

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公开说明书

[21] 申请号 200610001356.2

[51] Int. Cl.

C12N 9/00 (2006.01)

C12N 15/52 (2006.01)

C12N 15/33 (2006.01)

C12N 1/21 (2006.01)

C12N 15/63 (2006.01)

C12N 15/09 (2006.01)

[43] 公开日 2006年7月26日

[11] 公开号 CN 1807605A

[51] Int. Cl. (续)

C12R 1/19 (2006.01)

[22] 申请日 2006.1.19

[21] 申请号 200610001356.2

[71] 申请人 中国科学院生物物理研究所

地址 100101 北京市朝阳区大屯路中国科学院生物物理研究所

[72] 发明人 高光侠 刘树峰

[74] 专利代理机构 北京纪凯知识产权代理有限公司
代理人 关 畅

权利要求书 1 页 说明书 18 页 附图 2 页

[54] 发明名称

小鼠白血病病毒逆转录酶突变体及其表达方法与应用

[57] 摘要

本发明公开了小鼠白血病病毒逆转录酶突变体及其表达方法与应用。该小鼠白血病病毒逆转录酶突变体，是将小鼠白血病病毒逆转录酶的自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于谷氨酰胺侧链的氨基酸。本发明的小鼠白血病病毒逆转录酶突变体具有比 RT-F155V-H 更高的 RNA 聚合酶比活性和 RNA 合成能力。

1、小鼠白血病病毒逆转录酶突变体，是将小鼠白血病病毒逆转录酶的自 N 端第 155 位苯丙氨酸残基取代为缬氨酸、自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基和自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于谷氨酰胺侧链的氨基酸。

2、根据权利要求 1 所述的小鼠白血病病毒逆转录酶突变体，其特征在于：所述 X 为丙氨酸、天门冬酰胺、天门冬氨酸或丝氨酸。

3、根据权利要求 2 所述的小鼠白血病病毒逆转录酶突变体，其特征在于：所述 X 为丙氨酸；所述小鼠白血病病毒逆转录酶突变体具有序列表中序列 2 的氨基酸残基序列。

4、权利要求 1 至 3 中任一权利要求所述的小鼠白血病病毒逆转录酶突变体的编码基因。

5、根据权利要求 4 所述的编码基因，其特征在于：所述小鼠白血病病毒逆转录酶突变体具有序列表中序列 2 的氨基酸残基序列；所述小鼠白血病病毒逆转录酶突变体编码基因具有序列表中序列 1 的自 5' 端第 1515 位至第 3527 位脱氧核苷酸的核苷酸序列。

6、一种表达权利要求 1 所述的小鼠白血病病毒逆转录酶突变体的方法，是将含有权利要求 1 所述的小鼠白血病病毒逆转录酶突变体编码基因的表达载体转化到大肠杆菌中，培养阳性克隆，表达得到小鼠白血病病毒逆转录酶突变体。

7、根据权利要求 6 所述的方法，其特征在于：所述含有权利要求 1 所述的小鼠白血病病毒逆转录酶突变体编码基因的表达载体是具有序列表中序列 1 的核苷酸序列的质粒 pTacRT-Q84A-F155V-D524N。

8、根据权利要求 6 或 7 所述的方法，其特征在于：所述大肠杆菌为 *Escherichia coli* BL21。

9、含有权利要求 4 或 5 所述的小鼠白血病病毒逆转录酶突变体的编码基因的表达载体，细胞系或宿主菌。

10、权利要求 1 至 3 中任一权利要求所述的小鼠白血病病毒逆转录酶突变体在合成 RNA 中的应用。

小鼠白血病病毒逆转录酶突变体及其表达方法与应用

技术领域

本发明涉及一种逆转录酶突变体及其表达方法与应用，特别涉及小鼠白血病病毒逆转录酶突变体及其表达方法，与该小鼠白血病病毒逆转录酶突变体在合成RNA中的应用。

背景技术

根据对底物的选择特异性，传统上把核酸聚合酶分为DNA聚合酶和RNA聚合酶两大类。聚合酶对底物的选择通常是非常严格的，因为它们分别承担着遗传物质的复制和转录两个生物学最基本的功能。

然而对聚合酶的分类可能并非象想像的那么简单。DNA聚合酶和RNA聚合酶尽管在一级结构上没有同源性，但是它们的聚合酶结构域的高级结构却非常类似，呈一个半张开的右手状，包含有手掌，手指以及拇指亚结构域。另外，它们的催化机制也非常的类似。而对作为底物的脱氧核糖核苷酸（dNTPs）和核糖核苷酸（rNTPs）而言，它们的唯一区别就是rNTPs的2'-OH基团。这说明DNA聚合酶和RNA聚合酶之间可能存在着某种关联，或者说它们在进化上存在着某种相关性，有可能通过它们结构上的一个小的改变，来实现DNA聚合酶和RNA聚合酶的角色转换。

小鼠白血病病毒（Murine Leukemia Virus，简称MLV）逆转录酶（Reverse Transcriptase，简称RT，是由逆转录病毒编码的DNA聚合酶，能以RNA或DNA作为模板合成DNA。Guangxia Gao等于1997年发表在Proceedings of the National Academy of sciences of the United States of America第407-411页的文章（专利号：US6136582）表明，当将MLV-RT的自N-端155位的苯丙氨酸（F155）突变成缬氨酸（V）时，RT-F155V能够利用RNA或者DNA作为模板将rNTPs参入到产物，也就是说RT具有了RNA聚合酶的活性。RT-F155V-H带有F155V和D524N双突变的MMLV RT，F155V突变使RT具有了参入rNTPs的能力，D524N突变将消除MMLV RT的RNase H活性，这样可以避免RT的RNase H活性对RNA产物的降解。

但是RT-F155V-H催化RNA合成的效率仍然很低，而且合成的RNA片段也很短。

发明内容

本发明的一个目的是提供小鼠白血病病毒逆转录酶突变体。

本发明所提供的小鼠白血病毒逆转录酶突变体，是将小鼠白血病毒逆转录酶的自 N 端第 155 位苯丙氨酸残基取代为缬氨酸、自 N 端第 524 位天门冬氨酸残基取代为天门冬酰胺残基和自 N 端第 84 位谷氨酰胺残基取代为氨基酸残基 X 的蛋白质，其中 X 为侧链短于谷氨酰胺侧链的氨基酸。该小鼠白血病毒逆转录酶突变体的名称为 RT-Q84X-F155V-H。

