

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公开说明书

[21] 申请号 200610068150.1

[51] Int. Cl.

A61K 38/17 (2006.01)
A61K 31/711 (2006.01)
C07K 14/47 (2006.01)
C07K 17/02 (2006.01)
C07K 19/00 (2006.01)
C07K 16/18 (2006.01)

[43] 公开日 2006年10月4日

[11] 公开号 CN 1840178A

[51] Int. Cl. (续)

C12N 15/62 (2006.01)

C12N 15/12 (2006.01)

C12N 15/70 (2006.01)

C12N 15/85 (2006.01)

C12N 1/21 (2006.01)

G01N 33/50 (2006.01)

G01N 33/53 (2006.01)

C12Q 1/68 (2006.01)

[22] 申请日 2001.2.16

[21] 申请号 200610068150.1

分案原申请号 01808411.7

[30] 优先权

[32] 2000. 2. 23 [33] GB [31] 0004269.7

[32] 2000. 4. 20 [33] GB [31] 0009905.1

[32] 2000. 8. 25 [33] GB [31] 0021080.7

[71] 申请人 史密丝克莱恩比彻姆生物有限公司

地址 比利时里克森萨特

[72] 发明人 T·E·V·卡贝宗-斯尔瓦

J·-P·卡萨特 T·科彻

S·R·J·-T·高利斯

C·维纳尔斯 Y 德巴索尔斯

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 梁 谋

权利要求书4页 说明书91页 附图6页

[54] 发明名称

肿瘤特异性动物蛋白

[57] 摘要

本发明公开了 CASB7439 多肽和多核苷酸、包含它们的免疫原性组合物以及通过重组技术生产这样的多肽的方法。也公开了 CASB7439 多肽和多核苷酸在诊断学上的使用方法以及用于预防性和治疗性治疗癌症,尤其是结肠直肠癌、自身免疫病和相关病症的疫苗。

1. 一种免疫原性组合物，所述组合物包含安全有效量的CASB7439多肽或其免疫原性片段以及药学上可接受的载体，其中所述CASB7439多肽为SEQ ID NO: 2。

2. 一种免疫原性组合物，所述组合物包含安全有效量的编码CASB7439的多核苷酸或其片段以及药学上可接受的载体，其中所述编码CASB7439的多核苷酸为SEQ ID NO:1的核苷酸545-1126的多肽编码序列。

3. 一种权利要求1或2的免疫原性组合物，所述组合物还包含TH-1诱导型佐剂。

4. 一种权利要求3的免疫原性组合物，其中所述TH-1诱导型佐剂选自以下的佐剂：3D-MPL、QS21、QS21和胆固醇的混合物、CpG寡核苷酸、或者两种或两种以上所述佐剂的混合物。

5. 一种免疫原性组合物，所述组合物包含有效量的抗原呈递细胞和药学上的有效载体，所述抗原呈递细胞通过在体外加载CASB7439多肽而被修饰，或者在体外经遗传修饰以表达CASB7439多肽。

6. 一种分离的多肽，所述多肽包含的一段氨基酸序列在SEQ ID NO: 2的全长上与SEQ ID NO: 2中所示氨基酸序列具有至少70%同一性。

7. 一种权利要求6的分离的多肽，其中所述氨基酸序列与SEQ ID NO: 2具有至少95%同一性。

8. 权利要求7的多肽，所述多肽包含SEQ ID NO: 2的氨基酸序列。

9. SEQ ID NO: 2的分离的多肽。

10. 一种包含权利要求6-9中任一项的多肽的免疫原性片段的多肽，其中所述免疫原性片段的免疫原性活性与SEQ ID NO: 2的多肽

的免疫原性活性基本上相同。

11. SEQ ID NO: 2 的免疫原性片段, 其中所述片段包含 SEQ ID NO: 16 至 SEQ ID NO: 33 中的一个或多个的序列。

12. 一种权利要求 6-11 中任一项的多肽, 其中所述多肽是一种较大融合蛋白的一部分。

13. 一种权利要求 6-11 中任一项的多肽, 所述多肽与一种载体蛋白化学缀合。

14. 一种分离的多核苷酸, 所述多核苷酸编码权利要求 6-12 中任一项的多肽。

15. 权利要求 14 的分离的多核苷酸, 所述多核苷酸包含 SEQ ID NO:1 的核苷酸 545-1126 的序列。

16. 一种包含编码一种多肽的核苷酸序列的分离的多核苷酸, 所述多肽在 SEQ ID NO: 2 的全长上与 SEQ ID NO: 2 的氨基酸序列具有至少 70%同一性。

17. 一种权利要求 14-16 中任一项的分离的多核苷酸, 其中所述同一性为至少 95%。

18. 一种分离的多核苷酸, 所述多核苷酸选自:

(a) 包含编码 SEQ ID NO: 2 的多肽的核苷酸序列的多核苷酸;

(b) SEQ ID NO:1 的核苷酸 545-1126 的多肽编码序列。

19. 一种表达载体或一种活的重组微生物, 所述载体或微生物包含权利要求 14-18 中任一项的分离的多核苷酸。

20. 一种宿主细胞, 所述宿主细胞包含权利要求 19 的表达载体或权利要求 14-18 任一项的分离的多核苷酸。

21. 一种生产权利要求 1-5 中任一项的免疫原性组合物的方法, 所述方法包括将 CASB7439 多肽或编码 CASB7439 多肽的多核苷酸与合适的佐剂、稀释剂或其它药学上可接受的载体混合。

22. 一种生产权利要求 6-12 中任一项的一种多肽的方法, 所述方法包括在足以生产所述多肽的条件下培养权利要求 20 的宿主细

胞，然后从培养基中回收所述多肽。

23. 一种权利要求 6-13 和 14-18 中任一项的多肽或多核苷酸的用途，所述多肽或多核苷酸用于生产免疫治疗性治疗患有或易患癌症的患者的疫苗。

24. 一种权利要求 6-13 和 14-18 中任一项的多肽或多核苷酸的用途，所述多肽或多核苷酸用于生产免疫治疗性治疗患有或易患结肠癌或其它结肠相关肿瘤或疾病的患者的疫苗。

25. 一种免疫专一性抗体，所述抗体对于权利要求 6-13 中任一项要求保护的多肽或免疫学片段具有免疫专一性。

26. 一种筛选方法，所述方法用以鉴定刺激或抑制权利要求 6-13 中任一项的多肽的功能的化合物，所述方法包括选自以下的一种方法：

(a) 利用直接或间接结合候选化合物的标记，测定所述候选化合物与所述多肽(或与携带所述多肽的细胞或细胞膜)或与所述多肽融合蛋白的结合；

(b) 在标记竞争物的存在下，测定候选化合物与所述多肽(或与携带所述多肽的细胞或细胞膜)或与所述多肽融合蛋白的结合；

(c) 使用适合于携带所述多肽的细胞或细胞膜的检测系统，测试所述候选化合物是否引起由于所述多肽的活化或抑制而产生的信号；

(d) 将候选化合物与含权利要求 6-13 中任一项的多肽的溶液混合，形成混合物，然后测定所述混合物中所述多肽的活性，并且将所述混合物的活性与标准品的活性进行比较；或

(e) 采用例如 ELISA 测定，检测候选化合物对在细胞中编码所述多肽的 mRNA 和所述多肽的产生的影响。

27. 一种针对权利要求 6-13 的多肽的激动剂或拮抗剂。

28. 一种用于治疗的化合物，所述化合物是：

(a) 一种针对权利要求 6-13 的多肽的激动剂或拮抗剂；

(b) 权利要求 14-18 的分离的多核苷酸；或

(c) 一种调节编码权利要求 6-13 中任一项的多肽的核苷酸序列的表达的核酸分子。

29. 一种诊断方法，所述方法用于诊断受试者与权利要求 6-13 中任一项的多肽的表达或活性有关的疾病或对所述疾病的易感性，所述方法包括分析得自所述受试者的样品中的所述多肽的存在或含量。

30. 一种诊断方法，所述方法用于诊断受试者与权利要求 14-18 中任一项的多核苷酸的表达或活性有关的疾病或对所述疾病的易感性，所述方法包括分析得自所述受试者的样品中的所述多核苷酸的存在或含量。

31. 一种诊断方法，所述方法用于诊断受试者与权利要求 6-13 中任一项的多肽的表达或活性有关的结肠直肠癌的存在或对所述结肠直肠癌的易感性，所述方法包括分析得自所述受试者的样品中的所述多肽的存在或含量。

32. 一种诊断方法，所述方法用于诊断受试者与权利要求 14-18 中任一项的多核苷酸的表达或活性有关的结肠直肠癌的存在或对所述结肠直肠癌的易感性，所述方法包括分析得自所述受试者的样品中的所述多核苷酸的存在或含量。

肿瘤特异性动物蛋白

本申请是以下申请的分案申请：申请日：2001年2月16日；申请号：01808411.7(PCT/EP01/01779)；发明名称：“肿瘤特异性动物蛋白”。

发明领域

本发明涉及用于诱导针对肿瘤相关抗原的免疫应答的药用组合物和方法。更具体地说，本发明涉及多核苷酸(本文称为 CASB7439 多核苷酸)、由它们编码的多肽(本文称为 CASB7439 多肽)、用于生产它们的重组材料和方法。另一方面，本发明涉及这样的多肽和多核苷酸的使用方法，包括治疗癌症、更特别是结肠直肠癌和自身免疫病和其它相关病症的疗法。另一方面，本发明涉及含有 CASB7439 多肽和多核苷酸的药用组合物、涉及这种组合物的生产方法及其在医药领域的应用。再一方面，本发明涉及用本发明提供的材料鉴定激动剂和拮抗剂/抑制剂的方法以及用所鉴定的化合物治疗与 CASB7439 多肽不平衡相关的病症的方法。又一方面，本发明涉及用于检测与不适当 CASB7439 多肽活性或水平相关的疾病的诊断分析。

背景技术

本发明的多肽和多核苷酸被认为是针对肿瘤的专一性预防性或治疗性免疫接种的重要免疫原，因为与在正常细胞中的表达相比，它们在肿瘤中特异性表达或高度过量表达，因此可以通过抗原特异性免疫机制靶向它们，导致破坏肿瘤细胞。它们还可以用于诊断肿瘤细胞的存在。此外，它们在某些情况下不适当的表达可以引起诱导自身免疫、即不适当的免疫应答，通过用所述多肽或多核苷酸适当地接种疫苗可能矫正所述不适当的免疫应答。在这方面，对我们的目的而言，最重要的生物活性是本发明多肽的抗原活性和免疫原性活性。本发明的一种多肽还可能表现出一种 CASB7439 多肽的至少一种其它生物活性，它可能被看作是不同于与所述免疫应答相关的治疗性或预防性干涉的治疗性或预防性干涉的靶。

发明内容

第一方面,本发明涉及 CASB7439 多肽。这种肽包括分离的多肽,所述多肽包含的一段氨基酸序列在 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的全长上分别与 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的氨基酸序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性、最优选至少 97-99%同一性,前提是所述分离的多肽不是 SEQ ID NO:2、SEQ ID NO:12 或 SEQ ID NO:14。这种多肽包括含 SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 和 SEQ ID NO:11 的氨基酸的多肽。

本发明的其它肽包括分离的多肽,其中所述氨基酸序列在 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的全长上分别与 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的氨基酸序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性、最优选至少 97-99%同一性,前提是所述分离的多肽不是 SEQ ID NO:2、SEQ ID NO:12 或 SEQ ID NO:14。这种多肽包括 SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 和 SEQ ID NO:11 的多肽。

优选上述多肽通过重组产生。最优选按照本发明的多肽是经纯化的,并且基本上不含任何其它蛋白或污染性宿主来源的材料。

本发明的其它肽包括由包含在 SEQ ID NO:1 中所含序列的多核苷酸编码的分离的多肽。

本发明也提供 CASB7439 多肽的免疫原性片段,即 CASB7439 多肽的一个连续部分,所述免疫原性片段具有与包含下述氨基酸序列的多肽相同或相似的免疫原性特性: SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14。也就是说,所述片段(如有必要,与载体偶联或作为较大融合蛋白的一部分)能够产生识别所述 CASB7439 多肽的免疫应答。这样的免疫原性片段可以包括例如缺乏 N-末端前导序列、跨膜结构域或 C-

末端锚着结构域的 CASB7439 多肽。在一个优选的方面，按照本发明的 CASB7439 免疫原性片段包含多肽的基本上全部的胞外结构域，所述多肽在 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的全长上分别与 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 的氨基酸序列具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、最优选至少 97-99% 同一性。按照本发明的免疫原性片段最好包含至少一个表位。

掺入 CASB7439 表位的肽片段通常包含得自 SEQ ID NO:2 的至少 7 个、优选 9 个或 10 个连续氨基酸。优选的表位示于 SEQ ID NO:16 至 SEQ ID NO:33 中。

掺入这些表位的肽构成本发明的一个优选方面。与这些表位的特征相同的模拟表位 (mimotope) 和包含这些模拟表位、产生与在 CASB7439 分子环境中的表位有交叉反应的免疫应答的免疫原，也构成本发明的一部分。

因此，本发明包括包含这些表位本身及其任何模拟表位的分离的肽。模拟表位的含义被定义为与天然 CASB7439 表位足够相似、以致能够被识别所述天然分子的抗体识别的实体；(Gheysen, H.M. 等, 1986, *Synthetic peptides as antigens*. Wiley, Chichester, Ciba foundation symposium 119, 第 130-149 页；Gheysen, H.M., 1986, *Molecular Immunology*, 23,7, 709-715)；或者当与合适载体偶联时能够产生抗体的实体，其中所述抗体与所述天然分子有交叉反应。

以上所鉴定的表位的肽模拟表位可以通过添加、缺失或取代选定氨基酸，设计为供特定目的使用。因而，可以对本发明的肽进行修饰，以便易于与蛋白质载体缀合。例如，对于某些化学缀合方法，可能理想的是所述表位包括一个末端半胱氨酸。另外，对于与蛋白质载体缀合的肽，可能理想的是包括一个远离所述肽的缀合末端的疏水性末

端，以便所述肽的游离未缀合的末端保留与所述载体蛋白质的表面缔合。这降低了所述肽的构象自由度，因而增加了以与在所述完整分子的情况下存在的肽构象最为类似的构象呈递所述肽的机率。例如，所述肽可以被改变成具有一个 N-端半胱氨酸和一个 C-端疏水性酰胺化尾。另一方面，可以进行一个或多个所述氨基酸的 D-立体异构体形式的添加或取代，以产生有益的衍生物，例如增强所述肽的稳定性。本领域技术人员会认识到，这种经修饰的肽或模拟表位可以是全部或部分非肽模拟表位，其中所述组成残基不一定限于 20 种天然存在的氨基酸。另外，可以通过本领域已知的技术将其环化，以将所述肽限制在密切类似所述肽序列在完整分子环境中的形状的构象。将肽环化的一个优选方法包括添加一对半胱氨酸残基，以允许形成一个二硫桥。

此外，本领域技术人员会认识到，本发明的模拟表位或免疫原可以比上述鉴定的表位大，因此可以包含本文公开的序列。因此，本发明的模拟表位可以包括在一个末端或两个末端添加许多其它天然残基的 N 和/C 端延伸。所述肽模拟表位也可以是天然序列的反序列(retro sequence)，因为所述序列方向是相反的；或者所述序列可以全部或至少部分由 D-立体异构体氨基酸(倒位序列(inverso sequence))构成。另外，所述肽序列可以在特征上是反-倒位的(retro-inverso)，因为所述序列方向是相反的，并且所述氨基酸为 D-立体异构体形式。这样的反-倒位肽的优点是非自身的，因此可以克服免疫系统中的自体耐受的问题。

另一方面，采用诸如噬菌体展示技术(EP 0 552 267 B1)的多种技术，使用其自身能够与本发明的表位结合的抗体可以鉴定肽模拟表位。该技术产生大量的模拟天然肽结构的肽序列，因此能够与抗天然肽抗体结合，但不一定自身与所述天然肽共享显著的序列同源性。该方法可能是非常有利的，因为使得有可能鉴定免疫原性特性增强的肽，或者可以克服任何潜在的可能与应用所述天然肽序列相关的自身抗原耐受问题，另外，鉴于其在所识别肽模拟表位序列中有共享的化

学特性，该技术使得能够鉴定各种天然肽的识别模式。

可以用本领域众所周知的方法使所述肽与免疫原性载体进行共价偶联。因此，例如，对于直接共价偶联，有可能利用碳二亚胺、戊二醛或(N-[γ -马来酰亚胺基丁酰氧基]琥珀酰亚胺酯，利用普通市售的杂双官能接头例如 CDAP 和 SPDP (采用生产商的说明))。在偶联反应后，所述免疫原可以容易地通过透析法、凝胶过滤法、分级分离法等进行分离和纯化。

本发明免疫原中所用的载体类型是本领域技术人员容易知道的。所述载体的功能是提供细胞因子辅助，以便帮助诱导针对所述肽的免疫应答。可用于本发明的载体的非竭尽性一览表包括：匙孔瓣血蓝蛋白(KLH)、血清白蛋白例如牛血清白蛋白(BSA)、灭活细菌毒素(例如破伤风毒素或白喉毒素(TT 和 DT))或其重组片段(例如，TT 的片段 C 的结构域 1 或 DT 的易位结构域)或结核菌素的纯化蛋白衍生物(PPD)。另一方面，所述模拟表位或表位可以直接与脂质体载体缀合，所述脂质体载体另外包含能够提供 T-细胞辅助的免疫原。模拟表位与载体之比最好约为 1:1 至 20:1，每个载体最好应该携带 3-15 个肽。

在本发明的一个实施方案中，一种优选的载体是流感嗜血菌(*Haemophilus influenzae*)的 D 蛋白(EP 0 594 610 B1)。D 蛋白是一种得自流感嗜血菌的 IgD 结合蛋白，并且 Forsgren (WO 91/18926, 授权公告号(Granted) EP 0 594 610 B1)已获得专利权。在某些情况下，例如在重组免疫原表达系统中，可能最好利用 D 蛋白的片段，例如 D 蛋白的三分之一(包含 D 蛋白 N-末端 100-110 个氨基酸(GB 9717953.5))。

呈递本发明肽的另一优选的方法是在重组融合分子的环境中呈递。例如，EP 0 421 635 B 描述了应用嵌合嗜肝 DNA 病毒核心抗原颗粒呈递病毒样颗粒中的外源肽序列。因此，本发明的免疫原可以包含在由乙型肝炎核心抗原构成的嵌合颗粒中呈递的肽。另外，所述重组融合蛋白可以包含本发明的模拟表位和一种载体蛋白，例如流感病毒的 NS1。对于构成本发明部分的任何重组表达的蛋白，编码所述免疫

原的核酸也构成本发明的一个方面。

本发明中所用的肽可以通过本领域众所周知的固相方法容易地合成。适宜的合成可以通过利用“T-boc”或“F-moc”方法来进行。环肽可以利用众所周知的“F-moc”方法和全自动装置中的聚酰胺树脂通过固相方法来合成。或者，本领域技术人员会知道人工进行所述方法必需的实验室方法。固相合成的技术和方法描述于 E. Atherton 和 R.C. Sheppard 的“Solid Phase Peptide Synthesis: A Practical Approach”, Oxford University Press 的 IRL 出版(1989)。或者，所述肽可以通过重组方法产生，包括在细菌或哺乳动物细胞系中表达编码所述模拟表位的核酸分子，然后纯化所表达的模拟表位。肽和蛋白的重组表达技术是本领域已知的，并且描述于 Maniatis, T., Fritsch, E.F.和 Sambrook 等, *Molecular cloning, a laboratory manual*, 第二版; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)。

在本发明的再一实施方案中提供本文所述多肽的制备方法。本发明的方法可以通过常规重组技术来进行，所述常规重组技术例如描述于 Maniatis 等, *Molecular Cloning - A Laboratory Manual*; Cold Spring Harbor (1982-1989)中。因此，提供制备按照本发明多肽的方法，所述方法包括在足以产生所述多肽的条件下培养宿主细胞，从所述培养基中回收所述多肽。具体地说，本发明的方法可能最好包括下列步骤：

- i) 制备能够在宿主细胞中表达DNA聚合物的复制型或整合型表达载体，所述DNA聚合物包含编码所述蛋白或其免疫原性衍生物的核苷酸序列；
- ii) 用所述载体转化宿主细胞；
- iii) 在允许表达所述DNA聚合物的条件下培养所述转化的宿主细胞，以产生所述蛋白；和
- iv) 回收所述蛋白。

本发明的多肽或免疫原性片段可以是“成熟”蛋白形式，或者是更大的蛋白如前体蛋白或融合蛋白的一部分。包括包含以下序列

的额外氨基酸序列通常是有利的：分泌序列或前导序列、原序列(pro-sequences)、有助于纯化的序列如多组氨酸残基、或在重组生产过程中有利于稳定性的其它序列。此外，也考虑添加外源多肽或脂质尾或多核苷酸序列，以增加最终分子的免疫原性潜能。

一方面，本发明涉及遗传工程可溶性融合蛋白，所述融合蛋白包含本发明的多肽或其片段，以及不同亚类的免疫球蛋白(IgG、IgM、IgA、IgE)重链或轻链的恒定区的不同部分。作为免疫球蛋白，优选的是人类 IgG (特别是 IgG1)的重链的恒定部分，其中在绞链区发生融合。在一个特定实施方案中，可以通过加入可用凝血因子 Xa 切除的切割序列就可除去 Fc 部分。此外，本发明涉及通过遗传工程制备这些融合蛋白的方法，以及这些融合蛋白在药物筛选、诊断和治疗中的应用。本发明的一个特别优选的方面涉及应用多肽或多核苷酸生产免疫治疗性治疗患有或易患癌症、尤其是结肠癌或其它结肠相关性肿瘤或疾病的患者的疫苗。本发明的再一方面还涉及编码这样的融合蛋白的多核苷酸。在国际专利申请号 WO94/29458 和 WO94/22914 中可以找到融合蛋白技术的实例。

所述蛋白可以化学缀合或作为重组融合蛋白表达，以使得与未融合的蛋白相比，在表达系统中产生的水平提高。所述融合配偶体可以协助提供 T 辅助细胞表位(免疫融合配偶体)、最好是被人类识别的 T 辅助细胞表位，或者所述融合配偶体有助于以比天然重组蛋白更高产量表达所述蛋白(表达增强子)。最好所述融合配偶体既是免疫融合配偶体，又是表达增强配偶体。

