

[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 200510114234. X

[51] Int. Cl.

C12N 15/12 (2006.01)

C12N 15/63 (2006.01)

A61K 38/17 (2006.01)

A61K 48/00 (2006.01)

A61P 37/00 (2006.01)

[43] 公开日 2006 年 6 月 14 日

[11] 公开号 CN 1786167A

[22] 申请日 2005. 10. 21

[21] 申请号 200510114234. X

[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 刘利新 高光侠

[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 2 页 说明书 20 页 附图 3 页

[54] 发明名称

抗 HIV 病毒的多肽及其编码基因与应用

[57] 摘要

本发明公开了一种抗 HIV 病毒的多肽及其编码基因与应用。该抗 HIV 病毒多肽，选自下列多肽之一：所述多肽是将由序列表中序列 5 的氨基酸残基序列组成的多肽，自氨基端的第 453 位氨基酸残基开始向氨基端连续缺失 0 或自然数 1 至 100 中任意一个数值个氨基酸残基组成的 101 个多肽；或是将所述 101 个多肽的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的蛋白质。本发明的抗 HIV 病毒多肽可抑制 HIV 假病毒感染宿主细胞，可用于制备抗 HIV 病毒药物。

1、一种抗 HIV 病毒多肽，是将由序列表中序列 5 的氨基酸残基序列组成的多肽自氨基端的第 453 位氨基酸残基开始，向氨基端连续缺失 0 或自然数 1 至 100 中任意数值个氨基酸残基得到的 101 个多肽之一；或是将所述 101 个多肽的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽之一。

2、根据权利要求 1 所述的抗 HIV 病毒多肽，其特征在于：所述抗 HIV 病毒多肽为由序列表中序列 5 的氨基酸残基序列组成的多肽，或是将序列表中序列 5 的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽。

3、根据权利要求 2 所述的抗 HIV 病毒多肽，其特征在于：所述将序列表中序列 5 的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽是由自序列表中序列 3 的氨基端第 1 至 451 位氨基酸残基组成的多肽。

4、根据权利要求 1 所述的抗 HIV 病毒多肽，其特征在于：所述抗 HIV 病毒多肽为由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的多肽，或是将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽。

5、根据权利要求 4 所述的抗 HIV 病毒多肽，其特征在于：所述将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽，是由自序列表中序列 3 的氨基端第 1 至 351 位氨基酸残基组成的多肽。

6、权利要求 1-5 中任一所述的抗 HIV 病毒多肽的编码基因。

7、根据权利要求 6 所述的编码基因，其特征在于：所述由序列表中序列 5 的氨基酸残基序列组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 2 的自 5' 端第 1 位-第 1359 位脱氧核苷酸组成的核苷酸序列；所述由自序列表中序列 3 的氨基端第 1 至 451 位氨基酸残基组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 4 的自 5' 端第 1 位-第 1353 位脱氧核苷酸组成的核苷酸序列；所述由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 2 的自 5' 端第 1 位-第 1059 位脱氧核苷酸组成的核苷酸序列；由自序列表中序列 3 的氨基端第 1 至 351 位氨基酸残基组成的抗 HIV 病毒多肽，是由序列表中序列 4

的自 5' 端第 1 位-第 1053 位脱氧核苷酸组成的核苷酸序列。

8、含有权利要求 6 或 7 所述抗 HIV 病毒多肽的编码基因的工程菌、细胞系或载体。

9、权利要求 1-5 中任一所述的抗 HIV 病毒多肽在制备抗 HIV 病毒药物中的应用。

10、权利要求 1-5 中任一所述的抗 HIV 病毒多肽的编码基因在制备抗 HIV 病毒药物中的应用。

抗 HIV 病毒的多肽及其编码基因与应用

技术领域

本发明涉及抗 HIV 病毒的多肽及其编码基因与应用。

背景技术

HIV 病毒所引起的艾滋病是当今世界上对人类健康威胁最大的流行性传染病。在过去的十几年中，各国投入了巨大的努力和巨额资金致力于研制开发抗艾滋病的药物，目前世界上批准用于临床的抗艾滋病药物已有近二十种。但是，已有的任何一种药物都不能完全清除人体内的 HIV 病毒。几种药物组合使用的效果会优于单一药物治疗，但也不能完全清除 HIV 病毒。HIV 在复制过程中有很高的突变率，病人体内的病毒呈异源性，其中一些病毒具有抗药性，当治疗进行到一定阶段时，这些抗药病毒的数量会逐渐占据优势，使得现有的治疗无效。而且这些抗病毒药物的长期毒性已在临床上显示出来。不断地研制开发新的抗艾滋病药物，尤其是不同于现有药物作用机制的新一代抗艾滋病药物，是非常重要的而且十分必要的。

HIV 是一种逆转录病毒，在感染过程中首先通过病毒表面包膜蛋白 (envelope) 与其受体的相互作用吸附到细胞表面，然后病毒质膜与细胞质膜融合，裸露的病毒颗粒进入到细胞中。在细胞质中病毒的 RNA 被逆转录成双链 DNA，这段 DNA 存在于由病毒蛋白及宿主蛋白组成的整合前复合体 (Pre-Integration Complex, 简称 PIC) 中。PIC 由细胞质膜附近迁移到核膜附近并进入细胞核。在细胞核中病毒 DNA 由病毒的整合酶整合到宿主细胞的染色体中。上述过程被称为逆转录病毒生命周期的前期。整合的病毒 DNA 被转录成 RNA，其中大部分作为模板翻译病毒蛋白，小部分全长 RNA 作为病毒遗传物质被病毒蛋白包装成病毒颗粒出芽到细胞外成为未成熟病毒。在未成熟病毒中一些病毒蛋白以融合蛋白形式存在，经过病毒蛋白酶剪切后病毒颗粒结构重组成为有感染性的成熟病毒，此过程称为逆转录病毒生命周期的后期。

HIV 病毒基因只有 9kb，编码 15 种病毒蛋白。HIV 病毒以如此少的自身携带的功能完成一个非常复杂的复制过程必须利用宿主细胞的许多功能。另一方面，宿主在与病毒长期共存的过程中也进化产生了各种防御病毒的机制，其中一种机制是通过表达宿主细胞限制性因子抑制病毒的复制。近年来发现的 TRIM5、APOBEC3G 等都是能够有效抑制 HIV 复制的宿主细胞限制性因子。

CFIm (Cleavage Factor Im) 是真核生物 mRNA 前体 3' 加工复合体的一个组分。几乎所有的真核生物 mRNA 前体 3' 加工都包含两个密切相连的步骤，首先由加工复合

体识别初级转录物的多聚腺苷酸添加位点，并在此位点切割，然后上游切割片断进行聚腺苷酸化，下游片断则迅速降解。现在已知至少六种不同的蛋白参与真核生物 mRNA 前体 3' 加工复合体的组成，即 CPSF, CstF, PAP, PABPNI, CFIm 和 CFIIIm。鼠源 CFIm68 (rCFIm68)，由 552 个氨基酸残基组成，具有序列表中序列 1 的氨基酸序列，其编码基因具有序列表中序列 2 的核苷酸序列。rCFIm68 的序列已于 NCBI 的 GenBank 中发表，但未见相关研究论文报道。人源 CFIm68 (hCFIm68)，由 551 个氨基酸残基组成，具有序列表中序列 3 的氨基酸序列，其编码基因具有序列表中序列 4 的核苷酸序列。hCFIm68 与一个 25kd 小亚基形成异二聚体，该二聚体在 3' 加工复合体形成的早期就与 mRNA 前体结合，在 polyA 添加位点识别过程中起重要作用，能促进 3' 加工复合体的装配、PolyA 位点的切割和 PolyA 的添加。rCFIm68 和 hCFIm68 核苷酸序列同源性 93.2%，氨基酸序列同源性 99.5%，仅有两个氨基酸不同，分别是自氨基端的第 192 和 468 位氨基酸残基，另外 hCFIm (序列 3) 与 rCFIm (序列 1) 相比在 492 位处缺少一个精氨酸。序列 1 是序列 2 编码的蛋白质，序列 3 是序列 4 编码的蛋白质。

发明内容

本发明的目的是提供抗 HIV 病毒多肽及其编码基因。

本发明所提供的抗 HIV 病毒多肽，选自下列多肽之一：

所述多肽是将由序列表中序列 5 的氨基酸残基序列组成的多肽自氨基端的第 453 位氨基酸残基开始，向氨基端连续缺失 0 或自然数 1 至 100 中任意数值个氨基酸残基得到的 101 个多肽；或是将所述 101 个多肽的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽。

