



(12) 发明专利

(10) 授权公告号 CN 102985538 B

(45) 授权公告日 2015. 09. 09

(21) 申请号 201180031009. 3

C12N 15/55(2006. 01)

(22) 申请日 2011. 04. 23

A61K 38/46(2006. 01)

(30) 优先权数据

61/343, 177 2010. 04. 23 US

61/396, 376 2010. 05. 26 US

61/403, 011 2010. 09. 09 US

61/456, 014 2010. 10. 29 US

61/432, 372 2011. 01. 13 US

(85) PCT国际申请进入国家阶段日

2012. 12. 21

(86) PCT国际申请的申请数据

PCT/US2011/033699 2011. 04. 23

(87) PCT国际申请的公布数据

W02011/133960 EN 2011. 10. 27

(73) 专利权人 辛那杰瓦生物制药股份有限公司

地址 美国马萨诸塞州

(72) 发明人 A·奎因 A·J·哈维

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 陈文平

(51) Int. Cl.

C12N 9/20(2006. 01)

(56) 对比文件

Oliver Zschenker 等. Systematic Mutagenesis of Potential Glycosylation Sites of Lysosomal Acid Lipase. 《J. Biochem. 》. 2005, 第 137 卷摘要, 表 1.

Oliver Zschenker 等. Systematic Mutagenesis of Potential Glycosylation Sites of Lysosomal Acid Lipase. 《J. Biochem. 》. 2005, 第 137 卷摘要, 表 1.

审查员 田颖

权利要求书2页 说明书39页

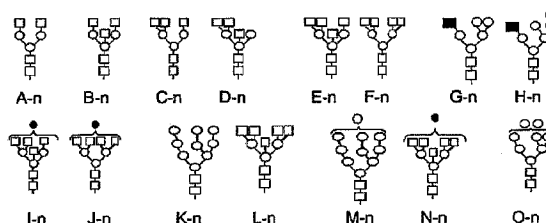
序列表27页 附图31页

(54) 发明名称

溶酶体贮积病酶

(57) 摘要

本发明提供了具有特定糖基化模式以内化至靶细胞中的重组人溶酶体酸性脂酶的组合物, 包含编码人溶酶体酸性脂酶的核酸的载体, 用所述载体转化的宿主细胞, 包含重组人溶酶体酸性脂酶的药物组合物, 和治疗与溶酶体酸性脂酶缺乏相关的病况的方法。



正方形 = N-乙酰葡萄糖胺

实心正方形 = 甘露糖-6-磷酸

圆圈 = 甘露糖

实心圆 = 半乳糖

实心三角形 = 岩藻糖

1. 一种组合物,其包含分离的、含有一种或多种 N- 聚糖结构的人重组溶酶体酸性脂酶(rhLAL),其中所述 rhLAL 在 SEQ ID NO :2 的 Asn¹⁵、Asn⁸⁰、Asn¹⁴⁰、Asn²⁵²和 Asn³⁰⁰上被 N- 连接糖基化,其中:

a) 所述在 Asn⁸⁰上的 N- 聚糖结构包含磷酸化甘露糖;

b) 所述 rhLAL 包含如下 N- 连接糖基化谱:

i) 在 Asn¹⁵上, GlcNAc4Man3GlcNAc2, 或

GallGlcNAc4Man3GlcNAc2;

ii) 在 Asn¹⁴⁰上, Phos1Man6GlcNAc2,

GlcNAc1Phos1Man6GlcNAc2;

Man3GlcNAc2;

GlcNAc2Man3GlcNAc2;

GlcNAc3Man3GlcNAc2;

GlcNAc4Man3GlcNAc2, 或

GallGlcNAc4Man3GlcNAc2;

iii) 在 Asn²⁵²上, Man7GlcNAc2,

Man8GlcNAc2,

Man9GlcNAc2,

Phos1Man8GlcNAc2, 或

Phos1Man9GlcNAc2; 和

iv) 在 Asn³⁰⁰上, GlcNAc2Man3GlcNAc2,

GlcNAc3Man3GlcNAc2,

GlcNAc4Man3GlcNAc2,

GallGlcNAc4Man3GlcNAc2,

GlcNAc5Man3GlcNAc2,

GallGlcNAc5Man3GlcNAc2,

GlcNAc6Man3GlcNAc2, 或

GallGlcNAc6Man3GlcNAc2

其中 Man = 甘露糖,

GlcNAc = N- 乙酰葡萄糖胺,

Phos = 磷酸,

Gal = 半乳糖。

2. 根据权利要求 1 所述的组合物,其中所述 rhLAL 包含在 Asn⁸⁰上包含双磷酸化甘露糖的 N- 聚糖结构。

3. 根据权利要求 1 所述的组合物,其中所述 rhLAL 包含在 Asn¹⁴⁰上的 Phos1 Man6GlcNAc2 或 GlcNAc1 Phos1 Man6GlcNAc2。

4. 根据权利要求 1 所述的组合物,其中所述 rhLAL 包含在 Asn²⁵²上的 Phos1 Man8GlcNAc2 或 Phos1 Man9GlcNAc2。

5. 根据权利要求 1 所述的组合物,其中所述 rhLAL 包含在 Asn⁸⁰上包含高-甘露糖基团的 N- 聚糖结构,并且所述高-甘露糖基团包含 6、7、8、9 或 10 个甘露糖。

6. 根据权利要求 1 所述的组合物,其中所述 rhLAL 不包含唾液酸和木糖。
7. 根据权利要求 1 所述的组合物,其中所述 rhLAL 不包含岩藻糖。
8. 根据权利要求 1 所述的组合物,其中在 Asn⁸⁰上的 N-聚糖结构包含 Phos2Man7GlcNAc2。
9. 权利要求 1-8 中任一项所述的组合物,其中所述 rhLAL 在 Asn⁵¹上是未糖基化的。
10. 根据权利要求 1-9 任一项所述的组合物,其中所述 rhLAL 在种系转基因禽类中产生。
11. 根据权利要求 10 所述的组合物,其中所述 rhLAL 从所述种系转基因禽类的输卵管细胞产生。
12. 根据权利要求 11 所述的组合物,其中所述禽类为鸡。
13. 一种制备权利要求 1 的 rhLAL 的方法,其包括产生包含编码可操作地连接于启动子的所述 rhLAL 的转基因并且在输卵管细胞中表达所述 rhLAL 的种系转基因禽类,其中所述 rhLAL 在所述输卵管细胞中被糖基化并且被储存至由所述转基因禽类所产的硬壳卵中,并且从所述卵的卵清获得所述 rhLAL。
14. 一种药物制剂,其包含与药学上可接受的载体、稀释剂或赋形剂组合的根据权利要求 1 所述的组合物。
15. 根据权利要求 14 所述的药物制剂,其中所述药物制剂包含至少一种选自脱水柠檬酸三钠、柠檬酸和人血清白蛋白的试剂。
16. 根据权利要求 15 所述的药物制剂,其中在维持于 5.6 至 6.2 的 pH 的水溶液中提供所述药物制剂。

溶酶体贮积病酶

[0001] 相关申请

[0002] 本申请要求 2010 年 4 月 23 日提交的美国临时申请 No. 61/343, 177、2010 年 5 月 26 日提交的美国临时申请 No. 61/396, 376、2010 年 9 月 9 日提交的美国临时申请 No. 61/403, 011、2010 年 10 月 29 日提交的美国临时申请 No. 61/456, 014、2011 年 1 月 13 日提交的美国临时申请 No. 61/432, 372 的权益。将上述申请的全部教导通过引用并入本文。

[0003] 发明背景

[0004] 溶酶体酸性脂酶 (LAL) 缺乏是极罕见的溶酶体贮积病 (LSD), 其特征在于因该酶的缺乏而不能在溶酶体中分解胆固醇酯 (CE) 和甘油三酯 (TAG)。LAL 缺乏与在许多组织和细胞类型具有底物积累的其它溶酶体贮积症相似。在 LAL 缺乏中, 底物积累在网状内皮系统的细胞 (包括肝中的肝巨噬细胞、脾中和小肠的粘膜固有层中的组织细胞) 中是最显著的。网状内皮细胞表达巨噬细胞甘露糖 /N- 乙酰葡萄糖胺受体 (也称为巨噬细胞甘露糖受体或 MMR、CD206), 其介导具有以 GlcNAc 或甘露糖终止的 N- 聚糖的蛋白质的结合、细胞摄取和溶酶体内化, 并且提供此类至关重要的细胞类型的酶缺乏的潜在校正的途径。

[0005] LAL 缺乏是多系统疾病, 其最常见地表现出胃肠、肝和心血管并发症, 并且与显著的发病率或死亡率相关。LAL 缺乏的临床效应因脂质材料在许多组织的溶酶体中大量积累以及胆固醇和脂质动态平衡机制的深度扰乱 (包括肝胆固醇合成的显著增加) 而引起。LAL 缺乏呈现至少两种表型沃尔曼病 (Wolman disease) (WD) 和胆固醇酯沉积病 (CESD)。

[0006] 沃尔曼病是 LAL 缺乏的最具侵袭性的呈现。该表型的特征在于胃肠和肝表现, 包括生长不足、吸收不良、脂肪泻、严重的体重减轻和肝肿大。沃尔曼病进展迅速并且在生命的第一年内通常是致命的。案例报告综述显示超过 12 月龄的存活对于在生命的第一年中呈现因 LAL 缺乏而引起的生长不足的患者是极罕见的。在该最具侵袭性的形式中, 生长不足是主要临床特征并且是早期死亡率的主要促因 (contributor)。如通过肝增大和转氨酶的升高证明的肝损害在婴儿中也是常见的。身体检查发现 (Physical finding) 包括具有肝肿大和脾肿大的腹胀, 并且放射照相检查通常显示肾上腺的钙化。实验室评估典型地显示升高的血清转氨酶的水平及缺乏或显著减小的 LAL 酶活性。还可在患者中看到胆固醇和甘油三酯的升高的血液水平。

[0007] 沃尔曼病的目前治疗选择极其有限。向具有发热和 / 或感染迹象的婴儿施用抗生素。可处方肾上腺功能减退的类固醇替代疗法和专门化营养支持, 然而无证据表明这些干预能预防死亡, 目前还不清楚它们是否对短期存活具有影响。在一系列 4 名利用骨髓移植治疗的具有 LAL 缺乏的患者中, 所有 4 名患者在移植的数月内死于该过程的并发症。

[0008] 具有 LAL 缺乏的患者还可在以后的生活中呈现显著的肝和心血管损害, 这通常称为胆固醇酯沉积病 (CESD)。在 CESD 中, 肝严重地受到显著的肝肿大、肝细胞坏死、转氨酶的升高、肝硬化和纤维化的影响。由于 CE 和 TG 的水平升高, 高脂血症和加速的动脉粥样硬化也见于 LAL 缺乏。特别地, 脂肪沉积物在动脉壁上的积累在生命早期已被描述。沉积使动脉腔变窄, 并且可导致血管阻塞, 从而增加包括心肌梗塞和中风的重大心血管事件的风险。CESD 的表现是高度可变的, 一些患者一直未确诊直至在成人晚期表现出并发症, 然而其他

患者在儿童早期已呈现肝功能障碍。CESD 与缩短的寿命和明显的不健康相关；患有 CESD 的患者的预期寿命取决于相关并发症的严重程度。

[0009] 用于 CESD 表型的目前治疗选择聚焦在通过排除富含胆固醇和甘油三酯的食物的饮食和通过施用降胆固醇药抑制胆固醇合成和载脂蛋白 B 产生来控制脂质积累。虽然可见一些临床改善,但潜在的疾病表现继续存在并且疾病进展仍然存在。

[0010] 在大多数案例中,LAL 缺乏的疗法需要终身治疗。此外,由于蛋白质治疗剂的高成本,期望施用最少的有效量的治疗剂来治疗 LAL 缺乏。然而,迄今为止,还没有治疗 LAL 缺乏,特别地患有沃尔曼病和 CESD 的患者的有效疗法。因此,存在对利用具有最低的施用频率以改善患者的生活质量的有效疗法的强烈需要。还存在对高表达且强劲的蛋白质生产平台的需要,所述蛋白质生产平台可产生稳定的并且高效地靶向患者的患病组织细胞中的溶酶体区室的 LAL 蛋白。

[0011] 发明概述

[0012] 本文公开了特别适合用于疗法、例如用于与 LAL 缺乏相关的病况的治疗的 LAL 的组合物。本文中描述的 LAL 分子包含特定的聚糖结构,当施用至受试者例如人受试者时该结构提供至细胞的溶酶体内的高效且快速的摄取。

[0013] 在一个方面,本文中公开的组合物包含人 LAL,其中相当高百分比的人 LAL 包含至少一个甘露糖-6-磷酸聚糖部分,其可用作用于通过在例如肝细胞上发现的细胞表面上的甘露糖-6-磷酸受体内化的配体。在一个实施方案中,组合物中包含的 LAL 的 30%或更多,例如至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 97%或至少 99%包含至少一个甘露糖-6-磷酸部分。甘露糖-6-磷酸部分可见于例如位于选自 SEQ ID NO:2 的 Asn¹⁵、Asn⁵¹、Asn⁸⁰、Asn¹⁴⁰、Asn²⁵²和 Asn³⁰⁰的一个或多个残基上的 N-聚糖结构上。

[0014] 在另一个方面,本文中公开的组合物包含人 LAL,其中相当高百分比的人 LAL 在任何其 N-聚糖结构中不包含唾液酸部分,这有时可干扰酶至细胞中的内化。在一个实施方案中,15%或更少,例如 10%或更少,5%或更少,2%或更少,1%或更少或基本上无包含于组合物中的 LAL 在任何其 N-聚糖结构中不包含唾液酸部分。

[0015] 在另一个方面,本文中公开的组合物包含人 LAL,其中相当高百分比的人 LAL 在任何其 N-聚糖结构中不包含岩藻糖部分。在一个实施方案中,包含于组合物中的 LAL 中的 50%或更少,例如,50%或更少,40%或更少,30%或更少,20%或更少,10%或更少,5%或更少,2%或更少,1%或更少包含岩藻糖部分,或包含于组合物中的 LAL 基本上在任何其 N-聚糖结构中不包含岩藻糖部分。

[0016] 在另一个方面,描述了适用于产生含 LAL 的组合物载体、宿主细胞、表达系统和相关方法。

[0017] 通常,本文中讨论的和公开的本发明的 LAL 为人 LAL。在一个实施方案中,包含 LAL 的组合物包括具有如下氨基酸序列的成熟 LAL:

[0018] SGGKLTAVDPETNMNVSEIIISYWGFPEEYLVETEDGYILCLNRIHPGRKNHSDKGPKPVVFLQHGL

[0019] LADSSNWVTNLANSSLGFILADAGFDVWMGNSRGNTWSRKHKTLSSVSQDEFWAFSYDEMAKYDLPAS

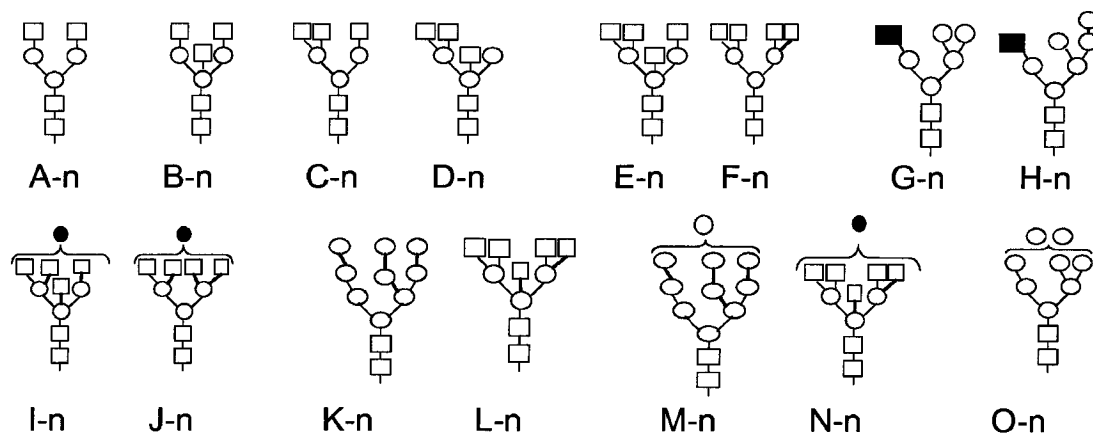
[0020] INFILNKTGQEQVYYVGHSGQTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDH

[0021] LIKDLFGDKFLPQSAFLKWLGTHTVCTHVLKELCGNLCFLLCGFNERNLMSRVDVYTTTHSPAGTS

- [0022] VQNMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPITYNVKMDLVPTAVWSGGHDWLADVYDVNILL
- [0023] LTQITNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ (SEQ ID NO :2) .
- [0024] 在另一个实施方案中,成熟 LAL 具有氨基酸序列 :
- [0025] GKLTAVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLA
- [0026] DSSNWVTNLANSSLGFILADAGFDVWMGNSRGNTWSRKHKTLSSVSQDEFWAFSYDEMAKYDLPASIN
- [0027] FILNKTGQEQQVYYVGHSGGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLI
- [0028] KDLFGDKEFLPOS AFLKWLGTHVCTHVILHEL CGNL CFLLCGFNERNLNMSRVDVYTTTHSPAGTSVQ
- [0029] NMLHWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPITYNVKMDLVPTAVWSGGHDWLADVYDVNILLT
- [0030] QITNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ (SEQ ID NO :3) .
- [0031] 在另一个实施方案中,成熟 LAL 具有氨基酸序列 :
- [0032] TAVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSS
- [0033] NWVTNLANSSLGFILADAGFDVWMGNSRGNTWSRKHKTLSSVSQDEFWAFSYDEMAKYDLPASINFIL
- [0034] NKTGQEQQVYYVGHSGGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDL
- [0035] FGDKEFLPQSAFLKWLGTHVCTHVILKELCGNL CFLLCGFNERNLNMSRVDVYTTTHSPAGTSVQNML
- [0036] HWSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPITYNVKMDLVPTAVWSGGHDWLADVYDVNILLTQIT
- [0037] NLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ (SEQ ID NO :4) .
- [0038] 在另一个实施方案中,成熟 LAL 具有氨基酸序列 :
- [0039] AVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSN
- [0040] WVTNLANSSLGFILADAGFDVWMGNSRGNTWSRKHKTLSSVSQDEFWAFSYDEMAKYDLPASINFILN
- [0041] KTGQEQQVYYVGHSGGTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLF
- [0042] GDKEFLPQSAFLKWLGTHVCTHVILKELCGNL CFLLCGFNERNLNMSRVDVYTTTHSPAGTSVQNMLH
- [0043] WSQAVKFQKFQAFDWGSSAKNYFHYNQSYPPITYNVKMDLVPTAVWSGGHDWLADVYDVNILLTQITN
- [0044] LVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ (SEQ ID NO :19) .
- [0045] 在另一个实施方案中,成熟 LAL 为选自 :SEQ ID NO :2、SEQ ID NO :3、SEQ ID NO :4 和 SEQ ID NO :19 的至少 2 种多肽的混合物。

[0046] 本发明还提供了下述组合物,其包含各类型的有用的蛋白质分子例如本文中公开的那些蛋白质的分离的混合物,其中混合物中包含的蛋白质分子的一种或多种具有连接的特殊寡糖结构,特别地,本文中公开的寡糖结构。例如,本发明提供了 LAL 分子的分离的混合物,例如包含利用下列结构 A-n 至 O-n 的一种或多种糖基化的 LAL 分子的人 LAL 分子 :

[0047]



[0048] 正方形 = N-乙酰葡萄糖胺

[0049] 实心正方形 = 甘露糖-6-磷酸

[0050] 圆圈 = 甘露糖

[0051] 实心圆圈 = 半乳糖

[0052] 实心三解形 = 岩藻糖

[0053] 根据本发明的一个方面,组合物包含任何分离的单一上述多肽或所述多肽的组合。在一个实施方案中,组合物可以是药物组合物,例如,还包含药学上可接受的载体的制剂,以便组合物例如适用于向受试者(例如,人,特别地患有或经诊断患有病况的患者)施用。可以以许多方式(包括通过静脉内施用)来施用组合物。在另一个实施方案中,组合物还包含第二试剂。这样的试剂可以是药剂,或当施用至受试者时可影响或改变生物过程的试剂。例如,第二试剂可以是免疫调节剂。此类免疫调节剂可包括下述任何试剂,所述试剂当与本文中描述的任何 LAL 组合物一起施用(即,同时或之前或之后立即施用)时,可具有减小 LAL 组合物在受试者中的免疫原性的作用(例如,利妥昔单抗或任何其它耗尽 B 细胞的抗体)。

[0054] 在最后的方面,公开了用于治疗与 LAL 缺乏相关的症状的方法和组合物。

[0055] 结合附图和序列,根据下文中所示的详细描述,本发明的另外的目的和方面将变得更加显而易见。

[0056] 附图简述

[0057] 图 1 描述了人 LAL 的氨基酸序列。重组 hLAL 的氨基酸序列显示与天然人 LAL 的氨基酸序列具有 100% 的同源性。hLAL 的成熟形式以下划线标示。

[0058] 图 2 描述了重组 hLAL, pALVIN-OVR1-I-hLAL-dSA 的 rhLAL 转基因的核苷酸序列。

[0059] 图 3A 和 3B 描述了 pALVIN-OVR1-I-hLAL-dSA 及其原病毒 (proviral) 区域的简图。图 3A 描述了用于产生图示的转导颗粒的人 LAL 逆转录病毒表达载体的简图(质粒的 DNA 序列位于附录 A 中)。图 3A 描述了已被整合入基因组的 pALVIN-OVR1-I-hLAL-dSA 原病毒区域。SIN LTR, 自身灭活长末端重复 (self-inactivating long terminal repeat); OVDHSIII 增强子, 卵清蛋白基因的 DNase 酶超敏位点 III; OV 内含子; 卵清蛋白 5' 非翻译区和内含子 1; hLAL, 人 LAL cDNA; OV 3' UTR, 卵清蛋白基因 3' 非翻译区; 部分 gag, 部分 gag 基因; LTR, 长末端重复。

[0060] 图 4 描述了 pALVIN-OVR1-I-hLAL-dSA 的核苷酸序列。

[0061] 图 5 描述了已被整合入基因组内的 pALVIN-OVR1-I-hLAL-dSA 原病毒区域的核苷酸序列。

[0062] 图 6 描述了 pALVIN-OV-1.1-I 载体的核苷酸序列。

[0063] 图 7 描述了 rhLAL 适体的核苷酸序列。

[0064] 图 8 描述了包括部分卵清蛋白启动子的 rhLAL 的核苷酸序列。

[0065] 图 9 描述了 OVR1 启动子的核苷酸序列。

[0066] 图 10 描述了用于构建 pALVIN-OVR1-I-hLAL-dSA 载体的步骤的示意图。

[0067] 图 11 描述了 XLL109 的半合子转基因 G1 后代的血液 DNA 样品的实时 PCR 分析。来自半合子 G1 后代 ILL7466 的一式两份 DNA 样品的信号由在第 22 个循环之前起始 ΔR_n 的增加的曲线指示。显示了两个非转基因后代的曲线; 这些曲线在至少 34 个循环中停留在基

线或基线附近。

[0068] 图 12A-D 描述了具有 ALVIN-OVR1-I-hLAL-dSA 转基因的 G1 鸡的 Southern 分析。图 12A 举例说明了整合的转基因的示意图,并且显示了侧翼基因组区域和转基因 B1pI 位点的已知位置以及侧翼基因组 B1pI 位点的预测位置。OV 启动子探针和 hLAL 编码序列探针(hLAL 探针)的位置通过黑色条棒标示。显示了 Southern 分析中检测到 4.3kb 和 10.6kb 条带的位置以及 4.3kb 和 10.6kb 条带的基因组和转基因部分的预测尺寸。图 12B 举例说明了利用 B1pI 降解并且利用 OV 探针探测的基因组 DNA 的 Southern 印迹。WT CTRL 为从非转基因鸡分离的基因组 DNA。G1 转基因的 ID 号标示于泳道上方。分子量标记物的位置和尺寸(kb)示于印迹的左边。检测到的转基因片段(4.3kb)和内源卵清蛋白基因(4.1kb)的位置和尺寸示于印迹的右边。图 12C 描述了利用 hLAL 探针探测的 Southern 印迹。检测到的转基因片段(10.6kb)的位置和尺寸示于印迹的右边。图 12D 描述了以更大比例描述图 12B 中显示的图像的部分以显示 4.1 和 4.3kb 条带的存在。

[0069] 图 13A 描述了 ALVIN-OVR1-I-hLAL-dSA 转基因的示意图。还显示了预计可被 OV 探针和 hLAL 探针检测的 ApaLI 条带的尺寸。图 13B 描述了用于确认转基因尺寸的 ALVIN-OVR1-I-hLAL-dSA 转基因的 Southern 印迹分析的示意图。利用 ApaLI 降解并且利用 OV 探针(左图)或 hLAL 探针(右图)探测的基因组 DNA 的 Southern 印迹。WT CTRL 为从非转基因鸡分离的基因组 DNA。G1 的 ID 号标示于每个泳道的上方。分子量标记物的位置和尺寸(kb)示于印迹的左边。检测到的转基因片段(OV 启动子探针,3.6kb;hLAL 探针,3.8kb)和内源卵清蛋白基因(7.7kb)的位置和尺寸示于印迹的右边。

[0070] 图 14 描述了转基因鸡的谱系。显示了每只鸡的代数(G0、G1 或 G2)、标识号、性别和孵化日期。其它 G1 鸡为其它谱系的鸡。

[0071] 图 15 描述了从卵清纯化 hLAL 的步骤。

[0072] 图 16 描述了作为按照本发明产生的 LAL 中的 N-连接糖基化结构被发现的 N-聚糖。正方形, N-乙酰葡萄糖胺;实心正方形,甘露糖-6-磷酸;圆圈,甘露糖;实心圆,半乳糖;和实心三角形,岩藻糖。

[0073] 图 17 描述了在 SEQ ID NO:1 中所示的 LAL 多肽(箭头)上标示的预测的 N-聚糖位点的相对位置。显示了在结构上代表在每个位点上检测到的 N-聚糖。正方形, N-乙酰葡萄糖胺;实心正方形,甘露糖-6-磷酸;圆形,甘露糖;实心圆圈,半乳糖;和实心三角形,岩藻糖。

[0074] 图 18 描述了通过 N-糖酰胺酶(PNGase)释放的和通过 MALDI-TOF 分析的磷酸化的 N-聚糖。显示了结构。

[0075] 图 19 描述了 LAL 的去磷酸化对 N-聚糖的 HPAEC-PAD 保留时间的作用。利用细菌碱性磷酸酶对按照本发明产生的 LAL 进行去磷酸化(上图)或不处理所述 LAL(下图)。利用 HPAEC-PAD 分析分析释放的 N-聚糖。

[0076] 图 20 描述了重组人 LAL(SBC-102)和溶酶体标记在此类细胞的溶酶体中的共定位,所述细胞使用顺序扫描模式通过共聚焦荧光显微镜检查。

[0077] 图 21 描述了使用巨噬细胞系 NR8383 通过竞争性结合检验估量的重组人 LAL(SBC-102)对 GlcNAc/甘露糖受体的结合特异性。

[0078] 图 22 描述了重组人 LAL 在体外在正常细胞和 LAL 缺乏细胞中的活性。

[0079] 图 23 描述了重组人 LAL (SBC-102) 处理对 LAL 缺乏大鼠的内脏质量的作用。器官尺寸表示为 8 周龄时测定的在每周以 5mg/kg 施用媒介物或 SBC-102 进行 4 周后 LAL^{-/-}大鼠和 LAL^{+/+}大鼠的体重的百分比。

[0080] 图 24 描述了在每周以 5mg. kg⁻¹施用媒介物或 SBC-102 进行 4 周后野生型和 LAL 缺乏大鼠的体重。剂量施用从第 4 周开始以菱形突出显示于 X 轴上。

[0081] 图 25 显示在 8 周龄时测定的在每周以 5mg. kg⁻¹施用媒介物或重组人 LAL (SBC-102) 进行 4 周后 WT 和 LAL 缺乏大鼠的肝胆固醇、胆固醇酯和甘油三酯的水平。

[0082] 图 26 描述了在 8 周龄时测定的在以指定的水平和方案施用重组人 LAL (SBC-102) 四周后的 LAL 缺乏大鼠的体重的百分比增加。

[0083] 图 27 显示在 8 周龄时测定的在以指定的水平和方案施用 SBC-102 四周后的 LAL 缺乏大鼠的肝重量,以体重的百分比表示。

[0084] 图 28 显示在 8 周龄时测定的在以指定的水平和方案施用 SBC-102 四周后的 LAL 缺乏大鼠的组织胆固醇酯水平。

[0085] 图 29 显示每周施用 1mg/kg 的 LAL 或每周施用 5mg/kg 的 LAL 或每 2 周施用 5mg/kg 的 LAL 的大鼠的体重增长的每日进展。

[0086] 图 30 描述了对处理的动物的总体病理学检查,该检查显示肝尺寸和颜色的大体正常化,如可在上图的解剖中看到的,来自 LAL 处理的大鼠的肝组织的组织病理学显示正常的肝组织学,与在下图中的安慰剂处理的动物的泡沫状巨噬细胞的大量积累形成鲜明对照。

[0087] 发明详述

[0088] 定义

[0089] 本文中提供了某些定义来举例说明和定义用于在本文中描述本发明的各种术语的含义和范围。

[0090] 如本文中所用,对于制剂、组合物或成分,术语“可接受的”意指对正在接受治疗的受试者的一般健康没有持久的有害作用。

[0091] 如本文中所用,术语“施用”或其其它词性形式是指向需要治疗的受试者提供本发明的重组人溶酶体酸性脂酶。

[0092] “核酸或多核苷酸序列”包括但不限于包含天然核苷碱基腺嘌呤、鸟嘌呤、胞嘧啶、胸苷和尿嘧啶的真核 mRNA、cDNA、基因组 DNA 以及合成的 DNA 和 RNA 序列。术语还包括具有一个或多个修饰碱基的序列。

[0093] 如本文中所用,术语“禽类”是指分类学种类 avia 的生物的任何种、亚种或种族,例如但不限于鸡、火鸡、鸭、鹅、鹌鹑、雉鸡、鸚鵡、雀、隼、乌鸦和包括鸵鸟、鸕鹚和鸕鸵的平胸禽类。术语包括原鸡或鸡的各种已知品系(例如,白色来亨鸡(White Leghorn)、棕色来亨鸡(Brown Leghorn)、横斑洛克鸡(Barred-Rock)、苏赛克斯鸡(Sussex)、新罕布什尔鸡(New Hampshire)、罗得岛鸡(Rhode Island)、澳洲黑鸡(Australorp)、米诺卡鸡(Minorca)、Amrox、加利福尼亚灰鸡(California Gray))以及火鸡、雉鸡、鹌鹑、鸭、鸵鸟的品系以及通常商业上大量繁殖的其它禽类。其还包括处于所有发育阶段(包括胚胎期和胎儿期)的单个禽类生物体。

[0094] “治疗性蛋白”或“药物蛋白”包括全部或部分组成药物的氨基酸序列。

[0095] “编码序列”或“开放阅读框”是指当置于适当的调控序列的控制之下时可在体外或体内被转录并且翻译（在DNA的情况下）或翻译（在mRNA的情况下）成多肽的多核苷酸或核酸序列。编码序列的边界通过5′（氨基）端上的翻译起始密码子和3′（羧基）端上的翻译终止密码子来确定。转录终止序列通常位于编码序列的3′。编码序列可在5′和/或3′末端上侧翼连接非翻译区。

[0096] “外显子”是指当转录成细胞核转录物时，在通过细胞核剪接除去内含子或间插序列（intervening sequences）后于细胞质mRNA中“表达”的基因的部分。

[0097] 核酸“控制序列”或“调控序列”是指启动子序列、翻译起始和终止密码子、核糖体结合位点、多腺苷酸化信号、转录终止序列、上游调节区（upstream regulatory domain）、增强子等，这对于给定的编码序列在确定的宿主细胞中的转录和翻译是必要和充分的。适用于真核细胞的控制序列的实例为启动子、多腺苷酸化信号和增强子。并非所有此类控制序列必需存在于重组载体中，只要对于期望的基因转录和翻译是必要和充分的控制序列存在即可。

[0098] “可操作地或操作地连接的”是指为了进行期望的功能而产生的编码序列和控制序列的配置。因此，可操作地连接于编码序列的控制序列能够影响编码序列的表达。当DNA聚合酶结合启动子序列并且将编码序列转录成可被翻译成编码蛋白质的mRNA时，编码序列被可操作地连接于细胞的转录调节区或处于所述转录调节区的控制之下。控制序列不一定与编码序列邻接，只要它们用于指导其表达即可。因此，例如，非翻译但仍转录的间插序列可存在于启动子序列与编码序列之间，并且启动子序列仍然可被认为“可操作地连接于”编码序列。

[0099] 术语“异源的”和“外源的”，当它们涉及核酸序列例如编码序列和控制序列时，是指通常不与重组构建体的区域或不与特定染色体基因座关联（associate）和/或通常不与特定细胞关联的序列。因此，核酸构建体的“外源”区域为另一个核酸分子内的或连接至其的可鉴定的核酸区段，所述核酸区段经发现在自然界中与所述另一个核酸分子无关联。例如，构建体的外源区域可包括侧翼连接有经发现在自然界中与编码序列无关联的序列的编码序列。外源编码序列的另一个实例为其中编码序列本身为非天然发现的（例如，具有与天然基因不同的密码子的合成序列）构建体。类似地，就本发明的目的而言，利用通常不存在于宿主细胞中的构建体或核酸转化的宿主细胞可被认为是外源的。

[0100] 如本文中所用，术语“N-聚糖”、“寡糖”、“寡糖结构”、“糖基化模式”、“糖基化谱”和“糖基化结构”在本质上具有相同的含义并且各自指从糖残基形成的并且连接于糖基化蛋白质的一种或多种结构。

[0101] 如本文中所用，“外源蛋白质”是指非天然存在于特定组织或细胞中的蛋白质，为外源表达构建体或转基因的表达产物的蛋白质，或并非以给定的量天然存在于特定组织或细胞中的蛋白质。对于卵为外源的蛋白质是通常未在卵中发现的蛋白质。例如，对于卵是外源的蛋白质可以是作为存在于产卵动物的转基因中的编码序列的表达的结果而存在于卵中的蛋白质。

[0102] “外源基因”是指通常与特定细胞相关的天然存在的基因或其片段。

[0103] “LAL”意指“人溶酶体酸性脂酶”、“SBC-102”或“人溶酶体酸性脂酶分子”，并且这些术语在整个本说明书中可互换使用。

[0104] 本文中描述的表达产物可由具有确定的化学结构的蛋白质性质的材料组成。然而,精确的结构取决于许多因素,特别地对于蛋白质是常见的化学修饰。例如,由于所有蛋白质包含可电离的氨基酸和羧基,因此蛋白质可以以酸性或碱性盐形式或以中性形式获得。可使用糖分子(糖基化)或利用其它化学衍生(包括与例如脂质、磷酸、乙酰基等的共价或离子连接)来衍生(通常通过与糖结合来进行)一级氨基酸序列。此类修饰可体外或体内发生,后者由宿主细胞通过翻译后加工系统来进行。此类修饰可增强或减弱分子的生物学活性,并且此类化学修饰分子也意欲在本发明的范围之内。

[0105] 克隆、扩增、表达和纯化的可选择的方法对于本领域技术人员来说是显而易见的。代表性方法公开于 Sambrook, Fritsch 和 Maniatis, *Molecular Cloning, a Laboratory Manual*, 第 2 版, Cold Spring Harbor Laboratory (1989) 中。

[0106] “载体”意指由单链、双链、环状或超螺旋 DNA 或 RNA 组成的多核苷酸。典型的载体可由以适当的距离操作地连接(以允许功能性基因表达)的下列元件组成:复制起点、启动子、增强子、5' mRNA 前导序列、核糖体结合位点、核酸盒、终止和多腺苷酸化位点以及可选择标记序列。在具体应用中可省去此类元件的一个或多个。核酸盒可包括用于插入待表达的核酸序列的限制位点。在功能性载体中,核酸盒包含待表达的核酸序列,包括翻译起始和终止位点。内含子任选地可包含在构建体中,例如,编码序列的 5'。构建载体以便特定编码序列位于具有适当的调控序列的载体中,编码序列相对于控制序列如此定位和定向使得编码序列在控制或调控序列的“控制”下转录。编码特定目标蛋白质的序列的修饰可期望用于实现该目的。例如,在一些情况下中,可能必须对序列加以修饰以便其可以以适当的方向连接于控制序列;或维持阅读框。可在插入载体之前将控制序列和其它调控序列连接于编码序列。或者,可将编码序列直接克隆入已包含控制序列和适当限制位点的表达载体,所述限制位点与控制序列在阅读框中和处于控制序列的调控之下。

[0107] “启动子”为 DNA 上 RNA 聚合酶与其结合以起始基因的转录的位点。在一些实施方案中,启动子可通过添加或缺失序列,或用可选择的序列(包括天然和合成序列以及可以为合成与天然序列的组合物)替代来进行修饰。许多真核启动子包含两种类型的识别序列:TATA 盒和上游启动子元件。前者,位于转录起始位点的上游,参与指导 RNA 聚合酶在正确的位点起始转录,而后者似乎决定转录率并且位于 TATA 盒的上游。增强子元件还可刺激从连接的启动子转录,但许多增强子元件专门地在特定细胞类型中起作用。衍生自病毒的许多增强子/启动子元件例如 SV40 启动子、巨细胞病毒(CMV)启动子、劳斯肉瘤病毒(RSV)启动子和鼠白血病病毒(MLV)启动子在广泛的细胞类型中都具有活性,并且称为“遍在的”。或者,非组成型启动子例如小鼠乳腺肿瘤病毒(MMTV)启动子也可用于本发明。插入克隆位点的核酸序列可具有任何编码目标多肽的开放阅读框,前提是其中编码序列编码目标多肽,其应当缺乏可阻断适当的 mRNA 分子产生和/或产生异常剪接的或异常 mRNA 分子的隐蔽剪接位点。

[0108] 如本文中所用,术语“药物组合物”是指本文中描述的化合物与其它化学组分例如载体、稳定剂、稀释剂、分散剂、悬浮剂、增稠剂和/或赋形剂的混合物。

[0109] 术语“家禽来源的”或“禽类来源的”是指由家禽产生的或从家禽获得的组合物或物质。“家禽”是指可作为家畜饲养的禽类,包括但不限于鸡、鸭、火鸡、鹌鹑和平胸鸟。例如,“家禽来源的”可指鸡来源的、火鸡来源的和/或鹌鹑来源的。

[0110] “逆转录病毒颗粒”、“转导颗粒 (transducing particle 或 transduction particle)”是指能够将非病毒 DNA 或 RNA 转导入细胞的复制缺陷型或具有复制能力的病毒。在一个特别有用的实施方案中,可如 2009 年 4 月 28 日发布的美国专利 No. 7, 524, 626 中公开的制备用于产生根据本发明的转基因禽类的逆转录病毒颗粒,将所述专利的公开内容通过引用整体并入本文。

[0111] 术语“转化”、“转导”和“转染”全都指多核苷酸至禽类胚盘细胞中的引入。“膨大部 (Magnum)”为漏斗状器官 (infundibulum) 与峡之间 (含有合成和分泌卵的卵清蛋白的管腺细胞) 的输卵管的这部分。

[0112] 术语“转基因”是指插入根据本发明的禽类基因组的异源核苷酸序列。“转基因”可以特别地指外源编码序列,与外源启动子或其它调控序列连接的外源编码序列,两个逆转录病毒 LTR 之间和 / 或逆转录病毒 LTR 与 LTR 之间的核苷酸序列之间的所有核苷酸序列,其中 LTR 是逆转录病毒用于引入转基因的部分。

[0113] 术语“优化的”用于“优化的编码序列”的背景中,其中卵清蛋白、溶菌酶、卵类粘蛋白和卵转铁蛋白中发现的每个特定氨基酸的最常使用的密码子被用于设计被插入本发明的载体的优化的人干扰素- α 2b (IFN- α 2b) 多核苷酸序列。更具体地,优化的人 IFN- α 2b 的 DNA 序列基于母鸡输卵管优化的密码子选择,并且利用从鸡 (原鸡) 卵清蛋白、溶菌酶、卵类粘蛋白和卵转铁蛋白编译的密码子选择表使用 Wisconsin Package, 9.1 版 (Genetics Computer Group Inc., Madison, Wis.) 的 BACKTRANSLATE 程序来产生。例如,在 4 种卵清蛋白中氨基酸丙氨酸的 4 个密码子的百分比使用为 34% (对于 GCU)、31% (对于 GCC)、26% (对于 GCA) 和 8% (对于 GCG)。因此, GCU 用作优化的编码序列中针对大部分丙氨酸的密码子。包含优化的人蛋白质的基因的载体用于产生在它们的组织和卵中表达转基因家禽来源的蛋白质的转基因禽类。

[0114] 如本文中所用,术语“受试者”包括哺乳动物和非哺乳动物。哺乳动物的实例包括但不限于人、黑猩猩、猿猴、牛、马、绵羊、山羊、猪、兔、狗、猫、大鼠、小鼠、豚鼠等。

[0115] 如本文中所用,术语“治疗有效量”是指与未曾接受这样的量的相应受试者相比较,导致疾病、病症或副作用的改善的治疗、愈合、预防或缓解,或疾病或病症的进展速度的减小的化合物的任何量。术语还在其范围内包括有效地增强正常生理功能的量。

[0116] 术语“治疗”或其其它词性形式是指减轻、消除或缓解疾病或病症症状,预防另外的症状,缓解或预防症状的潜在原因,抑制疾病或病况,缓解疾病或病况,阻止疾病或病况的发展,缓解疾病或病况,引起疾病或病况的消退,缓解由疾病或病症引起的病况,或预防性和 / 或治疗性终止疾病或病况的症状。

[0117] LAL 组合物

[0118] 本发明总体上涉及包含对于疗法 (例如在溶酶体贮积病的治疗中) 有用的酶的组合物。在一个方面,本发明涉及溶酶体贮积病酶 (lysosomal storage disease enzymes), 例如具有使得分子易于被某些细胞类型内化的糖基化模式的 LAL。本发明中还包括重组人蛋白质,包括以分离或纯化形式存在的 LAL。溶酶体贮积病酶 (例如 LAL) 的分离可通过对于蛋白质纯化领域中的技术人员来说是显而易见的方法来实现。

[0119] 在一个实施方案中,本发明涉及溶酶体贮积病酶,包括但不限于具有本文中描述的 N- 连接糖基化模式的 LAL。

[0120] 在一个方面,本文中公开的组合物包含人 LAL,其中相当高百分比的人 LAL 包含甘露糖-6-磷酸聚糖部分,其可用作通过在例如肝细胞上发现的细胞表面上的甘露糖-6-磷酸受体内化的配体。在一个实施方案中,包含于组合物中的 LAL 中的 30%或更多,例如,至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 97%或至少 99%包含至少一个甘露糖-6-磷酸部分。甘露糖-6-磷酸部分可见于例如位于选自 SEQ ID NO:2 的 Asn¹⁵、Asn⁵¹、Asn⁸⁰、Asn¹⁴⁰、Asn²⁵²和 Asn³⁰⁰的一个或多个残基上的 N-聚糖结构上。包含甘露糖-6-磷酸部分的聚糖结构包括例如图 16 中显示的 G-n 和 H-n。

[0121] 根据本发明的重组人 LAL 包含多个 N-连接糖链(例如,约 5 或 6 个糖链)。5 或 6 个位点的每一个位点上的 N-连接糖基化结构可选自图 16 中显示的 A-n、B-n、C-n、D-n、E-n、F-n、G-n、H-n、I-n、J-n、K-n、L-n、M-n、N-n 和 O-n 之一。

[0122] 本文中还提供了 LAL 分子的混合物(例如,超过一种 LAL 分子可存在于混合物中,例如 SEQ ID NO:2、3、4 和 19 中所示的 LAL 分子),其中 LAL 分子中的一些或全部具有一种或多种选自结构 A-n、结构 B-n、结构 C-n、结构 D-n、结构 E-n、结构 F-n、结构 G-n、结构 H-n、结构 I-n、结构 J-n、结构 K-n、结构 L-n、结构 M-n、结构 N-n 和结构 O-n 的糖基化结构(图 16)。在一个实施方案中,纯化或分离溶酶体酸性脂酶分子的混合物,例如从卵分离或从转基因禽类中产生的卵清纯化或分离所述混合物。

[0123] 本发明还包括包含结构 A-n 的单种 LAL 分子。本发明还包括包含结构 B-n 的单种 LAL 分子。本发明还包括包含结构 C-n 的单种 LAL 分子。本发明还包括包含结构 D-n 的单种 LAL 分子。本发明还包括包含结构 E-n 的单种 LAL 分子。本发明还包括包含结构 F-n 的单种 LAL 分子。本发明还包括包含结构 G-n 的单种 LAL 分子。本发明还包括包含结构 H-n 的单种 LAL 分子。本发明还包括包含结构 I-n 的单种 LAL 分子。本发明还包括包含结构 J-n 的单种 LAL 分子。本发明还包括包含结构 K-n 的单种 LAL 分子。本发明还包括包含结构 L-n 的单种 LAL 分子。本发明还包括包含结构 M-n 的单种溶酶体酸性脂酶分子。本发明还包括包含结构 N-n 的单种 LAL 分子。本发明还包括包含结构 O-n 的单种 LAL 分子。