融合配偶体包括流感嗜血菌 B 的 D 蛋白以及流感病毒的非结构蛋白 NS1 (血凝素)。另一种免疫融合配偶体是称为 LYTA 的蛋白。最好使用所述分子的 C-末端部分。Lyta 得自肺炎链球菌(*Streptococcus pneumoniae*)，它合成 N-乙酰-L-丙氨酸酰胺酶-酰胺酶 LYTA (由 *lytA* 基因编码{Gene, 43 (1986)第 265-272 页})，即特异性降解肽聚糖骨架中某些键的自溶素。LYTA 蛋白的 C-末端结构域负责对胆碱或一些胆碱

类似物如 DEAE 的亲水性。已经利用该特性开发可用于表达融合蛋白的大肠杆菌 C-LYTA 表达质粒。已经描述了在其氨基端包含 C-LYTA 片段的杂种蛋白的纯化{Biotechnology: 10, (1992)第 795-798 页}。可利用存在于 Lyta 分子 C-末端的起始于残基 178 的重复序列部分,例如残基 188-305。

本发明也包括前述多肽的异种形式(也称为直向同源物形式),所述异种形式是指与人类抗原(也称为自身抗原)具有大致序列同一性的抗原,所述抗原用作来源于不同的非人类物种的参比抗原。在该方面,所述大致同一性是指当序列在本领域已知的众多序列比对蛋白中的任一种中以最佳序列比对排列时一个氨基酸序列与另一个氨基酸序列或者一个多核苷酸序列与另一个多核苷酸序列的一致性。所谓的大致同一性是指比较序列之间的序列同一性至少为 70-95%、优选至少 85-95%、最优选至少 90-95%。因此,按照本发明,异种 CASB7439 多肽将是对于人 CASB7439 为异种的 CASB7439 多肽,换句话说,异种 CASB7439 多肽从非人类的物种中分离出来。在一个优选的实施方案中,所述多肽从小鼠、大鼠、猪或恒河猴中分离,最优选分离自小鼠或大鼠。因此,本发明也提供在人体内诱导针对人 CASB7439 的免疫应答的方法,所述方法包括给予所述受治疗者有效剂量的包含如本文中所述的所述人 CASB7439 的异种形式的组合物,所述人 CASB7439 的氨基酸序列如序列 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 或 SEQ ID NO:11 中的任一序列所示。一个优选的实施方案是使用从小鼠、大鼠、猪或恒河猴中分离的异种 CASB7439 诱导针对人 CASB7439 的免疫应答的方法。按照本发明诱导免疫应答的另一种优选方法是使用包括表达所述异种抗原的活病毒表达系统的抗原组合物。优选的异种 CASB7439 多肽的序列示于 SEQ ID NO:12 (小鼠)或 SEQ ID NO:14 (大鼠)中。

分离的异种 CASB7439 多肽通常共享大致序列相似性,包括包含在 SEQ ID NO:12 或 SEQ ID NO:14 的全长上与 SEQ ID NO:12 或 SEQ

ID NO:14 的氨基酸序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性、最优选至少 97-99%同一性的氨基酸序列的分离多肽。因此,所述异种多肽包含 SEQ ID NO:12 或 SEQ ID NO:14 的多肽的免疫原性片段,其中所述免疫原性片段的免疫原性活性基本上与 SEQ ID NO:12 或 SEQ ID NO:14 的多肽相同。另外,所述异种 CASB7439 多肽可以是这样的片段:选自 SEQ ID NO:12 或 SEQ ID NO:14 中所示的氨基酸序列的至少约 20 个连续氨基酸、优选约 30 个、更优选约 50 个、再更优选约 100 个、最优选约 150 个连续氨基酸。更具体地说,异种 CASB7439 片段会保留 SEQ ID NO:12 或 SEQ ID NO:14 中所示的较大分子的某些功能特性、最好是免疫活性,并且可用于本文所述的方法中(例如可用于药用组合物或疫苗组合物、诊断学中等)。具体地说,所述片段能够引起针对人类对应物的免疫应答,例如产生与任一 SEQ ID NO:2 所示的自身人类 CASB7439 形式反应的交叉反应性抗体。在一个具体的实施方案中,本发明的异种多肽可以是较大融合物的一部分,所述融合物包含异种 CASB7439 多肽或其片段以及如上所述的用作融合配偶体的异源蛋白或蛋白部分。

本发明还包括前述多肽的变异体,即通过保守氨基酸取代、从而一个残基被另一个具有相似特性的残基取代而与参比物不同的多肽。通常这样的取代发生在 Ala、Val、Leu 和 Ile 之间; Ser 和 Thr 之间; 酸性残基 Asp 和 Glu 之间; Asn 和 Gln 之间; 以及碱性残基 Lys 和 Arg 之间; 或芳香族残基 Phe 和 Tyr 之间。特别优选其中若干个、5-10 个、1-5 个、1-3 个、1-2 个或 1 个氨基酸以任何组合发生取代、缺失或添加的变异体。

可以以任何合适的方式制备本发明的多肽。这样的多肽包括分离的天然存在的多肽、重组产生的多肽、合成产生的多肽或通过这些方法的组合产生的多肽。用于制备这样的多肽的方法是本领域众所周知的。

另一方面,本发明涉及 CASB7439 多核苷酸。这样的多核苷酸包

括含编码多肽的核苷酸序列的分离的多核苷酸，所述多肽在 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 或 SEQ ID NO:11 的全长上分别与 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 或 SEQ ID NO:11 的氨基酸序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性。在这一方面，高度优选具有至少 97%同一性的所编码多肽，而更高度优选具有至少 98-99%同一性的所编码多肽，最优选具有至少 99%同一性的所编码多肽。

本发明的其它多核苷酸包括分离的多核苷酸，所述多核苷酸包含的一段核苷酸序列在完整编码区上与编码 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 或 SEQ ID NO:11 的多肽的核苷酸序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性。在这一方面，高度优选具有至少 97%同一性的多核苷酸，而更高度优选具有至少 98-99%同一性的多核苷酸，最优选具有至少 99%同一性的多核苷酸。

本发明的其它多核苷酸包括分离的多核苷酸，所述多核苷酸包含的一段核苷酸序列在所述序列的全长上与 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:9 或者在 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:9 的所述编码序列的全长上与 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:9 的编码序列具有至少 70%同一性、优选至少 80%同一性、更优选至少 90%同一性、再更优选至少 95%同一性。在这一方面，高度优选具有至少 97%同一性的多核苷酸，而更高度优选具有至少 98-99%同一性的多核苷酸，最优选具有至少 99%同一性的多核苷酸。这样的多核苷酸包括含 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:9 多核苷酸的多核苷酸以及 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、

SEQ ID NO:8、SEQ ID NO:9 的多核苷酸或者 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:9 的编码区。

本发明还提供编码本发明的前述异种蛋白的核酸及其在医药领域的应用。在一个优选的实施方案中，用于药用组合物的异种 CASB7439 多核苷酸的序列示于 SEQ ID NO:13 (小鼠)或 SEQ ID NO:15 (大鼠)。按照本发明的分离的异种 CASB7439 多核苷酸可以是单链(编码链或反义链)或双链的，并且可以是 DNA 分子(基因组分子、cDNA 分子或合成分子)或 RNA 分子。其它的编码序列或非编码序列可以但不一定存在于本发明的多核苷酸中。在其它相关实施方案中，本发明提供多核苷酸变体，所述多核苷酸变体与 SEQ ID NO:13 或 SEQ ID NO:15 中本文公开的序列有大致同一性，例如采用本文所述的方法(例如使用标准参数的 BLAST 分析)与本发明的多核苷酸序列相比包含至少 70% 序列同一性、优选至少 75%、80%、85%、90%、95%、96%、97%、98%或 99%或更高序列同一性的多核苷酸变体。在一个相关实施方案中，本发明的分离的异种多核苷酸包含编码在 SEQ ID NO:12 或 SEQ ID NO:14 的全长上与 SEQ ID NO:12 或 SEQ ID NO:14 的氨基酸序列具有至少 90%、最好是 95%和 95%以上同一性的多肽的核苷酸序列或者与所述分离的多核苷酸互补的核苷酸序列。

本发明还提供与所有上述多核苷酸互补的多核苷酸。

可以将所述多核苷酸插入到合适的质粒、重组微生物载体或活的重组微生物中，并且可以供免疫接种使用(参见例如 Wolff 等, Science 247:1465-1468 (1990); Corr 等, J. Exp. Med. 184:1555-1560 (1996); Doe 等, Proc. Natl. Acad. Sci. 93:8578-8583 (1996))。因此，本发明提供了包含如上定义的所述多核苷酸的表达载体或活的重组微生物。

本发明还提供 CASB7439 多核苷酸的片段，其中当给予受治疗者时所述片段的免疫原性特性与以下的多核苷酸的免疫原性特性相同：SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID

NO:8、SEQ ID NO:9、SEQ ID NO:13 或 SEQ ID NO:15。

本发明还提供编码如前面定义的 CASB7439 多肽的免疫性片段的多核苷酸。

所述片段的免疫原性活性水平为 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10 或 SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 中所示的多肽序列或者由 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:13 或 SEQ ID NO:15 中所示的多核苷酸序列编码的多肽序列的免疫原性活性水平的至少约 50%、优选至少约 70%、更优选至少约 90%。

按照本发明的多肽片段最好包含至少约 5 个、10 个、15 个、20 个、25 个、50 个或 100 个连续氨基酸或者 100 个以上的连续氨基酸，包括所有居中长度的本文所示的多肽组合物，例如 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14 中所示的多肽组合物，或者由 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:13 或 SEQ ID NO:15 的序列中所示的多核苷酸序列编码的多肽组合物。

SEQ ID NO:1 的核苷酸序列是包含编码 193 个氨基酸的多肽即 SEQ ID NO:2 的多肽的多肽编码序列(核苷酸 545-1126)的 cDNA 序列。编码 SEQ ID NO:2 的多肽的核苷酸序列可以与在 SEQ ID NO:1 中所含的多肽编码序列相同，或者它可以是不同于在 SEQ ID NO:1 中所含的多肽编码序列的序列，由于遗传密码的冗余性(简并性)，所述核苷酸序列也编码 SEQ ID NO:2 的多肽。SEQ ID NO:2 的多肽在结构上与 *Achaete Scute* 家族的其它蛋白相关，并且也命名为“人 *Achaete Scute* 同源物 2” (HASH2) (登记号 NP_005161 和 AAB86993)。

人 *Achaete Scute* 同源物 2 (HASH2) 基因—正式命名为人 ASCL2 (*Achaete Scute* 复合物样 2) 是果蝇属 (*Drosophila*) *Achaete* 和 *Scute* 基因的同源物。人 ASCL2 仅在发育中的胎盘的滋养层外绒毛 (extravillus

trophoblast)中表达,并且作图至染色体 11p15 上靠近 IGF2 和 H19。小鼠 *achaete-scute* 同系物-2 基因(MASH2)编码在滋养层的发育中起作用的转录因子。Mash2 基因在小鼠中是父性遗传的,并且在非恶性水泡状(单雌生殖(*androgenetic*))胎块中缺乏人 ASCL2 的表达,表明人 *Ascl2* 也在人类中遗传。

Ascl2 基因是转录因子的碱性螺旋-环-螺旋(BHLH)家族的成员。它们通过与 E 框(5'-CANNTG-3')结合而激活转录。与其它 BHLH 蛋白二聚体化是有效 DNA 结合所需的。它们在黑腹果蝇(*Drosophila melanogaster*)以及可能在哺乳动物中参与决定周围神经系统和中枢神经系统中的神经元前体。

SEQ ID NO:1 核苷酸序列的互补链是 SEQ ID NO:6 的多核苷酸序列。该链也包含两种其它多肽编码序列。第一种多肽编码序列(SEQ ID NO:1 的核苷酸 1184-399, SEQ ID NO:6 的核苷酸 608-1393)编码一种 262 个氨基酸的多肽,即 SEQ ID NO:3 的多肽。第二种多肽编码序列(SEQ ID NO:1 的核苷酸 840-262, SEQ ID NO:6 的核苷酸 952-1530)编码一种 193 个氨基酸的多肽,即 SEQ ID NO:11 的多肽。编码 SEQ ID NO:3 和 SEQ ID NO:11 的多肽的核苷酸序列可以与在 SEQ ID NO:6 中所含的多肽编码序列相同,或者它可以是不同于在 SEQ ID NO:6 中所含的多肽编码序列的序列,由于遗传密码的冗余性(简并性),所述核苷酸序列也编码 SEQ ID NO:3 和 SEQ ID NO:11 的多肽。SEQ ID NO:3 的多肽在结构上与剪接辅激活蛋白家族的其它蛋白相关,与人(*Homo sapiens*)剪接辅激活蛋白亚基 srm300 (Genbank 登记号 AAF21439)具有同源性和/或结构相似性。SEQ ID NO:11 的多肽与任何已知的蛋白不相关。SEQ ID NO:3 和 SEQ ID NO:11 中所示的多肽序列以及 SEQ ID NO:6 中所示的多核苷酸序列是新的序列,并且也构成本发明的一部分。

预期本发明的优选多肽和多核苷酸除了别的以外,与其同源多肽和多核苷酸的生物功能/特性相似。此外,本发明优选的多肽、免疫性

片段和多核苷酸具有 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3 或 SEQ ID NO:11 的适当的至少一种活性。

本发明还涉及部分或其它不完全的多核苷酸序列和多肽序列，所述序列是在测定 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3 和 SEQ ID NO:11 的相应全长序列之前首先鉴定的。

因此，再一方面，本发明提供分离的多核苷酸，所述多核苷酸：

(a) 包含在 SEQ ID NO:4 和 SEQ ID NO:5 的全长上与 SEQ ID NO:4 和 SEQ ID NO:5 具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、甚至更优选至少 97-99% 同一性的核苷酸序列；

(b) 具有在 SEQ ID NO:4 和 SEQ ID NO:5 的全长上分别与 SEQ ID NO:1 或 SEQ ID NO:6 具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、甚至更优选至少 97-99% 同一性的核苷酸序列；

(c) SEQ ID NO:4 和 SEQ ID NO:5 的多核苷酸；或

(d) 编码多肽的核苷酸序列以及 SEQ ID NO:4 和 SEQ ID NO:5 的多核苷酸，所述多肽在 SEQ ID NO:2 和 SEQ ID NO:7 的全长上分别与 SEQ ID NO:2 和 SEQ ID NO:7 的氨基酸序列具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、甚至更优选至少 97-99% 同一性。

本发明还提供这样的多肽，所述多肽：

(a) 包含在 SEQ ID NO:2 或 SEQ ID NO:7 的全长上与 SEQ ID NO:2 和 SEQ ID NO:7 的氨基酸序列具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、最优选至少 97-99% 同一性的氨基酸序列；

(b) 具有在 SEQ ID NO:2 或 SEQ ID NO:7 的全长上与 SEQ ID NO:2 或 SEQ ID NO:7 的氨基酸序列具有至少 70% 同一性、优选至少 80% 同一性、更优选至少 90% 同一性、再更优选至少 95% 同一性、最优选至

少 97-99%同一性的氨基酸序列;

(c) 包含 SEQ ID NO:2 或 SEQ ID NO:7 的氨基酸; 和

(d) 是 SEQ ID NO:7 的多肽;

以及本发明还提供由包含在 SEQ ID NO:4 和 SEQ ID NO:5 中所含的序列的多核苷酸编码的多肽。

可以采用标准克隆和筛选技术, 从人结肠癌细胞中的 mRNA 衍生的 cDNA 文库中获得本发明的多核苷酸, (例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual*, 第二版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989))。本发明的多核苷酸也可以得自诸如基因组 DNA 文库的天然来源, 或者可以采用众所周知的在商业上可获得的技术进行合成。

当用本发明的多核苷酸重组生产本发明的多肽时, 所述多核苷酸可以包括成熟多肽的编码序列本身; 或在读框中还有其它编码序列的成熟多肽的编码序列, 所述其它编码序列例如编码前导序列或分泌序列、前蛋白(pre-protein)序列或原蛋白(pro-protein)序列、或前原蛋白(prepro-protein)序列或其它融合肽部分的序列。例如, 可以编码有利于纯化所融合的多肽的标记序列。在本发明该方面的某些优选的实施方案中, 所述标记序列为在 pQE 载体(Qiagen, Inc.)中提供并在 Gentz 等, *Proc Natl Acad Sci USA* (1989) 86:821-824 中描述的六组氨基酸, 或者是 HA 标记。所述多核苷酸也可以含有 5'和 3'非编码序列, 例如转录但不翻译的序列、剪接信号和聚腺苷酸化信号、核糖体结合位点和稳定 mRNA 的序列。

本发明的其它实施方案包括编码多肽变异体的多核苷酸, 所述多肽变异体包含 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:11、SEQ ID NO:13 或 SEQ ID NO:15 的氨基酸序列, 其中若干个、例如 5-10 个、1-5 个、1-3 个、1-2 个或 1 个氨基酸以任何组合发生取代、缺失或添加。

与 SEQ ID NO:1 或 SEQ ID NO:6 中所含的核苷酸序列相同或足够

相同的多核苷酸可以用作 cDNA 和基因组 DNA 的杂交探针或者用作核酸扩增(PCR)反应的引物,以分离编码本发明多肽的全长 cDNA 和基因组克隆,以及分离与 SEQ ID NO:1 或 SEQ ID NO:6 具有高度序列相似性的其它基因(包括编码来源于人的共生同源物和来源于除人类以外的物种的直向同源物和共生同源物的基因)的 cDNA 和基因组克隆。通常,这些核苷酸序列与参比物的序列有 70%相同、优选 80%相同、更优选 90%相同、最优选 95%相同。所述探针或引物总的来讲包含至少 15 个核苷酸、最好是至少 30 个核苷酸,并且可以具有至少 50 个核苷酸。特别优选的探针具有 30-50 个核苷酸。特别优选的引物具有 20-25 个核苷酸。具体地说,从同种动物来源序列获得的多肽或多核苷酸可以用作免疫原,以获得与人基因有交叉反应的免疫应答。

编码本发明多肽包括来自除人以外的物种的同系物在内的多核苷酸,可以通过这样的方法获得:所述方法包括在严格杂交条件下,用具有 SEQ ID NO:1 或 SEQ ID NO:6 的序列或其片段的标记探针,筛选合适的文库;并且分离含有所述多核苷酸序列的全长 cDNA 和基因组克隆的步骤。这样的杂交技术是本领域技术人员众所周知的。优选的严格杂交条件包括在包含以下组分的溶液中于 42℃保温过夜:50%甲酰胺、5x SSC (150mM NaCl、15mM 柠檬酸三钠)、50 mM 磷酸钠(pH 7.6)、5x Denhardt 氏溶液、10%葡聚糖硫酸酯和 20 微克/ml 变性剪切的鲑精 DNA;然后在 0.1x SSC 中在约 65℃下洗涤滤膜。因此,本发明也包括在严格杂交条件下,用具有 SEQ ID NO:1 或 SEQ ID NO:6 的序列或其片段的标记探针,通过筛选合适的文库可获得的多核苷酸。

技术人员知道,在许多情况下,分离的 cDNA 序列是不完全的,因为所述多肽的编码区在所述 cDNA 的 5'端较短。

对于本领域技术人员而言,有数种获得全长 cDNA 或延伸短 cDNA 的方法可利用并且是众所周知的,例如基于 cDNA 末端快速扩增(RACE)方法的那些方法(参见例如 Frohman 等,PNAS USA 85, 8998-9002, 1988)。例如以 Marathon™技术(Clontech Laboratories Inc.)为例的

最新改进的该技术已大大简化了对较长 cDNA 的搜索。在 Marathon™ 技术中，用从选定组织中提取的 mRNA 制备 cDNA，并将“接头”序列连接到每个末端。然后使用基因特异性寡核苷酸引物和接头特异性寡核苷酸引物的组合，进行核酸扩增(PCR)，以扩增所述 cDNA “缺失的” 5'端。然后使用“嵌套”引物重复所述 PCR 反应，“嵌套”引物即是设计在所扩增的产物内退火的引物(通过是在接头序列中更远的 3'退火的接头特异性引物和已在已知的基因序列中更远的 5'退火的基因特异性引物)。然后通过 DNA 测序分析该反应的产物，随后通过使所述产物直接连接到现有 cDNA 而产生完整的序列，或者使用新的序列信息设计 5'引物来进行一个独立的全长 PCR，构建全长 cDNA。

通过本领域众所周知的方法，可以用包含表达系统的遗传工程改造的宿主细胞制备本发明的重组多肽。因此，再一方面，本发明涉及包含本发明多核苷酸的表达系统，涉及用这样的表达系统遗传工程改造的宿主细胞，还涉及通过重组技术产生本发明的多肽。也可以采用衍生自本发明的 DNA 构建体的 RNA，使用无细胞翻译系统生产这样的蛋白质。

为了重组生产，可以遗传工程改造宿主细胞以掺入本发明多核苷酸的表达系统或其部分。可以采用许多标准实验室手册中介绍的方法，实现将多核苷酸导入宿主细胞，所述实验室手册例如 Davis 等，Basic Methods in Molecular Biology (1986)和 Sambrook 等，Molecular Cloning: A Laboratory Manual, 第二版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)。优选的这类方法包括例如磷酸钙转染、DEAE-葡聚糖介导的转染、转位、微注射、阳离子脂质介导的转染、电穿孔、转导、擦伤加样(scrape loading)、基因枪引入(ballistic introduction)或感染。

本发明的蛋白最好与反式硫氧还蛋白(TIT)共表达。优选反式与顺式硫氧还蛋白共表达，以保持抗原无硫氧还蛋白，而不需要蛋白酶。硫氧还蛋白共表达使本发明的蛋白易于溶解。硫氧还蛋白共表达还对

蛋白纯化得率、对纯化蛋白的溶解度和质量产生显著影响。

合适宿主的代表性实例包括细菌细胞，如链球菌(*Streptococci*)、葡萄球菌(*Staphylococci*)、大肠杆菌、链霉菌属(*Streptomyces*)和枯草芽孢杆菌(*Bacillus subtilis*)细胞；真菌细胞，例如酵母细胞和曲霉属(*Aspergillus*)的细胞；昆虫细胞，例如果蝇属(*Drosophila*) S2 细胞和贪夜蛾(*Spodoptera*) Sf9 细胞；动物细胞，例如 CHO、COS、HeLa、C127、3T3、BHK、HEK 293 和 Bowes 黑素瘤细胞；和植物细胞。

可以使用各种各样的表达系统，例如染色体衍生的系统、附加体衍生的系统和病毒衍生的系统，例如衍生自以下的载体：细菌质粒、噬菌体、转座子、酵母附加体、插入元件、酵母染色体元件；衍生自以下病毒的载体：杆状病毒、乳多空病毒(如 SV40)、痘苗病毒、腺病毒、禽痘病毒、假狂犬病病毒和反转录病毒；以及衍生自其组合的载体，例如衍生自质粒和噬菌体遗传元件的那些载体，诸如粘粒和噬菌粒。所述表达系统可以包含调节以及产生表达的控制区。一般而言，可以使用能够在宿主中维持、繁殖或表达多核苷酸以产生多肽的任何系统或载体。可以通过多种众所周知的常规技术中的任何一种，将合适的核苷酸序列插入表达系统中，所述常规技术例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual* (参见上文)所介绍的技术。为了让翻译的蛋白分泌到内质网腔、壁膜间隙或胞外环境中，可以将合适的分泌信号掺入到所需的多肽中。这些信号对于所述多肽可以是内源的，或者它们可以是异源信号。