所述一个或几个氨基酸残基的取代和/或缺失和/或添加是指不多于十个氨基酸残基的取代和/或缺失和/或添加。

序列表中的序列 5 由 453 个氨基酸残基组成，是序列表中序列 1 的自氨基端第 1 至 453 位氨基酸残基。

所述抗 HIV 病毒多肽，优选为由序列表中序列 5 的氨基酸残基序列组成的多肽，或是将序列表中序列 5 的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽。

将由序列表中序列 5 的氨基酸残基序列组成的抗 HIV 病毒多肽命名为 rCFIm453。

所述将序列表中序列 5 的氨基酸残基序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽，是缺失自序列表中序列 5 的氨基端第 452 和第 453 位氨基酸残基且将自序列表中序列 5 的氨基端第 192 位的甘氨酸残基取代为丝氨酸残基得到的多肽，即由自序列表中序列 3 的氨基端第 1 至 451 位氨基酸

残基组成的多肽。

将由自序列表中序列 3 的氨基端第 1 至 451 位氨基酸残基组成的抗 HIV 病毒多肽命名为 hCFIm451。

所述抗 HIV 病毒多肽尤其优选为由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的多肽，或是将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽。

将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的抗 HIV 病毒多肽命名为 rCFIm353。

所述将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有抗 HIV 病毒活性的多肽，是 将由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的多肽中的自序列 5 的氨基端第 353 和 352 位氨基酸残基缺失，且将自序列表中序列 5 的氨基端第 192 位的甘氨酸残基取代为丝氨酸残基得到的多肽，即由自序列表中序列 3 的氨基端第 1 至 351 位氨基酸残基组成的多肽。

将由自序列表中序列 3 的氨基端第 1 至 351 位氨基酸残基组成的抗 HIV 病毒多肽命名为 hCFIm351。

上述抗 HIV 病毒多肽的编码基因也属于本发明的保护范围。

所述由序列表中序列 5 的氨基酸残基序列组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 2 的自 5' 端第 1 位-第 1359 位脱氧核苷酸组成的核苷酸序列；所述由自序列表中序列 3 的氨基端第 1 至 451 位氨基酸残基组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 4 的自 5' 端第 1 位-第 1353 位脱氧核苷酸组成的核苷酸序列；所述由自序列表中序列 5 的氨基端第 1 至第 353 位氨基酸残基组成的抗 HIV 病毒多肽的编码基因，是由序列表中序列 2 的自 5' 端第 1 位-第 1059 位脱氧核苷酸组成的核苷酸序列；由自序列表中序列 3 的氨基端第 1 至 351 位氨基酸残基组成的抗 HIV 病毒多肽，是由序列表中序列 4 的自 5' 端第 1 位-第 1053 位脱氧核苷酸组成的核苷酸序列。

上述抗 HIV 病毒多肽的编码基因的工程菌、细胞系或载体均属于本发明的保护范围。

细胞实验结果表明，本发明的抗 HIV 病毒多肽 rCFIm353，hCFIm351，rCFIm453 hCFIm451 均可抑制假病毒 HIV-luc 感染宿主细胞，在 293A 细胞中，rCFIm353 的抑制倍数达 38 倍，该 HIV 病毒多肽及其编码基因可用于制备抗 HIV 病毒药物，为 HIV 的

治疗开拓新的途径。

附图说明

图 1 为构建在 pcDNA4/to/myc-HisB 上的 CFIm68 截短体在 293A 细胞中的表达。

图 2 为构建在 pBabe-puro-myc 上的 CFIm68 截短体在 293A 细胞中的表达。

图 3 为在 Rat2 细胞中 CFIm68 及截短体对病毒感染的抑制。

图 4 为在 293A 细胞中 rCFIm353 和 hCFIm351 对病毒感染的抑制。

图 5 为表达 rCFIm353 的 Rat2 和 293A 细胞对 HIV-luc 和 MLV-luc 转染的抑制作用。

图 6 为病毒逆转录中间产物线性 DNA 和环状 DNA 的形成。

图 7 为 rCFIm353 和全长 rCFIm68 的细胞定位。

具体实施方式

下述实施例中的实验方法，如无特别说明，均为常规方法。

实施例 1、具有抗 HIV 功能的宿主细胞基因发现

按照 Guangxia Gao 等人于 2002 年 9 月发表在 *Science* 杂志中第 1703-1706 页的体细胞遗传学筛选方法获得宿主细胞基因。主要步骤如下：建立一个大鼠野生型 Rat2(购自 ATCC)细胞 cDNA 文库(该文库为 Guangxia Gao 等人于 2002 年 9 月发表在 *Science* 杂志中第 1703-1706 页的文章中所用文库)，将这个 cDNA 库引入宿主细胞 Rat2 中。设计引物(上游：aggatccatggc ttcgtacccc tgcca 下游：taggtacc tcagttagcctccccatct)从质粒 pHSV106(购自 Gibco/BRL 公司)上扩增得到编码胸苷激酶(thymidine kinase, 简称 TK)基因的 DNA 片段，扩增的体系为：dNTP 混合物(各 2.5mM)4 微升，10×PCR 缓冲液 5 微升，上游引物 (20 μM) 1 微升，下游引物 (20 μM) 1 微升，模板 10ng，Pyrobest DNA 聚合酶(购自 TaKaRa 公司) 1.25U，补加灭菌去离子水至总体积 50 微升。扩增的程序为：94℃ 预变性 4 分钟，然后扩增 25 个循环(每次 94℃ 30 秒，55℃ 30 秒，72℃ 2.0 分钟)，最后 72℃ 延伸 10 分钟。扩增产物经 *Bam*HI 和 *Kpn*I 酶切后，连入基于 HIV 的逆转录载体 pTrip-GFP (pTrip-GFP 的构建方法见 Zennou V 等于 2000 年 4 月发表在 *cell* 上第 173-185 页的文章)的 *Bam*HI, *Kpn*I 位点，用 TK 基因取代编码绿色荧光蛋白的 GFP 基因，得到逆转录载体 HIV-TK。将 1 微克表达 VSVG 的质粒 pMD.G (构建方法参见 Ori, DS. 等发表在 *Proc. Natl. Acad. Sci. USA*, 1996,93(21)第 11400-11406 页的文章)，1 微克表达 HIV-gag-pol 等蛋白的质粒 pCMV△R8.2(构建方法参见 Naldini, L 等发表在 *Proc. Natl. Acad. Sci. USA*, 1996,93(21), 第 11382-11388 页的文章)，1.5 微克逆转录载体 HIV-TK 共转染到 293T 细胞(购自 ATCC)，采用 FuGENE6 转染试剂(Roche 公司)在 60mm 培养皿中进行。48 小时后，收获含病毒

上清液，经 0.45 μ m 孔径的膜过滤后，得到 HIV-TK 病毒液。然后用 HIV-TK 假病毒侵染这些表达大鼠基因片段的宿主细胞，并利用三氟胸苷（TFT）筛选出对病毒侵染有抗性的宿主细胞。将这些宿主细胞中的 cDNA 用 PCR 方法分离出来并转入野生型的宿主细胞中，确证其中一个 cDNA 的表达能使宿主细胞对病毒侵染产生抗性。将所克隆的基因进行序列测定，经 GeneBank 序列比对，结果表明该基因编码一个 mRNA 前体剪切因子 CFIm68 的 N 端第 1 位-353 位氨基酸片段（自序列 1 的氨基端第 1 至 353 位氨基酸残基）。

