘤 F，（10）神经胶质瘤，和（11）淋巴瘤 G。这些样品的 mRNA 首先被逆转录为 cDNA，再用引物 HT1875F 和 HT2781R 扩增，最后用嵌套引物 HT2026F 和 HT2482R 扩增。在图 8 中发现了四个不同的扩增产物：220bp（谱带 1）、250bp（谱带 2）、400bp（谱带 3）和 430bp（谱带 4）。令人惊讶的是，所测试的肿瘤样品在扩增产物的总数和这些产物的定量分布上都有很大的差异。

这些产物中的三个从一些肿瘤组织中分离，并进行 DNA 序列分析。其中之一，一个 220bp 的片段与 LIM1215 文库的 53.2cDNA 相等。大约 250bp 的片段（谱带 2）含有 36bp 的框内插入片段，与从 LIM1215cDNA 文库的扩增产物中鉴定的插入片段相同。因为 RT-PCR 产物与 cDNA 文库的产物具有相同的序列，很明显 36bp 插入片段不是在文库构建时人工形成的。最大的产物（谱带 4）与 250bp 复制子相比含有 182bp 的插入片段（与前面从 LIM1215 RNA 扩增的较大产物相同）。400bp 谱带（谱带 3）的确切的序列还没有得到。根据它的大小，它可能含有 182bp 的插入片段，但是失去了谱带 2 和谱带 4 中存在而谱带 1 中没有的 36bp 插入片段。

为了检验该转录物存在的假设，设计了一个引物 HTM2028F，使得只有当 36bp 片段丢失时才能进行扩增。利用 HTM2028F 和 HT2026F 引物与 HT2356R 结合的扩增证明，含有 182bp 片段但是丢失了 36bp 片段的转录物存在于 LIM1215 RNA 中（图 9，a 和 b 泳道）。相同的上链引物（HTM2028F 和 HT2026F）与 HT2482R 引物结合从 LIM1215 RNA 中扩增了许多产物（图 9，c 和 d 泳道），根据 PCR 产物的直接序列分析其中大部分代表谱带 1-4。用 HTM2028F 和 HT2482R 引物扩增的一个 650bp 的片段代表另一个尚未完全表征的、逆转录酶-基元 A/逆转录酶基

元 B 区的以其它方式剪接的端粒酶变体。为了表示得更清楚，与 *Euplotes* 和酿酒酵母蛋白质匹配最好的蛋白质序列列于图 1 中作为对照序列。

特别地，至少有 7 个插入片段或内含子可以在端粒酶 RNA 中存在或从端粒酶 RNA 中消失。(1) 5' -最远端序列 (Y) 位于碱基 222 和 223 之间。(2) 插入序列 (X) 位于碱基 1766-1767 之间。测定了一个部分序列并示于图 10 中。终止密码子存在于所有三个阅读框中。因此，将产生一个没有任何逆转录酶基元的截短的蛋白质。(2) 一个序列，在图 7 中以“1”表示，位于碱基 1950-1951 之间。这个内含子为 38bp (图 10)，并且似乎存在于 ALT 和大部分肿瘤系中。该序列的存在增加了 13 个氨基酸并使阅读框移码，使得终止密码子 (TGA) 位于阅读框中 1973 位碱基处。(3) 一个序列，在图 7 中以“ α ”表示，位于碱基 2130 和 2167 之间。该插入序列为 36 碱基 (图 10)，它的丢失移去了逆转录酶基元“A”但是没有改变阅读框。(4) 一个序列，在图 7 中以“ β ”表示，位于碱基 2286 和 2469 之间。该插入序列为 182 碱基 (图 10)，并且它的丢失引起阅读框移码并且在逆转录酶基元 5 中 2604 位核苷酸出现终止密码子。

(5) 图 7 中的序列“2”存在于碱基 2823 和 2824 之间。其长度未定；其部分序列列于图 10 中。该插入片段的存在造成一个截短的端粒酶蛋白质，因为该插入片段的第一个密码子是终止密码子。(6) 序列“3”是一个位于碱基 3157 和 3158 之间 159bp 的插入片段 (图 10)。它的存在形成一个改变了羧基端的端粒酶蛋白质。该插入片段含有一个终止密码子。而且，序列“3”具有一个推测的 c-abl 的 SH3 结构域的结合位点 (PXXXXPXXP; PEMEPPRRP)。

与 *Euplotes* 和酵母的氨基酸相似性最接近的转录物含有序列 A 和 B，而不含有序列 C。由 mRNA 形成包括序列 A、B 和 C 的各种组合的八个变体的核苷酸和氨基酸序列列于图 8。

实施例 6

人端粒酶的的重组表达

人端粒酶被克隆到细菌表达载体中。从两个品种的 LIM 1215 mRNA

中扩增到图 1 所示的序列，并且将其连接起来。

扩增时，合成 cDNA 的第一条链并用于含有 DNA 聚合酶混合物的扩增反应中 (Titan 系统, Boehringer, IN)，这样一个校读热稳定的酶 (如 rTth) 与 Taq DNA 聚合酶一起使用。因为 LIM1215 中的许多 mRNA 缺少序列 B (图 9)，设计扩增引物，使得每对引物的一个引物位于序列 B 中，位于 2271 位核苷酸的 SacI 位点的任何一侧 (图 1)。用 HT2356R 和 HT0028F 首先从 cDNA 中扩增了 5' 部分 (循环条件: 70°C 2 分钟; 然后添加平衡至 50°C 的引物序列; 50°C 30 分钟; 95°C 2 分钟; 94°C 30 秒; 65°C 30 秒 2 个循环; 94°C 30 秒; 63°C 30 秒; 68°C 3 分钟 3 个循环; 94°C 30 秒; 60°C 30 秒; 68°C 3 分钟 32 个循环)。然后端粒酶基因的最远端 5' 部分连接在用 EcoRI/SacI 消化的 pTTQ18 (Amersham International plc, Buckinghamshire, 英国) 和 pBluescriptII KS+ 上，并且验证了该序列。

为了获得 3' 端，用 HT2230F 和 HT3292B 引物扩增 LIM1215 cDNA，该 cDNA 与编码端粒酶最 C-端的序列互补。扩增产物用 HindIII 和 SacI 消化，插入到 pTTQ18 和 pBluescript II KS+ 上。5' 端和 3' 端还被一起克隆在 pTTQ18 的天然 SacI 位点上作为六组氨酸融合体和非融合体蛋白质。

质粒 pTTQ18-Htel 被转染到细菌细胞中 (如 BL21 (DE3))。用 IPTG 诱导可以过量表达蛋白质。离心收集细菌并在裂解缓冲液中裂解 (20mM NaPO₄, pH7.0, 5mM EDTA, 5mM EGTA, 1mM DTT, 0.5 μg/ml 亮抑蛋白酶肽, 1 μg/ml 抑蛋白酶肽, 0.7 μg/ml 胃蛋白酶抑制剂)。用 Polytron 匀浆器均匀地悬浮此混合物，经过玻璃珠搅拌或通过微量流体器 (microfluidizer) 将细胞破碎。将裂解液 50,000 转/分离心 45 分钟。用 20mM NaPO₄, 1mM EDTA, pH7.0 (缓冲液 A) 稀释上清液。将稀释后的裂解液上清液上样于 SP-琼脂糖柱或与之相当的柱子，使用总计 6 倍柱体积的缓冲液 A 和以 0 至 30% 线性梯度加入缓冲液 A 中的缓冲液 B (1M NaCl, 20mM NaPO₄, 1mM EDTA, pH7.0)。混合含有端粒酶的级分。可以进行进一步的纯化。

对于六组氨酸融合体蛋白质，离心澄清裂解液，分批吸附到 Ni-IDA-琼脂糖柱上。将基质加入柱中并用缓冲液洗柱，一般缓冲液或是 50mM Tris pH 7.6, 1mM DTT; 50mM MES pH 7.0, 或是 IMAC 缓冲液（对于六组氨酸融合体）。结合到基质上的端粒酶蛋白质在含 NaCl 的缓冲液中被洗脱。

实施例 7

人端粒酶 RNA 组分的重组表达

人端粒酶 RNA 组分首先通过扩增从基因组 DNA 中分离。扩增引物是 telRNA T 和 telRNA 598B（图 5）。扩增条件是 95℃ 3 分钟；加聚合酶；80℃ 2 分钟；94℃ 30 秒；68℃ 2 分钟 35 个循环。

用 hTR TAC（具有 tac 启动子序列）和 hTR 3'Pst（具有顺式作用核酶序列）引物进行另一个扩增后，将扩增产物插入到 pBluescript 中。然后分离 pBluescript 插入片段并连接到 pACYC177。

实施例 8

人端粒酶亚区的表达

人端粒酶的逆转录酶结构域是通过与莫洛尼鼠类白血病病毒（Moloney MuLV）逆转录酶进行序列比较确定的。莫洛尼鼠类白血病病毒的手指/手掌区形成一个结晶稳定单元 (Georgiadis 等人, 结构 (Structure)3: 879, 1995)。许多残基和基元保留在两个蛋白质的活性位点上。设计引物扩增逆转录酶结构域和手指/手掌结构域，插入到表达载体中并分离蛋白质。

片段号	引物	氨基酸
I	BT-177/BT-178	AAEH... → ...VQMPAH
II	BT-177/BT-179	AAEH... → ...VGLGL
III	BT-182/BT-179	RATS... → ...VGLGL
IV	BT-183/BT-179	VQMPAH... → ...VGLGL

片段 I 编码相应于莫洛尼鼠类白血病病毒的“手指和手掌”结构域。

C-末端的“拇指”区和“连接”区(参考 Kohlstaedt 等人, 科学 256: 1783, 1992) 被删除了。片段 II 编码端粒酶的逆转录酶结构域, 以及 C-末端的“连接”区结构域。N-末端是通过与莫洛尼鼠类白血病病毒逆转录酶结构域比较大而选出的。片段 III 编码蛋白质的 C-末端。RATS 序列位于蛋白质的逆转录酶结构域(手掌区)内。片段 IV 编码含有“拇指”和“连接”结构域的 C-末端区并且可能起调节元件的作用。HIV-1 中的连接结构域在没有 RNase 结构域时能够堵塞 HIV 逆转录酶的催化裂缝(Kohlstaedt 等人, 同上)。C-末端区可能以一种类似的方式起调节(抑制)片段的作用。而且序列 C 具有一个推测的 SH3 结构域结合位点 c-abl (PXXXXPXXP; PEMEPPRRP, 参考图 8 的变体 2 序列)。c-abl 蛋白质直接与 ATM(共济失调毛细血管扩张)蛋白质相互作用(Shafman 等人, 自然 389: 520, 1997), ATM 蛋白质是一种明显参与细胞周期调控、减数分裂重组、端粒长度监测和 DNA 损伤应答的蛋白质。c-abl 蛋白质的结合可以通过标准的蛋白质-蛋白质相互作用方法加以测定。这样, 端粒酶与 c-abl 或其它含有 SH3 结构域的蛋白质(如 erb2)的相互作用, 以及通过端粒酶 C-末端在催化裂缝内或裂缝外的运动进行的调节是可以用此处描述的构件和产物进行控制的。在一种情况下, 这种调节可以通过磷酸化/去磷酸化反应介导的。

所有的引物都具有 HinIII 或 BamHI 位点。扩增反应是在 $1 \times$ Pfu 缓冲液、 $250 \mu\text{M}$ dNTP、 100ng 各种引物、克隆 53.2 模板 DNA 中进行的, 循环条件如下: 94°C 2 分钟; 55°C 、 60°C 或 65°C 2 分钟、 72°C 2 分钟、 94°C 1 分钟 25 个循环; 然后 72°C 10 分钟。获得了预期长度的产物 (BT-177/BT-178 966bp; BT-177/BT-179 1479bp; BT-182/BT-179 824bp; BT-183/BT-179 529bp)。用酚: 氯仿抽提扩增产物并用乙醇沉淀。重悬扩增产物并用在引物序列内有切点的合适的酶消化。

被消化的产物连接到经酶消化产生合适末端的 pBluescript 上, 将插入片段用 HindIII 消化并用 BamHI 部分消化以连接到 pGEX 上。质粒转染 BL21 (DE3) 细胞并在氨苄青霉素平板上选择。挑取克隆并在液体培养液中生长过夜。取一份培养液在加 $100 \mu\text{g/ml}$ 氨苄青霉素的 Terrific 液

中稀释。细胞在 37℃ 下生长并用 0.5mM IPTG 诱导至 O.D 值约为 0.8。继续生长 5 个小时。离心收集细胞并立即处理或冷冻在 -70℃ 备用。

从裂解细胞中纯化蛋白质。细胞菌体在 50mM Tris pH 8.0, 10mM 2-ME, 1mg/ml 溶菌酶, 0.5% Triton X-100, 1 μg/ml 胃蛋白酶抑制剂, 10 μg/ml 亮抑蛋白酶肽, 10 μg/ml 抑蛋白酶肽, 0.5mM PMSF, 和 2mM EDTA 中振荡混合, 冻融一次使之裂解。将裂解液离心澄清。将上清液加入到 50%GSH-琼脂糖浆液中, 在 4℃ 下搅拌 2 小时。将基质用裂解缓冲液和 50mM Tris pH8.0, 10mM 2-ME 洗两次。为了通过 SDS-PAGE 凝胶电泳进行分析, 加入含 150mM 2-ME 的样品缓冲液并将样品煮沸。

实施例 9

鼠端粒酶基因的分离

鼠端粒酶基因是从基因组文库或 cDNA 文库中分离的。鼠基因组文库是用细胞株 129 DNA 在 λ FIX II 载体上构建的。将文库涂平板, 把噬菌斑挑到尼龙膜上。尼龙膜用克隆 53.1 (1.9kb) 的插入片段在正常严谨条件下进行杂交。选出六个杂交斑点以作进一步的分析。

实施例 10

用 HT-1 和端粒酶变体说明端粒酶的活性

将全长的 hT-1 序列克隆至一表达载体, 分析产生的蛋白质的端粒酶活性。载体 pRc/CMV2(Invitrogen, Carlsbad, CA) 是一个真核表达载体, 在启动子 RSV LTR 之间具有多克隆位点, 以及聚腺苷酸化信号和来自牛生长激素基因的转录终止序列。将亮氨酸 49 密码子转变成甲硫氨酸密码子的端粒酶序列插入到 pRc/CMV2 中。从中选择一个克隆 pHTC51 进行进一步研究。测定了 5' 连接处的 DNA 序列并确定了插入片段的方向。随后, 测定了 3' 连接处的序列并且显示缺失了 polyA 信号, 但是没有缺失端粒酶编码序列。

将该克隆转染至第 44 代和第 68 代 HeLa GM847 细胞, 第 18 代 SUSM-1 细胞, 和第 40 代 RKF-T/A6 细胞。通过此处描述的 TRAP 分析

法分析细胞提取物的端粒酶活性。如图 12 中所示,以 1: 100 稀释的 SUSM-1 细胞的提取物中, 可以看到指示端粒酶活性的产物梯, 而在对照细胞中没有看到。在高浓度的提取物中不容易检测到产物带, 可能是由于提取物中的核酸酶活性的缘故。

构建了三个端粒酶变体: pAKI.4 是 β 区被剪接出的端粒酶 (图 13); pAKI.7 是 C-端插入片段 3 被改变的端粒酶 (图 14); 而 pAKI.14 是 α 区被剪接出的端粒酶 (图 15)。端粒酶基因的 5' 端分别插入至这三个载体中并将插入片段转移至 pCIneo 表达载体中。这些变体, 以及 pCIneo 中的对照端粒酶被瞬时转染至 GM847 细胞中, GM847 细胞中没有可检测到的端粒酶活性但是表达 RNA 亚单位。用 TRAP 法检测细胞提取物。对照端粒酶表现出活性, 具有插入片段 3 (pAKI.7 插入) 的端粒酶也有活性, 而其它的变体不表达活性。

从前面可以看出, 尽管为了进行说明已经在此描述了本发明的几个特殊的实施例, 但是可以在不背离本发明的精神和范围的前提下进行各种修正。因此, 除了附加的权利要求以外, 本发明没有加以限制。

说明书附图

人端粒酶

ATGCCGCGCGCTCCCGCTGCCGAGCCGTGCGCTCCCTGCTGGCAGCCACTACCGCGAG	60
MetProArgAlaProArgCysArgAlaValArgSerLeuLeuArgSerHisTyrArgGlu	20
GTGCTGCCGCTGGCCACGTTCTGTCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGCGAG	120
ValLeuProLeuAlaThrPheValArgArgLeuGlyProGlnGlyTrpArgLeuValGln	40
CGCGGGGACCCGGCGGCTTTCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCCTGG	180
ArgGlyAspProAlaAlaPheArgAlaLeuValAlaGlnCysLeuValCysValProTrp	60
GACGCAGGGCCGCCCCCGCGCCCCCTCCTCCGCCAGGTGCCTGCCTGAAGGAGCTG	240
AspAlaArgProProProAlaAlaProSerPheArgGlnValSerCysLeuLysGluLeu	80
GTGGCCCGAGTGTGCAGAGGCTGTGCGAGCGCGCGGAAGAACGTGCTGGCCTTCGGC	300
ValAlaArgValLeuGlnArgLeuCysGluArgGlyAlaLysAsnValLeuAlaPheGly	100
TTCGCGCTGCTGGACGGGGCCGCGGGGGCCCCCGAGGCCCTCACACCAGCGTGGCG	360
PheAlaLeuLeuAspGlyAlaArgGlyGlyProProGluAlaPheThrThrSerValArg	120
AGCTACCTGCCAACACGGTGACCGACGCACTGCGGGGAGCGGGCGTGGGGCTGCTG	420
SerTyrLeuProAsnThrValThrAspAlaLeuArgGlySerGlyAlaTrpGlyLeuLeu	140
TTGCGCCGCGTGGGGACGACGTGCTGGTTCACCTGCTGGCAGCTGCGCGCTCTTTGTG	480
LeuArgArgValGlyAspAspValLeuValHisLeuLeuAlaArgCysAlaLeuPheVal	160
CTGGTGGCTCCCAGCTGCGCCTACCAGGTGTGCGGGCCGCGCTGTACCAGCTCGGCGCT	540
LeuValAlaProSerCysAlaTyrGlnValCysGlyProProLeuTyrGlnLeuGlyAla	180
GCCACTCAGGCCCGCCCCGCCACACGCTAGTGGACCCGAAGGCGTCTGGGATGCGAA	600
AlaThrGlnAlaArgProProProHisAlaSerGlyProArgArgArgLeuGlyCysGlu	200
CGGGCCTGGAACCATAGCGTCAGGGAGGCCGGGTCCCCCTGGGCCTGCCAGCCCCGGT	660
ArgAlaTrpAsnHisSerValArgGluAlaGlyValProLeuGlyLeuProAlaProGly	220
GCGAGGAGGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTTGCCAAGAGGCCAGGCGT	720
AlaArgArgArgGlyGlySerAlaSerArgSerLeuProLeuProLysArgProArgArg	240

图 1A

GGCGCTGCCCTGAGCCGGAGCGGACGCCCGTTGGGCAGGGGTCTGGGCCACCCGGGC GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly	780 260
AGGACGCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGCACCTGCCAGACCCGCCGAA ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu	840 280
GAAGCCACCTCTTTGGAGGGTGGCTCTCTGGCACGGCCACTCCACCCATCCGTGGGC GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly	900 300
CGCCAGCACCCAGCGGGCCCCCATCCACATCGCGGCCACCACGTCCCTGGGACAGGCCT ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro	960 320
TGTCCCCGGTGTACGCCGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAG CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln	1020 340
CTGGCGCCCTCCTTCTACTCAGCTCTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTC LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu	1080 360
GTGGAGACCATCTTTCTGGGTCCAGGCCCTGGATGCCAGGGACTCCCCGAGGTTGCC ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro	1140 380
CGCCTGCCCCAGCGCTACTGGCAAATGCGGCCCTGTTTCTGGAGCTGCTGGGAACCAC ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis	1200 400
GCGCAGTGCCCTACGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCACC AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr	1260 420
CCAGCAGCCGGTGTCTGTGCCCGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCGAGGAG ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu	1320 440
GAGGACACAGACCCCGTCCGCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCTGGCAG GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln	1380 460
GTGTACGGCTTCGTGCGGGCCTGCCTGCGCGGCTGGTGCCCCAGGCCTCTGGGGCTCC ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer	1440 480
AGGCACAACGAACGCCGCTTCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCAT ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis	1500 500

图 1B

GCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGGCTGCGCTTGGCTG AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu	1560 520
CGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGAGAGCACCGTCTGCGTGAGGAGATC ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle	1620 540
CTGGCCAAGTTCCTGCACTGGCTGATGAGTGTACGTGCGAGCTGCTCAGGTCTTTC LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe	1680 560
TTTTATGTCACGGAGACCACGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTG PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal	1740 580
TGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAG TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu	1800 600
CTGTGGAAGCAGAGGTGAGGAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGA LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg	1860 620
CTCGCTTACCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTG LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal	1920 640
GGAGCCAGAAGCTTCGCGAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCA GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla	1980 660
CTGTTACAGGTGCTCAACTACGAGCGGGCGGGCCCGCCCTCCTGGGCGCCTCTGTG LeuPheSerValLeuAsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAlaSerVal	2040 680
CTGGGCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCAG LeuGlyLeuAspAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln	2100 700
GACCCGCCGCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATC AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle	2160 720
CCCCAGGACAGGCTCACGGAGGTCAATCGCCAGCATCATCAAACCCAGAACACGTACTION ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys	2220 740
GTGCGTCGGTATGCCGTGGTCCAGAAGGCCGCCATGGGCACGTCCGCAAGGCCTTCAAG ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys	2280 760

图 1C

AGCCACGTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCTGGCTCACCTG SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu	2340 780
CAGGAGACCAGCCCGCTGAGGGATGCCGTCGTATCGAGCAGAGCTCCTCCCTGAATGAG GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu	2400 800
GCCAGCAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCAGCCGTGGGCATC AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle	2460 820
AGGGGCAAGTCTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTG ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu	2520 840
CTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAGTGTGTTGCGGGGATTCGGCGGGAC LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp	2580 860
GGGCTGCTCCTGCGTTTGGTGGATGATTTCTTGTGGTGACACCTCACCTCACCCACGGC GlyLeuLeuLeuArgLeuValAspPheLeuLeuValThrProHisLeuThrHisAla	2640 880
AAAACCTCCTCAGGACCCTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGGTGAAGTGG LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu	2700 900
CGGAAGACAGTGGTGAAGTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTT ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal	2760 920
CAGATGCCGGCCACGGCCTATTCCTGGTGGGCTGCTGCTGGATACCCGGACCCTG GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuLeuAspThrArgThrLeu	2820 940
GAGGTGCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTC GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe	2880 960
AACCGCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGGCGCTG AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu	2940 980
AAGTGTACAGCCTGTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAAC LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn	3000 1000
ATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCA IleTyrLysIleLeuLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro	3060 1020

图 1D

TTTCATCAGCAAGTTTGGGAAGAAGCCACATTTTTCTGCGCGTCATCTCTGACACGGCC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGCCAAGGGC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGCCGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGCTGACTCGACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ACGCAGCTGAGTCGGAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGCCGCAGCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTgatggccacccgcccacagccag	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132
Gccgagagcagacaccagcagccctgtcacgccgggctctacgtcccagggagggagggg	3480
Cggcccacaccaggcccgcaccgctgggagtctgaggcctgagtgagtgtttggccgag	3540
gcctgcatgtccggctgaaggctgagtgtccggctgaggcctgagcgagtgtccagccaa	3600
gggctgagtgtccagcacacctgccgtcttcaactccccacaggctggcgctcggtcca	3660
ccccagggccagcttttctcaccaggagcccggcttccactccccacataggaatagtc	3720
catcccagattcgccattgttcacccctcgccctgccctcctttgccttccacccccac	3780
catccaggtggagaccctgagaaggaccctgggagctctgggaatttgagtgaccaaag	3840
gtgtgccctgtacacaggcgaggacctgcacctggatgggggtccctgtgggtcaaatt	3900
ggggggaggtgctgtgggagtaaaatactgaatataatgagttttcagttttgaaaaaaa	3960
aaaa	3964

图 1E

Euplotes 1 -----MEVDVDNQADNHGIHSALKTCEEIKEAKTLYSWIQKVIKRC--NQSQSHYKDLEDIK
 HT1 1 RRLGPOGWRLVQRGDPAAFRLVAQCLVCVPWDAR-PPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFGFALLDGA
 EST2 1 -----MKILFEFIQDKLDID--LQTNSTYKENLCKG

Euplotes 56 IFAQTNIVA7PRDYNEEDFKVIARK-----EVFSTGLMIELIDKCLVELLSSSDVSDRQKLCQCFGFOLKGNQ-LAK
 HT1 80 RGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLRRVGDVVLVHLLARCALFVLVAPSCAY---QVCGPPLYQLGAATQA
 EST2 30 HFNGLDEILTTCFALPNSRKIALP-----CLPGDLSHKAVIDHCIIYLLTGELYNN---VLTFGYKIARVEDVNN

Euplotes 126 THLLTALSTQKQYFFQDEWNOVRAMIGNELFRHLYTKYLIFQRTSEGTLVQFCGNNVFDHLKVNDKFDKQKGGGAADMNE
 HT1 157 RPPPHASGPRRRLGCEAAWHHSVREAGVPLGLPAGARRRGGASRSLPLPKRPRRGAAPERTPVQGGSWAHPGRTRG
 EST2 97 SLFCHSANVNVTLKGAAWKMFHSLVGYAFVDLLINYTVIQFNGQ-FFTQIVGNRCNEPHLPPKWVQRSSSS-----