[0124] 连接于根据本发明的人 LAL 的 N-连接寡糖缺少末端唾液酸和半乳糖残基。即,仅少量 N-连接寡核糖结构被末端唾液酸化,并且几乎不存在半乳糖残基。此外,末端 N-乙酰葡萄糖胺(GlcNAc)广泛地存在于本文中描述的 LAL 的 N-连接寡糖结构中。这样,根据本发明产生的 LAL 可被靶向至细胞例如单核巨噬细胞和肝巨噬细胞。

[0125] 本发明的一个方面提供了基本上无唾液酸的 LAL 的组合物。在另一个方面,本文中公开的组合物包含重组人 LAL,其中相当大百分比的人 LAL 在任何其 N-聚糖结构中不包含唾液酸部分,其可干扰酶至细胞内的内化。在一个实施方案中,15%或更少,例如,10%或更少,5%或更少,2%或更少,1%或更少或基本上无包含于组合物中的 LAL 在任何其 N-聚糖结构中不包含唾液酸部分。

[0126] 在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N-连接寡糖中的约 95%或更多不包含唾液酸。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N-连接寡糖中的约 90%或更多不包含唾液酸。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N-连接寡糖中的约 80%或更多不包含唾液酸。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N-连接寡糖中多于约 70%或更多不包含唾液酸。

[0127] 在另一个实施方案中,基本上无存在于本发明的 LAL 分子上的 N-连接寡糖结构类

型包含唾液酸。在另一个实施方案中,发现与本发明的 LAL 分子连接的 N- 连接寡糖结构类型中的约 90% 或更多不包含唾液酸。例如,如果存在 20 个寡糖结构类型,则 18 个或更多个结构类型不包含唾液酸。在另一个实施方案中,发现与本发明的 LAL 分子连接的 N- 连接寡糖结构类型中的约 80% 或更多不包含唾液酸。在另一个实施方案中,发现与本发明的 LAL 分子连接的 N- 连接寡糖结构类型中的约 70% 或更多不包含唾液酸。在另一个实施方案中,发现与本发明的 LAL 分子连接的 N- 连接寡糖结构类型中的约 60% 或更多不包含唾液酸。在另一个实施方案中,发现与本发明的 LAL 分子连接的 N- 连接寡糖结构类型中的约 50% 或更多不包含唾液酸。

[0128] 根据本发明的一个方面,本文中描述的 LAL 包含高水平的末端 N- 乙酰葡萄糖胺。在一个方面,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 95% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 90% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 80% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 70% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 60% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 50% 或更多包含末端 N- 乙酰葡萄糖胺。

[0129] 在一个实施方案中,所有存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型中的约 90% 或更多包含末端 N- 乙酰葡萄糖胺。例如,如果存在 20 个寡糖结构类型,那么 18 个或更多个结构类型不包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型中的约 80% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型中的约 70% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型中的约 60% 或更多包含末端 N- 乙酰葡萄糖胺。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖结构类型中的约 50% 或更多包含末端 N- 乙酰葡萄糖胺。

[0130] 在本发明的另一个方面,本文中公开的组合物包含人 LAL,其中相当高百分比的人 LAL 在任何其 N- 聚糖结构中不包含岩藻糖部分。在一个实施方案中,50% 或更少,例如,50% 或更少,40% 或更少,30% 或更少,20% 或更少,10% 或更少,5% 或更少,2% 或更少,1% 或更少或基本上无包含于组合物中的 LAL 在任何其 N- 聚糖结构中包含岩藻糖部分。

[0131] 在一个实施方案中,岩藻糖基本上不存在于根据本发明产生的 LAL 的 N- 连接寡糖结构上。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 95% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 90% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 85% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 80% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 70% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的单种 LAL 分子上的 N- 连接寡糖中的约 60% 或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N- 连接寡糖中的约 50% 或更多不

包含岩藻糖。

[0132] 在一个实施方案中,基本上无存在于本发明的 LAL 分子上的 N-连接寡糖结构包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N-连接寡糖结构类型中的约 95%或更多不包含岩藻糖。例如,如果存在 20 个寡糖结构类型,那么 19 个或更多个结构类型不包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N-连接寡糖结构类型中的约 90%或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N-连接寡糖结构类型中的约 85%或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N-连接寡糖结构类型中的约 80%或更多不包含岩藻糖。在另一个实施方案中,存在于本发明的 LAL 分子上的 N-连接寡糖结构类型中的约 70%或更多不包含岩藻糖。

[0133] 如上文中所讨论的,某些单糖大量地存在于根据本发明产生的 LAL 分子上。所分析的总的单糖种类包括岩藻糖、N-乙酰半乳糖胺、N-乙酰葡萄糖胺、半乳糖、葡萄糖、甘露糖、甘露糖-6-磷酸、N-乙酰神经氨酸和 N-羟乙酰神经氨酸。岩藻糖可占总单糖组成的约 0%至约 1%。N-乙酰半乳糖胺可占总单糖组成的约 0%至约 1%。N-乙酰葡萄糖胺可占总单糖组成的约 35%至约 50%。半乳糖可占总单糖组成的约 1-10%。葡萄糖占总单糖组成的 0%。甘露糖占总单糖组成的约 32%至约 50%。甘露糖-6-磷酸占总单糖组成的约 1%至约 11%。

[0134] 在一个实施方案中,根据本发明产生的 LAL 不包含任何木糖。此外,因为根据本发明产生的 LAL 中基本上不存在 N-乙酰半乳糖胺 (GalNac),因此本发明的一个方面包括不具有 O-连接糖基化的 LAL 的组合物。

[0135] LAL 在其氨基酸序列中具有 6 个用于 N-连接糖基化的潜在位点,例如,SEQ ID NO : 1 中的 Asn³⁶、Asn⁷²、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³和 Asn³²¹。这些位点中的 5 个位点 Asn³⁶、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³和 Asn³²¹被糖基化,然而 Asn⁷²可以未被糖基化或大体上未被糖基化(大体上未被糖基化意指在 LAL 分子的混合物中,比 Asn³⁶、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³和 Asn³²¹的任一个更少的 Asn⁷²被糖基化)(参见图 17)。因此,本发明的一个方面是在 Asn⁷²上未被糖基化和/或大体上未被糖基化的 LAL 的组合物。具有糖基化的 Asn⁷²的 LAL 在本发明的范围内。本文中描述的 Asn 的位置基于 SEQ ID NO : 1 中所示的 LAL 氨基酸序列。对于本领域技术人员来说很明显,Asn 的编号(即,天冬酰胺的位置)可根据单种 LAL 分子而变化,并且在其它 LAL 分子例如其氨基酸序列示于 SEQ ID NO : 2、3、4 和 19 中的那些分子中可容易地被确定。

[0136] 根据本发明产生的 LAL 分子包含含有以 N-乙酰葡萄糖胺、甘露醇和甘露糖-6-磷酸(M6P)作为主要糖的二-、三-和四天线结构的混合物的 N-聚糖结构(图 16 和 17)。根据本发明的一个方面,M6P-修饰的 N-聚糖至少存在于 Asn¹⁰¹、Asn¹⁶¹和 Asn²⁷³上。因此,本发明的一个实施方案包括具有存在于 Asn¹⁰¹、Asn¹⁶¹或 Asn²⁷³的任一个上的 M6P-修饰的 N-聚糖的 LAL 的组合物。在另一个实施方案中,本发明包括具有存在于 Asn²⁷³上的 M6P-修饰的 N-聚糖的 LAL 的组合物。在另一个实施方案中,本发明包括具有存在于 Asn¹⁰¹、Asn¹⁶¹或 Asn²⁷³的任一个上的单磷酸化的 N-聚糖(M6P)的 LAL 的组合物。在另一个实施方案中,本发明包括具有存在于 Asn¹⁶¹和 Asn²⁷³上的单磷酸化的 N-聚糖的 LAL 的组合物。在另一个实施方案中,本发明包括具有存在于 Asn¹⁰¹和 Asn²⁷³上的单磷酸化的 N-聚糖的 LAL 的组合物。在一个具体的实施方案中,根据本发明产生的 LAL 可在 Asn¹⁰¹上包含双磷酸化的甘露

糖（双-M6P）。

[0137] 根据本发明产生的 LAL 分子包含降低的水平半乳糖（例如，“Gal”）。本发明的一个方面包括在 Asn³⁶、Asn¹⁶¹或 Asn³²¹的任一个上具有末端半乳糖的 LAL 的组合物。在另一个实施方案中，本发明包括在 Asn³⁶和 Asn¹⁶¹上具有末端半乳糖的 LAL 的组合物。在另一个实施方案中，本发明包括在 Asn¹⁶¹和 Asn³²¹上具有末端半乳糖的 LAL 的组合物。在另一个实施方案中，本发明包括在 Asn³⁶和 Asn³²¹上具有末端半乳糖的 LAL 的组合物。在另一个实施方案中，本发明包括在 Asn³⁶、Asn¹⁶¹和 Asn³²¹上具有末端半乳糖的 LAL 的组合物。在另一个实施方案中，本发明包括不具有末端半乳糖的 LAL 的组合物。

[0138] 在 LAL 中在不同 N-连接糖基化位点上发现多种类型的 N-聚糖。N-聚糖结构包括以 N-乙酰葡萄糖胺、甘露糖和甘露糖-6-磷酸（M6P）作为主要糖的二-、三-和四天线结构的混合物。具体地，在本发明的一个实施方案中，LAL 在第一 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn³⁶）上包含选自 GlcNAc4Man3GlcNAc2 或 Gal1GlcNAc4Man3GlcNAc2 的 N-聚糖结构。在另一个实施方案中，LAL 在第二 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn⁷²）上不包含糖基化或大体上未被糖基化。在另一个实施方案中，LAL 在其第三 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn¹⁰¹）上包含 Phos2Man7GlcNAc2。在另一个实施方案中，LAL 在其第四 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn¹⁶¹）上包含选自 Phos1Man6GlcNAc2、GlcNAc1Phos1Man6GlcNAc2、Man3GlcNAc2、GlcNAc2Man3GlcNAc2、GlcNAc3Man3GlcNAc2、GlcNAc4Man3GlcNAc2 或 Gal1GlcNAc4Man3GlcNAc2 的 N-聚糖结构。在另一个实施方案中，LAL 在其第五 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn²⁷³）上包含选自 Man7GlcNAc2、Man8GlcNAc2、Man9GlcNAc2、Phos1Man8GlcNAc2 或 Phos1Man9GlcNAc2 的 N-聚糖结构。在另一个实施方案中，LAL 在其第六 N-连接糖基化位点（例如，如 SEQ ID NO:1 中的 Asn³²¹）上包含选自 GlcNAc2Man3GlcNAc2、GlcNAc3Man3GlcNAc2、GlcNAc4Man3GlcNAc2、Gal1GlcNAc4Man3GlcNAc2、GlcNAc5Man3GlcNAc2、Gal1GlcNAc5Man3GlcNAc2、GlcNAc6Man3GlcNAc2 或 Gal1GlcNAc6Man3GlcNAc2 的 N-聚糖结构。

[0139] 根据本发明的某些方面，LAL 的组合物包括在 SEQ ID NO:1 的 Asn³⁶、Asn⁷²、Asn¹⁰¹、Asn¹⁶¹或 Asn²⁷³和 Asn³²¹（或 SEQ ID NO:2、3、4 和 19 内的相应天冬酰胺残基）上糖基化的 LAL，指定的 Asn 位置上的一个 N-聚糖如下文中显示的：

[0140] a) 在 Asn³⁶上，GlcNAc4Man3GlcNAc2，或

[0141] Gal1GlcNAc4Man3GlcNAc2；

[0142] b) 在 Asn⁷²上，无糖基化；

[0143] c) 在 Asn¹⁰¹上，Phos2Man7GlcNAc2；

[0144] d) 在 Asn¹⁶¹上，Phos1Man6GlcNAc2，

[0145] GlcNAc1Phos1Man6GlcNAc2；

[0146] Man3GlcNAc2；

[0147] GlcNAc2Man3GlcNAc2；

[0148] GlcNAc3Man3GlcNAc2；

[0149] GlcNAc4Man3GlcNAc2，或

[0150] Gal1GlcNAc4Man3GlcNAc2；

- [0151] e) 在 Asn²⁷³ 上, Man7GlcNAc2,
[0152] Man8GlcNAc2,
[0153] Man9GlcNAc2,
[0154] Phos1Man8GlcNAc2, 或
[0155] Phos1Man9GlcNAc2 ; 和
[0156] f) 在 Asn³²¹ 上, GlcNAc2Man3GlcNAc2,
[0157] GlcNAc3Man3GlcNAc2,
[0158] GlcNAc4Man3GlcNAc2,
[0159] Gal1GlcNAc4Man3GlcNAc2,
[0160] GlcNAc5Man3GlcNAc2,
[0161] Gal1GlcNAc5Man3GlcNAc2,
[0162] GlcNAc6Man3GlcNAc2, 或
[0163] Gal1GlcNAc6Man3GlcNAc2,
[0164] 其中 Man = 甘露糖,
[0165] GlcNAc = N- 乙酰葡萄糖胺,
[0166] Phos = 磷酸, 和
[0167] Gal = 半乳糖。

[0168] 在一个实施方案中, 可发现 Gal1GlcNAc4Man3GlcNAc2 为根据本发明产生的 LAL 的 Asn³⁶、Asn¹⁶¹ 或 Asn³²¹ 的任一个上的聚糖成分。在一个具体的实施方案中, 可发现 Gal1GlcNAc4Man3GlcNAc2 为 Asn³⁶、Asn¹⁶¹ 和 Asn³²¹ 上的聚糖成分。

[0169] 在本发明的 LAL 中, Asn¹⁰¹ 和 Asn²⁷³ 展示具有约 6 至约 10 个甘露糖分子 (如本文中描述的 MAN6-MAN10) 作为主要组分的的高甘露糖类型。因此, 本发明的一个方面包括在 Asn¹⁰¹ 或 Asn²⁷³ 上具有高甘露糖结构的 LAL 的组合物。在另一个实施方案中, 本发明的 LAL 的组合物可包含在 Asn¹⁰¹ 或 Asn²⁷³ 上具有至少 6 个甘露糖的 N- 聚糖结构。在另一个实施方案中, LAL 的组合物包含在 Asn¹⁰¹ 或 Asn²⁷³ 上具有 7、8 或 9 个甘露糖的 N- 聚糖结构。在另一个实施方案中, 本发明包括在 Asn¹⁰¹ 和 Asn²⁷³ 上具有 7、8 或 9 个甘露糖的 LAL 的组合物。在另一个实施方案中, 本发明包括在 Asn¹⁰¹ 和 / 或 Asn²⁷³ 上具有 7、8 或 9 个甘露糖的 LAL 的组合物并且至少一个甘露糖被磷酸化。

[0170] 应当理解, 糖基化位点和与上述 Asn 相关的编号基于 SEQ ID NO : 1 中所示的 LAL 的氨基酸序列, 并且上文中所述的 SEQ ID NO : 1 的上下文中的糖基化谱也用于 SEQ ID NO : 2、3、4 和 19 中所示的 LAL 分子, 虽然相应 Asn 的编号可根据 LAL 分子而变化。例如, SEQ ID NO : 1 中的 Asn³⁶ 对应于 SEQ ID NO : 2 中的 Asn¹⁵、SEQ ID NO : 3 中的 Asn¹³、SEQ ID NO : 4 中的 Asn¹⁰ 和 SEQ ID NO : 19 中的 Asn⁹。SEQ ID NO : 1 中的 Asn⁷² 对应于 SEQ ID NO : 2 中的 Asn⁵¹、SEQ ID NO : 3 中的 Asn⁴⁹、SEQ ID NO : 4 中的 Asn⁴⁶ 和 SEQ ID NO : 19 中的 Asn⁴⁵。SEQ ID NO : 1 中的 Asn¹⁰¹ 对应于 SEQ ID NO : 2 中的 Asn⁸⁰、SEQ ID NO : 3 中的 Asn⁷⁸、SEQ ID NO : 4 中的 Asn⁷⁵ 和 SEQ ID NO : 19 中的 Asn⁷⁴。SEQ ID NO : 1 中的 Asn¹⁶¹ 对应于 SEQ ID NO : 2 中的 Asn¹⁴⁰、SEQ ID NO : 3 中的 Asn¹³⁸、SEQ ID NO : 4 中的 Asn¹³⁵ 和 SEQ ID NO : 19 中的 Asn¹³⁴。SEQ ID NO : 1 中的 Asn²⁷³ 对应于 SEQ ID NO : 2 中的 Asn²⁵²、SEQ ID NO : 3 中的 Asn²⁵⁰、SEQ ID NO : 4 中的 Asn²⁴⁷ 和 SEQ ID NO : 19 中的 Asn²⁴⁶。SEQ ID NO : 1 中的 Asn³²¹ 对应于 SEQ ID

NO :2 中的 Asn³⁰⁰、SEQ ID NO :3 中的 Asn²⁹⁸、SEQ ID NO :4 中的 Asn²⁹⁵和 SEQ ID NO :19 中的 Asn²⁹⁴。

[0171] 例如,在一个实施方案中,LAL 至少在一个选自 SEQ ID NO :2 的 Asn¹⁵、Asn⁵¹、Asn⁸⁰、Asn¹⁴⁰、Asn²⁵²和 Asn³⁰⁰的位置上被 N-连接糖基化。在另一个实施方案中,LAL 在 SEQ ID NO :2 的 Asn¹⁵、Asn⁸⁰、Asn¹⁴⁰、Asn²⁵²和 Asn³⁰⁰上被 N-连接糖基化。在另一个实施方案中,SEQ ID NO :2 的 LAL 的 N-聚糖结构不具有木糖,然而少于 15%、10%、5%或 1%的 N-聚糖结构包含唾液酸;少于 50%、40%、30%、20%、10%、5%或 1%的 N-聚糖结构包含岩藻糖;以及至少 30%、50%、60%、70%、80%、90%和 95%的 N-聚糖结构包含磷酸化的甘露糖 (M6P)。

[0172] 在一个实施方案中,LAL 至少在选自 SEQ ID NO :3 的 Asn¹³、Asn⁴⁹、Asn⁷⁸、Asn¹³⁸、Asn²⁵⁰和 Asn²⁹⁸的一个位置上被 N-连接糖基化。在另一个实施方案中,LAL 在 SEQ ID NO :3 的 Asn¹³、Asn⁷⁸、Asn¹³⁸、Asn²⁵⁰和 Asn²⁹⁸上被 N-连接糖基化。在另一个实施方案中,SEQ ID NO :3 的 LAL 的 N-聚糖结构不具有木糖,然而少于 15%、10%、5%或 1%的 N-聚糖结构包含唾液酸;少于 50%、40%、30%、20%、10%、5%或 1%的 N-聚糖结构包含岩藻糖;以及至少 30%、50%、60%、70%、80%、90%和 95%的 N-聚糖结构包含磷酸化的甘露糖 (M6P)。

[0173] 在一个实施方案中,LAL 至少在选自 SEQ ID NO :4 的 Asn¹⁰、Asn⁴⁶、Asn⁷⁵、Asn¹³⁵、Asn²⁴⁷和 Asn²⁹⁵的一个位置上被 N-连接糖基化。在另一个实施方案中,LAL 在 SEQ ID NO :4 的 Asn¹⁰、Asn⁷⁵、Asn¹³⁵、Asn²⁴⁷和 Asn²⁹⁵上被 N-连接糖基化。在另一个实施方案中,SEQ ID NO :4 的 LAL 的 N-聚糖结构不具有木糖,然而少于 15%、10%、5%或 1%的 N-聚糖结构包含唾液酸;少于 50%、40%、30%、20%、10%、5%或 1%的 N-聚糖结构包含岩藻糖;以及至少 30%、50%、60%、70%、80%、90%和 95%的 N-聚糖结构包含磷酸化的甘露糖 (M6P)。

[0174] 在一个实施方案中,LAL 至少在选自 SEQ ID NO :19 的 Asn⁹、Asn⁴⁵、Asn⁷⁴、Asn¹³⁴、Asn²⁴⁶和 Asn²⁹⁴的一个位置上被 N-连接糖基化。在另一个实施方案中,LAL 在 SEQ ID NO :19 的 Asn⁹、Asn⁷⁴、Asn¹³⁴、Asn²⁴⁶和 Asn²⁹⁴上被 N-连接糖基化。在另一个实施方案中,SEQ ID NO :4 的 LAL 的 N-聚糖结构不具有木糖,然而少于 15%、10%、5%或 1%的 N-聚糖结构包含唾液酸;少于 50%、40%、30%、20%、10%、5%或 1%的 N-聚糖结构包含岩藻糖;以及至少 30%、50%、60%、70%、80%、90%和 95%的 N-聚糖结构包含磷酸化的甘露糖 (M6P)。

[0175] 根据本发明的组合物可以以许多方式产生,包括通过使用转基因禽类、转基因鱼、转基因哺乳动物,例如转基因山羊或在转基因植物中,例如烟草和浮萍 (浮萍 (*Lemna minor*)) 和某些种类的细胞培养物。

[0176] 本发明还涵盖包含 PEG 化的 LAL 的组合物。可如例如 2007 年 4 月 26 日公布的美国专利公布 No. 20070092486 中所公开的,对本文中描述的 LAL 酶进行 PEG 化,将所述专利公布的公开内容通过引用整体并入本文。

[0177] 在一个实施方案中,通过例如来自禽类输卵管细胞例如管腺细胞的表达专门化表达系统获得衍生的糖基化模式。例如,已显示本文中公开的糖基化模式存在于根据本发明在禽类例如鸡的输卵管细胞中产生的溶酶体贮积病酶上。

[0178] 根据本发明产生的蛋白质可通过任何有用的方法例如对于蛋白质纯化领域的技术人员来说是显而易见的那些方法来从卵清纯化。例如,可通过对于蛋白质纯化领域的技术人员来说是显而易见的方法从卵清纯化根据本发明在转基因禽类中产生的人 LAL (hLAL)。用于存在于卵清中的 LAL 的纯化方案的实例描述于实施例中。

[0179] 本发明包括卵和卵清以及产卵和产生包含本发明的溶酶体酸性脂酶分子的卵清的禽类（例如鸡、火鸡和鹌鹑），所述溶酶体酸性脂酶包含一种或多种本文中公开的糖基化结构。

[0180] LAL 在禽类中的表达

[0181] 本文中公开了载体和用于将外源核酸序列稳定地引入禽类的基因组以表达期望的蛋白质的方法，所述蛋白质是例如从甘露糖-6-磷酸的添加获益（例如，获得增强的功效）的那些蛋白质，例如溶酶体酶，包括但不限于溶酶体酸性脂酶（LAL）和其它蛋白质例如本文中特别地公开的那些蛋白质。具体地，产生转基因禽类，所述转基因禽类在它们的输卵管中表达外源序列并且将外源蛋白质例如药物蛋白储存在它们的卵中。包含此类外源蛋白质的禽类的卵也在本文中进行了描述。本文中还公开了在转基因禽类的输卵管中高效表达并且储存在禽类卵中的 LAL 的新形式。

[0182] 本发明的一个方面涉及包含 LAL 即根据本发明产生的 LAL 分子的组合物。在特别有用的实施方案中，纯化或分离 LAL。例如，已从由转基因禽类产的硬壳卵的内容物取出 LAL。在一个特别有用的实施方案中，LAL 为人 LAL。在一个实施方案中，本发明的 LAL 具有因禽类的输卵管细胞中产生的 LAL 而导致的糖基化模式。例如，组合物可包含根据本发明在禽类例如鸡中产生的和从卵清分离的 LAL 分子的混合物。在一个有用的实施方案中，含 LAL 组合物为药物制剂。

[0183] 在一个方面，本发明涉及包含分离的 LAL 分子例如人 LAL 分子的组合物，其中 LAL 在包含编码 LAL 的转基因的禽类中产生。在一个实施方案中，LAL 在转基因禽类（例如，转基因鸡）的输卵管细胞（例如，管腺细胞）中产生并且 LAL 分离自转基因禽类的卵清。在一个实施方案中，LAL 在禽类例如鸡的输卵管细胞（例如，管腺细胞）中被糖基化。

[0184] 在另一个方面，提供了用于在禽类的特定组织中产生外源蛋白质例如溶酶体贮积病酶例如 LAL 的方法。可在禽类的输卵管、血液和 / 或其它细胞和组织中表达此类外源蛋白质。在一个实施方案中，将转基因引入胚部胚盘细胞（例如接近 X 期）以产生转基因禽类，以便目标蛋白在输卵管的膨大部的管腺细胞中表达，分泌至管腔中，然后储存在硬壳卵的卵清中。所产生的转基因禽类可在其种系中具有转基因，从而将外源转基因以孟德尔方式稳定地传递至禽类的后代。

[0185] 本发明包括在禽类输卵管中产生外源蛋白质例如 LAL 的方法。所述方法可包括提供包含编码序列和可操作地连接于编码序列的启动子的载体（以便启动子可在禽类输卵管中实现核酸的表达）的第一步骤。可产生转基因细胞和 / 或组织，其中将载体引入新分离的、培养中的或胚胎中的禽类胚部胚盘细胞，以便将载体序列插入禽类基因组。在其输卵管中表达外源蛋白质例如 LAL 的成熟转基因禽类可来源于转基因细胞和 / 或组织。

[0186] 在本发明的一个方面，转基因禽类的产生可通过利用在逆转录病毒载体的 5' 与 3' LTR 之间具有转基因的复制缺陷型或具有复制能力的逆转录病毒颗粒转导胚部胚盘细胞来实现。例如，可使用禽白血病病毒（ALV）逆转录病毒载体或鼠白血病病毒（MLV）逆转录病毒载体，所述载体包含含有被插入启动子区域的区段的下游的外源基因的经修饰的 pNLB 质粒。可将包装入病毒颗粒的经修饰的逆转录病毒载体的 RNA 拷贝用于感染发育成转基因禽类的胚部胚盘（embryonic blastoderms）。

[0187] 本发明的另一个方面提供了载体，所述载体以使编码序列在禽类输卵管中表达的

操作和位置关系包括编码序列和启动子。此类载体包括但不限于禽白血病病毒 (ALV) 逆转录病毒载体、鼠白血病病毒 (MLV) 逆转录病毒载体和慢病毒载体。此外,载体可以为包括禽白血病病毒 (ALV) 逆转录病毒载体、鼠白血病病毒 (MLV) 逆转录病毒载体或慢病毒载体的 LTR 的核酸序列。启动子对于实现编码序列在禽类输卵管中的表达是足够的。编码序列编码储存至硬壳卵的卵清中的外源蛋白质。这样,编码序列编码外源蛋白质例如转基因禽类来源的蛋白质例如转基因禽类来源的溶酶体酸性脂酶 (TPD LAL)。

[0188] 在一个实施方案中,本发明的方法中使用的载体包含特别适用于在禽类和它们的卵中表达外源蛋白质的启动子。这样,外源编码序列的表达可在转基因禽类的输卵管和血液中以及其禽类卵的卵清中发生。启动子包括但不限于巨细胞病毒 (CMV) 启动子、劳斯肉瘤病毒 (RSV) 启动子、 β -肌动蛋白启动子 (例如,鸡 β -肌动蛋白启动子)、鼠白血病病毒 (MLV) 启动子、小鼠乳腺肿瘤病毒 (MMTV) 启动子、卵清蛋白启动子、溶菌酶启动子、伴清蛋白启动子、卵类粘蛋白启动子、卵粘蛋白启动子和卵转铁蛋白启动子。任选地,启动子可以为至少一个启动子区域的区段,例如卵清蛋白、溶菌酶、伴清蛋白、卵类粘蛋白、卵粘蛋白和卵转铁蛋白启动子区域的区段。在一个实施方案中,启动子为一种或多种启动子的组合或融合物或一种或多种启动子例如卵清蛋白、溶菌酶、伴清蛋白、卵类粘蛋白、卵粘蛋白和卵转铁蛋白启动子的部分的融合物。

[0189] 在一个实施方案中,载体包括可操作地连接于编码序列的信号肽,以便在细胞中翻译后,信号肽指导由载体表达的外源蛋白质例如人 LAL 分泌至硬壳卵的卵清中。

[0190] 本发明的一个方面提供了如本文中所公开的产生的外源蛋白质的编码序列,其中编码序列经密码子优化以用于在禽类中例如在鸡中表达。可根据至少一种,优选超过一种在禽类细胞 (例如,鸡细胞) 中表达的蛋白质的密码子选择来确定密码子优化。例如,可根据编码鸡的蛋白质卵清蛋白、溶菌酶、卵粘蛋白和卵转铁蛋白的核酸序列来确定密码子选择。例如,编码外源蛋白质的序列的 DNA 可使用 Wisconsin Package, 9.1 版 (Genetics Computer Group, Inc., Madison, WI) 的 BACKTRANSLATE[®] 程序,利用从鸡 (原鸡) 卵清蛋白、溶菌酶、卵粘蛋白和卵转铁蛋白编译的密码子选择表来进行密码子优化。

[0191] 本发明的一个重要方面涉及禽类硬壳卵 (例如,鸡硬壳卵),其包含外源肽或蛋白质,包括但不限于人 LAL。外源肽或蛋白质例如人 LAL 可由转基因禽类的转基因编码。通常,外源肽或蛋白质 (例如, LAL) 被糖基化。蛋白质可以以任何有用的量的存在。在一个实施方案中,蛋白质以在约 0.01 μ g/ 硬壳卵至约 1 克 / 硬壳卵的范围内的量存在。在另一个实施方案中,蛋白质以在约 1 μ g/ 硬壳卵至约 1 克 / 硬壳卵的范围内的量存在。例如,蛋白质可以以在约 10 μ g/ 硬壳卵至约 1 克 / 硬壳卵的范围内 (例如,约 10 μ g/ 硬壳卵至约 400 毫克 / 硬壳卵的范围) 的量存在。

[0192] 在一个实施方案中,本发明的外源蛋白质存在于卵的卵清中。在一个实施方案中,蛋白质以在约 1ng/ 毫升的卵清至约 0.2 克 / 毫升的卵清的范围内的量存在。例如,蛋白质可以以在约 0.1 μ g/ 毫升的卵清至约 0.2 克 / 毫升的卵清的范围内的量存在 (例如,蛋白质可以以在约 1 μ g/ 毫升的卵清至约 100 毫克 / 毫升的卵清的范围内的量存在)。在一个实施方案中,蛋白质可以以在约 1 μ g/ 毫升的卵清至约 50 毫克 / 毫升的卵清的范围内的量存在。例如,蛋白质可以以在约 1 μ g/ 毫升的卵清至约 10 毫克 / 毫升的卵清的范围内的量存在 (例如,蛋白质可以以在约 1 μ g/ 毫升的卵清至约 1 毫克 / 毫升的卵清的范围内的量存在)。

存在)。在一个实施方案中,蛋白质以超过 $0.1 \mu\text{g}/\text{毫升}$ 的卵清的量存在。在一个实施方案中,蛋白质以超过 $0.5 \mu\text{g}/\text{毫升}$ 的卵清的量存在。在一个实施方案中,蛋白质以超过 $1 \mu\text{g}/\text{毫升}$ 的卵清的量存在。在一个实施方案中,蛋白质以超过 $1.5 \mu\text{g}/\text{毫升}$ 的卵清的量存在。

[0193] 从已向其中引入了载体的胚盘细胞发育的产生本文中公开的外源蛋白质(例如, LAL)的本发明的禽类为 G0 代并且可被称为“建立者”。建立者禽类通常是每个插入的转基因的嵌合体。即, G0 转基因禽类仅有一些细胞包含转基因。G0 代通常也是转基因的半合子。可将 G0 代与非转基因动物交配以产生对于转基因也是半合子的并且在基本上所有禽类细胞中包含转基因的 G1 转基因后代。可将 G1 半合子后代与非转基因动物交配,从而产生 G2 半合子后代或可将它们自己相互交配,从而产生对于转基因是纯合的 G2 后代。从 G1 后代产生的对于转基因为阳性的禽类的大体上所有细胞包含转基因。在一个实施方案中,可将来自相同品系的半合子 G2 后代交配以产生对于转基因是纯合的 G3 后代。在一个实施方案中,将半合子 G0 或 G1 动物自己相互交配以产生在动物的每个细胞中包含两个拷贝的转基因的纯合 G1 后代。这些仅仅为某些有用的育种方法的实例,并且本发明包括使用任何有用的育种方法例如,对于本领域技术人员来说是已知的那些方法。

[0194] 在一个实施方案中,本发明提供了待分离的 LAL。即,包含于组合物中的 LAL 可以是分离的 LAL。例如, LAL 可分离自卵清。分离的 LAL 可以是在 LAL 分子间具有多种糖基化结构的 LAL 分子。

[0195] 通过本发明的方法,可将转基因引入禽类胚部胚盘细胞以产生转基因鸡、转基因火鸡、转基因鹌鹑和其它禽类物种,所述转基因动物在其种系组织的遗传物质中具有转基因以产生本发明的蛋白质。胚盘细胞通常是 VII-XII 期细胞,或其等同物,在一个实施方案中,为接近 X 期。

[0196] 本文中描述了一些用于进行本发明的方法的载体。在一个实施方案中,编码序列和载体的启动子在被引入胚盘细胞之前都位于 $5'$ 与 $3'$ LTR 之间。在一个实施方案中,载体是逆转病毒并且编码序列和启动子都位于逆转录病毒载体的 $5'$ 与 $3'$ LTR 之间。在一个有用的实施方案中, LTR 或逆转录病毒载体来源于禽白血病病毒 (ALV)、鼠白血病病毒 (MLV) 或慢病毒。

[0197] 在一个实施方案中,用于转染胚盘细胞和产生至禽类基因组的稳定整合的载体以使编码序列在禽类输卵管的膨大部的管腺细胞中表达的操作和位置关系包含编码序列和启动子,其中外源蛋白质例如溶酶体酶(例如, LAL)被储存在硬壳卵的卵清中。

[0198] 启动子可任选地为卵清蛋白启动子区域中的区段,该区段足够大以指导编码序列在管腺细胞中表达。截短卵清蛋白启动子和/或压缩(condensing)卵清蛋白启动子的关键调控元件(以便其保留在输卵管的膨大部的管腺细胞中表达所需的序列,同时又足够小以使其可容易地整合入载体)包括在本发明的范围内。在一个实施方案中,可使用卵清蛋白启动子区域的区段。该区段包含卵清蛋白基因的 $5'$ - 侧翼区域。

[0199] 启动子还可为很大程度上但非完全地对于膨大部是特异性的启动子,例如溶菌酶启动子。启动子还可以是小鼠乳腺肿瘤病毒 (MMTV) 启动子。或者,启动子可以是组成型启动子(例如,巨细胞病毒 (CMV) 启动子、劳斯肉瘤病毒 (RSV) 启动子、鼠白血病病毒 (MLV) 启动子等)。在一个实施方案中,启动子为巨细胞病毒 (CMV) 启动子、MDOT 启动子、劳斯肉瘤病毒 (RSV) 启动子、鼠白血病病毒 (MLV) 启动子、小鼠乳腺肿瘤病毒 (MMTV) 启动子、卵清

蛋白启动子、溶菌酶启动子、伴清蛋白启动子、卵类粘蛋白启动子、卵粘蛋白启动子和 / 或卵转铁蛋白启动子。任选地,启动子可以为启动子区域的至少一个区段,例如卵清蛋白、溶菌酶、伴清蛋白、卵类粘蛋白、卵粘蛋白和卵转铁蛋白启动子区域的区段。

[0200] 在转染胚盘细胞的一种方法中,将包装的基于逆转录病毒的载体用于将载体递送入胚部胚盘细胞,以便载体被整合入禽类基因组。

[0201] 用于将转基因随机引入禽类基因组的有用逆转录病毒为复制缺陷型禽类白血病病毒 (ALV)、复制缺陷型鼠白血病病毒 (MLV) 或慢病毒。在一个实施方案中,pNLB 载体通过将卵清蛋白启动子的区域和一种或多种外源基因插入逆转录病毒基因组的 5' 与 3' 长末端重复 (LTR) 之间来进行修饰。本发明涉及:置于在管腺细胞中具有活性的启动子的下游的任何编码序列可在管腺细胞中表达。例如,卵清蛋白启动子可在输卵管膨大部的管腺细胞中表达,因为卵清蛋白启动子驱动卵清蛋白的表达并且在输卵管管腺细胞中具有活性。

[0202] 本文中描述的任何载体还可任选地包括编码指导由载体的编码序列表达的蛋白质从输卵管的管腺细胞分泌的信号肽的编码序列。该方面有效地扩展了可使用本文中描述的方法将其储存在禽类卵中的外源蛋白质的范围。在外源蛋白质本来不被分泌的情况下,修饰包含编码序列的载体以包含下述 DNA 序列,所述 DNA 序列包含来自溶菌酶基因的约 60bp 的编码信号肽。将编码信号肽的 DNA 序列插入载体,以便其位于由 DNA 编码的蛋白质的 N 端。

[0203] 本发明的另一个方面包括在本发明的任何载体中使用内部核糖体进入位点 (IRES) 元件以使得能够从双顺反子或多顺反子 mRNA 翻译两种或更多种蛋白质。将 IRES 单元融合至一个或多个另外的编码序列的 5' 末端,随后将其在原始编码序列的末端插入载体,以便可编码序列可通过 IRES 彼此分隔。

[0204] 在一个实施方案中,当使用 IRES 时,有助于产物的翻译后修饰,因为一个编码序列可编码能够修饰其它编码序列产物的酶。例如,第一编码序列可编码可被羟基化并且可被由第二编码序列编码的酶激活的胶原蛋白,其中如在本领域中所理解的那样使用 IRES。

[0205] 在另一个方面,提供了具有 3' 非翻译区 (3' UTR) (以向产生的 RNA 赋予稳定性) 的用于本发明的任一方法的载体的编码序列。当将 3' UTR 添加至逆转录病毒载体时,启动子、基因 X 和 3' UTR 的取向在构建体中是反向的,以便 3' UTR 的添加不干扰全长基因组 RNA 的转录。在一个实施方案中,3' UTR 可以为卵清蛋白或溶菌酶基因的 3' UTR,或在哺乳动物细胞中具有功能的任何 3' UTR,即 SV40 晚期区域。

[0206] 在一个实施方案中,组成型启动子用于在禽类中表达转基因的编码序列。在该情况下,表达不限于膨大部 (magnum);表达还可在禽类的其它组织 (例如,血液) 中发生。这样的转基因 (其包括组成型启动子和编码序列) 的使用特别适合于实现或驱动蛋白质在输卵管中表达,并且随后将蛋白质分泌至卵中。

[0207] 产生用于载体的转导颗粒 (transducing particle, 即, transduction particle), 对其进行滴定以确定可用于注射胚胎的适当浓度。按照 Speksnijder 方法 (美国专利 No. 5, 897, 998, 将其公开内容通过引用整体并入本文) 给禽卵开口,用转导颗粒注射卵。注射后将卵孵育约 21 天,选择雄性禽类用于交配。为了筛选在其精子中包含转基因的 G0 公鸡,从公鸡精子样品提取 DNA。通过人工授精将在其精子样品中具有最高转基因水平的 G0 公鸡与非转基因母鸡交配。筛选血液 DNA 样品的转基因的存在。测试转基因公鸡的血清的

外源蛋白质的存在。如果外源蛋白质得到确认,则转基因公鸡的精子用于非转基因母鸡的人工授精。一定百分比的后代则包含转基因(例如,超过 50%)。当外源蛋白质存在于根据本发明的产生的卵中时,可分离蛋白质。还可测试蛋白质的生物学活性。

[0208] 在禽类输卵管中产生外源蛋白质和产生包含外源蛋白质的卵的本发明的方法包括随后提供适当的载体和将所述载体引入胚部胚盘细胞以便载体被整合入禽类基因组的另外的步骤。随后的步骤包括从在先前步骤中产生的转基因胚盘细胞获得成熟转基因禽类。成熟转基因禽类可获自己用载体直接在胚胎内对其进行转染或转导的胚盘胚胎的细胞。使所得的胚胎发育并且使小鸡长大成熟。

[0209] 从胚盘细胞产生的转基因禽类称为建立者(founder)。一些建立者将在它们的输卵管的膨大部的管腺细胞中具有转基因。此类禽类将在它们的输卵管中表达由转基因编码的外源蛋白质。外源蛋白质还可在除了输卵管以外的其它组织(例如,血液)中表达。如果外源蛋白质包含适当的信号序列,其可被分泌至输卵管的管腔中和卵的卵清中。

[0210] 一些建立者为种系建立者。种系建立者为在其种系组织的遗传物质中具有转基因,以及还可在表达外源蛋白质的输卵管膨大部管腺细胞中具有转基因的建立者。因此,根据本发明,转基因禽类可具有表达外源蛋白质的管腺细胞,并且转基因禽类的后代还可具有表达外源蛋白质的输卵管膨大部管腺细胞。或者,后代表达通过外源基因在禽类的特定组织中表达决定的表型。在一个实施方案中,转基因禽类为鸡或火鸡。

[0211] 药物组合物和治疗方法

[0212] 虽然可能就疗法中的用途而言,可以以粗制形式施用如本文中所述产生的治疗性蛋白质,但优选将治疗性蛋白质作为药物制剂的一部分来进行施用。因此,还提供了包含家禽来源的糖基化治疗性蛋白质例如 LAL 或其药学上可接受的衍生物与一种或多种其药学上可接受的载体以及任选地,其它治疗性和/或预防性成分的药物制剂,和施用此类药物制剂的方法。载体在与制剂的其它成分相容并且对其接受者无害的意义上必须是“可接受的”。使用本发明的药物组合物治疗患者的方法(例如,施用的药物蛋白质的量,施用频率和治疗期的持续时间)可使用本领域技术人员已知的标准方法来确定。

[0213] 利用公知的常规方法(参见,例如,Remington's Pharmaceutical Sciences,第 14 版., Mack Publishing Co., Easton, Pa., 将其整个教导通过引用并入本文)配制包含载体,包括复合分子(composite molecules)的组合物。载体可包括赋形剂。在一个实施方案中,药物载体可以是液体并且重组人 LAL 可以以溶液的形式存在。药物载体可以为蜡、脂肪或酒精。在一个实施方案中,基于蜡或脂肪的载体不包含酯。在另一个实施方案中,药学上可接受的载体可以是以粉剂、冻干粉剂或片剂形式存在的固体。在一个实施方案中,载体可包含脂质体或微胶囊。

[0214] 药物制剂包括适合通过注射(包括肌内、皮下和静脉内施用)施用的那些制剂。药物制剂包括适合于口服、经直肠、经鼻、局部(包括含服和舌下)、经阴道或胃肠外施用的那些制剂。药物制剂还包括用于通过吸入或吹入施用的那些制剂。在适当的情况下,制剂可以方便地以分开的剂量单位提供,以及可利用制药领域内公知的任何方法来制备。产生药物制剂的方法通常包括步骤:将治疗性蛋白质与液体载体或精细固体载体或两者结合,随后,需要时,将产品塑造成期望的形式。

[0215] 适用于口服施用的药物制剂可方便地以分开的单位形式例如胶囊剂、扁囊剂

(cachet) 或片剂提供,所述单位各自包含预定量的活性成分;以粉剂或颗粒剂形式;以溶液形式;以悬浮液形式;或以乳剂形式提供。活性成分还可以以大丸剂(bolus)、药糖剂(electuary)或糊剂形式提供。用于口服施用的片剂和胶囊剂可包含常规赋形剂例如结合剂、填充剂、润滑剂、崩解剂或湿润剂。可按照本领域公知的方法包被片剂。口服液体制剂可以以例如水性或油性悬浮液、溶液、乳剂、糖浆剂或酏剂的形式存在,或可以以干燥产品(以在使用之前用水或其它适当的媒介物构建)提供。此类液体制剂可包含常规添加剂例如悬浮剂、乳化剂、非水性媒介物(其可包括食用油)或防腐剂。

[0216] LAL 还可被配制用于胃肠外施用(例如,通过注射,例如快速推注注射(bolus injection)或连续输注),以及可以在安瓿、预装注射器、小体积输注中以单位剂量形式提供或在添加有防腐剂的多剂量容器中提供。可以通过例如皮下注射、肌肉注射和静脉内输注或注射来注射治疗性蛋白质。

[0217] LAL 可采取如油性或水性媒介物中的悬浮液、溶液或乳剂等形式,以及可包含配制剂(formulatory agent)例如悬浮剂、稳定剂和/或分散剂。还涉及:治疗性蛋白质可以以粉剂形式存在,通过无菌固体的无菌分离或通过从溶液冻干(用于在使用之前利用适当的媒介物例如无菌无热原水构建)获得的。

[0218] 关于静脉内输注或注射,可根据本发明产生的LAL 配制为水性悬浮液或溶液。适用于用于静脉内输注或注射的制剂的赋形剂可包括下列试剂之一:脱水柠檬酸三钠、柠檬酸和人血清白蛋白。药物制剂还可包括本领域公知的用于溶酶体贮积症的其它产品的其它适当的赋形剂。将根据本发明产生的LAL 的pH 维持在约5.6 与约6.2 之间。优选地,将LAL 制剂的pH 维持在 5.9 ± 0.2 。