实施例 2、CFIm68 基因不同截短体对 HIV 的抗性

一、CFIm68 全长基因的克隆

根据序列表中序列 2 的大鼠 CFIm68 序列，设计引物（上游：cgg ggt acc atg gcg gac ggt gtg gac，下游：cgc gga tcc acg atg acg gta ttc tcg ctc）。采用 RNeasy® Mini kit RNA 提取试剂盒（购自 QIAGEN，货号 74104），并按照其说明书标准方法提取大鼠 Rat2 细胞总 RNA。采用 SuperScript™II 逆转录酶（购自 invitogen，货号 18064-022）合成第一链 cDNA，RT 反应体系：Oligo(dT)₁₈ (500 微克/微升) 1 微升，total RNA 1 微克，dNTP (各 10mM) 1 微升，补加灭菌去离子水至总体积 12 微升。65℃温浴 5 分钟，立即冰浴。在体系中加入 5×First-Strand buffer 4 微升，0.1M DTT 2 微升。42℃温浴 2 分钟。添加 1 微升逆转录酶，42℃反应 50 分钟，70℃热失活 15 分钟。以上述所得 cDNA 为模板，进行 PCR 扩增，PCR 扩增体系为：dNTP 混合物 (各 2.5mM) 4 微升，10×PCR 缓冲液 5 微升，上游引物 (20 μ M) 1 微升，下游引物 (20 μ M) 1 微升，第一链 cDNA 2 微升，Pyrobest DNA 聚合酶（购自 TaKaRa 公司）1.25U，补加灭菌去离子水至总体积 50 微升；扩增程序：94℃ 预变性 4 分钟，然后扩增 25 个循环（每次 94℃ 30 秒，55℃ 30 秒，72℃ 2.0 分钟），最后 72℃延伸 10 分钟；将得到的扩增产物进行测序，测序结果表明扩增得到具有序列表中序列 2 的核苷酸序列的鼠 CFIm68 基因（rCFIm68 基因）。根据序列表中序列 4 的人源 CFIm68 序列，设计引物（上游：cgg ggt acc atg gcg gac ggc gtg gac，下游：Cgc gga tcc acg atg acg ata ttc gcg ctc），以人胎肝 cDNA 库（购自 Invitrogen 货号 D8880-01）为模板，进行 PCR 扩增。其中，PCR 扩增体系为：dNTP 混合物 (各 2.5mM) 4 微升，10×PCR 缓冲液 5 微升，上游引物 (20 μ M) 1 微升，下游引物 (20 μ M) 1 微升，cDNA 库 100ng，Pyrobest DNA 聚合酶（购自 TaKaRa 公司）1.25U，补加灭菌去离子水至总体积 50 微升；扩增程序：94℃ 预变性 4 分钟，然后扩增 25 个循环（每次 94℃ 30 秒，55℃ 30 秒，72℃ 2.0 分钟），最后 72℃延伸 10 分钟。将得到的扩增产物进行测序，测序结果表明扩增得到具有序列表中序列 4 的核苷酸序列的人 CFIm68 基因（hCFIm68 基因）。hCFIm68 基

因编码具有 551 个氨基酸残基的序列表中序列 3 的蛋白质。rCFIm 编码具有 552 个氨基酸残基的序列表中序列 1 的蛋白质。两者核苷酸序列同源性 93.2%，氨基酸序列同源性 99.5%，仅有两个氨基酸不同，分别是自氨基端的第 192 和 468 位氨基酸残基，另外 hCFIm（序列 3）与 rCFIm（序列 1）相比在 492 位处缺少一个精氨酸。

二、CFIm68 基因不同截短体对 HIV 的抗性

1、以 pcDNA4/to/myc-HisB(购自 Invitrogen 公司)为载体的截短体构建

对应 rCFIm68 基因内部即序列表中序列 1 的氨基端第 46, 146, 309, 409 位氨基酸处分别设计引物作为上游引物（引物中带有下划线的序列表示 *KpnI* 酶切识别序列，方框中的序列为 *BamHI* 酶切识别序列）（46aa: cgg ggt acc atg gcc cca gaa gac cgg gac; 或 146aa: cgg ggt acc atg aaa aga gaa ctt cac ggt cag; 或 309aa: cgg ggt acc atg cca ccg cct cca cag cag; 或 409aa: cgg ggt acc atg ctg agt gaa gct gag ttt gaa;），根据序列表中序列 1 的 C 端设计引物 Pc: cgc gga tcc acg atg acg gta ttc tcg ctc 作为下游引物，分别以步骤一得到的 rCFIm68 基因做为模板，进行 PCR 扩增，分别获得自序列表中序列 1 的氨基端第 46—552 位、第 146—552 位、第 309—552 位、第 409—552 位氨基酸残基组成的 N 端缺失的氨基酸片段的编码基因，分别命名为 rCFIm Δ 45 基因、rCFIm Δ 145 基因、rCFIm Δ 308 基因、rCFIm Δ 408 基因。其中，扩增的体系分别均为：dNTP 混合物(各 2.5mM)4 微升，10×PCR 缓冲液 5 微升，上游引物（20 μM）1 微升，下游引物（20 μM）1 微升，模板 10ng，Pyrobest DNA 聚合酶（购自 TaKaRa 公司）1.25U，补加灭菌去离子水至总体积 50 微升。扩增的程序均为：94℃ 预变性 4 分钟，然后扩增 25 个循环（每次 94℃ 30 秒，55℃ 30 秒，72℃ 2.0 分钟），最后 72℃ 延伸 10 分钟。

同样对应自序列表中序列 1 的氨基端第 90, 190, 253, 353, 453 位氨基酸处设计引物作为下游引物（90aa: cgc gga tcc cca tgt tag att tcc aat ata; 190aa: cgc gga tcc tcc tgg agg acc agc ttt acc; 253aa: cgc gga tcc tgg agg acc tgg ggg gcc agg; 353aa: cgc gga tcc tgg tgg tgg ggc acc tgg agg; 453aa: cgc gga tcc aga aat tgc tgt tac caa tgt）；对应序列表中序列 1 氨基端第 1 位氨基酸处设计引物 Pn: cgg ggt acc atg gcg gac ggt gtg gac 作为上游引物，分别以步骤一得到的 rCFIm68 基因做为模板，进行 PCR 扩增，分别获得自序列表中序列 1 的氨基端第 1—90 位、第 1—190 位、第 1—253 位、第 1—353 位、第 1—453 位氨基酸残基组成的 C 端缺失的氨基酸片段的编码基因，分别命名为 rCFIm90 基因、rCFIm190 基因、rCFIm253 基因、rCFIm353 基因、rCFIm453 基因。其中，扩增的体系均为：dNTP 混合物(各 2.5mM)4 微升，10×PCR 缓冲液 5 微升，上游引物（20 μM）1 微升，下游引物

(20 μ M) 1 微升, 模板 10ng, Pyrobest DNA 聚合酶 (购自 TaKaRa 公司) 1.25U, 补加灭菌去离子水至总体积 50 微升。扩增的程序均为: 94 $^{\circ}$ C 预变性 4 分钟, 然后扩增 25 个循环 (每次 94 $^{\circ}$ C 30 秒, 55 $^{\circ}$ C 30 秒, 72 $^{\circ}$ C 2.0 分钟), 最后 72 $^{\circ}$ C 延伸 10 分钟。用上述 Pc 和 Pn 为引物, 分别以步骤一得到的 rCFIm68 基因做为模板, 进行 PCR 扩增, 得到序列表中序列 1 的 rCFIm68 编码基因 rCFIm552。

对应自序列表中序列 3 的氨基端第 351 位、451 位和 551 位氨基酸处设计了三条下游引物 h351: cgc gga tcc tgg ggc acc tgg agg tgg tcc agg; h451: cgc gga tcc tgc agt tac cag tgt ctc aat agc; h551: Cgc gga tcc acg atg acg ata ttc ggc ctc (带有下划线的序列为 *Bam*HI 的酶切识别序列), 对应自序列表中序列 3 的氨基端第 1 位氨基酸处设计了一条上游引物: cgg ggt acc atg gcg gac ggc gtg gac (方框中的序列为 *Kpn*I 的酶切识别序列), 分别以步骤一得到的 hCFIm68 基因做为模板, 进行 PCR 扩增反应, 得到自序列表中序列 3 的第 1—第 351 位和第 1—第 451 位氨基酸残基组成的 hCFIm351 基因、hCFIm451 基因和序列表中序列 3 的 hCFIm68 编码基因 hCFIm551。反应体系为: dNTP 混合物 (各 2.5mM) 4 微升, 10 \times PCR 缓冲液 5 微升, 上游引物 (20 μ M) 1 微升, 下游引物 (20 μ M) 1 微升, 模板 10ng, Pyrobest DNA 聚合酶 (购自 TaKaRa 公司) 1.25U, 补加灭菌去离子水至总体积 50 微升; 反应程序为: 94 $^{\circ}$ C 预变性 4 分钟, 然后扩增 25 个循环 (每次 94 $^{\circ}$ C 30 秒, 55 $^{\circ}$ C 30 秒, 72 $^{\circ}$ C 2.0 分钟), 最后 72 $^{\circ}$ C 延伸 10 分钟。其中, 上述引物中 5' 端引物引入 *Kpn*I 位点, 3' 端引物引入 *Bam*HI 位点, 上述 PCR 产物经 *Kpn*I 和 *Bam*HI 双酶切后分别连入载体 ncDNA4/to/myc-HisB 的 *Kpn*I-*Bam*HI 位点。将含有 rCFIm Δ 45 基因的载体命名为 prCFIm Δ 45, 将含有 rCFIm Δ 145 基因的载体命名为 prCFIm Δ 145、将含有 rCFIm Δ 308 基因的载体命名为 prCFIm Δ 308、将含有 rCFIm Δ 408 基因的载体命名为 prCFIm Δ 408、将含有 rCFIm90 基因的载体命名为 prCFIm90、将含有 rCFIm190 基因的载体命名为 prCFIm190、将含有 rCFIm253 基因的载体命名为 prCFIm253、将含有 rCFIm353 基因的载体命名为 prCFIm353、将含有 rCFIm453 基因的载体命名为 prCFIm453、将含有 hCFIm351 基因的载体命名为 phCFIm351、将含有 hCFIm451 基因的载体命名为 phCFIm451、将含有 rCFIm552 的载体命名为 prCFIm552、将含有 hCFIm551 的载体命名为 phCFIm551。在每个截短体、全长蛋白末端融合有 myc 标签, 将这些质粒转染入 293A 细胞 (购自 Invitrogen 公司), 48 小时后收获细胞, 用 anti-myc 抗体 (购自 Santa Cruz Biotechnology 公司) 进行免疫杂交, 表明分别得到大小为 60kDa 的 rCFIm Δ 45、大小为 49kDa 的 rCFIm Δ 145、大小为 33kDa 的 rCFIm Δ 308、大小为 23kDa 的 rCFIm Δ 408、大小为 15kDa 的 rCFIm90、大小为 26kDa 的 rCFIm190、大小为 32kDa 的

rCFIm253、大小为 42kDa 的 rCFIm353、大小为 52kDa 的 rCFIm453、大小为 65kDa 的 rCFIm552、大小为 42kDa 的 hCFIm351、大小为 52kDa 的 hCFIm451、大小为 65kDa 的 hCFIm551 的杂交条带，说明各片段正常表达。结果如图 1 所示。