Euplotes 206 PRCCSTCKYNVKNKDHFLWNI-----NVPNWNMKSRTIRIFYCTHFNRNNQFF
 HT1 237 PSDRGFCVSPARPAFEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAEKHFLYSSGDK--EQLR
 EST2 169 ----SATAAIKQLTEPVTN-----KQFLHKLNIIN-SSSFF

Euplotes 255 KKHEFVSNKNNISAM-DRAQTIFTNI-----FRNRIRKLLKDKVIEKIAYMLEKVKDFNFNYLLTKSCPLPENWRE
 HT1 315 PSFLLSSLRPSLTGARLVETIFLGSRPWPGTPRRLPRLPQRY-WQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTP
 EST2 200 PYSKILPSSSSIKKLTDLREAFIP-----TNLVKIPQRLKVRINLTQKLLKRHKRLNYSILNSICPPLEGT--

端粒酶结构域

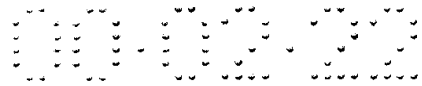
Euplotes 326 RK-----QKIENLINKTREEKS--KYEEELFSYTTDNKCVTQFINEFFYNILPKDFLTGR-MRKNFQKKVKKYVELNKHE
 HT1 394 AAGVCAREKPOGSVAAPAEEDTDPRLVQLRQHSPPWQYVGFVRACLRRLVPPGLWGSRRHWERRFLRNTKKFISLGKHA
 EST2 268 -----VLDLSHLSRQ-----SPKERVLFIVILQKLLPQEMFGSKKWKGIKKNLNLLSLPLNG

Euplotes 398 LIHKNLLEKINTREISWQVET-SAKHFYFDHENIYVLWKLRLRWIFEDLVVSLIRCFYVTEQQKSYSKTYYYRKNIW
 HT1 474 KLSLQELTWKMSVRDCAWLRSPGVCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTEFTFQKNRLLFFYRKSVM
 EST2 324 YLPFDLSLLKLRKDFRWFISD-IWFTKHNFENLN-QLAICFISWLFROLIPKIIQTFYFCTEIS-STVTIVYFRHDTW

基元 1 基元 2

Euplotes 477 DVIMKMSIADLKK-ETLAEVQEKEVEEWKSL-GFAPGKLRLIPKKT--FRPIMTFNKKIVNSDRK--TTKLTNTKLL
 HT1 554 SKLQSIGIRQHLKRVQLRELSEAEVROHREARPALLTSRLRFIPKPDG--LRPIVNM DYVVGARTFRREKRAERLTSRVK
 EST2 401 NKLITPFIVEYFK-TYLVENNVCRNHSNYTSL-NFNHSHKRIIPKKSNEFRIIAIPCRGADEEEFT--IYKENHKNIAIQ

图 2A



基元 A

Euplotes 551 NSHLMKTLKN-RMFKDPFGFAVFNDDVMKKYEEFVCKWKQVGOP-KLFFATMDIEKCYDSVNRKLSFLKTTKLLSS
 HT1 632 ALFVSLNYERARR--PGLLGASVGLDDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPN
 EST2 477 PTQKILEYLRNKRPTSFTKIYSPTQIADRIKEFKORLLKKFNNVLP-ELYFMKFDVKSCYDSIPRMECMRILKDALKNE

Euplotes 629 DFWIMTAQILKRKNIVIDSKNFRKKEKMDYFROKFQKIALEGGQYPTLFSVLENEQNOLNAKKTLLIVEAK-QRWYFKD
 HT1 710 TYCVRRYAVVQKAAHGHVRKAFKSHVS-----TLTDLQPYMRQFVAHLQETSPLRDVAVIEQSSSLVEASSG
 EST2 556 GFFVRSQYFFN-TNTGVLKLFNVVN-----A--SRVPKPYELYIDNVRTVHLSNQDVINNV-EMEIFKT-

基元 B

基元 C

Euplotes 708 NLLQPVINICQYNYINFNGKFYKQTKGIPOGLCVSSILSSFYATLEESSLGLRDESMNPENPNVLLMRLTDDYLLIT
 HT1 777 LFDVFLRFMCHHAVRIR-GKSYVQCQGIPOGSILSTLLCSLCYGD MEN---KLFAGIRRD-----GLLLRLVDDFLVLT
 EST2 616 --ALWVEDKCYIR-----EDGLFQSSLSAPIVDLVYDLELFYSEFKASPSQD-----TLILKLADDFLIIS

基元 D

基元 E

Euplotes 788 TQENNAVLFIKLIINVSRENGFKFMKQLQTSFPLSPSKFAKYGMDSVEEQNI VQDYCDWIGISIDMKTALMPNINLRI
 HT1 847 PHLTHAKTFLRTLVRGVPEYGCVVMLRKTVVNFPVEDEALGG-TAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYAR--
 EST2 677 TDQQ-QVINIKKLAMG----GFQKYNAPANRDKILAVS-----SQSDDDTVIQFCAMHIFVKELEVWKHSSTMV---

Euplotes 868 EGILCTLMLNMQTKKASMLKKKLKSFLMNNITHYFRKTITTEDFANKTLNKLFISSGGYKYMCAKEY--KDHFKKVLAM
 HT1 924 TSIRASLTFNRGFKAGRNMRRKLFGLVRLKCHSLFLDLQVNSLQTVCTNIYKILLLOAYRFHACVLQLPFHQQVWKNPTF
 EST2 741 -----NFHIRSKSS----KGIFRSLIALFNTRISYKTIDTNLNSTNTVLMQIDHVVKNISECYKSA--FKDLSIMVTO

Euplotes 946 SSMIDLEVSKJZYSVTRAFFKYLCNIKDTIFGEEHYDFFLSTLKHFIIEIFSTKKYIFNRVCMILKAKEAKLKSDQCQS
 HT1 1004 FLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLC-HQAFLLKLTRHRVTVYVPLLSLRTAQTQLSRKLPGT
 EST2 808 NMQFHSFLQRJIEMTVSG----CPITKCDPLIEYEV--FTI--LNGFLESLSNNTSKF-KDNIILLRKEIQHLOAYIYI

Euplotes 1026 LIQYDA-----
 HT1 1083 TLTALEAAANPALPSDFKTILD
 EST2 879 YZHIIVN-----

图 2B

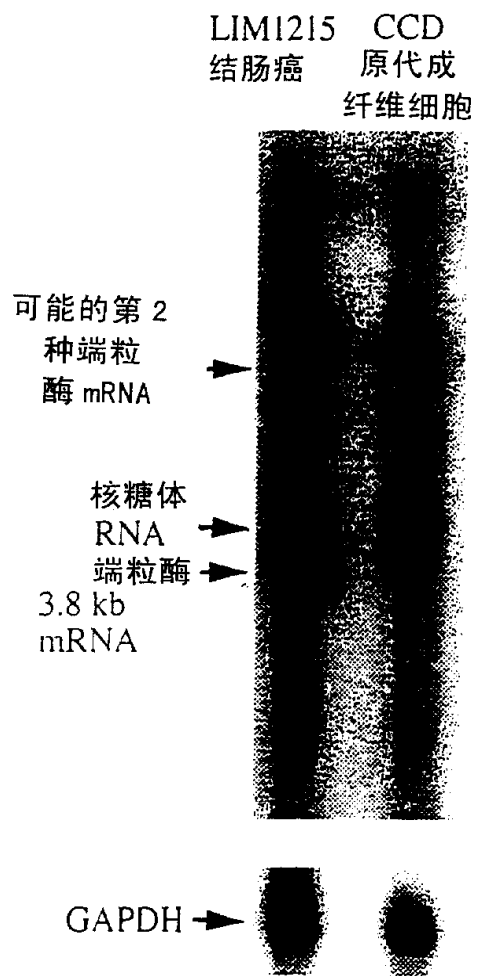


图 3

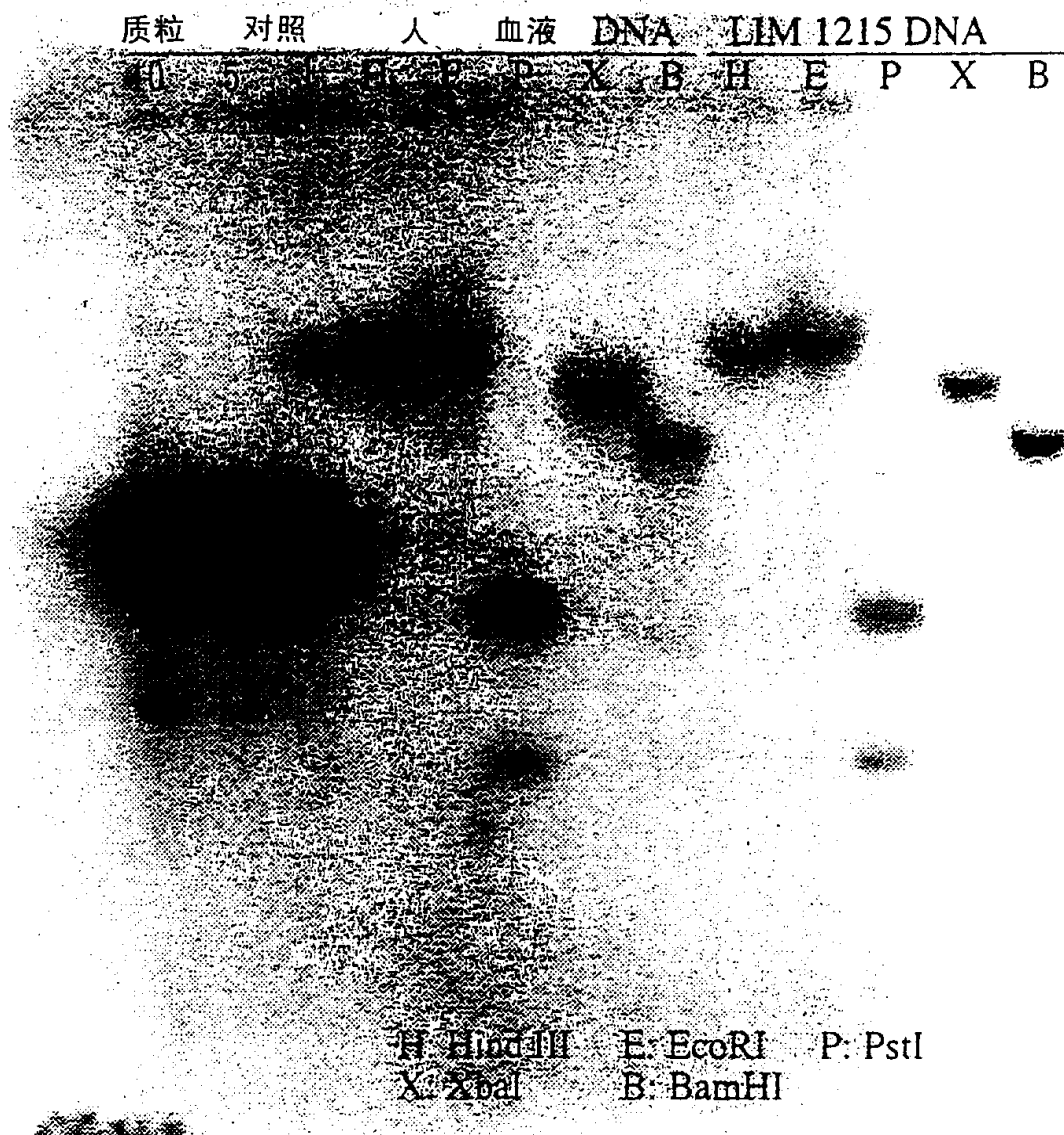


图 4

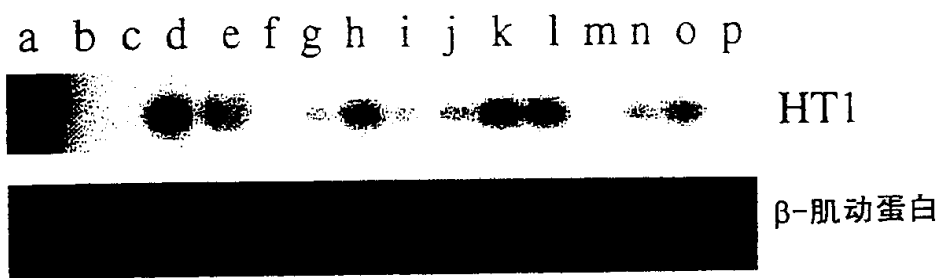


图 5

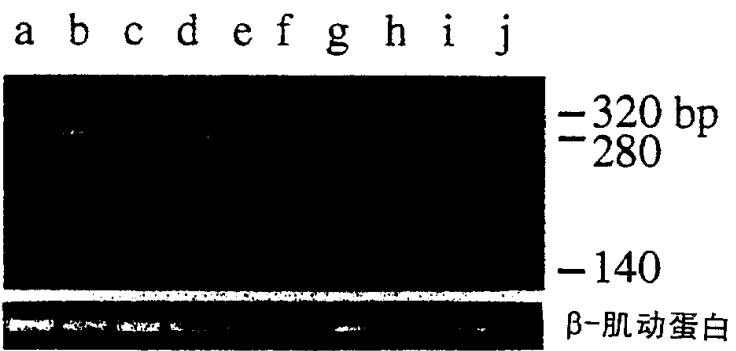


图 6

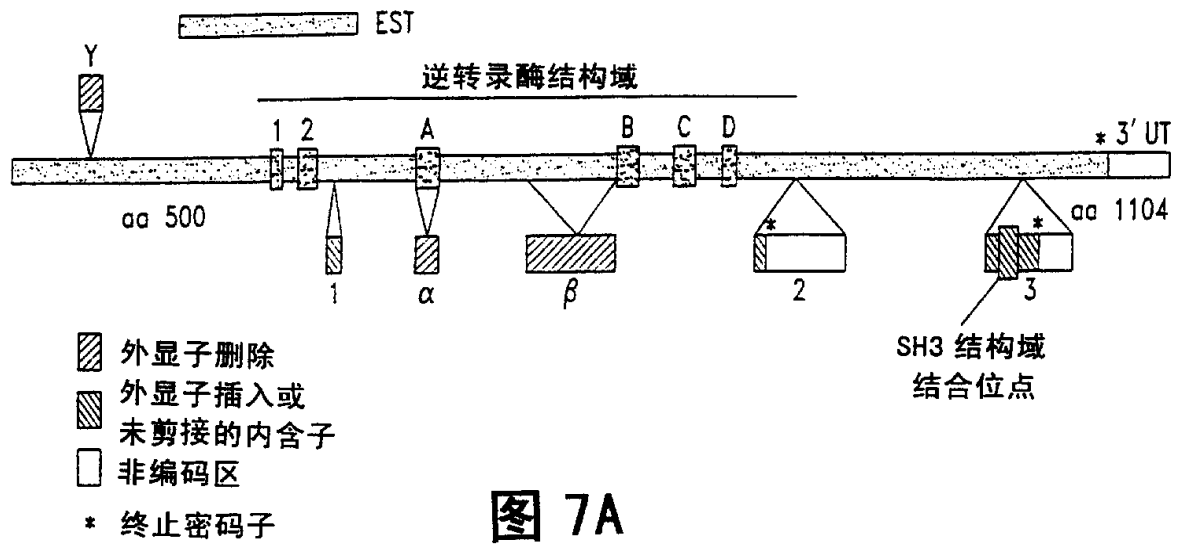


图 7A

变体:	1	α	β	2	3
RT-PCR 产物	未作	+	+	未作	+ & -
来自LIM1215文库的PCR产物	-	+	-	+	未作
RT-PCR 产物	未作	-	+	未作	+
53.2 cDNA	-	-	-	-	未作

图 7B

222
 Y 5'-CCAGGTG|ggcctc
 223
 gcaggtg|TCCTGCC-3'

1950
 1 5'-AAAGAGG|GTGGCTG.....AACAGAA|GCCGAGC-3'
 1952

2130
 α 5'-TGCAAG|gtggatg.....ccccag|GACAGGC-3'
 2167

2286
 β 5'-GAGCCAC|gtctcta.....ggggcaa|GTCCTAC-3'
 2468

2843
 2 5'-ACTCCAG|GTGAGCG.....XXXXXXXX|CTATGCC-3'
 2844

3157
 3 5'-AACGCAG|CCGAAGAAAACATTTCTGTCGTGACTCCTGCGGTGCTTGGGTCGGGACAGCCAGAGATGG
 T A A E E N I L V V T P A V L G S G Q P E M E
 AGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTCCGGTGTCTCCTGGGAGGGGAGTTG
 P P R R P S G V G S F P V S P G R G V G
 3158
 GGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCAG|GGATGTC-3'
 L G L *

图 7C

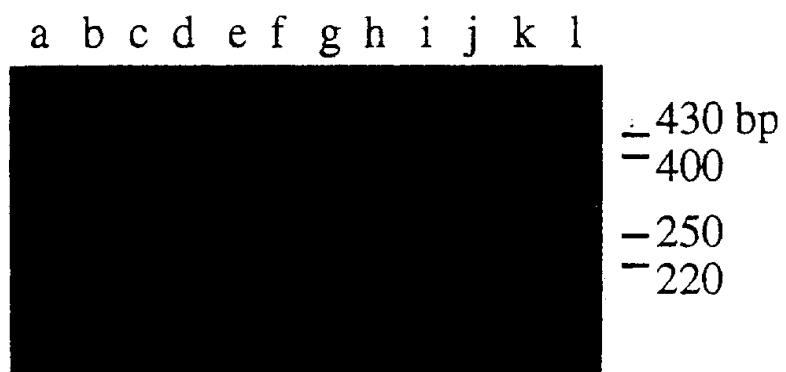


图 8

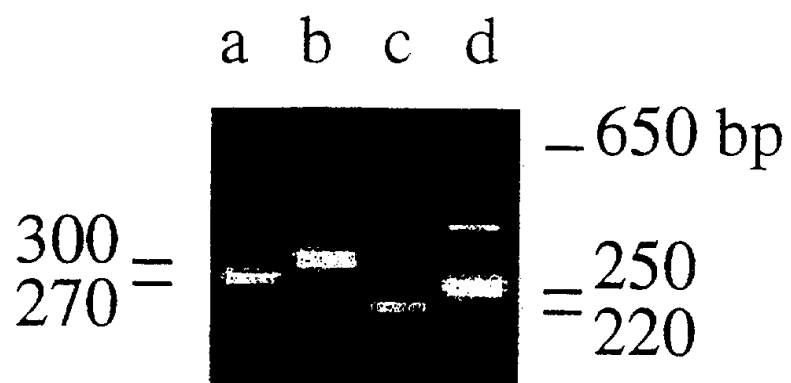


图 9

“ Y ” 104-105 序列碱基

GGCCTCCCCGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAG
GlyLeuProGlyValGlyValArgLeuGlyLeuArgAlaAlaGlyGlyAsnGln
AlaSerProGlySerAlaSerGlyTrpGly * GlyArgProGlyGlyThrSer
ProProArgGlyArgArgProAlaGlyValGluGlyGlyArgGlyGluProAla

CGACATGCGGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCCGCAGGTG
ArgHisAlaGluSerSerAlaGlyAspSerGlyArgPheProArgArg
AspMetArgArgAlaAlaGlnAlaThrGlnGlyAlaSerProAlaGly
ThrCysGlyGluGlnArgArgArgLeuArgAlaLeuProProGlnVal

“ 1 ” 38 序列碱基

GTGGCTGTGCTTTGGTTAACTTCCTTTTAAACCAGAA
ValAlaValLeuTrpPheAsnPheLeuPheAsnGlnLys

“ * ” 36 序列碱基

GTGGATGTGACGGGCGGTACGACACCATCCCCAG
ValAspValThrGlyAlaTyrAspThrIleProGln

“ * ” 182 序列碱基

GTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTG
ValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu

CAGGAGACCAGCCCCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTG
GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeu

AATGAGGCCAGCAGTGGCCTCTTCGACGTCTTCTACGCTTCATGTGCCACCAC
AsnGluAlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHis

GCCGTGCGCATCAGGGGCAA
AlaValArgIleArgGlyLys

“ 2 ” 部分序列长度未知

GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCGGCTGGGGCAGGTGCTGCTGCAG
Ter

GGCCGTTGCGTCCACCTCTGCTTCGGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGT
CAGATGCCACAGGGTGCCCTCGTCCCCTCTGGGGCTGAGCACAATGCATCTTTCTG
TGGGAGTGAGGGTGCCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAA...

图 10A

“ 3 ” 159 序列碱基

CCGAAGAAAACATTTCTGTCGTGACTCCTGCGGTGCTTGGGTGGGACAGCCAGAG
AlaGluGluAsnIleSerValValThrProAlaValLeuGlySerGlyGlnProGlu

ATGGAGCCACCCCGCAGACCGTCGGGTGTTGGGCAGCTTCCGGTGTCTCCTGGGAGG
MetGluProProArgArgProSerGlyValGlySerPheProValSerProGlyArg

GGAGTTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG
GlyValGlyLeuGlyLeu *

“ x ” 序列长度未知

_GACAGTCACCAGGGGGTTGACCGCCGACTGGGCGTCCCAGGGTTGACTATAGGA
CCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAGAGGCGTCTGGCTGGCATGG
GTGGACGTGGCCCCGGGCATGGCCTTCTGCGTGTGCTGCCGTGGGTGCCCTGAGCCCT
CACTGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCCCTAGTCTGTTGTCTGG
CTGAGCAAGCCTCCTGAGGGGCTCTATTG

基因组内含子部分序列 (约 2.7kb)

GTGGCTGTGCTTTGGTTTAACTTCCTTTTTAACCAGAAGTGCGTTTGAGCCCCACATT
TGGTATCAGCTTAGATGAAGGGCCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCAT
GGCACGGCGCCACCCATTTGTGCGCACAGTGAGGTGGCCGAGGTGCCGGTGCCTCCA
GAAAAGCAGCGTGGGGTGTAGGGGAGCTCCTGGGGCAGGGAC...