所述表达系统也可以是活的重组微生物，例如病毒或细菌。可以将感兴趣的基因插入活的重组病毒或细菌的基因组中。用该活的载体接种和体内感染，将使所述抗原在体内表达并诱导免疫应答。

因此，在某些实施方案中，采用众多已知基于病毒的系统中的任一系统，将编码本发明免疫原性多肽的多核苷酸导入用于表达的合适哺乳动物宿主细胞中。在一个说明性的实施方案中，反转录病毒提供用于基因传递系统的便利有效的平台。采用本领域已知的技术，可以

将编码本发明多肽的选定核苷酸序列插入到载体中，然后在反转录病毒颗粒中包装。然后可以分离所述重组病毒，将其给予受治疗者。许多说明性的反转录病毒系统已有介绍(例如美国专利第 5,219,740 号; Miller 和 Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa 等(1991) *Virology* 180:849-852; Burns 等(1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; 和 Boris-Lawrie 和 Temin (1993) *Cur. Opin. Genet. Develop.* 3:102-109)。

另外，许多说明性的基于腺病毒的系统也有介绍。与整合到宿主基因组中的反转录病毒不同，腺病毒一直位于染色体外，因而将与插入诱变相关的危险减至最小(Haj-Ahmad 和 Graham (1986) *J. Virol.* 57:267-274; Bett 等(1993) *J. Virol.* 67:5911-5921; Mittereder 等(1994) *Human Gene Therapy* 5:717-729; Seth 等(1994) *J. Virol.* 68:933-940; Barr 等(1994) *Gene Therapy* 1:51-58; Berkner, K. L. (1988) *BioTechniques* 6:616-629; 和 Rich 等(1993) *Human Gene Therapy* 4:461-476)。

还开发出各种腺伴随病毒(AAV)载体系统用以传递多核苷酸。采用本领域众所周知的技术，可以容易地构建 AAV 载体。参见，例如，美国专利第 5,173,414 号和第 5,139,941 号；国际公布号 WO 92/01070 和 WO 93/03769；Lebkowski 等(1988) *Molec. Cell. Biol.* 8:3988-3996；Vincent 等(1990) *Vaccines 90* (Cold Spring Harbor Laboratory Press)；Carter, B. J. (1992) *Current Opinion in Biotechnology* 3:533-539；Muzyczka, N. (1992) *Current Topics in Microbiol. and Immunol.* 158:97-129；Kotin, R. M. (1994) *Human Gene Therapy* 5:793-801；Shelling 和 Smith (1994) *Gene Therapy* 1:165-169；和 Zhou 等(1994) *J. Exp. Med.* 179:1867-1875。

可用于通过基因转移传递编码本发明多肽的核酸分子的其它病毒载体包括衍生自痘病毒科(例如痘苗病毒和禽痘病毒)的病毒载体。作为实例，可以如下构建表达所述新分子的痘苗病毒重组体。首先将编码多肽的 DNA 插入到合适的载体中，使得它邻接一个痘苗病毒启动子

和侧翼痘苗病毒 DNA 序列, 例如编码胸苷激酶(TK)的序列。然后用该载体转染同时用痘苗病毒感染的细胞。同源重组使痘苗病毒启动子加上编码感兴趣的多肽的基因插入到病毒基因组中。通过在 5-溴脱氧尿苷存下下培养细胞, 然后挑出抗性病毒空斑, 可以选择所产生的 TK.sup.(-)重组体。

可以方便地用基于痘苗病毒的感染/转染系统在某一生物的宿主细胞内提供本文所述的一种或多种多肽的诱导型瞬时表达或共表达。在这一特定系统中, 首先在体外用编码噬菌体 T7 RNA 聚合酶的痘苗病毒重组体感染细胞。该聚合酶表现出灵敏的特异性, 因为它只转录带有 T7 启动子的模板。感染后, 细胞用由 T7 启动子驱动的一种或多种感兴趣的多核苷酸转染。来自痘苗病毒重组体在胞质内表达的聚合酶将转染的 DNA 转录成 RNA, 然后通过宿主翻译机器将其翻译成多肽。所述方法可供用于高水平、瞬时、胞质内生产大量的 RNA 及其翻译产物。参见, 例如, Elroy-Stein 和 Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst 等, Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126。

或者, 也可用诸如禽痘病毒(fowlpox virus)和金丝雀痘病毒的禽痘病毒(avipoxvirus)传递感兴趣的编码序列。已知表达得自哺乳动物病原体的免疫原的重组禽痘病毒(avipox virus)当给予非禽类物种时提供保护性免疫。在人和其它哺乳动物物种中应用禽痘病毒载体是特别理想的, 这是由于禽痘病毒属的成员仅可以在易感的禽类物种中生产性地复制, 因此它们在哺乳动物细胞内是没有感染力的。制备重组禽痘病毒的方法在本领域是已知的, 并且使用如上所述的关于制备痘苗病毒的遗传重组。参见, 例如, WO 91/12882; WO 89/03429 和 WO 92/03545。

还可以用众多甲病毒载体中的任一种传递本发明的多核苷酸组合物, 例如在美国专利号 5,843,723、6,015,686、6,008,035 和 6,015,694 中介绍的那些载体。也可以使用基于委内瑞拉马脑炎(VEE)病毒的某些

载体，其说明性的实例可以在美国专利第 5,505,947 号和第 5,643,576 号中找到。

此外，还可以使用 Michael 等 *J. Biol. Chem.* (1993) 268:6866-6869 和 Wagner 等 *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103 中介绍的诸如腺病毒嵌合载体的分子缀合载体来进行本发明的基因传递。

有关这些和其它已知的基于病毒的传递系统的其它说明性资料可以在以下的文献中找到：例如 Fisher-Hoch 等, *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner 等, *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner 等, *Vaccine* 8:17-21, 1990; 美国专利号 4,603,112、4,769,330 和 5,017,487; WO 89/01973; 美国专利号 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld 等, *Science* 252:431-434, 1991; Kolls 等, *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler 等, *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman 等, *Circulation* 88:2838-2848, 1993; 和 Guzman 等, *Cir. Res.* 73:1202-1207, 1993.

上述活的重组微生物可以是有毒力的，或者是以各种方式减毒以获得活疫苗。这样的活疫苗也构成本发明的一部分。

在某些实施方案中，可以将多核苷酸整合到靶细胞的基因组中。这种整合可以在特定位置以特定方向通过同源重组进行(基因置换)，或者它可以在随机、非特定位置内整合(基因扩增)。在另外一些实施方案中，所述多核苷酸可以在细胞内作为独立的附加型 DNA 区段稳定地保持。这样的多核苷酸区段或“附加体”编码的序列足以允许其保持并且独立于宿主细胞周期进行复制或者与宿主细胞周期同步复制。表达构建体传递至细胞的方式以及所述多核苷酸保持在所述细胞内的位置取决于所用的表达构建体的类型。

在本发明的另一实施方案中，例如如在 Ulmer 等, *Science* 259:1745-1749, 1993 中的描述和 Cohen, *Science* 259:1691-1692, 1993 的综述，多核苷酸作为“裸”DNA 给予/传递。通过将裸 DNA 包被到被

有效地转运到细胞内的生物可降解的微珠上,可以增加所述 DNA 的摄入。

在再一实施方案中,本发明的组合物可以通过粒子轰击法传递,其中许多方法已有介绍。在一个说明性的实施例中,用诸如 Powderject Pharmaceuticals PLC (Oxford, UK)和 Powderject Vaccines Inc. (Madison, WI)生产的装置可以实现气动粒子加速,其中某些实施例描述于美国专利号 5,846,796、6,010,478、5,865,796、5,584,807 和欧洲专利号 0500 799。该方法提供一种无针传递方法,其中诸如多核苷酸或多肽粒子的微粒干粉制剂在由 hand held device 产生的氦气喷射流中被加速至高速,推动所述粒子到达感兴趣的靶组织。

在一个相关实施方案中,可以用于气动无针注射本发明组合物的其它装置和方法包括 Bioject, Inc. (Portland, OR)提供的装置和方法,其中某些实施例描述于美国专利号 4,790,824、5,064,413、5,312,335、5,383,851、5,399,163、5,520,639 和 5,993,412。

通过众所周知的方法,可以从重组细胞培养物中回收并纯化本发明的多肽,所述方法包括硫酸铵沉淀或乙醇沉淀、酸提取、阴离子交换层析或阳离子交换层析、磷酸纤维素层析、疏水作用层析、亲和层析、羟基磷灰石层析和凝集素层析。最优选使用离子金属亲和层析 (IMAC)进行纯化。当所述多肽在胞内合成、分离和/或纯化过程中变性时,可以使用熟知的重折叠蛋白技术重建活性构象。

本发明的另一重要方面涉及诱导、加强或调节哺乳动物体内免疫应答的方法,所述方法包括用足以产生抗体和/或 T 细胞免疫应答的本发明的完整多肽或多核苷酸或其片段接种哺乳动物,用于免疫预防或用于治疗性治疗癌症、更特别是结肠直肠癌和自身免疫病及相关病症。本发明的再一方面涉及诱导、加强或调节哺乳动物体内免疫应答的方法,所述方法包括通过在体内指导表达所述多核苷酸和编码所述多肽的载体或细胞来传递本发明的多肽,以便诱导这样的免疫应答,以产生预防或治疗所述动物免患疾病的免疫应答。

本发明的再一方面涉及免疫制剂/疫苗制剂(组合物)及其在医药领域的应用。当所述组合物被引入哺乳动物宿主后,在该哺乳动物体内诱导、加强或调节针对本发明多肽的免疫应答,其中所述组合物包含如本文前面所限定的本发明的多肽或多核苷酸或其免疫学片段。更具体地说,按照本发明的免疫原性组合物包含安全有效量的 CASB7439 多肽或其免疫原性片段,其中所述 CASB7439 多肽选自 SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12 或 SEQ ID NO:14。在另一个实施方案中,所述免疫原性组合物包含安全有效量的编码 CASB7439 的多核苷酸或其片段,其中所述编码 CASB7439 的多核苷酸选自 SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:13 或 SEQ ID NO:15。

按照本发明的疫苗制剂还可以包含一种合适的载体,即药学上可接受的载体。由于多肽可能在胃中分解,因此它们最好是胃肠外给予(例如皮下注射、肌肉注射、静脉内注射或皮内注射)。适合于胃肠外给药的制剂包括水性和非水无菌注射液,所述注射液可以包含抗氧化剂、缓冲剂、抑菌剂和使得所述制剂与受体血液等渗的溶质;以及可以包含悬浮剂或增稠剂的水性和非水无菌悬浮液。所述制剂可以盛装在单位剂量容器或多剂量容器(例如密封安瓿和管形瓶)中,并且可以在冻干条件下保存,只需要在临用前加入无菌液体载体即可。

本发明的再一方面涉及采用来自哺乳动物免疫系统的细胞,在体外诱导针对本发明的完整多肽或多核苷酸或其片段的免疫应答或者是针对包含本发明多肽或多核苷酸的分子的免疫应答,并且将这些活化免疫细胞回输给所述哺乳动物以治疗疾病。通过与本发明完整多肽或多核苷酸或者包含本发明多肽或多核苷酸的分子在有或无各种免疫调节分子的情况下体外孵育,达到活化来自所述免疫系统的所述细胞。本发明的再一方面涉及通过给予抗原呈递细胞免疫哺乳动物,所述抗原呈递细胞通过在体外加载本发明的完整多肽或其部分或者包含本发

明多肽的分子而被修饰，然后以免疫原性方式将所述抗原呈递细胞给予体内。或者，可以用含有本发明的完整多核苷酸或其片段或者包含本发明多核苷酸的分子的载体，体外转染抗原呈递细胞，例如以表达相应的多肽，然后以免疫原性方式将所述抗原呈递细胞给予体内。因此，本发明的药用组合物包含有效量的抗原呈递细胞和药学上有效的载体，所述抗原呈递细胞通过在体外加载 CASB7439 多肽而被修饰或者在体外经遗传修饰而表达 CASB7439 多肽。

按照另一个实施方案，除本发明的免疫原性多核苷酸、多肽、抗体、T 细胞和/抗原呈递细胞(APC)组合物外，本文所述的药用组合物/免疫原性组合物还包含一种或多种免疫刺激剂。因此，本文提供生产所述免疫原性组合物方法，所述方法包括将 CASB7439 多肽或编码 CASB7439 的多核苷酸与合适的佐剂/免疫刺激剂、稀释剂或其它药学上可接受的载体混合。免疫刺激剂实质上是指提高或增强对外源性抗原的免疫应答(抗体和/细胞介导)的任何物质。一种优选类型的免疫刺激剂包括佐剂。许多佐剂含有设计用以保护抗原免遭快速分解的物质，例如氢氧化铝或矿物油，以及含有免疫应答的刺激剂，例如脂质 A，百日咳博德特氏菌 (*Bordetella pertussis*) 或结核分枝杆菌 (*Mycobacterium tuberculosis*) 源性蛋白。某些佐剂是市售的，例如弗氏不完全佐剂和弗氏完全佐剂(Difco Laboratories, Detroit, MI); Merck 佐剂 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); 诸如氢氧化铝凝胶(明矾)或磷酸铝的铝盐; 钙盐、铁盐或锌盐; 酰化酪氨酸的不溶性悬浮剂; 酰化糖; 阳离子或阴离子衍生化的多糖; 聚磷腈; 生物可降解的微球体; 单磷脂酰脂质 A 和 quil A。诸如 GM-CSF、白介素-2、白介素-7、白介素-12 的细胞因子和其它类似的生长因子也可以用作佐剂。

在本发明的某些实施方案中，所述佐剂组合物最好是主要诱导 Th1 型免疫应答的佐剂组合物。高水平的 Th1 型细胞因子(例如 IFN- γ 、TNF α 、IL-2 和 IL-12)往往促进诱导针对所给予抗原的细胞介导的免疫

应答。相反，高水平的 Th2 型细胞因子(例如 IL-4、IL-5、IL-6 和 IL-10)往往促进诱导体液免疫应答。应用本文提供的疫苗后，患者将支持包括 Th1 型和 Th2 型应答的免疫应答。在应答主要为 Th1 型的一个优选实施方案中，Th1 型细胞因子的水平比 Th2 型细胞因子的水平增加的程度更大。采用标准测定可以容易地评价这些细胞因子的水平。有关细胞因子家族的综述，请参见 Mosmann 和 Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989。

供诱导主要为 Th1 型应答的某些优选佐剂包括例如单磷脂酰脂质 A、最好是 3-脱-O-酰化单磷脂酰脂质 A 与铝盐一起的组合。MPL[®]佐剂可得自 Corixa Corporation (Seattle, WA; 参见, 例如, 美国专利号 4,436,727、4,877,611、4,866,034 和 4,912,094)。含 CpG 的寡核苷酸(其中 CpG 二核苷酸未被甲基化)也诱导主要为 Th1 型的应答。这样的寡核苷酸是众所周知的, 描述于例如 WO 96/02555、WO99/33488 和美国专利号 6,008,200 和 5,856,642。免疫刺激 DNA 序列也描述于例如 Sato 等, *Science* 273:352, 1996。另一优选的佐剂包含诸如 Quil A 的皂苷或其衍生物, 包括 QS21 和 QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); 七叶皂苷; 毛地黄皂苷; 或丝石竹属(*Gypsophila*)或 *Chenopodium quinoa* 皂苷。在本发明的佐剂组合中, 其它优选的制剂包括一种以上的皂苷, 例如以下至少两种的组合: QS21、QS7、Quil A、 β -七叶皂苷或毛地黄皂苷。

或者, 可以将皂苷制剂与由脱乙酰壳多糖或其它聚阳离子聚合物、聚交酯和聚丙交酯-乙交酯共聚物颗粒、基于聚-N-乙酰葡糖胺的聚合物基质、由多糖或经化学修饰的多糖组成的颗粒、脂质体和基于脂质的颗粒、由甘油单酯组成的颗粒等组成的疫苗载体混合。皂苷也可以在胆固醇存在下进行配制, 以形成颗粒性结构, 例如脂质体或 ISCOM。此外, 皂苷可以与聚氧乙烯醚或酯一起在非颗粒性溶液或悬浮液或在诸如 paucilamellar 脂质体或 ISCOM 的颗粒性结构中进行配制。皂苷也可以与诸如 Carbopol^R 的赋形剂一起配制, 以增加粘度, 或

者可以以干粉形式与诸如乳糖的粉状赋形剂一起配制。

在一个优选的实施方案中，所述佐剂系统包括单磷脂酰脂质 A 和皂苷衍生物的组合，例如在 WO 94/00153 中描述的 QS21 和 3D-MPL[®] 佐剂的组合或者在 WO 96/33739 中描述的其中 QS21 用胆固醇猝灭的反应原性较低的组合物。其它优选的制剂包含水包油乳剂和生育酚。WO 95/17210 中描述了使用 QS21、3D-MPL[®] 佐剂和生育酚的水包油乳剂的另一特别优选的佐剂制剂。

另一增强的佐剂系统包括含 CpG 寡核苷酸和皂苷衍生物的组合，尤其是 WO 00/09159 中公开的 CpG 和 QS21 的组合。所述制剂最好另外包含水包油乳剂和生育酚。

可供用于本发明药用组合物中的其它说明性佐剂包括 Montanide ISA 720 (Seppic, 法国)、SAF (Chiron, California, 美国)、ISCOMS (CSL)、MF-59 (Chiron)、SBAS 系列的佐剂(例如 SBAS-2 或 SBAS-4, 可得自 SmithKline Beecham, Rixensart, 比利时)、Detox (Enhanzyn[®]) (Corixa, Hamilton, MT)、RC-529 (Corixa, Hamilton, MT)和其它氨基氨基葡萄糖苷 4-磷酸酯(AGP), 例如待审的美国专利申请顺序号 08/853,826 和 09/074,720 中描述的佐剂(所述专利申请的公开内容通过引用全部结合到本文中)以及聚氧乙烯醚佐剂，例如 WO 99/52549A1 中描述的佐剂。

其它优选的佐剂包括通式(I)的佐剂分子：



其中 n 为 1-50，A 为一个键或 -C(O)-，R 为 C₁₋₅₀ 烷基或苯基 C₁₋₅₀ 烷基。

本发明的一个实施方案包括含通式(I)的聚氧乙烯醚的疫苗制剂，其中 n 为 1-50、优选为 4-24、最优选为 9；R 组分为 C₁₋₅₀ 烷基、优选为 C₄₋₂₀ 烷基、最优选为 C₁₂ 烷基；A 为一个键。聚氧乙烯醚的浓度范围应该为 0.1-20%、优选为 0.1-10%、最优选为 0.1-1%。优选的聚氧乙烯醚选自以下组：聚氧乙烯-9-十二烷基醚、聚氧乙烯-9-硬脂酰基醚、

聚氧乙烯-8-硬脂酰基醚、聚氧乙烯-4-十二烷基醚、聚氧乙烯-35-十二烷基醚和聚氧乙烯-23-十二烷基醚。诸如聚氧乙烯十二烷基醚的聚氧乙烯醚在 Merck index (第 12 版: entry 7717) 中有描述。这些佐剂分子在 WO 99/52549 中有描述。

如果需要, 按照上述通式(I)的聚氧乙烯醚可以与另一佐剂组合。例如一种优选的佐剂组合最好是与待审的英国专利申请 GB 9820956.2 中描述的 CpG 组合。

按照本发明的疫苗组合物中最好还存在一种载体。所述载体可以是水包油乳剂, 或者是一种铝盐, 例如磷酸铝或氢氧化铝。

优选的水包油乳剂包含一种可代谢的油, 例如角鲨烯、 α -生育酚和吐温 80。在一个特别优选的方面, 按照本发明疫苗组合物中的抗原与这种乳剂中的 QS21 和 3D-MPL 混合。另外, 所述水包油乳剂可以含有 span 85 和/或卵磷脂和/或三辛酰甘油酯。

给予人时, 通常疫苗中的 QS21 和 3D-MPL 含量范围为每剂 1 μ g-200 μ g、例如 10 μ g-100 μ g、最好是 10 μ g-50 μ g。而水包油乳剂通常包含 2-10%角鲨烯、2-10% α -生育酚和 0.3-3%吐温 80。角鲨烯: α -生育酚的比例最好等于或小于 1, 因为这样的比例提供更稳定的乳剂。Span 85 的含量也可以为 1%。在某些情况下, 本发明的疫苗另外含有一种稳定剂可能是有利的。

无毒的水包油乳剂最好在含水载体中含有一种无毒的油例如角鲨烷或角鲨烯, 以及一种乳化剂例如吐温 80。所述含水载体可以是例如磷酸缓冲盐溶液。

WO 95/17210 中描述了一种特别有效的佐剂制剂, 该制剂为包含 QS21、3D-MPL 和生育酚的水包油乳剂。

本发明也提供多价疫苗组合物, 所述多价疫苗组合物包含本发明疫苗制剂以及其它抗原, 尤其是可用于治疗癌症、更特别是结肠直肠癌、自身免疫病和相关病症的抗原。这样的多价疫苗组合物可以包括如上文所述的 TH-1 诱导佐剂。

按照本发明的另一个实施方案，将本文所述的免疫原性组合物通过抗原呈递细胞(APC)传递至宿主，所述抗原呈递细胞(APC)例如树突细胞、巨噬细胞、B 细胞、单核细胞和其它可以被工程改造以成为有效的 APC 的细胞。这类细胞可以但不一定被遗传修饰，以增加呈递抗原的能力、改进 T 细胞应答的活化和/或保持、本身具有抗肿瘤效应和/或与受体在免疫学上是相容的(即匹配 HLA 单倍型)。APC 一般可以从各种各样的生物体液或器官(包括肿瘤和肿瘤周围组织)的任一种中分离出来，或可以是自体细胞、同种异体细胞、同种同基因细胞或异种细胞。