[0219] 关于至皮肤的局部施用,可将按照本发明产生的本发明的治疗性蛋白质配制为软膏、乳膏剂或洗剂,或配制为透皮贴剂。软膏和乳膏剂可以例如利用添加有适当的增稠剂和/或胶凝剂的水性或油性基质来配制。洗剂可利用水性或油性基质来配制,并且还可包含一种或多种乳化剂、稳定剂、分散剂、悬浮剂、增稠剂或着色剂。

[0220] 适用于口中局部施用的制剂包括在调味基质(通常蔗糖和阿拉伯树胶或黄蓍胶)中包含活性成分的锭剂;在惰性基质例如明胶和甘油或蔗糖和阿拉伯树胶中包含活性成分的软锭剂(pastille);和在适当的液体载体中包含活性成分的漱口药。

[0221] 适用于直肠施用的药物制剂(其中载体为固体)最优选地以单位剂量栓剂的形式提供。适当的载体包括可可脂和本领域通常使用的其它材料,并且栓剂可通过将活性化合物与软化的或熔化的载体混合,然后在模具中冷却和塑形来方便地形成。

[0222] 适用于阴道施用的制剂可以以阴道栓、止血栓(tampon)、乳膏剂、凝胶、糊剂、泡沫或喷雾剂形式提供,所述制剂除活性成分外还包含这样的载体如在本领域内已知是适当的载体。

[0223] 关于鼻内施用,本发明的治疗性蛋白质可用作液体喷雾剂或飞散性散剂或以滴剂的形式存在。滴剂可利用水性或非水性基质(还包含一种或多种分散剂、增溶剂或悬浮剂)来配制。液体喷雾剂可从加压包(Pressurized pack)方便地进行递送。

[0224] 关于通过吸入的施用,根据本发明的治疗性蛋白质可从吹入器、喷雾器或加压包或递送气雾喷雾剂的其它方便的工具来方便地递送。加压包可包含适当的推进剂例如二氯二氟甲烷、三氯氟甲烷、二氯四氟乙烷、二氧化碳或其它适当的气体。在加压气雾剂的情况

下,剂量单位可通过提供阀门以递送计量的量来确定。

[0225] 关于通过吸入或吹入的施用,根据本发明的治疗性蛋白质可采取干燥粉末组合物的形式,例如,化合物与适当的粉末基质例如乳糖或淀粉的粉末混合物。粉末组合物可以在例如胶囊或药筒或例如明胶或泡罩包装 (blister pack) (可借助于吸入器或吹入器从其施用粉剂) 中以单位剂量形式提供。

[0226] 需要时,可使用适合于提供活性成分的持续释放的上述制剂。

[0227] 本文中描述的药物组合物还可包含其它活性成分例如抗微生物剂或防腐剂。

[0228] 此外,本发明涉及,可将本文中公开的治疗性蛋白质与其它治疗剂组合使用。例如,本发明提供了用于利用药物有效剂量的抗组织胺预处理以使潜在的输注相关过敏反应降至最低或防止所述过敏反应的方法。例如,抗组织胺可以是本文中公开的和本领域已知的任何药学上可接受的抗组织胺 (例如,苯海拉明)。在一个实施方案中,以约 1mg 至约 10mg/ 千克体重的剂量施用抗组织胺。例如,可以以约 5mg/ 千克的剂量施用抗组织胺。在一个实施方案中,施用抗组织胺,进行约 10 分钟至约 90 分钟,例如约 30 分钟至 60 分钟,然后使用与血管入口 (vascular access port) 连接的移动系统 (ambulatory system) 施用溶酶体酸性脂酶。在一个实施方案中,苯海拉明的剂量有效地抵消了潜在的过敏性输注反应。

[0229] 如果患者将经历过敏反应或不利的免疫反应,则还可在 LAL 施用之前、期间或之后施用免疫抑制剂例如抗组织胺、皮质类固醇、西罗莫司、伏环孢素、环孢素、甲氨蝶呤、针对 IL-2 受体的抗体、针对 T 细胞受体的抗体、针对 TNF- α 的抗体或融合蛋白 (英夫利昔单抗、依那西普或阿达木单抗)、CTLA4-Ig (例如,阿巴他塞)、抗 -OX-40 抗体。

[0230] 本发明还涉及包括将含 LAL 的组合物与一种或多种降胆固醇试剂 (例如,HMG-CoA 还原酶抑制剂) 组合施用的疗法。此类试剂的非限定性实例包括:阿托伐他汀 (Lipitor[®] 和 Torvast[®])、氟伐他汀 (Lescol[®])、洛伐他汀 (Mevacor[®], Altocor[®], Altoprev[®])、匹伐他汀 (Livalo[®], Pitava[®])、普伐他汀 (Pravachol[®], Selektine[®], Lipostat[®])、罗舒伐他汀 (Crestor[®]) 和辛伐他汀 (Zocor[®], Lipex[®])。

[0231] 本文中描述的组合物或蛋白质可用于治疗多种病况。例如,存在用于其的治疗性疗法对于本领域技术人员来说是已知的病况。本发明涉及:在禽类系统中产生的包含家禽来源的糖基化模式的治疗性蛋白质 (例如,LAL) 可用于治疗此类疾病。即,还涉及通过使用如本文中所述产生的治疗性蛋白质治疗已知可利用常规产生的治疗性蛋白质治疗的病况。例如,如本文中所述产生的 LAL 可用于治疗因 LAL 缺乏或不足 (统称“LAL 缺乏”) 而引起的或与其相关的病况,例如沃尔曼病和胆固醇酯沉积病 (CESD)。如本文中所描述的,LAL 缺乏还涉及其中 LAL 的表达因导致体内产生的 LAL 的减少或不足的状况 (例如,基因突变)、生理或环境因素而减少的病况。如本文中所述产生的 LAL 还可用于治疗其它病况例如动脉粥样硬化、脂肪肝病、非酒精性脂肪肝、非酒精性脂肪性肝炎 (NASH) 和肝硬化。如本文中所述产生的 LAL 还可用于治疗其它病况例如 2005 年 2 月 1 日发布的美国专利 No. 6,849,257 (将所述专利的公开内容通过引用整体并入本文);2009 年 12 月 3 日公布的美国公布 No. 2009/0297496;2004 年 11 月 11 日公布的美国公布 No. 2004/0223960;2009 年 11 月 15 日公布的美国公布 No. 2007/0264249 (将所述公布的公开内容 (即,这 4 个专利公布的每一个的公开内容) 通过引用整体并入本文) 中公开的病况。

[0232] 还涉及：本文中公开的产生的 LAL 可用于治疗某些特殊病况，包括胰腺炎例如慢性胰腺炎和 / 或急性胰腺炎以及酒精诱导的胰腺损伤例如酒精诱导的胰腺炎。

[0233] 本发明涉及：通过任何有用的方法（例如本文中公开的方法）产生的 LAL 用于治疗因酒精诱导的细胞损伤（包括但不限于导致脂质酯 (lipid ester) 在身体组织包括但不限于肝、脾、肠和心血管组织中积累的那些酒精诱导的细胞损伤）而引起的疾病。本发明还涉及通过施用 LAL 治疗吸收不良。

[0234] 本发明的一个方面涉及治疗患者的方法，包括向患者施用治疗有效量的包含本文中描述的重组人 LAL 的组合物。患者可患有或经诊断患有许多病况，包括与 LAL 缺乏相关的病况。在一个实施方案中，治疗有效量为使患者的红细胞计数增加期望量的量。本发明涉及：根据本发明产生的 LAL 可用于治疗慢性肾病，例如，当组织不能支持溶酶体酸性脂酶的产生时。

[0235] 还涉及通过任何有用的方法产生的 LAL 可用于治疗患有丹吉尔病和家族性低 α 脂蛋白血症的患者。丹吉尔病 / 家族性低 α 脂蛋白血症与胆固醇酯在巨噬细胞中的积累相关，伴随肝脾肿大和 / 或淋巴结病以及低水平的高密度脂蛋白 (HDL)，其可通过施用 LAL 来治疗。例如，不希望将本发明限定于任何特定的操作理论或机制，据认为受损的 LAL 活性减少 ABCA1 表达，相反地通过向患有丹吉尔病 / 家族性低 α 脂蛋白血症的患者施用 LAL 获得的增强的 LAL 活性将增加 ABCA1 表达以克服因多态性而具有减小的功能活性的 ABCA1 基因的作用。

[0236] 关于病况的治疗，一般而言，取决于已知的因素例如接受者的年龄、健康和体重、现有治疗的类型、治疗的频率等，施用的剂量可变化。通常，活性成分的剂量可为约 0.0001 至约 10 毫克 / 千克体重。治疗的精确剂量、施用频率和时间跨度可由施用各自治疗性蛋白质的领域内的医生来确定。

[0237] 此外，已发现 1mg/kg 和更少的剂量在治疗 LAL 缺乏中可以是有效的。本发明提供了治疗病况的方法，包括每 5 天一次和每 25 天一次，例如每 7 天一次和每 14 天一次向哺乳动物（例如患者，优选人患者）施用治疗有效剂量的溶酶体酸性脂酶。在一个实施方案中，施用的溶酶体酸性脂酶的剂量为约 0.1mg 至约 50mg / 千克体重，例如剂量可以为约 1mg 至 5mg / 千克。

[0238] 在一个特别有用的实施方案中，本发明提供了通过按照任何治疗有效给药方案例如本文中描述的给药方案施用约 0.1mg 至 1.0mg / 千克体重的溶酶体酸性脂酶的剂量治疗病况的方法。

[0239] 本发明提供了用于治疗 LAL 缺乏的任何并发症的方法，所述并发症可从施用治疗有效剂量的 LAL 受益。在一个实施方案中，可按照本文中描述的方法治疗吸收不良和生长不足。在另一个实施方案中，可使用本文中提供的方法治疗 LAL 缺乏患者中看见的并发症包括但不限于肝肿大和肝功能障碍。

[0240] 本发明提供了利用重组 LAL（例如，重组人 LAL）的治疗，所述重组 LAL 可利用任何有用的蛋白质表达系统例如转基因哺乳动物和禽类（如本领域中所理解的）来产生。其它蛋白质表达系统可包括但不限于细胞培养物、细菌和植物系统。

[0241] 本发明包括将重组 LAL 作为药学上可接受的组合物的一部分通过任何途径进行的施用，所述途径可获得期望的疗效，如由本领域医生测定的。在一个实施方案中，可在约

5 小时的时间段内通过静脉内输注施用 LAL。例如,可利用连接于血管进出口(例如,一个端口-一个导管(aPort-a-Cath))的移动输注泵(ambulatory infusion pump)帮助输注。

[0242] 本发明还包括监测哺乳动物(例如,人患者)的病况例如沃尔曼病和 CESD 的临床和病理学表现。在一个实施方案中,评估由(但不限于)如下方面组成:脂质分析、胸部 X 光检查、肝功能试验、粪便记录(stool chart)、血浆甲羟戊酸(plasma mevalonic acid)、免疫原性、血浆溶酶体酸性脂酶、壳丙糖酶、PARC、门静脉高压、人体测量术(anthropometry)、肝、脾和胃肠道的体积和表征(使用例如成像技术)。例如,上述成像技术可由超声、磁共振成像以及核磁共振光谱学组成。

实施例

[0243] 本发明还可通过下列实施例进一步举例说明。所述实施例仅用于举例说明目的,而并非旨在也不应当将它们解释为以任何方式限定本发明。

[0244] 实施例 1

[0245] 具有重组人溶酶体酸性脂酶(rhLAL)编码序列的的载体(pALVIN-OVR1-I-hLAL-dSA)的构建

[0246] pALVIN-OVR1-I-hLAL-dSA 载体中的 hLAL 基因的核苷酸序列编码与由人溶酶体酸性脂酶基因产生的蛋白质的氨基酸序列相同的蛋白质(GenBank 登录号, NP_000226)(图 1)。该序列的转录以及所得 mRNA 的随后翻译产生 399 个氨基酸的前体蛋白质,其被加工成与 SEQ ID NO:1 中所示的人 LAL(GenBank 登录号, NP_000226)(图 1)相同的成熟的 378 个氨基酸的蛋白质。通过来源于卵清蛋白基因的非编码元件(包括增强子、启动子、内含子以及 5' 和 3' 非翻译序列)来控制本实施例中的 hLAL 基因(关于 cDNA 序列参见图 2)的表达。卵清蛋白基因产生卵清蛋白,卵清的主要蛋白质组分。鸡卵清蛋白启动子的活性对于鸡输卵管内产生卵清的细胞的特异性极高;在其它组织中的表达是最小程度的。

[0247] 质粒载体 pALVIN-OVR1-I-hLAL-dSA(图 3A;其核苷酸序列示于图 4 中)用于产生将 hLAL 转基因稳定地整合入建立者(XLL109)的基因组的复制缺陷型逆转录病毒(RDR)。该质粒载体包括病毒 RNA 包装、逆转录和整合所需的逆转录病毒核苷酸序列,但不包含病毒 gag、pol 和 env 基因的完整序列。本文中描述了用于产生逆转录病毒载体的方法及其在随后转基因(transgenesis)法中的用途。

[0248] pALVIN-OVR1-I-hLAL-dSA 的逆转录病毒部分基于 ALV 载体 pNLB。对 pNLB 进行修饰以便 LTR 可自身灭活(self-inactivating)(SIN)(图 3B)。为了实现该目的,缺失 3'LTR 的 273bp,其包括 U3 区域的增强子和 CAAT 盒。因为逆转录病毒序列的 3' 末端上的灭活的 U3 区域用作存在于整合的原病毒的 5' 末端上的新的 U3 区域的模板,所以 5' LTR 通常也被灭活。SIN 构建体的 LTR 序列的缺失减小了启动子对来自 LTR 的内部启动子的干扰,并且使序列重组以形成具有复制能力的逆转录病毒的概率降至最低。新的载体称为 pALVIN(对于 ALV 灭活载体)。

[0249] 5' LTR 的下游是部分 gag 和 env 编码序列,所述序列皆转自 pNLB 载体。在 pALVIN-OVR1-I-hLAL-dSA 中,gag 蛋白前体序列的小部分(12%)保留(55%的 p19 成熟肽序列)以及 RAV2 的 env 前体序列的小部分(1.7%)保留(GenBank 登录号, AF033808)。这些截短的 gag 和 env 区域不能产生用于产生具有复制能力的逆转录病毒所必需的功能性蛋

白质 (Cosset, 1991)。

[0250] 将鸡卵清蛋白基因的转录和翻译控制元件插入 pALVIN 以产生 pALVIN-OV-1.1-I (其序列示于图 6 中; SEQ ID NO:8)。pALVIN-OV-1.1-I 的第一部分由鸡卵清蛋白基因的连续部分组成, 所述连续部分包括 1.1kb 近端启动子区、第一外显子、第一内含子以及第二外显子的一部分。下一部分为填充插入物片段, 其替代了卵清蛋白编码序列。填充片段之后是鸡卵清蛋白基因的 3' 非翻译区 (UTR), 其包括有利于 mRNA 正确加工 (包括多腺苷酸化) 的序列。一般地, 填充片段被编码期望的蛋白质 (在该情况下 hLAL) 的 DNA 片段替代。结果为具有特殊元件的载体, 所述元件在转基因鸡的输卵管中促进 mRNA 的受控转录表达和翻译, 密切模拟内源卵清蛋白 mRNA 的调节, 以及使目标蛋白质能够在卵清中高表达。

[0251] pALVIN-OV-1.1-I 载体包括卵清蛋白基因的第一内含子。因为内含子在逆转录病毒 RNA 基因组的产生和包装过程中对剪接易感, 因此我们以相对于 LTR 相反的取向插入表达盒。这样, 内含子在逆转录病毒 RNA 中不能被识别, 从而在无剪接的情况下被包装。为了方便起见, 绘制本说明书中的所有图谱, LTR 以相反取向绘制, 表达盒以向前或顺时针方向取向绘制。

[0252] pALVIN-OV-1.1-I 为其中已插入了 hLAL 的编码序列 (CDS) 的基本载体。在 Integrated DNA Technologies, Coralville, Iowa 合成组成 hLAL CDS 和与 pALVIN-OV-1.1-I 相容的所需的序列的两个 DNA 片段 hLAL 适体和 Syn hLAL, (参见图 7 和 8; SEQ ID NO:9 和 10)。将 hLAL 适体的 229bp HpaI/BamHI 片段和 Syn hLAL 的 1113bp BamHI/BstBI 片段插入 pALVIN-OV-1.1-I 的 7882HpaI/BstBI 片段, 从而用 hLAL CDS 替代填充片段区域, 产生 pALVIN-OV-1.1-I-hLAL。

[0253] 已发现, 在 hLAL CDS 的反义链中存在阻止完整逆转录病毒 RNA 包装的隐蔽剪接位点。通过改变 hLAL 的 DNA 序列而不改变其氨基酸序列来除去隐蔽剪接位点。该改变通过用引物 5' -AGAACTGAGAGTGTCTTAT-3' (SEQ ID NO:12) 和引物 5' -TGACAGCTGTGGATCCAGAAACAAACATG-3' (SEQ ID NO:13) 对 pALVIN-OV-1.1-I-hLAL 的区域 232 至 534 进行聚合酶链式扩增来进行, 产生 329bp 扩增子。用 BamHI 和 SexAI 降解该扩增子, 将其与 pALVIN-OV-1.1-I-hLAL 的 8940bp BamHI/SexAI 片段连接以产生 pALVIN-OV-1.1-I-hLAL-dS A。

[0254] 将包含鸡卵清蛋白基因的 DNase 酶超敏位点 III (DHSIII) (相对于 OV 启动子起始位点 -3819 至 -2169) 的假定的启动子增强子 (Kaye, Bellard 等人, 1984) 插入 pALVIN-OV-1.1-I-hLAL-dSA 以产生 pALVIN-OVR1-I-hLAL-dSA。如下所述进行该过程: 通过用 XhoI 和 B1pI 降解分离包括 DHSIII 增强子和称为 OVR1 启动子的 1.1kb 近端 OV 启动子 (关于序列参见图 9; 和 SEQ ID NO:11) 的 DNA 片段。为了有利于亚克隆, 利用引物 5' -GCCGCTCGAGCGAGGAATATAAAAAATT-3' (SEQ ID NO:14) 和 5' -TCCGCGCACATTTCCCGAA-3' (SEQ ID NO:15) 通过 PCR 扩增 pALVIN-OV-1.1-I 的区域 6752 至 7974, 然后用 NgoMI 和 XhoI 降解来产生适体片段, pSIN-OV-1.1-I 的 PCR。将 OVR1 启动子的 2772bp XhoI/B1pI 片段和 pSIN-OV-1.1-I 的 PCR 的 1067bp NgoMI/XhoI 片段插入 pALVIN-OV-1.1-I-hLAL-dSA 的 7043bp NgoMI/B1pI 片段, 从而产生 pALVIN-OVR1-I-hLAL-dSA (关于 pALVIN-OVR1-I-hLAL-dSA 的构建示意图参见图 10)。在

美国专利申请 2008/0064862 中描述了称为 pALVIN(aka pAVIJCR-A395. 22. 3. 1-KM 或 pALV-SIN) 的载体的逆转录病毒载体区段的构建。

[0255] 此外,还包括使用 2008 年 3 月 13 日公布的美国专利公布 No. 2008/0064862(将其公开内容通过引用整体并入本文)中公开的启动子和 / 或载体产生根据本发明的 LAL。

[0256] 实施例 2

[0257] 病毒颗粒的产生

[0258] 通过如下使用逆转录病毒转基因法产生在其基因组中具有 hLAL 转基因的 G0 建立者转基因雄性 XLL109。通过永生化的鸡成纤维细胞系的瞬时转染产生具有 pALVIN-OVR1-I-hLAL-dSA 载体的复制缺陷型病毒颗粒。用三种质粒 pALVIN-OVR1-I-hLAL-dSA、pCMV-gag-pol 和 pCMV-VSV-G 同时转染此类鸡成纤维细胞。pCMV-gag-pol 表达禽白血病病毒的 RAV1 株的 gag 和 pol 基因。pCMV-VSV-G 表达水疱性口炎病毒的包膜蛋白。转染后 4 小时,用补充有 10% 胎牛血清、100 单位 /mL 青霉素和 100 μ g/mL 链霉素的 DMEM 替换培养基。转染后 48 小时收获培养基,将其通过 0.45 微米过滤器 (Millipore) 进行过滤,然后通过超速离心进行浓缩。收集浓缩的具有 ALVIN-OVR1-I-hLAL-dSA 转基因的逆转录病毒,将其用于早期胚胎的转导。注意:因为“p”是载体的质粒形式的标记符 (notation),因此一旦转基因以包装的载体或整合的转基因的形式存在,“p”就不存在于转基因名称中。

[0259] 实施例 3

[0260] 胚胎转基因

[0261] 通过早期胚胎的转导实现 ALVIN-OVR1-I-hLAL-dSA 表达盒至胚胎基因组的整合 (Speksnij der 和 Ivarie, 2000)。从繁殖集落 (breeding colony) 获得新产的受精白色来亨鸡的卵。在壳上产生开口以提供至胚胎的通路。将 7 微升浓缩的具有上述 ALVIN-OVR1-I-hLAL-dSA 表达盒的复制缺陷型逆转录病毒颗粒注射入胚胎的胚下腔。用热胶密封卵,随后将其在标准条件下进行孵育和孵化。从这些注射产生的后代在孵化时被赋予单个鉴定标志以进行鉴定和跟踪。当使用对于 hLAL 编码序列特异性的 PCR 引物,通过对 hLAL 转基因进行实时 PCR 来进行分析 (如下文中所述) 时,来自后代的血液样品是转基因阳性的。这表明转基因法是成功的。hLAL 转基因的实时 PCR 测定利用 Taqman[®] 化学制品 (Applied Biosystems)。正向和反应引物分别为 5' -ACGACTGGCTTGCAAGTGTCT-3' (SEQ ID NO:16) 和 5' -CCCCAAATGAAGTCAAGATGCT-3' (SEQ ID NO:17)。Taqman[®] 探针序列为 5' -CCGGAATGCTCTCATGGAACACCAA-3' (SEQ ID NO:18) 并且在 5' 末端用 FAM (作为发射剂) 标记以及在 3' 末端用 Iowa 黑 (作为淬灭剂) 标记。将引物、探针和 1 μ l 的提取 DNA 添加至 30 μ l Taqman[®] 通用预混合物 (Applied Biosystems) 中。对照反应包括具有 hLAL 序列和来自野生型鸡的 DNA 的质粒的不同稀释物 (数据未示出)。在 Applied Biosystems 7500Fast 实时 PCR 系统上使用标准循环参数。

[0262] 实施例 4

[0263] G0 建立者的鉴定

[0264] 从性成熟雄性收集精液,提取 DNA,使用 hLAL 实时 PCR 测定来进行测定。使用与阴性对照精液 DNA 混合的已知标准物 (具有 hLAL 基因的质粒) 估计每一个样品的转基因拷贝的数目。雄性 XLL109 中的转基因盒 DNA 含量在允许转基因被传递至其后代的水平上,如

通过实时 PCR 估计的。该 XLL109 雄性为 G0 转基因建立者,将其与非转基因鸡交配,以产生 G1 半合子转基因鸡。

[0265] 实施例 5

[0266] 半合子 G1 禽类的繁殖和表征

[0267] 使用 hLAL 实时 PCR 测定测试由转基因建立者 XLL109 产生的后代的血细胞 DNA 中转基因的存在。从 1-2 周龄后代收集血液,使用高通量技术 (Harvey 等人,2002) 提取 DNA。在进行 Taqman 测定以促进高通量筛选之前不定量 DNA 溶液。通常 1 μ l DNA 溶液包含 50 至 400ng DNA,所述 DNA 足以产生阳性扩增信号。测试总共 1,322 只由 XLL109 产生的小鸡,使阳性后代再交配,然后测试以进行确认。根据 PCR 结果,22 个后代对于 ALVIN-OVR1-I-hLAL-dSA 转基因为阳性。Taqman 结果的实例示于图 11 中。

[0268] 实施例 6

[0269] 高表达品系的鉴定和表征

[0270] G1 鸡之一 ILL7466 产下在卵清中具有与其它 G1 鸡相比较显著更高的水平的 rhLAL 蛋白质的卵。对 ILL7466 和同胞 G1 雄性进行 Southern 印迹分析以鉴定具有与所述高表达鸡相同的整合位点的同胞雄性。利用在转基因内仅切割一次的限制性内切酶 (B1pI) 进行降解,利用卵清蛋白启动子或 hLAL 编码序列的区段探测 Southern 印迹 (图 12A-D)。取决于整合的位点,存在于侧翼基因组区域中的第二限制性位点的位置可变化。因此,通过 OV 探针或 hLAL 探针检测到的 B1pI 条带的尺寸对于产生的每个品系是唯一的。

[0271] OV 探针在 B1pI- 降解的野生型鸡的 DNA 中检测到 4.1kb 的单个条带,该条带对应于鸡基因组的内源卵清蛋白基因内的 B1pI 区段的预期尺寸 (图 12B 和 12D)。对于鸡 ILL7466 检测到 4.3kb 的第二条带,该条带对应于转基因条带。3 个另外的雌性同胞 ILL10409、ILL10686 和 ILL12058 以及 3 个另外的雄性同胞 ILL8922、ILL9330 和 ILL11217 显示 4.3kb 的条带,表明这些同胞可能具有相同品系 (图 12B 和 12D)。

[0272] 如所预期的, hLAL 探针在野生型的 DNA 中未检测到条带如鸡溶酶体酸性脂酶基因的 DNA 序列,并且重组人溶酶体酸性脂酶的编码序列可被充分区分,从而不允许在这些 Southern 测定中使用的条件下杂交 (图 12C)。hLAL 探针在来自对于通过 OV 探针检测的 4.3kb 条带为阳性的相同鸡的 B1pI 降解的基因组 DNA 中检测到 ~ 10.6kb 的单个条带,这表明这 7 只 G1 鸡具有相同的整合位点,从而为相同品系。

[0273] 未检测到其它条带,这表明 ILL7466、ILL10409、ILL10686、ILL12058、ILL8922、ILL9330 和 ILL11217 都具有单个整合位点。

[0274] Southern 分析还表明转基因以被 OV 检测到的条带被整合, hLAL 探针具有不同尺寸并且在尺寸上大于来自单独的转基因的尺寸。显示整合的转基因的预测结构和 B1pI 位点在侧翼基因组区域中的位置的图谱示于图 12A 中。

[0275] 为了确认转基因是完整的,采用两个步骤。第一步,通过 PCR 从 ILL7466 分离 hLAL 编码序列。从 hLAL 起始密码子至终止密码子对 PCR 产物的两条链进行测序。DNA 序列完全如所预期的,这表明转基因中编码区的 DNA 序列没有改变。第二步,使用限制性内切酶 ApaLI 进行 Southern 印迹分析,所述酶将完整转基因降解成 2 个区段,3.6 和 3.8kb (图 13A)。在 ApaLI- 降解的 G1 的基因组 DNA 中同时检测到 3.6 和 3.8kb 条带,这表明转基因以完全完整的形式被整合 (图 13B)。

[0276] 实施例 7**[0277] G2 的繁殖和表征**

[0278] 图 14 显示从单个 G0 建立者 XLL109 起源的 hLAL G2 的谱系。在 G1 期,表征转基因的拷贝数、完整性、hLAL 序列和整合位点 - 以及鉴定和表征了 7 个 G1 转基因 (4 只小鸡和 3 只公鸡)。通过用从 G1 雄亲 ILL8922、ILL9330 和 ILL11217 收集的精液对非转基因鸡人工授精来实现 G2 的繁殖 (图 14)。将每只授精的鸡、其卵和随后的后代与其它后代分开饲养。使用 hLAL 实时 PCR 测定测试孵化的后代的 hLAL 转基因的存在。因为 G1 建立者对于转基因是半合子,因此一半的后代预期为转基因 G2。在迄今为止分析的 610 个 G2 后代中,330 个或 54% 为转基因的。

[0279] 实施例 8**[0280] hLAL 禽类的遗传分析**

[0281] 在通过血液 DNA 的 hLAL 实时 PCR 测定鉴定每只 G2 鸡后,将生产品系经历下列遗传测定:从血液 DNA PCR 扩增 hLAL 基因,对其进行测序以确认与人序列具有 100% 的同源性;如上所述,通过整合位点 PCR 确认转基因整合位点。在 < 10 只鸡的生产品系中对每一只鸡;对于 11-100 只鸡的生产品系对 10% 的鸡 (最少 10 只);对于 101-1000 只鸡的生产品系而言,对 5% 的鸡 (最少 10 只);对于 1001-10,000 只鸡的生产品系而言,对 1% 的鸡 (最少 50 只);对 > 10,001 只鸡的生产品系而言,对 0.1% 的鸡 (最少 100 只) 进行 PCR 测序和整合位点分析。在生长和生产期的每一个步骤保持详细记录。

[0282] 实施例 9**[0283] 从卵清纯化 hLAL**

[0284] 将含有 LAL 的卵清 (EW) 在 pH 6 下溶解过夜,通过利用 0.2 μ m 过滤的离心 (或深部过滤) 进行澄清。用 1M NaOAc 缓冲液 (pH 4) 将 EW 调整至 6。

[0285] 将澄清的 EW 加载至用 20mM 磷酸盐 /137mM NaCl 缓冲液 (pH 6) 平衡的苯基 -HIC 柱子 (EW : 柱子尺寸 = 2 : 1) 上。在完全加载后,用平衡缓冲液和 5mM 磷酸盐缓冲液 (pH 6) 洗涤柱子。用 30% 丙二醇和 5mM Tris 缓冲液 (pH 7.2) 洗脱 LAL。

[0286] 用 1M 酸将洗脱的 LAL 级分调整至 pH 5,随后将其加载至 GigaCap S 柱子 (EW : 柱子尺寸 = 10 : 1)。用 50mM NaOAc 缓冲液 (pH 5) 平衡柱子。在完成加载后,用平均缓冲液洗涤柱子。用 50mM NaOAc/60mM NaCl (pH 5) 洗脱 LAL。

[0287] 用 1M Tris 缓冲液将脱离 GigaCap S 柱的 LAL 级分调整至 pH6,随后将其加载至丁基 -HIC 柱子 (EW : 柱子尺寸 = 10 : 1) 上。用 20mM 磷酸盐 /137mM NaCl 缓冲液 (pH 6) 平衡柱子。在完成加载后,用平均缓冲液和 5mM 磷酸盐缓冲液 (pH 6) 洗涤柱子。用 50% 丙二醇和 5mM Tris 缓冲液 (pH 7.2) 洗脱纯 LAL。图 15 描述了从卵清纯化 hLAL 的步骤。

[0288] 实施例 10**[0289] 转基因禽类来源的 hLAL 的碳水化合物分析**

[0290] 通过使用下列对于本领域技术人员来说公知的分析技术测定禽类来源的人 LAL 的寡糖结构。

[0291] 用蛋白酶和胰凝乳蛋白酶在 37°C 下于 0.1M Tris-HCl, pH 8.2 (含有 1mM CaCl₂) 中降解 200 微克,进行 18 小时。富集降解产物,利用 Sep-Pak C18 筒柱 (cartridge column) 除去污染物。富集后,用 2 μ l 的 50 μ l 的 20mM 磷酸钠缓冲液 pH 7.5 中的 N-糖酰胺酶 F (7.5

单位 /ml) 在 37℃ 下降解糖肽 18 小时。通过将降解物通过 Sep-Pak C18 筒柱来将释放的寡糖与肽和酶分离。

[0292] 将聚糖级分溶解在二甲亚砜中,随后基于 Anumula 和 Taylor (Anumula 和 Taylor, 1992) 的方法对其进行全甲基化。通过加入水淬灭反应,利用二氯甲烷提取全 -O- 甲基化 (per-O-methylated) 碳水化合物。在氮流下干燥全 -O- 甲基化聚糖。

[0293] 使用 α -二羟苯甲酸 (DHBA, 50% 甲醇: 水中的 20mg/mL 溶液) 作为基质以反射体正离子模式 (reflector positive ion mode) 中进行 MALDI/TOF-MS (基质辅助激光解析电离飞行时间质谱)。通过使用 Microflex LRF (Bruker) 获得所有波谱。

[0294] 如本领域所理解的,在从肽主链释放和纯化后对寡糖进行 MALDI-TOF-MS 分析和 ESI MS/MS (电喷雾电离串联质谱 (electrospray ionization tandem mass spectrometry))。也用某些酶降解单个多糖种类的样品,如本领域中所理解的,利用 HPLC 分析降解产物。

[0295] 据认为在人 LAL 上存在约 6 个 N- 连接糖基化位点。参见, Zschenker 等人 (2005) J. Biochem., 第 137 卷, p 387-394, 将所述文献的公开内容通过引用整体并入本文。该参考资料还显示在人 LAL 上可存在 O- 连接糖基化位点。鉴定的 N- 连接寡糖结构示于图 16 中。

[0296] 数据显示在根据本发明产生的 LAL 中发现这些结构中有许多或全部为 N- 连接糖基化结构 (图 16)。例如,发现 A-n 连接于根据本发明产生的 LAL。例如,发现 O-n 连接于根据本发明产生的 LAL。例如,发现 B-n、C-n 和 D-n 的至少一个连接于根据本发明产生的 LAL。例如,发现 E-n 和 F-n 的至少一个连接于根据本发明产生的 LAL。例如,发现 I-n 和 J-n 的至少一个连接于根据本发明产生的 LAL。例如,发现 K-n 和 L-n 的至少一个连接于根据本发明产生的 LAL。例如,发现 M-n 和 N-n 的至少一个连接于根据本发明产生的 LAL。例如,发现 G-n 连接于根据本发明产生的 LAL。例如,发现 H-n 连接于根据本发明产生的 LAL。

[0297] 实施例 11

[0298] 转基因禽类来源的 LAL 的 N- 聚糖种类

[0299] 使用 Tube-O- 透析器 (4.0kDa 截断值的膜; G Biosciences) 在 4℃ 下将转基因禽类来源的 hLAL 的纯化的样品 (600 μ g/ 样品) 针对超纯水 (nanopure water) 透析 24 小时以除去盐和其它污染物。在整个透析期间更换超纯水 4 次。

[0300] 透析后,将每一个样品分成 3 等分: $\sim 1/4$ 的样品重量用于中性及氨基糖分析, $\sim 1/4$ 的样品重量用于甘露糖 -6- 磷酸分析, 以及 $\sim 1/2$ 的样品重量用于寡糖谱分析。在 100℃ 下用 2N 三氟乙酸 (TFA) 水解意欲用于中性及氨基糖分析的等分, 进行 4 小时, 在 100℃ 下用 6.75N TFA 水解用于甘露糖 -6- 磷酸分析的等分, 进行 1.5 小时。随后在 N_2 下干燥水解物, 用 50 μ l H_2O 将其再溶解, 在冰上超声 7 分钟, 随后转移至注射小瓶。然而, 将中性及氨基糖样品稀释 2 次, 因为从最初溶解的水解物产生的峰过大。

[0301] 以与样品相同的方式和在相同的时间水解具有已知摩尔数的用于中性及氨基糖以及用于甘露糖 -6- 磷酸的标准的混合物。制备 4 个浓度的中性及氨基糖标准混合物 (Fuc & GalNAc, 0.2、0.4、0.8 和 1.6 纳摩尔 /10 μ l; GlcNAc, 0.5、1.0、2.0 和 4.0 纳摩尔 /10 μ l; Gal & Man, 0.3、0.6、1.2 和 2.4 纳摩尔 /10 μ l; 和 Glc, 0.1、0.2、0.4 和 0.8 纳摩尔 /10 μ l) 和甘露糖 -6- 磷酸 (640、1280、2560、5120 皮摩尔 /10 μ l) 以建立校准方程。利用线性插值法, 根据校准方程定量样品中每种糖的摩尔数。

[0302] 使用配备有梯度泵、电化学检测器和自动采样品的 Dionex ICS3000 系统通过 HPAEC 分析中性及氨基糖以及甘露糖-6-磷酸。利用具有氨基捕捉器 (trap) 的 Dionex CarboPac PA20 (3x 150mm) 分析柱分离单独的中性及氨基糖与甘露糖-6-磷酸。梯度程序对于中性及氨基糖使用洗脱剂 A (脱气超纯水) 和 B (200mM NaOH), 对于甘露糖-6-磷酸使用 C (100mM NaOH) 和 D (1M 的 100mM NaOH 中的醋酸钠)。对于中性及氨基糖测定, 每 40 分钟进行一次注射 ($10 \mu\text{l}$ /注射), 对于甘露糖-6-磷酸测定, 每 35 分钟进行一次注射。所有方法基于由 Hardy 和 Townsend 描述的方案 (Hardy, M. R., 和 Townsend, R. R., " High-pH anion-exchange chromatography of glycoprotein-derived carbohydrates ", 1994, Methods Enzymol. 230 :208-225)。使用 Dionex chromeleon 软件实现仪器控制和数据获取。结果示于下面表 1 中。对照样品为从 EW 纯化的卵类粘蛋白。

[0303] 表 1 通过 HPAEC 分析的对照和 LAL 的单糖组成

样品 ID	分析物	纳摩尔	纳摩尔/ μg	摩尔%
对照	岩藻糖	nd	-	-
	N-乙酰半乳糖胺	5.066	0.020	9.6
	N-乙酰葡萄糖胺	26.947	0.108	51.4
	半乳糖	3.876	0.016	7.4
	葡萄糖	nd	-	-
	甘露糖	16.565	0.066	31.6
	甘露糖-6-磷酸	nd	-	-
	N-乙酰神经氨酸	ndm	-	-
	N-羟乙酰神经氨酸	ndm	-	-
[0304] 转基因禽 类来源的 hLAL	岩藻糖	nd	-	-
	N-乙酰半乳糖胺	nd	-	-
	N-乙酰葡萄糖胺	17.932	0.120	37.6
	半乳糖	0.879	0.006	1.8
	葡萄糖	nd	-	-
	甘露糖	23.290	0.155	48.8
	甘露糖-6-磷酸	5.642	0.038	11.8
	N-乙酰神经氨酸	ndm	-	-
	N-羟乙酰神经氨酸	ndm	-	-

[0305] nd = 未检测到 ; ndm = 未测定。

[0306] LAL 的结构特征

[0307] LAL 在其氨基酸序列中具有 6 个潜在的用于 N-连接糖基化的位点 Asn³⁶、Asn⁷²、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³ 和 Asn³²¹。这些位点中有 5 个 Asn³⁶、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³ 和 Asn³²¹ 经发现被糖基化, 然而 Asn⁷² 未被糖基化或大体上未被糖基化 (大体上未被糖基化意指在 LAL 分子的混合物中, 与 Asn³⁶、Asn¹⁰¹、Asn¹⁶¹、Asn²⁷³ 和 Asn³²¹ 的任一个相比较更少的 Asn⁷² 被糖基化)。因此, 本发明的一个方面为在 Asn⁷² 上未被糖基化和 / 或大体上未被糖基化的 LAL (例如, 人 LAL), 以及这样的 LAL 的产生和用途。然而, 具有糖基化 Asn⁷² 的 LAL 在本发明的范围内。N-聚糖结构主要由以 N-乙酰葡萄糖胺、甘露糖和甘露糖-6-磷酸 (M6P) 作为主要糖的二-、三-和四天线结构的混合物组成。每个位点似乎具有有利的一组结构 (表 2 和图 17), 所述结构为本发明的一个方面。例如, M6P-修饰的 N-聚糖至少存在于 Asn¹⁰¹、Asn¹⁶¹ 和 Asn²⁷³ 上。非磷酸化结构是内源卵清蛋白质上发现的 N-聚糖所具有的特征。未检测到 O-连接聚糖, 如通过 N-乙酰半乳糖胺 (GalNAc) 的缺乏测定的。未检测到唾液酸, 这与先前测定

的根据本发明产生的其它内源和外源蛋白质的 N- 聚糖结构一致。本发明包括利用一种或多种本文中公开的寡糖结构糖基化的 LAL。

[0308] 表 2. 如通过糖肽的 LC/MS 测定的 LAL 聚糖结构的存在位置

[0309]

位置	聚糖结构
Asn ³⁶	GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2
Asn ⁷²	没有被检测到
Asn ¹⁰¹	Phos2Man7GlcNAc2
Asn ¹⁶¹	Phos1Man6GlcNAc2 GlcNAc1Phos1Man6GlcNAc2 Man3GlcNAc2 GlcNAc2Man3GlcNAc2 GlcNAc3Man3GlcNAc2 GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2
Asn ²⁷³	Man7GlcNAc2 Man8GlcNAc2 Man9GlcNAc2 Phos1Man8GlcNAc2 Phos1Man9GlcNAc2
Asn ³²¹	GlcNAc2Man3GlcNAc2 GlcNAc3Man3GlcNAc2 GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2 GlcNAc5Man3GlcNAc2 Hex1GlcNAc5Man3GlcNAc2 GlcNAc6Man3GlcNAc2
[0310]	Hex1GlcNAc6Man3GlcNAc2

[0311] Hex, 半乳糖 ;Phos, 磷酸盐 ;Man, 甘露糖 ;GlcNAc2, N- 乙酰葡萄糖胺

[0312] 方法

[0313] 利用高 pH 阴离子交换色谱 - 脉冲安培检测法 (HPAEC-PAD) 定性和定量测定单糖组成, 包括中性、氨基和 M6P。

[0314] 利用来自若干质谱法 (MALDI-TOF、NSI-MS/MS 和糖肽 LC-MS) 的数据测定占优势的聚糖的结构。

[0315] MALDI-TOF 对于中性 N- 聚糖的测定是有用的并且能够检测磷酸化的 N- 聚糖 (图 18)。NSI-MS/MS 被用于测定 MALDI-TOF 波谱中的较小的峰的性质, 所述性质的一些性质归因于磷酸化的 N- 聚糖 (图 19)。提高 MALDI-TOF 检测磷酸化 N- 聚糖的能力的努力效果不佳。

[0316] 糖肽的 LC/MS 能够检测中性和磷酸化结构并且能够测定特定结构在 LAL 的氨基酸序列中的位置 (数据概述于图 17 和表 2 中)。

[0317] 为了确定 HPAEC-PAD 色谱图中哪个峰因磷酸化 N- 聚糖而引起, 利用磷酸酶处理 LAL, 然后进行分析 (图 3)。组 C 和 D 中的峰在曲线下面积 (AUC) 上减少, 然而组 A 中的峰变得更突出。组 B 中的峰未与其它峰成比例地改变。基于保留时间与电荷的程度 (归因于磷酸化或唾液酸化) 成正比的知识, 预期组 C 由具有一个磷酸 (单 M6P) 的 N- 聚糖组成并且组 D 由具有两个磷酸 (双 -M6P) 的 N- 聚糖组成。

[0318] 保留时间还受到中性和氨基单糖的组成和相对结构位置影响。此类实例包括半乳糖的存在、二等分 GlcNAc 的存在和 GlcNAc 置换的程度。此类因素促成了 HPAEC-PAD 色谱图中峰的多样性。

[0319] 实施例 12

[0320] 卵清中转基因禽类来源的 hLAL 的体外酶活性分析

[0321] 使用基本上如 Yan 等人 (2006), American Journal of Pathology, 第 169 卷, No. 3, p 916-926 (将其公开内容通过引用整体并入本文) 中所述的荧光底物 4- 甲基伞形酮基 - 油酸盐测定来测定卵清中的溶酶体酸性脂酶的活性。

[0322] 制备 4- 甲基伞形酮基油酸盐 (4-MUO) 的原液, 其由 2.5mM 的 4% Triton X-100 中的 4-MUO 组成。在微量滴定板上进行测定, 每个孔包含 62.5 μ l 的 0.2M 的 0.01% Tween80 中的柠檬酸钠 (pH 5.5)、12.5 μ l 的卵清样品和 25 μ l 的 2.5mM 4-MUO。使用 Bio-Tek Synergy HT 荧光微量板读数器 (激发 360nm 和发射 460nm) 在 37°C 下监测荧光的变化, 进行 30 分钟。在测定之前, 将包含 hLAL 的卵清稀释至导致反应线性持续至少 30 分钟的酶浓度。用 50 μ l 0.75M Tris-HCl, pH 8.0 终止反应, 在上面使用的相同板读数器中测量终点荧光信号 (激发 360nm 和发射 460nm)。

[0323] 使用 4- 甲基伞形酮基作为标准测定活性的单位。一个单位 (U) 定义为在上述测定条件下导致 1 微摩尔 4- 甲基伞形酮基 / 分钟的形成的酶的量。将含非 -hLAL 的蛋白用作阴性对照。

[0324] 对于 hLAL 为阳性的卵清样品包含 1U 至 100U 的活性 /ml 卵清。分析来自 21 只 G1 鸡的卵清。来自 10 只鸡的卵清经测试对于 hLAL 活性为阳性。

[0325] 实施例 13

[0326] 转基因禽类来源的 LAL 的体外分析

[0327] 使用巨噬细胞和成纤维细胞体外检查转基因禽类的输卵管细胞中产生的 LAL (在本文中称为“SBC-102”、“禽类来源的 LAL”、“LAL”或“hLAL”) 结合细胞和被内化至溶酶体区室的能力。当与巨噬细胞一起孵育时,发现荧光标记的 SBC-102 定位至溶酶体。可通过使用甘露糖多糖竞争剂减弱该作用,这暗示着 N-乙酰葡萄糖胺/甘露糖 (GlcNAc/甘露糖) 受体作为被此类细胞识别和摄取的机制。在体外孵育后,SBC-102 在 LAL- 缺乏人成纤维细胞和正常鼠成纤维细胞中增加细胞缔合的 LAL 活性,这表明对于 SBC-102 的暴露可导致对不足的酶促活性的充分替代。