图 10B

截短的蛋白质 1

ATGCCGCGCGTCCCGCTGCCAGCCGTGGGCTCCCTGCTGCCAGCCACTACCGGAGGTGCTGCCCTGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGGCTGGGGCCAGGGCTGGCGGCTGGTGCAGGCGGGACCCGGCGCTTCCGGCGCTGGTGGCCAGTCCCTGGTGTGGTGGCTGGGACGCACGGCCCGCCCGCGG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCCCTCCGCGAGGTGTCTGCTGAAGGAGCTGGTGGCCGAGTGGTGCAGAGGCTGTGGAGGGCGGCAAGAAGTGTGGCTTCGGCTTCGGCTGGACGGGCGCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGGCTTCCACCACAGGCTGCCAGTACCTGCCAACAGGTCACCGAGCACTGGGGGAGCGGGGCTGGGGCTGTGCTGCCGCGTGGGACAGGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTCACTGCTGGCAGCTGCCGCTCTTGTGTGGTCCAGCTGGGCTACCAGTGTGGGGCGCGCTGTACCAGCTGGGCTGCCACTCAGGCCCGCCCGCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCAAAGCGTCTGGGATGGACCGGGCTGGAACCATAGGCTCAGGAGCGGGGTCCCGCTGGGCTGCCAGCCCGGTTGGAGGAGGGCGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCAGTCTGCGTTGCCAAGAGGCCAGGGTGGCGTCCCGTGAGCCGAGCGGACGCCCTGGGAGGGTCTGGGCCACCGGGCAGGACGCTGGACCGAGTGCAG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTCTGTGTGTGTCACCTGCCAGACCGGCCAAGAGCCACCTTTGGAGGTCGCTCTGTGCCAGCGCCACTCCACCCATCCGTGGGCGCCAGCACCCAGCGGGCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGGACAGGCTTGTCCCGGTTACGGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGAGCTGGCCCTCCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGTAGGCCAGCCTGACTGGCGCTGGAGGCTGTGGAGCATTCTTCTGGTTCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGCCTGCCAGCGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGCGCCCTGTTTCTGGAGCTGCTTGGGAACACCGCAGTCCCGTACGGGTCCTCTCAAGACCACTGCCCGCTGGAGCTGGGTCACCCAGCAGCCGGTGTGTGCCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCAGGGCTCTGTGGCGGCCCGAGGAGGAGACACAGCCCGCTGGCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCTGGCAGGTGTAGGGCTTGTGGGGCTG
E X P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCGCGGCTGGTGCCTCCAGGCTTGGGCTCCAGGCACAACGAGCGCTTCTCAGGAACACCAAGAAGTTCACTCCCTGGGAAGCATGCCAAGCTCTGCTGCAGGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGCTGGGACTGGCTTGGCTGGCAGGAGCCAGGCGTGGCTGTGTTCCGGCCGAGAGCAGGCTGGGTGAGGAGATCCTGGCCAGTTCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACGCTGCGAGTGTGCTGAGTCTTTTATGTACCGAGACCGCTTCAAAGAACAGGCTCTTTTCTACGGGAAGTGTCTGGACCAAGTTCCAAAGCATGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

图 11B

AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTGGAAGCAGAGGTGAGGCAGCATCGGGAAGCCAAGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

GTGGCTGTGCTTTGGTTAACTTCCTTTTAAACCAGAA
V A V L W F T F L F N Q K

CGGCTGCGGCCGATTGTGAACATGGACTACGTGTGGGAGCCAGACGTTCCGCAGAGAAAAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTACGCGTGTCAACTACGA
G L R P I V N M D Y V V G A R T F R R E K R P S V S F R G *

图 11C

截短的蛋白质 2

ATGCCGCGCGCTCCCGCTGCCAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCAGGTGCTGCCGCTGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGCTCGGGGCCAGGGCTGGCGCTGGTGCAGCGGGGACCGGGGCTTTCCGCGCGCTGGTGGCCAGTGCTGGTGTGCGTGCCTGGGACGCAAGGCGCCCGCCCGCCCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTCCGCGAGGTGTCTGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCGAGCGCGCGGAAGAACGCTGCTGGCTTCGGCTTCGCGTCTGAGCGGGGCGG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTACCACCAGGTCGCGAGCTACCTGCCAACAGGTGACCGAGCACTGCGGGGAGCGGGCGTGGGGCTGCTGCTGCGCGCTGGGCGACGACGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGCGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGCCTACCAGGTGTGGGGCCGCGCTGTACCAGCTCGGGCTGCCACTCAGGCCGCGCCCGCC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAAAGGCGTCTGGGATGCGAACGGCTGSAACCATACGCTCAGGAGGCGGGGTCCCCCTGGGCTGCCAGCCCGGGTGGAGGAGGCGGGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAAGTCTGCGTTGCCAAGAGGCCAGGCTGGCGCTGCCCTGAGCCGAGGGAGCGCGGTTGGGCAGGGTCTGGGCCACCCGGCAGGACGCGTGGACCGAGTGAACC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTGCTCACCTGCGAGACCCGCGAAGAAGCCACCTCTTTGAGGGTGGCTCTCTGCGCAGCGCCACTCCGACCCATCCGTGGGCGCCAGCACCAGCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGCCACGACGCTCCCTGGGACAGGCTTGTCCCGCGTGTACGCGAGACCAAGCACTTCTCTACTCCTCAGGGACAGGAGCAGTGGCGCCCTCCTTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

图 11D

CTCTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTCGTGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGACTCCCCGAGGTTGCCCGCCTGCCAGCCTACTGGCA
 S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q
 AATGCGGCCCTGTTTCTGGAGCTGCTTGGGAACCACGGCAGTGCCTTACGGGTTGCTCCTCAAGACGCACTGCCGCTGGAGCTGCGGTCAECCAGCAGCCGGTGTCTGTGCCG
 M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R
 GGAGAAGCCCGAGGCTCTGTGGCGGCCCGAGGAGGAGACACAGACCCCGTGGCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCTGGCAGGTGTACGGCTTCTGGCGGCTG
 E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C
 CCTGCGCCGCTGTTGCCCCAGGCTCTGGGCTCCAGGCACAACAGCCGCTTCTCAGGAACACCAAGAAGTTATCTCCCTGGGAAGCATGCCAAGCTCTCGTGCAGGAGT
 L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L
 GAGTGGAGATGAGCGTGGGACTGCGCTTGGCTGGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGACAGCAGCCGCTGCGGTGAGGAGATCCTGGCAAGTCTCTGCAGTGGT
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L
 GATGAGTGTGTACGTGCGAGCTGCTCAGGCTTTCTTTTATGTACCGAGACCAGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTGCAAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AATCAGACAGCACTTGAAGAGGTTGACGCTCGGGAGCTGTGGAAAGCAGAGTCCAGCAGCATCGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTATCCCCAAGCCTGA
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
 CCGGCTGCGGCCGATTGTGAACATGGACTACGCTCGTGGGAGCCAGAACGTTCCGCAGAGAAAAGGGCCGAGCGTCTCAOCTCGAGGTTGAAGGCACTGTTGAGCGTCTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCGCGCCCGCCGCTCTGGGCGCTGTGTGGTGGGCTGGACATATCCACAGGCGCTGGCGCACCTTGTGCTGCGTGTGGGGCCAGGACCCCGCCCTGAGCTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAAGTGGATGTGACGGGCGGTACGACACCATCCCCAGGACAGGCTCAGGAGTTCATGCCAGCATCAAAACCCAGAACAGTACTGCGTGGTGGTATGCCGTGGTCCA
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGCCGCCATGGGCACGTCGCAAGGCTTCAAGAGCCAC
 K A A H G H V R K A F K S
 GTCTACGTCCAGT
 V L R P V
 CCAGGGATCCCGCAGGCTCCATCTCTCCACGCTGCTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTGGGGGATTGGCGGGAGGGCTGCTCTCGGTTTGGTGA
 P G D P A G L H P L H A A L Q P V L R R H G E Q A V C G D S A G R A A P A F V G
 TGATTTCTGTTGGTACACCTCACCTCACCCAGCGAAAAACCTTCTCAGGACCGTGGTCCGAGGTGCTCCTGAGTATGGCTGGTGGTGAACCTTGGGAAGACAGTGTGACTTCCC

图 11E

对照蛋白质

ATGCCGCGCGCTCCCGCTGCCGAGCGGTGGCTCCCTGCTGCGCAGCCACTACCGGGAG	60
MetProArgAlaProArgCysArgAlaValArgSerLeuLeuArgSerHisTyrArgGlu	20
GTEGTGCCGCTGGCCACGTTGCTGGGGCGCTGGGGCCCAGGGGTGGCGGTGGTGACG	120
ValLeuProLeuAlaThrPheValArgArgLeuGlyProGlnGlyTrpArgLeuValGln	40
CGCGGGGACCCGGCGGCTTCCGCGCGGTGGTGGCCAGTGCCCTGGTGTGGTGGCCGGS	180
ArgGlyAspProAlaAlaPheArgAlaLeuValAlaGlnCysLeuValCysValProTrp	60
GADGCACGGCCGCCCGCGCGCCCTCCCTCCGCGAGGTGCTGCTGAAGGAGCTG	240
AspAlaArgProProProAlaAlaProSerPheArgGlnValSerCysLeuLysGluLeu	80
GTGGCCCGAGTGTCTCAGAGGCTGTGCGAGCGGGCGGAAGAAGTGTGGCTTCGGC	300
ValAlaArgValLeuGlnArgLeuCysGluArgGlyAlaLysAsnValLeuAlaPheGly	100
TTCCGCGTGTGGACGGGGCCCGGGGGCCCCCGAGGCTTCACCACAGCGTGGCG	360
PheAlaLeuLeuAspGlyAlaArgGlyGlyProProGluAlaPheThrThrSerValArg	120
AGCTACCTGCCAACACGGTGACCGAGGCACTGCGGGGAGGGGGCGTGGGGGTGCTG	420
SerTyrLeuProAsnThrValThrAspAlaLeuArgGlySerGlyAlaTrpGlyLeuLeu	140
TTGGCGCGGTGGGCGACGAGTGTGTTACCTGCTGGCAGGCTGGCGCTCTTTGTG	480
LeuArgArgValGlyAspAspValLeuValHisLeuLeuAlaArgCysAlaLeuPheVal	160
CTGGTGGCTCCAGCTGCGCTACAGGTGTGGGGCCCGGCTGTACCAGCTGGGGCT	540
LeuValAlaProSerCysAlaTyrGlnValCysGlyProProLeuTyrGlnLeuGlyAla	180
GCCACTCAGGCCCGCCCCCACACGCTAGTGGACCCCGAAGGCGTCTGGGATGCGAA	600
AlaThrGlnAlaArgProProProHisAlaSerGlyProArgArgArgLeuGlyCysGlu	200
CGGGCTGGAACCATAGCGTCAAGGAGGGCGGGTCCCGCTGGGCTGCCAGCCCCGGT	660
ArgAlaTrpAsnHisSerValArgGluAlaGlyValProLeuGlyLeuProAlaProGly	220
GCGAGGAGCGCGGGCGAGTGCACGCCAAGTCTGCCGTTGCCAAGAGGCCAGGCGT	720
AlaArgArgArgGlyGlySerAlaSerArgSerLeuProLeuProLysArgProArgArg	240
GGCGTGCCTGAGCGGAGCGSACGCCGTTGGGAGGGGTCTGGGCCACCCGGGC	780
GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly	260
AGGACCGTGSACCGAGTGACCGTGGTTCTGTGTGGTGTACCTGCCAGACCCCGGAA	840
ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu	280
GAAGCCACTCTTTGGAGGTGCGCTCTCTGGCAGGGCCACTCCACCCATCCGTGGC	900
GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly	300
CGCCAGCACACGGGGCCCCCATCCACATCGCGCCACCAGTCCCTGGACACCGCT	960
ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro	320

图 11F

TGTCCCCGGTGTACGGCGAGACCAAGCACTTCCTACTCCTCAGGGACAAGGAGCAG	1020
CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln	340
CTGGGGCCCTCCTCCTACTCAGCTCTCTGAGGCCAGCCTGACTGGGGCTGGGAGGCTC	1080
LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu	360
GTGGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGACTCCCCGAGGTGGCC	1140
ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro	380
CGCCTGCCCCAGCGCTACTGCCAAATGGGGCCCTGTTTCTGGAGCTGCTGGGAACCAC	1200
ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis	400
GGCAGTGGCCCTACGGGGTCTCCTCAAGACGCACTGCCCGCTGGGAGCTGGGGTACC	1260
AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr	420
CCAGCAGCCGGTGTCTGTGCCGGGAGAAGCCCGAGGGCTCTGTGGGGCCCCGAGGAG	1320
ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu	440
GAGGACACAGACCCCGCTGGCTGGTGCAGCTGCTCCGGCAGCAGCAGCCCTGGCAG	1380
GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln	460
GTGTACGGCTTCGTGGGGCTGCCTGGCCGGCTGGTGGCCCCAGGCCCTCTGGGGCTCC	1440
ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer	480
AGGCACAACGAACGCCGCTTCCTCAGSAACACCAAGAAGTTCATCTCCCTGGGAAGCAT	1500
ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis	500
GCCAAGTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGGGGGCTGGGCTTGGCTG	1560
AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu	520
CSCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGAGAGCACCGTCTGGTGAGGAGATC	1620
ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle	540
CTGGCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTGTCGAGCTGCTCAGGTCTTC	1680
LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe	560
TTTTATGTACGGAGACCACGTTTCAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTG	1740
PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal	580
TGGAGCAAGTTGCAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTGGGGAG	1800
TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu	600
CTGTGGGAAGCAGAGGTGAGGAGCATCGGGAAGCCAGGCCCGCCCTGACGTCCAGA	1860
LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg	620
CTCCGCTTCATCCCAAGCCTGACGGGCTGGGGCGATTGTGAACATGGACTACGTGCTG	1920
LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal	640

图 11G

GGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCA GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla	1980 660
CTGTTGAGCGTCTCAACTACGAGCGCGCGCGCCCGCCCTCCTGGGCGCCTCTGTG LeuPheSerValLeuAsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAlaSerVal	2040 680
CTGGGCTGACGATATCCACAGGGCTGGCGCACCTTCGTGCTGCGTGTGGGGCCAG LeuGlyLeuAspAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln	2100 700
GACCCGCGCCTGAGCTGTACTTTGTCAAGGTGATGTGACGGCGCGTACGACACCATC AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle	2160 720
CCCCAGACAGGCTCACGAGGTCATCGCGAGCATCAAAACCCGAACACGTAAGTGC ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys	2220 740
GTGCGTCCGATGCCGTGGTCCAGAGGCGGCCATGGGCAGTCCGCAAGGCCTCAAG ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys	2280 760
AGCCACGTCTCTACCTTGACAGCCTCCAGCCGTACATGGGACAGTTCGTGGCTCACCTG SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu	2340 780
CAGGACACCAGCCGCTGAGGATGCCGTGTCATCGAGCAGGCTCCTCCCTGAATGAG GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu	2400 800
GCCAGCAGTGGCCTTCGACGCTTCCCTACGCTTCATGTGCCACCAGCGTCCGATC AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle	2460 820
AGGGGCAAGTCTACGTCAGTCCAGGGATCCCGAGGGCTCCATCCTCTCCAGCTG ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu	2520 840
CTCTGACGCTGTGCTACGGCGACATGGAGAACAAGCTGTTGCGGGATTGGCGGGAC LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp	2580 860
GGGCTGCTCCTGCGTTTGGTGGATGATTTCTTGTGGTGACACCTCACCTCACCCAGCG GlyLeuLeuLeuArgLeuValAspAspPheLeuLeuValThrProHisLeuThrHisAla	2640 880
AAAACCTTCTCAGGACCTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGGTGAACCTG LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu	2700 900
CGGAGACAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCAGCGCTTTTGTT ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal	2760 920
CAGATGCGGCCCCAGGCTATTCCCTGGTGGGCTGCTGCTGGATACCCGACCCCTG GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuLeuAspThrArgThrLeu	2820 940
GAGGTGCAGAGCGACTACTCCAGCTATGCCCGACCTCCATCAGAGCCAGTCTCACCTTC GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe	2880 960

图 11H

AACCGCGCTTCAAGGCTGGGAGGAACATCGGTGCAAACTCTTTGGGCTTGGCGGCTG	2940
AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu	980
AAGTGCACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAAC	3000
LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn	1000
ATCTACAAGATCCTCCTGCTGCAGGCGTACAGTTTCACGCATGTGTGCTGCAGCTCCCA	3060
IleTyrLysIleLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro	1020
TTTCATCAGCAAGTTTGAAGAACCCACATTTTCTGCGGTCATCTGACACGGCC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCCTGCTACTCCATCCTGAAAGCCAAGAAGCAGGGATGTCGCTGGGGCCAAAGGGC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGCCGGCCCTCGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGTGACTCGACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ACGCAGCTGAGTCGGAAGCTCCCGGGACGACGCTGACTGCCCTGGAGGCCGACCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CCGGCACTGCCCTCAGACTTCAAGACCATCCTGGAC	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132

图 111

截短的蛋白质 3

ATGCCGCGCGCTCCCGGTGCCGAGCCGCTGCGCTCCCTGCTGGCAGCCACTACCGCGAGGTGCTGCCGTGGCCACGTTCTGT
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCAGGGCTGGGGCTGGTGCAGCGGGGACCCGGCGCTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGCACGGCCGCCCGCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCCGCCAGGTCTCTGCCTGAAGAGCTGGTGGCCGAGTCTGCAGAGCTGTGCCAGCGCGCCGGAAGAACGTGCTGGCTTCGGCTTCGCGCTGCTGGACGGGGCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTACCACCAGCTGCGCAGCTACCTGCCAACAGCTGACCGAGCCACTGCGGGGAGCGGGGCTGGGGCTGCTGCTGCGCGCTGGGGACGAGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGCTGCGCGCTCTTTGCTGCTGGCTCCGAGCTGCCCTACCAGTGTGGGGCGCGCTGTACCAGCTCGCGCTGCCACTCAGGCCGGGCCCGCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCGAGGGCTGGEATGCCAACGGGCTGGAAACATAGGCTCAGGAGGCGGGGTCCCTCGGGCTGCCAGCCCGGGTGGGAGGCGGGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCGTTGCCAAGAGGCCAGGGCTGGCGTCCCTGAGCCGGAGCCGACCCGTTGGCAGGGGTCCTGGGCCACCCGGGACGAGCGGTGGACCGAGTACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTCTGTGTGCTCACCTGCCAGACCCGCGAAGAAGCCACCTCTTTGAGGGTGGCTGCTGGCAGCGCCACTCCACCCATCCGTGGGGCCGACGACGCGGGGCCCGCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q K H A G P P

ATCCACATCGGGCCACAGTCCCTGGGACAGCCTTGTCCCGGTGACGCGAGACCAAGCACTTCTCTACTCCTCAGGGACAAAGGAGCAGTGGGGCCCTCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCAGGGCCAGCTGACTGGCGTCCGAGGCTGTTGGAGACCATCTTTCTGGTTCCAGGCCCTGGATGCCAGGCACTCCCGCAGGTTGCCCGCCTGCCAGCGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATCGGCCCCGTTTCTGGAGCTGCTTGGGACCCAGCGCAGTGCCTACGGGCTGCTCCTCAAGACGCACTGCCCGCTGGGAGTGGGTACCCAGCAGCGGCTGTGTGGCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGCGGCCCGGAGGAGGACACAGCCCGCTGCTGCTGAGTGTGCTCGCCAGCAGCAGCCCTGGCAGGTGTACGGCTCTGTGCGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCGCGGCTGGTGCCTCCAGGCTCTGGGGCTCCAGGCACAACGAGCCCTTCTCAGGAACACCAAGTTCATCTCCCTGGGAAGCATGCCAGCTCTCGTGCAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACSTGGAGATGACCTCGGGACTCGCTTGGCTGGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGAGAGCAGCGTCTGGGTGAGGAGATCTGGCCAAGTCTCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGTAGTCTGAGCTGCTCAGGCTTTCTTTTATGTACGGAGACCCAGTTTCAAAGAAGCAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTCGAAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

图 11J

AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTCGSAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCCGCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
CGGGCTGCGGCCGATTGTGAACATGGACTACGTGCTGGGAGCCAGAACGTTCCGACAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCAGTGTTCAGCGTCTCAACTACGA
G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
GCGGGCGCGGCCCGCCGCTCCTGGGCGCCTCTGTGCTGGGCTGGACATATCCACAGGGCCTGGGCGACCTTCGTGCTGCGTGTGCGGGCCAGGACCCGCGCCTGAGCTGTACTT
R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
TGTCAAGGTGGATGTGACGGGGCGTACGACACCATCCCCAGGACAGGCTCACGGAGTTCATCGGCAGCATCATCAAACCCAGAACAGCTACTGCGTCCGTCGATGCCGTGTC
V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
GAAGGCCCGCATGGGCACGTCCGCAAGGCCCTCAAGAGCCAGTCTCTACCTTGACAGACCTCCAGCCGTACATGGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCCGTGAAGGA
K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
TGCCGTGCTCATCGAGCAGAGCTCCTCCTGAATGAGGCCAGCAGTGGCCTCTTCGACGCTTCTCAGCCTTCATGTGCCACCAGCCGTCGCGCATCAGGGCAAGTCTACGTCCAGTG
A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
CCAGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTTCGCGGGATTGCGGGGACGGGCTGCTCCTGCGTTTGGTGA
Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
TGATTTCTTGTGAGACCTCACCTCACCCACGCGAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACCTTGGGAAGACAGTGGTGAACCTCC
D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
TGTAGAAGCAGGCGCCTGGGTGGCAGGCTTTTGTTCAGATGCCGGCCACGGCCTATTCCCCTGGTGGGCGCTGCTGCTGATACCCGGACCTGGAGGTGCAGAGGACTACTCCAG
V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S R

GTGAGCGCACCTGGCCGGAAGTGGACCTGTGCCGGCTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCTCCGTGTGGGGCAGGGGACTGCCAATCCAAAAGGGTCAGA
*
TGCCACAGGGTGGCCCTGTCATCTGCGGCTGAGCACAATGCATCTTCTGTGGGAGTGAGGGTGCCTCACACGGGAGCAGTTTTCTGTGCTATTTTGTAA...

图 11K

改变的 C-末端蛋白质

ATGCGGGGCTCCCGCTGCGGAGCGGTGCGCTCCCTGCTGCGCAGCCACTACCGCAGGTGCTGCCGCTGGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGCCCCAGGCTGGCGCTGGTGCAGCGGGGACCCGGCGGCTTCCGCGCGCTGGTGGCCAGTGCCTGCTGCGTGCCCTGGCAGCAGCGCGCCCCCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CGCCTCCTCCGCGAGGTGCTGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCGAGCGCGCGGAGAGAACGTGCTGCCCTTCGGCTTCGCGCTGGTGGACGGGCGCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTCACCAGCAGGTGCGCAGTACCTGCCAACACGTTACCGAGCAGTGGGGGAGCGGGGCTGGGGCTGCTGCTGCCCGCTGGCGAGCAGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGGCTGGCGCTCTTGTGCTGGTGGCTCCAGCTGGCGCTACCAGTGTGGGGCCGCGCTGTACCAGCTCGCGCTGCCACTCAGGCGCGCCCCCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCCGAAGGCTGTGGATGCGAACGGCTTGAACCATAGCTCAGGAGCGCGGGTCCCGCTGGCGTGCAGCCCGGGTGGAGGAGCGCGGGGCGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCGTGGCCAGAGGCGCAGGCTGGCGCTGCCCTGAGCGGAGCGGCGGCTTGGCGAGGGTCTGGGCCACCGCGGCGAGGCGGTGGACCGAGTGCAGC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTCTGTGTGTGCTCACCTGCCAGACCCCGGAGAGAGCCACTCTTGGAGGGTGGCTCTGTCGACGCGCCACTCCACCCATCGTGGCGCGCAGCACCAGCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGCCACCAGTCCCTGGGACAGGCTTGTCCCGGCTGACGCGAGCAGCAGCACTTCTCTACTCCTCAGGCGACAGGAGCAGTGGCGCTCCTTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCGCAGGCTGACTGGCGCTGGGAGGCTGCTGGAGCACCCTTTCTGGTTCCAGGCGCTGGATGCCAGGACTCCCGCAGGTTGCCCGCTGCCAGCGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATCGGCCCCGTTTCTGGAGCTGCTGGGAACCCAGCGCAGTGCCCTACGGGCTGCTCTCAAGCGCACTGCCCGCTGGGAGTGGGTACCCAGCAGCGGCTGTGTGCCCC
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGCGGCCCCGAGGAGGACACAGACCCCCGTGCGCTGGTGCAGTGTCTCCGCGAGCAGCAGCCCTGGCAGGTGTACGGCTTCGTGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCSCGCTGTGCCCGCAGGCTCTGGGCTCCAGGACACGSAAGCGGCTTCTCAGGAACACGAAAGTTCATCTCCTGGGAAGCATGCCAAGCTCTGCTGCAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GAGTGGAGATGAGGCTGGGACTGCGCTGGCTGGCAGGAGCCAGGCGTGGCTGTGTTCCGGCCGAGAGCACCGTCTGCTGAGGAGATCCTGGCCAAGTTCCTGCAGTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACCTGCTCAGTGTCTCAGGCTTTCTTTTATGTACCGAGACCCAGTTCAAAAGAACAGGCTCTTTTCTACCGSAAGAGTGTCTGGAGCAAGTTCACCAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGCTGAGCTGGGGAGCTGTGGGAGGAGGCTCAGGACGATCGGGAGCCAGGCCCGCTGCTGACGCTCCAGACTCCGCTTCATCCCAAGGCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

图 11L

CCGGCTGGGCGGATTGTGAACATGGACTACGTGTTGGGAGCCAGAACGTTCCGGCAGAGAAAAGAGGGCCGAGCGTCCACCTCGAGGGTGAAGGCACTGTTACGGCTGCTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCGGGCGCCCGGCTCCTGGGCGCCTCTGTGCTGGGCTGGACGATATCCACAGGGCCTGGCCACCTTCGTGCTGCGTGTGGGGCCAGGACCCCGGCTGAGCTGACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAGGTGGATGTGACGGGCGGTACGACACCATCCCCAGGACAGGCTCACGGAGGTATCCGCCAGCATCAAACCCAGAACAGTACTGCCTGGTGGTATGCCGTGGTCCA
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGCGCCCATGGGACGTCGCGAAGGCCTTCAAGAGCCAGGTCTCTACCTTGACAGACCTCCAGCCGTACATGGACAGTTCGTGGCTCACCTGCAGGACAGCCCGGCTGAGGSA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGCCGTCGTCATCGAGCAGAGCTCCTCCGTAATGAGGCCAGCAGTGGCCTCTTCGACGTCTTCCACGCTTCATGTGCCACCACCGGTGGCCATCAGGGGCAAGTCTACGTCAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGCTACGGCAGATGGAGAACAAGCTGTTTGGGGGATTGGGGGACGGGCTGCTCCTGGTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTCTTGTGGTGACACCTCACCTCACCCAGCGAAAACCTTCTCAGGACCCCTGGTCCGAGGTGTCCCTGAGTATGGCTGGTGGTGAACCTGGGGAAGACAGTGGTGAACCTCC
 D F L L V T P H L T H A K T F L R L T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGCAGGCGCCTGGGTGGCAGGGCTTTTGTTCAGATGCCGGCCACGGCCTATCCCGTGGTGGGCGCTGCTGGATACCCGSAACCTGGAGGTGCAGAGGACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S
 CTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGGGCTTCAAGGCTGGGAGSAACATCGCTCGCAAACCTTTGGGCTTTGGGGTGAAGTGCACAGCCTGTTCTGGA
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D
 TTTGAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCAGCATGTGTGCTGCAGCTCCCATTTTCATCAGCAAGTTTGAAGA
 L Q V K S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N
 CCGCACATTTTCTGGGCGTCATCTCTGACAGGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACCGCAGGATGTGGTGGGGCCAAGGGCCCGGCGGCTCTGCCCTCGA
 P T F F L R V I S D T A S L C Y S I L K A K N A E
 |
 CCGAAGA-AAACATTTCTGTGCTGACTCCTGCGGTGCTTGGGTC
 E E N I L V V T P A V L G S
 GGSACAGCCAGATGGAGCCACCCCGCAGACCSTGGGTGTGGGACGTTTCCGGTGTCTCTGGGAGGGAGTTGGGCTGGGCTGCTGACTCCTCAGCCTCTGTTTCCCGCAG
 G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L *