本发明的某些优选的实施方案使用树突细胞或其祖细胞作为抗原呈递细胞。树突细胞是非常有效的 APC (Banchereau 和 Steinman, *Nature* 392:245-251,1998)，并且已表明作为诱导预防性或治疗性抗肿瘤免疫的生理佐剂是有效的(参见 Timmerman 和 Levy, *Ann. Rev. Med.* 50:507-529, 1999)。一般而言，树突细胞可能根据其典型的形状(原位为星状，体外可见明显的胞质突(树突))、其摄入、突起和以高效率呈递抗原的能力以及其活化原初 T 细胞应答的能力来鉴定。当然，可以对树突细胞进行工程改造，以表达通常在体内或离体树突细胞上不存在的特异性细胞表面受体或配体，本发明考虑了这类经修饰的树突细胞。作为树突细胞的替代物，在疫苗中可以使用加载分泌型小泡抗原的树突细胞(称为外来体)(参见 Zitvogel 等, *Nature Med* 4:594-600, 1998)。

树突细胞和祖细胞可以从外周血、骨髓、肿瘤浸润细胞、肿瘤周围组织浸润细胞、淋巴结、脾、皮肤、脐带血或任何其它合适的组织或体液获得。例如，可以通过将细胞因子例如 GM-CSF、IL-4、IL-13 和/或 TNF α 的组合加到从外周血收获的单核细胞培养物中，使树突细胞离体分化。或者，可以通过将 GM-CSF、IL-3、TNF α 、CD40 配体、LPS、flt3 配体和/或其它诱导树突细胞分化、成熟和增殖的化合物加入到所述培养基中，使从外周血、脐带血或骨髓收获的 CD34 阳性细胞分化成树突细胞。

树突细胞常规分为“未成熟”细胞和“成熟”细胞，可用简单的

方式将两种充分表征的表型区分开来。然而，这种命名不应该解释为排除所有可能的中间分化阶段。未成熟树突细胞的特征为与 Fc γ 受体和甘露糖受体高表达相关的抗原摄入和加工能力高的 APC。成熟表型通常的特征在于这些标记的表达较低，但负责 T 细胞活化的细胞表面分子(例如 I 类 MHC 和 II 类 MHC)、粘着分子(例如 CD54 和 CD11)和共同刺激分子(例如 CD40、CD80、CD86 和 4-1BB)的表达高。

一般可以用本发明的多核苷酸(或其部分或其它变异性)转染 APC，使得所编码的多肽或其免疫原性部分在所述细胞表面表达。可以离体进行这样的转染，然后包含这样的转染细胞的药用组合物可以用于本文所述的治疗目的。或者，可以将靶向树突细胞或其它抗原呈递细胞的基因传递载体给予患者，导致在体内发生转染。一般而言，采用本领域已知的任何方法，例如 WO 97/24447 中介绍的方法或 Mahvi 等, *Immunology and cell Biology* 75:456-460, 1997 介绍的基因枪方法，可以进行例如树突细胞的体内和离体转染。通过将树突细胞或祖细胞与肿瘤多肽、DNA (裸 DNA 或在质粒载体内的 DNA)或 RNA 一起孵育；或者与表达抗原的重组细菌或病毒(例如痘苗病毒、禽痘病毒、腺病毒或慢病毒属载体)一起孵育，可以达到树突细胞的抗原加载。在加载前，可以将多肽与提供 T 细胞辅助的免疫配偶体(例如载体分子)共价缀合。或者，树突细胞可以单独地或在所述多肽存在下用未缀合的免疫配偶体脉冲处理。

在本发明药用组合物中可以使用本领域技术人员已知的任何合适载体的情况下，载体类型通常随给药模式而变。本发明的组合物可以配制用于任何合适的给药方式，包括例如局部、口服、鼻腔、粘膜、静脉内、颅内、腹膜内、皮下和肌肉内给药。

供这类药用组合物使用的载体是生物相容的，并且也可是生物可降解的。在某些实施方案中，所述制剂最好提供以相对恒定水平释放有效成分。然而，在其它实施方案中，可能需要在给药后以更加快的速率立即释放有效成分。采用已知技术配制这类组合物是在本领域技

术人员的知识范围内。可用于该方面的说明性载体包括聚丙交酯-乙交酯共聚物、聚丙烯酸酯、胶乳、淀粉、纤维素、葡聚糖的微粒等。其它说明性延迟释放的载体包括超分子生物载体，它包含不流动的亲水核心(例如交联多糖或寡糖)以及任选的包含两亲化合物的外层(例如磷脂)(参见例如，美国专利第 5,151,254 号和 PCT 申请 WO 94/20078、WO 94/23701 和 WO 96/06638)。在缓释制剂中所含的活性化合物的量取决于植入部位、释放的速率和预期持续时间以及待治疗或预防的病症性质。

在另一个说明性实施方案中，生物可降解的微球体(例如聚乳酸酯聚乙醇酸酯)用作本发明组合物的载体。合适的生物可降解的微球体公开于例如美国专利号 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 和 5,942,252。经修饰的乙型肝炎核心蛋白载体系统，例如 WO/99 40934 和其中引用的参考文献中描述的载体系统，也可用于许多应用中。另一说明性载体/传递系统使用包含颗粒性蛋白复合物的载体，例如在美国专利第 5,928,647 号中描述的载体，所述载体能够在宿主体内诱导 I 类限制的细胞毒性 T 淋巴细胞应答。

本发明的药用组合物通常还包含一种或多种缓冲剂(例如中性缓冲盐溶液或磷酸缓冲盐溶液)、糖类(例如葡萄糖、甘露糖、蔗糖或葡聚糖)、甘露醇、蛋白质、多肽或氨基酸(例如甘氨酸)、抗氧化剂、抑菌剂、螯合剂(例如 EDTA)或谷胱甘肽、佐剂(例如氢氧化铝)、使制剂与受体的血液等渗、低渗或弱高渗的溶质、悬浮剂、增稠剂和/或防腐剂。或者，本发明组合物可以配制为冻干制剂。

本文所述的药用组合物可以盛装在单位剂量容器或多剂量容器(例如密封安瓿或管形瓶)中。这类容器通常的密封方式使得保持所述制剂在使用之前的无菌和稳定性。一般而言，制剂可以作为油性或水性溶媒中的混悬剂、溶液剂或乳剂保存。或者，药用组合物可以在冷冻-干燥条件下保存，只需要在临用前加入无菌液体载体即可。

对于在各种各样的治疗方案中使用本文所述的特定组合物而言，合适给药方案和治疗方案(包括例如口服、胃肠外、静脉内、鼻内和肌肉内给药和所用制剂)的开发是本领域众所周知的，为了说明，下文将扼要地论述其中的某些方案。

在某些应用中，可以将本文公开的药用组合物通过口服给药传递至动物。因此，这些组合物可以与惰性稀释剂或与可吸收的食用载体一起配制，或者它们可以包封在硬壳明胶胶囊或软壳明胶胶囊内，或者它们可以挤压成片剂，或者它们可以直接掺入到膳食的食物中。

所述活性化合物甚至可能与赋形剂一起掺入，以可食用的片剂、口含片、锭剂、胶囊剂、酏剂、混悬剂、糖浆、糯米纸囊剂形式等使用(参见，例如，Mathiowitz等，Nature 1997年3月27日;386(6623):410-4; Hwang等，Crit Rev Ther Drug Carrier Syst 1998; 15(3):243-84; 美国专利 5,641,515; 美国专利 5,580,579 和美国专利 5,792,451)。片剂、锭剂、丸剂、胶囊剂等也可以含有许多额外组分中的任一种，可以加入例如粘合剂，例如西黄蓍胶、阿拉伯树胶、玉米淀粉或明胶；赋形剂，例如磷酸二钙；崩解剂，例如玉米淀粉、马铃薯淀粉、藻酸等；润滑剂，例如硬脂酸镁；和甜味剂，例如蔗糖、乳糖或糖精或者矫味剂，例如薄荷、冬青油或樱桃矫味剂。当剂量单位形式为胶囊时，除上述类型物质以外它还可以含有液体载体。各种其它物质可能作为包衣剂存在，或者在其它方面改进剂量单位的物理形式。例如，片剂、丸剂或胶囊剂可以用虫胶、糖或者同时用虫胶和糖包衣。当然，在制备任何剂量单位形式中所用的任何材料都应该是药学纯的并且在所使用的量内基本上是无毒的。另外，所述活性化合物可以掺入到缓释制剂中。

通常，这些制剂含有至少约 0.1%活性化合物或更高剂量的活性化合物，当然尽管有效成分的百分率可以变化，但是可能常规为总制剂重量或体积的约 1%或 2%至约 60%或 70%或更高。当然，在可能制备的每种治疗上有效的组合物中活性化合物的用量的方式使得以所述化合物的任何给定单位剂量获得合适的剂量。在制备这种药用制剂时，

本领域技术人员会考虑诸如溶解性、生物利用度、生物半寿期、给药途径、制品贮藏期的诸多因素以及其它药理学考虑，因此各种各样的剂量和治疗方案可能都是理想的。

另一方面，对于口服给药，本发明的组合物可以在漱口药、洁牙剂、口含片、口腔喷雾剂或舌下口服给药制剂形式中掺入一种或多种赋形剂。或者，可以将所述有效成分掺入到口服溶液例如含有硼酸钠、甘油和碳酸氢钠的口服溶液中，或者在洁牙剂中分散，或以治疗有效量加入到可能包括水、粘合剂、磨擦剂、矫味剂、起泡剂和润湿剂的组合物中。或者，所述组合物可以精加工成可以放入舌下或者在口腔内溶解的片剂或溶液形式。

在某些情况下，理想的是胃肠外、静脉内、肌内、或者甚至腹膜内给予本文公开的药用组合物。这样的方法是本领域技术人员熟知的，其中某些方法进一步描述于例如美国专利 5,543,158；美国专利 5,641,515 和美国专利 5,399,363。在某些实施方案中，在水中适当与表面活性剂(例如羟丙基纤维素)混合，可以制备所述活性化合物作为游离碱或药理学上可接受的盐的溶液。在甘油、液体聚乙二醇及其混合物和在油中也可以制备分散剂。在通常的贮藏和使用条件下，这些制剂一般含有防腐剂，以防止微生物的生长。

适合于注射用的说明性药物形成包括无菌的水溶液或分散剂和用于临时制备无菌的注射液或分散液的无菌粉末(例如,参见美国专利 5,466,468)。在所有情况下,所述形式必须是无菌的且在易于注射方面必须是液体。在生产和贮藏的条件下它必须是稳定的且必须防止微生物(例如细菌和真菌)的污染作用。所述载体可以是溶剂或含有例如水、乙醇、多元醇(例如甘油、丙二醇和液体聚乙二醇等)、它们的适宜混合物和/或植物油的分散介质。例如通过利用诸如卵磷脂的包衣剂,在分散的情况下通过保持所需颗粒大小和/或利用表面活性剂,可以保持适当的流动性。可以利用各种抗细菌剂或抗真菌药(例如,对羟基苯甲酸酯类、三氯叔丁醇、苯酚、山梨酸、硫柳汞等)防止微生物的

作用。在许多情况下，最好包括等渗剂，例如糖或氯化钠。注射组合物的延长吸收可以通过在组合物中使用延迟吸收的药剂，例如单硬脂酸铝和明胶来实现。

在一个实施方案中，对于胃肠外给予水溶液，必要时所述溶液应该进行适当缓冲，所述液体稀释剂首先用足量的盐水或葡萄糖使其等渗。这些特定的水溶性尤其适合于静脉内、肌内、皮下和腹膜内给药。在这一方面，根据本发明的公开内容，可以使用的无菌水性介质对于本领域技术人员是已知的。例如，可以将一个剂量溶于 1 ml 等渗氯化钠溶液中，然后或者加入到 1000 ml 皮下灌注用液体或者在建议的输注部位注射(参见例如，“Remington’s Pharmaceutical Sciences”，第 15 版，第 1035-1038 页和第 1570-1580 页)。根据待治疗的受治疗者的病症，剂量的某些变化是必需的。此外，对于人类给药，当然制剂最好符合无菌、无热原并且依照 FDA Office of Biologics 标准所要求的一般性安全和纯度标准。

在本发明的另一个实施方案中，本文公开的组合物可以配制成中性形式或盐形式。说明性的药学上可接受的盐包括酸加成盐(与蛋白质的游离氨基生成的)，并且与诸如盐酸或磷酸的无机酸或者与诸如乙酸、草酸、酒石酸、扁桃酸等的有机酸生成的酸加成盐。与游离羧基形成的盐也可衍生自无机碱，例如氢氧化钠、氢氧化钾、氢氧化铵、氢氧化钙或氢氧化铁；和有机碱，例如异丙胺、三乙胺、组氨酸、普鲁卡因等。在配制时，溶液给予的方式需与给药制剂相容，而给予的剂量为治疗有效量。

所述载体还可以包含任何和所有的溶剂、分散介质、溶媒、包衣剂、稀释剂、抗细菌剂和抗真菌药、等渗剂和吸收延迟剂、缓冲剂、载体溶液、混悬剂、胶体等。供药用活性物质使用的这类介质和药剂是本领域众所周知的。除在任何常规介质和药剂与有效成分不相容的情况外，考虑其在治疗组合物中的应用。也可以将补充的有效成分掺入到所述组合物中。短语“药学上可接受的”是指当给予人类时不产

生变态反应或相似的不想要的反应的分子实体和组合物。

在某些实施方案中，所述药用组合物可以通过鼻内喷雾、吸入和/或其它气雾剂传递载体来传递。将基因、核酸和肽组合物通过鼻气雾剂喷雾直接给予肺部的方法描述于例如美国专利 5,756,353 和美国专利 5,804,212。同样，采用鼻内微粒树脂(Takenaga 等, J Controlled Release 1998 年 3 月 2 日;52(1-2):81-7)和溶血磷脂酰甘油化合物(美国专利 5,725,871)给药也是制药领域众所周知的。同样，聚四氟乙烯支持体骨架形式的说明性的经粘膜给药在美国专利 5,780,045 中有描述。

在某些实施方案，使用脂质体、纳米囊、微粒、脂质颗粒、小泡等将本发明的组合物引入合适的宿主细胞/生物中。具体地说，本发明的组合物可以配制成用于在脂质颗粒、脂质体、小泡、纳米球或纳米粒子等中包囊传递。或者，本发明的组合物可以与这样的载体表面或者共价结合或者非共价结合。

脂质体和脂质体样制剂作为潜在药物载体的制备和应用对于本领域技术人员一般是已知的(参见例如, Lasic, Trends Biotechnol 1998 年 7 月;16(7):307-21; Takakura, Nippon Rinsho 1998 年 3 月;56(3):691-5; Chandran 等, Indian J Exp Biol. 1997 年 8 月;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; 美国专利 5,567,434; 美国专利 5,552,157; 美国专利 5,565,213; 美国专利 5,738,868 和美国专利 5,795,587; 每个文献具体地通过引用全部结合到本文中)。

脂质体已成功地与许多通常难以通过其它方法转染的细胞类型一起使用，包括 T 细胞混悬剂、原代肝细胞培养物和 PC 12 细胞(Renneisen 等, J Biol Chem. 1990 年 9 月 25 日;265(27):16337-42; Muller 等, DNA Cell Biol. 1990 年 4 月;9(3):221-9)。另外，脂质体没有基于病毒给药系统中通常的 DNA 长度限制。脂质体已有效地用来将基因、各种药物、放射治疗药、酶、病毒、转录因子、别构效应剂等导入各种各样的培养细胞系和动物中。此外，看来脂质体的应用与系统给药后自身免疫性应答或不可接受的毒性无关。

在某些实施方案中，脂质体由在水性介质中分散的磷脂形成并且自动形成多层同心双层小泡(也称为多层脂质体(MLV))。

另一方面，在其它实施方案中，本发明提供本发明组合物的药学上可接受的纳米囊制剂。纳米囊一般可以以稳定和可重现的方式包封化合物(参见，例如，Quintanar-Guerrero等，*Drug Dev Ind Pharm.* 1998年12月;24(12):1113-28)。为了避免由于胞内聚合物负荷过多引起的副作用，这样的超微颗粒(大小约0.1 μm)可以采用能够体内降解的聚合物来设计。这样的颗粒可以按下述文献所述进行制备：例如 Couvreur等，*Crit Rev Ther Drug Carrier Syst.* 1988;5(1):1-20; zur Muhlen等，*Eur J Pharm Biopharm.* 1998年3月;45(2):149-55; Zambaux等，*J Controlled Release.* 1998年1月2日;50(1-3):31-40; 和美国专利5,145,684。

本发明也涉及得自本发明多核苷酸的引物形式的多核苷酸和特异性针对本发明多肽的抗体或试剂形式的多肽作为诊断试剂的用途。

鉴定能够检测出癌发生途径中极早期变化的血液或组织中遗传标记或生化标记，将有助于确定用于患者的最佳疗法。可以使用诸如多核苷酸表达的替代肿瘤标记，诊断不同形式和病期的癌症。对本发明多核苷酸表达水平的鉴定既可用于癌性疾病的病期分类，又可用于癌性组织的性质分级。病期分类方法监测癌症的进展，并且根据在活组织检查区内是否存在恶性肿瘤组织来确定。本发明的多核苷酸可以有助于通过鉴定癌症攻击力的标记、例如在机体不同部位中的存在来完善所述病期分类方法。癌症的分级描述肿瘤与其相同类型的正常组织多么密切相似，并且通过其细胞形态和其它分化标记进行评价。本发明的多核苷酸可用于确定肿瘤级别，因为它们可能有助于确定肿瘤细胞的分化状态。

通过包括确定得自受治疗者样品中多肽或mRNA水平的异常减少或增加在内的方法进行诊断，所述诊断分析提供一种用于诊断癌症、自身免疫病和相关病症或确定对所述病症的易感性的方法。该诊断方法称为差异表达。比较患病组织和正常组织中特定基因的表达。所比

较的这两种组织中所述多核苷酸相关基因、mRNA 或蛋白之间的差异，例如在分子量、氨基酸序列或核苷酸序列或相对丰度方面的差异，表明在怀疑患病的人的组织中所述基因或调节它的基因的变化。

可以在 RNA 水平上测定表达的减少或增加。首先从这两种组织中分离出 PolyA RNA，然后由对应于差异表达的本发明多核苷酸的基因编码的 mRNA 的检测可以通过以下方法进行：例如组织切片的原位杂交、反转录酶-PCR、使用含 PolyA+ mRNA 的 RNA 印迹或任何其他直接或间接 RNA 检测方法。与正常组织相比给定 RNA 在患病组织中表达的增加或减少，提示所述转录物和/或所表达的蛋白在所述疾病中起作用。因此，检测出相对应于 SEQ ID NO:1 的 mRNA 水平比正常水平高或低，表明所述患者患有癌症。

样品中的 mRNA 表达水平可以通过从所述样品产生已表达序列标志(EST)的文库来确定。可以用 EST 在所述文库中的相对表现度，评价所述基因转录物在原始样品中的相对表现度。然后将所述测试的 EST 分析与参比样品的 EST 分析进行比较，从而测定感兴趣的多核苷酸的相对表达水平。

可以采用基因表达的连续分析(SAGE)方法(Velculescu 等, Science (1995) 270:484)、差异展示法(例如, US 5,776,683)或依赖于核苷酸相互作用专一性的杂交分析，进行其它 mRNA 分析。

另一方面，可以在蛋白质水平上进行比较。使用抗体检测来自所述两种组织中的蛋白提取物的蛋白质印迹中的多肽，可以对两种组织中的蛋白质大小进行比较。使用抗相应蛋白的抗体，还可以在免疫学上检测表达水平和亚细胞定位。可以用来测定得自宿主的样品中的蛋白质水平(例如本发明的多肽水平)的其它分析技术，对于本领域技术人员而言是熟知的。在患病组织中多肽表达水平与正常组织中相同蛋白表达水平相比升高或降低，表明所表达的蛋白可能与所述疾病有关。

在本发明的测定中，通过检测由 SEQ ID NO:1 中所示的至少一种

序列编码的基因产物的表达水平，可以确定所述诊断。也可以用患病组织与正常组织中的 mRNA 水平或蛋白质水平的比较，来跟踪疾病的发展或缓解。

采用多核苷酸阵列，可以分析样品中的许多多核苷酸序列。这些多核苷酸可以用来检测基因的差异表达，从而确定基因的功能。例如，可以用多核苷酸序列 SEQ ID NO:1 的阵列来确定所述多核苷酸的任一种是否在正常细胞和癌症细胞间有差异表达。在本发明的一个实施方案中，可以构建一个包含 SEQ ID NO:1 核苷酸序列或其片段的寡核苷酸探针的阵列，以进行例如遗传突变的有效筛选。阵列技术方法是众所周知的并且具有普遍的可应用性，可以用来解决分子遗传学方面的各种各样的问题，包括基因表达、遗传连锁和遗传变异性(参见例如：M.Chee 等, *Science*, 第 274 卷, 第 610-613 页(1996))。

本文所用的“诊断”包括确定受治疗者对疾病的易感性、确定受治疗者目前是否患有所述疾病，还可以对受所述疾病影响的受治疗者作预后。

本发明还涉及用于进行诊断分析的诊断试剂盒，所述试剂盒包括：

(a) 一种本发明的多核苷酸或其片段，所述多核苷酸最好是 SEQ ID NO:1 的核苷酸序列；

(b) 一种与(a)的核苷酸序列互补的核苷酸序列，最好是 SEQ ID NO:6 的核苷酸序列；

(c) 一种本发明的多肽或其片段，所述多肽最好是 SEQ ID NO:2 或 SEQ ID NO:3 的多肽；或

(d) 一种抗本发明多肽的抗体，最好是抗 SEQ ID NO:2 或 SEQ ID NO:3 多肽的抗体。

对于染色体定位而言，本发明的核苷酸序列也是有价值的。所述序列特异性地靶向人单个染色体上的特定位置，并且可以与其杂交。

按照本发明将有关序列作图至染色体上,在使那些序列与基因相关疾病发生联系中是重要的第一步。一旦将序列作图至染色体上的精确位置,就可以使所述序列在所述染色体上的物理位置与基因图谱数据相联系。这样的数据可以在例如 V. McKusick, Mendelian Inheritance in Man (可通过 Johns Hopkins University Welch Medical Library 在线得到)中找到。然后通过连锁分析(物理上相邻的基因的共遗传)鉴定已作图至同一染色体区的基因和疾病之间的关系。也可以确定受影响个体和未受影响个体之间的 cDNA 序列或基因组序列中的差别。

本发明的多肽或其片段或其类似物、或表达它们的细胞还可以用作免疫原,以产生抗本发明多肽的免疫专一性抗体。术语“免疫专一性”是指所述抗体对本发明多肽的亲合性比它们对现有技术中其它相关多肽的亲合性高得多。

再一方面,本发明提供如上文所限定的按照本发明多肽的免疫专一性抗体或其免疫学片段。所述抗体最好是单克隆抗体。

采用常规方法,通过将所述多肽或携带表位的片段、类似物或细胞给予动物、最好是非人类动物,可以获得针对本发明多肽产生的抗体。为了制备单克隆抗体,可以使用提供由连续细胞系培养物生产的抗体的任何技术。实例包括杂交瘤技术(Kohler, G.和 Milstein, C., Nature (1975) 256:495-497)、trioma 技术-人 B 细胞杂交瘤技术(Kozbor 等, Immunology Today (1983) 4:72)和 EBV 杂交瘤技术(Cole 等, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985)。