所述小鼠白血病毒逆转录酶 (MLV-RT) 的氨基酸序列如序列 3 所示。

所述 X 可为丙氨酸、天门冬酰胺、天门冬氨酸或丝氨酸。

所述 X 为丙氨酸；所述小鼠白血病毒逆转录酶突变体具有序列表中序列 2 的氨基酸残基序列，该小鼠白血病毒逆转录酶突变体的名称为 RT-Q84A-F155V-H。

上述小鼠白血病毒逆转录酶突变体的编码基因也属于本发明的保护范围。

所述小鼠白血病毒逆转录酶突变体具有序列表中序列 2 的氨基酸残基序列；所述小鼠白血病毒逆转录酶突变体编码基因 (RT-Q84A-F155V-H) 具有序列表中序列 1 的自 5' 端第 1515 位至第 3527 位脱氧核苷酸的核苷酸序列。

本发明的另一个目的是提供一种表达 RT-Q84A-F155V-H 的方法。

该表达 RT-Q84X-F155V-H 的方法，是将含有 RT-Q84X-F155V-H 编码基因的表达载体转化到大肠杆菌中，培养阳性克隆，表达得到小鼠白血病毒逆转录酶突变体。

所述含有 RT-Q84X-F155V-H 编码基因的表达载体是具有序列表中序列 1 的核苷酸序列的质粒 pTacRT-Q84A-F155V-D524N。

所述大肠杆菌优选为 *Escherichia coli* BL21。

本发明利用蛋白质工程方法，将一种 DNA 聚合酶小鼠白血病毒逆转录酶 (MLV-RT) 改造成具有较高酶活性和 RNA 合成能力的 RNA 聚合酶。实验证明 RT-Q84X-F155V-H 具有比 RT-F155V-H 更高的 RNA 聚合酶比活性和 RNA 合成能力。

附图说明

图 1 是 RT-Q84A-F155V-H 纯化后的 SDS-PAGE 电泳检测图。

图 2 是 RT-Q84A-F155V-H 和 RT-F155V-H 的 RNA 聚合酶活性的比较图。

图 3 是 RT-Q84A-F155V-H 和 RT-F155V-H 合成 RNA 片段能力的比较图。

具体实施方式

本发明通过分析 MLV-RT 的晶体结构模型，发现该酶第 84 位的谷氨酰胺 (Q84) 位于酶的活性中心附近，参与调控酶的催化能力，其较长的侧链妨碍合成产物的

延伸,影响 RNA 聚合酶的活性和合成 RNA 的能力。本发明将这个位于酶的活性中心的大的氨基酸残基替换成小的氨基酸残基,协助这些 DNA 聚合酶更好的容纳 RNA 产物。本发明将 RT-F155V-H 基因进行突变,使 Q84 残基突变成侧链较短的残基,产生了新的突变酶 RT-F155V-Q84X,其中 X 为侧链短于谷氨酰胺侧链的氨基酸,如天冬酰胺(Asn)、丝氨酸(Ser)、丙氨酸(Ala)等。同时,本发明将 Q84X 突变引入 RT-F155V-H 中,使之成为 RT-Q84X-F155V-H。本发明比较了将 MLV-RT 中第 84 位的谷氨酰胺突变为丙氨酸后产生的 RT-Q84A-F155V-H 和 RT-F155V-H 的 RNA 聚合酶活性及 RNA 合成能力,实验证明 RT-Q84A-F155V-H 具有比 RT-F155V-H 更高的 RNA 聚合酶比活性和 RNA 合成能力。

本发明还提供了一种在大肠杆菌 *E. coli* 中表达重组 RT-Q84X-F155V-H 逆转录酶的方法。

下面通过优选实例对本发明做更详细的描述。下述实施例中的实验方法,如无特别说明,均为常规方法。下述实施例中的百分含量,如无特别说明,均为质量百分含量。

实施例 1、RT-Q84A-F155V-H 的合成及其功能

本实施例中将 RT-F155V-H (Guangxia Gao 等于 1997 年发表在 Proceedings of the National Academy of sciences of the United States of America 第 407-411 页的文章,专利号:US6136582)中的第 84 位的谷氨酰胺突变为丙氨酸,使之成为 RT-Q84A-F155V-H。

1、质粒 pTacRT-Q84A-F155V-D524N 的构建

RT-Q84A-F155V-H 的定点突变采用将两段 PCR 产物 Af1III-EcoRI 和 EcoRI-MfeI 替换掉 pTacRT-F155V-D524N 的 Af1III-MfeI 片段(nt1467-2280)。引物 Q84A-SP (5' CGGAATTCTGGTACCCTGCCAGTC)包含有一个由于同义突变产生的 EcoRI 酶切位点(用下划线表示),它和下游引物 PCR 扩增得到一个 300bp 的 EcoRI-MfeI 片段。引物 Q84A-AP (5' CGGAATTCCC**CGTCCAACAGTCTCTGTA**)同样包含有一个由于同义突变产生的 EcoRI 酶切位点以及突变的丙氨酸密码子(黑体表示),它和上游引物经 PCR 扩增得到一个 500bp 的 Af1III-EcoRI 片段。所得重组质粒通过酶切鉴定为阳性克隆。具体方法如下:

(1) EcoRI-MfeI 片段的合成

以 pTacRT-F155V-D524N (Guangxia Gao 等于 1997 年发表在 Proceedings of the National Academy of sciences of the United States of America 第 407-411 页的文章, 专利号: US6136582) 为模板, 利用上游引物 Q84A-SP 和下游引物: *Mfe* I- AP (5' -TGGGAGTCTGGTCCAGG-3' PCR 扩增得到 300bp 的 EcoRI-*Mfe*I 片段。其中, PCR 的温度条件为: 先 95°C 预变性 2 min; 然后 95°C 30 s, 55°C 1 min, 72°C 1 min, 20 个循环; 最后 72°C, 10 min 补平末端。