图 11M

缺少基元 A 的蛋白质

ATGCCGGCGCTCCCGCTGCCAGCCGCTGGCTCCCTGCTGCCAGCCACTACCGCGAGGTGCTGCCGCTGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGCGCCTGGGGCCAGGGCTGGCGCTGGTGCAGCGGGGACCGCGCGCTTTCGGCGCGTGGTGCAGTGCCTGGTGTGGTGGCTGGGACGACAGGGCCCGCCCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCCCGCAGTGTCTGCTGAAGAGCTGGTGGCCGAGTGTGCAGAGCTGTGCCAGCGCGCGGSAAGAAGTGTGTCCTTCCGCTTCCGCTGCTGGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTACACACAGGCTGGCAGCTACCTGCCACACAGGTGACGACGCACTGCGGGGAGCGGGGCTGGGGCTGCTGCTGCCCGCTGGCGACGAGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGCCGCTCTTTGTGCTGGTGGTCCAGCTGCCCTACAGGTGTGCCGCGCGCTGTACCAGCTGCGCGCTGCACTAGGCCCGCCCGCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAGGGCTCTGGATGGCAAGGGCTGGAACATAGCGTCAGGAGGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGAGGAGCGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCCGTTGCCAAGAGGCCAGCGTGGCGCTGCCCTGAGCGGAGCGCGCTGGCGTGGCAGGGTCTGGGCCACCGGGCAGGAGCGCTGGACCGAGTACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTGTACCTGCCAGACCGCGAAGAGCCACTTTTGGAGGTGGCTCTTGGCACGCGCACTCCACCCATCCGTGGGCGCGCAGCACCGCGGGCCCGC
G F C V Y S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGCCACCAGCTCCCTGGACACCGCTTGTCCCGGCTGACCGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAGTGGCGCTCCTTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGTAGGCGCAGCTGACTGGCGCTGGAGGCTCGTGGAGACCATTTTCTGGTTCCAGGCGCTGGATGCCAGGCACTCCCGCAGGTTGCCCGCGCTGCCCGCCTACTGCA
S L R P S L T G A R R L V E T I F L G S R P W W P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTTTCTGGAGCTGTTGGGAACACCGGAGTCCCGTACGGGCTGCTCCTCAAGAGCACTGCCGCTGGAGCTGGGTCAACCGAGCGCGGTGTGTGCGCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCGAGGCTGTGGCGGCCCCGAGGAGGAGACAGACCCCGTGGCTGGTGCAGCTGCTCCCGCAGCAGCAGCCCGTGGCAGGTGTACGCTTGTGGCGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGCTGCCCGCAGGCTTGGGGCTCCAGGCAACCAAGCGCTTCTCAGSAACCAAGAAGTTTCTCCCTGGGAAGCATGCCAAGCTCTCGCTGCAAGGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGCGTGGGGACTGCGCTGGCTGGCGAGSAGCCAGGGTGGCTGTGTCCGGCGCAGCAGCCGCTGCGTGAGGAGATCCTGGCAAGTCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATCAGTGTGTACGTGCTGAGGCTGCTCAGGCTTTCTTTTATGTACCGAGACCGTTCGAAAGAACAGGCTCTTTTCTACCGSAAGAGTGTCTGGAGCAAGTTCGAAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGTCAGCTGGGGAGCTGCGAAGCAGAGGTGAGGACGATCGGSAAGCAGGCGCGCGCTGCTGAGTCCAGACTCCGCTTCTCCCAAGGCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

图 11N

CGGGCTGCGGGCGATTGTGAACATGGACTACGTGGGAGCCAGAACGTTCCGACAGAAAAGAGGGCCGAGCGTCCACCTCGAGGTTGAAGGCAGTTCAGCGTCTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GGGGGCGGGCCCGCGGCTCCTGGGCGCTGTGTGGGGCTGGACGATATCCACAGGGCCCTGGCGACCTTCGTGCTGCGGTGCGGGCCAGGACCGCGCGCTGAGTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTC AAG V K GACAGGCTCAGGAGTTCATCGCCAGCATCAAAACCCAGAACGTAAGTGGTGGTGGTATGCGGTGGTCCA
 D R L T E V I A S I I K P O N T Y C V R R Y A V V Q
 GAAGGGCGCCATGGGACGTCGCAAGGCTTCAAGAGCCAGTCTCTACCTTGACAGACCTCCAGCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAGACCAGCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGCCGTCGTCATCGAGCAGAGTCTCCCTGAATGAGGCGAGGAGTGGCTCTTCCAGCTTCCTACGCTTCATGTGCCACACCGCGTGGCATCAGGGCAAGTCTACGTCGAGT
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGSAATCCCGAGGCTCCATCCTCTCCAGCGTGTGTGACGCTGTGTACGGGACATGGAGAACAAAGCTGTTTGGCGGSAATCCGGGGACGGGCTGCTCCTGGTGTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTGTGGTGCACCTCACCTCACCCAGGGAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGCTGGTGGTGAAGTTCGCGAAGACAGTGGTGAAGTCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGCAGGGCCCTGGGTGGCAGGCTTTTGTTCAGATGCCGGCCACGGCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGGACCTGGAGTGCAGAGCCACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S
 CTATGCCCGACCTCCATCAGAGCAGTCTCACCTTCAACCGGGCTTCAAGGCTGGGAGAACATGCGTCGCAAACTCTTGGGCTTGGGCTGAAGTGCACAGCTGTTCTGGA
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D
 TTTGAGGTGAACAGCCTCCAGAGGTGCACCAACATCTACAAGATCCTCCTGTGAGGGGTACAGGTTTCAAGCATGTGGTGGAGCTCCCATTTATCAGCAAGTTTGAAGAA
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N
 CCCCACATTTTCTGCGGTTCATCTGTACAGCGCTCCCTCTGCTACTCCATCCTGAAGGCCAAGACAGGATGTGGTGGGGCCAAAGGGCCCGCGCCCTGTGCCCTCCGA
 P T F F L R V I S D T A S L C Y S I L K A K N A G W S L G A K G A A G P L P S E
 GGGCTGAGTGGTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACACCGTGTGACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAGAGCTGAGTGGSAAGTCCC
 A V Q W L C H Q A F L L K L T R H R V T Y V P L L G S L R T A Q T Q L S R K L P
 GGGACGACGCTGACTGCGCTGGAGCGGAGCCAAACCGGCACTGCCCTCAGACTCAAGACCATCCTGGACTGATGGCCACCGCCACAGCCAGGCGGAGGACACACCAGCAGCC
 G T T L T A L E A A A N P A L P S D F K T I L D
 CTGTACCGCGGGCTACGTCCAGGGAGGGAGGGCCGACACCCAGCCCGCAAGGCTGGGAGTCTGAGGCTGAGTGTGTTGGCGAGGCTGCATGTCCGGCTGAAGGCT
 GAGTGTCCGGCTGAGGCTGAGGAGTGTCCAGCCAAGGCTGAGTGTCCAGCACCTGGCGTTCACCTCCCCACAGGCTGGCGCTGGCTCCACCCAGGGCCAGCTTTTCTCAC
 CAGGAGCCCGGCTTCCACTCCACATAGGAATAGTCCATCCCGAGATTGCGCATGTTCCAGCCCTCGCCCTGCCCTTTGCTTCCACCCCCACCATCCAGGTGGAGACCCTGAGAA
 GGACCTGGGAGCTCTGGAATTTGAGTGAACAAGGTGTGCCCTGTACACAGGCGAGGACCTGCACCTGATGGGGTCCCTGTGGTCAAAATGGGGGAGTCTGTGGGAGTAA
 AATACTGAATATAGAGTTTTTCAGTTTTGA

图 110

缺少基元 A 的截短的蛋白质

ATGCCCGCGCTCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCCAGCCACTACCGGAGGTGCTGCCGTGGCCAGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGACCCGGCGCTTCCGCGCTGGTGGCCAGTGCCTGGTGGCTGCCCTGGGAGGACGGGCCCCCGCGCG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCCCTCTCGCCAGGTGTCTGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGGAGCGCGCGGGAAGAAGTGTGCTGCTCGGCTTCGGCTTCGGCTGCTGAAGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGGCTTACCACAGCGTGGCAGCTACCTGCCAACACGGTCAACGCACTGGGGGAGCGGGGCTGGGGCTGCTGCTGCCCGCGTGGCGAGCAGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GGTGTTCACCTGCTGGCAGCTGGCGGCTCTTTGTGGTGGCTCCAGCTGGCCTACCAGTGTGGCGCGCGGCTGTACCAGTGGCGCTGCCACTCAGGCCGGGCCCCGGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAAGGGTCTGGGATGGACAGGGCTGGAAACATAGGCTCAGGAGGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGAGGAGCGGGGGGCGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTGTGCGTTGCCAAGAGGCCAGGCGTGGCGCTGCCCTGAGCGGAGCGGACCCGTTGGCAGGGTCTGGGCCCACCGGCGAGGACCGGTGGACCGAGTGGCC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTTCTGTGGTGTCACTGCCAGACCGCGGAGGAGGACCTCTTTGAGGCTGGGCTCTGTGGCAGCGCCACTCCACCCATCGGTGGCGCCAGCACCGGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACAGTCCCTGGGACAGCCCTGTCCCGGCTACGCCGAGCAAGCACTTCTCTACTCTCAGGGGACAAGGAGCAGTGGCGCCCTCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGGCCAGCTGACTGGCGCTGGAGGCTGGTGGAGACATCTTTCTGGTTCAGGCGCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCGAGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTTTCTGGAGCTGCTGGGACCGCGAGTGCCTACGGGTGCTCCTCAAGACCGACTGCCCGCTGGAGCTGGGTCAGCCGAGCAGCCGCTGTGTGCGCC
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCCAGGGCTCTGTGGCGGCCCGAGGAGGAGACAGACCCCGTGGCTGGTGCAGTGTCCCGCAGCACAGCAGCCCTGGCAGGTGTACGGCTTGTGGCGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGCCTCCAGGCTCTGGGCTCCAGGCACAACGACCGCTTCTCAGGAAACACCAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTGCTGCAGGAGT
L R R L V P P G L W G S R H N E R R F L R N T X K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGCGTGGGACTGGCTTGGCTGCGGAGGAGCCAGGGTGGCTGTGTTCCGGCCGACAGCAGCAGTGTGGTGGAGATCCTGGCAAGTTCCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACGTGCTGAGGCTGCTCAGGCTTTCTTTTATGTACCGGAGACCAGGTTTCAAGAACAGGCTCTTTTCTACCGGAGAGTGTGTGGAGCAAGTTCGAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTGAAGAGGTCAGCTGGGGAGCTGTGGAGCAGGTCAGGACGATCGGGAAGCCAGGCCCCCGCTGCTGACGTCCAGACTCCCTTCATCCCCAAGCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

图 11P

CGGGCTGCGGCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCCTGTTGAGCGTGCCTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCGGGCGCCCGGGCTCCTGGGGGCTCTGTGCTGGGCTGGACATATCCACAGGGCTGGCGACCTTCGTGCTGCGTGTGCGGGCCAGGACCCGCGGCTGAGCTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTC AAG GACAGGCTCAGGGAGTTCATCGCCAGCATCATCAAACCCAGAACAGTACTGCGTGCCTGGTATGCCGTGGTCCA
 V K D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGCCGCCATGGGCAGCTCCGCAAGGCTTCAAGAGCCAGCTCTACCTTGACAGACCTCCAGCGGTACATGCGACAGTTGCTGGCTCACCTGCAGGAGACCAGCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L O P Y M R Q F V A H L Q E T S P L R D
 TGCCGTGTCATCGAGCAGGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTCTTCGACGTCTTCTACGCTTCATGTGCCACCAGCCGCTGCGCATCAGGGBCAAGTCTACGTCCAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V V Q C
 CCAGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTGCGGGGATTGCGCGGACGGGCTGCTCCTGCGTTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTGTTGGTGACACCTCACCTCACCCACGGGAAAACCTTCTCAGGACGCTGGTCCGAGGTGCCTCGAGTATGGCTGCGTGGTGAACCTTGGGAAGACAGTGGTGAACCTCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGACGAGGCCCTGGGTGGCAGGCTTTTGTTCAGATGCCGGCCACGGCTATTCCCTGGTGGGCTGCTGCTGGATACCCGACCCCTGGAGGTGCAGAGGACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S R
 GTGAGGACCTGGCCGGAAGTGGAGCTGTGCCGGCTGGGGCAGTCTGTGACGGGCGTGGTCCACCTCTGCTTCCGTGTGGGGCAGGGCCTGCCAATCCCAAAGGGTCAGA
 TGCCACAGGGTCCCCTCGTCCATCTGGGGCTGAGCACAATGCATCTTCTGTGGGAGTGAGGCTGCCTCACACGGGAGCAGTTTTCTGTGCTATTTTGGTAA__

图 11Q

缺乏基元 A 和改变的 C-末端

ATGCCGCGCGCTCCCGGCTGCCGAGCGGTGGCTCCCTGCTGGCGAGCCACTACCGGAGGTGCTGCCGCTGGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCCTGGGGCCACAGGCTGGCGCTGGTGCAGCGGGGACCGGGGCTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCCTGGACGCACGGCCGCCCGCGCG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCGGCAGGTGCTGCTGCTGAAGGAGCTGGTGGCCGAGTGTGACAGAGGCTGTGGAGCGGGGGCAAGAAGTGTGCTGCTGCGCTTGGCTTGGCGTGTGACGGGGCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTACCACCGAGTGGCAGCTACCTGCCAACAGGCTACCGAGCCACTGGGGGGAGCGGGGCTGGGGCTGCTGCTGCCGCGTGGCGACGAGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGGCTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGGCCCTACCAGTGTGGGGGGCGCGGTACCAGCTCGGGCTGCCACTCAGCCCGGGCCCCGGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCCAAGGCGTGGGATGGGAACGGGCTGGAACCATAGCGTACGGGAGGGGCTCCCCCTGGGCTGCCAGCCCGGGTGGCAGGAGCCGGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCAGTCTGCGTGGCCAGAGGGCCAGGCGTGGCGCTGCCCTGAGCGGAGCGGACGCGGTTGGGAGGGTCTGGGCCACCCGGGAGGACGCGTGGACGAGTGAAGC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTCACTGCCAGACCGGCGAAGAAGCCACTCTTTGAGGCTGGCGTCTTGGCAGCGCCACTCCACCCATCGTGGGGCCAGCAGCAGCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGCCACAGCTCCCTGGGACAGCCTTGTCCCGGTTGACGGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAGTGGGGCCCTCCTCCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCAGGCGCCAGCTGACTGGCGCTGGGAGGCTGCTGGAGACCATCTTTGTGGTCCAGGCGCTGGATGCCAGGGACTCCCGCAGGTTGCCCGGCTGCCAGGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGCGCCTGTTTCTGGAGTGTGGGAACACCGGAGTGGCCCTACGGGCTGCTCCTCAAGAGCAGTGGCCGCTGGAGCTGGCGTGGCGGAGCAGCGGTTCTGTGCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGGGCCCCGAGGAGGACACAGACCCCGTGGCTGGTGCAGTGTCTCCGCGACAGCAGCCCGTGGCAGGTTGACGGCTCTGTGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCCCCAGGCTCTGGGGCTCCAGGCAACAAGCGGCTTCTCAGGAAACCAAGAAGTTCATCTCCCTGGGA .GCATGCCAAGCTCTGCTGCAGGAGT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAGATGAGCGTGGGGACTGGCTTGGCTGGCAGGAGCCAGGGTGGCTGTGTCCGGCGGACAGCAGCTGCTGGTGGAGAGATCCTGGCAAGTTCCTGCAGTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACGCTGAGGCTGCTCAGGCTTTCTTTATGTACGAGACACAGTTCAAAAGAAGGCTCTTTTCTACCGGAAGAGTGTGGAGCAAGTTCAAAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGTCAGCTGGGGAGCTGTGGAGGACAGGTCAGGAGCATGGGAAGCCAGGCCCCCTGCTGACGTCAGACTCCGCTTATCCCAAGCCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

图 11R

CGGGCTCGGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTACAGCTGCTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCGGGCCGCCCGGCTCCTGGGGCCCTGTGTGCTGGGCCCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCAGGACCCGCCGCTGAGCTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAGG GACAGGCTCACGGAGGTCATCGCCAGCATCATCAAACCCAGAACACGTA CTGCGTGGCTGGTATGCCGTGGTCCA
 V K D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGGCCCATGGGCAGCTCCGCAAGGCCCTTCAAGAGCCACGCTCTACCTTGACAGACCTCCAGCCGTACATGGGACAGTTGCTGGCTCACCTGCAGGAGACCAGCCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGCGTGTGTCATCGAGCAGAGCTCCTCCTGAATGAGGCCAGCAGTGGCCCTGTCGACGCTTCTCCTACGCTTCATGTGCCACCACGCGTGGCGCATCAGGGCAAGTCTACGTCGAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCCTGTGCTACGGGACATGGAGAACAGCTGTTTGGGGGATTCCGGGGGACGGGCTGCTCCTGCGTTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTTGTGGTGACACCTCACCTCACCCACGGAAAACCTTCTCAGGACCTGGTCCGAGTGTCCCTGAGTATGGCTGGTGGTGAACITGGCGAAGACAGTGGTGAACITCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGACAGGGCCCTGGGTGGCAGGGCTTTTGTTCAGATGCCGGCCACGGCCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGGACCTGGAGGTGCAGAGCCACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S
 CTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAGGCTGGGAGGAACATGGTCCGAAACTCTTTGGGGTCTTGGCGCTGAAGTGTACAGCCTGTTTCTGGA
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D
 TTTGCAGGTGAACAGCCTCCAGAGCGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGGGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTATCAGCAAGTTTGAAGAA
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N
 CCCCACATTTTCTGCGGTATCTGTGACAGGGCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAAGGATGTGCTGGGGCCAGGGCGCCGCGGCCCTCTGCGCTCCGA
 P T F F L R V I S D T A S L C Y S I L K A K N A E
 CCGAAGAAAACATTTCTGTGCTGACTCCTGCGGTGCTTGGGTG
 E E N I L V V T P A V L G S
 GGGACAGCCAGAGATGGAGCCACCCGAGACCGTGGGGTGTGGGACGTTTCCGGTGTCTCCTGGGAGGGGAGTTGGGCTGGGCCCTGTGACTCCTCAGCCTCTGTTTTCCCCAG
 G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L *

图 11S

N-末端截短的端粒酶 (版本 2)

ATGCCGCGGCTCCCGCTGCCAGCGTCCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCCTGGCCAGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGCGCTGGGGCCAGGGCTGGGGCTGGTGCAGCGGGGACCGGGCGCTTTCGGCGCGCTGGTGGCCAGTGCCTGGTGGTGGCTGGCCTGGGAGCAGCGCCGCCCCCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCCGGGTGGCGTCCGCTGGGTTGAGGGCGCCGGGGGAACAGCGACATGCGGAGAGCAGCGCAGCTCAGGGCGCTTCCCCCGAGSTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCTCCTCCCGCAGTGTCTGCTGAAGGAGTGGTGGCCGAGTGTGCACAGCTGTGCGAGCGGGCGGGAAGACGTGCTGGCTTCCGCTTCCGCTGCTGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTACGACACCGTGGCCAGCTACCTGCCAACAGGTGACCGAGCAGCTGCGGGGAGCGGGCGTGGGGCTGCTGCTCCGCGGTGGCGACGAGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGSCACCGTCCGCGCTCTTGTGCTGGTGGCTCCAGCTGCGCTACGAGTGTGGGGCCGCGCTGTACCAGCTGGCGCTGCCACTCAGGCCGGCCCCCGCG
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACCTAGTGGACCCGAAAGGCTCTGGATGCGAACGGGCTGGAACCATAGCGTCAGGGAGGGGGTCCCGCTGGGCTGCCAGCCCGGGTGGCAGGAGCGGGGAGCTG
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAGTCTGCGTTGCCAAGAGCCGAGGCTGGCGTGGCCCTGAGCCGAGGGACCGCGTTGGCAGGGTCTGGCCACCCGGGAGGAGCGGCTGACCGAGTGACCG
S R S L P L P K R P R R G A A P E P E R T P Y G Q G S W A H P G R T R G P S D R

TGTTTCTGTGTGTGTCACCTGCCAGACCGCGAAGAGCCACCTCTTTGAGGCTGGCTCTGCGCAGCCCACTCCACCCATCCGTGGGCGGCGAGCACCAGGGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATGCGGCCACCGCTCCCTGGGACAGGCTTGTCCCGGTGTAGCGGAGACCAAGCACTTCTCTACTCCTCAGCGCAAGGAGCAGTGGCGGCTCCTCCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTGTTGAGACCATCTTTCGGTCCAGGCGCTGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCGCCTACTGGA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTGTTCTGAGCTGCTTGGAAACCGGCGAGTGGCCGACCGGCTGCTCAAGCGCACTGCCGCTGGGAGCTGGGTACCCAGCAGCGGTTGCTGTCGGC
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGCGGCCCCGAGGAGGACACAGCCCGCTGCGCTGGTGCAGCTGCTCCGCGACACAGCAGCCCTGGCAGTGTACGGCTTCTGCGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGGCTGGTGGCCCCAGGCTCTGGGGCTCCAGGACAAAGACCGGCTTCTCAGGAACACCAAGAGTTATCTCCCTGGGAAGCATGCCAAGCTCTGCTGCAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGCTGCGGACTGCGCTTGGCTGCGCAGGAGCCAGGGTGGTGTGTTCCGGCGCAGAGCAGCTGTCGCTGAGGAGATCCTGGCAAGTCTCTGCACTGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

图 11T

GATGAGTGTGTACGTCGAGCTGCTCAGGCTCTTTCTTTATGTCAAGSAGACCAGCTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGSAGCAAGTTGCAAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AAT--NNN--GACAGTCACCAGGGGGTTGACCGCCGACTGGGCTCCCAAGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAGAGGCTGTGGCTGG
 CATGGGTGGACGTGGCCCGGGCATGGCTTCTGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCACTGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCGCTAGTCTGTGTCTG
 GCTGAGCAAGCCTCCTGAGGGGCTCTCTATTG.

图 11U

截短的蛋白质 1 (版本 2)

ATGCGCGCGCTCCCGCTGCCAGCCGTGGCTCCCTGCTGCCGAGCCACTACCGGAGGTGCTGCCGTGGCCADGTTGCTG
 M P R A P R C R A V R S L L R S H T R E V L P L A T F V
 CGGCGCTGGGGCCAGGGCTGGCGCTGGTGCAGCGGGGACCCGGCGCTTTCGGCGGCTGGTGGCCAGTGCCTGGTGTGGCTGCCCTGGGACGACGGGCGCCCGCCGCGC
 R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A
 GGCCTCCCGGGTGGCGTCCGGCTGGGTTGAGGGCGCGGGGGAACAGCGACATGCGGAGAGCAGCGAGGCACTCAGGGCGCTTCCCGCAGGTT
 G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
 A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
 P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V
 CCCTCCTTCCCGAGGTGCTGCTGCTGAAGSAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCAGCGCGCGGAAGAAGCTGCTGGCTTCCGCTTCCGCTGCTGGACGGGGCGG
 P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R
 CGGGGGCCCCCGAGGCTTACCACCAGGTCGCCAGCTACCTGCCAACAGGTCAGCCAGCACTGGGGGAGCGGGGCTGGGGCTGCTGCTGCCCGGCTGGCGACGACGT
 G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V
 GCTGGTTCACCTGCTGGCAGCTGGCGCTCTTGTGCTGGTGGTCCCACTGCCCTACCAGGTGTGCCGGCGCGGCTGTACCAGCTCGGCGCTGCCACTCAGGCCGGCCCGCGC
 L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P
 ACAGCTAGTGGACCCGAAGGCTCTGGATGCGAACGGGCTGSAACCATAGGTCAGGGAGCGGGGCTCCCGTGGGCTGCCAGCCCGGTTGCGAGGCGCGGGGAGTGG
 H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A
 CAGCCGAGTCTGCGTTGCCAAGAGGCCAGGCTGGCGTGGCGTGCAGCGGAGCGGCGGCTGGGCGAGGCTTGGGCGCACCGGGGAGGCGGAGTGGACCG
 S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R
 TGGTTCTGTGGTGTCACTGCCAGACCCGCGAAGAAGCCACCTCTTGGAGGTTGGCTCTGTGCACCGCCACTCCACCCATCGTGGGCGCGCAGCACACCGGGCCCCCG
 G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

图 11V

ATGCCGCGCGCTCCCGCTGCCAGCCGTGCGCTCCCTGCTGCCAGCCACTACCGCAGGTGCTGCCGCTGGCCAGTTCGTG
 M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCCTGGGGCCCAGGGCTGGCGGCTGGTGACGCGGGGACCCGGCGGCTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGCGTCCCTGGGACGCACGGCGGCCCGCGCGC
 R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGTCCGCTCCGCTGGGTTGAGGGGGCCGGGGGAACCAGCGACATCCGGAGCAGCGCAGGGACTCAGGGCGCTTCCCGCCAGGTG
 G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
 A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
 P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCCTCCTCCGCCAGGTGCTGCTGAAGGAGCTGGTGGCCGAGTGTGACAGGCTGTGCGAGCCGGCGCAAGAACCTGCTGGCCTTCGGCTTCGCGCTGCTGGACGGGGCCCG
 P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTACCACCAGCGTGGCAGCTACCTGCCAACACGGTGACCGCAGCTGCCGGGGAGCGGGCGTGGGGCTGCTGCTGCCCGCGTGGCCAGCAGT
 G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGCCGCTCTTTGTGCTGGTGGCTCCAGCTGCCCTACCAGGTGTGGGGCGCGGCTGTACCAGCTCGGGCTGCCACTCAGGCCCGGGCCCCGGC
 L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCGAAGGCGTCTGGGATGCGAAGGGCCTGGAACCATAGCGTCAGGAGGGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGCAGGAGCGGGGGCAGTGC
 H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCCTGCCAAGAGCCCGAGGCTGGCGCTGCCCTGAGCCGGAGCGGACGCCCTGGGGCAGGGTCTGGGCCACCCGGGCGAGCAGCGTGGACCGAGTGCAGC
 S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTTCTGTGGTGTACCTGCCAGACCCCGGAAGAGCCACCTCTTTGGAGGCTGGGCTCTCTGGCAGCGCCACTCCACCCATCCGTGGGCGCCAGCACCAGCGGGCCCCC
 G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGGACAGCCTTGTCCCGGCTGTACGCCAGACCAGCACTTCTCTACTCCTCAGGGCACAAGGAGCAGTGGCGCCTCCTTCTACTCAG
 S T S R P P R P W D T P C P P V Y A E T X H F L Y S S G D K E Q L R P S F L L S