产生单链抗体的技术,例如在美国专利第 4,946,778 号中描述的技术,可能也适用于产生抗本发明多肽的单链抗体。另外,可以用转基因小鼠或其它生物(包括其它哺乳动物)表达人源化抗体。

可以使用上述抗体分离或鉴定表达所述多肽的克隆、或者通过亲和层析纯化所述多肽。也可以使用本发明的抗体预防或治疗癌症、尤其是结肠直肠癌、自身免疫病和相关病症。

本发明的另一方面涉及诱导或调节哺乳动物体内免疫应答的方

法,所述方法包括用足以产生抗体和/或T细胞免疫应答的本发明多肽接种所述哺乳动物,以保护或改善所述疾病的症状或进程。本发明的再一方面涉及诱导或调节哺乳动物体内免疫应答的方法,所述方法包括通过在体内指导表达所述多核苷酸并编码所述多肽的载体传递本发明的多肽,以便诱导产生抗体的这种免疫应答,从而保护所述动物免患疾病。

因此人们会认识到,本发明提供治疗与CASB7439多肽活性的存在、过量表达或表达不足有关的异常病症例如癌症和自身免疫病、尤其是结肠直肠癌的方法。本发明试图治疗的其它与CASB7439表达有关的异常病症是慢性淋巴细胞白血病和生殖细胞肿瘤。

本发明还提供筛选化合物以鉴定刺激或抑制CASB7439多肽功能的那些化合物的方法。一般而言,对于上文所提到的这类疾病,为了治疗和预防目的可以使用激动剂或拮抗剂。可以从各种来源例如细胞、无细胞制备物、化学文库以及天然产物的混合物来鉴定化合物。根据具体情况,如此鉴定的这类激动剂、拮抗剂或抑制剂可以是所述多肽的天然或经修饰的底物、配体、受体、酶等,或者也可以是它们的结构模拟物或功能模拟物(参见Coligan等, *Current Protocols in Immunology* 1(2):第5章(1991))。筛选方法是本领域技术人员已知的。其它筛选方法还可以在例如D. Bennett等, *J Mol Recognition*, 8:52-58 (1995); 和K. Johanson等, *J Mol Biol*, 270(16):9459-9471 (1995)以及其中的参考文献中找到。

因此,本发明提供鉴定刺激或抑制本发明多肽功能的化合物的筛选方法,所述方法包括一种选自以下的方法:

(a) 利用直接或间接结合候选化合物的标记,测定所述候选化合物与所述多肽(或与携带所述多肽的细胞或细胞膜)或与其融合蛋白的结合;

(b) 在标记竞争物的存在下,测定候选化合物与所述多肽(或与携带所述多肽的细胞或细胞膜)或与其融合蛋白的结合;

(c) 使用适合于携带所述多肽的细胞或细胞膜的检测系统, 测试所述候选化合物是否引起由于所述多肽的活化或抑制而产生的信号;

(d) 将候选化合物与含权利要求 1 多肽的溶液混合, 形成混合物, 然后测定所述混合物中所述多肽的活性, 并且将所述混合物的活性与标准品的活性进行比较; 或

(e) 采用例如 ELISA 测定, 检测候选化合物对在细胞中编码所述多肽的 mRNA 和所述多肽的产生的影响。

通过本领域已知的标准受体结合技术, 本发明的多肽还可用来鉴定膜结合受体或可溶性受体(如果有的话)。还可以用众所周知的筛选方法鉴定与本发明多肽竞争与其受体结合的本发明多肽的激动剂和拮抗剂(如果有的话)。

因此, 另一方面, 本发明涉及用于鉴定本发明多肽的激动剂、拮抗剂、配体、受体、底物、酶等; 或者减少或增加这类多肽产生的化合物的筛选试剂盒, 所述试剂盒包含:

- (a) 一种本发明的多肽;
- (b) 一种表达本发明多肽的重组细胞;
- (c) 一种表达本发明多肽的细胞膜; 或
- (d) 抗本发明多肽的抗体;

所述多肽最好是 SEQ ID NO:2 或 SEQ ID NO:3 的多肽。

本领域技术人员会容易地认识到, 本发明的多肽还可以用于基于结构设计所述多肽的激动剂、拮抗剂或抑制剂的方法中, 使用方法包括:

- (a) 首先确定所述多肽的三维结构;
- (b) 推导激动剂、拮抗剂或抑制剂可能的活性部位或结合部位的三维结构;
- (c) 合成预测与推导的结合部位或活性部位结合或反应的候选化合物; 和
- (d) 测试所述候选化合物是否是真正的激动剂、拮抗剂或抑制

剂。

还可以使用基因治疗实现由受治疗者体内的有关细胞内源性产生 CASB7439 多肽。有关基因治疗的综述,参见 Human Molecular Genetics, 第 20 章, 基因治疗和其它基于分子遗传学的治疗方法(Gene Therapy and other Molecular Genetic-based Therapeutic Approaches), T Strachan 和 A P Read, BIOS Scientific Publishers Ltd (1996) (和其中引用的参考文献)。

在 Pharmaceutical Biotechnology, 第 61 卷, 疫苗设计—亚单位和佐剂方法(Vaccine Design - the subunit and adjuvant approach), Powell 和 Newman 编著, Plenum Press, 1995。New Trends and Developments in Vaccines, Voller 等编著, University Park Press, Baltimore, Maryland, U.S.A. 1978 中全面描述了疫苗制剂。用脂质体包封描述于例如 Fullerton 的美国专利 4,235,877。蛋白质与大分子缀合公开于例如 Likhite 的美国专利 4,372,945 和 Armor 等的美国专利 4,474,757。

每种疫苗剂量中的蛋白质量选定为在典型的疫苗中诱导免疫保护性应答而无明显毒副作用的量。这样的量随所使用的特定免疫原而变化。一般而言,预期每个剂量包含 1-1000 μ g 蛋白、优选 2-100 μ g 蛋白、最优选 4-40 μ g 蛋白。通过包括观察受治疗者体内抗体效价和其它反应的标准研究,可以确定特定疫苗的最佳量。初次疫苗接种后,受治疗者还可以在大约 4 周内接受一次加强接种。

“分离的”是指从天然状态“人为”改变的。如果一种“分离的”组合物或物质在自然界中存在,那么它的原始环境已改变或从其原始环境中取出、或两者兼而有之。作为本文使用的此术语,例如在活的动物中天然存在的多核苷酸或多肽并不是“分离的”,但与其天然状态共存的物质分开的所述相同多核苷酸或多肽是“分离的”。

“多核苷酸”一般是指任何多核糖核苷酸或多脱氧核糖核苷酸,它们可以是包括单链区和双链区的未修饰的 RNA 或 DNA 或者经修饰的 RNA 或 DNA。

“变异体”是指与参比多核苷酸或多肽不同的但保留基本特性的多核苷酸或多肽。典型多核苷酸变异体在核苷酸序列上与另一参比多核苷酸不同。所述变异体核苷酸序列的变化可能改变或不改变参比多核苷酸编码的多肽的氨基酸序列。如下文所论述，核苷酸的改变可能导致参比序列编码的多肽的氨基酸取代、添加、缺失、融合和截短。典型的多肽变异体在氨基酸序列上与另一参比多肽不同。一般而言，差异是有限的，致使参比多肽和变异体的序列在总体上非常相似并在许多区是相同的。变异体和参比多肽可能由于任何组合的一个或多个取代、添加、缺失而氨基酸序列不同。取代或插入的氨基酸残基可以是或不是由遗传密码编码的氨基酸残基。多核苷酸或多肽的变异体可以是天然存在的变异体如等位基因变异体，或者所述变异体可以是已知非天然存在的变异体。可以通过诱变技术或通过直接合成，制备多核苷酸和多肽的非天然存在的变异体。

“同一性”正如本领域所知的，是指通过比较序列而确定的两种或两种以上多肽序列、或两种或两种以上多核苷酸序列之间的关系。在本领域中，根据具体情况，“同一性”也指根据多肽或多核苷酸序列的字符串之间的匹配而确定的多肽或多核苷酸序列之间的序列相关性程度。采用已知方法可以容易地计算出“同一性”和“相似性”，所述方法包括但不限于以下文献中描述的方法：Computational Molecular Biology, Lesk, A.M. 编著, Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. 编著, Academic Press, New York, 1993; Computer Analysis of Sequence Data, 第 I 部分, Griffin, A.M. 和 Griffin, H.G. 编著, Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; 和 Sequence Analysis Primer, Gribskov, M. 和 Devereux, J. 编著, M Stockton Press, New York, 1991; 和 Carillo, H. 和 Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988)。设计测定同一性的优选方法用以给出所测试序列之间的最大匹配。测定同一性和相似性的方法已编

纂成公众可得到的计算机程序。用于测定两个序列之间的同一性和相似性的优选计算机程序方法包括但不限于 GCG 程序包(Devereux, J.等, *Nucleic Acids Research* 12(1):387 (1984))、BLASTP、BLASTN 和 FASTA (Atschul, S.F.等, *J. Molec. Biol.* 215:403-410 (1990))。公众可从 NCBI 和其它来源得到 BLAST X 程序(BLAST Manual, Altschul, S.等, NCBI NLM NIH Bethesda, MD 20894; Atschul, S.等, *J. Mol. Biol.* 215:403-410 (1990))。也可以使用众所周知的 Smith Waterman 算法测定同一性。

所用的优选算法是 FASTA。运用该算法进行多肽或多核苷酸序列比较的优选参数包括以下参数:

Gap Penalty (空位罚分): 12

Gap extension penalty (空位拓展罚分): 4

Word size (字串大小): 2, 最大为 6

用其它方法比较多肽序列的优选参数包括以下参数:

1) 算法: Needleman 和 Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix (比较矩阵): Hentikoff 和 Hentikoff 的 BLOSSUM62, *Proc. Natl. Acad. Sci. USA.* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length penalty (空位长度罚分): 4

可使用这些参数的程序为“gap”程序,公众可从 Genetics Computer Group, Madison WI 获得。上面所提到的参数是进行多肽比较的缺省参数(同时对末端空位没有罚分)。

用于多核苷酸比较的优选参数包括以下参数:

1) 算法: Needleman 和 Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix: matches (匹配) = +10, mismatch (错配) = 0

Gap Penalty: 50

Gap Length penalty: 3

可使用这些参数的程序为“gap”程序,公众可从 Genetics Computer Group, Madison WI 获得。上面所提到的参数是进行多核苷酸比较的缺

省参数。

作为实例，本发明的多核苷酸序列可与 SEQ ID NO:1 的参比序列相同，也就是 100% 相同，或者它可以包括与所述参比序列相比，多至某一整数的核苷酸改变。这样的改变选自至少一个核苷酸缺失、取代(包括碱基转换和碱基颠换)或插入，并且其中所述改变可以发生在所述对比核苷酸序列的 5' 或 3' 末端位置或这两个末端位置间的任何地方，或者各自单独散置于参比序列的核苷酸中，或者散置于参比序列的一个或多个连续组中。核苷酸改变的数目如下确定：SEQ ID NO:1 核苷酸总数乘以相应的同一性百分率的数值百分率(除以 100)，随后从所述 SEQ ID NO:1 核苷酸总数减去该乘积，或

$$n_n < x_n - (x_n \cdot y),$$

其中 n_n 是核苷酸改变的数目， x_n 是 SEQ ID NO:1 核苷酸总数， y 例如是 0.70 (70%)、0.80 (80%)、0.85 (85%)、0.90 (90%)、0.95 (95%) 等，并且其中将 x_n 和 y 的任何非整数乘积从 x_n 减去之前，将该乘积舍入到最近的整数。编码 SEQ ID NO:2 多肽的多核苷酸序列的改变可能在该编码序列中产生无义、错义或移码突变，从而在这样的改变后改变由所述多核苷酸编码的多肽。

同样，本发明的多肽序列可与 SEQ ID NO:2 的参比序列相同，也就是 100% 相同，或者它可以包括与所述参比序列相比，多至某一整数的氨基酸改变，以致同一性百分率小于 100%。这样的改变选自至少一个氨基酸缺失、取代(包括保守取代和非保守取代)或插入，并且其中所述改变可以发生在所述参比多肽序列的氨基末端或羧基末端位置或这两个末端位置间的任何地方，或者各自单独散置于参比序列的氨基酸中，或者散置于参比序列的一个或多个连续组中。对于特定的同一性百分率如下确定所述氨基酸改变的数目：SEQ ID NO:2 氨基酸总数乘以相应的同一性百分率的数值百分率(除以 100)，随后从所述 SEQ ID NO:2 氨基酸总数减去该乘积，或

$$n_a < x_a - (x_a \cdot y),$$

其中 n_a 是氨基酸改变的数目, x_a 是 SEQ ID NO:2 氨基酸总数, y 例如是 0.70 (70%)、0.80 (80%)、0.85 (85%) 等, 并且其中将 x_a 和 y 的任何非整数乘积从 x_a 减去之前, 将该乘积舍入到最近的整数。

“同系物”是本领域所用的一个类别的术语, 是指具有与主题序列高度序列相关性的多核苷酸序列或多肽序列。这样的相关性可以通过确定如上所述的所比较序列之间的同一性程度和/或相似性程度而定量。属于该类别术语范围内的是术语“直向同源物”和“共生同源物”, “直向同源物”是指作为一种多核苷酸或多肽在另一物种中的功能等同物的多核苷酸或多肽, 而“共生同源物”是指在同一物种内被认为功能上相似的序列。

附图说明

图 1: 显示采用 Taqman 探针的实时 PCR 数据。所述说明如下: 肾上腺: Ad_G1; 膀胱: B1; 骨髓: Bo_Ma; 宫颈: Ce; 结肠: Co; 输卵管: Fa_Tu; 回肠: Il; 肝: Li; 肺: Lu; 淋巴结: Ly_No; 食管: Oe; 甲状旁腺: Pa_Thy; 胎盘: P1; 前列腺: Pr; 直肠: Re; 皮肤: Sk; 骨骼肌: Sk_Mu; 小肠: Sm_In; 脾: Sp; 睾丸: Te; 甲状腺: Thy; 气管: Tr。

图 2 显示采用 Sybr 方案的实时 PCR 表达。所述说明如下: 肾上腺: Ad_G1; 膀胱: B1; 骨髓: Bo_Ma; 宫颈: Ce; 结肠: Co; 淋巴结: Ly_No; 食管: Oe; 甲状旁腺: Pa_Thy; 胎盘: P1; 前列腺: Pr; 直肠: Re; 皮肤: Sk; 骨骼肌: Sk_Mu; 小肠: Sm_In; 脾: Sp; 睾丸: Te; 甲状腺: Thy; 气管: Tr; 心脏: He。

图 3 显示得自表达 CASB7439 的菌株的细胞提取物的考马斯蓝染色的 SDS PAGE。泳道 1 显示分子量标准参照物, 泳道 2 显示于 39°C 诱导 5 小时的细胞提取物; 泳道 3 显示所诱导的细胞提取物的上清液; 泳道 4 显示所诱导的细胞提取物的沉淀。

图 4 显示 NS1-CASB7439 表达蛋白的蛋白质印迹分析。将表达

CASB7439 的菌株的细胞提取物加样至该凝胶，并且用抗 NS1 单克隆抗体揭示。

图 5 显示 CASB7439 在纯化后的考马斯蓝染色的 SDS-PAGE。泳道 1 和泳道 5 代表分子量标准参照物；泳道 2、3、4 分别加入 2 μ l、4 μ l 和 6 μ l 纯化蛋白。

图 6 显示根据抗多聚组氨酸单克隆抗体揭示的 CASB7439 在纯化后的蛋白质印迹。

实施例

实施例 1

实时 RT-PCR 分析

使用实时 RT-PCR (U. Gibson. 1996. Genome Research: 6,996) 比较来自多个患者的配对的肿瘤结肠组织和正常结肠组织中候选抗原的 mRNA 转录物的丰度。另外，通过该方法评价一组正常组织中所述候选基因的 mRNA 水平。

使用 TriPure 试剂(Boehringer)，从速冻活检组织中提取正常结肠和肿瘤结肠的总 RNA。正常组织的总 RNA 购自 InVitrogen 或者使用 TriPure 试剂从速冻活检组织中提取。使用寡脱氧胸苷酸磁珠(Dynal)，从经 DNA 酶处理后的总 RNA 中纯化 Poly-A⁺ mRNA。使用 SybrII 染料(Molecular Probes)，通过分光荧光测定法(VersaFluor, BioRad)进行 mRNA 的定量。采用 TaqMan 扩增条件的缺省选项，用 Perkin-Elmer Primer Express 软件设计用于实时 PCR 扩增的引物。

每次反应使用 2 ng 纯化 mRNA，按照标准 PCR 方法进行实时反应。为了实时检测，以终稀释度 1/75000 加入 SybrI 染料(Molecular Probes)。采用常规仪器设定值，在 Perkin-Elmer Biosystems PE7700 系统中进行扩增(40 个循环)和实时检测。使用 PE7700 Sequence Detector 软件计算 Ct 值。对于每个样品得到若干个 Ct 值：对于患者样品，候选 TAA 上的肿瘤 Ct (CtT)值和配对的正常结肠 Ct (CtN)值，而对于该

组正常组织样品,每个正常组织 XY 的 Ct_{XY}。对于所有样品也计算肌动蛋白基因的另一 Ct (Ct_A),作为所有样品的内部参照物。另一方面,可以采用 Taqman 探针监测实时 PCR 扩增。采用常规仪器设定值,在 Perkin-Elmer Biosystems PE7700 系统中进行扩增(40 个循环)和实时检测。使用 PE7700 Sequence Detector 软件计算 Ct 值。从每个组织样品获得靶 mRNA (Ct_X)和肌动蛋白 mRNA (Ct_A)的 Ct 值。

因为在普通实验条件下的 PCR 扩增效率接近于理论扩增效率,所以 $2^{(Ct_N/T/XY - Ct_A)}$ 值是所述样品的相对 TAA 转录水平的估计值,即相对于肌动蛋白转录水平进行标准化的值。因此,数值为 1 表示候选抗原和肌动蛋白的表达水平相同。

首先对来自 12 个患者的活检组织的肿瘤结肠和配对的正常结肠进行实时 PCR 反应。然后在总共 18 个患者的更全面的数据集上进行反应(前 12 个患者的数据集包括在内)。在该数据集中这 18 个患者中的 6 个患者的数据以复份制成。还测试了其它 6 个患者,结果与前 18 个患者的数据合并。对最终合并值进行的统计学分析结果示于表 3 中,并且在图 1 中加以说明。

也依照相同方法测试代表 29 种不同组织的一系列 48 个正常组织样品(所分析的正常组织示于表 3 中)。如上所述计算 TAA 转录水平。也根据该数据集计算出过量表达候选抗原的患者的比例以及相对于正常组织的转录物过量表达平均值。结果示于图 1 中。

表 1: CASB7439 实时 PCR 表达结果: 12 个患者的数据集

mRNA 水平在配对的肿瘤结肠中较高的患者百分率(阳性患者)	92%
mRNA 水平在配对的肿瘤结肠中至少高 3 倍的患者百分率	92%
mRNA 水平在配对的肿瘤结肠中至少高 10 倍的患者百分率	92%
mRNA 水平在配对的肿瘤结肠中至少低 3 倍的患者百分率	8%
配对的正常结肠 mRNA 水平平均值(用肌动蛋白标准化)	0.0026
在阳性患者中配对的正常结肠 mRNA 水平平均值(用肌动蛋白标准化)	0.265
mRNA 过量表达倍数平均值	2028
mRNA 过量表达倍数中值	115
正常组织 mRNA 水平平均值	0.0079
正常组织 mRNA 水平中值	0.0016
正常组织 mRNA 水平平均值	0.0064
正常组织 mRNA 水平中值	0.0017
mRNA 水平比正常组织平均值高的患者百分率	92%
mRNA 水平比正常组织平均值高 10 倍的患者百分率	75%
比正常组织的 mRNA 水平中值高的必需(non-dispensable)正常组织	无

表 2: CASB7439 实时 PCR 表达结果: 18 个患者的数据集

mRNA 水平在配对的肿瘤结肠中较高的患者百分率(阳性患者)	89%
mRNA 水平在配对的肿瘤结肠中至少高 3 倍的患者百分率	89%
mRNA 水平在配对的肿瘤结肠中至少高 10 倍的患者百分率	78%
mRNA 水平在配对的肿瘤结肠中至少低 3 倍的患者百分率	5%
配对的正常结肠 mRNA 水平平均值(用肌动蛋白标准化)	0.005
在阳性患者中配对的正常结肠 mRNA 水平平均值(用肌动蛋白标准化)	0.152
mRNA 过量表达倍数平均值	1100
mRNA 过量表达倍数中值	60
正常组织 mRNA 水平平均值	0.0065
正常组织 mRNA 水平中值	0.0015
正常组织 mRNA 水平平均值	0.005
正常组织 mRNA 水平中值	0.0015
mRNA 水平比正常组织平均值高的患者百分率	94%
mRNA 水平比正常组织平均值高 10 倍的患者百分率	94%
比正常组织的 mRNA 水平中值高的必需正常组织	无

表 3: CASB7439 实时 PCR 表达结果: 24 个患者的数据集

CASB7439 转录物水平在肿瘤结肠中比在相邻的正常结肠中高的患者百分率(阳性患者)	92%
CASB7439 转录物水平在肿瘤结肠中比在相邻的正常结肠中至少高 10 倍的阳性患者百分率	75%
在阳性患者的肿瘤结肠中转录物过量表达倍数平均值	1289
CASB7439 转录物水平在肿瘤结肠中比正常组织平均值高的患者百分率	96%
mRNA 水平在肿瘤结肠中比正常组织平均值至少高 10 倍的患者百分率	62.5%
其中 CASB7439 转录物表达等于肿瘤中的肿瘤转录物水平的正常组织	无

也采用 Taqman 方法(如上所述),对来自 6 个患者活检组织的肿瘤结肠和相邻的正常结肠进行实时 PCR 反应。对于每个样品采用三个复份测定,用平均值进行进一步的计算。结果示于图 1 中。此外,也依照相同方法测试代表 28 种不同组织的 36 个正常组织样品(参见表 5)。结果示于图 2 中。