2、转录载体 pBabe-puro 的改造

设计引物，上游引物：gaagatctgttttaaacttaagcttggtac（带有下划线的序列为 *Bg*III 酶切识别序列），下游引物：ggaattctca cagatcctcttctgagatgag（方框中的序列为 *Eco*RI 酶切识别序列），以 pcDNA4/to/myc-HisB 质粒为模板，获得 147bp 包含 myc 标签和多克隆位点的序列，经 *Bg*III/*Eco*RI 部分酶切后，连入 pBabe-puro（购自 clontech）的 *Ban*HI/*Eco*RI 位点，获得 pBabe-puro-myc。

3、pBabe-puro-myc 为载体的截短体的构建

将 prCFIm Δ 45、prCFIm Δ 145、prCFIm Δ 308、prCFIm Δ 408、prCFIm90、prCFIm190、prCFIm253、prCFIm353、prCFIm453、prCFIm552、phCFIm351、phCFIm451、phCFIm551 分别用 *Pme*I/*Ban*HI 双酶切，切下 rCFIm68 和 hCFIm68 基因截短体，将 pBabe-puro-myc 用 *Pme*I/*Ban*HI 双酶切，回收大片段作为载体，然后把 rCFIm68 和 hCFIm68 基因截短体分别与此载体片段连接，将连接产物分别转化大肠杆菌 TOP10，提取质粒并测序验证，将含有步骤 1 中 rCFIm Δ 45、rCFIm Δ 145、rCFIm Δ 308、rCFIm Δ 408、rCFIm90、rCFIm190、rCFIm253、rCFIm353、rCFIm453、hCFIm351、hCFIm451 基因、hCFIm551、rCFIm552 的质粒分别称为 pBrCFIm Δ 45、pBrCFIm Δ 145、pBrCFIm Δ 308、pBrCFIm Δ 408、pBrCFIm90、pBrCFIm190、pBrCFIm253、pBrCFIm353、pBrCFIm453、pBhCFIm351、pBhCFIm451、pBhCFIm551、pBrCFIm552。将这些质粒分别转染入 293A 细胞，用 anti-myc 抗体进行 Western 免疫杂交，结果表明分别得到大小为 59kDa 的 rCFIm Δ 45、大小为 48kDa 的 rCFIm Δ 145、大小为 32kDa 的 rCFIm Δ 308、大小为 22kDa 的 rCFIm Δ 408、大小为 14kDa 的 rCFIm90、大小为 25kDa 的 rCFIm190、大小为 31kDa 的 rCFIm253、大小为 41kDa 的 rCFIm353、大小为 51kDa 的 rCFIm453、大小为 64kDa 的 rCFIm552、大小为 41kDa 的 hCFIm351、大小为 51kDa 的 hCFIm451、大小为 64kDa 的 hCFIm551 的杂交条带，说明各片段正常表达。结果如图 2 所示。

4、抗性片段的确定

将 1 微克 表达 VSVG 的质粒 pMD.G，1 微克表达 MLV-gag-pol 的质粒 pHIT60（构建方法参见 Soneoka Y. 等发表在 *Nucleic Acids Res.* 1995 Feb; 23(4):第 628-633 页的文章），1.5 微克 MLV 假病毒载体 pFB-luc（购自 Stratagene）共转染到 293T 细胞，采用 FUGENE6 转染试剂（Roche 公司）在 60mm 培养皿中进行。48 小时后，收获含病毒上清液，经 0.45 μ m 孔径的膜过滤后，得到 MLV-luc 病毒液。

将 1 微克表达 VSVG 的质粒 pMD. G, 1 微克表达 HIV-gag-pol 等蛋白的质粒 pCMV Δ R8.2, 1.5 微克 HIV 假病毒载体 pHR' -CMV-luc (pHR' -CMV-luc 的构建方法见 Li SL 等, *Microbiol Immunol.* 2000 年 44(12), 第 1019-1025 页) 共转染到 293T 细胞, 采用 FuGENE6 转染试剂 (Roche 公司) 在 60mm 培养皿中进行。48 小时后, 收获含病毒上清液, 经 0.45 μ m 孔径的膜过滤后, 得到 HIV-luc 病毒液。

按照包装 MLV-luc 的方法, 将质粒 pMD. G 和 pHIT60 与下述 14 种质粒中的一种: pBabe-puro-myc、pBrCFIm Δ 45、pBrCFIm Δ 145、pBrCFIm Δ 308、pBrCFIm Δ 408、pBrCFIm90、pBrCFIm190、pBrCFIm253、pBrCFIm353、pBrCFIm453、pBrCFIm552、pBhCFIm351、pBhCFIm451、pBhCFIm551 共转染 293T 细胞, 获得相应的的假膜病毒。

下述 14 种假膜病毒: 假病毒 pBabe-puro-myc、假病毒 pBrCFIm Δ 45、假病毒 pBrCFIm Δ 145、假病毒 pBrCFIm Δ 308、假病毒 pBrCFIm Δ 408、假病毒 pBrCFIm90、假病毒 pBrCFIm190、假病毒 pBrCFIm253、假病毒 pBrCFIm353、假病毒 pBrCFIm453、假病毒 pBrCFIm552、假病毒 pBhCFIm351、假病毒 pBhCFIm451、假病毒 pBhCFIm551 分别感染 Rat2 细胞, 筛选得到嘌呤霉素 (4 μ g/ml) 的抗性细胞群。

将假病毒 MLV-luc 或 HIV-luc 分别感染上述得到的 14 种嘌呤霉素抗性的 Rat2 细胞群, 两天后用 Promega 公司的 luciferase 试剂盒测定嘌呤霉素抗性细胞群荧光素酶活性, 获得 14 种嘌呤霉素抗性细胞群荧光素酶的荧光值。以整合有空质粒 pBabe-puro-myc 的嘌呤霉素抗性细胞群的荧光值作为对照, 用它去除其他嘌呤酶素抗性细胞群荧光素酶的荧光值, 所得比值即为抑制倍数, 即: 抑制倍数=含空质粒嘌呤霉素抗性细胞群组值/其它嘌呤霉素抗性细胞群组值。

结果表明, 在 Rat2 细胞中, pBrCFIm353 组, pBrCFIm453 组, pBhCFIm351 组、pBhCFIm451 组均对 HIV-luc 具有抑制作用, 抑制倍数分别为 19 倍、6 倍、15 倍、8 倍, 而对 MLV-luc 则没有明显的抑制作用, 其他片段及全长基因对 HIV-luc 和 MLV-luc 感染无抗性。部分片段的抑制结果如图 3 所示, 其中 R90、R253、R353、R453、R552、R Δ 145、R Δ 308、R Δ 408、h351、h451、h551 分别表示 pBrCFIm90、pBrCFIm253、pBrCFIm353、pBrCFIm453、pBrCFIm552、pBrCFIm Δ 145、pBrCFIm Δ 308、pBrCFIm Δ 408、pBhCFIm351、pBhCFIm451、pBhCFIm551 组。

5、抗性片段在 293A 细胞中对 HIV-luc 和 MLV-luc 假病毒的抗性

按照步骤 4 的方法, 分别取假病毒 pBrCFIm353、假病毒 pBhCFIm351、假病毒 pBabe-puro-myc 三种假病毒中的一种, 感染 293A 细胞, 筛选得到嘌呤霉素 (4 μ g/ml) 抗性细胞群。分别用假病毒 MLV-luc 或 HIV-luc 感染这三个嘌呤霉素 (4 μ g/ml) 抗性细胞群, 两天后测定嘌呤霉素抗性细胞群荧光素酶活性。抑制倍数的确定方法同步