(2) AflIII-EcoRI 片段的合成

以 pTacRT-F155V-D524N (Guangxia Gao 等于 1997 年发表在 Proceedings of the National Academy of sciences of the United States of America 第 407-411 页的文章, 专利号: US6136582) 为模板, 利用上游引物: *Afl* II- SP

(5' -GTGGAATTGTGAGCGGA-3') 和下游引物 Q84A-AP PCR 扩增得到一个 500bp 的 AflIII-EcoRI 片段。其中, PCR 的温度条件为: 先 95°C 预变性 2 min; 然后 95°C 30 s, 55°C 1 min, 72°C 1 min, 20 个循环; 最后 72°C, 10 min 补平末端。

(3) 质粒 pTacRT-Q84A-F155V-D524N 的构建

将经 AflIII 和 EcoRI 酶切后的 AflIII-EcoRI 片段, EcoRI 和 *Mfe*I 酶切后的 EcoRI-*Mfe*I 片段和经 AflIII 和 *Mfe*I 酶切 pTacRT-F155V-D524N (Guangxia Gao 等于 1997 年发表在 Proceedings of the National Academy of sciences of the United States of America 第 407-411 页的文章, 专利号: US6136582) 后得到的 6.7kb 片段连接, 得到 AflIII-EcoRI 和 EcoRI-*Mfe*I 替换掉 pTacRT-F155V-D524N 的 AflIII-*Mfe*I 片段 (nt1467-2280) 的质粒 pTacRT-Q84A-F155V-D524N, 测序表明 pTacRT-Q84A-F155V-D524N (序列 1) 含有具有序列列表中序列 1 的自 5' 端第 1515 位至第 3527 位脱氧核苷酸的小鼠白血病病毒逆转录酶突变体编码基因 RT-Q84A-F155V-H, 编码具有序列列表中序列 2 的氨基酸序列的小鼠白血病病毒逆转录酶突变体 RT-Q84A-F155V-H。

2、RT-Q84A-F155V-H 在大肠杆菌中的诱导和表达

将 pTacRT-Q84A-F155V-D524N 转化 *Escherichia coli* BL21, 挑取单菌落接种于含 100 μg/ml 氨苄青霉素的 LB 液体培养基中, 37°C 培养至 OD₆₀₀ ≈ 0.5 时, 加入 IPTG 至终浓度 0.5mM, 37°C 继续培养 2-3 小时, 收集菌体, 将菌体用预冷的 50 mM Tris-HCl (pH 7.5) 洗涤一次后备用。

3、RT-Q84A-F155V-H 的纯化

将收集的菌体悬浮于缓冲液 A (20mM PBS, pH 7.4, 0.5M NaCl), 冻融后, 加入

溶菌酶至终浓度 0.5mg/ml, 4°C 消化 30min, 超声波破碎 3 次, 每次 20 秒, 离心, 收集上清。经过金属螯合层析柱 HiTrap chelating HP column(购自 Pharmacia) 分离纯化后, 纯度即达 80%以上, 收集活性蛋白峰, 经过离子交换层析柱 MonoS(购自 Pharmacia) 进一步分离纯化后, 电泳检测考马斯亮蓝染色, 结果显示该蛋白为一分子量约为 76KD 的均一条带, 未见杂质(图 1), 达到对 RT 进行酶学研究分析的要求。图 1 中各泳道分别为: M: 分子量标准; 1: RT-Q84A-F155V-H; 2: RT-F155V-H (Guangxia Gao 等于 1997 年发表在 Proceedings of the National Academy of sciences of the United States of America 第 407-411 页的文章; US6136582)。

4、RT-Q84A-F155V-H 的 RNA 聚合酶活性测定及动力学分析

50 μ l 反应体系中包括 40ng RT 纯酶样品, 60 mM Tris.HCl (pH 8.0), 75 mM NaCl, 0.7 mM MnCl₂, 5 mM DTT, 12 μ g/ml Poly(rA) 模板, 6 μ g/ml oligo(dT)₁₈ 引物, 10 μ Ci/ml (1Ci=37 GBq) ³²P 标记的 UTP 以及 12 μ M 未标记的 UTP, 反应于 37°C 进行。分别于反应开始后不同时间点取样 4 μ l 点于 DE81 滤纸, 终止反应。然后将 DE81 滤纸用 2 \times SSC 洗涤三次, 95%乙醇漂洗两次, 将滤纸晾干后, 放射自显影。定量分析采用液闪计数仪测定。

动力学分析采用公知的双倒数作图法进行。用每个样品中“保留在 DE81 滤纸的放射强度/同一样品总的放射强度”来表示参与到产物中的 UTP 的量。

选用 poly(rA)-oligo(dT)₁₈ 作为模板-引物, rUTP 为底物对这三个变体酶进行 RNA 聚合酶活性分析, 结果显示, 当将 Q84 替换成 A 后, 能提高 RT-F155V-H 的 RNA 聚合酶的活性(图 2, 比较 RT-Q84A-F155V-H 和 RT-F155V-H)。将 Q84 突变成天冬酰胺(N) (比 Q 少一个甲基基团)得到的 RT-Q84N-F155V-H (RT-Q84N-F155V-H 也由 671 个氨基酸残基组成, 除自氨基端第 84 位氨基酸残基为 Asn 外, 其它氨基酸残基与序列 2 相同, 可参照 RT-Q84A-F155V-H 的表达方法进行制备), 同样能提高 RT-F155V-H 的 RNA 聚合酶活性(图 2, RT-Q84N-F155V-H)。作为对照, 将 Q84 替换成大的氨基酸残基如 RT-Q84R-F155V-H (RT-Q84R-F155V-H 也由 671 个氨基酸残基组成, 除自氨基端第 84 位氨基酸残基为 Arg 外, 其它氨基酸残基与序列 2 相同, 可参照 RT-Q84A-F155V-H 的表达方法进行制备)时, 则不能提高 RT-F155V-H 的 RNA 聚合酶活性(图 2, RT-Q84R-F155V-H)。需要指出的是, 是 F155V 突变使得 MMLV RT 具有了 RNA 聚合酶活性, 只带有 Q84A 和 D524N 双突变的 RT-Q84A-H