表 4: 采用 Taqman 探针的 CASB7439 实时 PCR 表达结果

来自不同患者的肿瘤样品数	6
CASB7439 转录物水平在肿瘤结肠中比相邻的正常结肠高的患者百分率(阳性患者)	100%
CASB7439 转录物水平在肿瘤结肠中比相邻的正常结肠至少高 10 倍的阳性患者百分率	83%
在阳性患者的肿瘤中转录物过量表达倍数平均值	109
CASB7439 转录物水平在肿瘤结肠中比正常组织平均值高的患者百分率	100%
mRNA 水平在肿瘤结肠中比正常组织平均值至少高 10 倍的患者百分率	100%
其中 CASB7439 转录物表达等于肿瘤中的肿瘤转录物水平的正常组织	无

结果清楚地表明, 与相邻的正常结肠和所有的上述正常组织相比, CASB7439 转录物在结肠直肠肿瘤中过量表达。与相邻的正常结肠相比, 90%以上的所述患者的肿瘤中 CASB7439 转录物强过量表达。所述肿瘤中的平均过量表达倍数至少为 100。此外, 与其它正常组织相比, 90%以上的所述患者的结肠肿瘤中 CASB7439 转录物过量表达, 其中过量表达 CASB7439 转录物的所述患者中的 60%以上的过量表达倍数为至少 10 倍。

表 5: 用于 CASB7439 转录物表达分析的正常组织一览表

组织	缩写
肾上腺	Ad_Gl
主动脉	Ao
膀胱	Bl
骨髓	Bo_Ma
脑	Bra
宫颈	Ce
结肠	Co
输卵管	Fa_Tu
心脏	He
回肠	Il
肾	Ki
肝	Li
肺	Lu
淋巴结	Ly_No
食管	Oe
甲状旁腺	Pa_Thy
直肠	Re
皮肤	Sk
骨骼肌	Sk_Mu
小肠	Sm_In
脾	Sp
胃	St
甲状腺	Thy
气管	Tra
卵巢	Ov
胎盘	Pl
前列腺	Pr
睾丸	Te

实施例 2

cDNA 阵列的示差筛选

通过示差筛选完成对扣除 cDNA 文库中的肿瘤相关基因的鉴定。

从 100 μ l 过夜培养物中提取细菌总 DNA。用异硫氰酸胍裂解细菌，使用磁性玻璃(Boehringer)对细菌 DNA 进行亲和纯化。通过 Advantage PCR 扩增(Clontech)从细菌 DNA 回收质粒插入片段。采用 Biomek 96 HDRT 工具(Beekman)，将 PCR 产物点样至两个尼龙膜上，产生高密度 cDNA 阵列。斑点状 cDNA 经 UV 照射与膜共价结合。将第一张膜与从单个患者的肿瘤制备的混合 cDNA 探针杂交。将第二张膜与等量的从相同患者的正常结肠制备的混合 cDNA 探针杂交。如上所述通过 PCR 扩增制备探针 cDNA，并且采用 AlkPhos Direct System (Amersham)将其标记。杂交条件和严格洗涤如 AlkPhos Direct 试剂盒所述。杂交探针用化学发光进行检测。通过薄膜光密度测定法或直接测定法(BioRad Fluor-S Max)测定两个印迹上的每种 cDNA 片段的杂交强度。计算每个基因的肿瘤杂交强度与正常杂交强度的比率(T/N)，以评价肿瘤中的过量表达程度。对在结肠肿瘤中显著过量表达的基因进行跟踪研究。显著性任意地定义为 T/N 频率分布的一个标准偏差。采用来自多个患者供体(>18)的 RNA，重复示差筛选实验，以估计在该患者群体中过量表达的肿瘤的频率。另外，将 DNA 阵列与来自除结肠以外的正常组织(参见上文一览表)的混合 cDNA 探针杂交，以测定所述候选基因在这些组织中的表达水平。

实施例 3

DNA 微阵列

用 DNA 微阵列检测多个样品中基因大集合物的 mRNA 表达分布型。用该信息补充通过实时 PCR 获得的数据，并且提供在肿瘤组织和正常组织中基因表达水平的独立测量。

目前用于产生 DNA 微阵列的技术的实例包括 1) Affymetrix

“GeneChip”阵列，其中采用光刻法通过固相化学合成在所述芯片的表面合成寡核苷酸，2) DNA 点样(spotting)技术，其中由机器人将小体积 DNA 溶液沉积到固相(例如玻璃)的表面，然后将其固定化。在这两种情况下，将所述芯片与从目标组织(例如正常组织、肿瘤等……)提取并且用放射性或者用荧光报道分子标记的 cDNA 或 cRNA 杂交。使标记材料与所述芯片杂交，然后使用特殊扫描仪，测定与所述芯片上每一序列结合的探针量。所述实验可以用单一荧光报道分子(或放射性)建立，或者可以用两种荧光报道分子进行。在后一种情况下，所述两个样品中的每个用其中一种报道分子标记。然后将这两个已标记的样品与所述 DNA 芯片上的序列竞争性杂交。确定所述芯片上每一序列的两种荧光信号的比例。用该比例计算所述两个样品中所述转录物的相对丰度。详细方案可得自许多来源，包括“DNA 微阵列：一种实用方法(DNA Microarrays: A practical approach)。Schena M. Oxford University Press 1999 ” 和 万 维 网 (<http://cmgm.stanford.edu/pbrown/protocols/index.html>), <http://arrayit.com/DNA-Microarray-Protocols/>)和许多专业销售商(例如 Affymetrix)。

实施例 5

RNA-DNA 印迹分析

通过 Advantage PCR (参见上文)扩增有限量的混合的肿瘤结肠 cDNA 和配对的正常结肠 cDNA 。也使用相同方法扩增来自多个正常组织的信使 RNA。将所扩增的 cDNA (1 μ g)在 1.2%琼脂糖凝胶上电泳，然后转移至尼龙膜上。将所述膜与用候选 TAA cDNA 片段制备的探针杂交(AlkPhos Direct System)。RNA-DNA 印迹分析提供有关肿瘤组织和正常组织中转录物大小、是否存在剪接变异体以及转录物丰度的信息。

实施例 6

RNA 印迹分析

按照标准方法使用 1 μg poly A+ mRNA 产生 RNA 印迹。采用 Ready-to-Go 系统(Pharmacia)制备放射性探针。

实施例 7

全长 cDNA 序列的实验鉴定

采用 Lambda Zap II 系统(Stratagene)从 5 μg poly A+ mRNA 构建结肠肿瘤 cDNA 文库。采用所提供的方案，只是使用 SuperscriptII (Life Technologies)进行反转录步骤。构建寡聚 dT 引物文库和随机引物文库。对于文库的每次筛选，平板接种约 1.5×10^6 个独立噬菌体。将噬斑转移至尼龙滤膜上，并且采用 AlkPhos Direct 标记的 cDNA 探针杂交。通过化学发光检测阳性噬菌体。将阳性噬菌体从琼脂板上切下，在 500 μl SM 缓冲液中洗脱，通过基因特异性 PCR 进行证实。通过体内移除将洗脱的噬菌体转变成单链 M13 噬菌体。然后通过大肠杆菌的感染，将所述噬菌体转变成双链质粒 DNA。将受感染细菌平板接种，然后用所述 cDNA 探针进行第二轮筛选。从阳性细菌克隆纯化出质粒 DNA，然后对两条链测序。

当全长基因不能直接从 cDNA 文库中获得时，采用 RACE 技术(Marathon Kit, ClonTech)分离丢失序列。该方法依赖于将 mRNA 反转录成双链 cDNA，将接头与 cDNA 的两个末端连接，采用基因特异性引物和接头寡核苷酸的特异性引物扩增 cDNA 的所需端。将 Marathon PCR 产物克隆到质粒(pCRII-TOPO, InVitrogen)中，然后测序。

使用该方法获得 SEQ ID NO:1 的多核苷酸。

实施例 8.

EST 分布型

实验抗原组织表达表征的一种互补方法是搜索人 EST 数据库。

EST (“已表达序列标志”)是从特定组织或细胞系提取的 mRNA 集合物制备的 cDNA 的小片段。目前这样的数据库提供来自数千种 cDNA 组织文库(包括来自各种类型和病期的疾病的肿瘤组织)的大量的人 EST (2×10^6)。借助于信息工具(Blast), 进行 CASB7439 序列的比较搜索, 以便更加深入地了解组织表达。

CASB7439 的 EST 分布

EST GenBank 登记号	EST cDNA 组织文库
C00634	成人 (K. Okubo)
AA468668	NCI_CGAP_Co3
AA565752	NCI_CGAP_Co11
AA565766	NCI_CGAP_Co11
AA565767	NCI_CGAP_Co11
AI337239	NCI_CGAP_Co16
AI337448	NCI_CGAP_Co16
AI393930	NCI_CGAP_CLL1
AI473673	NCI_CGAP_Co14
AI632444	NCI_CGAP_GC6
AI861937	NCI_CGAP_Co16
AI825214	NCI_CGAP_GC6
AW080652	NCI_CGAP_Co19
AW083899	NCI_CGAP_Co19
AW206058	NCI_CGAP_Sub3
AW237006	NCI_CGAP_GC6
AW364626	DT0036
AW449612	NCI_CGAP_Sub5

这些 EST 与 CASB7439 完美匹配。该表包括 9 个来自 4 个不同肿

瘤结肠文库的 EST、1 个来自 1 个正常结肠文库的 EST、3 个来自 1 个肿瘤生殖细胞文库的 EST、1 个来自 1 个慢性淋巴细胞白血病细胞文库的 EST、2 个来自 2 个混合肿瘤文库的 EST、2 个来自未知类型文库的 EST。这清楚地表明，同预期的一样，与正常组织相比，CASB7439 在肿瘤组织中过量表达，尤其是在结肠直肠肿瘤组织中过量表达。

实施例 9:

9.1 肿瘤特异性抗原的表达和纯化

采用在微生物宿主中表达或者在体外转录/翻译，产生用于疫苗目的本发明抗原以及产生蛋白片段或完整蛋白，用于快速纯化以及产生通过免疫组织化学法鉴定天然已表达蛋白或者纯化跟踪所需的抗体。

可以在两种微生物宿主 - 大肠杆菌和酵母(例如酿酒酵母(*Saccharomyces cerevisiar*)或巴斯德毕赤酵母(*Pichia pastoris*))中表达重组蛋白。这允许选择具有最佳特征的表达系统，以供生产该特定抗原。一般而言，让重组抗原在大肠杆菌中表达，而让试剂蛋白在酵母中表达。

表达策略首先包括设计重组抗原的一级结构。一般将表达融合配偶体(EFP)置于 N 末端，以提高表达水平，所述表达融合配偶体也可以包括一个可用于调节所述抗原免疫原性特性的区域 - 免疫融合配偶体(IFP)。另外，在 C-末端包括可用于有助于进一步纯化的亲和融合配偶体(AFP)。

如上所述，可以对若干个构建体进行比较评价：

对于快速表达和纯化以及产生针对 CASB7439 的抗体，建议在大肠杆菌中产生具有 NS1 作为 EFP 和组氨酸尾作为 AFP 的全长 CASB7439 蛋白。

因此，提出两种构建体：

构建体 1: 与 NS1 cDNA (作为 EFP)以及组氨酸尾编码 cDNA (作为 AFP)融合的全长野生型 CASB7439 cDNA (SEQ ID NO:8)。所编码的

融合蛋白序列为 SEQ ID NO:10。

构建体 2: 与 NS1 cDNA (作为 EFP) 以及组氨酸尾编码 cDNA (作为 AFP) 融合的全长突变型 CASB7439 cDNA (SEQ ID NO:9)。建议在该构建体中具有被大肠杆菌密码子选择特异性的密码子取代天然 CASB7439 cDNA 的前 50 个密码子, 以增强 CASB7439 在其大肠杆菌宿主中的表达潜力。所编码的融合蛋白序列为 SEQ ID NO:10。

所述 CASB7439 蛋白设计如下所示:



“NS1”是流感病毒蛋白 NS1 的 N-端片段(80 个氨基酸)。“HIS”是多聚组氨酸尾。

所用的重组菌株是 AR58: 衍生于 N99 和 cI857 的隐蔽性 λ 溶原体, N99 是 *gal* E::Tn 10, Δ -8(*chlD-pgl*), Δ -H1(*cro-chlA*),N⁺ (Proc.Natl.Acad.Sci.USA 第 82 卷, 第 88-92 页, 1985 年 1 月 Biochemistry).

当重组菌株可得到时, 通过评价表达水平并且通过分析粗提物的行为进一步预测所述蛋白的溶解度来表征重组产物。

在合适培养基上生长并且诱导重组蛋白表达后, 通过 SDS-PAGE 分析总提取物。使重组蛋白在染色凝胶上显现, 然且使用特异性抗体通过蛋白质印迹分析进行鉴定。

质粒:

名称: TCM 281 pRIT..15143

复制子: pMB1

选择: Kan

启动子: PL long

插入片段: NS1-C74-39-His

用构建体 1 表达重组蛋白:

让细菌在 LB 培养基+ 50 $\mu\text{g/ml}$ Kan 中于 30 $^{\circ}\text{C}$ 生长

当培养物达到 OD = 0.5 (620nm)时, 将培养物加热到至多 39 $^{\circ}\text{C}$,
诱导 5 小时后, 收获细胞

提取物的制备:

细胞浓度: .50X..在缓冲液 PBS + 完全... 中

破碎: 弗氏压碎器 3X

离心: 于 14000t 30 分钟

备注: 在细胞提取物的上清液中 > 90%

将细胞提取物在 12.5% SDS PAGE 上电泳, 随后用考马斯蓝染色。也使用市售的针对多聚组氨酸尾的单克隆抗体(Quiagen)进行蛋白质印迹分析。所得凝胶(图 3 和图 4)显示该蛋白被表达, 并且在细胞提取物上清液中显现。

纯化方案采用基于重组蛋白中组氨酸亲和性尾存在的经典方法。在一个典型实验中, 过滤破碎的细菌, 然后将无细胞提取物加样到特异性地保留所述重组蛋白的离子金属亲和层析柱(IMAC; Ni^{++} NTA, 得自 Qiagen)上。在磷酸盐缓冲液中, 用 0-500 mM 咪唑梯度(可能在去垢剂存在下)洗脱所述保留的蛋白。

将得自收获培养物的上清液在 6M 尿素、100mM NaH_2PO_4 、10mM Tris、pH 8 中变性, 然后在以下条件下上样到层析柱 IMAC Qiagen NTA Ni^{++} :

平衡缓冲液:	NaH_2PO_4	100mM	pH 8
	Tris	10mM	
	尿素	6M	

样品: 尿素 6M、100mM NaH_2PO_4 、10mM Tris 中的上清液

洗涤缓冲液: 1)NaH₂PO₄ 100mM pH 8
 Tris 10mM
 尿素 6M
 咪唑 25mM

2)NaH₂PO₄ 100mM pH 8
 Tris 10mM
 尿素 6mM
 咪唑 50mM

洗脱缓冲液: NaH₂PO₄ 100mM pH 5.5
 Tris 10mM
 尿素 6M
 咪唑 500mM

在 500mM 咪唑+ 6M 尿素中的洗脱蛋白在以下条件下透析:

- PBS PH 7.2 + sarkosyl 0.5% + 4M 尿素
- 同前, 于 2M 尿素 2 小时
- 同前, 于 0M 尿素 2 小时

将最终的物质冷冻保存。蛋白质含量采用 Lowry 蛋白测定来定量 (0.9 mg/1.2 ml)。通过用考马斯蓝染色的 12.5% PAGE SDS 评价纯度(图 5), 通过蛋白质印迹、采用抗多聚组氨酸单克隆抗体检查重组蛋白的存在(图 6)。

对不同形式的所表达抗原的比较评价, 将使得能够选择出最有前途的候选物, 然后将其用于进一步纯化和免疫学评价。

9.2 抗体的产生和免疫组织化学

可以用少量相对纯化的蛋白产生免疫学工具, 以便

- a) 在正常或癌症组织切片中通过免疫组织化学检测所述表达;
- b) 检测所述表达, 并且在纯化过程期间跟踪所述蛋白(ELISA/蛋白质印迹); 或
- c) 鉴定/定量所述纯化蛋白(ELISA)。

9.2.1 多克隆抗体:

免疫

兔子用 100 μ g 配制在佐剂 3D-MPL/QS21 中的蛋白以 3 周间隔肌肉内(I.M.)免疫 3 次。每次免疫后 3 周, 抽取血样, 然后使用所述蛋白作为包被抗原, 根据标准方法, 通过 ELISA, 估计所述血清中的抗体效价。

ELISA

将 96 孔微量培养板(maxisorb Nunc)用 5 μ g 蛋白于 4 $^{\circ}$ C 包被过夜。用 1% PBS NCS 1%于 37 $^{\circ}$ C 饱和 1 小时后, 于 37 $^{\circ}$ C 加入连续稀释的兔血清(起始于 1/10)达 1 小时 30 分钟。用 PBS 吐温洗涤 3 次后, 加入抗兔生物素化抗血清(Amersham) (1/5000)。洗涤培养板, 然后于 37 $^{\circ}$ C 加入过氧化物酶偶联链霉抗生物素蛋白(1/5000)达 30 分钟。洗涤后, 加入 50 μ l TMB (BioRad)达 7 分钟, 然后用 0.2 M H₂SO₄ 终止反应。可以在 450 nm 测定 OD, 并且用 SoftmaxPro 计算中点稀释度。

9.2.2 单克隆抗体

免疫

用 5 μ g 纯化蛋白对 5 只 BALB/c 小鼠以 3 周间隔免疫 3 次。第 2 次免疫后 14 天以及第 3 次免疫后 1 周进行放血。用纯化蛋白作包被抗原, 通过 Elisa 测试所述血清。根据这些结果(中点稀释度 > 10000), 选择 1 只小鼠用于融合。

融合/HAT 选择

按照标准方法，使用 40% PEG 和 5% DMSO，将脾细胞与 SP2/0 骨髓瘤细胞融合。然后以 2.5×10^4 - 10^5 细胞/孔将细胞接种于 96 孔板，然后在 HAT 培养基中选择抗性克隆。测试这些杂交瘤上清液中特异性抗体的含量，当杂交瘤上清液为阳性时，将其进行 2 个循环的有限稀释。经过 2 轮筛选后，选择 3 个杂交瘤用于腹水生产。

9.2.3 免疫组织化学

当抗体可得到时，对正常组织切片或癌组织切片进行免疫染色，以便确定：

- ◇ 相对于正常组织而言癌组织中本发明抗原的表达水平或
- ◇ 表达所述抗原的某一类型癌的比例
- ◇ 是否其它癌类型也表达所述抗原
- ◇ 在一种癌组织中表达所述抗原的细胞比例

组织样品制备

解剖后，将组织样品用 OCT 化合物固定在软木盘上，然后在先前已在液氮(-160℃)中过冷的异戊烷中快速冷冻。使用之前一直将所述冷冻块在-70℃保存。7-10μm 切片在恒冷切片机室(-20, -30℃)中完成。

染色

将组织切片在室温(RT)下干燥 5 分钟，室温下在丙酮中固定 10 分钟，再次干燥，然后用 PBS 0.5% BSA 5% 血清饱和。于室温下 30 分钟后，用抗原特异性抗体进行直接或间接染色。直接染色产生特异性更好但强度较低的染色，而间接染色产生强度较高但特异性较低的染色。

9.3 针对本发明抗原的人细胞免疫应答的分析

通过在体外使人 T 细胞接触抗原, 可以评价本发明抗原的免疫相关性。所有 T 细胞淋巴细胞系和树突细胞均来源于健康供体(优选为 HLA-A2 亚型)的 PBMC (外周血单核细胞)。HLA-A2.1/Kb 转基因小鼠模型也用来筛选 HLA-A2.1 肽。

通过每周体外刺激, 产生并维持新发现的抗原特异性 CD8⁺ T 细胞系。采用标准方法, 测试所述 CD8⁺ 系应答所述抗原或抗原衍生肽的裂解活性和 γ -IFN 的产生。

使用两种策略来产生所述 CD8⁺ T 细胞系: 一种基于肽的方法和一种基于完整基因的方法。两种方法均要求将新发现抗原正确读框的全长 cDNA 或者在合适传递系统中克隆或者用来预测 HLA 结合肽的序列。

基于肽的方法

概述地说, 用含佐剂的 HLA-A2 肽免疫转基因小鼠, 不能诱导 CD8⁺ 应答(根据肽脉冲处理的自身脾细胞有效裂解来确定)的那些肽在人系统中进一步分析。

人树突细胞(按照 Romani 等所述方法培养)用肽进行脉冲处理, 并且用来刺激 CD8⁺ 分类的 T 细胞(通过 Facs)。通过数次每周刺激后, 首先在肽脉冲处理的自身 BLCL (EBV-B 转化细胞系)上测试所述 CD8⁺ 系。为了确证在体内正确加工所述肽, 在 cDNA 转染的肿瘤细胞 (HLA-A2 转染的 LnCap、Skov3 或 CAMA 肿瘤细胞)上测试所述 CD8⁺ 系。

基于完整基因的方法

使 CD8⁺ T 细胞系接触抗原, 然后用基因枪转染的树突细胞、反转录病毒转导的 B7.1 转染的成纤维细胞、重组痘病毒或腺病毒感染的树突细胞刺激。病毒感染的细胞非常有效地呈递抗原肽, 因为所述抗原

以高水平表达，而且可以只使用一次，以避免病毒 T 细胞系的过度生长。

改变刺激后，如上所述，在 cDNA 转染的肿瘤细胞上测试所述 CD8⁺ 系。测定肽特异性和同一性以证实免疫学检验。

CD4⁺ T 细胞应答

同样，也可以评价 CD4⁺ T 细胞免疫应答。采用加载用以刺激 T 细胞的重组纯化蛋白或肽的树突细胞产生特异性 CD4⁺ T 细胞。

预测结合 HLA 等位基因的表位(九聚体和十聚体):

根据 Parker 算法(Parker, K.C., M. A. Bednarek 和 J. E. Coligan. 1994. 根据个别肽单链的独立结合将潜在的 HLA-A2 结合肽分等级的方案 J. Immunol. 152:163 和 http://bimas.dcrn.nih.gov/molbio/hla_bind) 或 Rammensee 方法(Rammensee, Friede, Stevanovic, MHC 配体和肽基序: 第一列表, Immunogenetics 41, 178-228, 1995; Rammensee, Bachmann, Stevanovic: MHC 配体和肽基序. Landes Bioscience 1997 和 <http://134.2.96.221/scripts/hlaserver.dll/home.htm>), 预测 HLA I 类结合肽序列。然后在 HLA-A2.1/Kb 转基因小鼠模型(Vitiello 等)中筛选肽。

运用 Tepitope 算法，用截止计分设置为 6 (Sturniolo, Hammer 等, Nature Biotechnology. 1999. 17;555-561), 预测 HLA II 类结合肽序列。

下表搜集了 I 类和 II 类预测的表位序列:

HLA-A 0201 : 十聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	64	KLVN LGFQAL	142.060	SEQ ID NO:16

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-A 0201 : 九聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	182	ELLDFSSWL	507.976	SEQ ID NO:17
2	104	RLLAEHDAV	126.098	SEQ ID NO:18
3	64	KLVN LGFQA	100.850	SEQ ID NO:19

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-A 24 : 九聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	97	EYIRALQRL	360.000	SEQ ID NO:20

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-A 24 : 十聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	97	EYIRALQRL L	360.000	SEQ ID NO:21

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-B 7 : 十聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	111	AVRNALAGGL	600.000	SEQ ID NO:22

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-B 4403 : 十聚体				
等级	起始位置	亚序列残基列表	Parker 计分 ^o	SEQ ID :
1	156	SEPGSPRSAY	360.000	SEQ ID NO:23
2	89	VETLRSAVEY	180.000	SEQ ID NO:24

^o: 含有所述亚序列的分子解离半衰期的估计。

HLA-DRB1*1501 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	5.6	SEQ ID NO:25

HLA-DRB1*1502 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	4.6	SEQ ID NO:25

HLA-DRB1*0402 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	120	LRPQAVRPS	5.4	SEQ ID NO:26

HLA-DRB1*1101 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	4.8	SEQ ID NO:25

HLA-DRB1*1102 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	120	LRPQAVRPS	6.2	SEQ ID NO:26

HLA-DRB1*1104 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	5.8	SEQ ID NO:25

HLA-DRB1*1106 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	5.8	SEQ ID NO:25

HLA-DRB1*1301 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	120	LRPQAVRPS	6.6	SEQ ID NO:26
2	73	LRQHVPHGG	4.9	SEQ ID NO:27
3	31	LLRCSRERR	4.4	SEQ ID NO:33

HLA-DRB1*1302 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	120	LRPQAVRPS	5.6	SEQ ID NO:26

HLA-DRB1*1304 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	120	LRPQAVRPS	6.2	SEQ ID NO:26
2	73	LRQHVPHGG	4.8	SEQ ID NO:27
3	31	LGFQALRQH	4.6	SEQ ID NO:28

HLA-DRB1*1305 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	99	IRALQRLLA	4.8	SEQ ID NO:25

HLA-DRB1*0703 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	112	VRNALAGGL	5.1	SEQ ID NO:29
2	98	YRALQRL	4.8	SEQ ID NO:30
3	65	LVNLGFQAL	4.5	SEQ ID NO:31

HLA-DRB5*0101 : 九聚体				
等级	起始位置	亚序列残基列表	Tepitope 计分	SEQ ID :
1	96	VEYRALQR	4.3	SEQ ID NO:32

序列信息

SEQ ID NO:1

GTACCTTGCTTTGGGGCGCACTAAGTACCTGCCGGGAGCAGGGGGCGCACCGGGAACCTCGCAGATTCGCC
 AGTTGGGCGCACTGGGGATCTGTGGACTGCGTCCGGGGATGGGCTAGGGGGACATGCGCACGCTTTGGGCC
 TTACAGAATGTGATCGCGCGAGGGGGAGGGCGAAGCGTGGCGGGAGGGCGAGGCGAAGGAAGGAGGGCGTGA
 GAAAGGCGACGGCGGGCGGGGAGGAGGGTTATCTATACATTTAAAAACCAGCCGCCTGCGCCGCGCCTGC
 GGAGACCTGGGAGAGTCCGGCCGACGCGCGGGACACGAGCGTCCCACGCTCCCTGGCGCGTACGGCCTGCC
 ACCACTAGGCCTCCTATCCCCGGGCTCCAGACGACCTAGGACGCGTCCCTGGGGAGTTGCCTGGCGGGCC
 GTGCCAGAAGCCCCCTTGGGGCGCCACAGTTTTCCCGTTCGCTCCGGTTCTCTGCCTGCACCTTCTCGG
 GCGCGCCGGACCTGGAGCGGGCGGGTGGATGCAGGCGCGatggacggcggcacactgccaggtccgcgcc
 ccctgcgccccccgtccctgtcggctgcgctgccggcgagaccgcgctcccggaaactgttgcgctgcag
 ccggcgggcgaccggccaaccgcagagaccggagggcgagcggcggtagcgcggcgcaatgagcgcga
 gcqcaaccgcgtgaagctggtgaacttgggcttccaggcgctgcggcagcacgtgccgcagggcgccag
 caagaagctgagcaaggtggagacgctgcgctcagccgtggagtacatccgcgcgctgcagcgcctgctggc
 cyagcacgacgcccgtgcgcaacgcgctggcgggagggctgaggccgcaggccgtgcggccgtctgcgccccg
 cgggcccgcagggaccaccgccgtcgcgcctcgcctcccgcgcttcttctcgtccccgggcccgcgggggag
 ctcgagcccggctccccgcgttcgcctactcgtcggacgacagcggctgogaaggcgcgctgagtcctgc
 ggagcgcgagctactcgaactctccagctgggttagggggtactgaGCGCCCTCGACCTATGAGCCTCAGCC
 CCGGAAGCCGAGCGAGCGGCCCGCGCTCATCGCCGGGAGCCCGCCAGGTGGACCGCCCCGCGCTCCGCC
 CCCAGCGAGCCGGGACCCACCCACCCACCCCGCACCGCGACGCGCCTCGTTCGTCCGGCCAGCCTGA
 CCAATGCCGCGGTGGAAACGGGCTTGGAGCTGGCCCATAGGGCTGGCGGCTTCTCCGACGCGCCCTC
 CCCACAGCTTCTCGACTGCAGTGGGGCGGGGGCACCAACACTTGSAGATTTTCCGGAGGGGAGAGGATT
 TCTAAGGGCACAGAGAATCCATTTTCTACACATTAAGTGGAGCTGCTGGAGGGACACTGCTGGCAAACGGAG
 ACCTATTTTGTACAAAGAACCCTTGACCTGGGGCGTAATAAAGATGACCTGGACCCCTGCCCCACTATCT
 GGAGTTTTCCATGCTGGCCAAGATCTGGACACGAGCAGTCCCTGAGGGCGGGTCCCTGGCGTGGGCCCC
 CGTGACAGCCACCCTGGGGTGGGTTTGTGGGCACTGCTGCTCTGCTAGGGAGAAGCCTGTGTGGGGCACAC
 CTCTCAAGGGAGCGTGAACCTTTATAAATAAATCAGTTCGTTTTAAAAAAAAAAAAAAAAAAAA

SEQ ID NO:2

MDGGTLPRSAPPAPPVFPVVGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNRVKLVNLFQA
 LRQHVPHGGASKKLSKVETLRSAVEYIRALQRLLAHDAVRNALAGGLRPQAVRPSAPRGGPPGTTTPVAASPS
 RASSSPGRGGSSEPGSPRSAYSSDDSGCEGALS PAERELLD FSSWLGGY

SEQ ID NO:3

MSAPAARSASGAEAHRSRALSSPLTSWRSRVARAPQDSARLRSRCRPTSRRMAGSRAPSCFRGPGTKKGR
 ARRRPGNSLAARGAQTAAAPAAALPPARCARRRARPAGAAARGCTPRLSAASPPCSASCWRRAARAAAA
 PGSPSSPASRGCARAHCAALRPLRRLRSLRWPVAAAGCEATVPGTRVSAQQRERQGRGAQQARTWAVCRP
 SRLHPPARSRSRAAGRCRQRNRBRKLRPKGASCTAPFGNSPFGHAS

SEQ ID NO:4

GTACCTTGCTTTGGGGGCGCACTAAGTACCTGCCGGGAGCAGGGGGCGCACCGGGAACCTCGCAGATTTCGCC
AGTTGGGCGCACTGGGGATCTGTGGACTGCGTCCGGGGGATGGGCTAGGGGGACATGCGCACGCTTTGGGCC
TTACAGAATGTGATCGCGCCGAGGGGGAGGGCCGAAGCGTGGCGGGAGGGCGAGGCGAAGGAAGGAGGGCGT
GAGAAAGGCGACGGCCGGCGCGGGAGGAGGTTATCTATACATTTAAAAACCAGCCGCTGCGCCGCGCT
GCGGAGACCTGGGAGAGTCCGGCCGCACGCGCGGGACACGAGCGTCCCACGCTCCCTGGCGCGTACGGCCTG
CCACCACTAGGCCTCCTATCCCCGGGCTCCAGACGACCTAGGACGCGTGCCTTGGGGAGTTGCCCTGGCGGGC
CCGTGCCAGAAGCCCCCTTGGGGCGCCACAGTTTCCCCGTGCGCTCCGGTTCCTTGCCTGCACCTTCCTG
CGGCGCGCCGGGACCTGGAGCGGGCGGGTGGATGAGGCGCGatggacggcggcacactgcccaggtccgcg
ccccctgcgccccccgctccctgtcggtgcgctgcccggcgggagaccccggtccccggaactggtgcgctgc
agccggcggcggcgaaccggccaccgagagaccggaggcggcgcagcggccctagcgcggcgaatgagcgc
gagcgcbaaccgctgaagctggtgaacttgggcttccagggcgtgcgccagcacgtgccgcacggcggcggc
agcaagaagctgagcaaggtggagacgtgcgctcagccgtggagtacatccgcgctgcagcgcctgctg
gcccagcaagcgcctgcgcaacgcgctgggggagggctgaggccgcaggccgtgcgccgtctgccc
cgcgggcccgccagggaccaccccggctcgccgctcgccctccccgcttcttctgccccggcccgggggg
agctcggagcccggtcccccgcttccgctactcgtcggacgacagcggtgcyaggcgcgctgagctct
gcccagcgcgagctactcgaacttctccagctggttagggggctactgaGCGCCCTCGACCTAATAAGCCTCA
AGCCCCGGAACCCGAGCGAACGGGCCGGCGCTTCATCGCCGGGAAGCCCGCAAGGTGGACCGGGCCC
GCGCTCCGCCCCAGCGAGCCGGGACCCACCCACCCCGCACCGCCGACGCGCCTCGTTCGTCCGG
CCCAGCCTGACCAATGCCCGGTGGAAACGGGCTTGGAGCTGGCCCCATAGGGCTGGCGGCTTCCTCCGAC
GCGGCCCTCCCCACAGCTTCTCGACTGCAGTGGGGCGGGGGCACCACACTTGGAGATTTTTCCGGAGGG
GAGAGGATTTTCTAAGGGCACAGAGAATCCATTTCTACACATTAAGTGGAGCTGCTGGAGGGACACTGCTG
GCAAACGGAGACCTATTTTGTACAAAGAACCCTTGACCTGGGGCGTAATAAAGATGACCTGGACCCCTGCC
CCCCTATCTGGAGTTTCCATGCTGGCCAAGATCTGGACAGGAGTCCCTGAGGGGCGGGTCCCTGCC
GTGAGGCCCCCGTGACAGCCACCCTGGGGTGGGTTGTGGGCACTGCTGCTCTGCTAGGGAGAAGCCTGTG
TGGGGCACCCCTCTCAAGGGAGCGTGAACCTTTATAAATAAATCAGTTCGTTTAAAAAATAAATAAATAA
AAAACCGAGGGGGGGCCCGGAGCCAACAAA

SEQ ID NO:5

GGTAAACAGAACTGATTTATTTATAAAGTTCACGCTCCCTTGAAGAGGTGTGCCCCACACAGGCTTCTCCC
TAGCAGAGCAGCAGTGGCCACAAACCCACCCAGGGTGGGCTGTACGGGGCCCTCACGCCAGGGACCCCG
CCCCCTCAGGACTGCTCGTGTCCAGATCTTGGCCAGCATGGAAAAC'TCCAGATAGTGGGGGAGGGGTCCA
GGTCATCTTATACGCCCCAGGTCAAGGGTCTTTGTACAAAATAGGTCTCCGTTTGCCAGCAGTGTCC
CTCCAGCAGCTCAAGTTAATGTGTAGAAAATGGATTCTCTGTGCCCTTAGAAAATCCTCTCCCCTCCGGAA
AAATCTCCAAGTGTGGTGGCCCCCGCCCCACTGCAGTCGAGAAGCTGTGGGGAGGGGCGGCGTGGAGGA
AGCCGAGCCATTATGGGGCCAGCTCCAAGCCCGTTTCCACCGCGCATTTGGTCAAGGCTGGGCGGACGAA
CGAGGCGCGTCCGGCGGTGCGGGGGTGGTGGGTGGGTCCCCGGCTCGCTGGGGGCGGAGCAGCGGGCCGG
TCCACCTGGCGGGCTCCCC

SEQ ID NO:6

TTTTTTTTTTTTTTTTTAAACAGAACTGATTTATTTATAAAGTTCAGGCTCCCTTGAAGAGGTGTGCCCC
ACACAGGCTTCTCCCTAGCAGAGCAGCAGTGCCACAAACCCACCCAGGGTGGGCTGTCACGGGGGCTCA
CGCCAGGGACCCCGCCCTCAGGGACTGCTCGTGTCCAGATCTTGGCCAGCATGGAAAACCTCCAGATAGTGG
GGGCAGGGGTCCAGGTCACTTTATTACGCCCCAGGTCAAGGGTTCCTTGTACAAAAATAGGTCTCCGTTTGG
CCAGCAGTGTCCCTCCAGCAGCTCAAGTTAATGTGTAGAAAATGGATTCTCTGTGCCCTTAGAAAATCCTCT
CCCCCCGGAAAAATCTCCAAGTGTGGTGGCCCCCGCCCACTGCAGTCGAGAAGCTGTGGGGAGGGGCGG
CGTCCGAGGAAGCCGCCAGCCCTTATGGGGCCAGCTCCAAGCCCGTTTCCACCGCGGCATTGGTCAGGCTGG
GCCGGACGAACGAGGCGGCGTCCGGCGGTGCGGGGGTGGTGGGTGGGTCCCCGGCTCGCTGGGGGCGGAGCG
CGGGCCGGTCCACCTGGCGGGCTCCCCGGCGATGAGCGCGCCGGCCGCTCGCTCGGCTTCCGGGGCTGAGGC
TCATAGGTCGAGGGCGCTCAGTAGCCCCCTAACCCAGCTGGAGAAGTCGAGTAGCTCGCGCTCCGCAGGACTC
AGCGCGCCTTCGCAGCCGCTGTCTCGACGAGTAGGCGGAACGCGGGGAGCCGGGCTCCGAGCTGCCCCCG
CGCCCCGGGGACGAAGAAGCCCGGGAGGGCGAGGCGGCGACCGGGGTGGTCCCTGGCGGCCCGGGGGCGCA
GACGGCCGCACGGCCTGCGGCCTCAGCCCTCCCGCCAGCGCTTGCSCACGGCGTCTGTCTCGGCCAGCAGG
CGCTGCAGCGCGCGGATGTACTCCACGGCTGAGCGCAGCGTCTCCACCTTGCTCAGCTTCTTGGTGGCGCCG
CCGTGCGGCACGTGCTGCCGCAGCGCTGGAAGCCCAAGTTCACCAGCTTACGCGGTTGCGCTCGCGCTCA
TTGCGCCGCGCTACGGCCGCTGCGCCGCTCCGGTCTCTGCGGTGGCCGCTCGCCGCCCGGGCTGCAGCCG
AACAGTTCGGGGACCGGGTCTCCGCCGGGCAGCGCAGCCGACAGGGACGGGGGCGCAGGGGGCGCGGAC
CTGGGCAGTGTGCCCGCTCCATCGCGCCTGCATCCACCGCCCGCTCCAGGTCCCGGCGCGCCGCAGGAAG
GTGCAGGCAGAGGAACCGGAGGCGACGGGGAAAACCTGTGGCGCCCAAGGGGGCTTCTGGCACGGCGCCGCC
AGGCAACTCCCCAGGGCAGCGTCTTAGGTCTGTGGAGCCCGGGGATAGGAGCCCTAGTGGTGGCAGGCCG
TACCGCCAGGGAGCGTGGGACGCTCGTGTCCCGCGCTGCGGCCGACTCTCCAGGTCTCCGCAGGCGCG
GCGCAGGCGGCTGTTTTAAATGTATAGATAACCTCCTCCGCGCCCGCGCGTCCGCTTCTCAGGCCCT
CCTTCCTTCGCCTCGCCCTCCCGCCACGCTTCGCCCTCCCCCTCGCGGATCACATTCTGTAAGGCCCAAAG
CGTGGCATGTCCCCCTAGCCATCCCCCGGACGCAGTCCACAGATCCCAGTGGCCCAACTGGCGAAATC
TGCGAGTCCCCTGCGCCCCCTGCTCCCGGCAGGTACTTAGTGGCCCCCAAAGCAAGGTAC

SEQ ID NO:7

MCRKWILCALRKSSPLRKNLQVLVPPAPLQSRSCGEGRRRRKPPALMGPAPSPFFPPRHWSGWAGRTRRRR
CGGWWVGPRLAGGARARSTLAGFPDEARRPVRSGFRLRLIRSRALSSPLTSWRSRVARAPQDSARLRS
RCRPTSRNAGSRAPSCPRGPGTKKRGRARRRPGWLAARGAQTAAARPAASALPPARCARRRARPAGAAAR
GCTPRLSAASPPCSASCWERRAARAAAAPGSPSSPASRGCARAHCAALRPLRRLRSLRWPVAAAGCSATVP
GTRVSAGQSRQGRGAQGARTWAVCRRPSRLHPPARSRSRRAAGRCBQRNRRRRGKLWRPKGASGTAPPGN
SPGHAS

SEQ ID NO: 8

ATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTCTTTGGCATGTCCGCAAACGAGTTGCAGAC
 CAAGAACTAGGTGATGCCCCATTCTTGATCGGCTTCGCCGAGATCAGAAATCCCTAAGAGGAAGGGCAGC
 ACcCTcGGTCTGGACATCGAGACAGCCACACGTGCTGGAAAGCAGATAGtGGAGCGGAttctGAAAGAAAGAA
 TCCGATGAGGCACTTAAATGACCATGGACGGCGGCACACTGCCAGGTCCGCGCCCCCTGCGCCCCCGTC
 CCTGTCCGCTGCGCTGCCCGGGGAGACCCCGCTCCCGGAAGTGTTCGCTGCAGCCGGCGGGCGGCGACC
 GCCACCCGAGAGACCCGGAGGGCGGCAGCGCCGTAGCGCGGCCCAATGAGCGCGAGCGCAACCGGTGAAG
 CTGGTGAAGTTGGGCTTCCAGGCGCTGCGGCAGCACGTGCCGCACGGCGGGCCAGCAAGAAGCTGAGCAAG
 GTGGAGACGCTGCGCTCAGCCGTGGAGTACATCCGCGCGCTGCAGCGCTGCTGGCCGAGCAGCAGCCGTG
 CGCAACCGCTGCGGGGAGGGCTGAGGCCGAGGCCGTGCGGCCGTCTGCGCCCCGCGGGCCGCCAGGGACC
 ACCCGGTGCGCCGCTCGCCCTCCCGCGCTTCTTCGTCCCCGGGCCGCGGGGGCAGCTCGGAGCCCGGCTCC
 CCGCGTTCGCGCTACTCGTCGGACGACAGCGGCTGCGAAGCGCGCTGAGTCTGCGGAGCGCGAGCTACTC
 GACTTCTCCAGCTGGTTAGGGGGCTACactagtggccaccatcaccatcaccattaa

SEQ ID NO:9

ATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTCTTTGGCATGTCCGCAAACGAGTTGCAGA
 CCAAGAACTAGGTGATGCCCCATTCTTGATCGGCTTCGCCGAGATCAGAAATCCCTAAGAGGAAGGGCA
 GCACCCTCGGTCTGGACATCGAGACAGCCACACGTGCTGGAAAGCAGATAGTGGAGCGGATTCTGAAAGAA
 GAATCCGATGAGGCACTTAAATGACCATGGACGGCGGCACCCTGCGCGTTCGCGCCGCGCGCGCCGCC
 AGTTCGGTTGGCTGCGCTGCCCGTCCCGCTCCCGCTCCCGGAACTGCTGCGCTGCAGCCGTCCCGCTC
 GCCCGCCACCGCAGAGACCCGAGGCGGGCGCAGCGGCCGTAGCCCGCGCAATGAGCGCGAGCGCAACCGC
 GTGAAGCTGTTGAAGTTGGGCTTCCAGGCGCTGCGGCAGCACGTGCCGCACGGCGGGCCAGCAAGAAGCT
 GAGCAAGGTGGAGACGCTGCGCTCAGCCGTGGAGTACATCCGCGCGCTGCAGCGCTGCTGGCCGAGCACG
 ACGCGTTCGCAACCGCTGCGGGGAGGGCTGAGGCCGAGGCCGTGCGGCCGTCTGCGCCCCGCGGGCCG
 CCAGGGACCACCCCGGTGCGCCGCTCGCCCTCCCGCGCTTCTTCGTCCCCGGGCCGCGGGGGCAGCTCGGA
 GCCCGCTCCCGCGTTCGCGCTACTCGTCGGACGACAGCGGCTGCGAAGCGCGCTGAGTCTGCGGAGC
 GCGAGCTACTCGACTTCTCCAGCTGGTTAGGGGGCTACACTAGTGGCCACCATCACCATCACCATTAA

SEQ ID NO:10

MDPNTVSSFQVDCFLWHVRKRVDQELGDAPFLDRLRRDQKSLRGRGSLGLDIETATRAGKQIVERILKEE
 SDEALKMTMDGGTLERSAPPAPPVFGCAARRRPASPELLRCSRRRRPATAETGGGAAAVARRNERERNVK
 LVNLGFQALRQHVPHGASKKLSKVETLRSVEYIRALQRLLAEHDAVRNALAGGLRPQAVRPSAPRGGPGT
 TPVAASPSRASSSPGRGGSSEPGSPRSAYSSDDSGCEGALSPAERELDFSSWLGGYTSGHHHHH

SEQ ID NO:11

MYSTAERSVSTLLSFLAPPCGTCCRSANKPKFTSFTLRSRLRRATAAAPPVSAVAGRRRRLQRNSSG
 DAGLRRAAQPTGTGGAGGADLGSVPPSIAPASTRPLQVPARRRKVQAEPEATGKTVAPOGGFWHGAARQL
 PRARVLGRLEPGDRRPSGGPFYAPGSVGRSCPARAAGLSQVSAAGAAQAGF

SEQ ID NO:12

MEAHLDWYGVPLQEQASDACPRESCESSALPEAREGANVHFPHPVPREHFSCAPELVAGAQLNASLMDG
GALPRLMPTSSGVAGACAARRRQASPELLRCSRRRRSGATEASSSSAAVARENERERNRVKLVNLGFQALR
QHVPHGANKKLSKVETLRSVEYIYRALQRLLAZHDVRAALAGLLTPATPPSDECAQPSASPASASLSC
ASTSPSPDRLGCSEPTSPRSAYSSEESSCEGELSPMEQELLDPFSSWLGYY