[0328] 甘露糖 -6- 磷酸 (M6P) 存在于 SBC-102 的寡糖结构中,已显示所述寡糖结构通过遍在的 M6P 受体参与溶酶体酶至多种细胞类型的递送。

[0329] 从转基因母鸡的卵清纯化 LAL。俄勒冈绿 NHS 获自 Invitrogen™(#0-10241)。大鼠肺泡巨噬细胞系 NR8383 和小鼠成纤维细胞系 NIH-3T3 获自 ATCC。LAL- 缺陷型沃尔曼成纤维细胞获自 Coriell 医学研究所 (Coriell Institute for Medical Research) 以及 LysoTracker®红获自 Invitrogen™。

[0330] 酶标记:按照制造商的推荐,用俄勒冈绿标记 4mg PBS 中的转基因禽类来源的 LAL,随后将反应物对 PBS 进行透析,随后浓缩。

[0331] 巨噬细胞摄取:将荧光标记的转基因禽类来源的 LAL (5 μ g/mL) 和 LysoTracker®红与 NR8383 细胞一起孵育 2 小时。利用共聚焦荧光显微镜在 488nm 和随后 514nm 使用连续扫描模式检查细胞。

[0332] 与甘露聚糖的竞争性抑制:将荧光标记的 SBC-102 (5 μ g/mL) 和甘露聚糖与 NR8383 细胞一起孵育 2 小时。对细胞进行胰蛋白酶处理,使用中位荧光强度作为终点利用荧光活化细胞分选术测量 LAL 摄取。

[0333] 使用巨噬细胞系 NR8383 检查转基因禽类来源的 LAL 被摄取和随后整合入靶细胞的溶酶体的能力。将荧光标记的转基因禽类来源的 LAL 和溶酶体标记“LysoTracker®红”(Invitrogen™) 与细胞一起孵育 2 小时。随后利用共聚焦荧光显微镜,使用连续扫描模式检查转基因禽类来源的 LAL 和溶酶体标记在这些细胞中的共定位 (图 20)。LAL 显示至溶酶体的定位,这与使用来自多种来源的 rhLAL 的类似体外研究一致。

[0334] 已使用巨噬细胞系 NR8383 通过竞争性结合测定评估转基因禽类来源的 LAL 对 GlcNAc/甘露糖受体的结合特异性 (图 21)。将 5 μ g/mL 的荧光标记的 (俄勒冈绿 (Oregon Green)) 转基因禽类来源的 LAL 和不同浓度的含甘露糖寡糖、甘露聚糖与细胞共孵育 2 小时。使用中位荧光强度作为终点,利用荧光活化细胞分选分析定量与无甘露聚糖对照相比的转基因禽类来源的 LAL 摄取的相对抑制。观察到转基因禽类来源的 LAL 结合 / 摄取的甘露聚糖剂量依赖性抑制,这与转基因禽类来源的 LAL :GlcNAcR 相互作用一致。

[0335] 此外,通过利用甘露糖 -6- 磷酸的竞争性实验显示成纤维细胞中甘露糖 -6- 磷酸介导的摄取 (结果未示出)。

[0336] 已使用正常和 LAL 缺乏细胞在体外检查了转基因禽类来源的 LAL 暴露增强细胞中 LAL 活性的能力。在浓度为 0、0.16 或 0.5 μ g/mL 的转基因禽类来源的 LAL 存在的情况下孵育从沃尔曼患者分离的成纤维细胞和正常鼠成纤维细胞 (NIH-3T3),进行 5 小时。随后洗涤细胞以除去非特异性信号,使用 4-MUO 底物测定细胞裂解物的 LAL 活性。内源性细胞 - 缔合的 LAL 活性在沃尔曼成纤维细胞中更低 (与在 NIH-3T3 中相比较),在用转基因禽类来源

的 LAL 繁育后,在两种细胞类型中都观察到活性的剂量依赖性增强(图 22)。

[0337] 实施例 14

[0338] 转基因禽类来源的 LAL 的体内分析

[0339] 始于 4 周龄,每周一次用 SBC-102(5mg/kg, IV) 或安慰剂处理 LAL- 缺乏 Yoshida 大鼠(即,纯合的)(参见 Kuriyama 等人(1990),Journal of Lipid Research,第 31 卷,p 1605-1611;Nakagawa 等人,(1995)Journal of Lipid Research,第 36 卷,p 2212-2218;以及 Yoshida 和 Kuriyama(1990)Laboratory Animal Science,第 40 卷,p 486-489),进行 4 周。对于每次施用,以两个相同的剂量(2.5mg/kg)间隔 30 分钟将 SBC-102 注射入大鼠尾静脉。在最终给药后 1 周检查大鼠和年龄匹配的野生型对照。以一式三份重复进行分析。

[0340] SBC-102 处理的动物的总体病理学检查显示除器官尺寸减小外肝的颜色正常化。SBC-102 处理的小鼠显示基本上正常的肝组织学,与媒介物处理的动物的泡沫状巨噬细胞的大量积累(数据未示出)形成鲜明对照。在 LAL^{-/-}大鼠中升高的血清丙氨酸和天冬氨酸转移酶水平在 SBC-102 处理的大鼠中也下降了(未示出)。

[0341] 测定每只大鼠的内脏和组织的质量,数据示于图 23 中。在每周以 5mg/kg 施用一次媒介物或 SBC-102,进行 4 周后, LAL^{-/-}和 LAL^{+/+}大鼠的器官尺寸表示为 8 周龄时测定的体重的百分比。

[0342] 将 SBC-102 处理的或媒介物处理的 Yoshida 大鼠的体重与野生型大鼠相比较,这示于图 24 中。以单次剂量或以分份剂量(在 4 小时的时间段内给出)通过 IV 注射给 LAL^{-/-}大鼠施用 SBC-102(5mg/kg) 或媒介物。LAL^{+/+}大鼠为年龄匹配的同窝对照。

[0343] 实施例 15

[0344] 甘油三酯分析

[0345] 对来自野生型动物、纯合安慰剂处理的动物和纯合 SBC-102 处理的动物的肝和脾组织进行甘油三酯分析。使用标准方法(即,MBL 国际甘油三酯定量试剂盒, Catalog#JM-K622-100)进行甘油三酯分析,以一式三份重复进行该分析。

[0346] 表 3:野生型和 LAL 缺乏大鼠的肝和脾甘油三酯的水平

[0347]

甘油三酯($\mu\text{g}/\text{mg}$ 湿重)			
	野生型(n=3)	安慰剂(n=3)	SBC-102(n=3)
肝	48	84	57
脾	3	22	4

[0348] 肝底物水平

[0349] 图 25 显示在每周以 $5\text{mg} \cdot \text{kg}^{-1}$ 施用一次媒介物或 SBC-102,进行 4 周后,在 8 周龄时测定的 WT 和 LAL 缺乏大鼠的肝胆固醇、胆固醇酯和甘油三酯水平。

[0350] 实施例 16

[0351] 剂量反应研究

[0352] 基于上面进行的研究,在 LAL^{-/-}大鼠中检查许多 LAL(“SBC-102”)的给药和给药安排(qw 和 qow)的药代动力学(PD)效应。在这些研究中,始于 4 周龄,以 0.2、1、3 和

5mg/kg, qow 的剂量, 或 0.35、1 和 5mg/kg, qw 的剂量, 通过 IV 注射施用 SBC-102, 进行 1 个月。结果显示体重 (BW) 增加 (图 26)、器官巨大症 (图 27) 和组织底物水平 (图 28) 的改善。血清转氨酶水平也随着 SBC-102 剂量增加而降低, 在更高的剂量上水平基本上达到野生型水平。

[0353] 实施例 17

[0354] 大鼠模型中重组 LAL 的施用

[0355] 在于 Yoshida 和 Kuriyama(1990)Laboratory Animal Science, 第 40 卷, p 486-489 中描述的 LAL 缺乏 Donryu 大鼠中评估利用重组人溶酶体酸性脂酶 (LAL) 的重复给药对体重、组织甘油三酯和胆固醇、肝肿大、脾肿大、淋巴结病、肠重量以及其它参数的作用 (也参见 Kuriyama 等人, (1990)Journal of Lipid Research, 第 31 卷, p 1605-1611; Nakagawa 等人, (1995)Journal of Lipid Research, 第 36 卷, p 2212-2218), 将所述文献的公开内容通过引用整体并入本文。

[0356] 在 4 周龄时, 将对于 LAL 缺失是纯合的 Donryu 大鼠 ($LAL^{-/-}$) 分配至待用在转基因鸡输卵管系统中产生的重组人 LAL 或盐水安慰剂给药的组中。野生型年龄匹配的同窝出生大鼠用作对照。每周一次进行 4 周 (总共 4 个剂量) 或每 2 周一次进行 4 周 (总共两个剂量) 以单次剂量或以分份剂量 (相隔 30 分钟给出的) 通过尾-静脉注射向 $LAL^{-/-}$ 大鼠给药。LAL 的剂量为 1mg/kg 或 5mg/kg。给药方案示于表 4 中。用苯海拉明 (5mg/kg) 预处理大鼠以抵消潜在的过敏反应, 基于先前在用于治疗溶酶体贮积病的酶替代疗法的动物模型中的经验的方法 (Shull 等人, (1994)Proceedings of the National Academy of Science, 第 91 卷, p. 12937; Bielicki 等人, (1999)The Journal of Biological Chemistry, 274, p. 36335; Vogler 等人, (1999)Pediatric Research, 45, p. 838.), 将所述文献的公开内容通过引用整体并入本文。

[0357] 图 29 显示每周施用 1mg/kg 的 LAL 或每周施用 5mg/kg 的 LAL 或每 2 周施用 5mg/kg 的 LAL 的大鼠的体重增长的日进展。可在图中看到在两个剂量大小和频率之间几乎不存在或不存在的疗效的差异。

[0358] 表 4: 称重和给药方案

[0359]

出生后天数	评估/进行的注射	
第 13 天	W	
第 14 天	E	

[0360]

第 20 天	I	
第 21 天	G	幼崽断奶
第 24 天	H	
第 25 天	E	
第 27 天	D	
第 28 天		第一次注射，对于每周一次和每 2 周一次的施用
第 31 天		
第 32 天		
第 34 天		
第 35 天		第二次注射，对于每周一次的施用
第 38 天		
第 39 天		
第 41 天		
第 42 天		第 3 次注射，对于每周一次的施用；第二次施用，对于每 2 周一次
第 45 天		
第 48 天		
第 49 天		第 4 次注射，对于每周一次的施用
第 55 天		
第 56 天		尸检

[0361] 用重组 LAL 处理的 LAL^{-/-}大鼠的病理学检查

[0362] 在实施例 1 描述的研究结束时，人道地使研究动物安乐死，进行尸检以检查总体病理学、组织病理学和临床化学。肉眼尸检包括身体的外表面、所有孔口以及颅腔、胸腔和腹腔以及它们的内容物的检查。测定大鼠的内脏和组织的质量，采集器官和组织，将其在 10% 中性缓冲福尔马林中进行固定。固定后，处理组织，制备和评估苏木精和伊红染色的切片的组织学载片。

[0363] 分析的处理的动物的总体病理学检查显示肝尺寸和颜色大体正常化，如可在图 30 中显示的解剖中看到的。测定器官体重比，所述器官体重比显示与安慰剂处理的大鼠相比较，被解剖的成功处理的动物的肝、脾、肠系膜组织、十二指肠、空肠和回肠的相对器官尺寸减小（图 23）。来自分析的 LAL 处理的大鼠的肝组织的组织病理学显示基本上正常的肝组织学，与安慰剂处理的动物的泡沫状巨噬细胞的大量积累形成鲜明对照（图 30）。

[0364] 实施例 18

[0365] 通过施用重组 LAL 治疗沃尔曼病 (WD)

[0366] 在 7 周龄时,由于自出生以来体重增长的困难和不良进展,女性患者被医院接收。在最初体格检查时,患者称重为 3.6kg(出生体重 3.7kg)并且非常瘦,皮肤松弛皱褶。腹部膨胀,具有 6cm 的坚硬肝肿大和约 4cm 的坚硬脾肿大。在腹股沟中淋巴结肿大非常明显并且肌肉活动非常弱。

[0367] 初血红蛋白水平为 9.2gm,血小板 506,000 以及白细胞 11,550。尿分析是正常的,骨髓穿刺涂片显示形成空泡的淋巴细胞和许多泡沫细胞。血清化学测量:总脂质 834mg/100ml,磷脂 176mg/100ml,甘油三酯 141mg/100ml,胆固醇 129mg/100ml,胆红素 0.3mg/100ml,碱性磷酸酶 9.0BU%,SGOT 90 个单位,SGPT 50 个单位,胆碱酯酶 20 个单位,尿素氮 8.3mg,空腹血糖(fasting sugar)45mg/100ml。腹部的 CT 扫描显示肝脾肿大和具有钙化的双侧对称肾上腺增大。

[0368] 通过手术向患者植入静脉血管入口(venous vascular access port)以便给药。在将入口连接至移动输注机器后,在 LAL 输注前 20 分钟用 1mg/kg 苯海拉明预处理患者以抵消潜在的过敏性输注反应。随后在 5 个小时的过程中通过静脉内输注以 1mg/kg 向患者施用 LAL。无限期地每 7 天重复该疗法一次。

[0369] 在施用 LAL 的第一剂量的 2 周内,患者经历体重增长的显著改善和关键腹部器官尺寸的正常化,如通过超声测定的。实验室结果显示 LAL 的输注恢复了患者的溶酶体酸性脂酶活性并且导致相关异常的校正。

[0370] 实施例 19

[0371] 通过施用重组 LAL 治疗胆固醇酯沉积病(CESD)

[0372] 一名 3 岁男孩由其儿科医生检查患有腹部皮疹瘙痒。在腹部检查后,医生注意到肝肿大,并且通过超声得到确认。此刻未进行诊断,定期监测患者。

[0373] 在 8 岁时,他因胃肠炎而被医院接收。肝脏活组织光学显微镜检查显示肝细胞中增加的胞质内糖原和小脂滴。电子显微镜显示具有小的电子致密颗粒的膜结合脂滴。产生了糖原贮积病 III 型(DeBrancher 病)的工作诊断,但皮肤成纤维细胞 Debrancher 活性是正常的。

[0374] 10 岁时,肝肿大持续并且进行第二次肝脏活组织检查。光学显微镜显示改变的肝实质的小叶结构,膨胀的肝细胞包含细胞质颗粒和空泡,具有轻度 portal 静脉纤维化(mild periportal fibrosis)。发现成纤维细胞酸性脂酶活性为正常的 7%,从而确认了 CESD 的诊断。总胆固醇(TC)、甘油三酯(TG)、低密度脂蛋白胆固醇(LDL-C)的血浆浓度各自分别高于与年龄和性别匹配的 7.51、3.24 和 5.58mmol/L 的第 95 百分位,然而血浆高密度脂蛋白胆固醇(HDL-C)低于 0.47mmol/L 的第 5 百分位;他患有混合性高脂血症(高胆固醇血症、高甘油三酯血症、低 α -脂蛋白血症和高 β -脂蛋白血症)。

[0375] 通过手术向患者植入静脉血管入口以便给药。在将入口连接至移动输注机器后,在 LAL 输注前 20 分钟用 5mg/kg 苯海拉明预处理患者以抵消潜在的过敏性输注反应。随后在 5 个小时的过程中通过静脉内输注以 5mg/kg 向患者施用 LAL。无限期地每 14 天重复该疗法一次。

[0376] 在施用 LAL 的第一剂量的 2 周内,患者经历体重增长的显著改善和关键器官尺寸的正常化,如通过超声测定的。实验室结果显示 LAL 的输注恢复了患者的溶酶体酸性脂酶活性并且导致相关异常的校正。

[0377] 实施例 20

[0378] 医学产品的描述和组成

[0379] 本文中描述的 LAL 的药品 ("SBC-102") 是从转基因原鸡属产生的卵清纯化的重组人溶酶体酸性脂酶 (rhLAL)。用于 SBC-102 的赋形剂与用于目前市场上用于溶酶体贮积症 (LSD) 的其它产品的赋形剂相似, 并且已被选择来维持药品的稳定性。

[0380] SBC-102 是在具有非天然胶乳 (丁基)FluroTec® - 包被的塞子和铝质钳口盖 (aluminum crimp seal) 的干净的 I 型硼硅玻璃小瓶中提供的澄清、无色、无菌液体。SBC-102 以水溶液形式提供, 该溶液由 SBC-102 (2mg/mL)、柠檬酸三钠二水合物 (13.7mg/mL, USP)、柠檬酸一水合物 (1.57mg/mL, USP)、人血清白蛋白 (10mg/mL, USP) 和注射用水 (至终体积, USP) 组成。SBC-102 的 pH 为 5.9 ± 0.2 。SBC-102 不包含防腐剂并且小瓶仅打算单次使用。

[0381] 表 5 SBC-102 (LAL) 中的赋形剂

[0382]

赋形剂	CAS 编号	等级	功能
柠檬酸三钠二水合物	6132-04-03	USP	缓冲液
柠檬酸一水合物	5949-29-1	USP	缓冲液
人血清白蛋白	70024-90-7	USP	稳定剂

[0383] 药品的组分

[0384] 表 6 SBC-102 的制剂

[0385]

组分	浓度
SBC-102 (rhLAL)	2mg/mL*
柠檬酸三钠二水合物	13.7mg/mL
柠檬酸一水合物	1.57mg/mL
人血清白蛋白	10mg/mL
注射用水, 加至	1.0mL

[0386] 通过解释本发明 (但非限制本发明) 来提供上述说明书中的每一个实施例。事实上, 对于本领域技术人员来说显而易见是可在不背离本发明的范围或精神的情况下进行各种修饰、组合、添加、缺失和变型。例如, 作为一个实施方案的部分举例说明或描述的特征可用于另一个实施方案中来产生另外的实施方案。本发明意欲涵盖此类修饰、组合、添加、缺失和变型。

[0387] 上述说明书中引用的所有文献 (例如, 美国专利、美国专利申请、公布) 通过引用并入本文。本发明的各种修饰和变型对于本领域技术人员来说是显而易见的, 而不背离本

发明的范围和精神。

[0388] 虽然已参考其示例性实施方案特别地显示和描述了本发明,但本领域技术人员将理解,可在不背离由所附权利要求包括的本发明的范围的情况下在其中进行许多形式和内容上的改变。

[0001]

序列表

<110> 辛那杰瓦生物制药股份有限公司
<120> 溶酶体贮积病酶
<130> SYN-079PCT
<140> PCT/US2011/033699
<141> 2011-04-23
<150> US 61/343,177
<151> 2010-04-23
<150> US 61/396,376
<151> 2010-05-26
<150> US 61/403,011
<151> 2010-09-09
<150> US 61/456,014
<151> 2010-10-29
<150> US 61/432,372
<151> 2011-01-13
<160> 20
<170> PatentIn version 3.5
<210> 1
<211> 399
<212> PRT
<213> 人
<400> 1
Met Lys Met Arg Phe Leu Gly Leu Val Val Cys Leu Val Leu Trp Thr
1 5 10 15
Leu His Ser Glu Gly Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu
20 25 30
Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr Trp Gly Phe Pro Ser
35 40 45
Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr Ile Leu Cys Leu Asn
50 55 60
Arg Ile Pro His Gly Arg Lys Asn His Ser Asp Lys Gly Pro Lys Pro
65 70 75 80
Val Val Phe Leu Gln His Gly Leu Leu Ala Asp Ser Ser Asn Trp Val
85 90 95
Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly
100 105 110
Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn Thr Trp Ser Arg Lys
115 120 125
His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe Trp Ala Phe Ser Tyr
130 135 140
Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser Ile Asn Phe Ile Leu
145 150 155 160

[0002]

Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val Gly His Ser Gln Gly
 165 170 175
 Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile Pro Glu Leu Ala Lys
 180 185 190
 Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val Ala Ser Val Ala Phe
 195 200 205
 Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu Pro Asp His Leu Ile
 210 215 220
 Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro Gln Ser Ala Phe Leu
 225 230 235 240
 Lys Trp Leu Gly Thr His Val Cys Thr His Val Ile Leu Lys Glu Leu
 245 250 255
 Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe Asn Glu Arg Asn Leu
 260 265 270
 Asn Met Ser Arg Val Asp Val Tyr Thr Thr His Ser Pro Ala Gly Thr
 275 280 285
 Ser Val Gln Asn Met Leu His Trp Ser Gln Ala Val Lys Phe Gln Lys
 290 295 300
 Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala Lys Asn Tyr Phe His Tyr
 305 310 315 320
 Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val Lys Asp Met Leu Val Pro
 325 330 335
 Thr Ala Val Trp Ser Gly Gly His Asp Trp Leu Ala Asp Val Tyr Asp
 340 345 350
 Val Asn Ile Leu Leu Thr Gln Ile Thr Asn Leu Val Phe His Glu Ser
 355 360 365
 Ile Pro Glu Trp Glu His Leu Asp Phe Ile Trp Gly Leu Asp Ala Pro
 370 375 380
 Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu Met Arg Lys Tyr Gln
 385 390 395
 <210> 2
 <211> 378
 <212> PRT
 <213> 人
 <400> 2
 Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu Thr Asn Met Asn Val
 1 5 10 15
 Ser Glu Ile Ile Ser Tyr Trp Gly Phe Pro Ser Glu Glu Tyr Leu Val
 20 25 30

[0003]

Glu	Thr	Glu	Asp	Gly	Tyr	Ile	Leu	Cys	Leu	Asn	Arg	Ile	Pro	His	Gly	35	40	45	
Arg	Lys	Asn	His	Ser	Asp	Lys	Gly	Pro	Lys	Pro	Val	Val	Phe	Leu	Gln	50	55	60	
His	Gly	Leu	Leu	Ala	Asp	Ser	Ser	Asn	Trp	Val	Thr	Asn	Leu	Ala	Asn	65	70	75	80
Ser	Ser	Leu	Gly	Phe	Ile	Leu	Ala	Asp	Ala	Gly	Phe	Asp	Val	Trp	Met	85	90	95	
Gly	Asn	Ser	Arg	Gly	Asn	Thr	Trp	Ser	Arg	Lys	His	Lys	Thr	Leu	Ser	100	105	110	
Val	Ser	Gln	Asp	Glu	Phe	Trp	Ala	Phe	Ser	Tyr	Asp	Glu	Met	Ala	Lys	115	120	125	
Tyr	Asp	Leu	Pro	Ala	Ser	Ile	Asn	Phe	Ile	Leu	Asn	Lys	Thr	Gly	Gln	130	135	140	
Glu	Gln	Val	Tyr	Tyr	Val	Gly	His	Ser	Gln	Gly	Thr	Thr	Ile	Gly	Phe	145	150	155	160
Ile	Ala	Phe	Ser	Gln	Ile	Pro	Glu	Leu	Ala	Lys	Arg	Ile	Lys	Met	Phe	165	170	175	
Phe	Ala	Leu	Gly	Pro	Val	Ala	Ser	Val	Ala	Phe	Cys	Thr	Ser	Pro	Met	180	185	190	
Ala	Lys	Leu	Gly	Arg	Leu	Pro	Asp	His	Leu	Ile	Lys	Asp	Leu	Phe	Gly	195	200	205	
Asp	Lys	Glu	Phe	Leu	Pro	Gln	Ser	Ala	Phe	Leu	Lys	Trp	Leu	Gly	Thr	210	215	220	
His	Val	Cys	Thr	His	Val	Ile	Leu	Lys	Glu	Leu	Cys	Gly	Asn	Leu	Cys	225	230	235	240
Phe	Leu	Leu	Cys	Gly	Phe	Asn	Glu	Arg	Asn	Leu	Asn	Met	Ser	Arg	Val	245	250	255	
Asp	Val	Tyr	Thr	Thr	His	Ser	Pro	Ala	Gly	Thr	Ser	Val	Gln	Asn	Met	260	265	270	
Leu	His	Trp	Ser	Gln	Ala	Val	Lys	Phe	Gln	Lys	Phe	Gln	Ala	Phe	Asp	275	280	285	
Trp	Gly	Ser	Ser	Ala	Lys	Asn	Tyr	Phe	His	Tyr	Asn	Gln	Ser	Tyr	Pro	290	295	300	
Pro	Thr	Tyr	Asn	Val	Lys	Asp	Met	Leu	Val	Pro	Thr	Ala	Val	Trp	Ser	305	310	315	320
Gly	Gly	His	Asp	Trp	Leu	Ala	Asp	Val	Tyr	Asp	Val	Asn	Ile	Leu	Leu	325	330	335	

[0004]

Thr Gln Ile Thr Asn Leu Val Phe His Glu Ser Ile Pro Glu Trp Glu
 340 345 350
 His Leu Asp Phe Ile Trp Gly Leu Asp Ala Pro Trp Arg Leu Tyr Asn
 355 360 365
 Lys Ile Ile Asn Leu Met Arg Lys Tyr Gln
 370 375
 <210> 3
 <211> 376
 <212> PRT
 <213> 人
 <400> 3
 Gly Lys Leu Thr Ala Val Asp Pro Glu Thr Asn Met Asn Val Ser Glu
 1 5 10 15
 Ile Ile Ser Tyr Trp Gly Phe Pro Ser Glu Glu Tyr Leu Val Glu Thr
 20 25 30
 Glu Asp Gly Tyr Ile Leu Cys Leu Asn Arg Ile Pro His Gly Arg Lys
 35 40 45
 Asn His Ser Asp Lys Gly Pro Lys Pro Val Val Phe Leu Gln His Gly
 50 55 60
 Leu Leu Ala Asp Ser Ser Asn Trp Val Thr Asn Leu Ala Asn Ser Ser
 65 70 75 80
 Leu Gly Phe Ile Leu Ala Asp Ala Gly Phe Asp Val Trp Met Gly Asn
 85 90 95
 Ser Arg Gly Asn Thr Trp Ser Arg Lys His Lys Thr Leu Ser Val Ser
 100 105 110
 Gln Asp Glu Phe Trp Ala Phe Ser Tyr Asp Glu Met Ala Lys Tyr Asp
 115 120 125
 Leu Pro Ala Ser Ile Asn Phe Ile Leu Asn Lys Thr Gly Gln Glu Gln
 130 135 140
 Val Tyr Tyr Val Gly His Ser Gln Gly Thr Thr Ile Gly Phe Ile Ala
 145 150 155 160
 Phe Ser Gln Ile Pro Glu Leu Ala Lys Arg Ile Lys Met Phe Phe Ala
 165 170 175
 Leu Gly Pro Val Ala Ser Val Ala Phe Cys Thr Ser Pro Met Ala Lys
 180 185 190
 Leu Gly Arg Leu Pro Asp His Leu Ile Lys Asp Leu Phe Gly Asp Lys
 195 200 205
 Glu Phe Leu Pro Gln Ser Ala Phe Leu Lys Trp Leu Gly Thr His Val
 210 215 220
 Cys Thr His Val Ile Leu Lys Glu Leu Cys Gly Asn Leu Cys Phe Leu

[0005]

225	230	235	240
Leu Cys Gly Phe	Asn Glu Arg Asn Leu	Asn Met Ser Arg Val	Asp Val
	245	250	255
Tyr Thr Thr His	Ser Pro Ala Gly Thr	Ser Val Gln Asn Met	Leu His
	260	265	270
Trp Ser Gln Ala	Val Lys Phe Gln Lys Phe	Gln Ala Phe Asp	Trp Gly
	275	280	285
Ser Ser Ala Lys	Asn Tyr Phe His Tyr	Asn Gln Ser Tyr	Pro Pro Thr
	290	295	300
Tyr Asn Val Lys	Asp Met Leu Val Pro Thr	Ala Val Trp Ser	Gly Gly
305	310	315	320
His Asp Trp Leu	Ala Asp Val Tyr Asp	Val Asn Ile Leu	Leu Thr Gln
	325	330	335
Ile Thr Asn Leu	Val Phe His Glu Ser	Ile Pro Glu Trp	Glu His Leu
	340	345	350
Asp Phe Ile Trp	Gly Leu Asp Ala Pro	Trp Arg Leu Tyr	Asn Lys Ile
	355	360	365
Ile Asn Leu Met	Arg Lys Tyr Gln		
370	375		
<210>	4		
<211>	373		
<212>	PRT		
<213>	人		
<400>	4		
Thr Ala Val Asp	Pro Glu Thr Asn Met	Asn Val Ser Glu	Ile Ile Ser
1	5	10	15
Tyr Trp Gly Phe	Pro Ser Glu Glu Tyr	Leu Val Glu Thr	Glu Asp Gly
	20	25	30
Tyr Ile Leu Cys	Leu Asn Arg Ile	Pro His Gly Arg	Lys Asn His Ser
	35	40	45
Asp Lys Gly Pro	Lys Pro Val Val Phe	Leu Gln His Gly	Leu Leu Ala
	50	55	60
Asp Ser Ser Asn	Trp Val Thr Asn Leu	Ala Asn Ser Ser	Leu Gly Phe
65	70	75	80
Ile Leu Ala Asp	Ala Gly Phe Asp	Val Trp Met Gly	Asn Ser Arg Gly
	85	90	95
Asn Thr Trp Ser	Arg Lys His Lys Thr	Leu Ser Val Ser	Gln Asp Glu
	100	105	110
Phe Trp Ala Phe	Ser Tyr Asp Glu	Met Ala Lys Tyr	Asp Leu Pro Ala
	115	120	125

[0006]

Ser Ile Asn Phe Ile Leu Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr
 130 135 140
 Val Gly His Ser Gln Gly Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln
 145 150 155 160
 Ile Pro Glu Leu Ala Lys Arg Ile Lys Met Phe Phe Ala Leu Gly Pro
 165 170 175
 Val Ala Ser Val Ala Phe Cys Thr Ser Pro Met Ala Lys Leu Gly Arg
 180 185 190
 Leu Pro Asp His Leu Ile Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu
 195 200 205
 Pro Gln Ser Ala Phe Leu Lys Trp Leu Gly Thr His Val Cys Thr His
 210 215 220
 Val Ile Leu Lys Glu Leu Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly
 225 230 235 240
 Phe Asn Glu Arg Asn Leu Asn Met Ser Arg Val Asp Val Tyr Thr Thr
 245 250 255
 His Ser Pro Ala Gly Thr Ser Val Gln Asn Met Leu His Trp Ser Gln
 260 265 270
 Ala Val Lys Phe Gln Lys Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala
 275 280 285
 Lys Asn Tyr Phe His Tyr Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val
 290 295 300
 Lys Asp Met Leu Val Pro Thr Ala Val Trp Ser Gly Gly His Asp Trp
 305 310 315 320
 Leu Ala Asp Val Tyr Asp Val Asn Ile Leu Leu Thr Gln Ile Thr Asn
 325 330 335
 Leu Val Phe His Glu Ser Ile Pro Glu Trp Glu His Leu Asp Phe Ile
 340 345 350
 Trp Gly Leu Asp Ala Pro Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu
 355 360 365
 Met Arg Lys Tyr Gln
 370

<210> 5

<211> 1150

<212> DNA

<213> 人

<400> 5

atgaaaatgc ggttcttggg gttggtggtc tgtttggttc tctggaccct gcattccgag 60

gggtccggag ggaaactgac agctgtggat cctgaaacaa acatgaatgt cagtgaatt 120

[0007]

atctcttact ggggattccc tagtgaggaa tacctagttg agacagaaga tggatatatt	180
ctgtgcctta accgaattcc tcatggggagg aagaaccatt ctgacaaagg tcccaaacca	240
gttgctttcc tgcaacatgg cttgctggca gattctagta actgggtcac aaaccttgcc	300
aacagcagcc tgggcttcat tcttgctgat gctggttttg acgtgtggat gggcaacagc	360
agaggaaata cctggctcgc gaaacataag acactctcag ttctcagga tgaattctgg	420
gctttcagtt atgatgagat ggcaaaatat gacctaccag ctccattaa cttcattctg	480
aataagactg gccaaagaaca agtgtattat gtgggtcatt ctcaaggcac cactataggt	540
tttatagcat ttccacagat cctgagctg gctaaaagga ttaaatggt ttttgcctg	600
ggctcgtggt cttccgtcgc cttctgtact agccctatgg ccaaactggg acgactgcc	660
gatcatctca ttaaggacct ctttgagac aaagaatttc tccccagag tgcgtttttg	720
aagtggctgg gtacccacgt ttgcactcat gtcatactga aggagctctg tggaaatctc	780
tgttttcttc tgtgtggatt taatgagaga aatttaaata tgtctagagt ggatgtgtat	840
acaacacatt ctctgctgg aacttctgtg caaaacatgt tacactggag ccaggctgtt	900
aaattccaaa agtttcaagc ctttgactgg ggaagcagt ccaagaatta ttttcattac	960
aaccagagtt atcctccac atacaatgtg aaggacatgc ttgtgccgac tgcagtctgg	1020
agcgggggtc acgactggct tgcagatgac tacgacgtca atatcttact gactcagatc	1080
accaacttgg tgttccatga gagcattccg gaatgggagc atcttgactt catttggggc	1140
ctggatgccc	1150

<210> 6
 <211> 10882
 <212> DNA
 <213> 人工序列

<220>
 <223> pALVIN-OVR1-I-SBC102-dSA

<400> 6	
ctttcccct caagctctaa atcgggggct ccttttaggg ttccgattta gtgetttacg	60
gcacctcgac cccaaaaaac ttgattaggg tgatggttca cgtagtgggc catcgccctg	120
atagacggtt tttcgccctt tgacgttgga gtccacgttc tttaatagtg gactcttggt	180
ccaaactgga acaacactca accctatctc ggtctattct ttgatttat aagggtttt	240
gccgatttcg gcctatttgt taaaaatga gctgatttaa caaaaattta acgcgaattt	300
taacaaaata ttaacgctta caatttccat tcgccattca ggctgcgcaa ctgttgggaa	360
gggcgatcgg tgcgggcctc ttgcttatta cgccagctgg cgaaaggggg atgtgctgca	420
aggcgattaa gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc	480
agtgagcgcg tattccctaa cgtacacgtc ggggtcacca aatgaagcct tcgtcttcat	540
gcatgtgctc gtagtcgtca gggaatcaac ggtccggcca tcaaccagg tgcacaccaa	600
tgtggtgaat ggtcaaatgg cgtttattgt atcgagctag gcacttaaat acaatatctc	660
tgcaatggg aattcagtgg ttcttccaat ccgtccccct cctatgcaa aagcgaaact	720
actatactct gaggggactc ctaaccgcgt acaaccgaag ccccgctttt cgcctaaaca	780
tgctattgtc cctcagtcac agccttgccc gttacaacc gattcgcaag ctttgcctc	840
cccacattat ccgttagcatt atttcttagc agtcacaga gctacagaag atactctatg	900

[0008]

ctgtagccaa gcttacaagt ttactattca gcgacctcct atattccgcg tgccagccga	960
tcaattacca atccaaccag ctatcacacg gaatacaaga actcgccctac gctcttcttt	1020
cgggctgctt ataagcctcc tgtaatTTTT ttatatctct cgctcgagtc tcttcagaaat	1080
ggcacagcac cgttcagaaa aatgccagg tggactatga actcacatcc aaaggagctt	1140
gacctgatac ctgattttct tcaaacaggg gaaacaacac aatcccaaa zacagctcag	1200
agagaaacca tcactgatgg ctacagcacc aaggtatgca atggcaatcc attcgacatt	1260
catctgtgac ctgagcaaaa tgattttatct ctccatgaat ggttgcttct tccctcatg	1320
aaaaggcaat tcccacactc acaatatgca acaaagacaa acagagaaca attaatgtgc	1380
tccttcctaa tgttaaaatt gtagtggcaa agaggagaac aaaatctcaa gttctgagta	1440
ggttttagtg attggataag aggctttgac ctgtgagctc acctggactt catatccttt	1500
tggataaaaa gtgcttttat aactttcagg tctccgagtc tttattcatg agactgttgg	1560
tttagggaca gaccacaat gaaatgcctg gcataggaaa gggcagcaga gccttagctg	1620
accttttctt gggacaagca ttgtcaaaca atgtgtgaca aaactatttg tactgctttg	1680
cacagctgtg ctgggcaggg caatccattg ccacctatcc caggtaacct tccaactgca	1740
agaagattgt tgccttactct ctctagacct ccaagtcaaa ccaactatgc aggtatgctg	1800
acaacactat gatgacagcc tgttctgac aagatctcat ttgttcatgg acaatttttg	1860
ttgcttgacg ctggctctcc attgggaaag agtgtagtat atccttctca tctgacagaa	1920
aagcagaaat tctcatgctc cacacttaat ctacattgtt ttaaaccacc ggctacttct	1980
tggagaggaa aatggcttt tataagactc acaaacaaca gctctgcaag tcaaatgcat	2040
acaaaactgt tctgtaggtc tggatcagg acactatgtg gaagtcaaat agagcagctt	2100
taaaaagcct ttgggatcat tctcatctta tatttgacg acgatactat gacagtgata	2160
actgacataa ctgcatcaat ttccttgata ttttatttgt cttaaagtac aagacataga	2220
gatggacgta aagatggaca tatgactcag gtctggacag gtccgtggtc catgtatgat	2280
aaaagagatg aagggaagga gaattgagac tgtctaagaa gggttcagg gacgttctga	2340
aggcagattt gactgaatca gatgtactgt ccaagtctca tatgtagcaa tggaaggctg	2400
atattggaga aatataaaga aatggctgtg aactcaaagt gaccctgaac agaaaaggga	2460
tatggagtta aaataatgtc acagaactga ggtttatatg atataccatg ggctgcagag	2520
ggtcagagtg ctccaccatg ggctctctt gggtgcagg gaacttctgt tctacacctg	2580
gaacacctcc tgcctcctc cgcactgacc tcagtgtcat cagggtgtt tctctcacat	2640
tttctcactc acctctccca actaccattg tacagcagtt gttcttcat attgctcctc	2700
ctgaggtaaa tctagcatcg ttaagtctc agacttgga aggagaatgt agatttccac	2760
agtatatatg tttcacaaa aggaaggaga gaaacaaaag aaaatggcac tgaclaaact	2820
tcagctagtg gtataggaaa gtaattctgc ttaacagaga ttgcagtgat ctctatgtat	2880
gtcctgaaga attatgttgt acttttttcc cccattttta aatcaaacag tgccttacag	2940
aggtcagaat ggtttcttta ctgtttgtca attctattat ttcaatacag aacaatagct	3000
tctataactg aaatatattt gctatgtat attatgattg tccctcgaac catgaacact	3060
cctccagctg aatttcacaa ttcctctgtc atctgccagg ccattaagtt attcatggaa	3120
gatctttgag gaacactgca agttcatatc ataaacacat ttgaaattga gtattgtttt	3180

[0009]

gcattgtaig gagctatggt ttgctgtatc ctcagaataa aagtttgtta taaagcattc	3240
acacccataa aaagatagat ttaaataatc caactatagg aaagaaagtg tgtctgtctc	3300
tcactctagt ctgagttggc tccttcacat gcacgcttct ttattttctc tattttgtca	3360
agaaaataat aggtaagtc ttgtttctcat ttatgtcctg tctagcgtgg ctcagatgca	3420
cattgtacat acaagaagga tcaaatgaaa cagacttctg gtcgtttact acaaccatag	3480
taataagcac actaactaat aattgctaata tatgttttcc atctccaagg ttcccacatt	3540
tttctgtttt cttaaagatc ccattatctg gttgtaactg aagctcaatg gaacatgagc	3600
aatatttccc agtcttctct cccatccaac agtcctgatg gattagcaga acaggcagaa	3660
aacacattgt taccagaat taaaaactaa tatttgcctc ccattcaatc caaaatggac	3720
ctattgaaac taaaatctaa cccaatccca ttaaatagatt tctatggtgt caaagggtcaa	3780
acttctgaag ggaacctgtg ggtgggtcac aattcagact atatattccc cagggtcag	3840
ccagtgtctg tacatacagc tagaaagctg tattgccttt agcagtcaag ctcgaaaggt	3900
aagcaactct ctggaattac cttctctcta tattagctct tacttgacc taaacttta	3960
aaaattaaca attattgtgc tatgtgtgt atctttaagg gtgaagtacc tgcgtgatac	4020
cccctataaa aacttctcac ctgtgtatgc attctgcaat attttattat gtgtaaaagc	4080
tttgtgtttg ttttcaggag gcttattctt tgtgcttaaa atatgttttt aatttcagaa	4140
catcttatcc tgcgttcac tatctgatat gctttgcagt ttgcttgatt aacttctagc	4200
cctacagagt gcacagagag caaaatcatg gtgttcagt aattctgggg agttatttta	4260
atgtgaaaat tctctagaag tttaattcct gcaaagtga gctgctgac actacacaag	4320
ataaaaatgt ggggggtgca taaacgtata ttcttacaat aatagatata tgtgaactta	4380
tatacagaaa agaaaatgag aaaaatgtgt gtgtgtatac tcacacacgt ggtcagtaaa	4440
aacttttgag gggtttaata cagaaaatcc aatcctgagg cccagcact cagtacgcat	4500
ataaagggtt gggtcttgaa ggaactctga ctttcacaga ttatataaat ctcaggaaag	4560
caactagatt catgctggct ccaaaagctg tgctttatat aagcacactg gctatacaat	4620
agttgtacag ttcagctctt tataatagaa acagacagaa caagtataaa tcttctattg	4680
gtctatgtca tgaacaagaa ttcattcagt ggctctgttt tatagtaaac attgctattt	4740
tatcatgtct gcatttctct tctgtctgaa tgtcaccact aaaatttaac tccacagaaa	4800
gtttatacta cagtacacat gcatacttt gagcaaagca aaccatacct gaaagtgcaa	4860
tagagcagaa tatgaattac atgctgtctt ttctctaga ctacatgacc ccatataaat	4920
tacattcctt atctattctg ccatacccaa acaaaagga aaaatacttt tgaagatcta	4980
ctcatagcaa gtagtgtgca acaaacagat atttctctac atttattttt agggaataaa	5040
aataagaaat aaaatagtca gcaagcctct gctttctcat atatctgtcc aaacctaaag	5100
tttactgaaa tttgtctttt gaatttccag ttgtgcaagc ctatcagatt gtgtttta	5160
cagaggtact gaaaagtatc aatgaattct agctttcact gaacaaaaat atgtagaggc	5220
aactggcttc tgggacagtt tgctacccaa aagacaactg aatgcaaata cataaataga	5280
tttatgaata tggttttgaa catgcacatg agagggtgat atagcaacag acacattacc	5340
acagaattac tttaaaacta cttgttaaca tttaattgcc taaaaactgc tcgtaattta	5400
ctgttgtagc ctacataga gtacctgca tgggtactatg tacagcattc catccttaca	5460

[0010]

ttttactgt tctgctgttt gctctagaca actcagagtt caccatgaaa atgcggttct	5520
tggggtttgt ggtctgtttg gttctctgga cctgcattc cgaggggtcc ggagggaaac	5580
tgacagctgt ggatccagaa acaaacatga atgtcagta aattatctct tactggggat	5640
tccttagtga ggaataccta gttgagacag aagatggata tattctgtgc cttaacggaa	5700
ttcctcatgg gaggaagaac cattctgaca aaggatccaa accagttgtc ttcctgcaac	5760
atggcttgct ggcagattct aglaactggg tcacaaacct tgccaacagc agcctgggct	5820
tcattcttgc tgatgctggt tttgacgtgt ggatgggcaa cagcagagga aatacctggt	5880
ctcggaacaa taagacactc tcagtttctc aggatgaatt ctgggcttgc agttatgatg	5940
agatggcaaa atatgaccta ccagcttcca ttaacttcat tctgaataag actggccaag	6000
aacaagtgtg ttatgtgggt cattctcaag gcaccactat aggttttata gcattttcac	6060
agatccctga gctggctaaa aggattaaaa tgttttttgc cctgggtcct gtggcttccg	6120
tcgcctcttg tactagccct atggccaaac tgggacgact gccagatcat ctcatgaagg	6180
acctctttgg agacaaagaa tttcttcccc agagtgcgtt tttgaagtgg ctgggtaccc	6240
acgtttgcac tcatgicaca ctgaaggagc tctgtggaaa tctctgtttt cttctgtgtg	6300
gatttaatga gagaaattta aatatgtcta gagtggatgt gtatacaaca cattctcttg	6360
ctggaacttc tgtgcaaaac atgttacact ggagccaggc tgttaaattc caaaagtttc	6420
aagcctttga ctggggaagc agtgccaaga attattttca ttacaaccag agttatectc	6480
ccacatacaa tgtgaaggac atgcttgtgc cgactgcagt ctggagcggg ggtcacgact	6540
ggcttgacga tgtctacgac gtcaatatct tactgactca gatcaccaac ttggtgttcc	6600
atgagagcat tccggaatgg gagcatcttg acttcatttg gggcctggat gcccttggga	6660
ggctttataa taagattatt aatctaataa ggaaatatca gtgattcgaa gcggccgcaa	6720
gaagaaagct gaaaaactct gtcccttcca acaagaccca gagcactgta gtatcagggg	6780
taaaatgaaa agtatgttat ctgctgcac cagacttcat aaaagctgga gcttaatcta	6840
gaaaaaaat cagaaagaaa ttacactgtg agaacagggt caattcattt ttcctttaca	6900
cagagtaata ctggaactc atggatgaag gcttaaggga atgaaattgg actcacagta	6960
ctgagtcac acactgaaaa atgcaacctg atacatcagc agaaggttta tgggggaaaa	7020
atgcagcctt ccaattaagc cagatatctg tatgaccaag ctgctccaga attagtcact	7080
caaaatctct cagattaaat tatcaactgt caccaacctt tcctatgctg acaaggcaat	7140
tgtgtgttct ctgtgttctt gatactacaa ggctcttctt gacttcttaa agatgcatta	7200
taaaaatctt ataattcaca ttcttcccta aactttgact caatcatggt atgttgcaa	7260
atatggtata ttactattca aattgttttc cttgtacca tatgtaatgg gtcttgtgaa	7320
tgtgtctttt tgttcttita atcataataa aaacatgttt aagcaaacac ttttacttgc	7380
tagtatttga aggtaccgga tctcgagccg ctttcaatgc ccccaaaacc aatccccagg	7440
tttttaactc tcccgatatt ccaagtacca tagcccgtg agagagcgcc gcggtaatgg	7500
gatccaggga ccccggggaa tataagtctg agggggacgt aagcaacctt tccttttgta	7560
acagggacaa catagccctt atttcttctt tagaaggaga ggttttcccg caataggtct	7620
tacacgcgga cgaaatcacc ttatgacgg ctcccatgct tgatccaccg ggcgaccgga	7680
atcacgcaga gcaaccggaa tcacgcctgg ggtggaccgc tcagtcgtcg ggcttcttcc	7740