(RT-Q84A-H 也由 671 个氨基酸残基组成, 除自氨基端第 84 位氨基酸残基为 Ala、第 155 位氨基酸残基为 Phe、第 524 位氨基酸残基为 Asn 外, 其它氨基酸残基与序

列 2 相同, 可按常规方法制备) 与 RT-WT-H (RT-WT-H 也是由 671 个氨基酸残基组成, 除自氨基端第 84 位氨基酸残基为 Gln、第 155 位氨基酸残基为 Phe、第 524 位氨基酸残基为 Asn 外, 其它氨基酸残基与序列 2 相同, 可按常规方法制备) 一样, 不具有 RNA 聚合酶活性。

动力学分析显示, 在以 UTP 为底物时, RT-Q84A-F155V-H 和 RT-F155V-H 具有相似的 K_m (分别为 $2.98 \mu\text{M}$ 和 $2.88 \mu\text{M}$), 而 RT-Q84A-F155V-H 的最大反应速率 V_{max} 值则是 RT-F155V-H 的 4.3 倍 (分别为 $9.94 \text{ nmol} \cdot \text{min}^{-1} \cdot \text{ng}^{-1}$ 和 $2.31 \text{ nmol} \cdot \text{min}^{-1} \cdot \text{ng}^{-1}$) (见表 1)。

表 1. RT-Q84A-F155V-H 与 RT-F155V-H RNA 聚合酶动力学参数比较。

酶	rUTP	
	V_{max} ($\text{nmol} \cdot \text{min}^{-1} \cdot \text{ng}^{-1}$)	K_m (μM)
RT-F155V-H	2.31 ± 0.30	2.88 ± 0.67
RT-Q84A-F155V-H	9.94 ± 0.90	2.98 ± 0.72

5、RT 合成 RNA 实验

将寡核苷酸 P21 (5' -AAGCCCCACATACAGAGACTG-3') 用 [γ - ^{32}P] ATP 进行末端标记。此标记产物 *P21 先过 G25 柱纯化, 然后与作为模板的一段寡核苷酸 T36 (3' -TTCGGGGTGTATGTCTCTGACAACCTGGTCGTCCT-5') 退火。退火反应先在 65°C 保温 10min, 然后缓慢降温至室温。RNA 合成反应 $60 \mu\text{l}$ 反应体系中包括 $3 \mu\text{g}$ RT 纯酶样品, 60 mM Tris.HCl (pH 8.0), 75 mM NaCl, 0.7 mM MnCl_2 , 5 mM DTT, $0.1 \mu\text{M}$ *P21/T36 以及 $500 \mu\text{M}$ 未标记的超纯的 rNTP (Amersham)。反应于 37°C 进行, 反应开始后于不同时间点取样中止反应。然后通过 23% 尿素 PAGE 分离产物, 放射自显影检测, 采用 PhosphorImager 定量。

和 Gguangxia Gao 等 1997 年发表的结果一致, RT-F155V-H 在反应 120 min 后仍然只能合成 7 个核苷酸的一段 RNA (图 3, 泳道 6 箭头示)。与之相比较, 经过 30 min 反应 RT-Q84A-F155V-H 就能合成出更长的 RNA 产物。并且 RT-Q84A-F155V-H 合成 RNA 的速率要比 RT-F155V-H 快得多 (图 3, 比较第 2 和 7, 3 和 8, 4 和 9, 5 和 10, 6 和 11 泳道)。同样的反应条件下, 以 dNTPs 作为底物充分延伸得到的全长 DNA (36nt) 的迁移速率明显比 RT-Q84A-F155V-H 合成的最长的 RNA 快, 表明

RT-Q84A-F155V-H 的合成产物不是由于 dNTPs 的污染造成的。RT-Q84A-F155V-H 合成的最长 RNA 产物是一段全长的 15-nt RNA (图 3, 第 11 泳道)。图 3 中, 泳道 1 为 *P21 的电泳结果, 泳道 2、3、4、5、6 分别为 RT-F155V-H 反应 5、10、30、60 和 120 分钟的产物电泳结果, 泳道 7、8、9、10、11 分别为 RT-Q84A-F155V-H 反应 5、10、30、60 和 120 分钟的产物电泳结果, 泳道 12 为同样的反应条件下, 以 dNTPs 作为底物 RT-Q84A-F155V-H 合成的 DNA 电泳结果。

序列表

<160> 3

<210> 1

<211> 7488

<212> DNA

<213> 人工序列

<220>

<223>

<400> 1

ccgacacccat cgaatggtgc aaaacctttc gcggtatggc atgatagcgc ccggaagaga	60
gtcaattcag ggtggtgaat gtgaaaccag taacgttata cgatgtcgca gagtatgccg	120
gtgtctctta tcagaccgtt tcccgcgtgg tgaaccaggc cagccacgtt tctgcaaaa	180
cgcgggaaaa agtggaagcg gcgatggcgg agctgaatta cattcccaac cgcggtggcac	240
aacaactggc gggcaaacag tcgttgctga ttggcgttgc cacctccagt ctggcctgc	300
acgcgccgtc gcaaattgtc gcggcgatta aatctcgcgc cgatcaactg ggtgccagcg	360
tggtggtgtc gatggtagaa cgaagcggcg tcgaagcctg taaagcggcg gtgcacaatc	420
ttctcgcgca acgcgtcagt gggctgatca ttaactatcc gctggatgac caggatgcca	480
ttgctgtgga agctgcctgc actaatgttc cggcgttatt tcttgatgtc tctgaccaga	540
cacccatcaa cagtattatt ttctcccatg aagacggtag gcgactgggc gtggagcatc	600
tggtcgcatt gggtcaccag caaatcgcgc tgttagcggg ccattaagt tctgtctcgg	660
cgcgtctcgc tctggctggc tggcataaat atctcactcg caatcaaatt cagccgatag	720
cggaacggga aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgctga	780
atgagggcat cgttcccact gcgatgctgg ttgccaacga tcagatggcg ctgggcgcaa	840
tgcgcgccat taccgagtcc gggctgcgcg ttggtgcgga tatctcggtg gtgggatacg	900
acgataccga agacagctca tgttatatcc cgccgtaac caccatcaaa caggattttc	960
gcctgctggg gcaaaccagc gtggaccgct tgctgcaact ctctcagggc caggcggtga	1020
agggcaatca gctgttgccc gtctcactgg tgaaaagaaa aaccaccctg gcgccaata	1080
cgcaaaccgc ctctccccgc gcgttggcgg attcattaat gcagctggca cgacaggttt	1140

cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag	1200
gcacaattct catgtttgac agcttatcat cgactgcacg gtgcaccaat gcttctggcg	1260
tcaggcagcc atcgggaagct gtggtatggc tgtgcaggtc gtaaatact gcataattcg	1320
tgtegctcaa ggcgcaactcc cgttctggat aatgtttttt gcgccgacat cataacggtt	1380
ctggcaaata ttctgaaatg agctgttgac aattaatcat cggctcgat aatgtgtgga	1440
attgtgagcg gataacaatt tgaattctta agatttgtga ggggataaca atttcacaca	1500
ggaaacagaa tatgacccta aatatagaag atgagcatcg gctacatgag acctcaaag	1560
agccagatgt ttctctaggg tccacatggc tgtctgattt tcctcaggcc tgggcggaaa	1620
ccgggggcat gggactggca gttcgccaag ctctctgat catacctctg aaagcaacct	1680
ctacccccgt gtccataaaa caatacccca tgtcacaaga agccagactg gggatcaagc	1740
cccacataca gagactgttg gacgcgggaa tcctgggtacc ctgccagtcc ccctggaaca	1800
cgccccctgt acccgtaag aaaccaggga ctaatgatta taggcctgtc caggatctga	1860
gagaagtcaa caagcgggtg gaagacatcc accccacct gccaacct tacaacct	1920
tgagcgggct cccaccgtcc caccagtgtt aactgtgct tgatcttaag gatgccgttt	1980
tctgcctgag actccacccc accagtcagc ctctcttcgc ctttgagtgg agagatccag	2040
agatgggaat ctcaggacia ttgacctgga ccagactccc acagggtttc aaaaacagtc	2100
ccaccctgtt tgatgaggca ctgcacagag acctagcaga cttccggatc cagcaccag	2160
acttgatcct gctacagtac gtggatgact tactgctggc cgccaattct gagctagact	2220
gccaacaagg tactcgggcc ctgttataaa ccctagggaa cctcgggtat cgggcctcgg	2280
ccaagaaagc ccaaatttgc cagaaacagg tcaagtatct ggggtatctt ctaaagagg	2340
gtcagagatg gctgactgag gccagaaaag agactgtgat ggggcagcct actccgaaga	2400
cccctcgaca actaaggag ttcttaggga cggcaggctt ctgtcgctc tggatccctg	2460
ggtttgcaga aatggcagcc ccctgtacc ctctaccaaa aacggggact ctgtttaatt	2520
ggggcccaga ccaacaaaag gcctatcaag aatcaagca agctcttcta actgccccag	2580
ccctgggggt gccagatttg actaagcct ttgaactctt tgtcgacgag aagcagggt	2640
acgcaaagg tgtcctaacg caaaaactgg gaccttggcg tcggccgtg gcctacctgt	2700
caaaaagct agaccagta gcagctgggt ggcccccttg cctacggatg gtagcagcca	2760
ttgccgtact gacaaaggat gcaggcaagc taacatggg acagccacta gtcattctgg	2820
cccccatgc agtagaggca ctagtcaaac aacccccga ccgctggctt tccaacgccc	2880
ggatgactca ctatcaggcc ttgcttttgg acacggaccg ggtccagttc ggaccgggtg	2940
tagccctgaa cccggctacg ctgctccac tgcctgagga agggctgcaa cacaactgcc	3000

ttgatatcct ggccgaagcc cacggaaccc gacccgacct aacggaccag ccgctcccag	3060
acgccgacca cacctggtat acgaatggaa gcagtctctt acaagaggga cagcgtaagg	3120
cgggagctgc ggtgaccacc gagaccgagg taatctgggc taaagccctg ccagccggga	3180
catccgctca gcgggctgaa ctgatagcac tcaccaggc cctaaagatg gcagaaggta	3240
agaagctaaa tgtttatact gatagccggt atgcttttgc tactgccccat atccatggag	3300
aatatacag aaggcgtggg ttgctcacat cagaaggcaa agagatcaaa aataaagacg	3360
agatcttggc cctactaaaa gccctctttc tgcccaaaag acttagcata atccattgtc	3420
caggacatca aaaggacac agcgccgagg ctagaggcaa ccggatggct gaccaagcgg	3480
cccgaaggc agccatcaca gagactccag acacctctac cctcctccat caccatcacc	3540
atcactagtc tagagtcgac ctgcaggcaa gcttggcact ggccgctcgtt ttacaacgtc	3600
gtgactggga aaacctggc gttacccaac ttaatcgctt tgcagcacat cccctttcg	3660
ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc	3720
tgaatggcga atggcagctt ggctgttttg gcggatgaga taagattttc agcctgatac	3780
agattaaatc agaacgcaga agcggctctga taaaacagaa tttgcctggc ggcagtagcg	3840
cggtgggtccc acctgacccc atgccgaact cagaagtga acgccgtagc gccgatggta	3900
gtgtggggtc tccccatgcg agagtaggga actgccaggc atcaaataaa acgaaaggct	3960
cagtcgaaag actgggcctt tcgttttatac tgttgtttgt cggatgaacgc tctcctgagt	4020
aggacaaatc cgccgggagc ggatttgaac gttgcgaagc aacggcccgg agggtggcgg	4080
gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat cctgacggat	4140
ggcctttttg cgtttctaca aactcttttt gttatttttt ctaaatacat tcaaatatgt	4200
atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta	4260
tgagtattca acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttctg	4320
tttttgcctc cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac	4380
gagtgggtta catcgaactg gatctcaaca gcggtaagat ccttgagagt ttcgccccg	4440
aagaacgttc tccaatgatg agcactttta aagtctctgct atgtggcgcg gtattatccc	4500
gtgttgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg	4560
ttgagtactc accagtcaca gaaaagcadc ttacggatgg catgacagta agagaattat	4620
gcagtgctgc cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg	4680
gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg	4740
atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc	4800
ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt	4860

cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 4920
cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc 4980
gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 5040
cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct 5100
cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt 5160
taccocggtt gataatcaga aaagcccaa aaacaggaag attgtataag caaatattta 5220
aattgtaaac gttaatat tttgtaaaatt cgcgttaaatt ttttgtaaa tcagctcatt 5280
ttttaaccaa taggccgaaa tcggcaaaat cccttataaa tcaaaagaat agcccagat 5340
agggttgagt gttgttccag tttggaacaa gagtccacta ttaaagaacg tggactccaa 5400
cgtcaaaggc cgaaaaaccg tctatcaggg cgatggccca ctacgtgaac catcacccaa 5460
atcaagtttt ttggggtcga ggtgccgtaa agcactaaat cggaacccta aaggagccc 5520
ccgatttaga gcttgacggg gaaagccggc gaacgtggcg agaaaggaag ggaagaaagc 5580
gaaaggagcg ggcgctaggg cgctggcaag tgtagcggtc acgctgcgcg taaccaccac 5640
accgcccgcg cttaatgccc cgctacaggg cgcgtaaaag gatctaggtg aagatccttt 5700
ttgataatct catgaccaa atccctaac gtgagttttc gttccactga gcgtcagacc 5760
ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct 5820
tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt gccgatcaa gagctaccaa 5880
ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag 5940
tgtagccgta gttagccac cacttcaaga actctgtagc accgcctaca tacctcgctc 6000
tgctaatect gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttg 6060
actcaagacg atagttaccg gataaggcgc agcggtcggg ctgaacgggg gtttcgtgca 6120
cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat 6180
gagaaagcgc cacgcttccc gaaggagaaa agcggacag gtatccggta agcggcaggg 6240
tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcttggtat ctttatagtc 6300
ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc 6360
ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc 6420
cttttgetca catgttcttt cctgcgttat ccctgattc tgtggataac cgtattaccg 6480
cctttgagtg agctgatacc gctcgcgca gccgaacgac cgagcgcagc gagtcagtga 6540
gagaggaagc ggaagagcgc ctgatgcggt attttctcct tacgcatctg tgcggtattt 6600
cacaccgat atggtgcact ctcagtacaa tctgctctga tgccgatag ttaagccagt 6660
atacactccg ctatcgtac gtgactgggt catggctgcg ccccgacacc cgccaacacc 6720

cgctgacgcg ccctgacggg cttgtctgct cccggcatcc gttacagac aagctgtgac 6780
 cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgaggca 6840
 gctgcggtaa agctcatcag cgtggtcgtg cagcgattca cagatgtctg cctgttcatc 6900
 cgcgtccagc tcgttgagtt tctccagaag cgттаатgtc tggcttctga taaagcgggc 6960
 catgttaagg gcggtttttt cctgtttggt cacttgatgc ctccgtgtaa gggggaattt 7020
 ctgttcatgg gggtaatgat accgatgaaa cgagagagga tgctcacgat acgggttact 7080
 gatgatgaac atgcccggtt actggaacgt tgtgagggta aacaactggc ggtatggatg 7140
 cggcgggacc agagaaaaat cactcagggt caatgccagc gtttcgttaa tacagatgta 7200
 ggtgttccac agggtagcca gcagcatcct gcgatgcaga tccggaacat aatgggtgcag 7260
 ggcgctgact tccgcgtttc cagactttac gaaacacgga aaccgaagac cattcatgtt 7320
 gttgctcagg tcgcagacgt tttgcagcag cagtcgcttc acgttcgctc gcgtatcggg 7380
 gattcattct gctaaccagt aaggcaacct cggcagccta gccgggtcct caacgacagg 7440
 agcagatca tgcgcacctg tggccaggac ccaacgctgc ccgaaatt 7488

<210> 2

<211> 671

<212> PRT

<213> 人工序列

<220>

<223>

<400> 2

Thr Leu Asn Ile Glu Asp Glu His Arg Leu His Glu Thr Ser Lys Glu
 1 5 10 15
 Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
 20 25 30
 Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
 35 40 45
 Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
 50 55 60

Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80
Leu Leu Asp Ala Gly Ile Leu Val Pro Cys Gln Ser Pro Trp Asn Thr
 85 90 95
Pro Leu Leu Pro Val Lys Lys Pro Gly Thr Asn Asp Tyr Arg Pro Val
 100 105 110
Gln Asp Leu Arg Glu Val Asn Lys Arg Val Glu Asp Ile His Pro Thr
 115 120 125
Val Pro Asn Pro Tyr Asn Leu Leu Ser Gly Leu Pro Pro Ser His Gln
 130 135 140
Trp Tyr Thr Val Leu Asp Leu Lys Asp Ala Val Phe Cys Leu Arg Leu
145 150 155 160
His Pro Thr Ser Gln Pro Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu
 165 170 175
Met Gly Ile Ser Gly Gln Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe
 180 185 190
Lys Asn Ser Pro Thr Leu Phe Asp Glu Ala Leu His Arg Asp Leu Ala
 195 200 205
Asp Phe Arg Ile Gln His Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp
210 215 220
Asp Leu Leu Leu Ala Ala Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr
225 230 235 240
Arg Ala Leu Leu Gln Thr Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala
 245 250 255
Lys Lys Ala Gln Ile Cys Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu
 260 265 270
Leu Lys Glu Gly Gln Arg Trp Leu Thr Glu Ala Arg Lys Glu Thr Val
 275 280 285
Met Gly Gln Pro Thr Pro Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu
290 295 300
Gly Thr Ala Gly Phe Cys Arg Leu Trp Ile Pro Gly Phe Ala Glu Met