骤 4。结果如图 4 所示，表明在 293A 细胞中，rCFIm353 和 hCFIm351 均对 HIV-luc 具有抑制作用，抑制倍数分别为在 38 倍和 9 倍，而对 MLV-luc 则没有明显的抑制作用。

实施例 3、rCFIm353 对 HIV-luc 抑制位点的确定

rCFIm353 对 HIV-luc 的抑制作用可能发生在感染的早期阶段，即整合入染色体形成原病毒之前，或发生晚期阶段，即原病毒的表达。为区分这两种可能性，用脂质体转染的方法分别将载体 HIV-luc 和载体 MLV-luc 转入表达 rCFIm353 的 Rat2 和 293A 细胞，同时将载体 HIV-luc 和 MLV-luc 转染入整合有空载体 pBabe-puro-myc 的 Rat2 和 293A 细胞。48 小时后荧光素酶活性，确定抑制倍数，结果如图 5 所示，含 rCFIm353 的细胞对 HIV-luc 和 MLV-luc 都没有抑制作用。表明 rCFIm353 对 HIV-luc 的抑制应该发生在感染的早期阶段。

为确定阻断 HIV 感染的确切位置，检测了 HIV-luc 感染后的病毒 DNA 合成情况。将转入 pBrCFIm353 得到的表达 rCFIm353 的 Rat2 细胞（实验组）与整合有空载体 pBabe-puro-myc 的 Rat2（对照组细胞）在用不同浓度 HIV-luc 假病毒感染 24 小时后，提取 Hirt DNA，该方法可提取细胞中的低分子量 DNA，其中包括逆转录病毒逆转录后生成的 DNA（具体方法见 Hirt B, *J. Mol. Boil.* 1967, 26:365-369）。以 Hirt DNA 为模板，PCR 检测 Luc DNA 的合成，（上游引物：gag gat gga acc gct gga gag caa c；下游引物：atc cag agg aat tca tta tca gtg c）PCR 的反应体系为：dNTP 混合物（各 2.5mM）4 微升，10×PCR 缓冲液 5 微升，上游引物（20 μM）1 微升，下游引物（20 μM）1 微升，Hirt DNA 10 微升，Pyrobest DNA 聚合酶（购自 TaKaRa 公司）1.25U，补加灭菌去离子水至总体积 50 微升；反应程序为：94℃ 预变性 4 分钟，然后扩增 30 个循环（每次 94℃ 30 秒，55℃ 30 秒，72℃ 1.0 分钟），最后 72℃ 延伸 10 分钟。凝胶电泳表明实验组与对照组中 HIV-Luc 病毒逆转录产生 Luc DNA 的 PCR 产物的量无明显差异。结果如图 6 所示，表明阻断位点在逆转录完成以后的阶段。

病毒完成逆转录后，线性的病毒 DNA 进入核中，部分形成有两个长末端重复序列（2-LTR）连接的环状 DNA 形式。这些环状 DNA 不是形成正常整合原病毒过程的中间体，但可作为 DNA 进入核中的标志。同样以上述试验中的 Hirt DNA 为模板，PCR 检测环状 DNA 量，（上游引物为：gta act aga gat acc ctc aac 下游引物为：cag atc tgg tct aac cag aga）。PCR 的反应体系为 dNTP 混合物（各 2.5mM）4 微升，10×PCR 缓冲液 5 微升，上游引物（20 μM）1 微升，下游引物（20 μM）1 微升，Hirt DNA 10 微升，Pyrobest DNA 聚合酶（购自 TaKaRa 公司）1.25U，补加灭菌去离子水至总体积 50 微升；反应程序为：94℃ 预变性 4 分钟，然后扩增 35 个循环（每次 94℃ 30 秒，55℃ 30 秒，72℃ 1.0 分钟），最后 72℃ 延伸 10 分钟。凝胶电泳结果如图 6 所示，表

明与对照相比, 含有 rCFIm353 的细胞的病毒环状 DNA 的 PCR 产物显著减少。说明 rCFIm353 对 HIV-1uc 的抑制位点在 DNA 入核之前。图 6 中, 2-LTR 指细胞内病毒环状 DNA 的扩增产物。

由上可确定, 阻断 HIV-1uc 的位点应在逆转录完成后, 入核之前。即整合前复合体 (PIC) 的形成, 及 PIC 向核膜的迁移和入核过程。

实施例 4、rCFIm353 和全长 rCFIm68 的细胞定位

将 rCFIm353 从实施例 2 中构建的 pBrCFIm353 用 *KpnI* 和 *BamHI* 双酶切下, 连入载体 pEGFP-N1 (购自 clontech 公司) 的 *KpnI*, *BamHI* 位点。将含 GFP 标签的 rCFIm353 和含 myc 标签的全长 rCFIm68 共转染入 Hala 细胞 (购自 ATCC) 中, 荧光显微镜下观测, 发现 rCFIm353 定位于细胞质和细胞核中, 而全长 rCFIm68 仅定位于细胞核中, 结果如图 7 所示。图 7 中, rCFIm-FL 表示全长 rCFIm68, 右侧图片为左侧图片和中间图片的叠加结果。

序列表

<160> 5

<210> 1

<211> 552

<212> PRT

<213>大鼠 (*Rattus norvegicus*)

<400> 1

Met	Ala	Asp	Gly	Val	Asp	His	Ile	Asp	Ile	Tyr	Ala	Asp	Val	Gly	Glu	1	5	10	15
Glu	Phe	Asn	Gln	Glu	Ala	Glu	Tyr	Gly	Gly	His	Asp	Gln	Ile	Asp	Leu	20	25	30	
Tyr	Asp	Asp	Val	Ile	Ser	Pro	Ser	Ala	Asn	Asn	Gly	Asp	Ala	Pro	Glu	35	40	45	
Asp	Arg	Asp	Tyr	Met	Asp	Thr	Leu	Pro	Pro	Thr	Val	Gly	Asp	Asp	Val	50	55	60	
Gly	Lys	Gly	Ala	Ala	Pro	Asn	Val	Val	Tyr	Thr	Tyr	Thr	Gly	Lys	Arg	65	70	75	80
Ile	Ala	Leu	Tyr	Ile	Gly	Asn	Leu	Thr	Trp	Trp	Thr	Thr	Asp	Glu	Asp	85	90	95	
Leu	Thr	Glu	Ala	Val	His	Ser	Leu	Gly	Val	Asn	Asp	Ile	Leu	Glu	Ile	100	105	110	
Lys	Phe	Phe	Glu	Asn	Arg	Ala	Asn	Gly	Gln	Ser	Lys	Gly	Phe	Ala	Leu	115	120	125	
Val	Gly	Val	Gly	Ser	Glu	Ala	Ser	Ser	Lys	Lys	Leu	Met	Asp	Leu	Leu	130	135	140	
Pro	Lys	Arg	Glu	Leu	His	Gly	Gln	Asn	Pro	Val	Val	Thr	Pro	Cys	Asn	145	150	155	160
Lys	Gln	Phe	Leu	Ser	Gln	Phe	Glu	Met	Gln	Ser	Arg	Lys	Thr	Thr	Gln	165	170	175	
Ser	Gly	Gln	Met	Ser	Gly	Glu	Gly	Lys	Ala	Gly	Pro	Pro	Gly	Gly	Gly	180	185	190	
Ser	Arg	Ala	Ala	Phe	Pro	Gln	Gly	Gly	Arg	Gly	Arg	Gly	Arg	Phe	Pro	195	200	205	
Gly	Ala	Val	Pro	Gly	Gly	Asp	Arg	Phe	Pro	Gly	Pro	Ala	Gly	Pro	Gly	210	215	220	