[0011]

ccgtcttcca acgactctct gagttctcgg tagggatatgt tggcccccctg cagtagggct	7800
ccctccgacg ccactcagct tctgccctcc taagccgcag cccctctac tagggcacc	7860
gtccgcctccc cgaataagcg agacggatga ggacaggatc gccacgccgc ctgtggccga	7920
ccactattcc ctaacgatca cgtcggggtc accaaatgaa gccttctgct tcatgcatgt	7980
gctcgtagtc gtcaggaat caacggctcg gccatcaacc caggtgcaca ccaatgtggt	8040
gaatggtcaa atggcggtta ttgtatcgag ctaggcactt aaatacaata tctctgcaat	8100
gcggaattca gtggttcgtc caatccgtgt tagaccctgc tgttgccctc ctaacaaggc	8160
acgatcalac cacgatcata ccaccttact cccaccaatc ggcatgcacg gtgctttttc	8220
tctccttata aggcatgttg ctaactcacc gttacataag catgttgcaa gactacaaga	8280
gtattgcata agactacatt tccccctccc tatgcaaaag cgaaactact atatcctgag	8340
gggactccta accgcgtaca accgaagccc cgtttttcgc ctaaacatgc tattgtcccc	8400
tcagtcaagc cttgcccggtt acaaccgatc tcgcaagcct tgcctctccc acattatccg	8460
tagcattatt tcctagcagt catcagagct acagaagata ctctatgctg tagccaagtc	8520
tacaagttta ctattcagcg acctcctata ttccgcgtgc cagccgatca attaccaatg	8580
cgcgcttgcc gtaatcatgg tcatagctgt ttctgtgtg aaattgttat ccgctcacia	8640
ttccacacia catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga	8700
gctaactcac attaatgtcg ttgcgtcac lgcccgttt ccagtcggga aacctgtcgt	8760
gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgt	8820
cttcgccttc ctcgctcact gactcgtgc gctcggctgt tcggctgcgg cgagcggat	8880
cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga	8940
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt	9000
ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgtca agtcagaggt	9060
ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc	9120
gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc cttcgggaa	9180
gcgtggcgct ttctcatagc tcacgtgtga ggtatctcag ttcgggttag gtcgttcgt	9240
ccaagctggg ctgtgtgcac gaacccccg ttccagccga ccgctgcgcc ttatccggta	9300
actatcgtct tgagtccaac ccggttaagc acgacttacc gccactggca gcagccactg	9360
gtaacaggat tagcagagcg aggtatgtag gcggtgtac agagtcttg aagtgtggc	9420
ctaactacgg ctacactaga aggacagtat ttggtatctg cgtctgtctg aagccagtta	9480
ccttcggaaa aagagttggt agctcttgat ccggcaaaaa aaccaccgct ggtagcggtg	9540
gtttttttgt ttgcaagcag cagattacgc gcagaaaaa aggatctcaa gaagatcctt	9600
tgatcttttc tacggggctt gacgctcagt ggaacgaaaa ctcacgttaa gggattttg	9660
tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta	9720
aatcaatcta aagtatatat gagtaaacct ggtctgacag ttaccaatgc ttaatcagt	9780
aggcacctat ctcagcgatc tgtctatttc gtctatccat agttgcctga ctccccgtcg	9840
tgtagataac tacgatacgg gagggcttac catctggccc cagtgtgca atgataccgc	9900
gagaccacg ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaaggccg	9960
agcgcagaag tggctcctgca actttatccg cctccatcca gtctattaat tgttgccggg	10020

[0012]

```

aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgtacag 10080
gcatcgttgg gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat 10140
caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggtagactcc ttcggtcctc 10200
cgatcgttgt cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc 10260
ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtgactggg gactactcaa 10320
ccaagtcatc ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac 10380
gggataatac cgcgccacat agcagaactt taaaagtgtc catcatigga aaacgttctt 10440
cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg taaccactc 10500
gtgaccccaa ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa 10560
caggaagcca aaatgccgca aaaaagggaa taaggcgac acggaaatgt tgaatactca 10620
tactcttctt tttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat 10680
acatatattg atgtatttag aaaaataaac aaataggggt tccgcgcaca tttcccga 10740
aagtgccacc tgacgcgcc tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc 10800
gcagcgtgac cgctacactt gccagcgccc tagcggccgc tcctttcgct ttcttcctt 10860
cctttctcgc cacgttcgcc gg 10882

```

<210> 7
 <211> 7780
 <212> DNA
 <213> 人工序列

<220>
 <223> pALVIN-OVR1-I-SBC102-dSA的原病毒序列

```

<400> 7
aatgaagcct tctgttcat gcatgtgctc gtagtcgtca gggaatcaac ggtccggcca 60
tcaaccagg tgacaccaa tgtggatgaat ggtcaaatgg cgtttattgt atcgagctag 120
gcacttaaat acaatctc tcgaatgcgg aattcagttg ttcgtccaat ccgtccccc 180
ccctatgcaa aagcgaaact actatctctt gaggggactc ctaaccgctt acaaccgaag 240
ccccgtttt cgcctaaaca tgcattgttc cctcagtc aagccttgcg gttacaaccc 300
gattcgcaag ccttgcctc cccacattat ccgtagcatt atttcctagc agtcacaga 360
gctacagaag atactctatg ctgtagccaa gtctacaagt ttactattca gcgacctct 420
atattccgcg tgcagccga tcaattacca atccaaccag ctatcacacg gaatacaaga 480
actcgctac gctctctttt cgggctgctt ataagcctcc tgtaattttt ttataattct 540
cgctcgagtc tcttcagaat ggcacagcac cgtgcagaa aaatgccagg tggactatga 600
actcacatcc aaaggagctt gacctgatac ctgattttct tcaaacaggg gaaacaacac 660
aatcccacaa aacagctcag agagaaacca tcaatgatgg ctacagcacc aaggtatgca 720
atggcaatcc attcgacatt catctgtgac ctgagcaaaa tgatttatct ctccatgaat 780
ggttgcttct ttcctcatg aaaaggcaat ttccacactc acaatatgca acaaagacaa 840
acagagaaca attaatgtgc tcttccctaa tgtttaaatt gtagtggcaa agaggagaac 900
aaaatctcaa gttctgagta ggttttagtg attggataag aggccttgac ctgtgagctc 960
acctggactt catatccttt tggataaaaa gtgcttttat aactttcagg tctccgagtc 1020
tttattcatg agactgttgg tttagggaca gaccacaaat gaaatgcctg gcataggaaa 1080

```

[0013]

gggcagcaga gccttagctg accttttctt gggacaagca ttgtcaaaca atgtgtgaca	1140
aaactatttg tactgtcttg cacagctgtg ctgggcaggg caatccattg ccacctatcc	1200
caggtaacct tccaactgca agaagattgt tgcttactct ctctagacce ccaagtcaaa	1260
ccaactatgc aggtatgctg acaacactat gatgacagcc tgttctgac aagatctcat	1320
ttgttcatgg acaatttttg ttgcttgacg ctggctctcc attgggaaag agttagtlat	1380
atccttctca tctgacagaa aagcagaaat tctcatgctc cacacttaat ctacattgtt	1440
ttaaaccacc ggctacttct tggagaggaa aaatggcttt tataagactc acaaaacaaa	1500
gctctgcaag tcaaatgcat acaaaactgt tctgtaggtc tggaatcagg aactatgtg	1560
gaagtcaaat agagcagctt taaaaagcct ttgggatcat tctcatctta tatttcagc	1620
acgatactat gacagtata actgacataa ctgcatcaat ttccttgata ttttatttgt	1680
cttaaagtac aagacataga gatggacgta aagatggaca tatgactcag gtctggacag	1740
gtccglggtc catgtatgat aaaagagatg aagggaagga gaattgagac tgtctaagaa	1800
gggcttcagg gacgttctga aggcagattt gactgaatca gatgtactgt ccaagtctca	1860
tatgtagcaa tggaaggctg atattggaga aatataaaga aatggctgtg aactcaaagt	1920
gacctgaac agaaaaggga tatggagtta aaataatgtc acagaactga ggtttatatg	1980
atataccatg ggctgcagag ggtcagatg ctccaccatg ggctctctt gggtgcagg	2040
gaacttctgt tctacacctg gaacacctcc tgccctctc cgcactgacc tcagtgtcat	2100
cagggtgtt tctctcatat tttctcactc acctctcca actaccattg tacagcagtt	2160
gttcttcatat attgtctctc ctgaggtaca tctagcatcg ttaagtctc agacttgcca	2220
aggagaatgt agatttccac agtatatatg tttcacaaa aggaaggaga gaaacaaaag	2280
aaaatggcac tgactaaact tcagctagtg gtataggaaa gtaattctgc ttaacagaga	2340
ttgcagtgal cctctatgtat gtctgaaga attatgttgt acttttttcc cccattttta	2400
aatcaaacag tgccttacag aggtcagaat gggttcttta ctgtttgtca attctattat	2460
ttcaatacag aacaatagct tctataactg aaatatattt gctattgtat attatgattg	2520
tcctctgaac catgaacact cctccagctg aatttcacaa ttcctctgtc atctgccagg	2580
ccattaagtt attcatggaa gatctttgag gaacactgca agttcatatc ataaacacat	2640
ttgaaattga gtattgtlll gcattgtatg gagctatgtt ttgctgtatc ctgagaataa	2700
aagtttgta taaagcattc acaccataa aaagatagat ttaaataatc caactatagg	2760
aaagaaagtg tgtctgtctt tcaacttagt ctgagttggc tccttcacat gcacgcttct	2820
ttatttctcc tattttgtca agaaaataat aggtcaagtc ttgttctcat ttatgtctg	2880
tctagcgtgg ctgagtga cattgtacat acaagaagga tcaaatgaaa cagacttctg	2940
gtctgttact acaaccatag taataagcac actaactaat aattgctaatt tatgttttcc	3000
atctccaagg ttcccacatt tttctgtttt cttaaagatc ccattatctg gttgtaactg	3060
aagctcaatg gaacatgagc aatatttccc agtcttctct cccatccaac agtcctgatg	3120
gattagcaga acaggcagaa aacacattgt taccagaat taaaaactaa tatttgctct	3180
ccattcaatc caaaatggac ctattgaaac taaaactaa ccaatccca ttaaagatt	3240
tctatggtgt caaaggtcaa acttctgaag ggaacctgtg ggtgggtcac aattcagact	3300
atatattccc cagggtcag ccagtgtctg tacatacagc tagaaagctg tattgccttt	3360

[0014]

agcagtc	caag	ctcgaaaggt	aagcaactct	ctggaattac	cttctctcta	tattagctct	3420
tacttg	cacc	taaacttta	aaaattaaca	allattgtgc	tatgtgtgt	atctttaagg	3480
gtgaagt	tacc	lgcgtgatac	ccccataaaa	aacttctcac	ctgtgtatgc	attctgcact	3540
attttatt	at	gtgtaaaagc	tttggtttg	ttttcaggag	gcttattctt	tgtgcttaaa	3600
atatgtttt	aat	ttcagaa	catcttatcc	tgtcgttcac	tatctgatat	gctttgcagt	3660
ttgcttgatt	aacttctagc	cctacagagt	gcacagagag	caaaatcatg	gtgttcagt		3720
aattctgggg	agttatttta	atgtgaaaat	tctctagaag	tttaattcct	gcaaagtga		3780
gctgctg	atc	actacacaag	ataaaaaigt	ggggggtgca	taaactgata	ttcttacaat	3840
aatagataca	tgtgaactta	tatacagaaa	agaaaatgag	aaaaatgtgt	gtgtgtatac		3900
tcacacacgt	ggtcagtaaa	aacttttgag	gggtttaata	cagaaaaatcc	aatcctgagg		3960
ccccagcact	cagtacgcat	ataaagggt	gggctctgaa	ggacttctga	ctttcacaga		4020
ttatataaat	ctcaggaaag	caactagatt	catgctggct	ccaaaagctg	tgctttatat		4080
aagcacactg	gctatacaat	agttgtacag	ttcagctctt	tataatagaa	acagacagaa		4140
caagtataaa	tcttctattg	gtctatgtca	tgaacaagaa	ttcattcagt	ggctctgttt		4200
tatagtaaac	attgctat	ttt	catgtct	gcatttctct	tctgtctgaa	tgtcaccact	4260
aaaatttaac	tccacagaaa	gtttatacta	cagtacacat	gcatactttt	gagcaaagca		4320
aaccatacct	gaaagtga	tagagcagaa	tatgaattac	atgcgtgtct	ttctcctaga		4380
ctacatgacc	ccatataaat	tacattcctt	atctattctg	ccalcaccaa	aacaaaggta		4440
aaaatacttt	tgaagalcia	ctcatagcaa	gtagtgtgca	acaaacagat	atttctctac		4500
attttatttt	aggaataaaa	aataagaaat	aaaatagtca	gcaagcctct	gctttctcat		4560
atatctgtcc	aaacctaaag	tttactgaaa	tttgcctttt	gaatttcag	ttttgcaagc		4620
ctatcagatt	gtgtttta	aat	cagaggtact	gaaaagtatc	aatgaattct	agctttcact	4680
gaacaaaaat	atgtagaggc	aactggcttc	tgggacagtt	tgtacccaa	aagacaactg		4740
aatgcaata	cataaataga	tttatgaata	tggttttgaa	catgcacatg	agagggtgat		4800
atagcaacag	acacattacc	acagaattac	tttaaaacta	cttgtaaca	tttaattgcc		4860
taaaaactgc	tcgtaattta	ctgttgtagc	ctacataga	gtaccctgca	tgttactatg		4920
tacagcattc	catccttaca	ttttcactgt	tctgtgtttt	gctctagaca	actcagagtt		4980
caccatgaaa	atgcggttct	tggggttggt	ggtctgtttg	gttctctgga	ccctgcattc		5040
caggggtcc	ggagggaac	tgacagctgt	ggatccagaa	acaaacatga	atgtcagtga		5100
aattatctct	tactggggat	tccttagtga	ggaataccta	gttgagacag	aagatggata		5160
tattctgtgc	cttaaccgaa	ttcctcatgg	gaggaagaac	cattctgaca	aaggtcccaa		5220
accagttgtc	ttcctgcaac	atggcttgct	ggcagattct	agtaactggg	tcacaaacct		5280
tgtccaacagc	agcctgggct	tcattcttgc	tgatgtggt	tttgacgtgt	ggatgggcaa		5340
cagcagagga	aatacctggt	ctcgaaaca	taagacactc	tcagtlctc	aggatgaatt		5400
ctgggcttcc	agllaigatg	agatggcaaa	atatgacctc	ccagcttcca	ttaacttcat		5460
tctgaataag	actggccaag	aacaagtgtg	ttatgtgggt	cattctcaag	gcaccactat		5520
aggttttata	gcattttcac	agatccctga	gctggctaaa	aggattaaaa	tgttttttgc		5580
cctgggtcct	gtggcttccg	tcgcttctg	tactagccct	atggccaac	tgggacgact		5640

[0015]

```

gccagatcat ctcattaagg acctctttgg agacaaagaa ttcttcccc agagtgcgtt 5700
tttgaagtgg ctgggtaccc acgtttgcac lcatgtcata ctgaaggagc tctgtggaaa 5760
tctctgtttt cttctgtgtg gatttaatga gagaaattta aatatgtcta gagtggatgt 5820
gtatacaaca cattctcctg ctggaacttc tgtgcaaaac atgttacact ggagccagge 5880
tgttaaattc caaaagtffc aagcctttga ctggggaagc agtgccaaga attattttca 5940
ttacaaccag agttatcctc ccacatacaa tgtgaaggac atgcttgtgc cgactgcagt 6000
ctggagcggg ggtcacgact ggcttcgaga tgtctacgac gtcaatatct tactgactca 6060
gatcaccaac ttggtgttcc atgagagcat tccggaatgg gagcatcttg acttcatttg 6120
gggcctggat gcccttggga ggctttataa taagattatt aatctaataa ggaaatatca 6180
gtgattcgaa gcggccgcaa gaagaaagct gaaaaactct gtcccttcca acaagaccca 6240
gagcactgta gtatcagggg taaaaatgaa agtatgttat ctgctgcac cagacttcat 6300
aaaagctgga gcttaatcta gaaaaaaaat cagaaagaaa ttacactgtg agaacaggtg 6360
caattcactt ttcctttaca cagagtaata ctggtaactc atggatgaag gcttaaggga 6420
atgaaattgg actcacagta ctgagtcac acactgaaaa atgcaacctg atacatcagc 6480
agaaggttta tgggggaaaa atgcagcctt ccaattaagc cagatatctg tatgaccaag 6540
ctgctccaga attagtact caaatctct cagattaaat tatcaactgt caccaacct 6600
tcttatgctg acaaggcaat tgcttgttct ctgtgttctt gatactaca ggctcttct 6660
gacttcctaa agatgcatta taaaaatctt ataattcaca ttcttcccta aactttgact 6720
caatcatggt atgttgcaa atatggtata ttactattca aattgttttc cttgtaccca 6780
tatgtaatgg gtcttgtgaa tgtgctcttt tgttccctta atcataataa aaacatgttt 6840
aagcaaacac ttttacttgg tagtatttga aggtaccgga tctcagccg cttcaatgc 6900
cccaaaaacc aatccccagg tttttaactc tcccatttt ccaagtacca tagcccgtg 6960
agagagcgcc gcggaatgg gatccagga cccgggggaa tataagtcig agggggacgt 7020
aagcaaccct tccttttgta acagggacaa catagccctt atttcttct tagaaggaga 7080
ggttttcccg caataggctt tacacggga cgaatcacc tttatgacgg cttccatgt 7140
tgatccaccg ggcgaccgga atcacgcaga gcaaccggaa tcacgccitg ggtggaccgc 7200
tcagtcgtcg ggcttcttc ccgtcttcca acgactctct gatttctcgg tagggtatgt 7260
tggccccctg cagtagggtt cctccgacg ccactcagct tctgccctcc taagccgcag 7320
ccccctctac tagggtcac gtccgctccc cgaataagcg agacggatga ggacaggatc 7380
gccacgccgc ctgtggccga ccactattcc ctaacgatca cgtcggggtc accaaatgaa 7440
gccttctgct tcatgcatgt gctcgtagt gtcagggaat caacgggtcc gccatcaacc 7500
caggtgcaca ccaatgtggt gaatgggtcaa atggcggtta ttgtalcgag ctaggcactt 7560
aaatacaata tctctgcaat gcggaattca gtggttcgtc caatccgtgt tagaccgctc 7620
tgttgcttc ctaacaagc acgatcatac cagatcata ccaccttact cccaccaatc 7680
ggcatgcacg gtgcttttct tctcttata aggcattgtg ctaactcacc gttacataag 7740
catgttgcaa gactacaaga gtattgcata agactacatt 7780

```

<210> 8
 <211> 10762
 <212> DNA

[0016]

nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	1980
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2040
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2100
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2160
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2220
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2280
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2340
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2400
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2460
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2520
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2580
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2640
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	2700
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnttcga	agcggccgca	agaagaaagc	2760
tgaaaaactc	tgtcccttcc	aacaagacc	agagcactgt	agtatcaggg	gtaaaatgaa	2820
aagtatgtta	tctgtcgc	ccagacttca	taaaagctgg	agcttaatct	agaaaaaaa	2880
tcagaaagaa	attacactgt	gagaacaggt	gcaattcact	tttctttac	acagagtaat	2940
actggtaact	catggatgaa	ggcittaaggg	aatgaaattg	gactcacagt	actgagtc	3000
cacactgaaa	aatgcaacct	gatacatcag	cagaaggttt	atgggggaaa	aatgcagcct	3060
tccaattaag	ccagatatct	gtatgaccaa	gctgctccag	aattagtcac	tcaaaatctc	3120
tcagattaaa	ttatcaactg	tcaccaacca	ttctatgct	gacaaggcaa	ttgcttgttc	3180
tctgtgttcc	tgatactaca	aggctcttcc	tgacttccta	aagatgcatt	ataaaaatct	3240
tataattcac	atttctccct	aaactttgac	tcaatcatgg	tatgttgcca	aatatggtat	3300
attactattc	aaattgtttt	ccttgtacc	atatgtaatg	ggtcttgtga	atgtgctctt	3360
ltgttccttt	aatcataata	aaaacatgtt	taagcaaaca	cttttcaact	gtagtatttg	3420
aaggtagcgg	atctcgagcc	gccttcaatg	cccccaaaac	caatccccag	gtttttaact	3480
ctcccgatit	tccaagtacc	atagcccgtc	gagagagcgc	cgcggtaatg	ggatcccagg	3540
accccgggga	atataagtct	gaggggggacg	taagcaaccc	ttcttttgt	aacagggaca	3600
acatagcccc	tatttctctc	ttagaaggag	aggttttccc	gcaataggtc	ttacacgcgg	3660
acgaaatcac	cttatgacg	gcttccatgc	ttgatccacc	ggcgagaccg	aatcacgcag	3720
agcaaccgga	atcacgcctg	gggtggacgg	ctcagtcgtc	gggtctctct	cccgctcttc	3780
aacgactctc	tgagttctcg	gtagggtatg	ttggccccct	gcagtagggc	tcctctcgac	3840
gccactcagc	ttctgccctc	ctaagccgca	gccccctcta	ctagggtcat	cgctccgtcc	3900
ccgaataagc	gagacggatg	aggacaggat	cgccacgcgc	cctgtggccg	accactattc	3960
cctaacgac	acgtcggggt	caccaaatga	agccttctgc	ttcatgcatg	tgctcgtagt	4020
cgtcagggaa	tcaacgggtc	ggccatcaac	ccaggtgcac	accaatgtgg	tgaatgggtca	4080
aatggcggtt	attgtatcga	gctaggcact	taaatacaat	atctctgcaa	tgcggaattc	4140
agtggttcgt	ccaatccgtg	ttagaccgtg	ctgttgccct	cctaacaagg	cacgatcata	4200

[0018]

ccaagatcat accacattac tcccaccaat cggcatgcac ggtgcttttt ctctccttat	4260
aaggcatgtt gctaactcat cgttacataa gcatgttgca agactacaag agtattgcat	4320
aagactacat ttcccccctc ctatgcaaaa gcgaacttac tatatcctga ggggactcct	4380
aaccgcgtac aaccgaagcc ccgcttttcg cctaaacatg ctattgtccc ctcagtcaag	4440
ccttgcccgt tacaaccgga ttgcgaagcc ttgccctccc cacattatcc gtagcattat	4500
ttcctagcag tcacagagc tacagaagat actctatgct gtagccaagt ctacaagttt	4560
actattcagc gacctcctat attccgcgtg ccagccgacg aattaccaat gcgcgcttgg	4620
cgtaatcatg gtcatactg tttcctgtgt gaaattgtta tccgcicaca attccacaca	4680
acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca	4740
cattaattgc gttgcgtca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc	4800
attaatgaat cggccaacgc gcggggagag gcggttttgc tattgggcgc tcttccgctt	4860
cctcgtcac tgactcgtg cgctcgtcg ttccgctcgc gcgagcggta tcagctcact	4920
caaaggcggc aatacgttta tccacagaat caggggataa cgcaggaaag aacatgtgag	4980
caaaaggcca gcaaaaggcc aggaaccgtt aaaaggccgc gttgctggcg tttttccata	5040
ggctccgccc cctgacgag catcacaaaa atcgacgctc aagtcagagg tggcgaacc	5100
cgacaggact ataaagatac caggcgcttc cccctggaag ctccctcgtg cgctctcctg	5160
ttccgacctt gccgcttacc ggatacctgt ccgcctttct ccttccggga agcgtggcgc	5220
tttctcatag ctacacgtgt aggtatctca gtccggtgta ggtcgttcgc tccaagctgg	5280
gctgtgtgca cgaaccccc gttcagcccg accgctgcgc cttatccggt aactatcgtc	5340
ttgagtccaa cccggttaaga cagacttat cgcactggc agcagccact ggtaacagga	5400
ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtag cctaactacg	5460
gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt acctcggaa	5520
aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggg ggttttttg	5580
tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct ttgatctttt	5640
ctacggggtc tgacgtcag tggaacgaaa actcacgtta agggattttg gtcattgagat	5700
tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt aatcaatct	5760
aaagtatata tgagtaaact tggcttgaca gttaccaatg cttaatcagt gaggcaccta	5820
tctcagcgt ctgtctatct cgttcaccca tagttgcctg actccccgtc gtgtagataa	5880
ctacgatacg ggagggccta ccatctggcc ccagtgtgc aatgataccg cgagaccac	5940
gctcaccggc tccagattta tcagcaataa accagccagc cgggaaggcc gagcgcagaa	6000
gtggtctcgc aactttatcc gcctccatcc agtctattaa ttgttgccgg gaagctagag	6060
taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca ggcatcgtgg	6120
tgtaacgctc gtcgttttgt atggttctat tcagctccgg ttcccaacga tcaaggcgag	6180
ttacatgac ccccatgttg tgcaaaaaag cggttagctc ctccggtcct ccgacgttg	6240
tcagaagtaa gttggccgca gtgttatcac tcattgttat ggcagcactg calaattctc	6300
ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca accaagtcac	6360
tctgagaata gtgtatgcgg cgaccgagtt gctcttgcgc ggcgtcaata cgggataata	6420
ccgcgccaca tagcagaact ttaaaagtgc tcattatggg aaaacgttct tcggggcgaa	6480

[0019]

aactctcaag gatcttacg ctgttgagat ccagttcgat gtaacccact cgtgcaccca	6540
actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa acaggaaggc	6600
aaaatgccgc aaaaaagga ataaggcgca cacggaaatg ttgaatactc atactcttcc	6660
ttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg	6720
aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac	6780
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga	6840
ccgtacact tgccagcgcc ctagegccgc ctcttttcgc ttcttccct tcctttctcg	6900
ccacgttcgc cgcttttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat	6960
ttagtctttt acggcacctc gaccccaaaa aacttgatta gggtagtggt tcacgtagtg	7020
ggccatcgcc ctgatagacg gtttttcgcc ctttgacggt ggagtccacg ttctttaata	7080
gtggactctt gtccaaaact ggaacaacac tcaaccctat ctccgtctat tcttttgatt	7140
tataagggat ttgcccatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat	7200
ttaacgcgaa ttttaacaaa atattaacgc ttacaatttc caticgccat tcaggtcgcg	7260
caactgttgg gaaggcgat cgggtcgggc ctcttcgcta ttacgccage tggcgaaagg	7320
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cagcagcttg	7380
taaaacgacg gccagtgage gcgtattccc taacgatcac gtcgggtgca ccaaatgaag	7440
ctttctgctt catgcattg ctctagtcg tcagggaatc aacggtcgg ccatcaaccc	7500
aggtgcacac caatgtggtg aatggtaaaa tggcgtttat tgtatcagc taggcactta	7560
aatacaatat ctctgcaatg cggaattcag tggttcgtcc aatccgtccc cctccctaig	7620
caaaagcgaa actactatat cctgagggga ctccaaaccg cgtacaaccg aagccccgct	7680
tttcgcctaa acatgctatt gtccctcag tcaagccttg cccgttaca cccgattcgc	7740
aagccttgcc ctccccacat tatccgtagc attatttctt agcagtcac agagctacag	7800
aagatactct atgctgtage caagtctaca agtttactat tcagcgacct cctatatcc	7860
gcgtgccage cgatcaattia ccaatccaac cagctatcac acggaataga agaactcgcc	7920
tacgctcttc tttcgggctg cttataagcc tctgttaatt tttttatatt cctcgttaag	7980
tcctcagact tggcaaggag aatgtagatt tccacagtat atatgttttc acaaaaggaa	8040
ggagagaaac aaaagaaat ggcaactgact aaacttcagc tagtggtata ggaaagtaat	8100
tctgcttaac agagattgca gtgatctcta tgtatgtcct gaagaattat gttgtacttt	8160
tttcccccat ttttaaatca aacagtgtt tacagaggtc agaattggtt ctttactgtt	8220
tgtaattctt attatttcaa tacagaacaa tagcttctat aactgaaata tatttgctat	8280
tgtatattat gattgtccct cgaacctga acactcctcc agctgaattt cacaattcct	8340
ctgtcatctg ccaggccatt aagttattca tggaagatct ttgaggaaca ctgcaagttc	8400
atatcataaa cacatttgaa attgagtatt gttttgcatt gtatggagct atgttttgct	8460
gtatcctcag aataaaagtt tgttataaag cattcacacc cataaaaaga tagatttaaa	8520
tattccaact ataggaaaga aagtgtgtct gctcttcact ctagtctcag ttggtcctt	8580
cacatgcacg cttcttttatt tctcttattt tgtcaagaaa ataataggtc aagtcctgtt	8640
ctcatttatg tctgtctag cgtggctcag atgcacattg tacatacaag aaggatcaaa	8700
tgaacagac ttctgtctg ttactacaac catagtaata agcacactaa ctaataattg	8760

[0020]

ctaattatgt tttccatctc caaggttccc acatttttct gttttcttaa agatcccatt	8820
atctggttgt aactgaagct caatggaaca tgagcaatat ttcccagctt tctctcccat	8880
ccaacagtc c tgaaggatta gcagaacagg cagaaaaacac attgttacct agaattaaaa	8940
actaatattt gctctccatt caatccaaaa tggacctatt gaaactaaaa tctaacccaa	9000
tcccattaaa tgatttctat ggtgtcaaag gtcaaacttc tgaagggaac ctgtgggtgg	9060
gtcacaaatc agactatata ttccccaggg ctccagccagt gtctgtacat acagctagaa	9120
agctgtattg cctttagcag tcaagctcga aaggttaagca actctctgga attaccttct	9180
ctctatatta gctcttactt gcacctaaac tttaaaaaat taacaattat tgtgctatgt	9240
gttgtatctt taagggtgaa gtacctgcgt gataccccct ataaaaactt ctccactgtg	9300
tatgcattct gcactatttt attatgtgta aaagctttgt gtttgttttc aggaggctta	9360
ttctttgtgc ttaaaatatg tttttaattt cagaacatct tatcctgtcg ttactatct	9420
gatatgcttt gcagtttget tgattaactt ctageccctac agagtgcaca gagagcaaaa	9480
tcatggtgtt cagtgaatc tggggagtta tttaatgtg aaaattctct agaagtttaa	9540
ttctgcaaa gtgcagctgc tgatcactac acaagataaa aatgtggggg gtgcataaac	9600
gtatattctt acaataatag atacatgtga acttatatac agaaaagaaa atgagaaaaa	9660
tgtgtgtgtg tatactcaca cacgtgggtc gtaaaaactt ttgaggggtt taatacagaa	9720
aatccaatcc tgaggcccca gcactcagta cgcatalaaa gggctgggct ctgaaggact	9780
tctgacttcc acagattata taaatctcag gaaagcaact agattcatgc tggctccaaa	9840
agctgtgctt tatataagca cactggctat acaatagttg tacagttcag ctctttataa	9900
tagaaacaga cagaacaagt ataaalcctc tatgtgtcta tgtcatgaac aagaattcat	9960
tcagtggctc tgttttatag taaacattgc tattttatca tgtctgcatt tctcttctgt	10020
ctgaatgtca ccaactaaaat ttaactccac agaaagttaa tactacagta cacatgcata	10080
tctttgagca aagcaaacca tacclgaaag tgcaatagag cagaatatga attacatgcg	10140
tgtctttctc ctgactaca tgaccccata taaattacat tctttatcta tctgccatc	10200
acaaaaacaa aggtaaaaat acttttgaag atctactcat agcaagtagt gtgcaacaaa	10260
cagatatttc tctacattta tttttaggga ataaaaataa gaaataaaat agtcagcaag	10320
cctctgcttt ctcatatata tgtccaaacc taaagtctac tgaatttgc tctttgaatt	10380
tccagttttg caagcctatc agattgtgtt ttaatcagag gtactgaaaa gtatcaatga	10440
attctagctt tcacigaaca aaaatatgta gaggcaactg gcttctggga cagtttgcta	10500
cccaaaagac aactgaatgc aaatacataa atagatttat gaatatggtt ttgaacatgc	10560
acatgagagg tggatatagc aacagacaca ttaccacaga attactttaa aactacttgt	10620
taacatttaa ttgcctaaaa acigctcgta atttactgtt gtagcctacc atagagtacc	10680
ctgcatggta ctatgtacag cattccatcc ttacatttcc actgttctgc tgtttgctct	10740
agacaactca gagttcacca tg	10762

<210> 9
 <211> 242
 <212> DNA
 <213> 人工序列

<220>
 <223> 适体序列

[0021]

<400> 9	
cccgggttgt taacatttaa ttgcctaaaa actgctcgta atttactgtt gtagcctacc	60
atagagtacc ctgcatggta ctatgtacag cattccatcc ttacattttc actgttctgc	120
tgtttgtctt agacaactca gagttcacca tgaatatgcg gttcttgggg ttggtgggtct	180
gtttggttct ctggaccctg cattccgagg ggtccggagg gaaactgaca gctgtggatc	240
ct	242
<210> 10	
<211> 1575	
<212> DNA	
<213> 人工序列	
<220>	
<223> Syn SBC102	
<400> 10	
ccattatctg gttgtaactg aagctcaatg gaacatgagc aatatttccc agtcttctct	60
cccattcaac agtctctgat gattagcaga acaggcagaa aacacattgt taccagaat	120
taaaaactaa tatttgcctt ccattcaatc caaaatggac ctattgaaac taaaatctaa	180
cccaatccca ttaaatgatt tctatggtgt caaaggctca acttctgaag ggaacctgtg	240
ggtagggcac aattcagact atatatccc cagggtctag ccagtgtctg tacatacagc	300
tagaaagctg tattgccttt agcagtcaag ctgaaagac aactcagagt tcaccatgaa	360
aatgcggttc ttgggggttg tggctctgtt ggttctctgg accctgcatt ccgaggggtc	420
cggagggaaa ctgacagctg tggatcctga aacaacatg aatgtcagt aaattatctc	480
ttactgggga ttccctagt aggaatacct agttgagaca gaagatggat atattctgtg	540
ccttaaccga attcctcatg ggaggaagaa ccattctgac aaaggtccca aaccagttgt	600
cttctgcaa catggcttgc tggcagattc tagtaactgg gtcacaaacc ttgccaacag	660
cagcctgggc ttcatcttg ctgatgtcgg ttttgacgtg tggatgggca acagcagagg	720
aaatacctgg tctcggaac ataagacact ctcatgttct caggatgaat tctgggcttt	780
cagttatgat gagatggcaa aatatgacct accagcttcc attaacttca ttctgaataa	840
gactggccaa gaacaagtgt attatgtggg tcattctcaa ggcaccacta taggttttat	900
agcattttca cagatccctg agctggctaa aaggattaaa atgttttttg ccttgggtcc	960
tgtggcttcc gtcgcttct gtactagccc tatggccaaa ctgggacgac tgccagatca	1020
tctcattaag gacctctttg gagacaaaga atttcttccc cagagtgcgt ttttgaagtg	1080
gctgggtacc cagtttgca ctcatgtcat actgaaggag ctctgtggaa atctctgttt	1140
tcttctgtgt ggatttaatg agagaaattt aaatatgtct agagtggatg tgtatacaac	1200
acattctcct gctggaactt ctgtgcaaaa catgttacac tggagccagg ctgttaaatt	1260
ccaaaagttt caagcctttg actggggaag cagtgccaa aattattttc attacaacca	1320
gagttatcct cccacataca atgtgaagga catgcttgtg ccgactgcag tctggagcgg	1380
gggtcacgac tggcttgcag atgtctacga cgtcaatatc ttactgactc agatcaccaa	1440
cttgggtgtc catgagagca ttccggaatg ggagcatctt gacttcattt ggggcctgga	1500
tgccccttgg aggccttata ataagattat taatctaata aggaaatata agtgattcga	1560
agcggccgcc ccggg	1575

[0022]

<210> 11
 <211> 2789
 <212> DNA
 <213> 人工序列

<220>
 <223> OVR1启动子

<400> 11
 ctcgagtctc ttcagaatgg cacagcaccg ctgcagaaaa atgccagggtg gactatgaac 60
 tcacatccaa aggagcttga cctgatacct gattttcttc aaacagggga aacaacacaa 120
 tcccacaaaa cagctcagag agaaaccatc actgatggct acagcaccaa ggtatgcaat 180
 ggcaatccat tcgacattca tctgtgacct gagcaaaatg atttatctct ccatgaatgg 240
 ttgcttcttt ccttcataaa aaggcaatit ccacactcac aatatgcaac aaagacaaaac 300
 agagaacaat taatgtgctc ctctctaattg ttaaaattgt agtggcaaaag aggagaacaa 360
 aatctcaagt tctgagtagg ttttagtgat tggataagag gctttgacct gtgagctcac 420
 ctggacttca tctccttttg gataaaaaagt gcttttataa ctttcaggtc tccgagtctt 480
 tattcatgag actgttgggt tagggacaga cccacaatga aatgcctggc ataggaaagg 540
 gcagcagagc cttagctgac cttttcttgg gacaagcatt gtcaaacaaat gtgtgacaaa 600
 actattttgta ctgctttgca cagctgtgct gggcagggca atccattgcc acctatccca 660
 ggtaaccttc caactgcaag aagattgttg ctactctctc ctagaacccc aagtcaaacc 720
 aactatgcag gtatgtgac aacactatga tgacagcctg ttctgatcaa gatctcattt 780
 gttcatggac aatttttggt gcttgagct ggtcttccat tgggaaagag tgtagtatat 840
 ccttctcctc tgacagaaaa gcagaaattc tcatgctcca cacttaatct acattgtttt 900
 aaaccaccgg ctacttcttg gagaggaaaa atggctttta taagactcac aaaacaaagc 960
 tctgcaagtc aaatgcatac aaaactgttc ttaggtctg gaatcaggac actatgtgga 1020
 agtcaaatag agcagcttta aaaagccttt gggatcattc tcatcttata ttgacgacac 1080
 gatactatga cagtataaac tgacataact gcatcaattt ccttgatatt ttatttgtct 1140
 taaagtacaa gacatagaga tggacgtaaa gatggacata tgactcagggt ctggacagggt 1200
 ccgtgggtcca tgtatgataa aagagatgaa gggaaggaga attgagactg tctaagaagg 1260
 gcttcaggga cgttctgaag gcagatttga ctgaatcaga tgtactgtcc aagtctcata 1320
 tgtagcaatg gaaggtgat attggagaaa tataaagaaa tggtgtgaa ctcaaagtga 1380
 cctgaacag aaaaggata tggagttaaa ataatgtcac agaactgagg tttatatgat 1440
 ataccatggg ctgcagaggg tcagagtgtt ccaccatggg cctctcttgg gctgcaggga 1500
 acttctgttc tacacctgga acacctctg ccctctctcg cactgacctc agtgtcatca 1560
 gggtgttttc tctcacattt tctactcac ctctcccaac taccattgta cagcagttgt 1620
 tcttacatat tgctctctc gaggtacatc tagcatcgtt aagtcctcag acttggcaag 1680
 gagaatgtag atttcacag tatatatgtt ttcacaaaag gaaggagaga aacaaaagaa 1740
 aatggcactg actaaacttc agctagtgggt ataggaaagt aattctgctt aacagagatt 1800
 gcagtgtct ctatgtatgt cctgaagaat tatgtgtac tttttccccc catttttaa 1860
 tcaaacagtg ctttacagag gtcagaatgg tttctttact gttgtcaat tctattattt 1920
 caatacagaa caatagcttc tataactgaa atatatgtgc tattgtatat tatgattgtc 1980

[0023]

```

cctcgaacca tgaacactcc tccagctgaa ttccacaatt cctctgtcat ctgccaggcc 2040
attaagttat tcatggaaga tctttgagga acactgcaag ttcatatcat aaacacattt 2100
gaaattgagt attgttttgc attgtatgga gctatgtttt gctgtatcct cagaataaaa 2160
gtttgttata aagcattcac acccataaaa agatagattt aaatattcca actataggaa 2220
agaaagtgtg tctgtcttcc actctagtct cagttggtc cttcacatgc acgtttcttt 2280
atttctecta ttttgtcaag aaaataatag gtcaagtctt gttctcattt atgtcctgtc 2340
tagcggtggt cagatgcaca ttgtacatac aagaaggatc aaatgaaca gacttctggt 2400
ctgttactac aaccatagta ataagcacac taactaataa ttgctaatta tgttttccat 2460
ctccaaggtt cccacatttt tctgttttct taaagatccc attatctggt tgaactgaa 2520
gtcfaatgga acatgagcaa tatttcccag tcttctctcc catccaacag tcctgatgga 2580
ttagcagaac aggcagaaaa cacattgita cccagaatta aaaactaata tttgctctcc 2640
attcaatcca aaatggacct attgaaacta aaatctaacc caateccatt aaatgatttc 2700
tatgggtgca aaggtaaac ttctgaaggg aacctgtggg tgggtcaca ttcagactat 2760
atattcccca gggctcagcc agtgtctgt 2789

```

<210> 12
 <211> 20
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 12
 agaaactgag agtgtcttat 20

<210> 13
 <211> 29
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 13
 tgacagctgt ggatccagaa acaaacatg 29

<210> 14
 <211> 29
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 14
 gccgctcgag cgaggaatat aaaaaaatt 29

<210> 15
 <211> 20
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 15
 tccgcgcaca tttccccgaa 20

[0024]

<210> 16
 <211> 21
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 16
 acgactggct tgcagatgtc t 21

<210> 17
 <211> 22
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 17
 ccccaaata agtcaagatg ct 22

<210> 18
 <211> 25
 <212> DNA
 <213> 人工序列

<220>
 <223> 合成寡核苷酸

<400> 18
 ccggaatgct ctcatggaac accaa 25

<210> 19
 <211> 372
 <212> PRT
 <213> 人

<400> 19

Ala Val Asp Pro Glu Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr
 1 5 10 15

Trp Gly Phe Pro Ser Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr
 20 25 30

Ile Leu Cys Leu Asn Arg Ile Pro His Gly Arg Lys Asn His Ser Asp
 35 40 45

Lys Gly Pro Lys Pro Val Val Phe Leu Gln His Gly Leu Leu Ala Asp
 50 55 60

Ser Ser Asn Trp Val Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile
 65 70 75 80

Leu Ala Asp Ala Gly Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn
 85 90 95

Thr Trp Ser Arg Lys His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe
 100 105 110

Trp Ala Phe Ser Tyr Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser
 115 120 125

[0025]

Ile Asn Phe Ile Leu Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val
 130 135 140
 Gly His Ser Gln Gly Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile
 145 150 155 160
 Pro Glu Leu Ala Lys Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val
 165 170 175
 Ala Ser Val Ala Phe Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu
 180 185 190
 Pro Asp His Leu Ile Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro
 195 200 205
 Gln Ser Ala Phe Leu Lys Trp Leu Gly Thr His Val Cys Thr His Val
 210 215 220
 Ile Leu Lys Glu Leu Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe
 225 230 235 240
 Asn Glu Arg Asn Leu Asn Met Ser Arg Val Asp Val Tyr Thr Thr His
 245 250 255
 Ser Pro Ala Gly Thr Ser Val Gln Asn Met Leu His Trp Ser Gln Ala
 260 265 270
 Val Lys Phe Gln Lys Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala Lys
 275 280 285
 Asn Tyr Phe His Tyr Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val Lys
 290 295 300
 Asp Met Leu Val Pro Thr Ala Val Trp Ser Gly Gly His Asp Trp Leu
 305 310 315 320
 Ala Asp Val Tyr Asp Val Asn Ile Leu Leu Thr Gln Ile Thr Asn Leu
 325 330 335
 Val Phe His Glu Ser Ile Pro Glu Trp Glu His Leu Asp Phe Ile Trp
 340 345 350
 Gly Leu Asp Ala Pro Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu Met
 355 360 365
 Arg Lys Tyr Gln
 370
 <210> 20
 <211> 399
 <212> PRT
 <213> 人
 <400> 20
 Met Lys Met Arg Phe Leu Gly Leu Val Val Cys Leu Val Leu Trp Thr
 1 5 10 15
 Leu His Ser Glu Gly Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu

[0026]

20	25	30
Thr Asn Met Asn Val Ser Glu 35	Ile Ile Ser Tyr Trp Gly 40	Phe Pro Ser 45
Glu Glu Tyr Leu Val Glu Thr 50	Glu Asp Gly Tyr Ile Leu Cys Leu Asn 55	60
Arg Ile Pro His Gly Arg Lys Asn His Ser Asp Lys Gly Pro Lys Pro 65	70	75 80
Val Val Phe Leu Gln His Gly Leu Leu Ala Asp Ser Ser Asn Trp Val 85	90	95
Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly 100	105	110
Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn Thr Trp Ser Arg Lys 115	120	125
His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe Trp Ala Phe Ser Tyr 130	135	140
Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser Ile Asn Phe Ile Leu 145	150	155 160
Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val Gly His Ser Gln Gly 165	170	175
Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile Pro Glu Leu Ala Lys 180	185	190
Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val Ala Ser Val Ala Phe 195	200	205
Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu Pro Asp His Leu Ile 210	215	220
Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro Gln Ser Ala Phe Leu 225	230	235 240
Lys Trp Leu Gly Thr His Val Cys Thr His Val Ile Leu Lys Glu Leu 245	250	255
Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe Asn Glu Arg Asn Leu 260	265	270
Asn Met Ser Arg Val Asp Val Tyr Thr Thr His Ser Pro Ala Gly Thr 275	280	285
Ser Val Gln Asn Met Leu His Trp Ser Gln Ala Val Lys Phe Gln Lys 290	295	300
Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala Lys Asn Tyr Phe His Tyr 305	310	315 320
Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val Lys Asp Met Leu Val Pro		

[0027]

325										330					335				
Thr	Ala	Val	Trp	Ser	Gly	Gly	His	Asp	Trp	Leu	Ala	Asp	Val	Tyr	Asp				
			340				345						350						
Val	Asn	Ile	Leu	Leu	Thr	Gln	Ile	Thr	Asn	Leu	Val	Phe	His	Glu	Ser				
			355				360						365						
Ile	Pro	Glu	Trp	Glu	His	Leu	Asp	Phe	Ile	Trp	Gly	Leu	Asp	Ala	Pro				
			370				375						380						
Trp	Arg	Leu	Tyr	Asn	Lys	Ile	Ile	Asn	Leu	Met	Arg	Lys	Tyr	Gln					
			385				390						395						

rhLAL	1	60
天然的	MKMRFLGLVVCLVLWTLHSEG <u>SGGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI</u> MKMRFLGLVVCLVLWTLHSEG <u>SGGKLTAVDPETNMNVSEIISYWGFPSEEYLVETEDGYI</u>	
rhLAL	61	120
天然的	<u>LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS SLGFILADAGFDVWMGNS</u> <u>LCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS SLGFILADAGFDVWMGNS</u>	
rhLAL	121	180
天然的	<u>RGNTWSRKHKTL SVSQDEFWAFSYDEMAKYDLPASINFILNKTGOEQVYYVGHSQGT TIG</u> <u>RGNTWSRKHKTL SVSQDEFWAFSYDEMAKYDLPASINFILNKTGOEQVYYVGHSQGT TIG</u>	
rhLAL	181	240
天然的	<u>FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL</u> <u>FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL</u>	
rhLAL	241	300
天然的	<u>KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTTHSPAGTSVQNMLHWSQAV</u> <u>KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTTHSPAGTSVQNMLHWSQAV</u>	
rhLAL	301	360
天然的	<u>KFQKFQAFDWGSSAKNYFHYNQSYPTYNVKDMLVPTAVWSGGHDWLADVDVNILLTQI</u> <u>KFQKFQAFDWGSSAKNYFHYNQSYPTYNVKDMLVPTAVWSGGHDWLADVDVNILLTQI</u>	
rhLAL	361	399
天然的	<u>TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ</u> (SEQ ID NO:1) <u>TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ</u> (SEQ ID NO:20)	

图 1

```
1   ATGAAAATGC GGTTCCTGGG GTTGGTGGTC TGTTTGGTTC TCTGGACCCT
51  GCATTCCGAG GGGTCCGGAG GGAAACTGAC AGCTGTGGAT CCTGAAACAA
101 ACATGAATGT CAGTGAAATT ATCTCTTACT GGGGATTCCC TAGTGAGGAA
151 TACCTAGTTG AGACAGAAGA TGGATATATT CTGTGCCTTA ACCGAATTCC
201 TCATGGGAGG AAGAACCATT CTGACAAAGG TCCCAAACCA GTTGTCTTCC
251 TGCAACATGG CTTGCTGGCA GATTCTAGTA ACTGGGTCAC AAACCTTGCC
301 AACAGCAGCC TGGGCTTCAT TCTTGCTGAT GCTGGTTTTG ACGTGTGGAT
351 GGGCAACAGC AGAGGAAATA CCTGGTCTCG GAAACATAAG ACACTCTCAG
401 TTTCTCAGGA TGAATTCTGG GCTTTCAGTT ATGATGAGAT GGCAAAATAT
451 GACCTACCAG CTTCCATTAA CTTCAATTCTG AATAAGACTG GCCAAGAACA
501 AGTGTATTAT GTGGGTCATT CTCAAGGCAC CACTATAGGT TTTATAGCAT
551 TTTACAGAT CCCTGAGCTG GCTAAAAGGA TTAAATGTT TTTTGCCCTG
601 GGTCTGTGG CTTCCGTCGC CTTCTGTACT AGCCCTATGG CCAAACCTGGG
651 ACGACTGCCA GATCATCTCA TTAAGGACCT CTTTGAGAC AAAGAATTC
701 TTCCCAGAG TCGTTTTTG AAGTGGCTGG GTACCCACGT TTGCACTCAT
751 GTCATACTGA AGGAGCTCTG TGGAAATCTC TGTTTTCTTC TGTGTGGATT
801 TAATGAGAGA AATTTAAATA TGTCTAGAGT GGATGTGTAT ACAACACATT
851 CTCCTGCTGG AACTTCTGTG CAAAACATGT TACACTGGAG CCAGGCTGTT
901 AAATTCCAAA AGTTTCAAGC CTTTGACTGG GGAAGCAGTG CCAAGAATTA
951 TTTTCATTAC AACCAGAGTT ATCCTCCAC ATACAATGTG AAGGACATGC
1001 TTGTGCCGAC TGCACTCTGG AGCGGGGGTC ACGACTGGCT TGCAGATGTC
1051 TACGACGTCA ATATCTTACT GACTCAGATC ACCAACTTGG TGTTCCATGA
1101 GAGCATTCCG GAATGGGAGC ATCTTGACTT CATTTGGGGC CTGGATGCCC
(SEQ ID NO:5)
```

图 2

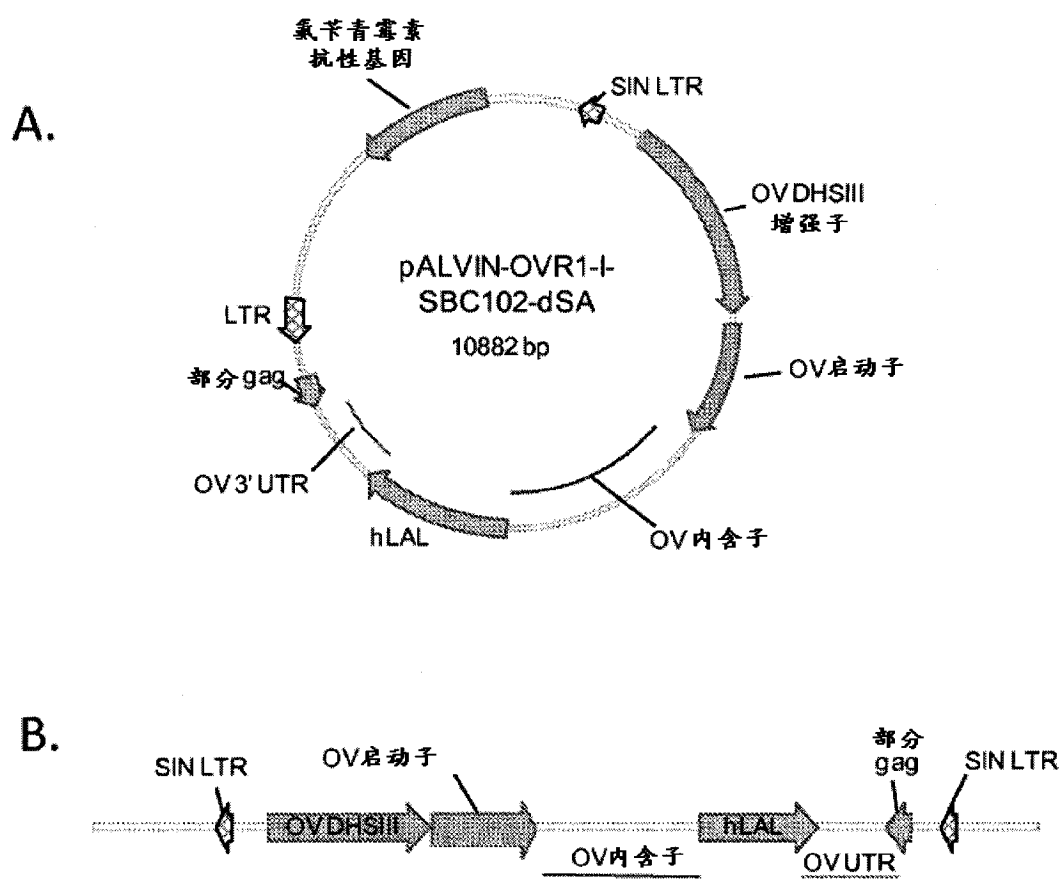


图 3

pALVIN-OVR1-I-SBC102-dSA 的序列 (10882 bp)

特征	位置 (bp)
SIN LTR	521-693 (互补链)
DHSIII 增强子	1069-2720
OV 启动子	2720-3851
内含子	3899-5487
hLAL CDS	5505-6704
OV 3'UTR	6719-7392
部分 gag CDS	7404-7657 (互补链)
LTR	7955-8300 (互补链)
氯霉素抗性基因	9764-10621 (互补链)

```

1   CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT CCCTTTAGGG TTCCGATTTA GTGCTTTACG
61  GCACCTCGAC CCCAAAAAAC TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG
121 ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTC TTTAATAGTG GACTCTTGTT
181 CCAAACCTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT
241 GCCGATTTCG GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTT
301 TAACAAAATA TTAACGCTTA CAATTTCAT TCGCCATTCA GGCTGCGCAA CTGTTGGGAA
361 GGGCGATCGG TCGGGGCCTC TTCGCTATTA CGCCAGCTGG CGAAAGGGGG ATGTGCTGCA
421 AGGCGATTAA GTTGGGTAAC GCCAGGGTTT TCCCAGTCAC GACGTTGTAA AACGACGGCC
481 AGTGAGCGCG TATCCCTTAA CGATCACGTC GGGGTCACCA AATGAAGCCT TCTGCTTCAT
541 GCATGTGCTC GTAGTCGTCA GGGAAATCAAC GGTCCGGCCA TCAACCCAGG TGCACACCAA
601 TGTGGTGAAT GGTCAAATGG CGTTTATTGT ATCGAGCTAG GCACTTAAAT ACAATATCTC
661 TGCAATGCGG AATTCACTGG TTCGTCCAAT CCGTCCCCCT CCCTATGCAA AAGCGAAACT
721 ACTATATCCT GAGGGGACTC CTAACCGCGT ACAACCGAAG CCCCCTTTT CGCCTAAACA
781 TGCTATTGTC CCTCAGTCA AGCCTTGCCC GTTACAACCC GATTGCAAG CTTGCCCTC
841 CCCACATTAT CCGTAGCATT ATTCCTAGC AGTCATCAGA GCTACAGAAG ATACTCTATG
901 CTGTAGCCAA GTCTACAAGT TTACTATTCA GCGACCTCCT ATATTCCCGG TGCCAGCCGA
961 TCAATTACCA ATCCAACCAG CTATCACACG GAATACAAGA ACTCGCCTAC GCTCTTCTTT
1021 CGGGCTGCTT ATAAGCCTCC TGTAATTTTT TTATATTCTT CGCTCGAGTC TCTTCAGAA
1081 GGCACAGCAC CGCTGCAGAA AAATGCCAGG TGGACTATGA ACTCACATCC AAAGGAGCTT
1141 GACCTGATAC CTGATTTTCT TCAACAGGG GAAACAACAC AATCCCACAA AACAGCTCAG
1201 AGAGAAACCA TCACTGATGG CTACAGCACC AAGGTATGCA ATGGCAATCC ATTCGACATT
1261 CATCTGTGAC CTGAGCAAAA TGATTATCT CTCCATGAAT GGTGCTTCT TTCCCTCATG
1321 AAAAGGCAAT TTCCACACTC ACAATATGCA ACAAAGACAA ACAGAGAACA ATTAATGTGC
1381 TCCTTCCTAA TGTTAAATTT TGTAGTGCAA AGAGGAGAAC AAAATCTCAA GTTCTGAGTA
1441 GGTTTTAGTG ATTGGATAAG AGGCTTTGAC CTGTGAGCTC ACCTGGACTT CATATCCTTT
1501 TGGATAAAAA GTGCTTTTAT AACTTTCAGG TCTCCGAGTC TTTATTCTATG AGACTGTTGG
1561 TTTAGGGACA GACCACAAT GAAATGCCTG GCATAGGAAA GGGCAGCAGA GCCTTAGCTG
1621 ACCTTTTCTT GGGACAAGCA TTGTCAAACA ATGTGTGACA AAACATTTTG TACTGCTTTG
1681 CACAGCTGTG CTGGGCAGGG CAATCCATTG CCACCTATCC CAGGTAACCT TCCAACCTGA
1741 AGAAGATTGT TGCTTACTCT CTCTAGACCC CCAAGTCAAA CCAACTATGC AGGTATGCTG
1801 ACAACACTAT GATGACAGCC TGTTCTGATC AAGATCTCAT TTGTTTCATGG ACAATTTTTG
1861 TTGCTTGCAG CTGCTCTTCC ATTGGGAAAG AGTGTAGTAT ATCCTTCTCA TCTGACAGAA
1921 AAGCAGAAAT TCTCATGCTC CACACTTAAT CTACATTGTT TTAAACCACC GGCTACTTCT
1981 TGGAGAGGAA AAATGGCTTT TATAAGACTC ACAAACAAA GCTCTGCAAG TCAAATGCAT
2041 ACAAAACCTG TCTGTAGGTC TGGAAATCAGG ACACATGTG GAAGTCAAT AGAGCAGCTT
2101 TAAAAAGCCT TTGGGATCAT TCTCATCTTA TATTTGCAGC ACGATACTAT GACAGTGATA
2161 ACTGACATAA CTGCATCAAT TTCCTTGATA TTTTATTTGT CTTAAAGTAC AAGACATAGA
2221 GATGGACGTA AAGATGGACA TATGACTCAG GTCTGGACAG GTCCGTGGTC CATGTATGAT
2281 AAAAGAGATG AAGGGAAGGA GAATTGAGAC TGTCTAAGAA GGGCTTCAGG GACGTTCTGA
2341 AGGCAGATTT GACTGAATCA GATGTACTGT CCAAGTCTCA TATGTAGCAA TGGAAAGGCTG
2401 ATATTGGAGA AATATAAAGA AATGGCTGTG AACTCAAAGT GACCCTGAAC AGAAAAGGGA
2461 TATGGAGTTA AAATAATGTC ACAGAACTGA GGTTTATATG ATATACCATG GGCTGCAGAG
2521 GGTGAGAGTG CTCCACCATG GGCCTCTCTT GGGCTGCAGG GAACTTCTGT TCTACACCTG
2581 GAACACCTCC TGCCCTCCTC CGCACTGACC TCAGTGTCAT CAGGGCTGTT TCTCTCACAT
2641 TTTCTCACTC ACCTCTCCCA ACTACCATTG TACAGCAGTT GTTCTTACAT ATTGCTCCTC
2701 CTGAGGTACA TCTAGCATCG TTAAGTCTC AGACTTGGCA AGGAGAAATG AGATTTCAC
2761 AGTATATATG TTTTCACAAA AGGAAGGAGA GAAACAAAAG AAAATGGCAC TGACTAAACT

```

2821	TCAGCTAGTG	GTATAGGAAA	GTAATTCTGC	TTAACAGAGA	TTGCAGTGAT	CTCTATGTAT
2881	GTCCTGAAGA	ATTATGTTGT	ACTTTTTTCC	CCCATTTTTA	AATCAAACAG	TGCTTTACAG
2941	AGGTCAGAAT	GGTTTCTTTA	CTGTTTGTCA	ATTCTATTAT	TTCAATACAG	AACAATAGCT
3001	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	TCCCTCGAAC	CATGAACACT
3061	CCTCCAGCTG	AATTTACAAA	TTCTCTGTCT	ATCTGCCAGG	CCATTAAGTT	ATTTCATGGAA
3121	GATCTTTGAG	GAACACTGCA	AGTTCATATC	ATAAACACAT	TTGAAATTGA	GTATTGTTTT
3181	GCATTGTATG	GAGCTATGTT	TTGCTGTATC	CTCAGAATAA	AAGTTTGTTA	TAAAGCATTTC
3241	ACACCCATAA	AAAGATAGAT	TTAAATATTC	CAACTATAGG	AAAGAAAAGTG	TGTCTGCTCT
3301	TCACTCTAGT	CTCAGTTGGC	TCCTTCACAT	GCACGCTTCT	TTATTTCTCC	TATTTTGTCA
3361	AGAAAATAAT	AGGTCAAGTC	TTGTTCTCAT	TTATGTCCTG	TCTAGCGTGG	CTCAGATGCA
3421	CATTGTACAT	ACAAGAAGGA	TCAAATGAAA	CAGACTTCTG	GTCTGTTACT	ACAACCATAG
3481	TAATAAGCAC	ACTAACTAAT	AATTGCTAAT	TATGTTTTCC	ATCTCCAAGG	TTCCCACATT
3541	TTTCTGTTTT	CTTAAAGATC	CCATTATCTG	GTTGTAAGTG	AAGCTCAATG	GAACATGAGC
3601	AATATTTCCT	AGTCTTCTCT	CCCATCCAAC	AGTCCTGATG	GATTAGCAGA	ACAGGCAGAA
3661	AACACATTGT	TACCCAGAAT	TAAAAACTAA	TATTTGCTCT	CCATTCAATC	CAAAATGGAC
3721	CTATTGAAAC	TAAATCTAA	CCCAATCCCA	TTAAATGATT	TCTATGGTGT	CAAAGGTCAA
3781	ACTTCTGAAG	GGAACCTGTG	GGTGGGTCAC	AATTCAGACT	ATATATTCCC	CAGGGCTCAG
3841	CCAGTGTCTG	TACATACAGC	TAGAAAGCTG	TATTGCCCTT	AGCAGTCAAG	CTCGAAAGGT
3901	AAGCAACTCT	CTGGAATTAC	CTTCTCTCTA	TATTAGCTCT	TACTTGCACC	TAAACTTTAA
3961	AAAATTAACA	ATTATTGTGC	TATGTGTTGT	ATCTTTAAGG	GTGAAGTACC	TGCGTGATAC
4021	CCCTATATAA	AACCTTCTAC	CTGTGTATGC	ATTCTGCACT	ATTTTATTAT	GTGTAAAAGC
4081	TTTGTGTTTG	TTTTCAGGAG	GCTTATTCTT	TGTGCTTAAA	ATATGTTTTT	AATTTTCAGAA
4141	CATCTTATCC	TGTCGTTTAC	TATCTGATAT	GCTTTGCAGT	TTGCTTGATT	AACTTCTAGC
4201	CCTACAGAGT	GCACAGAGAG	CAAAATCATG	GTGTTTCAGT	AATTCCTGGG	AGTTATTTTA
4261	ATGTGAAAAT	TCTCTAGAAG	TTTAATTCCT	GCAAAGTGCA	GCTGCTGATC	ACTACACAAG
4321	ATAAAAATGT	GGGGGGTGCA	TAAACGTATA	TTCTTACAAT	AATAGATACA	TGTGAACCTA
4381	TATACAGAAA	AGAAAATGAG	AAAAATGTGT	TGTGTATATC	TCACACACGT	GGTCAGTAAA
4441	AACTTTTGAG	GGGTTTAATA	CAGAAAAATC	AATCCTGAGG	CCCCAGCACT	CAGTACGCAT
4501	ATAAAGGGCT	GGGCTCTGAA	GGACTTCTGA	CTTTCACAGA	TTATATAAAT	CTCAGGAAAG
4561	CAACTAGATT	CATGCTGGCT	CCAAAAGCTG	TGCTTTATAT	AAGCACACTG	GCTATACAAT
4621	AGTTGTACAG	TTCAGCTCTT	TATAATAGAA	ACAGACAGAA	CAAGTATAAA	TCTTCTATTG
4681	GTCTATGTCA	TGAACAAGAA	TTCATTACAGT	GGCTCTGTTT	TATAGTAAAC	ATTGCTATTT
4741	TATCATGTCT	GCATTCTCTT	TCTGTCTGAA	TGTCACCACT	AAAATTTAAC	TCCACAGAAA
4801	GTTTATACCTA	CAGTACACAT	GCATATCTTT	GAGCAAAGCA	AAACCATACCT	GAAAGTGCAG
4861	TAGAGCAGAA	TATGAATTAC	ATGCGTGTCT	TTCTCCTAGA	CTACATGACC	CCATATAAAT
4921	TACATTCCTT	ATCTATTCTG	CCATCACCAA	AACAAAGGTA	AAAATACTTT	TGAAGATCTA
4981	CTCATAGCAA	GTAGTGTGCA	ACAAACAGAT	ATTTCTCTAC	ATTTATTTTT	AGGGAATAAA
5041	AATAAGAAAT	AAAAATAGTCA	GCAAGCCTCT	GCTTTCTCAT	ATATCTGTCC	AAACCTAAAG
5101	TTTACTGAAA	TTTGCTCTTT	GAATTTCCAG	TTTTGCAAGC	CTATCAGATT	GTGTTTTAAT
5161	CAGAGGTACT	GAAAAGTATC	AATGAATTTCT	AGCTTTCACT	GAACAAAAAT	ATGTAGAGGC
5221	AACCTAGTTC	TGGGACAGTT	TGCTATCCAA	AAGACAACCTG	AATGCAAAAT	CATAAATAGA
5281	TTTATGAATA	TGGTTTGAA	CATGCACATG	AGAGGTGGAT	ATAGCAACAG	ACACATTACC
5341	ACAGAATTAC	TTTAAACTA	CTTGTTAACA	TTTAATTGCC	TAAAAACTGC	TCGTAATTTA
5401	CTGTTGTAGC	CTACCATAGA	GTACCCTGCA	TGGTACTATG	TACAGCATTC	CATCCTTACA
5461	TTTTCACTGT	TCTGCTGTTT	GCTCTAGACA	ACTCAGAGTT	CACCATGAAA	ATGCGGTTCT
5521	TGGGGTTGGT	GGTCTGTTTG	GTTCTCTGGA	CCCTGCATTC	CGAGGGGTCC	GGAGGGAAAC
5581	TGACAGCTGT	GGATCCAGAA	ACAAACATGA	ATGTCAGTGA	AATTATCTCT	TACTGGGGAT
5641	TCCCTAGTGA	GGAATACCTA	GTTGAGACAG	AAGATGGATA	TATTCTGTCT	CTTAACCGAA
5701	TTCCTCATGG	GAGGAAGAAC	CATTCTGACA	AAGGTCCCAA	ACCAGTTGTC	TTCTTGCAAC
5761	ATGGCTTGCT	GGCAGATTCT	AGTAACTGGG	TCACAAACCT	TGCCAACAGC	AGCCTGGGCT
5821	TCATTCTTGC	TGATGCTGGT	TTTGACGTGT	GGATGGGCAA	CAGCAGAGGA	AATACCTGGT
5881	CTCGGAAACA	TAAGACACTC	TCAGTTTCTC	AGGATGAATT	CTGGGCTTTC	AGTTATGATG
5941	AGATGGCAAA	ATATGACCTA	CCAGCTTCCA	TTAACTTCAT	TCTGAATAAG	ACTGGCCAAG
6001	AACAAGTGTA	TTATGTGGGT	CATTCTCAAG	GCACCACTAT	AGGTTTATA	GCATTTTCAC
6061	AGATCCCTGA	GCTGGCTAAA	AGGATTAATA	TGTTTTTTGC	CCTGGGTCCT	TGGGCTTCCG
6121	TCGCCTTCTG	TACTAGCCCT	ATGGCCAAAC	TGGGACGACT	GCCAGATCAT	CTCATTAAGG
6181	ACCTCTTTGG	AGACAAAGAA	TTTCTTCCCC	AGAGTGCGTT	TTTGAAGTGG	CTGGGTACCC
6241	ACGTTTGCAC	TCATGTCATA	CTGAAGGAGC	TCTGTGGAAA	TCTCTGTTTT	CTTCTGTGTG
6301	GATTTAATGA	GAGAAATTTA	AATATGTCTA	GAGTGGATGT	GTATACAACA	CATTCTCCTG
6361	CTGGAACCTC	TGTGCAAAAC	ATGTTACACT	GGAGCCAGGC	TGTTAAATTC	CAAAAGTTTC
6421	AAGCCTTTGA	CTGGGGAAGC	AGTGCCAAGA	ATTATTTTCA	TTACAACCAG	AGTTATCCTC

6481	CCACATACAA	TGTGAAGGAC	ATGCTTGTGC	CGACTGCAGT	CTGGAGCGGG	GGTCACGACT
6541	GGCTTGCAGA	TGTCTACGAC	GTCAATATCT	TACTGACTCA	GATCACCAAC	TTGGTGTTC
6601	ATGAGAGCAT	TCCGGAATGG	GAGCATCTTG	ACTTCATTTG	GGGCCTGGAT	GCCCCCTTGA
6661	GGCTTTTATA	TAAGATTATT	AATCTAATGA	GGAAATATCA	GTGATTTCGAA	GCGGCCGCAA
6721	GAAGAAAGCT	GAAAAACTCT	GTCCCTTCCA	ACAAGACCCA	GAGCACTGTA	GTATCAGGGG
6781	TAAATGAAA	AGTATGTTAT	CTGCTGCATC	CAGACTTCAT	AAAAGCTGGA	GCTTAATCTA
6841	GAAAAAAAT	CAGAAAGAAA	TTACACTGTG	AGAACAGGTG	CAATTCACCT	TTCTTTTACA
6901	CAGAGTAATA	CTGGTAACTC	ATGGATGAAG	GCTTAAGGGA	ATGAAATTGG	ACTCACAGTA
6961	CTGAGTCATC	ACACTGAAAA	ATGCAACCTG	ATACATCAGC	AGAAGGTTTA	TGGGGGAAAA
7021	ATGCAGCCTT	CCAATTAAGC	CAGATATCTG	TATGACCAAG	CTGCTCCAGA	ATTAGTCACT
7081	CAAAATCTCT	CAGATTAAAT	TATCAACTGT	CACCAACCAT	TCCTATGCTG	ACAAGGCAAT
7141	TGCTTGTCTT	CTGTGTTCC	GATACTACAA	GGCTCTTCCT	GACTTCCTAA	AGATGCATTA
7201	TAAAAATCTT	ATAATTCACA	TTCTCCCTTA	AACCTTGACT	CAATCATGGT	ATGTTGGCAA
7261	ATATGGTATA	TTACTATTCA	AATTGTTTTT	CTTGTACCCA	TATGTAATGG	GTCTTGTGAA
7321	TGTGCTCTTT	TGTTCTTTTA	ATCATAATAA	AAACATGTTT	AAGCAAACAC	TTTTCACTTG
7381	TAGTATTTGA	AGGTACCGGA	TCCTCAGCCG	CCTTCAATGC	CCCCAAAACC	AATCCCCAGG
7441	TTTTTAACTC	TCCCGATTTT	CCAAGTACCA	TAGCCCGCTG	AGAGAGCGCC	GCGGTAATGG
7501	GATCCCAGGA	CCCCGGGGAA	TATAAGTCTG	AGGGGGACGT	AAGCAACCTT	TCCTTTTGTA
7561	ACAGGGACAA	CATAGCCCTT	ATTTCTTCTT	TAGAAGGAGA	GGTTTTCCCG	CAATAGGTCT
7621	TACACGCGGA	CGAAATCACC	TTTATGACGG	CTTCCATGCT	TGATCCACCG	GCGGACCGGA
7681	ATCACGCAGA	GCAACCGGAA	TCACGCCTGG	GGTGGACCGC	TCAGTCGTCG	GGCTTCCTTC
7741	CCGTCTTCCA	ACGACTCTCT	GAGTTCTCGG	TAGGGTATGT	TGGCCCCCTG	CAGTAGGGCT
7801	CCCTCCGACG	CCACTCAGCT	TCTGCCCTCC	TAAGCCGCAG	CCCCCTCTAC	TAGGGTCATC
7861	GTCCGCTCCC	CGAATAAGCG	AGACGGATGA	GGACAGGATC	GCCACGCCGC	CTGTGGCCGA
7921	CCACTATTCC	CTAACGATCA	CGTCGGGGTG	ACCAAATGAA	GCCTTCTGCT	TCATGCATGT
7981	GACTCGTAGT	GTCAGGGAAT	CAACGGTCCG	GCCATCAACC	CAGGTGCACA	CCAATGTGGT
8041	GAATGGTCAA	ATGGCGTTTA	TTGTATCGAG	CTAGGCACCT	AAATACAATA	TCTCTGCAAT
8101	GCGGAATTCA	GTGGTTCGTC	CAATCCGTGT	TAGACCCGTC	TGTTGCCTTC	CTAACAAGGC
8161	ACGATCATAC	CACGATCATA	CCACCTTACT	CCCACCAATC	GGCATGCACG	GTGCTTTTTT
8221	TCTCCTTATA	AGGCATGTTG	CTAACTCATC	GTTACATAAG	CATGTTGCAA	GACTACAAGA
8281	GTATTGCATA	AGACTACATT	TCCCCCTCCC	TATGCAAAAG	CGAAACTACT	ATATCCTGAG
8341	GGGACTCCTA	ACCGCGTACA	ACCGAAGCCC	CGCTTTTCGC	CTAAACATGC	TATTGTCCCC
8401	TCAGTCAAGC	CTTGCCCGTT	ACAACCCGAT	TCGCAAGCCT	TGCCCTCCCG	ACATTATCCG
8461	TAGCATTATT	TCCTAGCAGT	CATCAGAGCT	ACAGAAGATA	CTCTATGCTC	AGTCAAGTGC
8521	TACAAGTTTA	CTATTACAGC	ACCTCCTATA	TTCCGCGTGC	CAGCCGATCA	ATTACCAATG
8581	CGCGCTTGGC	GTAATCATGG	TCATAGCTGT	TTCTGTGTGT	AAATTGTTAT	CCGCTCACAA
8641	TTCCACACAA	CATACGAGCC	GGAAGCATAA	AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA
8701	GCTAACTCAC	ATTAATTGCG	TTGCGCTCAC	TGCCCCGCTT	CCAGTCGGGA	AACCTGTGCT
8761	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	CGGGGAGAGG	CGGTTTGCGT	ATTGGGCGCT
8821	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	GCTCGGTGCT	TCGGCTGCGG	CGAGCGGTAT
8881	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	CCACAGAATC	AGGGGATAAT	GCAGGAAAGA
8941	ACATGTGAGC	AAAAGGCCAG	CAAAAGGCCA	GGAACCGTAA	AAAGGCCGCG	TTGCTGGCCT
9001	TTTTCCATAG	GCTCCGCCCC	CCTGACGAGC	ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT
9061	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC
9121	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	GATACCTGTC	CGCCTTTCTC	CCTTCGGGAA
9181	GCGTGGCGCT	TTCTCATAGC	TCACGCTGTA	GGTATCTCAG	TTCCGGTGTAG	GTGTTTCGCT
9241	CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	TTACGCCCCG	CCGCTGCGCC	TTATCCGGTA
9301	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG
9361	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC
9421	CTAACTACGG	CTACACTAGA	AGGACAGTAT	TTGGTATCTG	CGCTCTGCTG	AAGCCAGTTA
9481	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	CCGGCAAACA	AACCACCGCT	GGTAGCGGTG
9541	GTTTTTTTTG	TTGCAAGCAG	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT
9601	TGATCTTTTC	TACGGGGTCT	GACGCTCAGT	GGAACGAAAA	CTCACGTTAA	GGGATTTTGG
9661	TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	AGATCCTTTT	AAATTAAAAA	TGAAGTTTTA
9721	AATCAATCTA	AAGTATATAT	GAGTAAACTT	GGTCTGACAG	TTACCAATGC	TTAATCAGTG
9781	AGGCACCTAT	CTCAGCGATC	TGTCTATTTC	GTTCACTCCAT	AGTTGCCTGA	CTCCCCGTCG
9841	TGTAGATAAC	TACGATACGG	GAGGGCTTAC	CATCTGGCCC	CAGTGCTGCA	ATGATACCGC
9901	GAGACCCACG	CTCACCGGCT	CCAGATTTAT	CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG
9961	AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	CCTCCATCCA	GTCTATTAAT	TGTTGCCGGG
10021	AAGCTAGAGT	AAGTAGTTCG	CCAGTTAATA	GTTTGCGCAA	CGTTGTTGCC	ATTGCTACAG
10081	GCATCGTGGT	GTCACGCTCG	TCGTTTGCTA	TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT

```
10141 CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC
10201 CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTATCACT CATGGTTATG GCAGCACTGC
10261 ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTC TGTGACTGGT GAGTACTCAA
10321 CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC
10381 GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT
10441 CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTGAGATC CAGTTCGATG TAACCCACTC
10501 GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA
10561 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA
10621 TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT
10681 ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA
10741 AAGTGCCACC TGACGCGCCC TGTAGCGGCG CATTAAAGCGC GGCGGGTGTG GTGGTTACGC
10801 GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCGC TCCTTTCGCT TTCTTCCCTT
10861 CCTTCTCGC CACGTTCCGC GG
      (SEQ ID NO: 6)
```

图 4

原病毒区域的DNA序列(7780 bp)

特征	位置(bp)
SIN LTR	1-173 (互补链)
DHSIII 增强子	549-2200
OV启动子	2200-3331
内含子	3379-4967
hLAL CDS	4985-6184
OV 3'UTR	6199-6872
部分gag CDS	6884-7137 (互补链)
LTR	7435-7780 (互补链)

1	AATGAAGCCT	TCTGCTTCAT	GCATGTGCTC	GTAGTCGTCA	GGGAATCAAC	GGTCCGGCCA
61	TCAACCCAGG	TGCACACCAA	TGTGGTGAAT	GGTCAAATGG	CGTTTATTGT	ATCGAGCTAG
121	GCACTTAAAT	ACAATATCTC	TGCAATGCGG	AATTCAAGTGG	TTCGTCCAAT	CCGTCCCCCT
181	CCCTATGCAA	AAGCGAACT	ACTATATCCT	GAGGGGACTC	CTAACCGCGT	ACAACCGAAG
241	CCCCGCTTTT	CGCCTAAACA	TGCTATTGTC	CCCTCAGTCA	AGCCTTGCCC	GTTACAACCC
301	GATTTCGCAAG	CCTTGCCCTC	CCCACATTAT	CCGTAGCATT	ATTTCTTAGC	AGTCATCAGA
361	GCTACAGAAG	ATACTCTATG	CTGTAGCCAA	GTCTACAAGT	TTACTATTCA	GCGACCTCCT
421	ATATTCCGCG	TGCCAGCCGA	TCAATTACCA	ATCCAACCAG	CTATCACACG	GAATACAAGA
481	ACTCGCCTAC	GCTCTTCTTT	CGGGCTGCTT	ATAAGCCTCC	TGTAATTTTT	TTATATTCTT
541	CGCTCGAGTC	TCTTCAGAAT	GGCACAGCAC	CGCTGCAGAA	AAATGCCAGG	TGGACTATGA
601	ACTCACATCC	AAAGGAGCTT	GACCTGATAC	CTGATTTTCT	TCAAACAGGG	GAAACAACAC
661	AATCCCACAA	AACAGCTCAG	AGAGAAACCA	TCACTGATGG	CTACAGCACC	AAGGTATGCA
721	ATGGCAATCC	ATTCGACATT	CATCTGTGAC	CTGAGCAAAA	TGATTATCT	CTCCATGAAT
781	GGTTGCTTCT	TTCCCTCATG	AAAAGGCAAT	TCCACACTC	ACAATATGCA	ACAAAGACAA
841	ACAGAGAACA	ATTAATGTGC	TCCTTCCTAA	TGTTAAAATT	GTAGTGGCAA	AGAGGAGAAC
901	AAAATCTCAA	GTTCTGAGTA	GGTTTTAGTG	ATTGGATAAG	AGGCTTGAC	CTGTGAGCTC
961	ACCTGGACTT	CATATCCTTT	TGGATAAAAA	GTGCTTTTAT	AAC'TTCAGG	TCTCCGAGTC
1021	TTTATTCATG	AGACTGTTGG	TTTAGGGACA	GACCCACAAT	GAAATGCCTG	GCATAGGAAA
1081	GGGCAGCAGA	GCCTTAGCTG	ACCTTTTCTT	GGGACAAGCA	TTGTCAAACA	ATGTGTGACA
1141	AAACTATTTG	TACTGCTTTG	CACAGCTGTG	CTGGGCAGGG	CAATCCATTG	CCACCTATCC
1201	CAGGTAACCT	TCCAAC TGCA	AGAAGATTGT	TGCTTACTCT	CTCTAGACCC	CCAAGTCAAA
1261	CCAAC TATGC	AGGTATGCTG	ACAACACTAT	GATGACAGCC	TGTTC TGATC	AAGATCTCAT
1321	TTGTT CATGG	ACAATTTT TG	TTGCTTGCAG	CTGGTCTTCC	ATTGGGAAAG	AGTGTAGTAT
1381	ATCCTTCTCA	TCTGACAGAA	AAGCAGAAAT	TCTCATGCTC	CACACTTAAT	CTACATTGTT
1441	TTAAACCAAC	GGCTACTTCT	TGGAGAGGAA	AAATGGCTTT	TATAAGACTC	ACAAAACAAA
1501	GCTCTGCAAG	TCAAATGCAT	ACAAAAC TGT	TCTGTAGGTC	TGGAATCAGG	ACACTATGTG
1561	GAAGTCAAAT	AGAGCAGCTT	TAAAAAGCCT	TTGGGATCAT	TCTCATCTTA	TATTTGCAGC
1621	ACGATACTAT	GACAGTGATA	ACTGACATAA	CTGCATCAAT	TTCTTGATA	TTTTATT TGT
1681	CTTAAAGTAC	AAGACATAGA	GATGGACGTA	AAGATGGACA	TATGACTCAG	GTCTGGACAG
1741	GTCCGTGGTC	CATGTATGAT	AAAAGAGATG	AAGGGAAGGA	GAATTGAGAC	TGTCTAAGAA
1801	GGGCTTCAGG	GACGTTC TGA	AGGCAGATTT	GACTGAATCA	GATGTACTGT	CCAAGTCTCA
1861	TATGTAGCAA	TGGAAGGCTG	ATATTGGAGA	AATATAAAGA	AATGGCTGTG	AACTCAAAGT
1921	GACCTGAAC	AGAAAAGGGA	TATGGAGTTA	AAATAATGTC	ACAGAACTGA	GGTTTATATG
1981	ATATACCATG	GGCTGCAGAG	GGTCAGAGTG	CTCCACCATG	GGCCTCTCTT	GGGCTGCAGG
2041	GAACTTCTGT	TCTACACCTG	GAACACCTCC	TGCCCTCCTC	CGCACTGACC	TCAGTGT CAT
2101	CAGGGCTGTT	TCTCTCAGAT	TTTCTCACTC	ACCTCTCCCA	ACTACCATTG	TACAGCAGTT
2161	GTTCTTACAT	ATTGCTCCTC	CTGAGGTACA	TCTAGCATCG	TTAAGTCCTC	AGACTTGGCA
2221	AGGAGAA TGT	AGATTTCCAC	AGTATATATG	TTTTCACAAA	AGGAAGGAGA	GAAACAAAAG
2281	AAAATGGCAC	TGACTAAACT	TCAGCTAGTG	GTATAGGAAA	GTAATTCTGC	TTAACAGAGA
2341	TTGCAGTGAT	CTCTATGTAT	GTCTGAAGA	ATTATGTTGT	ACTTTTTTCC	CCCATTTTTA
2401	AATCAACAG	TGCTTTACAG	AGGTCAGAAT	GGTTTCTTTA	CTGTTTG TCA	ATTCTATTAT
2461	TTCAATACAG	AACAATAGCT	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG
2521	TCCCTCGAAC	CATGAACACT	CCTCCAGCTG	AATTTACAAA	TTCTCTGTG	ATCTGCCAGG
2581	CCATTAAGTT	ATTCATGGAA	GATCTTTGAG	GAACACTGCA	AGTTCATATC	ATAAACACAT
2641	TTGAAAT TGA	GTATTGTTTT	GCATTGTATG	GAGCTATGTT	TTGCTGTATC	CTCAGAAATA
2701	AAGTTTGTTA	TAAAGCATTC	ACACCCATAA	AAAGATAGAT	TTAAATATTC	CAACTATAGG
2761	AAAGAAAGTG	TGTCTGCTCT	TCACTCTAGT	CTCAGTTGGC	TCCTTCACAT	GCACGCTTCT
2821	TTATTTCTCC	TATTTTGTCA	AGAAAATAAT	AGGTCAAGTC	TTGTTCTCAT	TTATGTCTGT