305				310						315					320
Ala	Ala	Pro	Leu	Tyr	Pro	Leu	Thr	Lys	Thr	Gly	Thr	Leu	Phe	Asn	Trp
				325						330					335
Gly	Pro	Asp	Gln	Gln	Lys	Ala	Tyr	Gln	Glu	Ile	Lys	Gln	Ala	Leu	Leu
				340						345					350
Thr	Ala	Pro	Ala	Leu	Gly	Leu	Pro	Asp	Leu	Thr	Lys	Pro	Phe	Glu	Leu
				355						360					365
Phe	Val	Asp	Glu	Lys	Gln	Gly	Tyr	Ala	Lys	Gly	Val	Leu	Thr	Gln	Lys
				370						375					380
Leu	Gly	Pro	Trp	Arg	Arg	Pro	Val	Ala	Tyr	Leu	Ser	Lys	Lys	Leu	Asp
385															400
Pro	Val	Ala	Ala	Gly	Trp	Pro	Pro	Cys	Leu	Arg	Met	Val	Ala	Ala	Ile
				405						410					415
Ala	Val	Leu	Thr	Lys	Asp	Ala	Gly	Lys	Leu	Thr	Met	Gly	Gln	Pro	Leu
				420						425					430
Val	Ile	Leu	Ala	Pro	His	Ala	Val	Glu	Ala	Leu	Val	Lys	Gln	Pro	Pro
				435						440					445
Asp	Arg	Trp	Leu	Ser	Asn	Ala	Arg	Met	Thr	His	Tyr	Gln	Ala	Leu	Leu
				450											460
Leu	Asp	Thr	Asp	Arg	Val	Gln	Phe	Gly	Pro	Val	Val	Ala	Leu	Asn	Pro
465															480
Ala	Thr	Leu	Leu	Pro	Leu	Pro	Glu	Glu	Gly	Leu	Gln	His	Asn	Cys	Leu
				485						490					495
Asp	Ile	Leu	Ala	Glu	Ala	His	Gly	Thr	Arg	Pro	Asp	Leu	Thr	Asp	Gln
				500						505					510
Pro	Leu	Pro	Asp	Ala	Asp	His	Thr	Trp	Tyr	Thr	Asn	Gly	Ser	Ser	Leu
				515											520
Leu	Gln	Glu	Gly	Gln	Arg	Lys	Ala	Gly	Ala	Ala	Val	Thr	Thr	Glu	Thr
				530						535					540
Glu	Val	Ile	Trp	Ala	Lys	Ala	Leu	Pro	Ala	Gly	Thr	Ser	Ala	Gln	Arg
545															560

Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys
565 570 575
Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590
Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605
Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620
Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640
Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655
Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu
660 665 670

<210> 3

<211> 671

<212> PRT

<213> 小鼠白血病病毒 (Murine Leukemia Virus)

<400> 3

Thr Leu Asn Ile Glu Asp Glu His Arg Leu His Glu Thr Ser Lys Glu
1 5 10 15
Pro Asp Val Ser Leu Gly Ser Thr Trp Leu Ser Asp Phe Pro Gln Ala
20 25 30
Trp Ala Glu Thr Gly Gly Met Gly Leu Ala Val Arg Gln Ala Pro Leu
35 40 45
Ile Ile Pro Leu Lys Ala Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr
50 55 60
Pro Met Ser Gln Glu Ala Arg Leu Gly Ile Lys Pro His Ile Gln Arg
65 70 75 80

Leu	Leu	Asp	Gln	Gly	Ile	Leu	Val	Pro	Cys	Gln	Ser	Pro	Trp	Asn	Thr		
				85					90					95			
Pro	Leu	Leu	Pro	Val	Lys	Lys	Pro	Gly	Thr	Asn	Asp	Tyr	Arg	Pro	Val		
			100					105					110				
Gln	Asp	Leu	Arg	Glu	Val	Asn	Lys	Arg	Val	Glu	Asp	Ile	His	Pro	Thr		
			115				120					125					
Val	Pro	Asn	Pro	Tyr	Asn	Leu	Leu	Ser	Gly	Leu	Pro	Pro	Ser	His	Gln		
		130				135					140						
Trp	Tyr	Thr	Val	Leu	Asp	Leu	Lys	Asp	Ala	Phe	Phe	Cys	Leu	Arg	Leu		
145					150					155					160		
His	Pro	Thr	Ser	Gln	Pro	Leu	Phe	Ala	Phe	Glu	Trp	Arg	Asp	Pro	Glu		
				165				170						175			
Met	Gly	Ile	Ser	Gly	Gln	Leu	Thr	Trp	Thr	Arg	Leu	Pro	Gln	Gly	Phe		
			180					185					190				
Lys	Asn	Ser	Pro	Thr	Leu	Phe	Asp	Glu	Ala	Leu	His	Arg	Asp	Leu	Ala		
			195				200						205				
Asp	Phe	Arg	Ile	Gln	His	Pro	Asp	Leu	Ile	Leu	Leu	Gln	Tyr	Val	Asp		
			210				215					220					
Asp	Leu	Leu	Leu	Ala	Ala	Thr	Ser	Glu	Leu	Asp	Cys	Gln	Gln	Gly	Thr		
225					230					235					240		
Arg	Ala	Leu	Leu	Gln	Thr	Leu	Gly	Asn	Leu	Gly	Tyr	Arg	Ala	Ser	Ala		
				245						250					255		
Lys	Lys	Ala	Gln	Ile	Cys	Gln	Lys	Gln	Val	Lys	Tyr	Leu	Gly	Tyr	Leu		
			260					265						270			
Leu	Lys	Glu	Gly	Gln	Arg	Trp	Leu	Thr	Glu	Ala	Arg	Lys	Glu	Thr	Val		
			275					280						285			
Met	Gly	Gln	Pro	Thr	Pro	Lys	Thr	Pro	Arg	Gln	Leu	Arg	Glu	Phe	Leu		
			290				295						300				
Gly	Thr	Ala	Gly	Phe	Cys	Arg	Leu	Trp	Ile	Pro	Gly	Phe	Ala	Glu	Met		
305					310						315				320		
Ala	Ala	Pro	Leu	Tyr	Pro	Leu	Thr	Lys	Thr	Gly	Thr	Leu	Phe	Asn	Trp		