图 11W

截短的蛋白质 2 (版本 2)

ATGCCGGCGCTCCCGCTGCCAGCCGCTGGCTCCCTGCTGCGCAGCCACTACCGGAGGTGCTGCCGTGCCACGTTCTGTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGCCCCAGGGCTGGCGCTGGTGCAGCGGGGACCCGGCGCTTCCGCGGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGCACGGCCGCCCCCGCGC
R R L G P Q G W R L V O R G D P A A F R A L V A Q C L V C V P W : D A R P P P A A

GGCTCCCGGGTGGCGCTCCGCTGGGTTGAGGGGGCCGGGGGAACCGCAGCATGCCGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCGCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCTCCTTCCCGCAGGTGCTCCTGCTGAAGGAGCTGGGCGGAGTGCAGAGGCTGTGCCAGCGCGCGGSAAGAAGCTGCTGGCTTCGGCTTCGCGCTGCTGGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTCCACCACGAGTGGCAGCTACCTGCCAACAGGTCACCGCAGCACTGCCGGGAGCGGGGGTGGGGCTGCTGCTGCCCGCTGGGACACGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGCTGGCGCTTTTGTGCTGGTGGCTCCAGCTGCCCTACAGGTGTGCCGGCGCGCTGTACCAGCTGGCGCTGCCACTCAGGCGCGCCCCCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAAGGCTCTGGGATGCGACGGGCTGGAACATAGGTCAGGGAGCGGGTCCCGCTGGGCTGCCAGCCCGGTGGAGGCGGGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCAGTGTCCGTTGCCAAGAGCCAGGCTGGGCTGCCCTGAGCGGAGGGACGCCGTTGGGAGGGTCTGGGCCACCGGGCAGGACGCTGGAACGAGTGCAGC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTTGTGTGGTGCACCTGCCAGACCCGGGAAGAGGCACCTCTTGGAGGTCGCGCTCTGCGCAGCGCCACTCCACCCATCCGTGGGCGCCAGCACCAGCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGGACAGGCTTGTCCCGGCTGACCGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAGCTGGCGCCCTCCTCCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGCCTCTTTCTGGGTTCCAGGCGCTGGATGCCAGGACTCCCCGAGGTTGCCCGCTGCCAGGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGCGCCCTGTTCTGAGCTGCTGGGAACACGGCAGTGCCTACGGGCTCTCAAGACGCACTGCCCGCTGGGAGCTGGGTCACCCAGCAGCGGTGTGTGCCCC
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCGGCTCTGTGGCGCCCCGAGGAGGACACAGACCCCGTGGCTGGTGCAGCTGCTCCGCGACAGCAGCCCGCTGGCAGGTGTACGGCTTGTGGCGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCGCGGCTGGTGGCCCCAGGCTCTGGGCTCCAGGCACAACGACCGCTTCTCAGGAACCAAGAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCCTGCGAGGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

图 11X

GACGTGAAGATGAGCGTCCGGGACTGCGCTTGGCTGGCAGGAGCCAGGGTTGGCTGTGTTCCGGCCGACAGGACCCTGCGGTGAGGAGATCCTGGCCAAGTTCCTGCACTGGCT
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I - L A K F L H W L

GATGAGTGTACGTGTCGAGCTGCTCAGGCTTTCTTTATGTCACGGAGACCAGTTTCAAAGAACAGGCTCTTTTCTACGGGAAGAGTGTCTGGAGCAAGTTCGCAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGCGGAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGA
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

CGGCTGCGGCCGATTGTGAACATGACTACGTGTTGGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGTGAAGCCACTGTTGAGCGTCTCAACTAGGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

GCGGGCCGCGCCCGCCGCGCTCCTGGGCGCTCTGTGCTGGGCGCTGACGATATCCACAGGGCTGGCCACCTTCGTGCTGGGTGCGGGCCAGGACCCGCGCCTGAGGTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAAGTGGATGTGACGGGCGGTACGACACCATCCCCAGGACAGGCTCAGCGAGGTGATCGCCAGCATCATCAAACCCAGAACGTAAGTGGGTGGGTGGTATGCGGTGGTCCA
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q

GAAGCCCGCCATGGGCAGTCCGCAAGGCCCTCAAGAGCCAC
 K A A H G H V R K A F K S H

GTCCTACGTCCAGTG
 V L R P V

CCAGGGATCCCGCAGGGCTCCATCCTCTCCAGCTGCTGTGAGCCTGTGCTACGGCCACATGGAGAACAAGCTGTTTCCGGGGATTCCGGGGACGGGCTGCTCCTGCGTTTGGTGA
 P G D P A G L H P L H A A L Q P V L R R H G E Q A V C G D S A G R A A P A F V G

TGATTTCTTGTGGTACACCTCACCTCACCCAGCGGAAAACCTTCTCAGGACCTGTCGAGGTGTCCTGAGTATGGCTGGGTGGTGAAGTTCGGGAAGACAGTGGTGAAGTCCC

图 11Y

对比蛋白质（版本 2）

ATGCCGGCGCTCCCGCTGCCGAGCCGTGCCGCTCCCTGCTGGCGAGCCACTACC3CGAG	60
MetProArgAlaProArgCysArgAlaValArgSerLeuLeuArgSerHisTyrArgGlu	20
GTGCTGCCGCTGGCCAGTTGCTGGGGCCCTGGGGCCCGAGGGCTGGGGCTGGTGCA	120
ValLeuProLeuAlaThrPheValArgArgLeuGlyProGlnGlyTrpArgLeuValGln	40
CGGGGGACCCGGCGGCTTCCGGCGGTGGTGGCCAGTGCCTGGTGTGGTGGCCCTGG	180
ArgGlyAspProAlaAlaPheArgAlaLeuValAlaGlnCysLeuValCysValProTrp	60
GACGCACGGCCCGCCCGCCCGCCCTCCCTCCGCCAGGTG	
AspAlaArgProProProAlaAlaProSerPheArgGlnVal	
GGCCTCCCGGGTCCGGTCCGGTGG3TTGAGGGCGCCGGGGGAACCGCCACATCGGGAGAGCAGCCAGGCGACTCAGGGCGCTCCCGCCAGGTG	
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R	
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G	
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V	
TCCTGCCTGAAGGAGCTG	240
SerCysLeuLysGluLeu	80
GTGGCCCGAGTGTGCAGAGGCTGTGGAGCCGCGCGAAGAAGCTGTGGCCTTCGGC	300
ValAlaArgValLeuGlnArgLeuCysGluArgGlyAlaLysAsnValLeuAlaPheGly	100
TTCGGCTGCTGGACGGGCCCCGGGGCCCCCGAGGCTTCACCACCAGCGTGGC	360
PheAlaLeuLeuAspGlyAlaArgGlyGlyProProGluAlaPheThrThrSerValArg	120
AGCTACCTGCCAACCGGTGACCGAGCCACTGGGGGAGGGGGCTGGGGCTGGTG	420
SerTyrLeuProAsnThrValThrAspAlaLeuArgGlySerGlyAlaTrpGlyLeuLeu	140
TTGGCCGGGTGGCCAGCAGTGTCTGCTGCTGCTGGCAGGCTGGCGCTCTTTGTG	480
LeuArgArgValGlyAspAspValLeuValHisLeuLeuAlaArgCysAlaLeuPheVal	160
CTGGTGGCTCCAGCTGGCCTACCAGGTGTGGGGCCGCGCTGTACCAGCTCGGCGCT	540
LeuValAlaProSerCysAlaTyrGlnValCysGlyProProLeuTyrGlnLeuGlyAla	180
GCCACTCAGGCCCGCCCGCCACCGCTAGTGGACCCCGAAGGGCTCTGGATGGGA	600
AlaThrGlnAlaArgProProProHisAlaSerGlyProArgArgArgLeuGlyCysGlu	200
CGGGCTGGAAACCATAGCGTCAGGGAGCCGGGTCCCGCTGGCCCTGCCAGCCCGGGT	660
ArgAlaTrpAsnHisSerValArgGluAlaGlyValProLeuGlyLeuProAlaProGly	220
GCGAGAGGCCCGGGCAGTGCAGCCGAAGTCTGCGGTTGCCAAGAGGCCAGGGGT	720
AlaArgArgArgGlyGlySerAlaSerArgSerLeuProLeuProLysArgProArgArg	240
GGCGCTGCCCTGAGCCGAGCGGACGCCCGTTGGGAGGGCTCTGGGCCACCCGGGC	780
GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly	260

图 11Z

AGGACCGCGTGGACCGAGTGACCCTGGTTTCTGTGTGGTGCACCTGCCAGACCCGCGAA	840
ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu	280
GAAGCCACCTCTTTGGAGGTGCGCTCTGTGCACGCGCCACTCCACCCATCCGTGGGC	900
GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly	300
CGCCAGCACGACGCGGGCCCCCATCCACATCGCGGCCACCGTCCCTGGCACAGCCT	960
ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro	320
TGTCGCCCGGTGTACGCGAGACCAAGCACTTCCCTACTCCTCAGGGACAAGGAGGAG	1020
CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln	340
CTGCGCCCTCCTTCCCTACTCAGCTCTCTGAGGCCAGCCGACTGGCGCTGGGAGGCTC	1080
LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu	360
GTGGAGCCATCTTTCTGGTTCCAGGCCCTGGATGCCAGGCACTCCCGCAGGTTGCCC	1140
ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro	380
CGCCTGCCCGACGCGCTACTGGCAATCGCGCCCTGTTTCTGGAGCTGCTGGGAACCAC	1200
ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis	400
GCGCAGTGCCCTACGGGGTCTCCTCAAGACGCACTGCCCGCTGGGAGCTGCCGTACC	1260
AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr	420
CCAGCAGCGGTGTCTGTGCCGGGAGAAGCCCGAGGCTCTGTGGGGCCCCGAGGAG	1320
ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu	440
GAGCACAGACCCCGTCCCTGGTGCAGTGTCTCCCGCAGCAGCAGCCCTGGCAG	1380
GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln	460
GTGTACGGTTCTGTGGGGCTGCCCTGCCCGGCTGGTGGCCCCAGCCCTCTGGGGCTCC	1440
ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer	480
AGGCACAACGAACGCGCTTCCCTCAGGAACACCAAGATTTCATCTCCCTGGGAAGCAT	1500
ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis	500
GCCAGCTCTCGCTGCAGGAGCTGACGTGGAGATCAGCGTGGGGGCTGGCTTGGCTG	1560
AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu	520
CGCAGGAGCCAGGGTTGGCTGTGTTCCGGCCGACAGCACCGCTCGGTGAGGAGATC	1620
ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle	540
CTGGCAAGTTCCTGCAGTGGCTGATGAGTGTGTACGTCTCGAGCTGCTCAGGTCTTTC	1680
LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe	560
TTTTATGTCAGGAGACCAGCTTTCAAAAGACAGGCTCTTTTTCTACCGGAAGGTGTC	1740
PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal	580

图 11AA

TGBAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGCTCGGGAG	1800
TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu	600
CTGTGGAAAGCAGAGGTACGGCAGCATCGGAAGCCAGCCCGCCCTGCTGACGTCCAGA	1860
LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg	620
CTCGGCTTCATCCCAAGCCTGAGGGGCTGGGGCCGATTGTGAACATGGAAGTACGTCTGTG	1920
LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal	640
GGAGCCACAACGTTCCGCAGAGAAAAGAGGGCCGAGCGCTCACCTCGAGGGTGAAGCA	1980
GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla	660
CTGTTGAGCGTCAACTACGAGCGGGCCGGGCCCGCCCGCTCCTGGGCCCTCTGTG	2040
LeuPheSerValLeuAsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAlaSerVal	680
CTGGGCTGGACGATATCCACAGGGCTGGCCGACCTTCGTGCTGCCTGTGGGGCCCGAG	2100
LeuGlyLeuAspAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln	700
GACCCCGCCCTGAGCTGTACTTTGTCAAGGTGSATGTGACGGGCGGTACGACACCATC	2160
AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle	720
CCCCAGSACAGGCTCAGGGAGTCAATCCGACATCATCAACCCABAACAGTACTGC	2220
ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys	740
GTGCGTCGATGCGTGGTCCAGAAGGGCCCATGGCCAGTCCGCAAGGCTTCAAG	2280
ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys	760
AGCCACGTCTCTACCTTGACAGACCTCCAGCGTACATGGCAGTTCGTGGCTCACCTG	2340
SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu	780
CAGGAGACCAGCCCGCTGAGGGATGCCGTGGTATCGAGCAGAGCTCCTCCCTGAATGAG	2400
GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu	800
GCCAGGAGTGGCTCTTCSAGCTCTCCTACGGTTCATGTGCCACCAGCCCGTGGCCATC	2460
AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle	820
AGGGCAAGTCTACGTCCAGTGCCAGGGATCCCGCAGGGCTCCATCCTCACCAGCTG	2520
ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu	840
CTCTGCAGCCTGTGCTACGGGACATGSAGAACAAGCTGTTGGGGGATTGGGGGGAC	2580
LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp	860
GGGCTGCTCCTCGCTTGGTGGATGATTTCTGTTGGTGACACCTCACCTCACCCAGCG	2640
GlyLeuLeuLeuArgLeuValAspAspPheLeuLeuValThrProHisLeuThrHisAla	880
AAAACCTTCCTCAGSACCTGGTCCAGGTGTCCCTGAGTATGCTGCGTGGTGAACCTG	2700
LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu	900

图 11AB

CGGAAGACAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCAGGGCTTTTGT	2760
ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal	920
CAGATGCCGGCCACGGCCTATCCCTGGTCCGGCCTGCTGCTGGATACCCGGACCCTG	2820
GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuLeuAspThrArgThrLeu	940
GAGGTGCAGAGGACTACTCCAGCTATGCCCGACCTCCATCAGAGCCAGTCTCACCTC	2880
GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe	960
AACCCGGCTTCAAGGCTGGGAGGAACATGGCTCGCAAACCTTTGGGGTCTTGGGGCTG	2940
AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu	980
AAGTGTACAGCCTGTTTCTGSATTTGCAGGTGAACAGCCTCCAGCGGTGTGCCAAC	3000
LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn	1000
ATCTACAAGATCCTCTGCTGCAGGGTACAGGTTTCACGCATGTGTGCTGCAGCTCCA	3060
IleTyrLysIleLeuLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro	1020
TTTCATCAGCAAGTTTGAAGAAGCCACATTTTTCTGCGCGTCATCTCTGACAGGGC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCCTCTGCTACTCCATCCTGAAAGCCAAGAAGCAGGGATGTCGCTGGGGCCAAGGGC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGGCCGGCCTCTGCCCTCCGAGGCGGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGCTGACTCGACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ACGCAGGTGAGTGGGAAGCTCCCGGGACGACGCTGACTGCCCTGGAGGCCGACGCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CCGGCACTGCCCTCAGACTTCAAGACCATCCTGGAC	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132

图 11AC

截短的蛋白质 3 (版本 2)

ATGCCGGGGCTCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCCGAGCCACTACCGCAGGTGCTGCCCTGCCACGTTCTGT
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCCTCGGGCCAGGGCTGGCGCTGGTGCAGCGGGGACCGCGCGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGAGCAGCGCGCCCGCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGTCCGGCTCCGGCTGGGGTTGAGGGCGCCGGGGGAACCGGACATGCGSAGAGCGCCAGGGACTCAGGGCGCTTCCCGCGCAGGTG
G L P G V G V R L G L R A A G G N O R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCTCCTCCGCCAGGTGTCTGCTGAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGGAGCGCGGCGAAGAACGTCTGCCCTCCGGCTTCCGGCTGCTGGACGGGCGCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTACCACCAGGCTGGCAGGTACCTGCCAACAGGTTGACCGAGCACTCGGGGAGCGGGGCGTGGGGCTGCTGCTGCCCGCGTGGCGACGAGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGCGCTCTTGTGCTGGTGGCTCCAGCTGGCGCTACCAGGTGTGGGGCGCGGCTGTACCAGCTCGGCGCTGCCACTCAGGCCCGCCCGCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAAAGCGCTGTGGATGGCAAGGGCTGGAAACATAGCGTACGGAGGGCGGGGTCCCGCTGGGCTGCCAGCCCGGGTGGAGGAGGGCGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCAAAGTCTGCCGTTGCCAAGAGGCCAGGCGTGGCGCTGCCCTGAGCGGAGCGGACGCGCTTGGCAGGGTCTCGGGCCACCGGGCAGGACCGCTGGACCGAGTGACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTCTGTGTGTGTGCTGCTGCCAGACCGCGAAGAGCCACTCTTGGAGGTGGCTCTCTGGCAGCGCCACTCCACCCATCGTGGGGCCAGCAGCAGCGGGCGCGCG
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGGACAGGCTTGTCCCGGGTGTAGCGGAGACCAAGCACTTCTCTACTCCTCAGGGGACAAGGAGCAGCTGGCGCCCTCCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGTCAGGCCAGGCTGACTGGCGCTGGAGGCTGTGGAGACCATCTTCTGGGTTCCAGGCGCTGGATGCCAGGACTCCCGCAGGTGGCGCGCTGCCAGCGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTTTCTGGAGCTGCTGGAAACCACGCGAGTCCCGTACGGGTTGCTCTCAAGAGCACTGCCCGCTGGGAGCTGGGTCACCCAGCAGCGGTTGTGTGCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGCGCGCGGAGGAGGAGACACAGCCCGTGGCTGGTGGAGCTGCTCGGCCAGCACAGCAGCCCTGGCAGGTGTACGGCTTGTGGGGCGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGGGCTGGTCCCGCAGGCTCTGGGCTCCAGGCAACCAACCGCTTCTCAGGAACACCAAGATTCACTCCCTGGGAAGCATGCCAAGCTCTGCTGCAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T X K F I S L G K H A K L S L Q E L

图 11AD

GACGTGGAAGATGAGCGTGGGGACTGGCCTTGGCTGGCCAGGAGCCAGGGGTGGCTGTGTTCCGGCCGACAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTTCCTGCACTGGCT
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L
 GATGAGTGTGTAGCTGCTGAGTGTCTGAGTCTTTTATGTACCGAGACCAGCTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTGGAGCAAGTTCCAAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AATCAGACAGCACTTGAAGAGGGTGCAGCTGGGGAGCTGTGGAAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGA
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
 CGGGCTGGGGCGATTGTGAACATGGAAGTACGTGGGAGCCAGAACGTTCCCGAGAGAAAAGAGGGCCGAGGCTCTCACCTCGAGGGTGAAGGCACTGTTGAGCGTGTCAACTACCA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCCGGCCCGCCCGGCTCCTGGGGCCTCTGTGTGGGCTGGACGATCCACAGGGCCTGGGGCACCTCGTGTGCGGTGTGGGGCCAGGACCCCGCCCTGAGCTGTACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAAGTGTGATGACGGGGCGTACGACACCATCCCGCAGGACAGGCTCAGCGAGGTCATCGCCAGCATCAAAACCCAGAACAGCTACTGCGTGGCTCGGTATCCCGTGGTCCA
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGCCCGCATGGGCACGTCGCAAGGCCCTCAAGAGCCAGCTCTACCTTGACAGACCTCCAGCGGTACATGGCAGAGTTCGTGGCTCACCTGCAGGAGACCCAGCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGGCTGCTCATGAGGAGAGGCTCCTCCGTAATGAGGCCAGGAGTGGCCTCTCGAGCTTCTCAGGCTTCATGTGCCACCACCGGTGGCCATCAGGGCAAGTCCCTACGTCAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGGATCCCGCAGGGCTCCATCCTCTCCAGGCTGCTGCGAGCCTGTGCTAGGGCAGATGGAGAACAAGCTGTTTGGGGGATTCGGGGGACGGGCTGCTCCTGCTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTCTGTGTGGTACACCTCACCTCAGCCAGCCGAAAGCCTTCTCAGGACCTGGTGGAGGTGCCCTGAGTATGGCTCCGTGGTGAAGTTCGGGAAGACAGTGGTGAAGTCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGCAGGCGCTGGGTGGCAGGGCTTTTGTTCAGATGCCGGCCACGGCCTATCCCTGGTGGGGCTGCTGCTGATACCCGACCCCTGGAGGTGCAGAGCCACTACTCCAG
 V E D E A L G G T A F V Q W P A H G L F P W C G L L L D T R T L E V Q S D Y S R
 GTGAGCCACCTGGCCGGAAGTGGAGCCTGTGCCCGGCTGGGACAGTGTGCTGCAGGGCCGTTCCGTCCACCTCTGCTTCCGTGTGGGGCAGGGCACTGCCAATCCCAAAGGTCAGA
 *
 TCCACAGGGTGGCCCTGTCCTCATCTGGGGCTGAGCAGAAATGCATCTTTCTGTGGAGTGAGGGTGCCTCAGAACGGGAGCAGTTTTCTGTGCTATTTTGGTAA_

图 11AE

改变的 C-端蛋白质 (版本 2)

ATGCCGCGCGCTCCCGCTGCCAGCCGCTGCGCTCCCTGCTGCGCAGCCACTACCGGAGGTGCTGCCGCTGGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGCGCTGGGGCCCGAGGCTGGGGCTGGTGCAGCGGGGACCGGGGCTTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGACCGCCCGCCCGCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCTCCCGGGGTCCGGCTCCGGCTGGGTTGAGGGCGCGCGGGGSAACCAGCGACATGCCGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCGCCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P O V

CCCCCTCCCGCAGGTGTCTGCTGAAAGGAGCTGGTGGCCCGAGTCTGCAGAGCTGTGGAGCCGGCGCAAGAAGTGTGCTGCGCTTCCGCTTCCGCTGCTGACCGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTACCACAGCGTGGCGAGCTACCTGCCCCAACCGGTGACCGCAGCAGTGGGGGAGCGGGCGTGGGGCTGCTGTGCGCGGCTGGCGACGACGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTCACTGCTGGCAGGCTCCGCGCTTTGTGCTGGTGGCTCCAGCTGGCCCTACCAGGTGTGGGGCGCGGCTGTACCAGCTCGCGCTGCCACTCAGGCCCGCCCGCCCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGACCCCGAAGCGCTCTGGGATGGGAACGGGCTGGAAGCATAGCGTCAGGAGGGCGGGGTCCCGCTGGGCTGCCAGCCCGGGTGGAGGAGCGGGGGCAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCCGTTGCCAAGAGCCCGAGGCTGGGCTGCCCTGAGCCGAGCGGACGCCCGTTGGGAGGGTCTGGGCCACCCGGCAGGACCGGTGAGCCGAGTGACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTTGTCACCTGCCAGACCCCGCAAGAGCCACTCTTTGGAGGTTGGCTCTCTGTCAGCGCCACTCCACCCATCCGTGGGCGGTCAGCAGCGGGCCCGC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATGCACATCGGGCCACGCTCCCTGGGACAGGCTTGTCCCGGGTGTAGCCGAGACCAAGCACTTCTCTACTCTCAGGGACAAGGAGCAGCTGGGGCCCTCCTCTACTAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D X E Q L R P S F L L S

CTGCTGAGGCCAGCCTGACTGGGCTCGGAGGCTGGTGGAGACCATCTTTCTGGGTTCCAGGCCCTGATGGCAGGACTCCCGCAGSTTGCCCGGCTGCCAGCCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGCGGCCCTGTTTCTGAGGTGCTTGGGAAACCAGCGGAGTGGCCCTAGGGGCTGCTCTCAAGCGGACTGCCCGCTGGGAGCTGGGTCACCCCAAGAGCGGCTGTGTGCGCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCGAGGCTGTGTGGCGGCCCGGAGGAGGAGACAGACCCCGCTGGCTGGTGGAGCTGCGGCGAGCAGCAGCCCTGGCAGGTGTACGGTTCGTTGGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCCCGGCTGGTGGCCCGAGGCTGTGGGCTCCAGGACAAGCAAGCGGCTTCTCAGGACCAAGAGTTCACTCTCCCTGGGAAAGCATGCCAAGCTCTGCTGCGAGGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGGCTGGGGACTGCCCTGGCTGGCAGGAGCCAGGGTGGCTGGTGTTCGGCCCGCAGAGCAGCGTCTGGTGGAGATCTGGCCAGTTCCTGCACTGGCT
T W X M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