Gly Pro Pro Pro Pro Phe Pro Ala Gly Gln Thr Pro Pro Arg Pro Pro
 225 230 235 240
 Leu Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Pro Pro Gly
 245 250 255
 Gln Val Leu Pro Pro Pro Leu Ala Gly Pro Pro Asn Arg Gly Asp Arg
 260 265 270
 Pro Pro Pro Pro Val Leu Phe Pro Gly Gln Pro Phe Gly Gln Pro Pro
 275 280 285
 Leu Gly Pro Leu Pro Pro Gly Pro Pro Pro Pro Val Pro Gly Tyr Gly
 290 295 300
 Pro Pro Pro Gly Pro Pro Pro Pro Gln Gln Gly Pro Pro Pro Pro Pro
 305 310 315 320
 Gly Pro Phe Pro Pro Arg Pro Pro Gly Pro Leu Gly Pro Pro Leu Thr
 325 330 335
 Leu Ala Pro Pro Pro His Leu Pro Gly Pro Pro Pro Gly Ala Pro Pro
 340 345 350
 Pro Ala Pro His Val Asn Pro Ala Phe Phe Pro Pro Pro Thr Asn Ser
 355 360 365
 Gly Met Pro Thr Ser Asp Ser Arg Gly Pro Pro Pro Thr Asp Pro Tyr
 370 375 380
 Gly Arg Pro Pro Pro Tyr Asp Arg Gly Asp Tyr Gly Pro Pro Gly Arg
 385 390 395 400
 Glu Met Asp Thr Ala Arg Thr Pro Leu Ser Glu Ala Glu Phe Glu Glu
 405 410 415
 Ile Met Asn Arg Asn Arg Ala Ile Ser Ser Ser Ala Ile Ser Arg Ala
 420 425 430
 Val Ser Asp Ala Ser Ala Gly Asp Tyr Gly Ser Ala Ile Glu Thr Leu
 435 440 445
 Val Thr Ala Ile Ser Leu Ile Lys Gln Ser Lys Val Ser Ala Asp Asp
 450 455 460
 Arg Cys Lys Gly Leu Ile Ser Ser Leu Gln Asp Cys Leu His Gly Ile
 465 470 475 480
 Glu Ser Lys Ser Tyr Gly Ser Gly Ser Arg Arg Arg Glu Arg Ser Arg
 485 490 495
 Glu Arg Asp His Ser Arg Ser Arg Glu Lys Ser Arg Arg His Lys Ser
 500 505 510
 Arg Ser Arg Asp Arg His Asp Asp Tyr Tyr Arg Glu Arg Ser Arg Glu
 515 520 525
 Arg Glu Arg His Arg Asp Arg Asp Arg Asp Arg Asp Arg Glu Arg Asp
 530 535 540
 Arg Glu Arg Glu Tyr Arg His Arg

545

550

<210> 2

<211> 1659

<212> DNA

<213>大鼠 (*Rattus norvegicus*)

<400> 2

atggcggacg gtgtggacca catagacatt tacgcggatg tgggggaaga gttcaaccag	60
gaagcggaat atggtgggca tgatcagata gatttgtatg atgatgtcat ctctccatct	120
gcaaataatg gcgatgcccc agaagaccgg gactacatgg atactcttcc accaactggt	180
ggtgatgatg tgggtaaagg ggcagcacca aacgttgtct atacttacac tggaaagaga	240
atcgcattgt atattgaaa tctaacatgg tggacaacag atgaggacct aactgaagca	300
gttcattctt tgggagtaaa tgatattttg gagataaaat tttttgaaa tggggcaaat	360
ggacaatcaa agggatttgc ccttgttggg gttggatctg aagcatcttc caaaaagtta	420
atggatcttt tgcctaaaag agaacttcac ggtcagaatc ctggtgtaac tccatgcaat	480
aaacagttcc tgagtcagtt tgaaatgcaa tccaggaaaa ctacacaatc aggacaaatg	540
tctggggaag gtaaagctgg tctccagga ggcggttcac gcgcagcatt tccacaaggt	600
ggtagaggac ggggccgttt tccaggggct gttcctggtg gggacagatt tcttgggcca	660
gcaggaccag gagggccacc tccaccttc ccagctggac agactccacc acgtccacct	720
ctaggtccac ctggcccacc tggccccca ggtcctccac ctcccgtca ggttctgcca	780
cctcctctag caggacctcc taatcgagga gaccgcctc caccaccagt tcttttctct	840
ggacaacctt ttgggcagcc tccgctgggt ccaactctc ctgggctcc acctccagtt	900
ccaggctacg gccccctcc aggtccaccg cctccacagc agggaccacc tccacctcca	960
ggcccccttc cacctcgacc accaggtcca ctagggcctc cccttacct tgctcctcct	1020
ccacatcttc cgggaccacc tccaggtgcc ccaccaccag ctccacatgt gaatcctgct	1080
ttttccctc caccaactaa tagtggcatg ccaacatcag atagtcgagg tccaccgcca	1140
acagaccat atggccgacc tccaccatat gataggggtg actatggtcc ccctgggagg	1200
gaaatggata ctgcaagaac acctctgagt gaagctgagt ttgaagaaat catgaataga	1260
aatcgggcaa tctctagcag tgctatttca agagctgtat ctgatgctag tgctggtgat	1320
tatgggagtg ctattgaaac attggttaaca gcaatttctc taattaaaca atccaaagta	1380
tctgccgacg atcgtgcaa aggtcttatt agctctttgc aagactgect tcatggaatt	1440

gagtccaagt cctatgggtc tggatcaaga agacgtgaac gatcaagaga aaggaccat 1500
 agtagatcac gggaaaagag tcgacgtcat aaatctcgga gtagagatcg ccatgatgat 1560
 tattacagag agagaagcag agagcgagag agacaccggg atcgcgatcg ggatcgtgac 1620
 cgggagcgcg accgagagcg agaataccgt catcgttag 1659

<210> 3

<211> 551

<212> PRT

<213> 人属人 (*Homo sapiens*)

<400> 3

Met Ala Asp Gly Val Asp His Ile Asp Ile Tyr Ala Asp Val Gly Glu
 1 5 10 15
 Glu Phe Asn Gln Glu Ala Glu Tyr Gly Gly His Asp Gln Ile Asp Leu
 20 25 30
 Tyr Asp Asp Val Ile Ser Pro Ser Ala Asn Asn Gly Asp Ala Pro Glu
 35 40 45
 Asp Arg Asp Tyr Met Asp Thr Leu Pro Pro Thr Val Gly Asp Asp Val
 50 55 60
 Gly Lys Gly Ala Ala Pro Asn Val Val Tyr Thr Tyr Thr Gly Lys Arg
 65 70 75 80
 Ile Ala Leu Tyr Ile Gly Asn Leu Thr Trp Trp Thr Thr Asp Glu Asp
 85 90 95
 Leu Thr Glu Ala Val His Ser Leu Gly Val Asn Asp Ile Leu Glu Ile
 100 105 110
 Lys Phe Phe Glu Asn Arg Ala Asn Gly Gln Ser Lys Gly Phe Ala Leu
 115 120 125
 Val Gly Val Gly Ser Glu Ala Ser Ser Lys Lys Leu Met Asp Leu Leu
 130 135 140
 Pro Lys Arg Glu Leu His Gly Gln Asn Pro Val Val Thr Pro Cys Asn
 145 150 155 160
 Lys Gln Phe Leu Ser Gln Phe Glu Met Gln Ser Arg Lys Thr Thr Gln

	165		170		175
Ser Gly Gln Met Ser Gly Glu Gly Lys Ala Gly Pro Pro Gly Gly Ser					
	180		185		190
Ser Arg Ala Ala Phe Pro Gln Gly Gly Arg Gly Arg Gly Arg Phe Pro					
	195		200		205
Gly Ala Val Pro Gly Gly Asp Arg Phe Pro Gly Pro Ala Gly Pro Gly					
	210		215		220
Gly Pro Pro Pro Pro Phe Pro Ala Gly Gln Thr Pro Pro Arg Pro Pro					
	225		230		235
Leu Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Pro Pro Gly					
	245		250		255
Gln Val Leu Pro Pro Pro Leu Ala Gly Pro Pro Asn Arg Gly Asp Arg					
	260		265		270
Pro Pro Pro Pro Val Leu Phe Pro Gly Gln Pro Phe Gly Gln Pro Pro					
	275		280		285
Leu Gly Pro Leu Pro Pro Gly Pro Pro Pro Pro Val Pro Gly Tyr Gly					
	290		295		300
Pro Pro Pro Gly Pro Pro Pro Pro Gln Gln Gly Pro Pro Pro Pro Pro					
	305		310		315
Gly Pro Phe Pro Pro Arg Pro Pro Gly Pro Leu Gly Pro Pro Leu Thr					
	325		330		335
Leu Ala Pro Pro Pro His Leu Pro Gly Pro Pro Pro Gly Ala Pro Pro					
	340		345		350
Pro Ala Pro His Val Asn Pro Ala Phe Phe Pro Pro Pro Thr Asn Ser					
	355		360		365
Gly Met Pro Thr Ser Asp Ser Arg Gly Pro Pro Pro Thr Asp Pro Tyr					
	370		375		380
Gly Arg Pro Pro Pro Tyr Asp Arg Gly Asp Tyr Gly Pro Pro Gly Arg					
	385		390		395
Glu Met Asp Thr Ala Arg Thr Pro Leu Ser Glu Ala Glu Phe Glu Glu					
	405		410		415
Ile Met Asn Arg Asn Arg Ala Ile Ser Ser Ser Ala Ile Ser Arg Ala					