	325		330		335
Gly Pro Asp Gln Gln Lys Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu					
	340		345		350
Thr Ala Pro Ala Leu Gly Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu					
	355		360		365
Phe Val Asp Glu Lys Gln Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys					
	370		375		380
Leu Gly Pro Trp Arg Arg Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp					
385		390		395	400
Pro Val Ala Ala Gly Trp Pro Pro Cys Leu Arg Met Val Ala Ala Ile					
	405		410		415
Ala Val Leu Thr Lys Asp Ala Gly Lys Leu Thr Met Gly Gln Pro Leu					
	420		425		430
Val Ile Leu Ala Pro His Ala Val Glu Ala Leu Val Lys Gln Pro Pro					
	435		440		445
Asp Arg Trp Leu Ser Asn Ala Arg Met Thr His Tyr Gln Ala Leu Leu					
	450		455		460
Leu Asp Thr Asp Arg Val Gln Phe Gly Pro Val Val Ala Leu Asn Pro					
465		470		475	480
Ala Thr Leu Leu Pro Leu Pro Glu Glu Gly Leu Gln His Asn Cys Leu					
	485		490		495
Asp Ile Leu Ala Glu Ala His Gly Thr Arg Pro Asp Leu Thr Asp Gln					
	500		505		510
Pro Leu Pro Asp Ala Asp His Thr Trp Tyr Thr Asp Gly Ser Ser Leu					
	515		520		525
Leu Gln Glu Gly Gln Arg Lys Ala Gly Ala Ala Val Thr Thr Glu Thr					
	530		535		540
Glu Val Ile Trp Ala Lys Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg					
545		550		555	560
Ala Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys					
	565		570		575

Lys Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His
580 585 590

Ile His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly
595 600 605

Lys Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu
610 615 620

Phe Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys
625 630 635 640

Gly His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala
645 650 655

Arg Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu
660 665 670

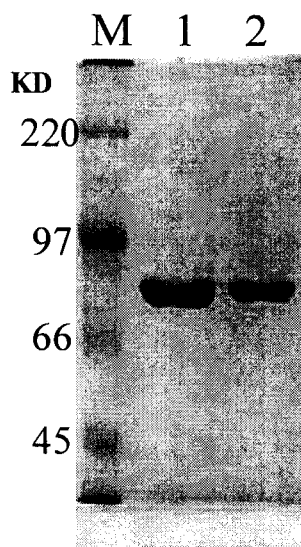


图 1

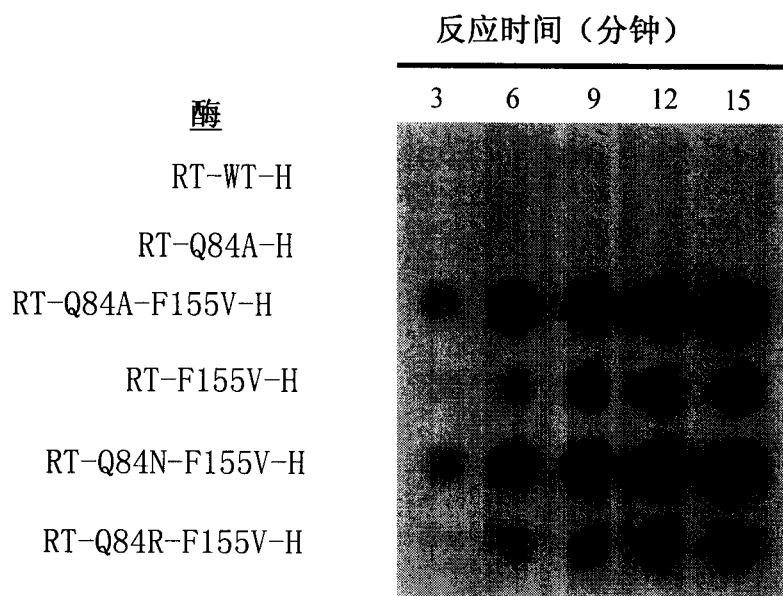


图 2

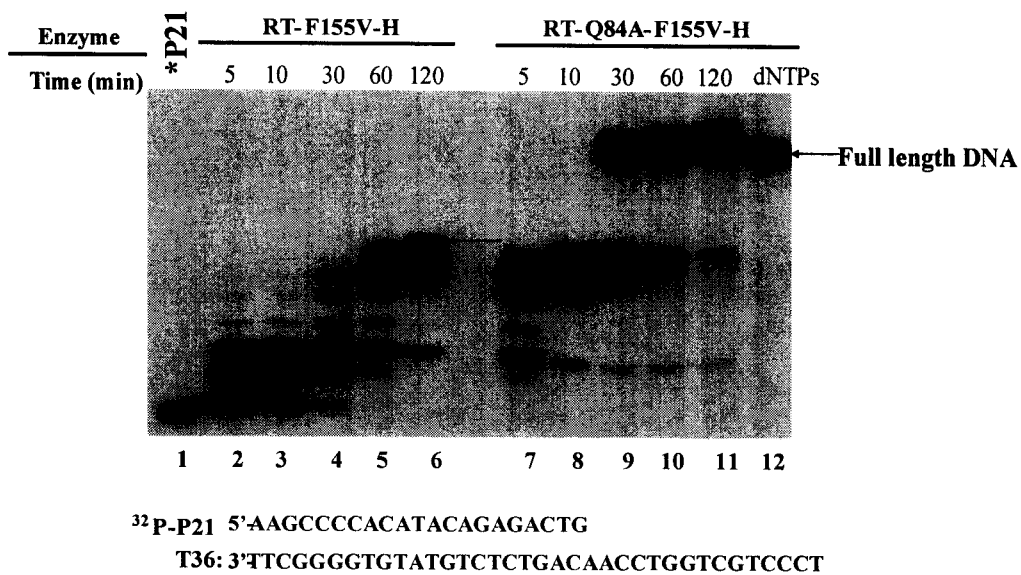


图 3