图 11AF

GACGTGGAAGATGAGCGTGCAGGACTGCGCTGGCTGCGCAGGAGCCAGCGTGGCTGTGTTCCGCGCCAGAGCACCCTGCGGTGAGGAGATCCTGGCCAGTTCTGCACTGGCT
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L
 GATGAGTGTGTACCTGCTGAGCTGCTCAGGCTTTTCTTTATGTACGGAGACCAGCTTTCAAAGAACAGGCTCTTTTCTACGGGAAGAGTGTCTGAGCAAGTTGCAAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGAGCTGTGCGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGA
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
 CCGGCTCCGCGGATTGTGAACATGACTAGTGTGCGAGCCAGAACGTTCCGAGAGAAAAGAGGGCCGAGGCTCTCACCTCGAGGAGAGGCACTGTTACGCTGCTCAACTAGCA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGCGCGGCGCCCGGCTCCTGGCGCTCTGTGCTGGGCTGAGCATATCCACAGGGCTGGCGCACCTTGTGCTGCGGTGCGCGCCAGGACCCGCGCTGAGCTGACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAGGTGGATGTGACGGGGGTAGGACACCATCCCGCAGGACAGGCTCAGCGAGGTCATGCCAGCATCAAAACCCAGAACAGCTACTGCGGTGCGGTATGCGGTGGTCCA
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q H T Y C V R R Y A V V Q
 GAAGCGCGCCATGGGACGTCGCGAAGCCCTCAAGAGCCAGTCTCTACCTTGACAGCCTCCAGCCGACATGGACAGTTCGTGGCTCACGTGCAAGGACACCAGCCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGCGCTGCTCATGAGCAGAGCTCCTCCGTAATGAGGCGCAGTGGGCTCTTCGACGCTTCTCTACGCTTCAATGTCGCCACCAGCCGCTGGCGCATCAGGGCAAGTCTACGCTCCAGT
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGATCCCGCAGGCTCCATCCTCTCCACGCTGCTGTCAGGCTGTCTACGGGACATGGAGAACAGCTGTTGCGGGATTGCGGGGACGGGCTCCTCCTGCTTGGTGGG
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTGTGGTGACACCTCACCTCACCCAGGGAAGCCTTCTCAGGACCTGCTCCGAGGTGCTCCCTGAGTATGGCTGCGGTGAGTTCGGGAAGACAGTGGTGAACCTCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGACGAGCCCTGGGTGGCAGGCTTTTGTTCAGATGCCGCGCCAGGCTATTCCCTGGTGGCGCTGCTGCTGGATACCCGAGCCCTGGAGGTGACAGGCACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S
 CTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAGGCTGGGAGAACATGGCTCGCAAACTCTTTGGGCTTGGGGTGAAGTGTACAGGCTGTTTCTGGA
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D
 TTTGACGGTGAAGACCCCTCCAGAGGCTGTGCACCAACATCTACAAGATCCTCCTGCTGACGGGCTACAGGTTTCACGCATGTGTGCTGACGCTCCCATTTTCATCAGCAAGTTTGAAGAA
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N
 CCCCACATTTTCTGCGGCTCATCTGTGACAGGCTCCCTCTGCTACTCCATCCTGAAAGCCAAAGAACCGAGGGAATGTGCTGCGGGCCAAAGGGCCCGCGGCTCTGCGCTCCGA
 P T F F L R V I S D T A S L C Y S I L K A K N A E
 CCGAAGAAAACATTTCTGTGCTGACTCCTGCGGTGCTTGGGTC
 E E N I L V V T P A V L G S
 GGGACAGCAGAGATGGAGCCACCCGACAGCCGTGGGTGTGGGCACTTTCCGGTGTCTCCTGGGAGGGGAGTTGGGCTGGGCTGTGACTCCTCAGCCTGTGTTTCCCGCAG
 G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L *

图 11AG

缺少基元 A 的蛋白质 (版本 2)

ATGCCGCGCGCTCCCGCTGCCGAGCGTGGCTCCCTGCTGCCGAGCCACTACCGGAGGTGCTGCCGCTGGCCAGTTCGTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCAGGGCTGGGGCTGGTGCAGGGGGGACCCGGGGCTTTCCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACCGACGGCCCGCCCGCCCG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCTCCCGGGTCCGGTCCGGTGGGGTTGACGGCGGGGGGAAACCAGCCACATGCGGAGAGCAGCGAGGCGACTCAGGGCTTCCCGCCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T O G A S P A G
P P R G R R P A G V E G G R G E P A T C G E O R R R L R A L P P Q V

CCCTCCTCCGCGAGGTGCTGCCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCCAGCGCGCGGAAAGAACGTGCTGGCTTCCGCTTCCGCTGCTGGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGGCTTACCACCGAGTGGGAGCTAGCTGCCAACAGGTTGACCGCACTGCGGGGAGCGGGGGTGGGGCTGCTGCTGCCGCGTGGGAGCAGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTCACTGCTGGCAGGCTGCGGCTCTTGTGCTGGTGGCTCCAGCTGCGGCTACCAGGTGTGCCGGCGCGGCTGTACCAGTCCGCGCTGCCACTCAGGCCCGCCCCCGCC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCGAGGGCTCTGGATGGAAAGGGCTGGAACCATAGGTCAGGAGGGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGCAGGAGGGCGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAGTCTGCCGTTGCCAAGAGCCAGGGTGGGGCTGCCCTGAGCGGAGCGGACGCCCTTGGCAGGGTCTTGGCCACCCGGGAGGAGCGGTGGAACGAGTACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTGTCACCTGCCAGACCCGCGAAGAAGCCACTCTTGGAGGGTGGGCTCTGTGGCAGCGCCACTCCACCATCCGTTGGGCGCCAGCACCAGCGGGCCCGCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGCCACCACTCCCTGGGACAGCCCTTGTCCCGGGTGTAGCGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAGCTGCGGCTCCTTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTGCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTCGTGAGACCATCTTTCTGGTTCAGGCGCTGATGCCAGGACTCCCGCAGGTTGCCCGCTGCCAGCGGTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P O R Y W Q

AATGGGCCCCCTGTTCTGAGCTGCTTGGGACGACCGGAGTGGCCCTACGGGGTCTCCTCAAGACCGACTGCCCGCTGGGAGTGGGTACCCGAGCAGCGGTGTCTGTGCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCGAGGGCTCTGTGGCGGCCCGGAGGAGGAGACAGACCCCGTGGCTGCTGCACTGCTCCGCGAGCAGCAGCCCGTGGCAGGTGTAGGCTTGTGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCGCGCGCTGGTCCCGCAGGCTTGGGGCTCCAGGCACACCAACCGCTTCTCAGAACACCAAGAAGTTTCATCTCCCTGGGGAAGCATGCCAAGCTGTGCTGCAGGAGT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

图 11AH

GACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAGGAGGCCAGGGGTTGGGTGTGTTCCGGCCGACAGCACCGCTGCGGTGAGGAGATCCTGGCCAAAGTTCCTGCACTGGCT
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L
 GATGAGTGTGTACGTCGAGCTGCTCAGGTCTTTCTTTTATGTCACGGAGACCAGTTCAAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTCGCAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTGGAAGCAGAGGTGAGGCAGCATCGGAAGCCAGGGCCGCGCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGA
 I R Q H L K R Y Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
 CGGGCTGCGGGCATTGTGAACATGGACTAGCTGCTGGGAGCCAGAACGTTCCGACAGAAAAGAGGGCCGAGGCTCACTCGAGGGTGAAGGCACTGTTGAGCGTGTCAACTACGA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGCGCGGGCCCGCGCTCCTGGGCGCTCTGTGCTGGGCTGGACGATATCCACAGGGCTGGCGACCTTCGTGCTGGGTGCGGGCCAGGACCGCGCGCTGAGCTGTACTT
 R A R R P G L L G A V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAAG GACAGGCTCACGGAGGTGATGCCAGCATCATCAAACCCAGAACAGCTACTGCTGCGTCCGATGCCGTGGTCCA
 V K D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGGCCGCCATGGGCAGTCCCGAAGGCTTCAAGAGCCAGTCTCTACCTGACAGACCTCCAGCCGTACATCCGACAGTTCGTGCTCACCTGCAGGAGACCAGCCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R O F V A H L Q E T S P L R D
 TGCCGTGCTATCGCAAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTCTTGACGCTTCTCCTACGCTTCAATGTGCCACCCAGCCGTGCGCATCAGGGCCAGTCTACGTCCAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGATCCCGAGGGCTCCATCCTCTCCACGCTCTGTCAGCCTGTGCTACGGGACATGAGAACAAAGCTGTTTCCGGGATTCCGGGGACGGGCTGCTCCTGCGTTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTGTTGGTACACCTCACCTCACCCAGGAAAACCTTCTCAGGACCTGGTCCGAGGTGCTCCCTGAGTATGGCTGCGTGGTGAAGTTCGCGAAGCAGTGGTGAAGTTCAC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAGAAGCAGGGCCCTGGTGGCAGGCTTTTGTTCAGATGCCGGCCAGGGCTATTCCCTGGTGGGCGCTGCTGCTGATACCCGGACCTGGAGTGGCAGGCACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S
 CTATGCCCGACCTCCATCAGAGCAGTCTCACCTTCAACCGGGCTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGCTCTGGCGCTGAAGTGTACACGCTGTTTCTGGA
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D
 TTTGAGGTGACAGCCTCCAGAGGTTGTGACCAACATCTACAAGATCCTCCTGCTGCAGGCTACAGGTTTCAGCATGTGTGCTGCAGCTCCCATTTTCATCAGCAAGTTTGGAGAA
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W X N
 CCCACATTTTCTGCGGTGCTCTGTGACAGGCTCCTCTGCTACTCCATCCTGAAAGCCAAAGACCGAGGATGTGCTGAGGGCCAGGGCCCGCGGCGCTCTGCGCTCCGA
 P T F F L R V I S D T A S L C Y S I L K A K N A G M S L G A K G A A G P L P S E
 GCGCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTCGACACCGTGTACCTAGTGCAGCTCCTGGGCTCACTCAGSACAGCCAGCAGGCTGAGTCCGAAGCTCCC
 A V Q W L C H O A F L L K L T R H R V T Y V P L L G S L R T A Q T Q L S R K L P

图 11A1

GGGGACGACGCTGACTGCCCTGGAGGGCCGACGCCAACCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCGGCCACAGCCAGGCCGACAGCAGACACCAGCAGCC
G T T L T A L E A A A N P A L P S D F K T I L D
CTGTACGCCGGGCTCTACGTCCCAGGSAGGGAGGGCCGCCACCCAGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTGGCCGAGGCCTGCATGTCCGGCTGAAGGCT
GAGTGTCCGGCTGAGGCCTGAGCCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCCGTCTTCACTTCCCACAGGCTGGCGCTCCGGCTCCACCCGAGGCCAGCTTTTCTCAC
CAGGAGCCGGGTTCCACTCCCCACATAGGAATAGTCCATCCCAGATTGCCATTGTTACCCCTGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGACCCCTGAGAA
GGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAGGTGTGCCCTGTACACAGCCGAGGACCCTGCACCTGGATGGGGTCCCTGTGGGTCAAATTTGGGGGAGGTGCTGTGGGAGTAA
AATACTGAATATATGAGTTTTTCAGTTTTGA

图 11AJ

缺少基元 A 的截短的蛋白质 (版本 2)

ATGCCGGCGCCTCCCGCTGCCAGCGGTGGCTCCCTGCTGGCGAGCCACTACCGGAGGTGGTCCGCTGGCCAGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCCTGGGGCCAGGGCTGGGGCTGGTGCAGCGGGGACCCGGGGTTCCGGCGCTGGTGGCCAGTGCCTGGTGGCTGGCTGGGACGACGGCCGGCCCGGGCGG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGTGGCTCCGGCTGGGTTGAGGGCGCCGGGGGAACACGGACATGCCGAGAGCAGCGCAGGCTCAGGGCGCTCCCGCCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCCCTCTCCGCGAGTGTCTGCCTGAAGGAGTGGTGGCCGAGTGTGCAGAGGCTGTGCGAGCGGGCCGAAGAAGCTGCTGGCTTCGGCTTCGGCTTCGGCTGGACGGGGCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTCAGCACAGGGTGGCAGCTAGCTGCCCAACAGGTGACCGACGCACTGGGGGGAGCGGGCTGGGGCTGCTGCTGGCCGCTGGCGACGACGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGCGCTCTTTGTGTGGTGGCTCCAGCTGGCGCTACCAGGTGTGCGGGCCGCGCTGTACCAGCTGGCGTGGCACTCAGGCCGGCCCGCC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACCTAGTGGACCCGAGGGCTGCGGATGGACGGGCTGGAACATAGCGTCAGGGAGGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGCAGGAGGGGGGGGCGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAGTGGCTGGCCAGAGGGCCAGGGTGGCGTGGCCGTGAGCCGAGGGACCGCCGTTGGCAGGGTCCGGGCGCACCGGGCAGGAGCGGTGGACCGAGTGACCG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTTCTGTGGTGTGCTGCGCAGACCGGGCAGAGAGCCACCTTTTGGAGGTTGGCTCTCTGCCAGCGCCACTCCACCCATCGGTGGGCGCAGCACCAGCGGGGGCCCG
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGACAGGCTTGTCCCGGGTGTAGCCGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAGCTGGCGCCCTCCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGCCAGCCTGACTGGCGCTGGGAGGCTGGTGGAGACCATCTTTCTGGGTTCCAGGCGCTGGATGCCAGGGACTCCCGCAGGTTGGCCCGCTGGCCAGGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGGCCCTGTTCTGGAGTGCCTGGGACCGGCACTGCCCTACGGGTGCTCCTCAAGACGCACTGCCCGTGGCAGCTGGCGTACCCAGCAGCGGCTGTGTGGCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTGTGTGGGGCCCGGAGGAGGACACAGACCCCGTGGCTGGTGCAGCTGGTCCGCGAGCACAGCAGCCCTGGCAGGTGTACGGCTGTGGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCCCGAGGCTGTGGGGTCCAGGACAAAGACCGGCTCCTCAGGACACCAAGATTCACTCCCTGGGGAGCATGCCAGCTCTCCTGAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GAGTGGAGATGAGGTGGGGACTGGCTTGGCTGGCAGGAGCCAGGGTGGCTGTGTTCCGGCCGAGAGCACCGTCTGGGTGAGGAGATCCTGGCCAGTTCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

图 11AK

GATGAGTGTGTACGTGTCGAGCTGCTCAGGTCTTTCTTTTATGTACGGAGACCAGSTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGACCAAGTTGCAAAGCATTGG
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
 AATCAGACAGCACTTGAAGAGGGTGCAGTGCAGGAGCTGTCGGAAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGA
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D
 CGGGCTGCGGGCATTGTGAACATGGACTACGTGCTGGGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTACGCGTCTCAACTAGCA
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E
 GCGGGCGGGCGCCCGGCTCCTGGGGCCTCTGTGCTGGGCCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCAGGACCCCGCGCTGAGGTGACTT
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F
 TGTCAGG GACAGGCTCAGGGAGGTGATCGCCAGCATCATCAAACCCAGAACACGTACTGCGTGGCTCGGTATGCCGTGGTCCA
 V K D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q
 GAAGCCCGCCATGGGCACGTCCGCAAGGCCTCAAGAGCCACGTCTACCTTGACAGACCTCCAGCCGTACATGCCACAGTTCGTGGCTCACCTGACAGSAGACCAGCCCGCTGAGGGA
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D
 TGGCTGCTGATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCACCGGTGGCATCAGGGGAAGTCTACGTCCAGTG
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C
 CCAGGGATCCCGAGGGTCCATCCTCTCCACGGTGTCTGGACGCTGTGCTACGGGACATGGAGAACAAGCTGTTGCGGGATCCGGCGGACGGGCTGCTCCTGGGTTTGGTGA
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D
 TGATTTCTGTTGGTACACCTCACCTCACCCACGCGAAAACCTTCCTCAGGACCCGTGTCGAGGTGTCCTGAGTATGGCTGCGTGGTGAACCTGCGGAAGACAGTGGTGAACCTCCC
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P
 TGTAAGACGAGGCCCTGGGTGGCAGGCTTTGTTTCAGATGCCGGCCACGGCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGGACCCCTGGAGGTGCAGAGGCACTACTCCAG
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S R
 GTGAGCCACCTGGCCGSAAGTGGAGCCTGTGCCGGCTGGGGCAGGTGCTGCTGCAGGGCGGTTGCGTCCACCTCTGCTTCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGA
 TGCCACAGGSGTGCCTCCTCCATCTGGGGCTGAGCACAATGCATCTTTGTGGGAGTGAGGSGTGCCTCACAAAGGGAGCAGTTTTCTGTGCTATTTTGGTAA

图 11AL

缺少基元 A 和改变的 C-末端 (版本 2)

ATGCCCGCGCTCCCGCTGCCGAGCGTGCCTCCCTGCTGCCGAGCCACTACCGGAGGTGCTGCCGCTGCCACGTTGCTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGCCGCTGGGGCCCAAGGCTGGCGGTGGTGCAGCGGGGACCCGGCGGCTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGGTGGCTGGGACGACGGCGCCCGCCCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GCCCTCCCGGGGTGGCGTCCGCTGGGGTTGAGGGCGGCGGGGGGAACAGCGACATCGGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCGCGCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCCCTCCCGCAGGTGCTGCTGCTGAAAGAGCTGGTGGCGGAGTGTGCAGAGGCTGTGGGAGGCGGGCGGAAGAAGCTGCTGGCTTCGGCTTCGGCTGCTGGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCGGAGGCGCTCACCACAGGCTGCGGAGCTACCTGCCAACACCGTGACCGAGCCACTGCCGGGAGCGGGCGTGGGGCTGCTGCTGCCCGGCTGGGCGACGACGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGCCTGCCGCTGTTGTGCTGGTGGCTCCAGCTGCCCTACCAGGTGTGGGGCGCGCGTGTACCAGCTGGCGCTGCCACTCAGGCCCGGGCCCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCCGAAGGCGTCTGGATGGGACCGGCTGGAACCATAGCGTACGGGAGGCGGGGTCCCGCTGGGCTGCCAGCCCGGGTGGGAGGCGGGGGCAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCGTTGCCCAAGAGGCGCAGCGTGGCGTGCCTGAGCGGAGCGGACCGCGTTGESCAGGGTCTGGGCCACCCGGGAGGACCGCTGGAACGAGTGCAGG
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGTGCTCACCTGCCAGACCCCGGAAGAAGCCACTCTTTGAGGGTGGGCTCTGTGCACGGCCACTCCACCCATCCGTGGCGGCGCAGCACCAGCGGGCCCGC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGACACGCTCCCTGGGACAGGCTGTGCCCGGTGTACCGGAGACCAAGCACTTCCTCTACTCCTCAGGGACAGGAGCAGTGCGGCCTCCTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCAGCCTGACTGGCGTGGGAGCTGCTGGAGACCATTTTGTGGTCCAGGCGCTGATGCCAGGACTCCCGCAGGTGGCCCGCTGCCACCGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCGCCCTGTTCTGGAGCTGCTGGGACCGGCGAGTGGCCCTACGGGCTGCTCAAGACCGCACTGCCCGTGGAGTGGGTGACCCAGCAGCGGCTGTGTGCCCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAGCCCGAGGCTGTGTGGCGCCCGCCAGGAGGAGACAGACACCCCGTGGCTGGTGCAGCTGCTCGGCGAGCAGGACCCCGTGGCAGGTGACCGCTTCTGCGGGCGTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGCCTCCAGGCTGTGGGCTCCAGGACAAAGACCGGCTTCTCAGGACACCAAGAGTTTCATCTCCCTGGGGAAGCATGCCAAGCTCTCGTGCAGGAGCT
L R R L V P P G L W G S R H K E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGCGTGGGACTGGGCTGGGTCGGCAGGACCCAGCGGTTGCTGTGTTCCGGCGGAGAGCAGCGTGTGGGAGATCCTGGCAAGTCTCTGCACTGGCT
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

图 11AM
— 5 4 —

GATGAGTGTACGTCGTCGAGCTGCTCAGGTCCTTTCTTTTATGTCACGGAGACCACGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTCAAAGCATTGG
M S V Y V V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGCAGCACTTGAAGAGGTCAGCTGCGGAGCTGTCGSAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGAGCTCAGACTCGGTTTATCCCAAGCCTGA
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P X P D

CGGCTGCGGCCATTGTGACATGACTACGTCGTCGAGCCAGAACGTTCCGACAGAAAAGAGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTACGCTGCTCAACTACGA
G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

CGGCGCGCGCCCGCCCTCTGCGGCGCTCTGTGCTGGSCCTGGACATATCCAGAGGCCCTGGCGCACCTTCGTGCTGCGTGTGGCGCCAGGACCCGCGCCCTGAGCTGACTT
R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAGG GACAGGCTCAGGAGGTCATCGCCAGCATCAAAACCCAGAACACGTAAGTGGTGGTGGTATGCCGTGGTCCA
V K D R L T E V I A S I I K P Q N T Y C V R R Y A V Y Q

GAGGCCGCCATGGGACGTCGCAAGGCTTCAAGAGCCAGTCTGTACCTTGACAGACCTCCAGCCGTACATGGGACAGTTCGTGGCTCACCTGCAAGSAGCCAGCCCGCTGAGGGA
K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D

TGCCCTGTCATCGAGCAGACTCTCCCTGAATGAGGCCAGCAGTGGCTCTTCGACGCTTCTCAGCCTTCATGTGCCACCACCGCTGCGCATCAGGGCAAGTCTACGTCAGTG
A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C

CCAGGGATCCCGCAGGCTCCATCCTCTCCACGCTGCTGTGACGCTGTGGTACGGCAGACATGAGAACAAAGCTGTTTGGGGGATTCCGGCGGACGGGCTGCTGCTGGTGGTGGSA
Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D

TGATTTCTGTTGGTACACTCACCTCACCCACGGCAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCTGAGTATGGCTGCGTGGTGAAGTTCGGAAGCAGTGGTGAAGTCTCC
D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P

TGTAGAAGCAGGCCCTGGGTGGCAGGGCTTTTGTTCAGATGCCGGCCACGGCCATTCCCTGGTGGGGCTGCTGCTGGATACCCGACCTGGAGGTGACAGGCACTACTCCAG
V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S

CTATGCCGACCTCCATCAGAGCAGTCTCACCTTCAACCGCGCTTCAAGGCTGGGAGGAAACATCGGTCGCAAACTCTTTGGGTCCTGGGCTGAAGTGTACAGCCTGTTTGTGSA
Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D

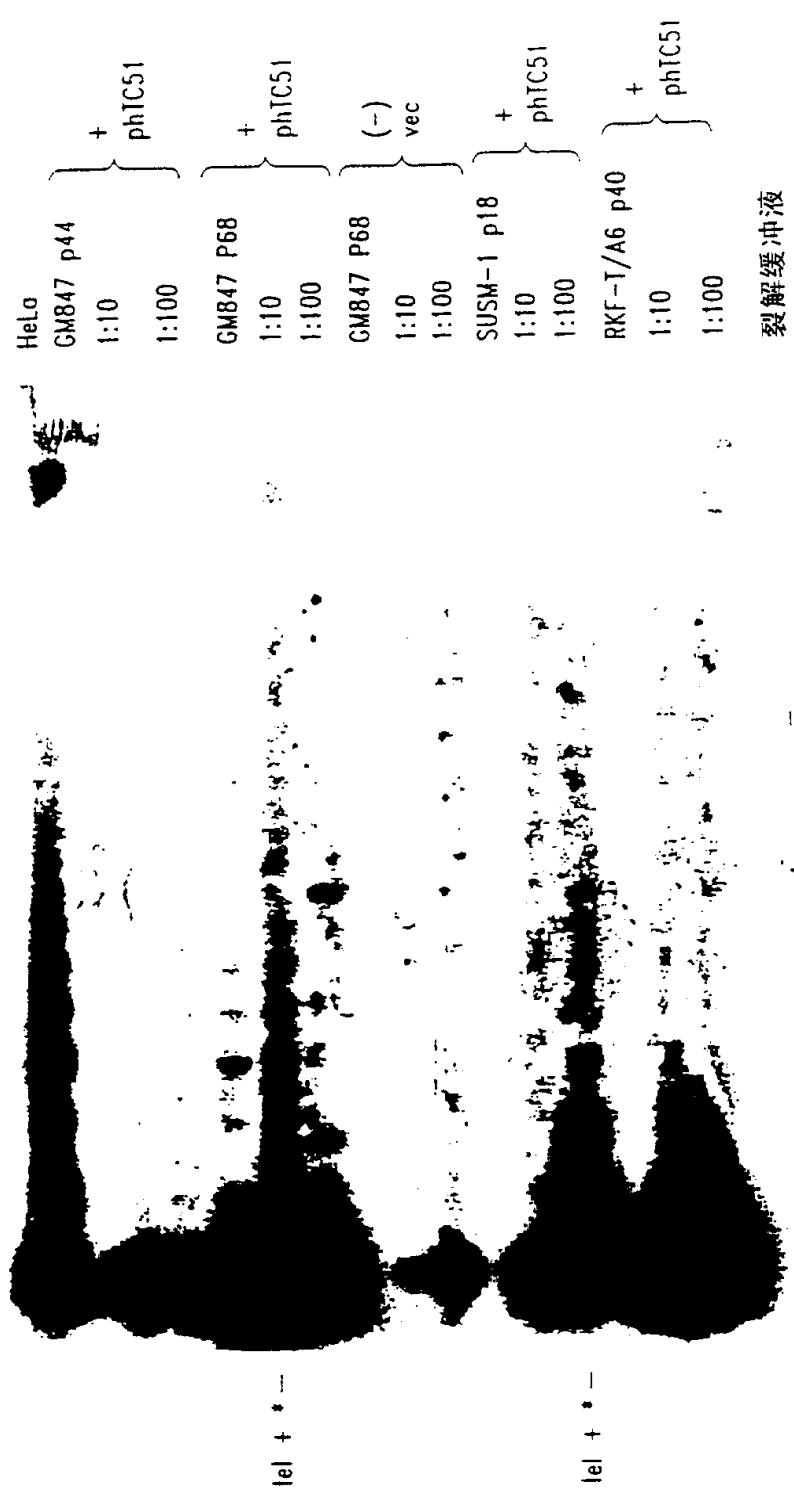
TTTGCAGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCTACAGGTTTCAAGCATGTGTGCTGCAGGTCCTCATCAGCAAGTTTGAAGAA
L Q V N S L O T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N

CCCCAGATTTTCTGCGGCTCATCTCTGACAGGCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGAGGSA TGTGCTGGGGCCAAAGGCGCCCGCCGCTCTGCTCCGA
P T F F L R V I S D T A S L C Y S I L K A K N A E

CGGAAGAAACATTTCTGTGCTGACTCCTGCGGTGCTGGSTC
E E N I L V V T P A V L G S

GGGACAGCCAGAGATGGAGCCACCCGACAGCCTGCGGCTGTGGCAGCTTCCGCTGCTCCTGGGAGGGGAGTTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTCCCCAG
G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L *