	420		425		430
Val Ser Asp Ala Ser Ala Gly Asp Tyr Gly Ser Ala Ile Glu Thr Leu					
	435		440		445
Val Thr Ala Ile Ser Leu Ile Lys Gln Ser Lys Val Ser Ala Asp Asp					
	450		455		460
Arg Cys Lys Val Leu Ile Ser Ser Leu Gln Asp Cys Leu His Gly Ile					
465		470		475	480
Glu Ser Lys Ser Tyr Gly Ser Gly Ser Arg Arg Glu Arg Ser Arg Glu					
	485		490		495
Arg Asp His Ser Arg Ser Arg Glu Lys Ser Arg Arg His Lys Ser Arg					
	500		505		510
Ser Arg Asp Arg His Asp Asp Tyr Tyr Arg Glu Arg Ser Arg Glu Arg					
	515		520		525
Glu Arg His Arg Asp Arg Asp Arg Asp Arg Asp Arg Glu Arg Asp Arg					
	530		535		540
Glu Arg Glu Tyr Arg His Arg					
545		550			

<210> 4

<211> 1656

<212> DNA

<213>人属人 (*Homo sapiens*)

<400> 4

atggcggacg gcggtggacca catagacatt tacgcggatg tggcgaaga gttcaaccag	60
gaagctgaat atggtgggca tgatcagata gatttgtatg acgatgtcat atctccatct	120
gcaaataatg gagatgcccc agaagaccga gattacatgg atactctccc accaactggt	180
ggtgatgatg tgggtaaagg agcagcacca aatgttgtct atacatatac tggaaagaga	240
attgcattat atattggaaa tctaacatgg tggacaacag atgaagactt aactgaagca	300
gttcattctt tgggagtaaa tgatattttg gagataaaat tttttgaaaa tgggcaaat	360
ggccagtcaa aggggtttgc cttgtttggt gttggatctg aagcatcttc aaaaaagtta	420
atggatctgt tacctaaaag agaacttcat ggtcagaatc ctgttgtaac tccatgcaat	480

aaacagttcc tgagtcaatt tgaaatgcag tccaggaaaa ctacacaatc aggacaaatg 540
 tctggggaag gtaaagctgg tcctccagga ggcagttccc gtgcagcatt tccacaaggt 600
 ggtagaggac ggggccgttt tccaggggct gttcctgggtg gggacagatt tcctgggcca 660
 gcaggaccag gagggccacc cccacctttt ccagctggac agactccacc acgtccaccc 720
 ttaggtcctc caggcccacc tgggccacca ggtcctccac ctctctggta ggttctgctt 780
 cctcctctag ctgggcctcc taatcgagga gatcgccctc caccaccagt tctttttcct 840
 ggacaacctt ttgggcagcc tccattgggt ccaacttctc ctggccctcc acctccagtt 900
 ccaggctacg gccccctcc tggcccacca cctccacaac agggaccacc tccacctcca 960
 ggcccccttc cacctcgtcc acccgtcca cttggggccac cccttacct agctcctcct 1020
 ccgcatcttc ctggaccacc tccaggtgcc ccaccgccag ctccgcatgt gaaccagct 1080
 ttctttcctc caccaactaa cagtggcatg cctacatcag atagccgagg tccaccacca 1140
 acagatccat atgggcgacc tccaccatat gataggggtg actatggccc ccctggaagg 1200
 gaaatggata ctgcaagaac gccattgagt gaagctgaat ttgaagaaat catgaataga 1260
 aatagggcaa tctcaagcag tgctatttcg agagctgtgt ctgatgccag tgctgggtgat 1320
 tatgggagtg ctattgagac actggtaact gcaatttctt taattaaaca atccaaagta 1380
 tctgctgatg atcgttgcaa agttcttatt agttctttgc aagattgcct tcatggaatt 1440
 gagtccaagt cttatggttc tggatcaaga cgtgaacgat caagagagag ggaccatagt 1500
 agatcacgag aaaagagtcg acgtcataaa tcccgtagta gagaccgtca tgacgattat 1560
 tacagagaga gaagcagaga acgagagagg caccgggatc gtgaccgaga ccgtgaccga 1620
 gagcgtgacc gagagcgcga atatcgtcat cgttag 1656

<210> 5

<211> 453

<212> PRT

<213> 人工序列

<400> 5

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

50						55						60					
Gly	Lys	Gly	Ala	Ala	Pro	Asn	Val	Val	Tyr	Thr	Tyr	Thr	Gly	Lys	Arg		
65					70					75					80		
Ile	Ala	Leu	Tyr	Ile	Gly	Asn	Leu	Thr	Trp	Trp	Thr	Thr	Asp	Glu	Asp		
				85					90					95			
Leu	Thr	Glu	Ala	Val	His	Ser	Leu	Gly	Val	Asn	Asp	Ile	Leu	Glu	Ile		
			100					105					110				
Lys	Phe	Phe	Glu	Asn	Arg	Ala	Asn	Gly	Gln	Ser	Lys	Gly	Phe	Ala	Leu		
			115				120					125					
Val	Gly	Val	Gly	Ser	Glu	Ala	Ser	Ser	Lys	Lys	Leu	Met	Asp	Leu	Leu		
			130				135					140					
Pro	Lys	Arg	Glu	Leu	His	Gly	Gln	Asn	Pro	Val	Val	Thr	Pro	Cys	Asn		
145					150						155				160		
Lys	Gln	Phe	Leu	Ser	Gln	Phe	Glu	Met	Gln	Ser	Arg	Lys	Thr	Thr	Gln		
				165					170						175		
Ser	Gly	Gln	Met	Ser	Gly	Glu	Gly	Lys	Ala	Gly	Pro	Pro	Gly	Gly	Gly		
			180					185						190			
Ser	Arg	Ala	Ala	Phe	Pro	Gln	Gly	Gly	Arg	Gly	Arg	Gly	Arg	Phe	Pro		
			195				200						205				
Gly	Ala	Val	Pro	Gly	Gly	Asp	Arg	Phe	Pro	Gly	Pro	Ala	Gly	Pro	Gly		
			210				215						220				
Gly	Pro	Pro	Pro	Pro	Phe	Pro	Ala	Gly	Gln	Thr	Pro	Pro	Arg	Pro	Pro		
225					230						235				240		
Leu	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Pro	Pro	Gly		
				245							250				255		
Gln	Val	Leu	Pro	Pro	Pro	Leu	Ala	Gly	Pro	Pro	Asn	Arg	Gly	Asp	Arg		
			260					265						270			
Pro	Pro	Pro	Pro	Val	Leu	Phe	Pro	Gly	Gln	Pro	Phe	Gly	Gln	Pro	Pro		
			275				280						285				
Leu	Gly	Pro	Leu	Pro	Pro	Gly	Pro	Pro	Pro	Pro	Val	Pro	Gly	Tyr	Gly		
			290				295						300				
Pro	Pro	Pro	Gly	Pro	Pro	Pro	Pro	Gln	Gln	Gly	Pro	Pro	Pro	Pro	Pro		
305						310					315				320		
Gly	Pro	Phe	Pro	Pro	Arg	Pro	Pro	Gly	Pro	Leu	Gly	Pro	Pro	Leu	Thr		
				325							330				335		
Leu	Ala	Pro	Pro	Pro	His	Leu	Pro	Gly	Pro	Pro	Pro	Gly	Ala	Pro	Pro		
				340				345					350				
Pro	Ala	Pro	His	Val	Asn	Pro	Ala	Phe	Phe	Pro	Pro	Pro	Thr	Asn	Ser		
			355				360						365				
Gly	Met	Pro	Thr	Ser	Asp	Ser	Arg	Gly	Pro	Pro	Pro	Thr	Asp	Pro	Tyr		
			370				375						380				
Gly	Arg	Pro	Pro	Pro	Tyr	Asp	Arg	Gly	Asp	Tyr	Gly	Pro	Pro	Gly	Arg		