2881	TCTAGCGTGG	CTCAGATGCA	CATTGTACAT	ACAAGAAGGA	TCAAATGAAA	CAGACTTCTG
2941	GTCTGTTACT	ACAACCATAG	TAATAAGCAC	ACTAACTAAT	AATTGCTAAT	TATGTTTTCC
3001	ATCTCCAAGG	TTCCACATT	TTTCTGTTT	CTTAAAGATC	CCATTATCTG	GTTGTAAGTG
3061	AAGCTCAATG	GAACATGAGC	AATATTCCC	AGTCTTCTCT	CCCATCCAAC	AGTCCTGATG
3121	GATTAGCAGA	ACAGGCAGAA	AACACATTGT	TACCCAGAAT	TAAAACTAA	TATTTGCTCT
3181	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAATCTAA	CCCAATCCCA	TTAAATGATT
3241	TCTATGGTGT	CAAAGGTCAA	ACTTCTGAAG	GGAACCTGTG	GGTGGGTCAC	AATTCAGACT
3301	ATATATTCCC	CAGGGCTCAG	CCAGTGTCTG	TACATACAGC	TAGAAAGCTG	TATTGCCTTT
3361	AGCAGTCAAG	CTCGAAAGGT	AAGCAACTCT	CTGGAATTAC	CTTCTCTCTA	TATTAGCTCT
3421	TACTTGACAC	TAAACTTTAA	AAATTTAACA	ATTATTGTGC	TATGTGTTGT	ATCTTTAAGG
3481	GTGAAGTACC	TGCGTGATAC	CCCCATAAAA	AACTTCTCAC	CTGTGTATGC	ATTCTGCACT
3541	ATTTTATTAT	GTGTAAAAGC	TTTGTGTTTG	TTTTCAGGAG	GCTTATTCTT	TGTGCTTAAA
3601	ATATGTTTTT	AATTTTCAGAA	CATCTTATCC	TGTCGTTTAC	TATCTGATAT	GCTTTGCACT
3661	TTGCTTGATT	AACTTCTAGC	CCTACAGAGT	GCACAGAGAG	CAAAATCATG	GTGTTCACTG
3721	AATTCTGGGG	AGTTATTTTA	ATGTGAAAAT	TCTCTAGAAG	TTTAATTCCT	GCAAAGTGCA
3781	GCTGCTGATC	ACTACACAAG	ATAAAAATGT	GGGGGGTGCA	TAAACGTATA	TTCTTACAA
3841	AATAGATACA	TGTGAACTTA	TATACAGAAA	AGAAAATGAG	AAAAATGTGT	GTGTGTATAC
3901	TCACACACGT	GGTCAGTAAA	AACTTTTGAG	GGGTTTAAAT	CAGAAAATCC	AATCCTGAGG
3961	CCCCAGCAGT	CAGTACGCAT	ATAAAGGGCT	GGGCTCTGAA	GGACTTCTGA	CTTTCACAGA
4021	TTATATAAAT	CTCAGGAAAG	CAACTAGATT	CATGCTGGCT	CCAAAAGCTG	TGCTTTATAT
4081	AAGCACACTG	GCTATACAAT	AGTTGTACAG	TTCAGCTCTT	TATAATAGAA	ACAGACAGAA
4141	CAAGTATAAA	TCTTCTATTG	GTCTATGTCA	TGAACAAGAA	TTCATTGAGT	GGCTCTGTTT
4201	TATAGTAAAC	ATTGCTATTT	TATCATGTCT	GCATTTCTCT	TCTGTCTGAA	TGTCACCACT
4261	AAAATTTAAT	TCCACAGAAA	GTTTATACTA	CAGTACACAT	GCATATCTTT	GCAATAAGCA
4321	AACCATACCT	GAAAGTGCAA	TAGAGCAGAA	TATGAATTAC	ATGCGTGTCT	TTCTCTAGTA
4381	CTACATGACC	CCATATAAAT	TACATTCCTT	ATCTATTCTG	CCATCACCAA	AACAAAGGTA
4441	AAAATACCTT	TGAAGATCTA	CTCATAGCAA	GTAGTGTGCA	ACAAACAGAT	ATTTCTCTAC
4501	ATTTATTTTT	AGGGAATAAA	AATAAGAAAT	AAAATAGTCA	GCAAGCCTCT	GCTTTCATAT
4561	ATATCTGTCC	AAACCTAAAG	TTTACTGAAA	TTTGCTCTTT	GAATTTCCAG	TTTTGCAAGC
4621	CTATCAGATT	GTGTTTTAAT	CAGAGGTACT	GAAAAGTATC	AATGAATTC	AGCTTTCCTT
4681	GAACAAAAT	ATGTAGAGGC	AACTGGTCTC	TGGGACAGTT	TGCTACCCAA	AAGACACTG
4741	AATGCAAATA	CATAAATAGA	TTTATGAATA	TGGTTTTGAA	CATGCACATG	AGAGGTGGAT
4801	ATAGCAACAG	ACACATTACC	ACAGAATTAC	TTTAAACTA	CTTGTTAACA	TTTAATTGCC
4861	TAAAACTGC	TCGTAATTTA	CTGTTGTAGC	CTACCATAGA	GTACCCTGCA	TGGTACTATG
4921	TACAGCATT	CATCCTTACA	TTTTCACTGT	TCTGCTGTTT	GCTCTAGACA	ACTCAGAGTT
4981	CACCATGAAA	ATGCGGTTCT	TGGGGTTGGT	GGTCTGTTTG	GTTCTCTGGA	CCCTGCATTC
5041	CGAGGGGTCC	GGAGGGAAAC	TGACAGCTGT	GGATCCAGAA	ACAAACATGA	ATGTCAGTGA
5101	AATTATCTCT	TACTGGGGAT	TCCCTAGTGA	GGAATACCTA	GTTGAGACAG	AGATTGGATA
5161	TATTCTGTGC	CTTAACCGAA	TTCTCATGG	GAGGAAGAAC	CATTCTGACA	AAGGTCCCAA
5221	ACCAGTTGTC	TTCTTGCAAC	ATGGCTTGCT	GGCAGATTCT	AGTAACTGGG	TCACAAACCT
5281	TGCCAACAGC	AGCCTGGGCT	TCATTCCTGC	TGATGCTGGT	TTTGACGTGT	GGATGGGCAA
5341	CAGCAGAGGA	AATACCTGGT	CTCGGAAACA	TAAGACACTC	TCAGTTTCTC	AGGATGAATT
5401	CTGGGCTTTC	AGTTATGATG	AGATGGCAAA	ATATGACCTA	CCAGCTTCCA	TTAATTCAT
5461	TCTGAATAAG	ACTGGCCAAG	AACAAGTGTA	TTATGTGGGT	CATTCTCAAG	GCACCACTAT
5521	AGGTTTTATA	GCATTTTCAC	AGATCCCTGA	GCTGGCTAAA	AGGATTAAAA	TGTTTTTTGC
5581	CCTGGGTCTC	GTGGCTTCCG	TCGCCTCTCG	TACTAGCCCT	ATGGCCAAAC	TGGGACGACT
5641	GCCAGATCAT	CTCATTAAGG	ACCTCTTTGG	AGACAAAGAA	TTTCTTCCCC	AGAGTGGGTT
5701	TTTGAAGTGG	CTGGGTACCC	ACGTTTGAC	TCATGTCATA	CTGAAGGAGC	TCTGTGGAAA
5761	TCTCTGTTTT	CTTCTGTGTG	GATTTAATGA	GAGAAATTTA	AATATGTCTA	GAGTGGATGT
5821	GTATACAAAC	CATTCTCCTG	CTGGAACCTC	TGTGCAAAAC	ATGTTACACT	GGAGCCAGGC
5881	TGTTAAATTC	CAAAAGTTTC	AAGCCTTTGA	CTGGGGAAGC	AGTGCCAAGA	ATTATTTTCA
5941	TTACAACCAG	AGTTATCCTC	CCACATACAA	TGTGAAGGAC	ATGCTTGTGC	CGACTGCAGT
6001	CTGGAGCGGG	GGTCACGACT	GGCTTGACAG	TGTCTACGAC	GTCAATATCT	TACTGACTCA
6061	GATCACCAAC	TTGGTGTTCC	ATGAGAGCAT	TCCGGAATGG	GAGCATCTTG	ACTTCATTTG
6121	GGGCCTGGAT	GCCCCTTGGA	GGCTTTATAA	TAAGATTATT	AATCTAATGA	GGAAATATCA
6181	GTGATTCGAA	GCGGCCGCAA	GAAGAAAGCT	GAAAACTCT	GTCCCTTCCA	ACAAGACCCA
6241	GAGCACTGTA	GTATCAGGGG	TAAAATGAAA	AGTATGTTAT	CTGCTGCATC	CAGACTTCAT
6301	AAAAGCTGGA	GCTTAATCTA	GAAAAAAAT	CAGAAAGAAA	TTACACTGTG	AGAACAGGTG
6361	CAATTCACTT	TTCTTTTACA	CAGAGTAATA	CTGGTAACCT	ATGGATGAAG	GCTTAAGGGA
6421	ATGAAATTGG	ACTCACAGTA	CTGAGTCATC	ACACTGAAAA	ATGCAACCTG	ATACATCAGC
6481	AGAAGGTTTA	TGGGGGAAAA	ATGCAGCCTT	CCAATTAAGC	CAGATATCTG	TATGACCAAG

```

6541 CTGCTCCAGA ATTAGTCACT CAAAATCTCT CAGATTAAAT TATCAACTGT CACCAACCAT
6601 TCCTATGCTG ACAAGGCAAT TGCTTGTTCT CTGTGTTCC TATCTACAA GGCTCTTCCT
6661 GACTTCCTAA AGATGCATTA TAAAAATCTT ATAATTCACA TTTCTCCCTA AACTTTGACT
6721 CAATCATGGT ATGTTGGCAA ATATGGTATA TTACTATTCA AATTGTTTTT CTTGTACCCA
6781 TATGTAATGG GTCTTGTGAA TGTGCTCTTT TGTTCTTTA ATCATAATAA AAACATGTTT
6841 AAGCAAACAC TTTTCACTTG TAGTATTTGA AGGTACCGGA TCTCGAGCCG CCTTCAATGC
6901 CCCCAAAACC AATCCCAGG TTTTAACTC TCCCGATTTT CCAAGTACCA TAGCCCGCTG
6961 AGAGAGCGCC GCGGTAATGG GATCCCAGGA CCCCAGGGAA TATAAGTCTG AGGGGGACGT
7021 AAGCAACCC TCTTTTGTA ACAGGGACAA CATAGCCCCT ATTTCTTCT TAGAAGGAGA
7081 GGTTCCTCCG CAATAGGTCT TACACGCGGA CGAAATCACC TTTATGACGG CTTCCATGCT
7141 TGATCCACCG GGCGACCGGA ATCACGCAGA GCAACCGGAA TCACGCCTGG GGTGGACCGC
7201 TCAGTCGTCG GGCTTCCTTC CCGTCTTCCA ACGACTCTCT GAGTTCTCGG TAGGGTATGT
7261 TGGCCCCCTG CAGTAGGGCT CCCTCCGACG CCACTCAGCT TCTGCCCTCC TAAGCCGCAG
7321 CCCCCTCTAC TAGGGTCATC GTCCGCTCCC CGAATAAGCG AGACGGATGA GGACAGGATC
7381 GCCACGCCGC CTGTGGCCGA CCACTATTC CTAACGATCA CGTCGGGGTC ACCAAATGAA
7441 GCCTTCGTCT TCATGCATGT GCTCGTAGTC GTCAGGGAAT CAACGGTCCG GCCATCAACC
7501 CAGGTGCACA CCAATGTGGT GAATGGTCAA ATGGCGTTTA TTGTATCGAG CTAGGCACCT
7561 AAATACAATA TCTCTGCAAT GCGGAATTCA GTGGTTCGTC CAATCCGTGT TAGACCCGTC
7621 TGTTGCCTTC CTAACAAGGC ACGATCATA CACGATCATA CCACCTTACT CCCACCAATC
7681 GGCATGCACG GTGCTTTTTC TCTCCTTATA AGGCATGTTG CTAATCATC GTTACATAAG
7741 CATGTTGCAA GACTACAAGA GTATTGCATA AGACTACATT

```

(SEQ ID NO:7)

图 5

特征	位置(bp)
填充片段	1-2734
OV 3'UTR	2750-3423
部分gag CDS	3435-3688 (互补链)
LTR	3986-4331 (互补链)
氟苄青霉素抗性基因	5795-6652 (互补链)
SIN LTR	7434-7606 (互补链)
OV启动子	7975-9106
内含子	9154-10742
卵清蛋白的ATG或POI	10760-10762

80

2821	AAGTATGTTA	TCTGCTGCAT	CCAGACTTCA	TAAAAGCTGG	AGCTTAATCT	AGAAAAAAAA
2881	TCAGAAAGAA	ATTACACTGT	GAGAACAGGT	GCAATTCACT	TTTCCTTTAC	ACAGAGTAAT
2941	ACTGGTAAC	CATGGATGAA	GGCTTAAGGG	AATGAAATTG	GACTCACAGT	ACTGAGTCAT
3001	CACACTGAAA	AATGCAACCT	GATACATCAG	CAGAAGGTTT	ATGGGGGAAA	AATGCAGCCT
3061	TCCAATTAA	CCAGATATCT	GTATGACCAA	GCTGCTCCAG	AATTAGTCAC	TCAAATCTC
3121	TCAGATTAAA	TTATCAACTG	TCACCAACCA	TTCCCTATGCT	GACAAGGCAA	TTGCTTGTTT
3181	TCTGTGTTCC	TGATACTACA	AGGCTCTTCC	TGACTTCCTA	AAGATGCATT	ATAAAAAATCT
3241	TATAATTAC	ATTTCTCCCT	AAACTTTGAC	TCAATCATGG	TATGTTGGCA	AATATGGTAT
3301	ATTACTATTC	AAATGTGTTT	CCTTGTAACC	ATATGTAATG	GGTCTGTGTA	ATGTGCTCTT
3361	TTGTTCCCTT	AATCATAATA	AAAACATGTT	TAAGCAAACA	CTTTTCACTT	GTAGTATTTG
3421	AAGGTACCGG	ATCTCGAGCC	GCCTTCAATG	CCCCCAAAC	CAATCCCCAG	GTTTTTAACT
3481	CTCCCGATTT	TCCAAGTACC	ATAGCCCCTG	GAGAGAGCGC	CGCGGTAATG	GGATCCCAGG
3541	ACCCCGGGGA	ATATAAGTCT	GAGGGGGACG	TAAGCAACCC	TTCCCTTTTG	AACAGGGACA
3601	ACATAGCCCC	TATTTCCCTT	TTAGAAGGAG	AGGTTTTCCC	GCAATAGGTC	TTACACGCGG
3661	ACGAAATCAC	CTTTATGACG	GCTTCCATGC	TTGATCCACC	GGGCGACCGG	AATCACGCAG
3721	AGCAACCGGA	ATCACGCCTG	GGGTGGACCG	CTCAGTCGTC	GGGCTTCCTT	CCCGTCTTCC
3781	AACGACTCTC	TGAGTTCTCG	GTAGGGTATG	TTGGCCCCCT	GCAGTAGGGC	TCCCTCCGAC
3841	GCCACTCAGC	TTCTGCCCTC	CTAAGCCGCA	GCCCCCTCTA	CTAGGGTCAT	CGTCCGCTCC
3901	CCGAATAAGC	GAGACGGATG	AGGACAGGAT	CGCCACGCCG	CCTGTGGCCG	ACCCTATTTC
3961	CCTAACGATC	ACGTGCGGGT	CACCAAAATG	AGCCTTCTGC	TTCAATGCATG	TGCTCGTAGT
4021	CGTCAGGGAA	TCACCGGTCC	GGCCATCAAC	CCAGGTGCAC	ACCAATGTGG	TGAATGGTCA
4081	AATGGCGTTT	ATTGTATCGA	GCTAGGCACT	TAAATACAAT	ATCTCTGCAA	TGCGGAATTC
4141	AGTGGTTTCG	CCAATCCGTG	TTAGACCCGT	CTGTTGCCTT	CCTAACAAGG	CACGATCATA
4201	CCACGATCAT	ACCACCTTAC	TCCCACCAAT	CGGCATGCAC	GGTGTCTTTT	CTCTCCTTAT
4261	AAGGCATGTT	GCTAACTCAT	CGTTACATAA	GCATGTTGCA	AGACTACAAG	AGTATTGCAT
4321	AAGACTACAT	TTCCCCCTCC	CTATGCAAAA	GCGAAACTAC	TATATCCTGA	GGGGACTCCT
4381	AACCGCGTAC	AACCGAAGCC	CCGCTTTTCG	CCTAAACATG	CTATTGTCCC	CTCAGTCAAG
4441	CCTTGCCCCG	TACAACCCGA	TTGCGCAAGC	TTGCCCTCCC	CACATTATCC	GTAGCATTAT
4501	TTCTTAGCAG	TCATCAGAGC	TACAGAAGAT	ACTCTATGCT	GTAGCCAAAG	CTACAAGTTT
4561	ACTATTCAGC	GACCTCCTAT	ATTCCGCGTG	CCAGCCGATC	AATTACCAAT	GCGCGCTTGG
4621	CGTAATCATG	GTCATAGCTG	TTTCTGTGTG	GAAATTGTTA	TCCGCTCACA	ATTCCACACA
4681	ACATACGAGC	CGGAAGCATA	AAGTGTAAG	CCTGGGGTGC	CTAATGAGTG	AGCTAACTCA
4741	CATTAATTGC	GTTGCGCTCA	CTGCCCCGCT	TCCAGTCGGG	AAACCTGTCT	TGCCAGCTGC
4801	ATTAATGAA	CGGCCAAGCG	GCGGGGAGAG	GCGGTTTGCG	TATTGGGCGC	TCTTCCGCTT
4861	CCTCGCTCAC	TGACTCGCTG	CGCTCGGTGC	TTGCGCTGCG	GCGAGCGGTA	TCAGCTCACT
4921	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA	CGCAGGAAAG	AACATGTGAG
4981	CAAAAGGCCA	GCAAAAGGCC	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA
5041	GGCTCCGCCC	CCCTGACGAG	CATCACAATA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC
5101	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG
5161	TTCCGACCC	GCCGCTTACC	GGATACCTGT	CCGCTTTTCT	CCCTTCGGGA	AGCGTGGCGC
5221	TTTCTCATAG	CTCACGCTGT	AGGTATCTCA	GTTTCGGTGT	GGTCTGTCTG	TCCAAGCTGG
5281	GCTGTGTGCA	CGAACCCCCC	GTTACGCCCC	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC
5341	TTGAGTCCAA	CCCGGTAAGA	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA
5401	TTAGCAGAGC	GAGGTATGTA	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG
5461	GCTACACTAG	AAGGACAGTA	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA
5521	AAAGAGTTGG	TAGCTCTTGA	TCCGGCAAA	AAACCACCGC	TGGTAGCGGT	GGTTTTTTTG
5581	TTTGCAAGCA	GCAGATTACG	CGCAGAAAA	AAGGATCTCA	AGAAGATCCT	TTGATCTTTT
5641	CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	GTCATGAGAT
5701	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	AAATCAATCT
5761	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	GAGGCACCTA
5821	TCTCAGCGAT	CTGTCTATTT	CGTTCATCCA	TAGTTGCCTG	ACTCCCCGTC	GTGTAGATAA
5881	CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	AATGATACCG	CGAGACCCAC
5941	GCTCACC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	CGGAAGGGCC	GAGCGCAGAA
6001	GTGGTCTTGC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	TTGTTGCCGG	GAAGCTAGAG
6061	TAAGTAGTTC	GCCAGTTAAT	AGTTTGCGCA	ACGTTGTTGC	CATTGCTACA	GGCATCGTGG
6121	TGTCACGCTC	GTCGTTTGGT	ATGGCTTCAT	TCAGCTCCGG	TTCCCAACGA	TCAAGGCGAG
6181	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGCTC	CTTCGGTCTT	CCGATCGTTG
6241	TCAGAAGTAA	GTTGGCCGCA	GTGTTATCAC	TCATGGTTAT	GGCAGCACTG	CATAATTCTC
6301	TTACTGTTCAT	GCCATCCGTA	AGATGCTTTT	CTGTGACTGG	TGAGTACTCA	ACCAAGTCAT
6361	TCTGAGAATA	GTGTATGCGG	CGACCCAGTT	GCTCTTGCCC	GGCGTCAATA	CGGGATAATA
6421	CCGCGCCACA	TAGCAGAACT	TTAAAAGTGC	TCATCATTTG	AAAACGTTCT	TCCGGGCGAA

6481	AACCTCTCAAG	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	GTAACCCACT	CGTGCACCCA
6541	ACTGATCTTC	AGCATCTTTT	ACTTTACCA	GCGTTTCTGG	GTGAGCAAAA	ACAGGAAGGC
6601	AAAAAGCCGC	AAAAAAGGGA	ATAAGGGCGA	CACGGAAATG	TTGAATACTC	ATACTCTTCC
6661	TTTTTCAATA	TTATTGAAGC	ATTTATCAGG	GTTATTGTCT	CATGAGCGGA	TACATATTTG
6721	AATGTATTTA	GAAAAATAAA	CAAATAGGGG	TTCCGCGCAC	ATTTCCCCGA	AAAGTGCCAC
6781	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	CGCAGCGTGA
6841	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTTCG	TTTCTTCCTT	TGCTTTCTCG
6901	CCACGTTTCG	CGGCTTTCCC	CGTCAAGCTC	TAAATCGGGG	GCTCCCTTTA	GGGTTCGGAT
6961	TTAGTGCTTT	ACGGCACCTC	GACCCCAAAA	AACTTGATTA	GGGTGATGGT	TCACGTAGTG
7021	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTTGACGTT	GGAGTCCACG	TTCTTTAATA
7081	GTGGACTCTT	GTTCCAAACT	GGAACAACAC	TCAACCCTAT	CTCGGTCTAT	TCTTTTGATT
7141	TATAAGGGAT	TTTGCCGATT	TCGGCCTATT	GGTTAAAAAA	TGAGCTGATT	TAACAAAAAT
7201	TTAACGCGAA	TTTTAACAAA	ATATTAACGC	TTACAATTTT	CATTCGCCAT	TCAGGCTGCG
7261	CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGGCAGC	TGGCGAAAGG
7321	GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG
7381	TAAAACGACG	GCCAGTGAGC	GCGTATTCCC	TAACGATCAC	GTCGGGGTCA	CCAAATGAAG
7441	CCTTCTGCTT	CATGCATGTG	CTCGTAGTCG	TCAGGGAATC	AACGGTCCGG	CCATCAACCC
7501	AGGTGCACAC	CAATGTGGTG	AATGGTCAAA	TGGCGTTTAT	TGTATCGAGC	TAGGCACTTA
7561	AATACAATAT	CTCTGCAATG	CGGAATTCAG	TGGTTCGTCC	AATCCGTCCC	CCTCCCTATG
7621	CAAAAGCGAA	ACTACTATAT	CCTGAGGGGA	CTCCTAACCG	CGTACAAACG	AAGCCCCGCT
7681	TTTCGCCTAA	ACATGCTATT	GTCCCTCAG	TCAAGCCTTG	CCCGTTACAA	CCCGATTTCG
7741	AAGCCTTGCC	CTCCCCACAT	TATCCGTAGC	ATTATTTTCT	AGCAGTCATC	AGAGCTACAG
7801	AAGATACTCT	ATGCTGTAGC	CAAGTCTACA	AGTTTACTAT	TCAGCGACCT	CCTATATTCC
7861	GCGTGCCAGC	CGATCAATTA	CCAAATCCAA	CAGCTATCAC	ACGGAATACA	AGAAGCTCGC
7921	TACGCTCTTC	TTTCGGGCTG	CTTATAAGCC	TCCTGTAATT	TTTTTATATT	CCTCGTTAAG
7981	TCTTCAGACT	TGGCAAGGAG	AATGTAGATT	TCCACAGTAT	ATATGTTTTT	ACAAAAGGAA
8041	GGAGAGAAAC	AAAAGAAAAA	GGCACTGACT	AAACTTCAGC	TAGTGGTATA	GGAAAGTAAT
8101	TCTGCTTAAC	AGAGATTGCA	GTGATCTCTA	TGTATGTCCT	GAAGAATTAT	GTTGTACTTT
8161	TTTCCCCCAT	TTTTAAATCA	AACAGTGCTT	TACAGAGGTC	AGAATGGTTT	CTTTACTGTT
8221	TGTCATTCTT	ATTATTTCAA	TACAGAACAA	TAGCTTCTAT	AACTGAAATA	TATTTGCTAT
8281	TGTATATTAT	GATTGTCCCT	CGAACCATGA	ACACTCCTCC	AGCTGAATTT	CACAATTCCT
8341	CTGTCATCTG	CCAGGCCATT	AAGTTATTCA	TGGAAGATCT	TTGAGGAACA	CTGCAAGTTC
8401	ATATCATAAA	CACATTTGAA	ATTGAGTATT	GTTTTGCATT	GTATGGAGCT	ATGTTTTGCT
8461	GTATCCTCAG	AATAAAAGTT	TGTTATAAAG	CATTCACACC	CATAAAAAGA	TAGATTTAAA
8521	TATTCCAAC	ATAGGAAAGA	AAGTGTGTCT	GCTCTTCACT	CTAGTCTCAG	TTGGCTCCTT
8581	CACATGCACG	CTTCTTTATT	TCTCCTATTT	TGTCAAGAAA	ATAATAGGTC	AAGTCTTGTT
8641	CTCATTTATG	TCCTGTCTAG	CGTGGCTCAG	ATGCACATTG	TACATACAAG	AAGGATCAAA
8701	TGAAACAGAC	TTCTGGTCTG	TTACTACAAC	CATAGTAATA	AGCACACTAA	CTAATAATTG
8761	CTAATTATGT	TTTCCATCTC	CAAGGTTCCT	ACATTTTCTT	GTTTTCTTAA	AGATCCCATT
8821	ATCTGGTTGT	AACTGAAGCT	CAATGGAAAC	TGAGCAATAT	TTCCAGTCTT	TCTCTCCCAT
8881	CCAACAGTCC	TGATGGATTA	GCAGAACAGG	CAGAAAACAC	ATTGTTACCC	AGAATTAAAA
8941	ACTAATATTT	GCTCTCCATT	CAATCCAAAA	TGGACCTATT	GAACTAAAA	TCTAACCCAA
9001	TCCCATTA	TGATTTCTAT	GGTGTCAAAG	GTCAAACCTC	TGAAGGGAAC	CTGTGGGTGG
9061	GTCACAATTC	AGACTATATA	TTCCCGAGGG	CTCAGCCAGT	GTCTGTACAT	ACAGCTAGAA
9121	AGCTGTATTG	CCTTTAGCAG	TCAAGCTCGA	AAGGTAAGCA	ACTCTCTGGA	ATTACCTTCT
9181	CTCTATATTA	GCTCTTACTT	GCACCTAAAC	TTTAAAAAAT	TAACAATTAT	TGTGCTATGT
9241	GTTGTATCTT	TAAGGGTGAA	GTACCTGCCT	GATACCCCTT	ATAAAAACCT	CTCACCTGTG
9301	TATGCATTCT	GCACTATTTT	ATTATGTGTA	AAAGCTTTGT	GTTTGTTTTC	AGGAGGCTTA
9361	TTCTTTGTGC	TTAAAATATG	TTTTTAATTT	CAGAACATCT	TATCCTGTCT	TTCACTATCT
9421	GATATGCTTT	GCAGTTTGCT	TGATTAACCT	CTAGCCCTAC	AGAGTGACAC	GAGAGCAAAA
9481	TCATGGTGTT	CAGTGAATTC	TGGGGAGTTA	TTTTAATGTG	AAAATTCTCT	AGAAGTTTAA
9541	TTCTGCAAAA	GTGCAGCTGC	TGATCACTAG	ACAAGATAAA	AATGTGGGGG	GTGCATAAAA
9601	GTATATTCTT	ACAATAATAG	ATACATGTGA	ACTTATATAC	AGAAAAGAAA	ATGAGAAAAA
9661	TGTGTGTGTG	TATACTCACA	CACGTGGTCA	GTAAAAACTT	TTGAGGGGTT	TAATACAGAA
9721	AATCCAATCC	TGAGGCCCCA	GCACTCAGTA	CGCATATAAA	GGGCTGGGCT	CTGAAGGACT
9781	TCTGACTTTC	ACAGATTATA	TAAATCTCAG	GAAAGCAACT	AGATTCATGC	TGGCTCCAAA
9841	AGCTGTGCTT	TATATAAGCA	CACCTGGCTAT	ACAATAGTTG	TACAGTTCAG	CTCTTTATAA
9901	TAGAAACAGA	CAGAACAAAGT	ATAAATCTTC	TATTGGTCTA	TGTCATGAAC	AAGAATTCAT
9961	TCAGTGGCTC	TGTTTTATAG	TAAACATTGC	TATTTTATCA	TGCTCGCAT	TCTCTTCTGT
10021	CTGAATGTCA	CCACTAAAAT	TTAACTCCAC	AGAAAAGTTA	TACTACAGTA	CACATGCATA
10081	TCTTTGAGCA	AAGCAAACCA	TACCTGAAAG	TGCAATAGAG	CAGAATATGA	ATTACATGCG

```

10141 TGTCTTTCTC CTAGACTACA TGACCCCATATA TAAATTACAT TCCTTATCTA TTCTGCCATC
10201 ACCAAAACPA AGGTAAAAAT ACTTTTGAAG ATCTACTCAT AGCAAGTAGT GTGCAACAAA
10261 CAGATATTTT TCTACATTTA TTTT TAGGGA ATAAAAATAA GAAATAAAAT AGTCAGCAAG
10321 CCTCTGCTTT CTCATATATC TGTCCAAACC TAAAGTTTAC TGAAATTTGC TCTTTGAATT
10381 TCCAGTTTTG CAAGCCTATC AGATTGTGTT TTAATCAGAG GTACTGAAAA GTATCAATGA
10441 ATTCTAGCTT TCACTGAACA AAAATATGTA GAGGCAACTG GCTTCTGGGA CAGTTTGCTA
10501 CCCAAAAGAC AACTGAATGC AAATACATAA ATAGATTTAT GAATATGGTT TTGAACATGC
10561 ACATGAGAGG TGGATATAGC AACAGACACA TTACCACAGA ATTACTTTAA AACTACTTGT
10621 TAACATTTAA TTGCCTAAAA ACTGCTCGTA ATTTACTGTT GTAGCCTACC ATAGAGTACC
10681 CTGCATGGTA CTATGTACAG CATTCCATCC TTACATTTTC ACTGTTCTGC TGTGCTCT
10741 AGACAACTCA GAGTTCACCA TG

```

(SEQ ID NO:8)

图 6

SBC102 逆体的DNA序列(242 bp)

特征	位置(bp)
来自 OV 基因	1-149
部分 hLAL CDS	150-242

```

1 CCCGGGTGTT TAACATTTAA TTGCCTAAAA ACTGCTCGTA ATTTACTGTT GTAGCCTACC
61 ATAGAGTACC CTGCATGGTA CTATGTACAG CATTCCATCC TTACATTTTC ACTGTTCTGC
121 TGTGTTGCTCT AGACAACTCA GAGTTCACCA TGAAAATGCG GTTCTTGGGG TTGGTGGTCT
181 GTTTGGTCTCT CTGGACCTG CATTCCGAGG GGTCCGAGG GAAACTGACA GCTGTGGATC
241 CT

```

(SEQ ID NO:9)

图 7

Syn SBC102 的DNA 序列 (1575 bp)

特征		位置(bp)				
部分OV启动子		1-355				
hLAL CDS		356-1575				
1	CCATTATCTG	GTTGTAAC	TG AAGCTCA	ATG GAACATG	AGC AATATTT	CCC AGTCTTCTCT
61	CCCATCCAAC	AGTCCTGAT	G GATTAGCA	GAC ACAGGCAG	AA AACACATT	GT TACCCAGAAT
121	TAAAAACTAA	TATTTGCTCT	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAATCTAA
181	CCCAATCCCA	TTAAATGATT	TCTATGGTGT	CAAAGGTCAA	ACTTCTGAAG	GGAACTCTGTG
241	GGTGGGTCAC	AATTCAGACT	ATATATTCCC	CAGGGCTCAG	CCAGTGTCTG	TACATACAGC
301	TAGAAAGCTG	TATTGCCCTT	AGCAGTCAAG	CTCGAAAGAC	AACTCAGAGT	TCACCATGAA
361	AATGCGGTTC	TTGGGGTTGG	TGGTCTGTTT	GGTCTCTGG	ACCCTGCATT	CCGAGGGGTC
421	CGGAGGGAAA	CTGACAGCTG	TGGATCCTGA	AACAAACATG	AATGTCAGTG	AAATTATCTC
481	TTACTGGGGA	TTCCCTAGTG	AGGAATACCT	AGTTGAGACA	GAAGATGGAT	ATATTCTGTG
541	CCTTAACCGA	ATTCCTCATG	GGAGGAAGAA	CCATTCTGAC	AAAGGTCCCA	AACCAAGTTGT
601	CTTCCTGCAA	CATGGCTTGC	TGGCAGATTG	TAGTAACTGG	GTCACAAACC	TTGCCAACAG
661	CAGCCTGGGC	TTCATTCTTG	CTGATGCTGG	TTTTGACGTG	TGGATGGGCA	ACAGCAGAGG
721	AAATACCTGG	TCTCGGAAAC	ATAAGACACT	CTCAGTTTCT	CAGGATGAAT	TCTGGGCTTT
781	CAGTTATGAT	GAGATGGCAA	AATATGACCT	ACCAGCTTCC	ATTAACCTCA	TTCTGAATAA
841	GACTGGCCAA	GAACAAGTGT	ATTATGTGGG	TCATTCTCAA	GGCACCCTA	TAGGTTTTAT
901	AGCATTTCAT	CAGATCCCTG	AGCTGGCTAA	AAGGATTAAA	ATGTTTTTTG	CCCTGGGTCC
961	TGTGGCTTCC	GTCGCCCTCT	GTACTAGCCC	TATGGCCAAA	CTGGGACGAC	TGCCAGATCA
1021	TCTCATTAA	GACCTCTTTG	GAGACAAAGA	ATTTCTTCCC	CAGAGTGCCT	TTTTGAAGTG
1081	GCTGGGTACC	CACGTTTGCA	CTCATGTGAT	ACTGAAGGAG	CTCTGTGGAA	ATCTCTGTTT
1141	TCTTCTGTGT	GGATTTAATG	AGAGAAATTT	AAATATGTCT	AGAGTGGATG	TGTATACAAC
1201	ACATTCTCCT	GCTGGAACCT	CTGTGCAAAA	CATGTTACAC	TGGAGCCAGG	CTGTTAAATT
1261	CCAAAAGTTT	CAAGCCTTTG	ACTGGGGAAG	CAGTGCCAAG	AATTATTTTC	ATTACAACCA
1321	GAGTTATCCT	CCCACATACA	ATGTGAAGGA	CATGCTTGTG	CCGACTGCAG	TCTGGAGCGG
1381	GGGTCACGAC	TGGCTTGACG	ATGTCTACGA	CGTCAATATC	TTACTGACTC	AGATCACCAA
1441	CTTGGTGTTC	CATGAGAGCA	TTCCGGAATG	GGAGCATCTT	GACTTCATTT	GGGGCCTGGA
1501	TGCCCCCTTG	AGGCTTTATA	ATAAGATTAT	TAATCTAATG	AGGAAATATC	AGTGATTCTGA
1561	AGCGGCCGCC	CCGGG				

(SEQ ID NO:10)

图 8

OVR1启动子的DNA序列(2789 bp)

特征		位置(bp)	
来自OV基因的DHSIII区域		7-1658	
OV启动子		1658-2789	
1	CTCGAGTCTC	TTCAGAATGG	CACAGCACCG
61	TCACATCCAA	AGGAGCTTGA	CCTGATACCT
121	TCCCACAAA	CAGCTCAGAG	AGAAACCATC
181	GGCAATCCAT	TCGACATTCA	TCTGTGACCT
241	TTGCTTCTTT	CCCTCATGAA	AAGGCAATTT
301	AGAGAACAAT	TAATGTGCTC	CTTCCTAATG
361	AATCTCAAGT	TCTGAGTAGG	TTTTAGTGAT
421	CTGGACTTCA	TATCCTTTTG	GATAAAAAGT
481	TATTCATGAG	ACTGTTGGTT	TAGGGACAGA
541	GCAGCAGAGC	CTTAGCTGAC	CTTTTCTTGG
601	ACTATTGTGA	CTGCTTTGCA	CAGCTGTGCT
661	GGTAACCTTC	CAACTGCAAG	AAGATTGTTG
721	AACATGTCAG	GTATGCTGAC	AACACTATGA
781	GTTTCATGGAC	AATTTTGTGT	GCTTGCAGCT
841	CCTTCTCATC	TGACAGAAAA	GCAGAAATTC
901	AAACCACCGG	CTACTTCTTG	GAGAGGAAAA
961	TCTGCAAGTC	AAATGCATAC	AAAACCTGTC
1021	AGTCAAAATAG	AGCAGCTTTA	AAAAGCCTTT
1081	GATACTATGA	CAGTGATAAC	TGACATAACT
1141	TAAAGTACAA	GACATAGAGA	TGGACGTAAG
1201	CCGTGGTCCA	TGTATGATAA	AAGAGATGAA
1261	GCTTCAGGGA	CGTTCTGAAG	GCAGATTGTA
1321	TGTAGCAATG	GAAGGCTGAT	ATTGGAGAAA
1381	CCCTGAACAG	AAAAGGGATA	TGGAGTTAAA
1441	ATACCATGGG	CTGCAGAGGG	TCAGAGTGCT
1501	ACTTCTGTTT	TACACCTGGA	ACACCTCCTG
1561	GGGCTGTTTC	TCTCACATTT	TCTCACTCAC
1621	TCTTACATAT	TGCTCCTCCT	GAGGTACATC
1681	GAGAATGTAG	ATTTCCACAG	TATATATGTT
1741	AATGGCACTG	ACTAAACTTC	AGCTAGTGGT
1801	GCAGTGATCT	CTATGTATGT	CCTGAAGAAT
1861	TCAAACAGTG	CTTTACAGAG	GTCAGAAATG
1921	CAATACAGAA	CAATAGCTTC	TATAACTGAA
1981	CCTCGAACCA	TGAACACTCC	TCCAGCTGAA
2041	ATTAAGTTAT	TCATGGAAGA	TCTTTGAGGA
2101	GAAATTGAGT	ATTGTTTTGC	ATTGTATGGA
2161	GTTTGTATATA	AAGCATTCAC	ACCCATAAAA
2221	AGAAAGTGTG	TCTGCTCTTC	ACTCTAGTCT
2281	ATTTCTCCTA	TTTTGTCAAG	AAAATAATAG
2341	TAGCGTGGCT	CAGATGCACA	TTGTACATAC
2401	CTGTTACTAC	AACCATAGTA	ATAAGCACAC
2461	CTCCAAGGTT	CCCACATTTT	TCTGTTTTCT
2521	GCTCAATGGA	ACATGAGCAA	TATTTCCCAG
2581	TTAGCAGAAC	AGGCAGAAAA	CACATTGTTA
2641	ATTCAATCCA	AAATGGACCT	ATTGAAACTA
2701	TATGGTGTCA	AAGGTCAAAC	TTCTGAAGGG
2761	ATATTCCCCA	GGGCTCAGCC	AGTGTCTGT

(SEQ ID NO: 11)