图 11AN



25/10/97
接触 5 日
2:6

图 12

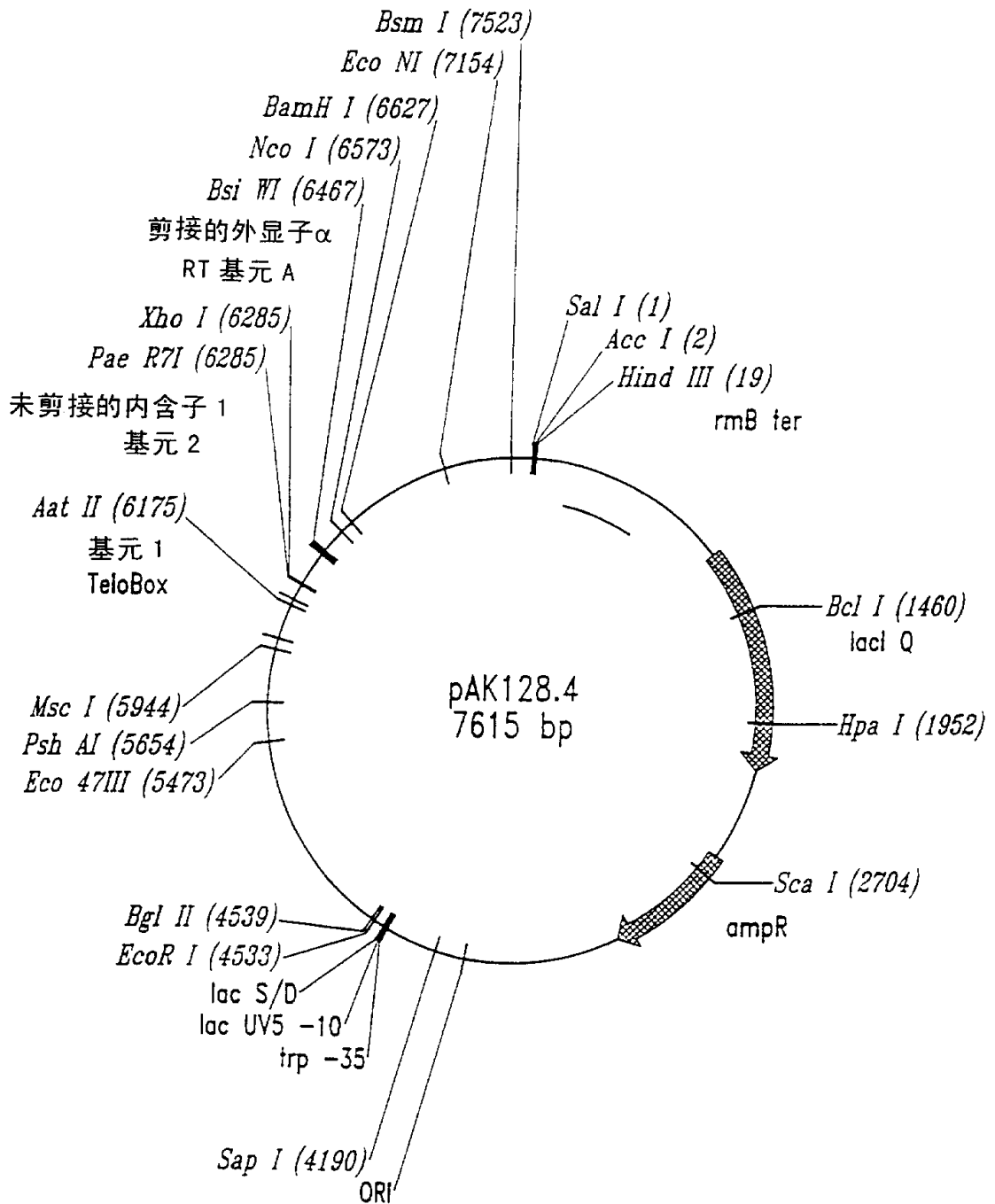


图 13A

基因座 - pAKI28.4 7615 bp dsDNA 环状
 说明 带有剪接出外显子β的人端粒酶克隆

```

1 tgcacctgca ggcattgcaag cttggcactg gccgtcgttt tacaacgctg tgactgggaa
61 aaccttggcg ttacccaact taatgcctt gcagcacatc cccctttcgc cagctggcgt
121 aatagcgaag aggcccgcac cgatcgccct tcccacagt tgcgcagcct gaatggcgaa
181 tggcgctga tgcggtattt tctccttacg catctgtgcg gtatttcaca ccgcataaat
241 tccctgtttt ggcggatgag agaagatttt cagcctgata cagattaaat cagaacgcag
301 aagcggctcg ataaaacaga atttgctgg cggcagtagc gcggtggctc cacctgaccc
361 catgccgaac tcagaagtga aacgccgtag cggcagtagt agtgtggggt ctccccatgc
421 gagagtaggg aactgccagg catcaaataa aacgaaaggc tcagtcgaaa gactgggcct
481 ttcgttttat ctgtgtttg tcggtgaacg ctctcctgag taggacaaat ccgcccggag
541 cggatttgaa cgttgcaag caacggcccg gagggtggcg ggcaggacgc ccgccataaa
601 ctgccaggca tcaaatgaag cagaaggcca tcctgaocga tggccttttt gcgtttctac
661 aaactcttcc tgtcgtcata tctacaagcc atccccccac agatacggta aactagcctc
721 gtttttgcac caggaaagca ggaatttat ggtgcactct cagtacaatc tgcctgatg
781 ccgcatagtt aagccagccc cgacaccgcc caacaccgcc tgacgcgcc tgacgggctt
841 gtctgctccc ggcattccgt tacagacaag ctgtgaccgt ctccgggagc tgcattgttc
901 agaggttttc accgtcatca ccgaaacgcg cgagacgaaa gggcctcgtg ataccgctat
961 ttttataggt taatgtcatg ataataatgg tttcttagac gtgaggttct gtacccgaca
1021 ccattcgaatg gtgcaaaacc tttcgcggtg tggcattgata gcgcccggaa gagagtcaat
1081 tcagggtggt gaattgtgaa ccagtaacgt tatacagatg cgcagagtat gccggtgtct
1141 cttatcagac cgtttcccgc gtggtgaacc agccagcca cgtttctcgc aaaacgcggg
1201 aaaaagtgga agcggcgatg gcggagctga attacattcc caaccgcgtg gcacaacaac
1261 tggcgggcaa acagtcgttg ctgattggcg ttgccacctc cagtcaggcc ctgcacgcgc
1321 cgtcgcaaat tgtcgcggcg attaaatctc gcgccgatca actgggtgcc agcgtggtgg
1381 tgtcagatgtt agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgcac aatcttctcg
1441 cgcaacgcgt cagtggtctg atcattaact atccgctgga tgaccaggat gccattgtctg
1501 tggaaagctg ctgcactaat gttccggcgt tatttcttga tgtctctgac cagacaccca
1561 tcaacagtat tattttctcc catgaagacg gtacgcgact gggcgtggag catctggtcg
1621 cattgggtca ccagcaaatc gcgctggttag cgggccatt aagtctctgc tcggcgcgtc
1681 tgcgtctggc tggctggcat aaatatctca ctcgcaatca aattcagccg atagcggaac
1741 ggaaggcga ctggagtgc atgtccggtt ttcaacaac catgcaaatg ctgaatgagg
1801 gcattcgtcc cactgcgatg ctggttgcca acgacagat ggcgctggc gcaatgcgcg
1861 ccattaccga gtcgggctg cgcgttggtg cggatatctc ggtagtggga tacgacgata
1921 ccgaagacag ctcatgttat atcccgcgt taaccaccat caaacaggat tttcgcctgc
1981 tggggcaaac cagcgtggac cgttgctgc aactctctca gggccaggcg gtgaaggcca
2041 atcagctggt gccctctca ctggtgaaaa gaaaaaccac cctggcgcgc aataccgaaa
2101 ccgctctcc ccgcgcttg gccgattcat taatgcagct ggcacgacag gtttcccgc
2161 tggaaagcgg gcagtgagcg caacgcaatt aatgtaagtt agctcactca ttaggcacc
2221 caggctttac actttatgct tccgacctgc aagaacctca cgtcagggtg cacttttcgg
2281 ggaaatgtgc gcggaacccc tattgttta tttttctaaa tacattcaaa tatgtatccg
2341 ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt
2401 attcaacatt tccgtgtcgc ccttattccc ttttttcggc cattttgcct tctgttttt
2461 gctcaccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg
2521 ggttacatcg agaactggat ctcaacagcg gtaagatcct tgagagtttt cggcccgaag
2581 aacgttttcc aatgatgagc acttttaag ttctgtatg tggcgcggta ttatcccgta
2641 ttgacgcgg gcaagagcaa ctcgctgcgc gcatacacta ttctcagaat gacttggttg

```

图 13B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca
2761 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag
2821 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc
2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg
2941 tagcaatggc aacaacgctg cgcaactat taactggcga actacttact ctgacttccc
3001 ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcgctcgg
3061 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggctctcgg
3121 gtatcattgc agcactgggg ccagatggra agccctccc tatcgtagt atctacacga
3181 cggggagtcg gcaactatg gatgaacgaa atagacagat cgcctgagata ggtgcctcac
3241 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatitaa
3301 aacttcattt ttaatttaa aggatctagg tgaagatcct ttttgataat ctcatgacca
3361 aaatccctta acgtgagitt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag
3421 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac
3481 cgctaccagc ggtggttgt ttgcggatc aagagctacc aactctttt cogaaggtaa
3541 ctggcttcag cagagcgcag atacaaata ctgtccttct agttagcgc tagttagcc
3601 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaate ctgttaccag
3661 tggctgctgc cagtggcgtg aagtcgtgic ttaccgggtt ggactcaaga cgatagttac
3721 cggataaggc gcagcggctg ggtgaacgg ggggttcgtg cacacagccc agcttggagc
3781 gaacgacctc caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcttc
3841 ccgaaggag aaaggcggac aggtatccgg taagcggcag ggtcggaaaca ggagagcgc
3901 cgagggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcgcacc
3961 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg cgggagccta tggaaaaacg
4021 ccagcaacgc ggctttttt cggttcctgg ccttttctg gccttttct cacatgttct
4081 ttctgtcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata
4141 ccgctcggcg cagccgaacg accgagcgcg gcgagtcagt gagcgaggaa gcggaagagc
4201 gcccaatacg caaacgcct ctccccgcgc gttggccgat tcattaatgc agaattaatt
4261 ctcatgtttg acagcttacc atcgactgca cggtgacca atgcttctgg cgtcaggcag
4321 ccacgcgaag ctgtggtatg gctgtgcagg tcgtaaatca ctgcataatt cgtgtcgcctc
4381 aaggcgcact ccgcttctgg ataattttt ttgcggcgac atcataacgg ttctggcaaa
4441 tattctgaaa tgagctgttg acaattaatc atcggctcgt ataattgtgt gaattgtgag
4501 cggataacaa tttcacacag gaacagcga tgaattcaga tctcaccatg aaggagctgg
4561 tggcccaggt gctgcagagg ctgtgcgagc gcggcgcgaa gaacgtgctg gccttcggct
4621 tcgctgctgt ggacggggcc cgcgggggccc cccccgaggc cttcaccacc agcgtgcgca
4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtgg gggctgctgc
4741 tgcgcccgtt gggcgacgac gtgctggttc acctgctggc acgctgcgcg ctctttgtgc
4801 tgggtgctcc cagctgcgccc taccaggtgt gcgggcccgc cgtgtaccag ctccggcctg
4861 ccactcaggc ccggcccccg ccacacgcta gtggaccccc aaggcgtctg ggatgcgaac
4921 gggcctggaa ccatagcgtc agggaggcgg ggtccccct gggcctgcca gccccgggtg
4981 cgaggrggcg cgggggcagt gccagccgaa gtctgccgtt gccaagagg ccagggcgtg
5041 gcgctgcccc tgagccggag cggacgccc ttgggcaggg gtcttgggccc caccgggca
5101 ggacgcgtgg accgagtgac cgtggtttct gtgtggtgic acctgccaga ccgcccgaag
5161 aagccacctc ttggaqggt gcgctctctg gcacgcgcca ctcccacca tccgtgggccc
5221 gccagacca cgcggycccc ccateccat cgcggccacc acgtccctgg gacacgcctt
5281 gtccccgggt gtacgcccag accaagcact tcctctactc ctccaggcag aaggagcagc
5341 tgcggccctc ctctctactc agctctctga ggcccagcct gactggcgtc cggaggctcg
5401 tggagaccat ctctctgggt tccaggccct ggatgccagg gactccccgc aggttgcctc
5461 gcctgccccca gcgctactgg caaatgcggc ccctgtttct ggagctgctt gggaaaccag
5521 cgcagtgccc ctacgggggt ctctcaaga cgcactgccc gctgcgagct gcggtcacc

图 13C

5581 cagcagccgg tgtctgtgcc cgggagaagc cccagggctc tgtggcggcc cccgaggagg
5641 aggacacaga cccccgtcgc ctgggtcagc tgctccgccg gcacagcagc ccttggcagg
5701 tgtacggcct cgtgcggggc tgcctgcgcc ggctgggtgc cccaggcctc tggggctcca
5761 ggcacaacga acgccgcttc ctcaggaaca ccaagaagt catctccctg ggaagcatg
5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt gcgggactgc gcttggctgc
5881 gcaggagccc aggggttggc tgtgttccgg ccgcagagca ccgtctcgtg gaggagatcc
5941 tggccaagt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct
6001 ttatgtcac ggagaccacg tttcaaaaga acaggctctt tttctaccgg aagagtgtct
6061 ggagcaagt gcaaagcatt ggaatcagac agcacttgaa gagggtgcag ctgcgggagc
6121 tgcggaagc agaggtcagg cagcatcggg aagccaggcc cgccctgctg acgtccagac
6181 tccgcttcat ccccagcct gacgggtgc gcccgattgt gaacatggac tacgtcgtgg
6241 gagccagaac gttccgcaga gaaaagagg ccgagcgtct cacctcgagg gtgaaggcac
6301 tttcagcgt gctcaactac gagcggggcg gccgccccgg cctcctgggc gcctctgtgc
6361 tgggcttga cgatatccac agggcctggc gcaccttctg gctcgtgtg cggggccagg
6421 acccgccgcc tgagctgtac tttgtcaagg tggatgtgac gggcgcgtac gacaccatcc
6481 cccaggacag gctcacggag gtcacgccg gcacatcaa accccagaac acgtactcgg
6541 tgcgtcggta tgccgtggc cagaaggccg cccatgggca cgtccgcaag gccttcaaga
6601 gccacgtcct acgtccagt ccaggggatc ccgcagggt ccatctctc cacgtgctc
6661 tgcagcctgt gctacggcga catggagaac aagctgtttg cggggattcg gcgggacggg
6721 ctgctcctgc gtttgggtga tgattcttg ttggtgacac ctcacctcac ccacgcgaaa
6781 acttccctcag gacctggtcc gaagtgtcct gagtatggct gcgtggtgaa cttgcggaag
6841 acagtgggtga acttccctgt agaagacgaa gccctgggtg gcacggcttt tttcagatg
6901 cgggccacg gcctattccc ctgggtcggc ctgctgctgg ataccggac cctggagggtg
6961 cagagcgact actccageta tgcccggacc tccatcagag ccagtctcac cttcaaccgc
7021 ggcttcaagg ctgggaggaa catgcgtcgc aaactctttg gggctctgag gctgaagtgt
7081 cacagcctgt ttctggatt gcagggtgac agcctccaga cgggtgtgac caacatctac
7141 aagatcctcc tgctgcaggc gtacaggttt cacgcattgt tgctgcagct cccatttcat
7201 cagcaagttt ggaagaacc cacattttc ctgcgcgtca tctctgacac ggctccctc
7261 tgctactcca tctgaaagc caagaacgca gccgaagaaa acatttctgt cgtgactcct
7321 gcggtgcttg ggtcgggaca gccagagatg gagccacccc gcagaccgtc ggggtgtggc
7381 agctttccgg tgtctcctgg gaggggagtt gggctgggcc tgtgactcct cagcctctgt
7441 tttccccag ggatgtcgt gggggccaag ggcgcgccg gccctctgcc ctccgaggcc
7501 gtgcagtggc tgtgccacca agcattcctg ctcaagctga ctcgacacc tgtcacctac
7561 gtgccactcc tggggctact caggacagcc aagtgtgggt ggaggccagt gcggg

图 13D

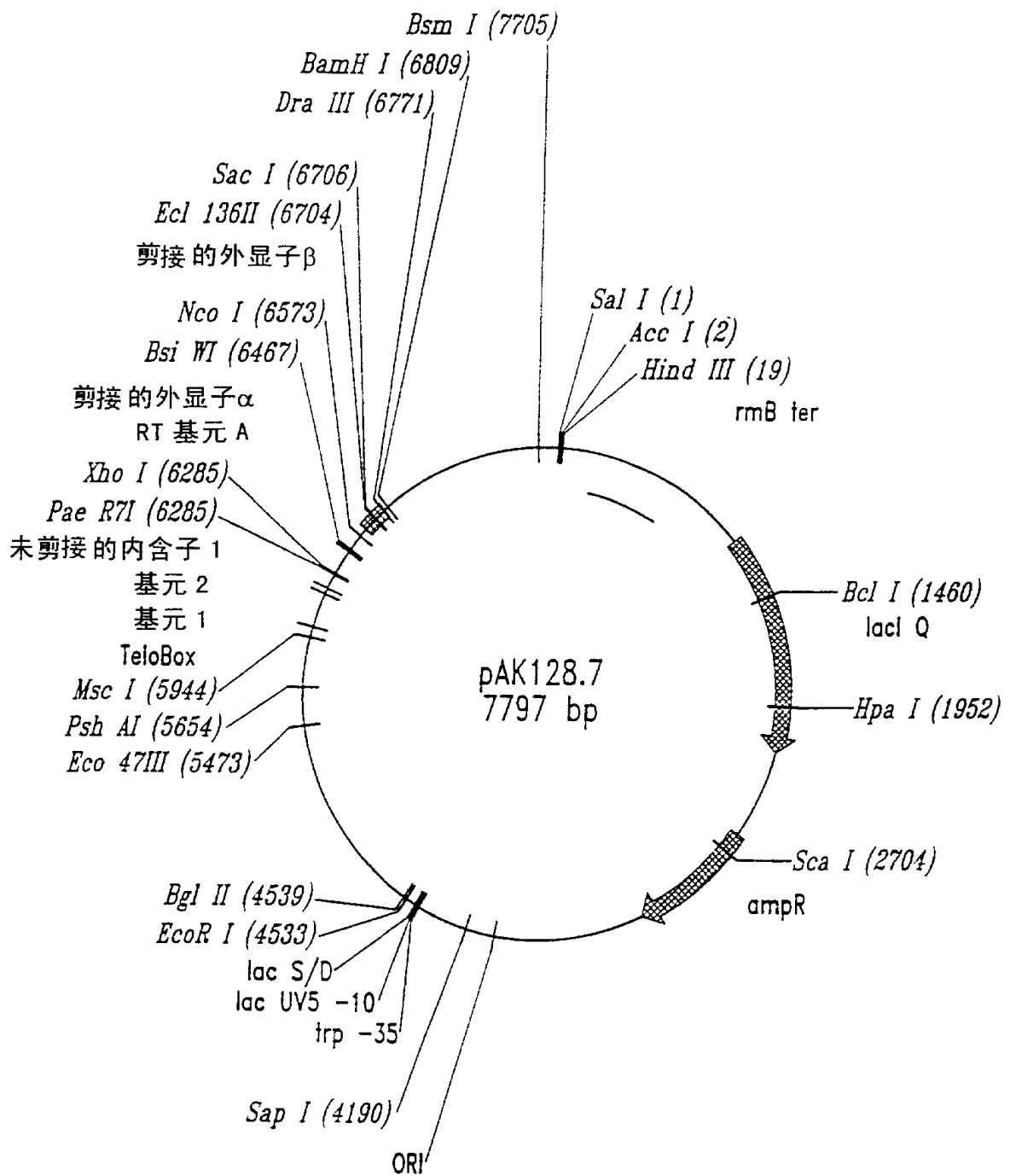


图 14A

基因座 - pAKI28.7 7797 bp dsDNA 环状
说明 带有改变的 C-末端的人端粒酶克隆

```

1  tcgacctgca  ggcattgcaag  cttggcactg  gccgtcgttt  tacaacgtcg  tgactgggaa
61  aaccctggcg  ttacccaact  taatcgccct  gcagcacatc  cccctttcgc  cagctggcgt
121  aatagcgaag  aggcccgcac  cgatcgccct  tccaacagct  tgccgagcct  gaatggcgaa
181  tggcgctga  tgcggtatct  tctccttacg  catctgtgcg  gtatttcaca  ccgcataaat
241  tccctgtttt  ggcggatgag  agaagatctt  cagcctgata  cagattaaat  cagaacgcag
301  aagcggctct  ataaaacaga  atttgcctgg  cggcagtagc  gcgggtggtc  cacctgaccc
361  catgccgaac  tcagaagtga  aacgcccgtg  cgcggatggt  agtgtggggt  ctccccatgc
421  gagagtaggg  aactgccagg  catcaaataa  aacgaaaagg  tcagtcgaaa  gactgggctt
481  ttcgttttat  ctgttgtttg  tcggtgaacg  ctctcctgag  taggacaaat  ccgcccggag
541  cggatttgaa  cgttgcgaag  caacggcccg  gaggggtggc  ggcaggacgc  ccgcataaaa
601  ctgccaggca  tcaaattaag  cagaaggcca  tcctgacgga  tggccttttt  gcgtttctac
661  aaactcttcc  tgtcgtcata  tctacaagcc  atccccccac  agatacggta  aactagcctc
721  gtttttgcat  caggaaagca  gggaaatctt  ggtgcaactc  cagtacaatc  tgctctgatg
781  ccgcatagtt  aagccagccc  cgacaccgcc  caacaccgcc  tgaccgcgcc  tgaccgggctt
841  gtctgctccc  ggcattccgc  tacagacaag  ctgtgaccgt  ctccgggagc  tgcatgtgtc
901  agaggtttcc  accgtcatca  ccgaaacgcg  cgagacgaaa  gggcctcgtg  ataccgctat
961  ttttataggt  taatgtcatg  ataataatgg  tttcttagac  gtgaggttct  gtacccgaca
1021  ccatcgaatg  gtgcaaaacc  tttcgggta  tggcatgata  gcgcccggaa  gagagtcaat
1081  tcagggtggt  gaatgtgaaa  ccagtaacgt  tatacagatg  cgcagagtat  gccggtgtct
1141  cttatcagac  cgtttcccgc  gtgtgtaacc  agccagacca  cgtttctgcg  aaaacggggg
1201  aaaaagtgga  agcggcgatg  gcggagctga  attacattcc  caaccgcgtg  gcacaacaac
1261  tggcgggcaa  acagtcgttg  ctgattggcg  ttgccacctc  cagtctggcc  ctgcacgcgc
1321  cgtcgcgaaa  tgctcggcgc  attaaatctc  gcgccgatca  actgggtgcc  agcgtggtgg
1381  tgtcagatgt  agaacgaagc  ggcgtcgaag  cctgtaaagc  ggcgggtcac  aatcttctcg
1441  cgcaacgcgt  cagtgggctg  atcattaact  atccgctgga  tgaccaggat  gccattgctg
1501  tggagagctg  ctgcactaat  gttccggcgt  tatttcttga  tgtctctgac  cagacacca
1561  tcaacagtat  tattttctcc  catgaagacg  gtacgcgact  gggcgtggag  catctggtcg
1621  cattgggtca  ccagcaaatc  gcgctgtag  cgggccatt  aagttctgtc  tcggcgcgtc
1681  tgcgtctggc  tggctggcat  aaatatctca  ctgcgaatca  aattcagccc  atagcggaac
1741  gggaaggcga  ctggagtgcc  atgtccggtt  ttcaacaac  catgcaaatg  ctgaatgagg
1801  gcatcgttcc  cactgcgatg  ctggttgcca  acgatcagat  ggcgctgggc  gcaatgcgcg
1861  ccattaccga  gtccgggctg  cgcgttggtg  cggatatctc  ggtagtggga  tacgacgata
1921  ccgaagacag  ctcatgttat  atcccgccgt  taaccaccat  caaacaggat  tttcgcctgc
1981  tggggcaaac  cagcgtggac  cgcttgcctg  aactctctca  gggccaggcg  gtgaagggca
2041  atcagctggt  gcccgtctca  ctggtgaaaa  gaaaaaccac  cctggcggcc  aatacgcgaa
2101  ccgcctctcc  ccgcccgttg  gccgattcat  taatgcagct  ggcaacgacg  gtttcccgcg
2161  tggaaagcgg  gcagtgagcg  caacgcaatt  aatgtaagt  agctcactca  ttaggacccc
2221  caggctttac  actttatgct  tccgacctgc  agaacctca  cgtcaggtgg  cacttttcgg
2281  ggaaatgtgc  gcggaacccc  tatttgttta  tttttctaaa  tacattcaaa  tatgtatccg
2341  ctcatgagac  aataaccctg  ataaatgctt  caataatatt  gaaaaaggaa  gagtatgagt
2401  attcaacatt  tccgtgtcgc  ccttattccc  tttttgccc  cattttgcct  tcctgttttt
2461  gctcaccag  aaacgctggt  gaaagtaaaa  gatgctgaag  atcagttggg  tgcaacgagt
2521  ggttacatcg  agaactggat  ctcaacagcg  gtaagatcct  tgagagtttt  ccgcccgag
2581  aacgttttcc  aatgatgagc  acttttaaa  tctcgtatg  tggcgcggta  ttatcccgtg
2641  ttgacgccgg  gcaagagcaa  ctccgctgcc  gcatacacta  ttctcagaat  gacttggttg