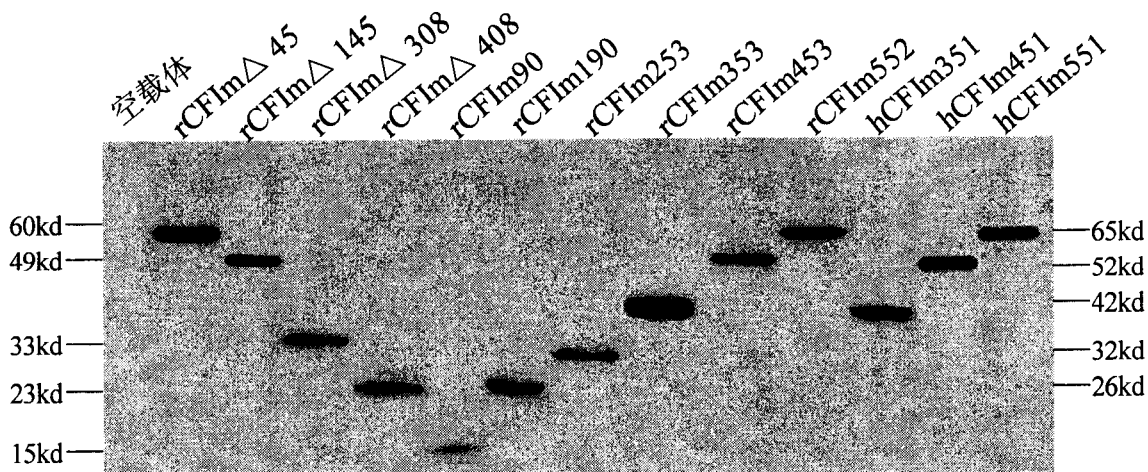


图 1

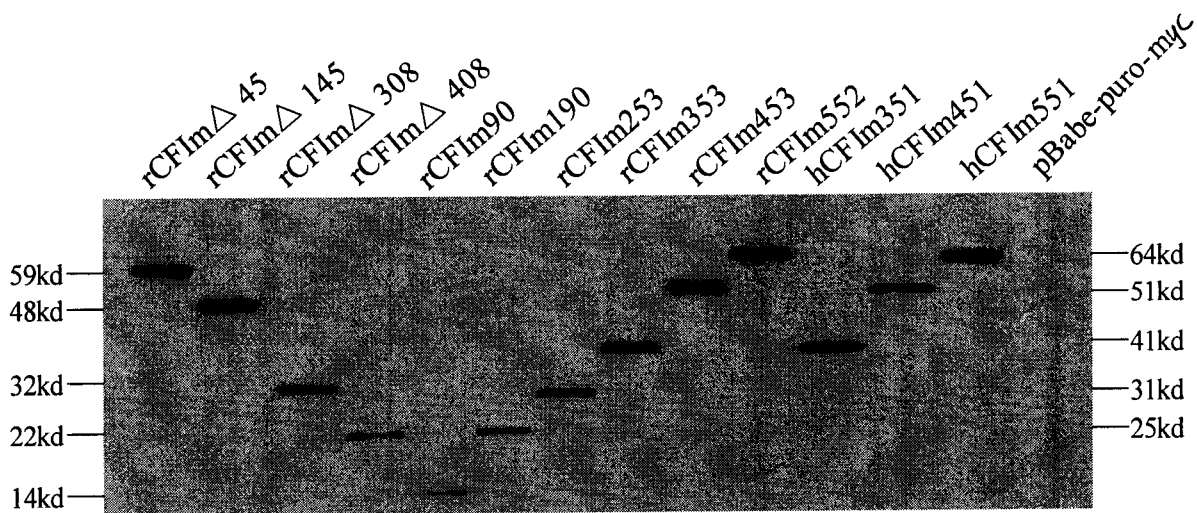


图 2

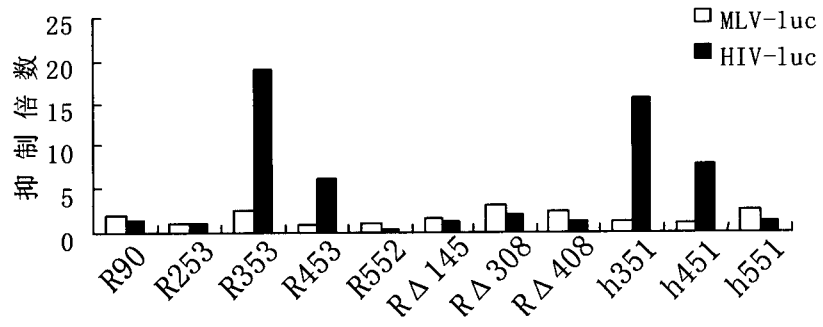


图 3

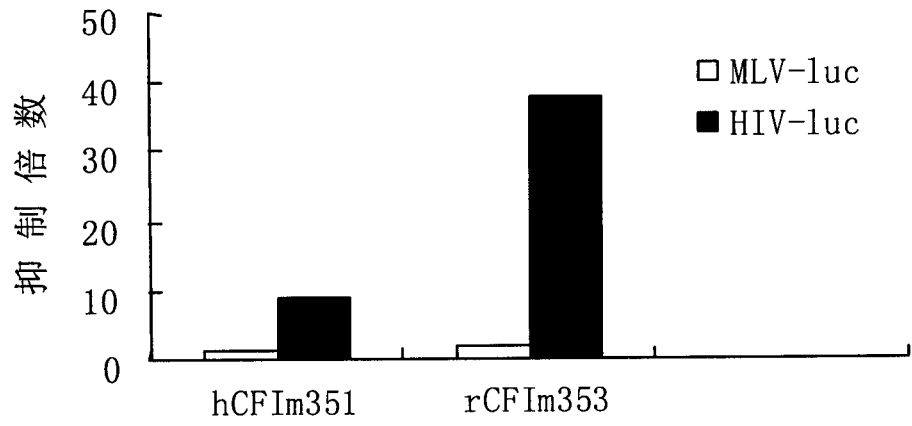


图 4

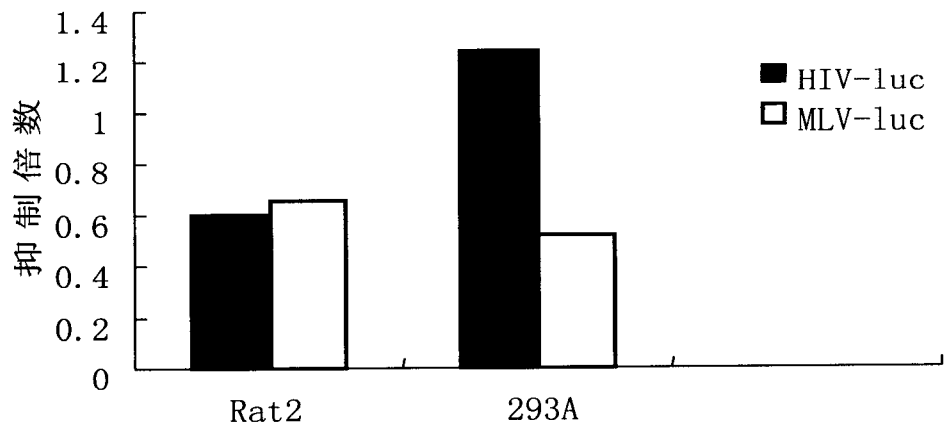


图 5

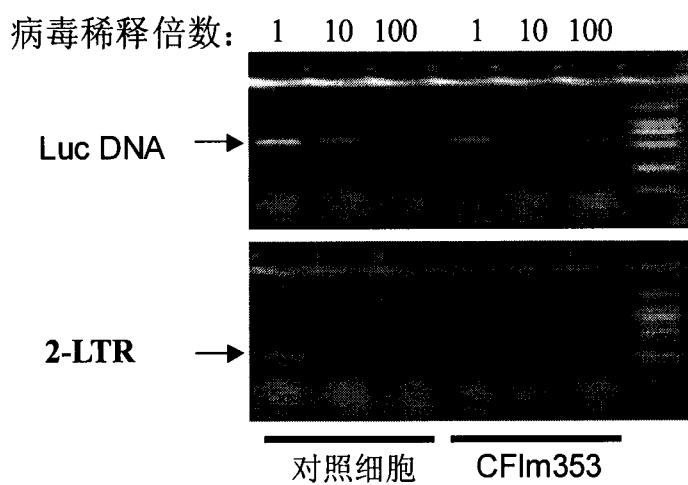


图 6

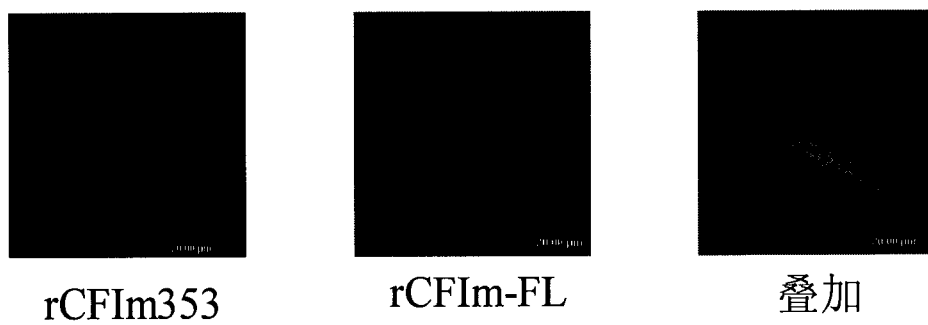


图 7