图 9

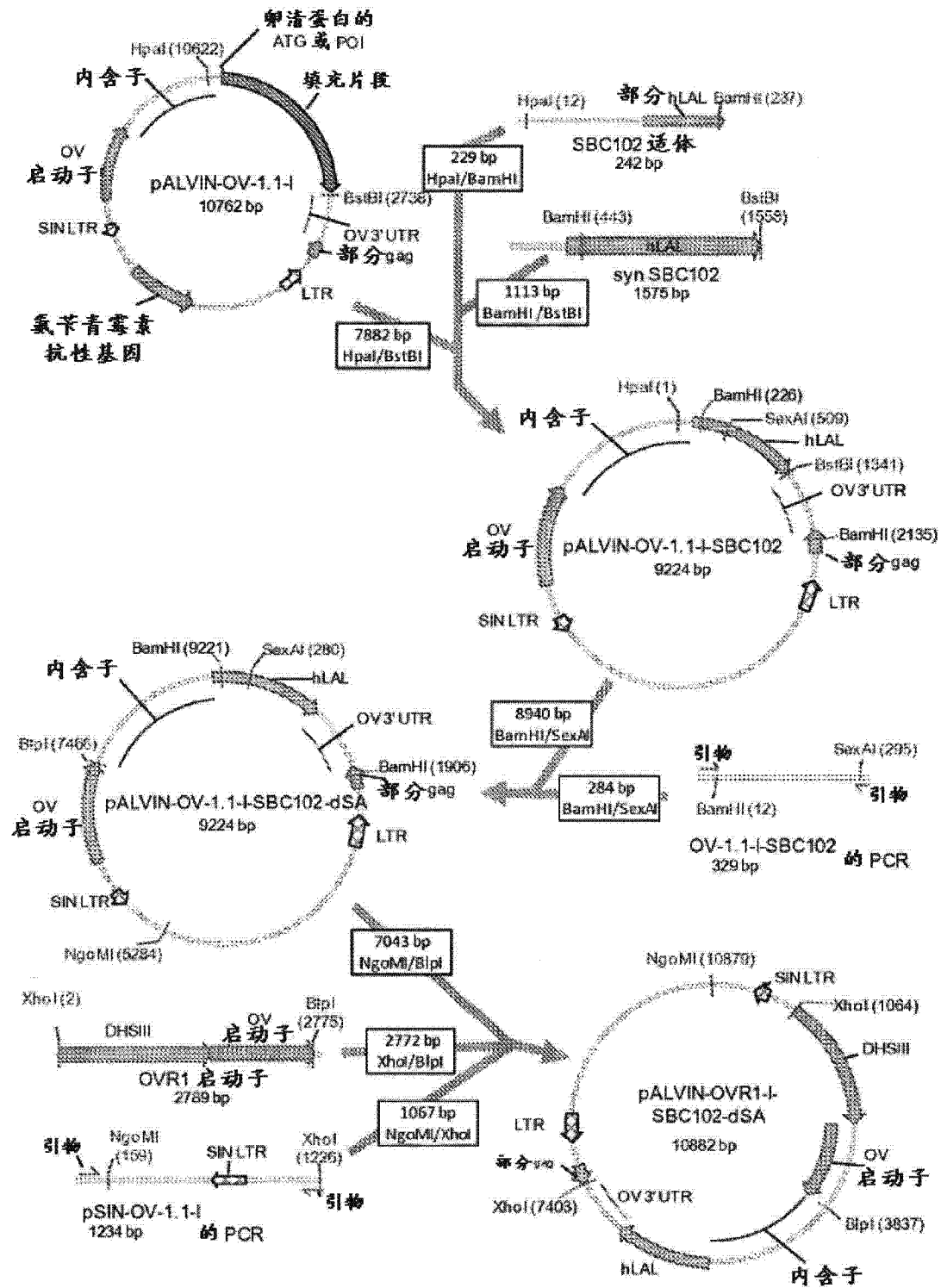


图 10

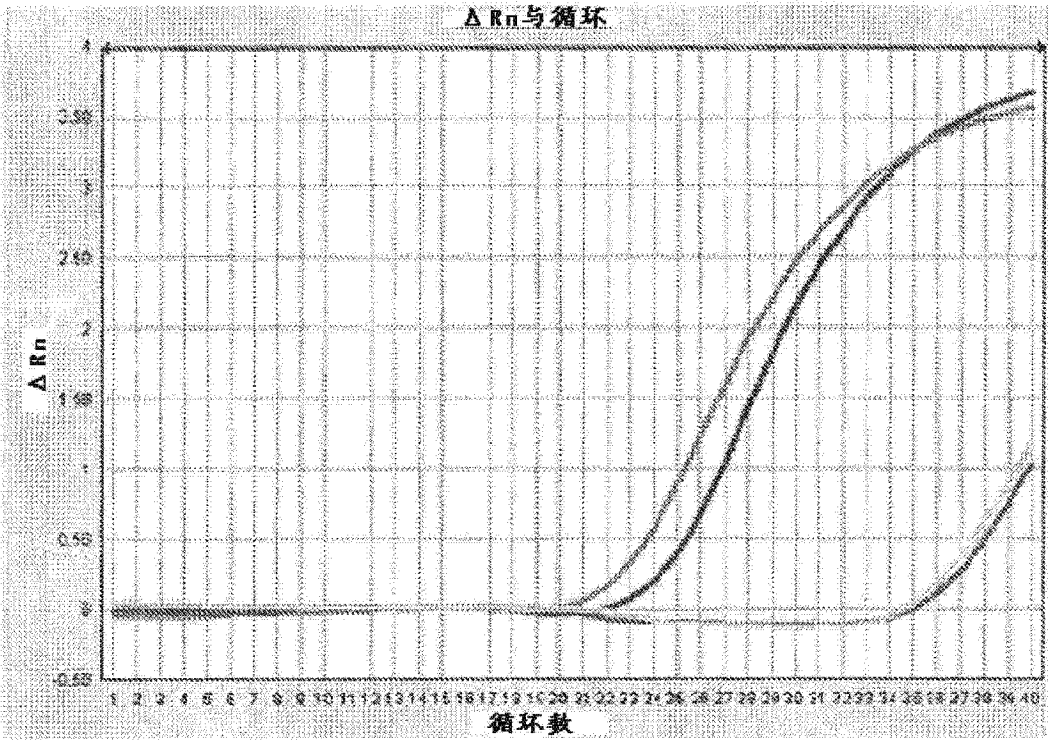


图 11

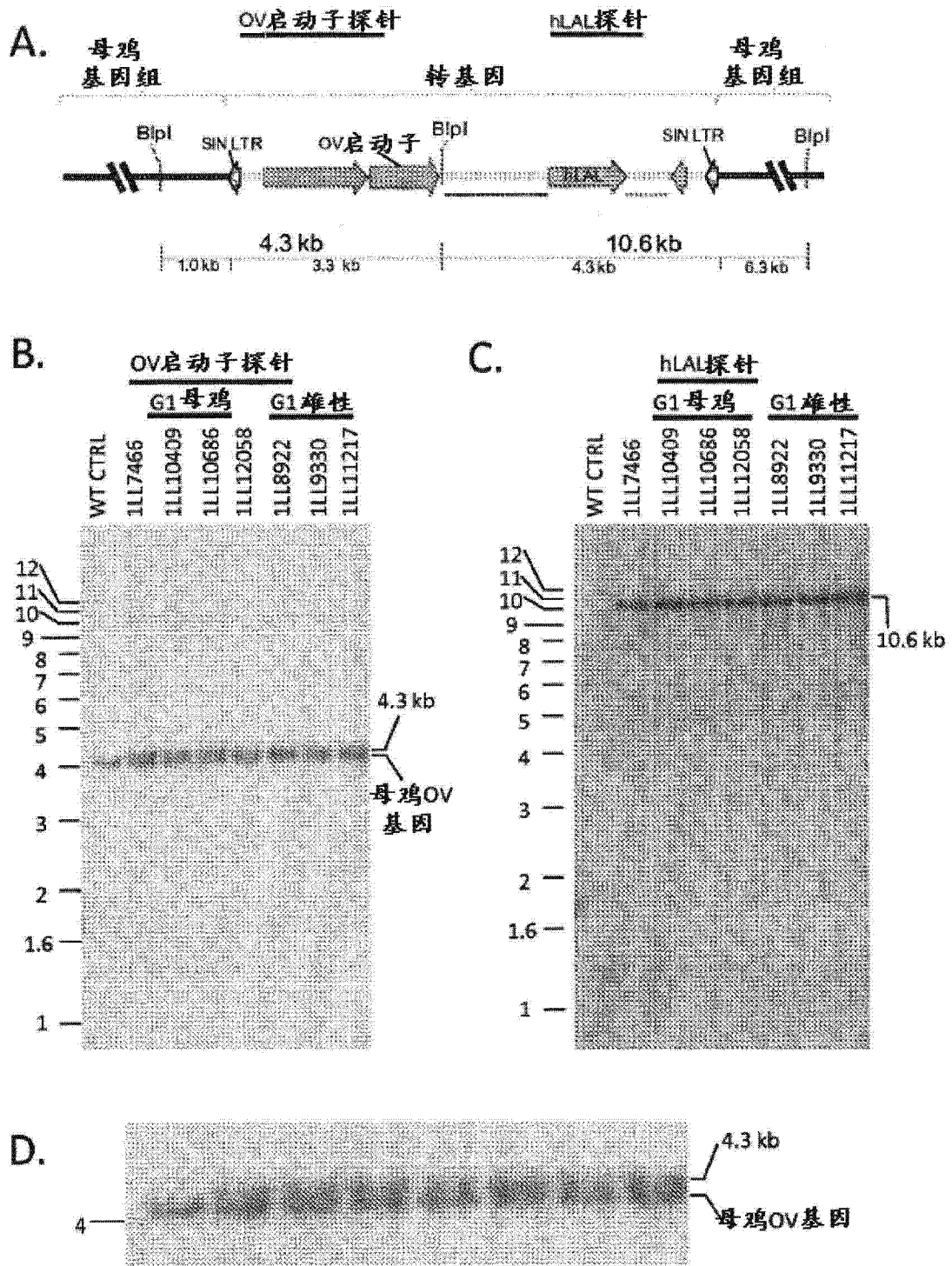


图 12

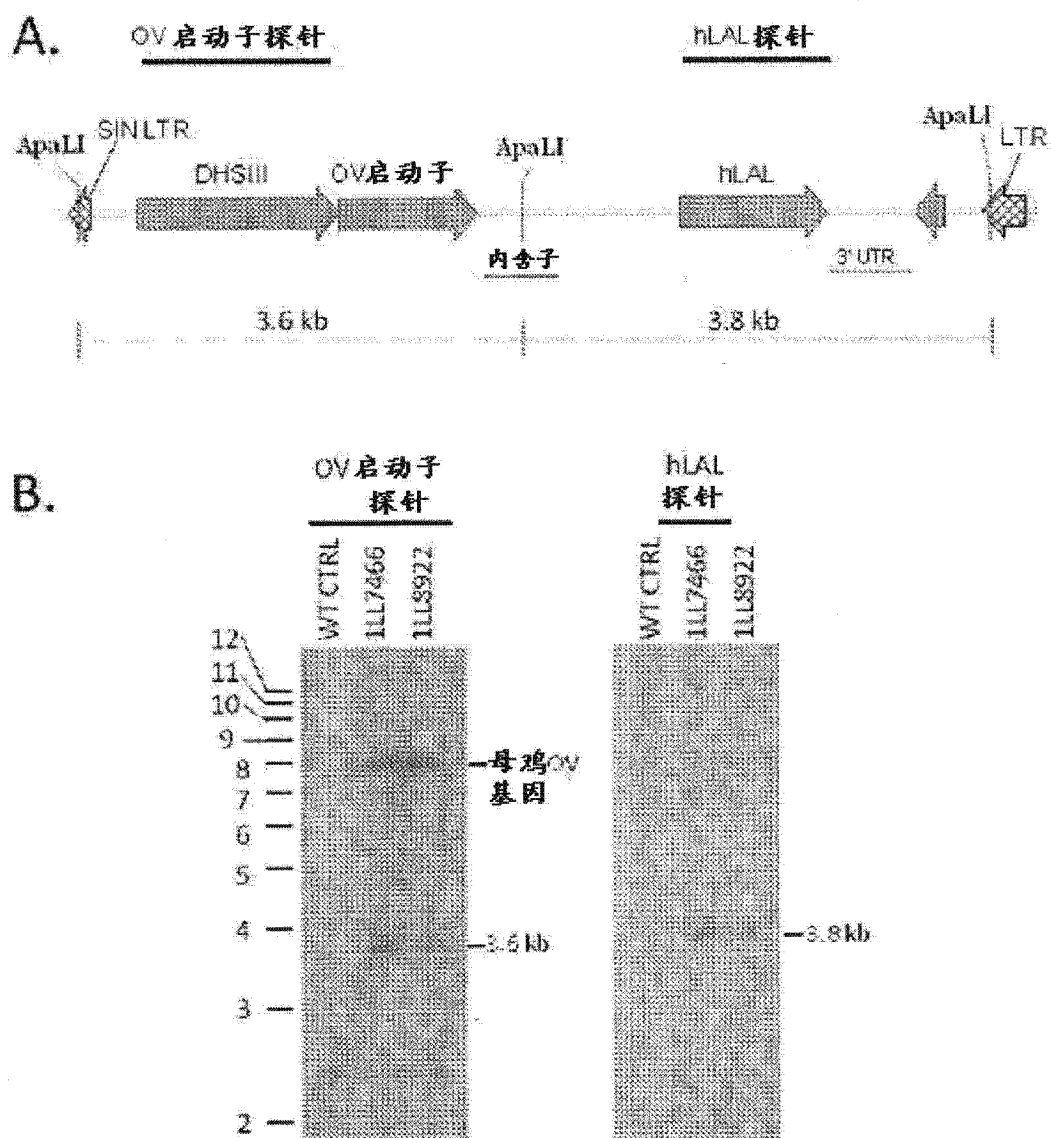


图 13

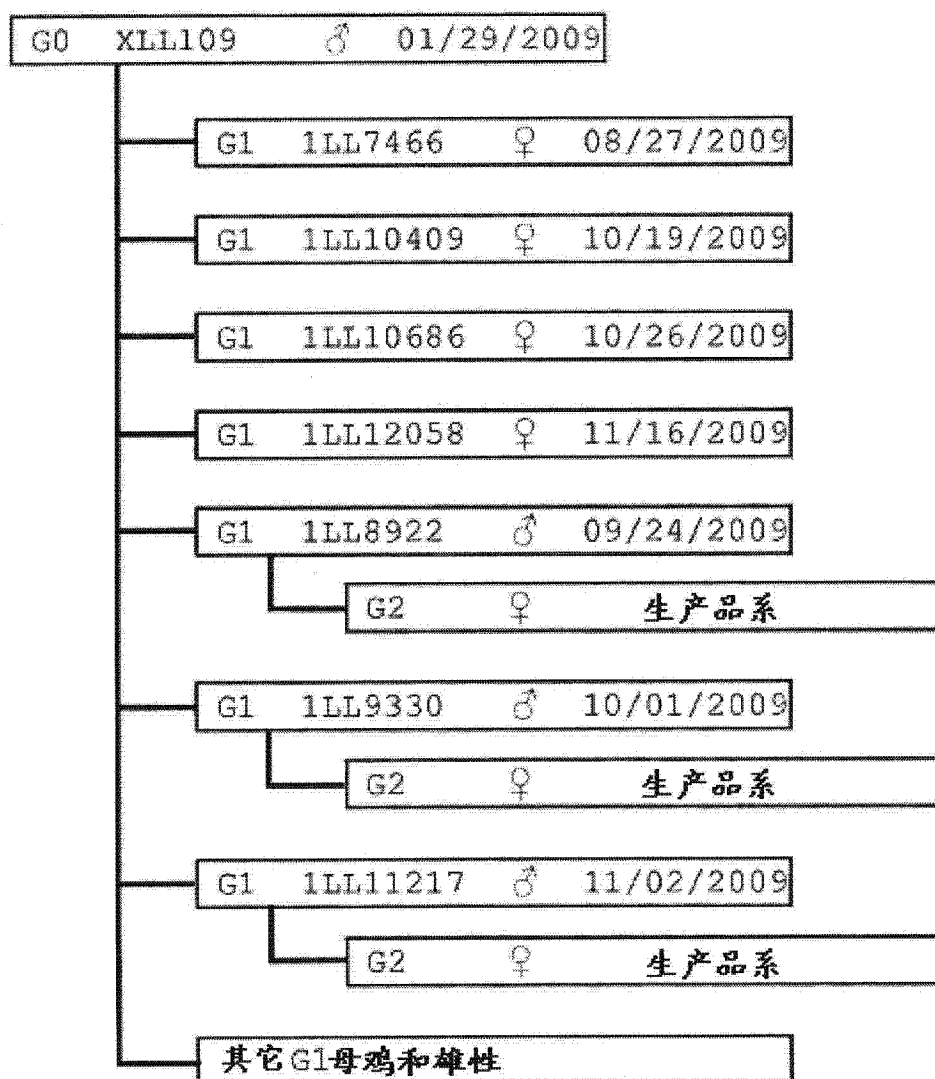


图 14

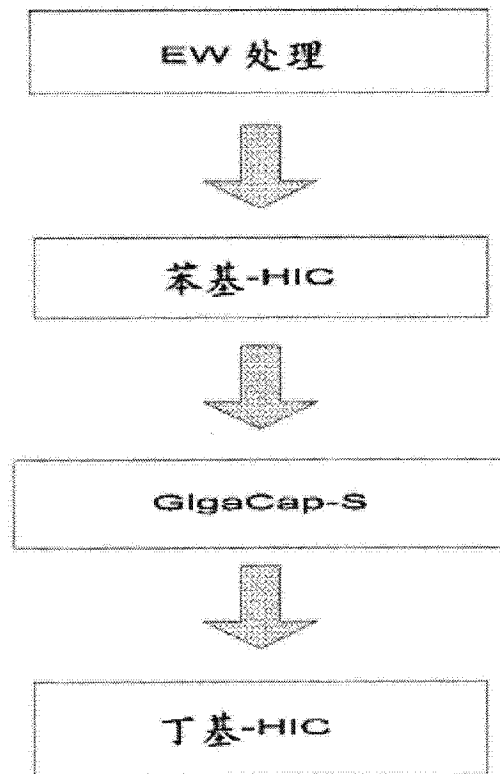
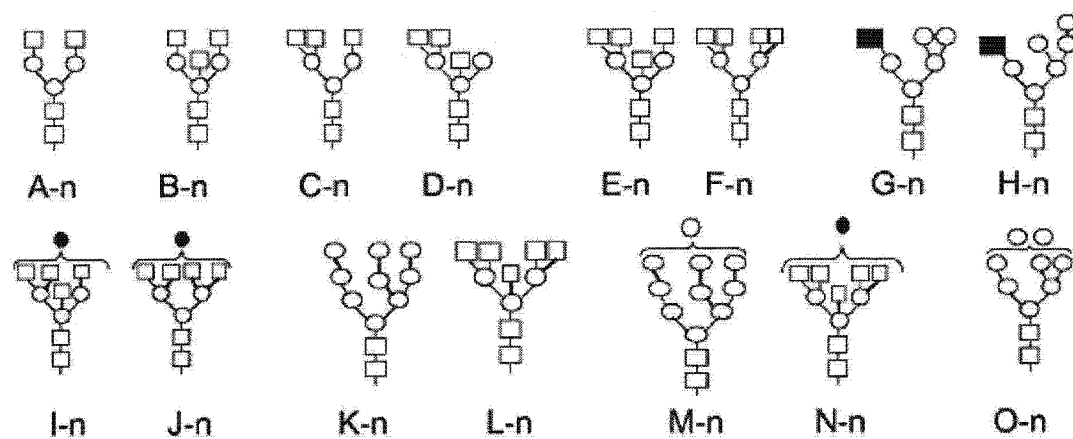


图 15



正方形 = N-乙酰葡萄糖胺

实心正方形 = 甘露糖-6-磷酸

圆圈 = 甘露糖

实心圆 = 半乳糖

实心三角形 = 岩藻糖

图 16

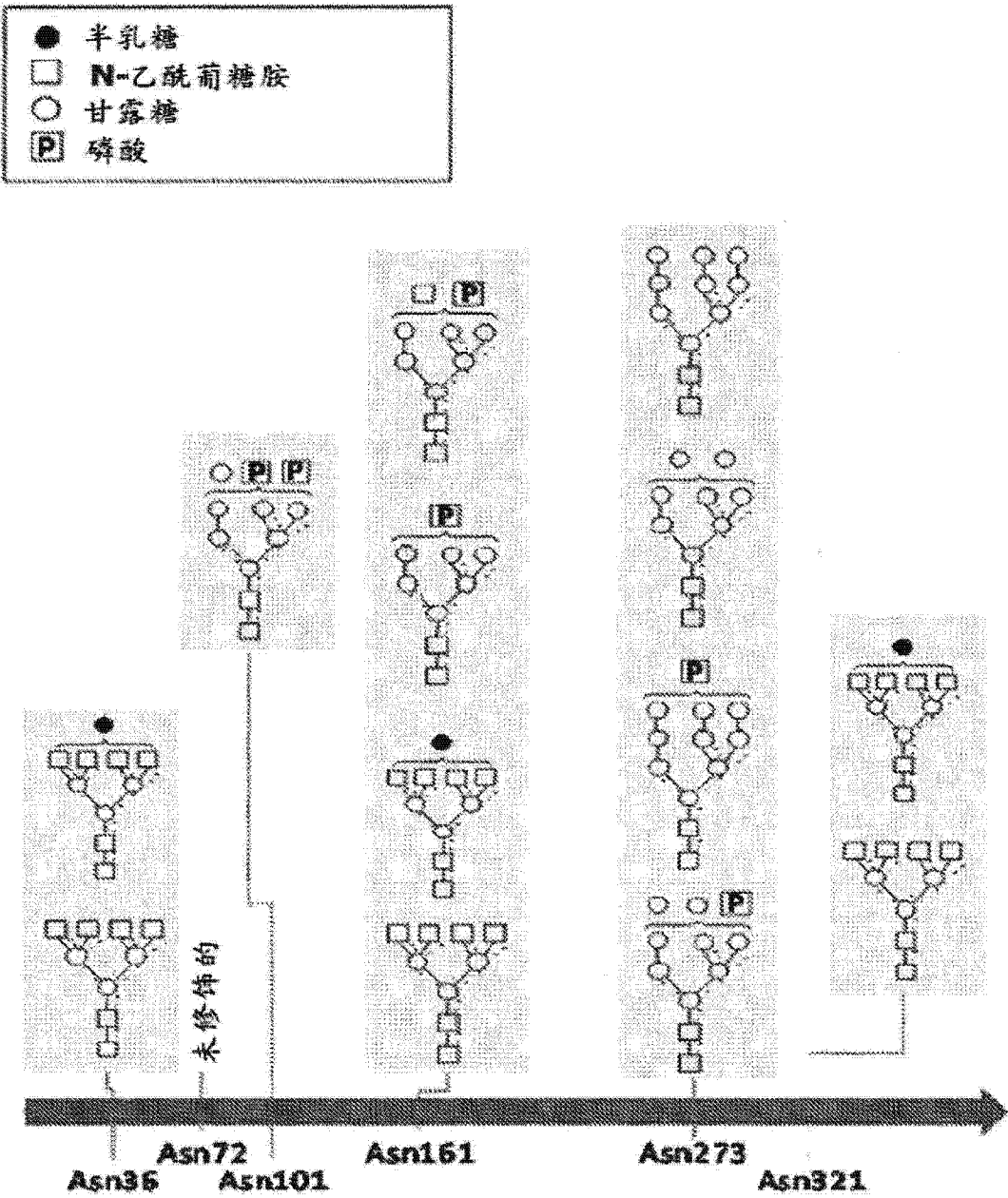


图 17

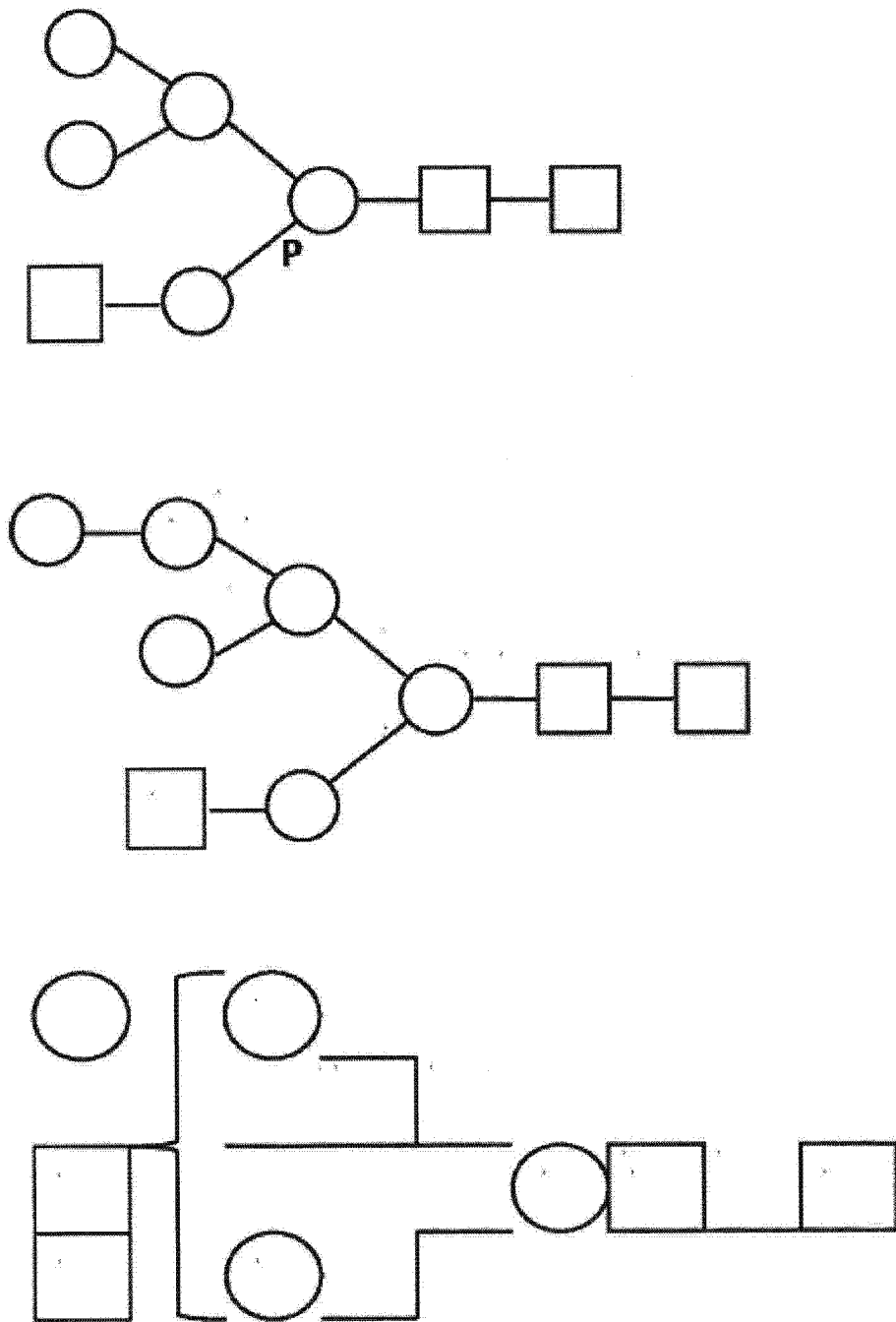


图 18

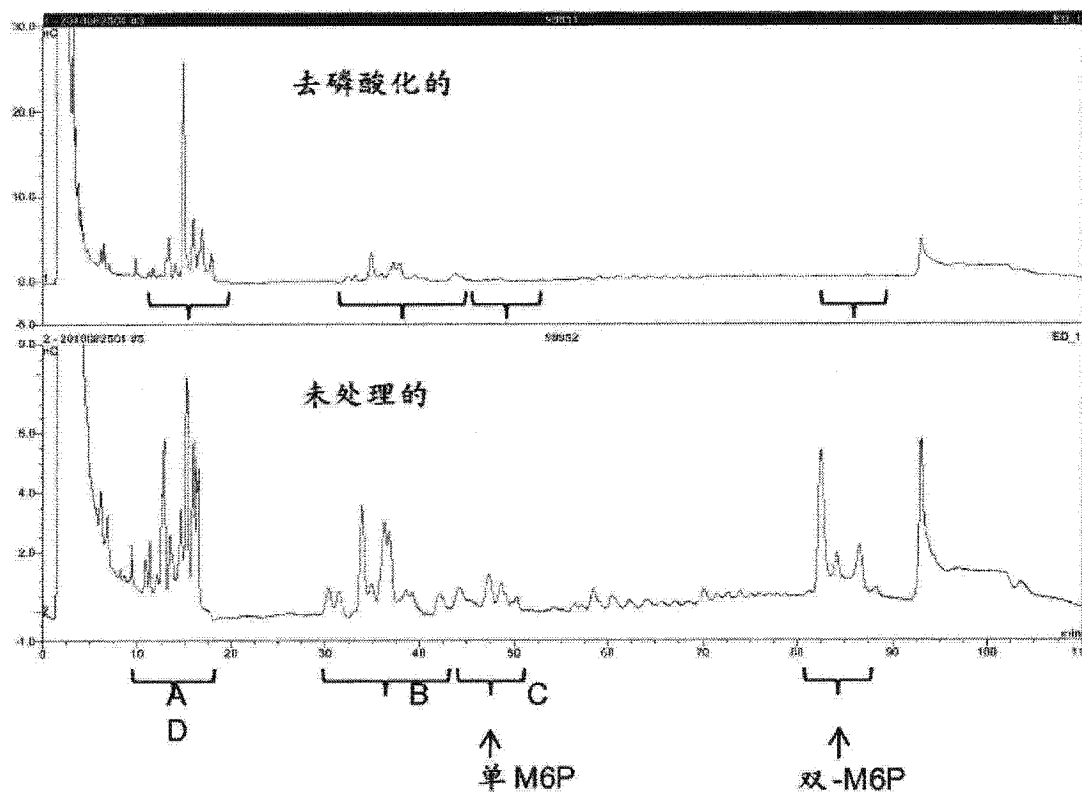


图 19

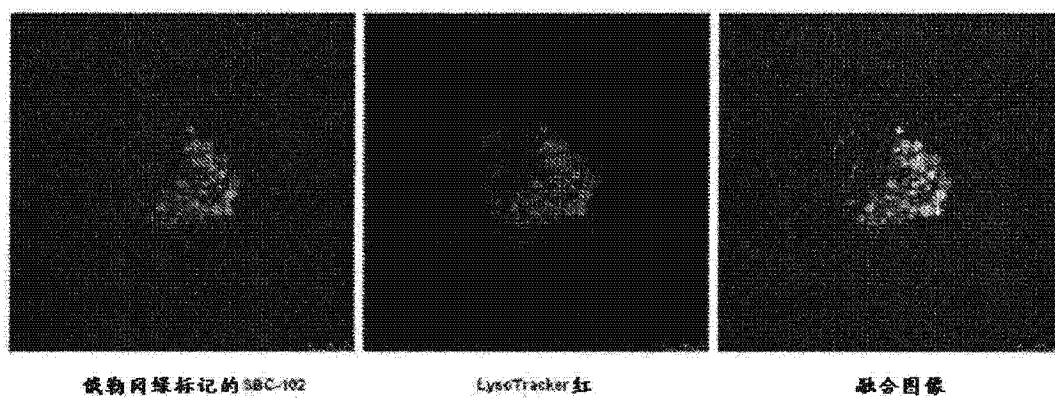


图 20

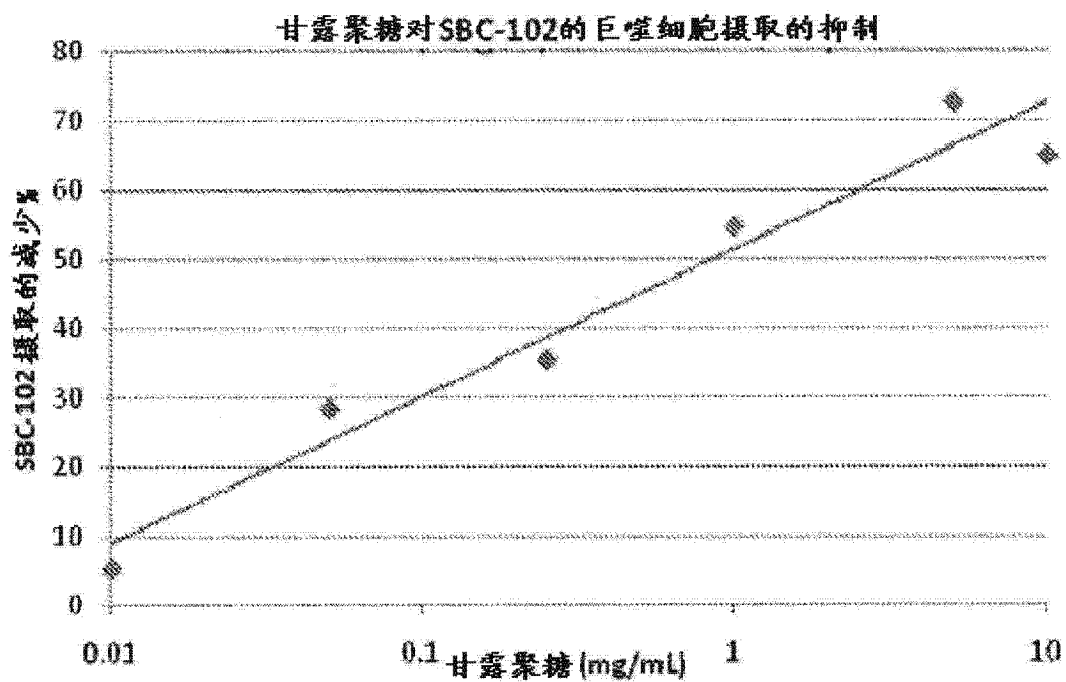


图 21

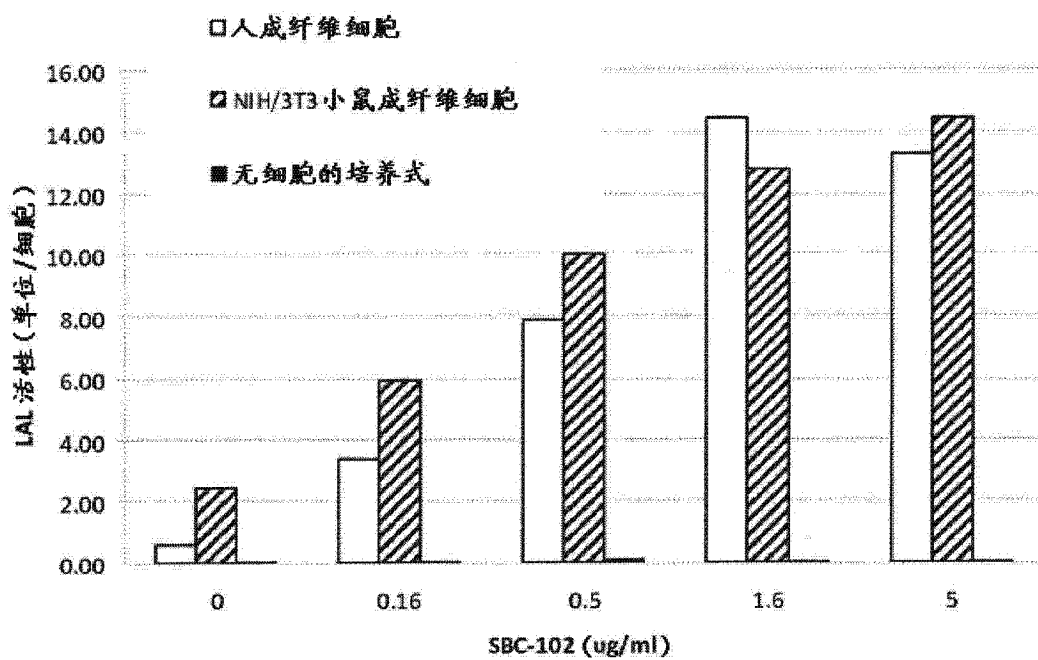


图 22

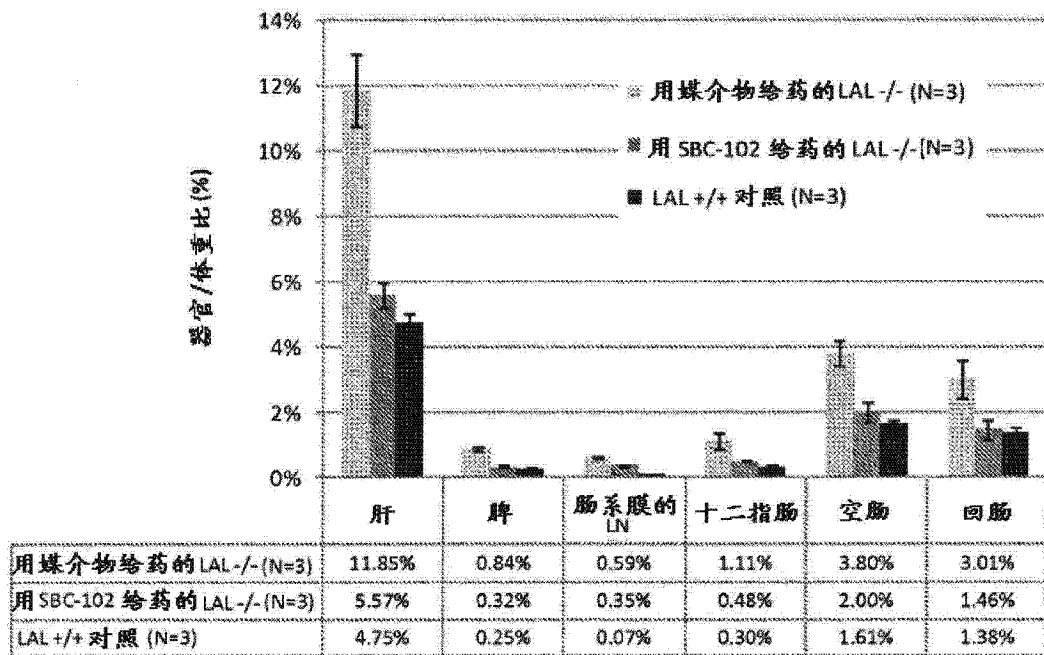


图 23

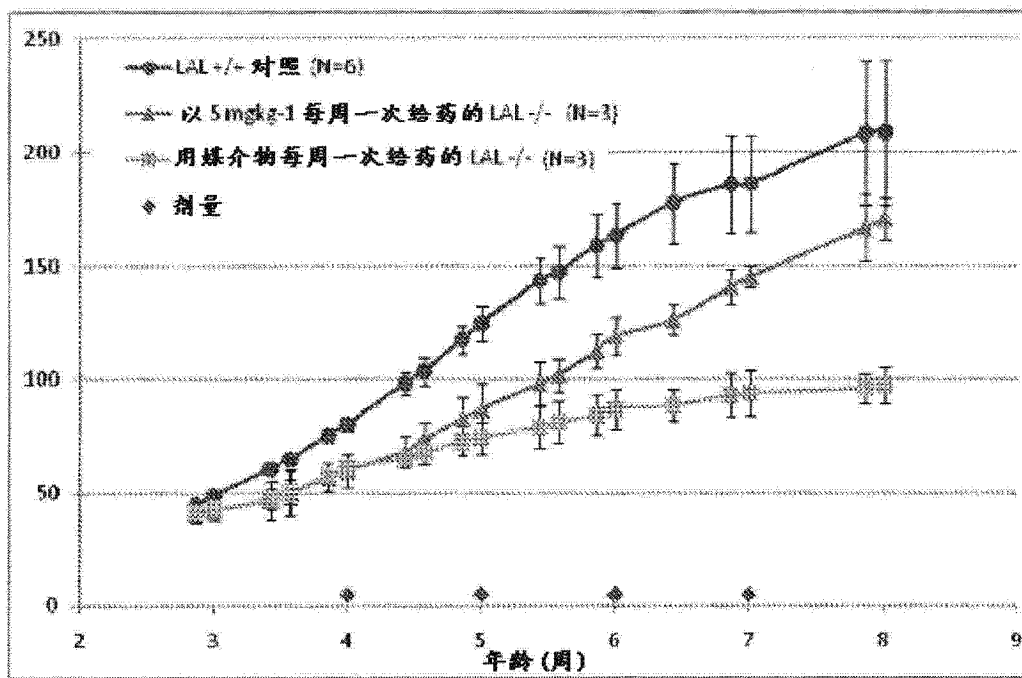


图 24

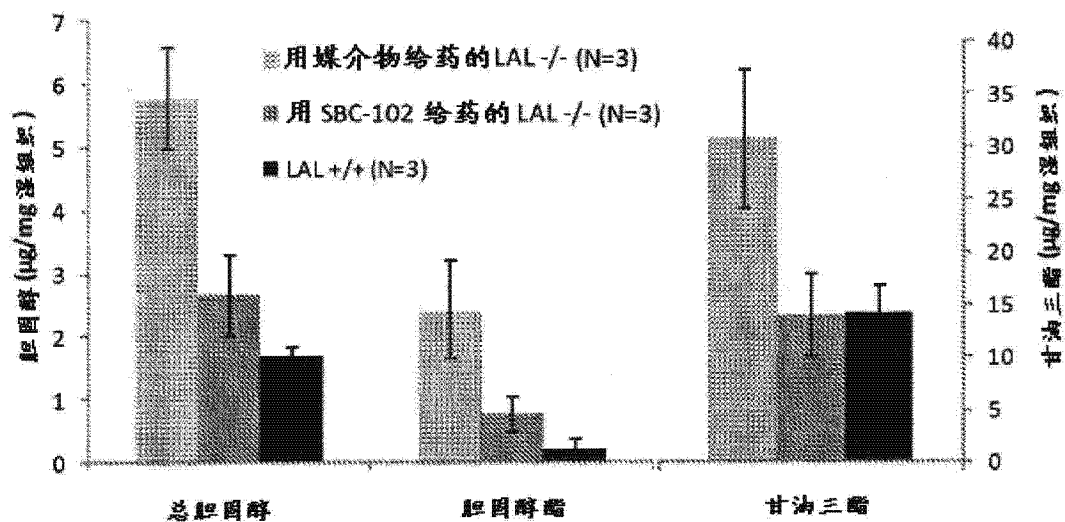


图 25

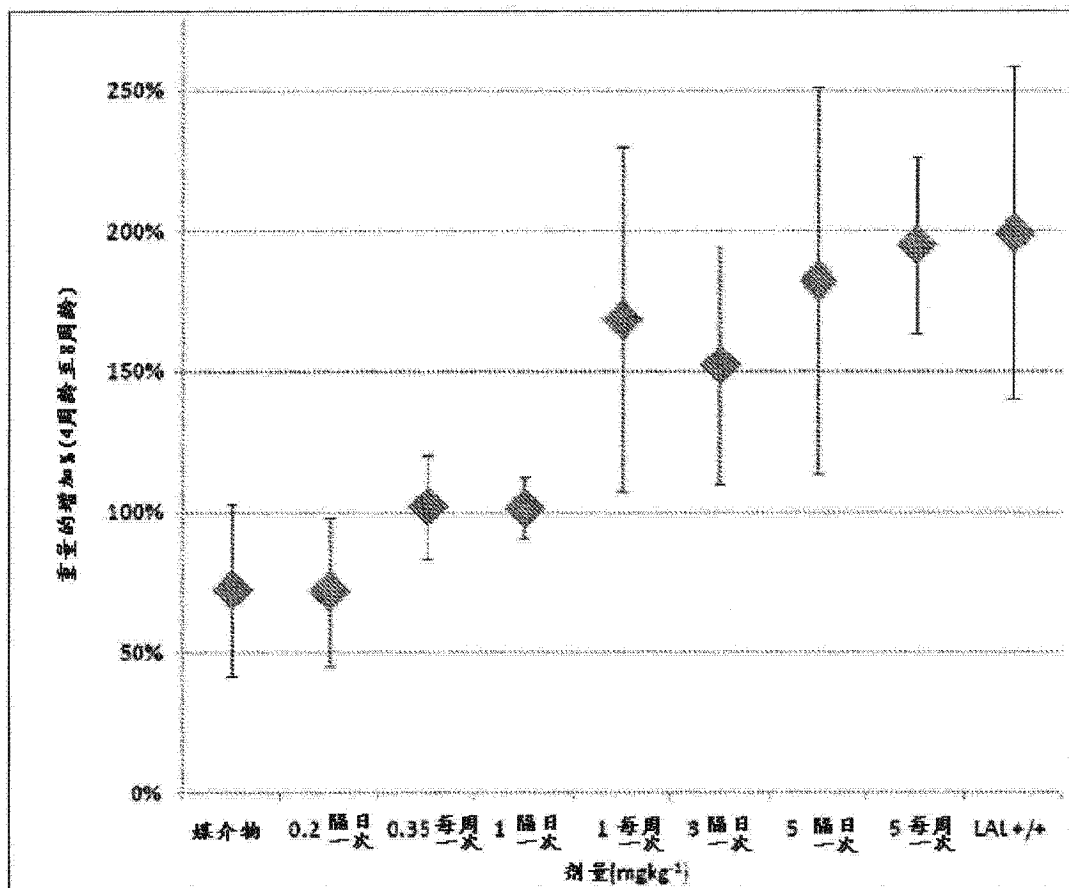


图 26

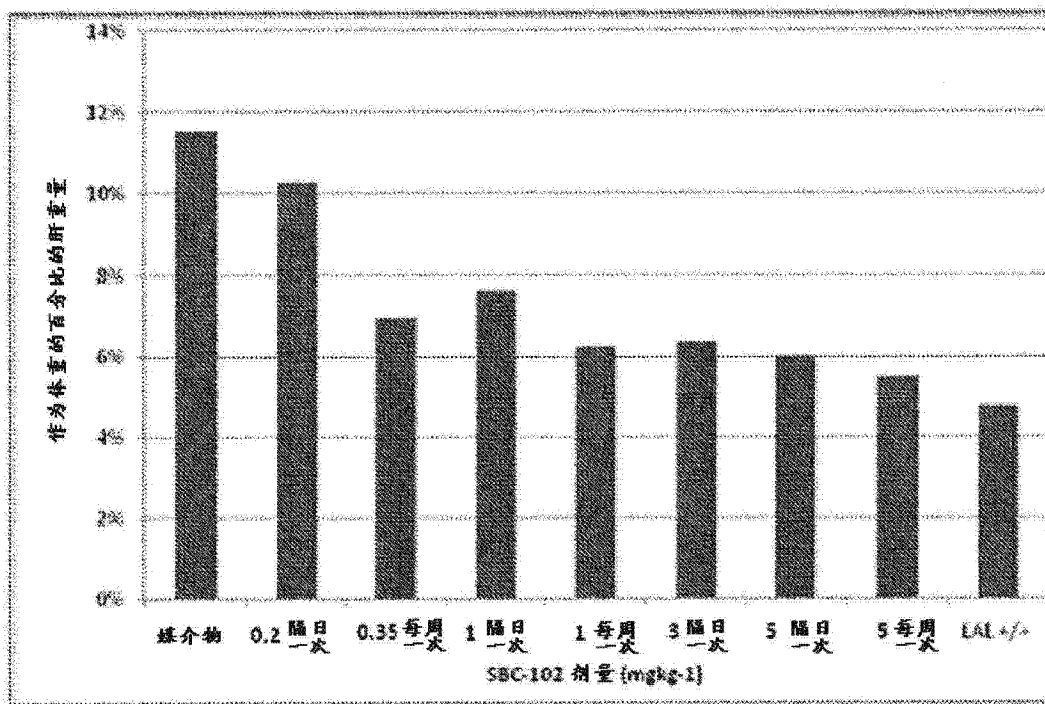


图 27

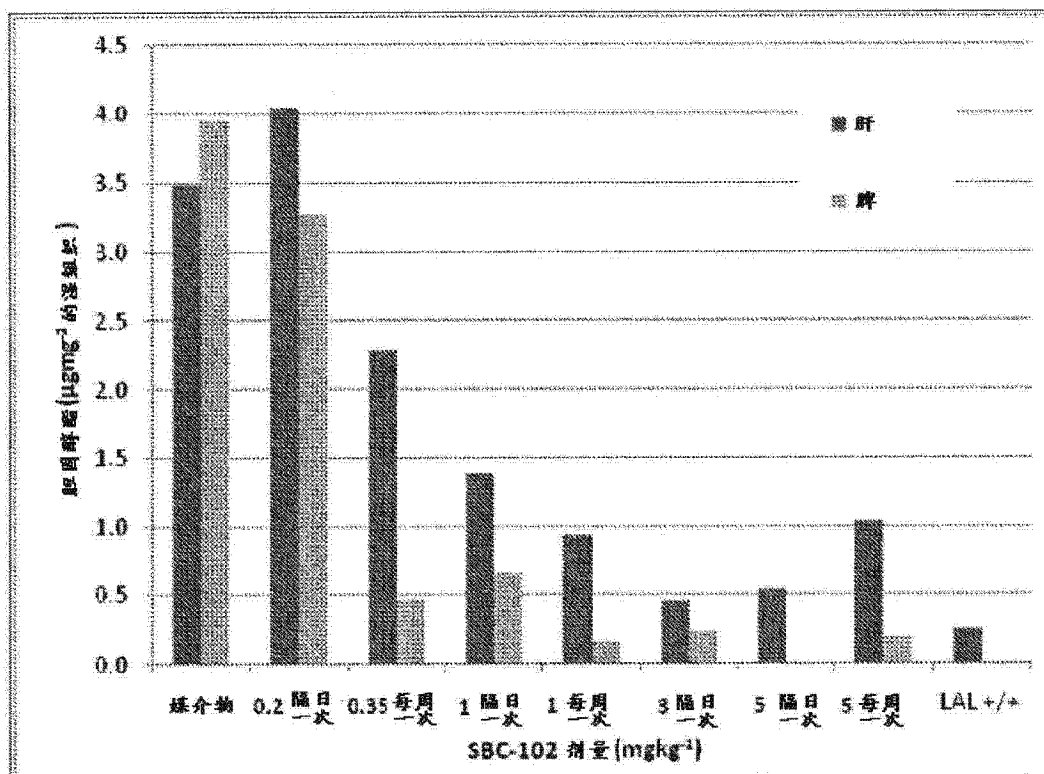


图 28

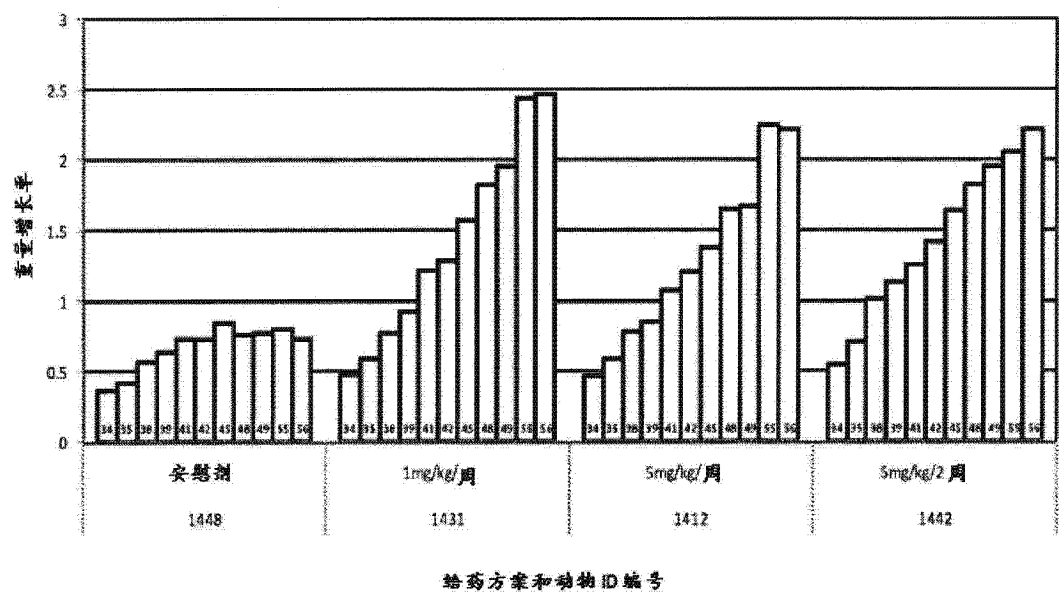


图 29

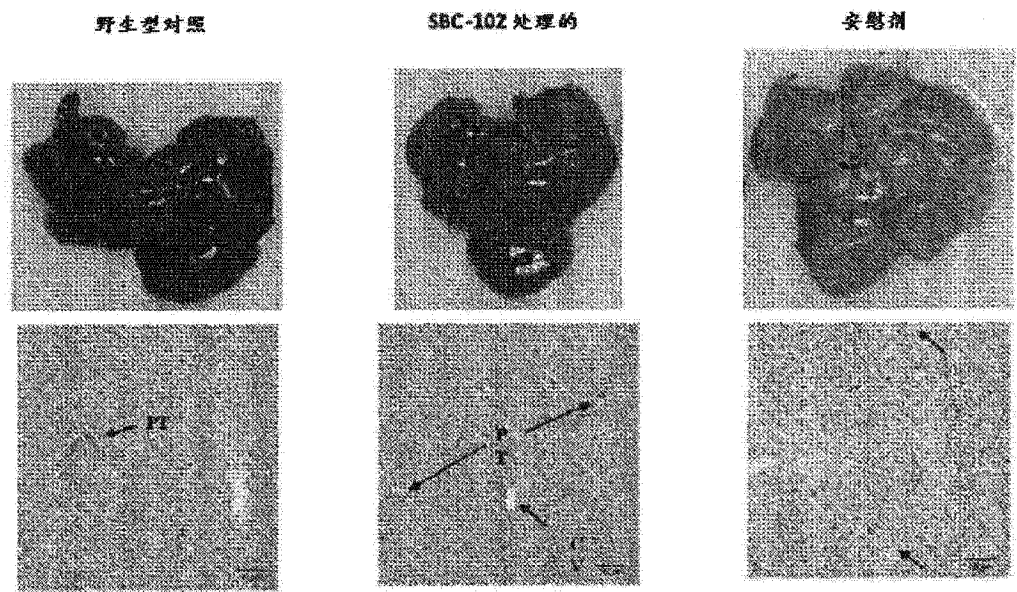


图 30