```

图 14B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca
2761 gtgtgccat aaccaatgagt gataacactg cggccaactt acttctgaca acgatcggag
2821 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc
2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgctg
2941 tagcaatggc aacaacgttg cgcaactat taactggcga actacttact cttagcttccc
3001 ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgocctcgg
3061 cccttccggc tggttggttt attgctgata aatctggagc cggtgagcgt gggctctcgg
3121 gtatcattgc agcactgggg ccagatggta agccctcccg tctcgtagtt atctacacga
3181 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac
3241 tgattaagca ttggtaactg tcagaccaag tttactcata tatacttttag attgatttaa
3301 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca
3361 aatccctta acgtgagttt tctgtccact gagcgtcaga ccccgtagaa aagatcaaa
3421 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaaca aaaaaaccac
3481 cgtaccagc ggtggttgt ttgccggatc aagagctacc aactctttt ccgaaggtaa
3541 ctggctcag cagagcgcag ataccaaata ctgtccttct agttagcgc tagttaggcc
3601 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaate ctgttaccag
3661 tggtctcgc cagtggcgat aagctgtgtc ttaccgggtt ggactcaaga cgatagttac
3721 cggataaggc gcagcggcgc ggctgaacgg ggggttcgtg cacacagccc agcttgagc
3781 gaacgacctc caccgaactg agatacctac agcgtgagca ttgagaaaag gccacgcttc
3841 ccgaaggag aaaggcggac aggtatccgg taagcggcag ggtcggaca ggagagcga
3901 cgaggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcgcacc
3961 tctgacttga gcgtcgattt ttgtgatgct cgtcagggg gcggagccta tggaaaaacg
4021 ccagcaacgc ggcttttta cggttcctgg ccttttgctg gccttttct cacatgttct
4081 ttcctgcgtt atcccctgat tctgtggata accgtattac gcctttgag tgagctgata
4141 ccgctcgcgc cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc
4201 gcccaatag caaacgcct ctcgccgcgc gttggccgat tcattaatgc agaattaatt
4261 ctcatgtttg acagcttacc atcactgca cgggtcacca atgcttctgg cgtcagcgag
4321 ccacggaag ctgtggtatg gctgtgcagg tctgaaatca ctgcataatt cgtgtcgtc
4381 aaggcgcact cccgttctgg ataattttt ttgcgccgac atcataacgg ttctggcaaa
4441 tttctgaaa tgagctgtg acaattaatc atcggctcgt ataattgtg gaattgtgag
4501 cggataaaaa tttcacacag gaaacagcga tgaattcaga tctaccatg aaggagctgg
4561 tggcccagat gctgcagagg ctgtgcgagc gcggcgcgaa gaactgtctg gccttcggct
4621 tgcgctgct ggacggggcc cgcgggggccc ccccgaggc cttcaccacc agcgtgcgca
4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtgg gggctgctgc
4741 tgcgccgct gggcgcgac gtgctggtc acctgctgac acgtgcgcg ctctttgtgc
4801 tggtggtccc cagctgcgcc taccaggtgt gcgggccgccc gctgtaccag ctgcggcgtg
4861 ccaactcaggc cggccccccg ccacacgcta gtggaccccg aaggcgtctg ggatgcgaac
4921 gggcctggaa ccatagcgtc agggaggccc gggccccct gggcctgcca gcccccgggtg
4981 cgaggaggcg cgggggcagt gccagccgaa gtctgccgtt gcccaagagg cccagcgtg
5041 gcgctgcccc tgagccggag cggacgcccg ttgggcaggg gtcttgggccc caccgggca
5101 ggacgcgtgg accgagtgc cgtggtttct gtgtggtgtc acctgccaga cccgccgaag
5161 aagccacctc ttgagggtg gcgctctctg gcacgcgcca ctcccacca tccgtgggccc
5221 gccagcacc cgcgggcccc ccatccacat cgcggccacc acgtccctgg gacacgcctt
5281 gtcccccggt gtacgcccag accaagcact tctctactc ctgagcgcac aaggagcagc
5341 tgcggccctc ctctctactc agctctctga ggcccagcct gactggcgtc cggaggctcg
5401 tggagaccat ctttctgggt tccagccctt ggatgccagg gactccccgc aggttgcgcc
5461 gcctgcccga gcgctactgg caaatgcggc cctgtttct ggagctgctt ggaaccacg
5521 cgcagtgcgc ctacgggggt ctctcaaga cgcactgccc gctgcgagct gcggtcacc

图 14C

5581 cagcagccgg tgtctgtgcc cgggagaagc cccagggctc tgtggcggcc cccgaggagg
5641 aggacacaga cccccgtcgc ctggtgcagc tgctccgcca gcacagcagc ccctggcagg
5701 tgtacggcct cgtgcggggc tgccctgcgc ggctgggtgcc cccaggcctc tggggctcca
5761 ggcacaacga acgcccgttc ctcaggaaca ccaagaagt catctccctg gggagcatg
5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt gcgggactgc gcttggctgc
5881 gcaggagccc aggggttggc tgtgttccgg ccgcagagca ccgtctgcgt gaggagatcc
5941 tggccaagtt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct
6001 tttatgtcac ggagaccacg tttcaaaaga acaggctctt tttctaccgg aagagtgtct
6061 ggagcaagtt gcaaagcatt ggaatcagac agcactttaa gaggggtgcag ctgcgggagc
6121 tgtcgggaagc agaggtcagg cagcatcggg aagccaggcc cgccctgctg acgtccagac
6181 tccgcttcat cccaagcct gacgggctgc ggccgattgt gaacatggac tacgtcgtgg
6241 gagccagaac gttccgcaga gaaaagaggg ccgagcgtct cacctcgagg gtgaaggcac
6301 tgttcagcgt gctcaactac gagcggggcg ggcgccccgg cctcctgggc gcctctgtgc
6361 tgggcctgga cgatatccac agggcctggc gcacctcgt gctgcgtgtg cgggcccagg
6421 acccgccgcc tgagctgtac tttgtcaagg tggatgtgac gggcgcgtac gacaccatcc
6481 cccaggacag gctcacggag gtcacgcca gcatcatcaa accccagaac acgtactgcg
6541 tgcgctcggta tgccgtggtc cagaaggccg cccatgggca cgtccgcaag gccttcaaga
6601 gccacgtctc taccttgaca gacctccagc cgtacatgcg acagttcgtg gctcacctgc
6661 aggagaccag cccgtgagg gatgccgtcg tcacgagca gagctcctcc ctgaatgagg
6721 ccagcagtgg cctcttcgac gtcttcctac gcttcattgt ccaccagccg gtgocgatca
6781 ggggcaagtc ctacgtccag tgccagggga tcccgcaggg ctccatctc tccacgtgc
6841 tctgcagcct gtgtacggc gacatggaga acaagctgtt tgggggatt cggcgggagc
6901 ggctgtcctc gcgtttggtg gatgattct tgttggtagc acctcacctc acccacgga
6961 aaacttctc aggacctggt ccgaagtgtc ctgagtatgg ctgcgtggtg aacttgcgga
7021 agacagtggg gaacttccct gtagaagaag aagccctggg tggcacggct tttgttcaga
7081 tgccggccca cggcctattc ccctgggtgcg gcctgctgct ggatacccg accctggagg
7141 tgacagcga ctactccagc tatgcccgga cctccatcag agccagtctc accttcaacc
7201 gcggcttcaa ggctgggagg aacatgcgtc gcaaaactct tggggctctg cggtgaagt
7261 gtcacagcct gttctggat ttgcaggta acagcctca gacggtgtgc accaacatct
7321 acaagatcct cctgctgcag gcgtacaggt ttcacgatg tgtgtgcag ctcccattc
7381 atcagcaagt ttggaagaac cccacatttt tcttgcgct catctctgac acggcctccc
7441 tctgctactc catcctgaaa gccaaagaac cagccgaaga aaacatttct gctgtgactc
7501 ctgcggtgct tgggtcggga cagccagaga tggagccacc ccgcagaccg tcgggtgtgg
7561 gcagctttcc ggtgtctcct gggaggggag ttgggctggg cctgtgactc ctcagcctct
7621 gttttcccc agggatgtcg ctgggggcca agggcgcgc cggccctctg cctccgagg
7681 ccgtgcagtg gctgtgccac caagcattcc tgctcaagct gactcgacac cgtgtcacct
7741 acgtgccact cctggggtca ctcaggacag gcaagtgtgg gtggaggcca gtgcggg

图 14D

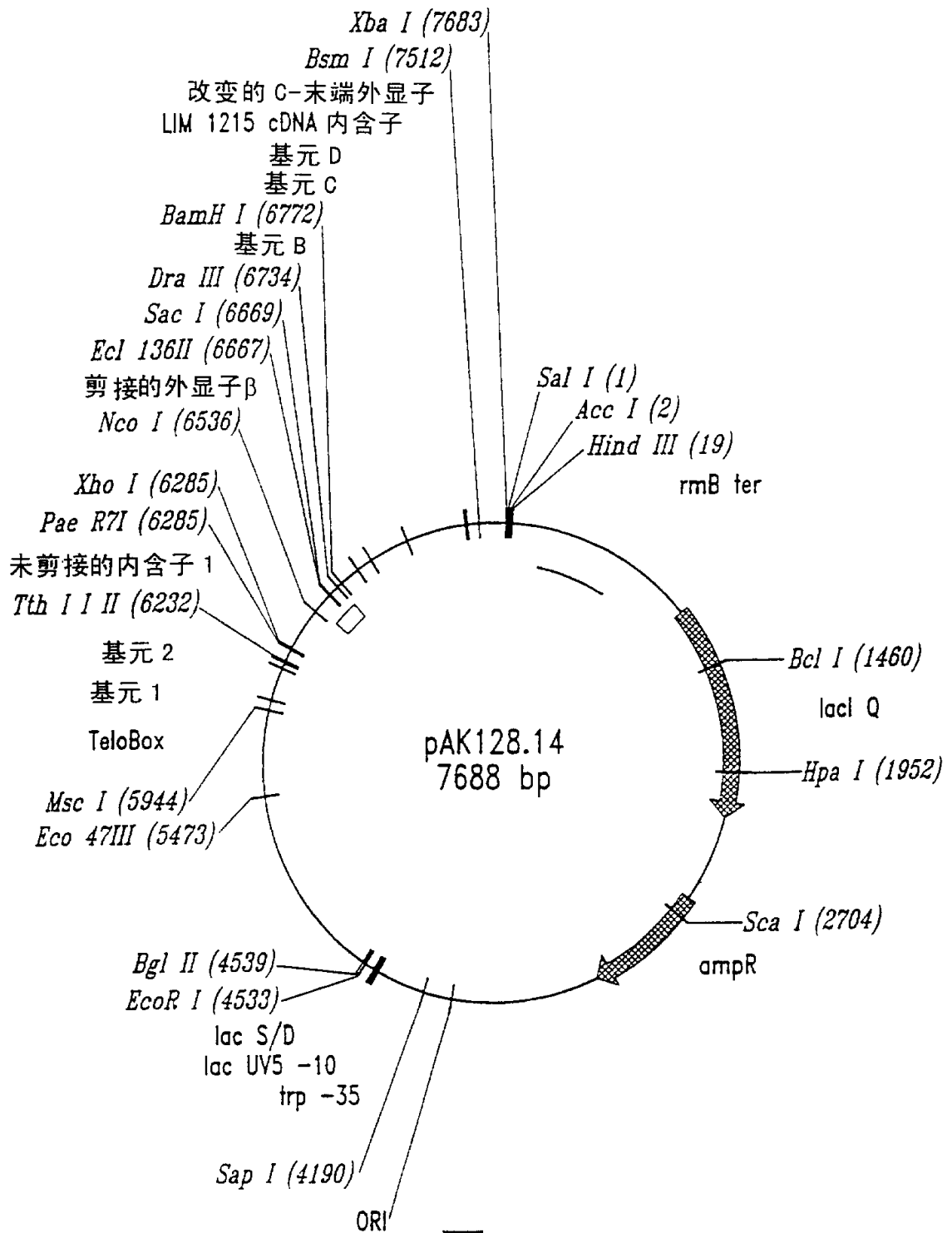


图 15A

基因座 _pAKI28.14 7688 bp dsDNA 环状
 说明 带有剪接出外显子 α 的人端粒酶

```

1  tcgacctgca  ggcatgcaag  cttggcactg  gccgtcgttt  tacaacgtcg  tgactgggaa
61  aaccttgccg  ttaccaact  taatgcctt  gcagcacatc  ccccttcgc  cagctggcgt
121 aatagcgaag  aggcccgcac  cgatgccct  toccaacagt  tgcgcagcct  gaatggcgaa
181 tggcgccctg  tgcggtatit  tctccttacg  catctgtgcg  gtattcaca  cgcataaat
241 tccctgtttt  ggcggatgag  agaagatit  cagcctgata  cagattaaat  cagaacgcag
301 aagcggctcg  ataaaacaga  atttgcctgg  cggcagtagc  gcggtggctc  caccgaccc
361 catgccgaac  tcagaagtga  aacgccgtag  cgccgatggt  agtgtgggt  ctccccatgc
421 gagagtaggg  aactgccagg  catcaataa  aacgaaaggc  tcagtcgaaa  gactgggctt
481 ttctgtttat  ctgttgtttg  tcggtgaacg  ctctcctgag  taggacaaat  ccgccgggag
541 cggatttgaa  cgttgcaag  caacggccc  gaggtggcg  ggcaggacgc  ccgccataaa
601 ctgccaggca  tcaaatgaag  cagaaggcca  tcctgacgga  tggcctttt  gcgtttctac
661 aaactcttc  tgtcgtcata  tctacaagcc  atccccccac  agatacggta  aactagcctc
721 gtttttgc  caggaaagca  gggaatttt  ggtgcaactc  cagtacaatc  tgctctgatg
781 ccgcatagtt  aagccagccc  cgacaccgc  caacaccgc  tgacgcgccc  tgacgggctt
841 gtctgtctcc  ggcacccgct  tacagacaag  ctgtgaccgt  ctccgggagc  tgcattgttc
901 agaggttttc  accgtcatca  ccgaaacgcg  cgagacgaaa  gggcctcgtg  ataccctat
961 ttttataggt  taatgtcatg  ataataatgg  tttcttagac  gtgaggtct  gtaccgaca
1021 ccatcgaatg  gtgcaaac  tttcgggta  tggcatgata  gcgccggaa  gagagtcaat
1081 tcagggtggt  gaatgtgaaa  ccagtaacgt  tatacagatg  cgcagagtat  gccggtgtct
1141 cttatcagac  cgtttccgc  gtgtgaacc  aggccagcca  cgtttctgcg  aaaacgggg
1201 aaaaagtgga  agcggcgatg  gcggagctga  attacattcc  caaccgctg  gcacaacaac
1261 tggcgggcaa  acagtgttg  ctgattggcg  ttgccacctc  cagtctggcc  ctgacgcgc
1321 cgtcgcaaat  tgcgcggcg  ataaatctc  gcgccgatca  actgggtgcc  agcgtggtg
1381 tgcgatggt  agaacgaagc  ggcgtcgaag  cctgtaaagc  ggcggtgac  aatcttctcg
1441 cgcaacgcgt  cagtggctg  atcattaact  atccgctgga  tgaccaggat  gccattgctg
1501 tggaaactgc  ctgcactaat  gttccggcgt  tatttcttga  tgtctctgac  cagacaccca
1561 tcaacagtat  tatttctcc  catgaagacg  gtacgcgact  ggcgtggag  catctggtcg
1621 cattgggtca  ccagcaaatc  gcgctgttag  cgggccatt  aagttctgtc  tcggcgcgtc
1681 tgcgtctggc  tggctggcat  aaatatctca  ctgcaatca  aattcagccg  atagcggaac
1741 ggaagggcga  ctggagtgc  atgtccggtt  ttcaacaac  catgcaaatg  ctgaatgagg
1801 gcategttcc  cactgcgatg  ctggttgcca  acgatcagat  ggcgctggc  gcaatgcgcg
1861 ccattaccga  gtccggctg  cgcgttggtg  cggatatctc  ggtagtggga  tacgacgata
1921 ccgaagacag  ctcatgttat  atcccccggt  taaccacat  caaacaggat  tttcgcctgc
1981 tggggcaaac  cagcgtggac  cgcttgctgc  aactctctca  gggccaggcg  gtgaagggca
2041 atcagctgtt  gccgtctca  ctggtgaaa  gaaaaaccac  cctggcgc  aatacgcaaa
2101 ccgcctctcc  ccgcgcttg  gccgattcat  taatgcagct  ggcacgacag  gttccccgac
2161 tggaaagcgg  gcagtgagcg  caacgcaatt  aatgtaagt  agtctactca  ttaggcaccc
2221 caggctttac  actttatgct  tccgacctgc  aagaacctca  cgtcagggtg  cacttttogg
2281 ggaaatgtgc  gcggaacccc  tattgttta  ttttctaaa  tacattcaaa  tatgtatccg
2341 ctcatgagac  aataaccctg  ataatgctt  caataatatt  gaaaaaggaa  gagtatgagt
2401 attcaacatt  tccgtgtcgc  cttattccc  tttttgcg  cttttgcct  tctgttttt
2461 gctcaccag  aaacgctggt  gaaagtata  gatgctgaag  atcagttggg  tgcacgagtg
2521 gttacatcg  agaactggat  ctcaacagcg  gtaagatcct  tgagagttt  cccccgag
2581 aacgtttcc  aatgatgagc  acttttaaag  ttctgctatg  tggcgggta  ttatccgta
2641 ttgacgccg  gcaagagcaa  ctccgtgcc  gcatacacta  ttctcagaat  gacttgggtg

```

图 15B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca
 2761 gtgctgccat aaccatgagt gataaactg cggccaactt acttctgaca acgatcggag
 2821 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccctgatc
 2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg
 2941 tagcaatggc aacaacgttg cgcaactat taactggcga actacttact ctagcttccc
 3001 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg
 3061 cccctccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggctctcgg
 3121 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagt tcttacacga
 3181 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac
 3241 tgattaagca ttgtaactg tcagaccaag tttactcata tatactttag attgatttaa
 3301 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca
 3361 aaatccctta acgtgagttt tgcgtccact gagcgtcaga ccccgtagaa aagatcaaa
 3421 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaaca aaaaaaccac
 3481 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactctttt cgaaggtaa
 3541 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc
 3601 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaact ctgttaccag
 3661 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagtac
 3721 cggataaggc gcagcggcgc ggctgaacgg ggggttcgtg cacacagccc agcttgagc
 3781 gaacgacctc caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgctc
 3841 ccgaaggag aaaggcggac aggtatccgg taagcggcag ggtcggaaaca ggagagcga
 3901 cgaggagcct tccaggggga aacgcctggt atctttatag tccgtcggg ttcgccacc
 3961 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gccggagccta tggaaaaacg
 4021 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgc tcatgttct
 4081 ttctgcggtt atcccctgat tctgtggata accgtattac gccttttgag tgagctgata
 4141 ccgctcgcgc cagccgaacg accgagcgcga gcgagtcagt gagcggaggaa gcggaagagc
 4201 gcccaatacg caaacgcct ctcccgcgc gttggccgat tcattaatgc agaattaatt
 4261 ctcatgtttg acagcttatc atcgaactga cggtgaccaca atgcttctgg cgtcaggcag
 4321 ccatcggag ctgtggtatg gctgtgcagg tcgtaaatca ctgcataatt cgtgtcgtc
 4381 aaggcgcact cccgttctgg ataatgtttt ttgcgccgac atcataacgg tcttgcaaa
 4441 tattctgaaa tgagctgtg acaattaatc atcggctcgt ataatgtgtg gaattgtgag
 4501 cggataacaa tttcacacag gaaacagcga tgaattcaga tctcaccatg aaggagctgg
 4561 tggcccagat gctgcagagg ctgtgcgagc gcggcgcgaa gaacgtgctg gccttcggct
 4621 tcgctcgtct ggacggggcc cgcgggggccc cccccgaggc cttcaccacc agcgtgcga
 4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtgg gggctgctg
 4741 tgcgcccgcgt gggcgcgac gctgtggttc acctgctggc acgctgcgcg ctctttgtg
 4801 tgggtgctcc cagctgcgcc taccaggtgt gcgggccgcc gctgtaccag ctccggcctg
 4861 ccaactcaggc cggccccccg ccacacgcta gtggaccccg aaggcgtctg ggatgcgaac
 4921 gggcctggaa ccatagcgtc agggaggccg gggccccct gggcctgcca gccccgggtg
 4981 cgaggaggcg cgggggcagt gccagccgaa gtctgccgtt gcccaagagg cccaggcgtg
 5041 gcgctgcccc tgagccggag cggacgcccg ttgggcaggg gtccctgggccc cccccggca
 5101 ggacgcgtgg accgagtgac cgtggtttct gtgtggtgtc acctgccaga cccgcgaag
 5161 aagccacctc tttggagggt ggcctctctg gcacgcgcca ctcccacca tccgtgggccc
 5221 gccagacca cgcgggcccc ccatccacat cgcggccacc acgtccctgg gacacgcctt
 5281 gtccccgggt gtacgccgag accaagcact tctctactc ctcaggcgac aaggagcagc
 5341 tgcggccctc cttctactc agctctctga ggcccagcct gactggcgtc cggaggctcg
 5401 tggagaccat cttctgggt tccaggccct ggatgccagg gactccccgc aggttgcccc
 5461 gcctgccccca gcgctactgg caaatgcggc ccctgtttct ggagctgctt gggaccacg
 5521 cgcagtgccc ctacggggtg ctctcaaga cgcactgccc gctgcgagct gcggtcacc

图 15C

5581 cagcagccgg tgtctgtgcc cgggagaagc cccagggctc tggggggcc cccgaggagg
5641 aggacacaga cccccgtgc ctggtgcagc tgctccgcca gcacagcagc ccctggcagg
5701 tgtacggcct cgtgcggggc tgctgcgcc ggctgggtgcc cccaggcctc tggggctcca
5761 ggcacaacga acgccgcttc ctcaggaaca ccaagaagt catctcccig gggaaagcatg
5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt gcgggactgc gcttggctgc
5881 gcaggagccc aggggttggc tggttccgg ccgcagagca ccgtctgcgt gaggagatcc
5941 tggccaagt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct
6001 tttatgtcac ggagaccag ttcaaaaaga acaggctctt tttctaccgg aagagtgtct
6061 ggagcaagt gcaaagcatt ggaatcagac agcacttgaa gagggtgag ctgcgggagc
6121 tgtcgggaag agaggtcagg cagcatcggg aagccaggcc cgccctgctg acgtccagac
6181 tccgcttcat cccaagcct gacgggctgc ggccgattgt gaacatggac tacgtcgtgg
6241 gagccagaac gttccgcaga gaaaagaggg ccgagcgtct cacctcgagg gtgaaggcac
6301 tgttcagcgt gctcaactac gagcggggcg ggcgccccgg cctcctgggc gcctctgtgc
6361 tgggcctgga cgatatccac agggcctggc gcaccttcgt gctgcgtgtg cgggcccagg
6421 acccgccgcc tgagctgtac tttgtcaagg acaggctcac ggaggtcatc gccagcatca
6481 tcaaaccag aacacgtact gcgtgcgtcg gtatgccgtg gtccagaagg ccgcccatgg
6541 gcacgtccgc aaggccttca agagccacgt ctctacctg acagacctc agccgtacat
6601 gcgacagttc gtggctcacc tgcaggagac cagcccgtg agggatgcc tcgtcatcga
6661 gcagagctcc tccctgaatg aggccagcag tggcctcttc gacgtcttc tacgttcat
6721 gtgccaccac gccgtgcgca tcaggggcaa gtcctacgtc cagtgccagg ggatcccga
6781 gggctccatc ctctccacgc tgctctgcag cctgtgtac ggcgacatgg agaacaagct
6841 gtttgccggg attcggcggg acgggctgct cctgcgttg gtggatgatt tcttgttgg
6901 gacacctcac ctacccacg cgaaaacct cctcaggacc ctggtccgag gtgtccctga
6961 gtatggctgc gtggtgaact tgcggaagac agtggtgaac ttccctgtag aagacgaggc
7021 cctgggtggc acggctttg ttcagatgcc ggcccacggc ctattcccct ggtgcggcct
7081 gctgtggat acccgacc caggagtgca gagcgactac tccagctatg cccggacctc
7141 catcagagcc agtctcacct tcaaccgagg cttcaaggct gggaggaaca tgcgtcgcaa
7201 actcttggg gtcttgccgc tgaagtgtca cagcctgtt ctggatttgc aggtgaacag
7261 cctccagacg gtgtgcacca acatctacaa gatcctctg ctgcaggcgt acaggtttca
7321 cgcatgtgtg ctgcagctcc catttcatca gcaagtgtg aagaacccca catttttct
7381 gcgctcatc tctgacacgg cctccctctg ctactccatc ctgaaagcca agaacgcagg
7441 gatgtcgtg ggggccaagg gcgccgccg cctctgccc tccgaggccg tgcagtggct
7501 gtgccaccaa gcattcctgc tcaagctgac tgcacaccgt gtcacctac tgccactcct
7561 ggggtcactc aggacagccc agacgcagct gagtccgaag ctcccgggga cgacgtgac
7621 tgccctggag gccgcagcca accgggcact gccctcagac ttcaagacca tcttgactg
7681 atctagag

图 15D