SEQ ID NO:13

GCCCGGAGCATGGAAGCACGTCAGCTAGGCCATGAACTGCACCCGGGAGGGGTGGGGGTGGAAGCGCACGG
TGTCAGCTTTGCAGAATGTGTACCCAAGGGGAGGGCGAGGCGAAGGAABGAGGGCGTAAGAAAGGAGGCG
GTGGCGGGGGCGGAGGAGATTATCTATACTTTTTAAAAAAAAGGAGCCTCTTAGCCGCGTAAAGGAGACTTG
GGGAGCGCCTGACAGCACGCGCGGGACACGAGAGTACCACGCTTCCCTACTCTTTTCAGACCTTGACTGGT
ACGGGGTCCCAGGACTGCAGGAGCCAGCGACCGCTGCCCCTAGGGAGTCTGTCAGCAGTGCCTGCTGAG
GCCCCGTAAGGTGCAAACGTCCTCCACCCGACCCGGTTCCTCGCGAGCACFTTTCCTGTGCCGACC
AGAACTCGTAGCAGGGGCCAGGGCTGAATGCAAGCTTGATGGACGGCGCGCTGCCAGACTCATGC
CCACCTCGTCTGGAGTCGCTGGAGCCTGCGCTGCTCGGCGGAGACAAGCGTCTCCGGAATTGCTGCGCTGC
AGCCGGCGGCGCGATCTGAGCAACCGAGGCCAGCAGCAGCTCGGCGTCCGTGGCACGCGCAATGAGCG
CGAGCGCAACCGCTAAAGCTGGTAAACTTGGGCTTCCAGGCGCTGCGGCAGCACGTGCCGACGGGGCG
CCAACAAGAAGCTGAGTAAGGTGGAGACGCTGCGCTCCGCGTAGAGTACATTCGTGCGCTGCAGCGGCTG
CTCGCAGACACGACACGGTGCAGCCGGNGCTCGCTGGGGGCTGTTAACACCCGCTACTCCGCCGTCCGA
TGAGTGCACGACCCCTGCTCCCTGCCAGCGGGTCTCTGTCTGCGCTCTACGTCTCCGTCCCGGA
CCCTGGGCTGCTCTGAGCCTACCTCCCCGCGCTCCGCCACTCTGTCGGAGGAAAGCAGCTGCGAGGGAGAG
CTAAGCCCGATGGAGCAGGAGCTGCTTGACTTTTCCAGTTGGTTAGGGGGCTACTGA

SEQ ID NO:14

MESHFNWYGVPRQLQKASDACPRESCESSALPEAREGANVHFPHPVPREHFS CGAPKPVAGAPALNASLMDG
GALPRLVPTSSGVAGACTARRRPPPELLRCSRRRRSGATEASSSSAAVARRNERERNRVKLVNLGFQALR
QHVPHGANKKLSKVETLRSVEYIYRALQRLLAZHDVRAALSGLLTPATRPSDVCTQPSASPASASLSC
TSTSPDRLGCSEPASPRAYSSEEDSSCEGRTYPMGQMFDFSNWLGYY

SEQ ID NO:15

TTCACCCGGCTGCAAGCGCTAGGTGTACGGAGACCTGGCAGCTCTTGGGGCTTAAGGACTGAGCRCCAGAG
CCGGTGGAGGTTCTGTGGAGTACATTCGGACCCTCTCACAGCCCCGAGAGTGCGGGACGTGCCGAGCGC
AGTTCGGGATCTGCACTCGAGGACTTGTCGAGGACGCATTAAGCTAAGCATCTGCTCGGAGCATGGAATCG
CACTTTAACTGGTACGGGGTCCCAAGGCTCCAGAAGGCTAGCGACCGTGCCTAGGGAATCCTGCAGCAG
TGCCCTGCTGAGGCCCGTGAAGGTGCGAACGTCCTCCACCCGACCCCGTTCCTCGCGAGCACFTTTC
CCTGTGGCGCACCGAAACCCGTAGCGGGGGCCCCGGCGCTGAATGCAAGCTTGATGGACGGGGCGCGCTG
CCCAGACTCGTGCACCTCGTCTGGAGTCGCTGGAGCCTGCACCTGCTCGGCGGAGACCCCCGTCCCGGA

ACTGCTTCGCTGCAGCCGACGGCGGCGATCGGGAGCAACCGAGGCCAGCAGCAGCTCGGCGGCCGTGGCAC
 GCCGCAATGAGCGTGAGCGCAACCGCGTAAAGCTGGTAAACTTGGGCTTCCAGGCGCTGCGGCAGCACGTG
 CCGCACGGCGGCGCCRA.CAAGAAGCTGAGTAAAGGTGGAGACGCTGCGCTCCGCGGTAGAGTACATCCGTGC
 GCTGCAGCGGCTGCTAGCAGAGCACGACGCGGTGCGTGCCTGCGCTCTCTGGGGGTCTATTAAACACCCGCTA
 CTCGGCCGTCCGATGTGTGCACGCGAGCCCTCCGCTCCCTGCCAGCGCTCTCTGTCTGCACCTCTACA
 TCCCCAGACCGCTAGGCTGCTCCGAGCCTGCTCTCCGCGCTCCGCTACTCGTCCGAGGACAGCAGCTG
 CGAGGGAGAGACTTACCCGATGGGGCAGATGTTTGACTTTTCCAATTGGTTAGGGGGCTACTGAGCACCCC
 ACACCCCTAAGCTGCGTCCCTGGGTGTCCCTGGTGGACCTACCTGCGTTTCTTGCCCAGGAAACCTGGGC
 CCATGCCTTACCCATGCTGTCTAGTGCAGCCTGACCAAATGCCAAGTACTGACCTCTGCTCGGCCTCCAG
 CCGCGGAATGACATCTTCCATCTCCAGTCTTTGCCGAACCAGGACTTGGAATTTCTCAGGAGAAAGAAT
 TTTACAATGACAATCTGCTTTTTATCAATTAACCTTGAACCTGCTGGAGGACTCTGCTGAAAATATGAAGAAT
 TATTTTTATACAAAGGATCCTTAAGCTTGGAGCACAATAAAGATGACCTCTGTCTCTCACCCCCACTGTCT
 AGAACTTTCCAACCTGGCCAAAGTGTGGACGGTTCGGGCCCTGAGGGCAAGATGCCTGGCTGCACCCCTTCT
 TCCTCTTCCGAAGCCTATCCTGACGCTGATGTTTGGCCAGTGTGGGAACCCCTGCTATTGCAAAGTGTACTA
 TTCTATAAAAGTTGTTTTTCATTGGAAAGGAATTC

SEQ ID NO:16

KLVN1GFQAL

SEQ ID NO:17

ELLDFFSSWL

SEQ ID NO:18

RLLAEHDAV

SEQ ID NO:19

KLVN1GFQA

SEQ ID NO:20

EYIRALQRL

SEQ ID NO:21

EYIRALQRL

SEQ ID NO:22

AVRNALAGGL

SEQ ID NO:23

SEPGSPRSAY

SEQ ID NO:24

VETLRSAVEY

SEQ ID NO:25

IRALQRLLA

SEQ ID NO:26

LRPQAVRPS

SEQ ID NO:27

LRQHVPHGG

SEQ ID NO:28

LGFOALRQH

SEQ ID NO:29

VRNALAGGL

SEQ ID NO:30

YIRALQRL

SEQ ID NO:31

LVNLLGFOAL

SEQ ID NO:32

VEYIRALQR

SEQ ID NO:33

LLRCSRRRR

<110> 史密丝克莱恩比彻姆生物有限公司
(SmithKline Beecham Biologicals s.a.)

<120> 新化合物

<130> BC45300

<160> 33

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 1791

<212> DNA

<213> 人类

<400> 1

```

gtaccttget ttgggggcgc actaagtaac tgcogggagc agggggcgca cogggaacte    60
gcagatttgc ccagttgggc gcaactggga tctgtggact gcgtcogggg gatgggctag    120
ggggacatgc gcacgctttg ggccttacag aatgtgatcg cgcgaggggg agggcgaagc    180
gtggcgggag ggcgagggca aggaaggagg gcgtgagaaa ggcgacggcg gcggcgggga    240
ggagggttat ctatacattt aaaaaccagc cgcctgogcc gcgootgogg agacctggga    300
gagtcogggc gcacgcgcgg gacacgagcg tcccacgctc octggcgcgt aoggcctgcc    360
accactaggc ctccatcccc ogggctccag acgacctagg acgctgccc tggggagtgt    420
cctggcggcg cctgcccaga agccccctt gggcgcaca gtttccccg togcctccgg    480
ttcctctgcc tgcaccttcc tgcggcgcgc cgggacctgg agcggcggg tggatgcagg    540
cgcgatggac ggcggcacac tgcccaggtc cgcgccccct gcgccccccg tccctgtcgg    600
ctgcgctgcc cggcggagac ccgcgtcccc ggaactgttg cgtgcagcc ggogggcgcg    660
accggccacc gcagagaccg gaggcggcgc agcggcogta gcgcgcgca atgagcgcga    720
gcgcaaccgc gtgaagctgg tgaacttggg cttccaggcg ctgcggcagc acgtgccgca    780
cggcggcgcc agcaagaagc tgagcaaggt ggagacgctg cgtcagcog tggagtacat    840
ccgcgogctg cagcgcctgc tggcggagca cgacgcctg cgcaacgcgc tggcgggagg    900
gctgaggcgc caggcctgic ggccgtctgc gcccogggg ccgocaggga ccacccgggt    960
cgccgcctcg ccctccogcg cttcttcgtc cccgggcgcg gggggcagct cggagcccgg    1020
ctccccogct tccgcctact cgtcgggaga cagcggctgc gaaggcgcgc tgagtctcgc    1080
ggagcgcgag ctactcgact tctccagctg gttagggggc tactgagcgc cctcgacctc    1140
tgacctcag ccccggaagc cgagcgcgog gccggcgcgc toatcgccgg ggagcccgcc    1200
aggtggaccg gcccgcgctc cgcccccagc gagccgggga cccaccacc accccccgca    1260
ccgccgacgc cgcctcgttc gtccggccca gcctgaccaa tgccgcggtg gaaacgggct    1320

```

```

tggagctggc cccataaggg ctggcggett cctcgaagc cgcacctcc caccagcttct 1380
cgactgcagt gggcgggggg gcaccaacac ttggagattt tccggaggg gagaggattt 1440
tetaagggca cagagaatcc attttttaca cattaacttg agctgctgga gggacactgc 1500
tggcaaacgg agaactattt ttgtacaaag aaccottgac ctggggcgta ataaagatga 1560
cctggacccc tgcacctact atctggagtt tccatgctg gccaaagatct ggacacgagc 1620
agtccctgag gggcgggggtc cctggcgtga ggccccgtg acagcccacc ctggggtggg 1680
tttgtgggca ctgctgctct gctagggaga agcctgtgtg gggcacacct cttaaggga 1740
gctgaactt tataaataaa tcagttotgt ttaaaaaaaaa aaaaaaaaaa a 1791

```

<210> 2

<211> 193

<212> PRT

<213> 人类

<400> 2

```

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

```

<210> 3

<211> 262

<212> PRT

<213> 人类

<400> 3

Met Ser Ala Pro Ala Ala Arg Ser Ala Ser Gly Ala Glu Ala His Arg
 1 5 10 15
 Ser Arg Ala Leu Ser Ser Pro Leu Thr Ser Trp Arg Ser Arg Val Ala
 20 25 30
 Arg Ala Pro Gln Asp Ser Ala Arg Leu Arg Ser Arg Cys Arg Pro Thr
 35 40 45
 Ser Arg Arg Asn Ala Gly Ser Arg Ala Pro Ser Cys Pro Arg Gly Pro
 50 55 60
 Gly Thr Lys Lys Arg Gly Arg Ala Arg Arg Arg Pro Gly Trp Ser Leu
 65 70 75 80
 Ala Ala Arg Gly Ala Gln Thr Ala Ala Arg Pro Ala Ala Ser Ala Leu
 85 90 95
 Pro Pro Ala Arg Cys Ala Arg Arg Arg Ala Arg Pro Ala Gly Ala Ala
 100 105 110
 Ala Arg Gly Cys Thr Pro Arg Leu Ser Ala Ala Ser Pro Pro Cys Ser
 115 120 125
 Ala Ser Cys Trp Arg Arg Arg Ala Ala Arg Ala Ala Ala Pro Gly
 130 135 140
 Ser Pro Ser Ser Pro Ala Ser Arg Gly Cys Ala Arg Ala His Cys Ala
 145 150 155 160
 Ala Leu Arg Pro Leu Arg Arg Leu Arg Ser Leu Arg Trp Pro Val Ala
 165 170 175
 Ala Ala Gly Cys Ser Ala Thr Val Pro Gly Thr Arg Val Ser Ala Gly
 180 185 190
 Gln Arg Ser Arg Gln Gly Arg Gly Ala Gln Gly Ala Arg Thr Trp Ala
 195 200 205
 Val Cys Arg Arg Pro Ser Arg Leu His Pro Pro Ala Arg Ser Arg Ser
 210 215 220
 Arg Arg Ala Ala Gly Arg Cys Arg Gln Arg Asn Arg Arg Arg Arg Gly
 225 230 235 240
 Lys Leu Trp Arg Pro Lys Gly Ala Ser Gly Thr Ala Pro Pro Gly Asn
 245 250 255
 Ser Pro Gly His Ala Ser
 260

<210> 4

<211> 1830

<212> DNA

<213> 人类

<400> 4

```

gtaccttgct ttgggggocg actaagtacc tgccgggagc agggggcgca ccgggaactc      60
gcagatttcg ccagttgggc gcaactgggga tctgtggact gcgtccgggg gatgggctag      120
ggggacatgc gcaocgtttg ggccttacag aatgtgatcg cgcogagggg gagggcogaa      180
gcgtggcggg agggcgagge gaaggaagga gggcgtgaga aaggcgacgg cggcggcgcg      240
gaggagggtt atctatacat ttaaaaacca gccgcctgcg ccgcgcctgc ggagacctgg      300
gagagtccgg ccgcacgcgc gggacacgag cgtcccacgc tccttggcgc gtacggcctg      360
ccaccactag gcctcctate cccgggctcc agacaccta ggacgcgtgc cctggggagt      420
tgctggcggc cgcctgcca gaagcccctc tggggcgcca cagttttccc cgtcgcctcc      480
ggttcctctg cctgcacctt cctgcggcgc gccgggacct ggagcgggcg ggtggatgca      540
ggcgcgatgg acggcggcac actgccaggc tccgcgcccc ctgoccccc cgtccctgtc      600
ggctgcgctg cccggcggag acccgcgtcc ccggaactgt tgogctgcag ccggcggcgg      660
cgaccggcca ccgcagagac cggaggcggc gcagcggccg tagcgcggcg caatgagcgc      720
gagcgcaacc gcgtgaagct ggtgaacttg ggcttcacag cgtgcggca gcacgtgccg      780
cacggcggcg ccagcaagaa gctgagcaag gtggagacgc tgogctcagc cgtggagtac      840
atccgcgcgc tgcagcgcct gctggccgag cacgacgcgc tgcgcaacgc gctggcggga      900
gggctgaggc cgcaggccgt gggccctctt gcgccccgcg ggccgcacag gaccaccccg      960
gtcgcgcctt cgcctcccg cgttctctg tccccggccc gcgggggcag ctcggagccc     1020
ggctccccgc gttccgccta ctctoggac gacagcggct gcgaaggcgc gctgagtcct     1080
gcggagcgcg agctactoga cttctccagc tggtagggg gctactgagc gccctcgacc     1140
taataagcct caagccccg aaaccgagc gaacgggccc gcgcgcttca tcgcgggga     1200
agcccccaa ggtggacogg gcccgcctc cgcccccagc gagccgggga cccaccacc     1260
acccccgca ccgcgcagc cgcctcgttc gtccggcca gcctgaccaa tgccgcggtg     1320
gaaacgggct tggagctggc ccataaggg ctggcggctt cctccgacgc cgcctctccc     1380
cacagcttct cactgcagt gggcggggg gcaccaaac ttggagattt ttccggaggg     1440
gagaggattt tctaagggca cagagaatcc atttctaca cattaacttg agctgctgga     1500
gggacactgc tggcaaacgg agacctattt ttgtacaaag aacccttgac ctggggcgta     1560
ataaagatga cctggacccc tgccccact atctggagt ttcoctgctg gccaaagatct     1620
ggacacgagc agtcctgag gggcggggtc cctggcgtga ggccccctg acagcccacc     1680
ctggggtggg tttgtgggca ctgctctct gctagggaga agcctgtgtg gggcacacct     1740
cttcaaggga gcgtgaactt tataaataaa tcagttctgt taaaaaaaa aaaaaaaaaa     1800
aaaaccgagg gggggcggg agccaacaaa     1830

```

<210> 5

<211> 587

<212> DNA

<213> 人类

<400> 5

ggtaaacaga actgatttat ttataaagtt cacgctccct tgaagagggtg tgccccacac	60
aggettctcc ctagcagage agcagtgecc acaaacccac cccagggtgg gctgtcacgg	120
gggcctcaag ccagggaccc cgcacctcag ggactgctcg tgtccagatc ttggccagca	180
tggaaaactc cagatagtgg gggcaggggt ccaggtcac tttattacgc cccagggtcaa	240
gggttctttg tacaaaaata ggtctccgtt tgccagcagt gtcctccag cagctcaagt	300
taatgtgtag aaaatggatt ctctgtgccc ttagaaaaac ctctccccctc cggaaaaatc	360
tccaagtgtt ggtgcccccc gccccactgc agtcgagaag ctgtggggag gggcggcgtc	420
ggaggaagcc gcagcccatt atggggccag ctccaagccc gtttccacog cggcatlgtt	480
caggctgggc ggaagaaaga ggcggcgctc gcgggtcggg ggggtggtggg tgggtccccg	540
gctcgtggg ggcggagcag cgggcccgtc cacctggcgg gctcccc	587

<210> 6

<211> 1791

<212> DNA

<213> 人类

<400> 6

ttttttttt tttttttta aacagaactg atttatttat aaagttcacg ctccctgaa	60
gagggtgtgc ccacacagge ttctccctag cagagcagca gtgcccacaa acccaccaca	120
gggtgggctg tcaogggggc ctacgccag ggaccocgc cctcaggagc tgctcgtgtc	180
cagatcttgg ccagcatgga aaactccaga tagtgggggc aggggtccag gtoatcttta	240
ttacgcccc ggtcaagggt tctttgtaca aaaataggtc tccgtttgc agcagtgtcc	300
ctccagcagc tcaagttaat gtgtagaaaa tggattctct gtgcccttag aaaatcctct	360
cccctccgga aaaatctoca agtgttgggt cccccgcgc cactgcagtc gagaagctgt	420
ggggaggggc ggcgtcggag gaagcogcoa gcccttatgg ggcoagctcc aagcccgttt	480
ccaccgggce attggtcagg ctgggcccga cgaacgaggc ggcgtcggcg gtgogggggg	540
tgggtgggtg gtccccggt ctctgggggc ggagcgcggg ccggtccacc tggcgggctc	600
cccggcgatg agcgcgcgg ccgctcgtc ggcttcggg gctgaggctc ataggctcag	660
ggcgtcagc agccccctaa ccagctggag aagtcagta gctcgcgctc cgcaggactc	720
agcgcgcctt cgcagccgct gtctcgcag gagtagggg aacgcgggga gccgggctcc	780
gagctgcccc cgcggcccgg ggacgaagaa gcgcgggagg gcgaggcggc gaccggggtg	840
gtccctggcg gcccgcgggg cgcagacggc cgcacggcct gggcctcag cctcccgc	900
agcgcgttgc gcacggcgtc gtctcggcc agcaggcgtc gcagcgcgg gatgtactcc	960
acggctgagc gcagcgtctc caccctgtct agctcttgc tggcgcgcgc gtgcggcaag	1020
tgctcgcga ggccttgaa gcccaagttc accagcttca cgcggttgcg ctccgcgtca	1080
ttgcgcgcg ctacggccgc tgcgcgcct ccggtctctg cgggtggcgg tgcgcgcgc	1140
cggctgcagc gcaacagttc cggggacggg ggtctccgc gggcagcgc gccgacaggg	1200
acggggggcg cagggggcgc ggacctgggc agtgtgccc cgtccatcgc gcctgcctcc	1260
accgcgcgc tccaggtccc ggcgcgcgc aggaaggtgc aggcagagga accggaggcg	1320
acggggaaaa ctgtggcgc ccaagggggc ttctggcaag gcgcgcgcag gcaactcccc	1380
agggcacgcg tcttaggtcg tctggagccc ggggatagga ggocctagtg tggcaggccc	1440

```

tacgcgccag ggagcgtggg acgctogtgt ccgcgcgtg cggccggaot ctcccaggtc 1500
fccgcaggcg cggcgcagge ggctggtttt taaatgtata gataacoctc ctccgcgccg 1560
ccgcgctcgc ctttctcaag cctccttcc ttcgcctcgc cctcccgcca cgettcgccc 1620
tccccctcgc gcgatecaat tetgtaagge ccaaagcgtg cgcattgtccc cctagcccat 1680
cccccgagc cagtccacag atccccagtg cgcccaactg gcgaaatctg cgagttcccg 1740
gtgcgcccc tgctccoggc aggtacttag tgogcccca aagcaaggta c 1791

```

<210> 7

<211> 361

<212> PRT

<213> 人类

<400> 7

```

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

```


<212> DNA

<213> 流感病毒和人类

<400> 9

```

atggatccaa acaactgtgtc aagctttcag gtagattgct ttctttggca tgtccgcaaa      60
cgagttgcag accaagaact aggtgatgcc ccattccttg atcggttcg ccgagatcag      120
aaatccctaa gaggaagggg cagcacctc ggtctggaca tcgagacagc cacacgtgct      180
ggaaagcaga tagtggagcg gattctgaaa gaagaatcog atgaggoact taaaatgacc      240
atggacggcg gcaccctgcc gcgttcocog ccgccggcgc cgcocagttcc ggttggctgc      300
gctgccocgc gcocgtccocg gtccccggaa ctgctgcgct gcagcogtcg ccgtcccccg      360
gccaccgcag agaccggagg cggcgcagcg gccgtagcgc ggcgcaatga gcgcgagcgc      420
aaccgcgtga agctggtgaa cttgggcttc cagcgcctgc ggcagcacgt gccgcacgge      480
ggcgcacgca agaagctgag caaggtggag acgctgcgct cagcctgga gtacatccgc      540
gcgctgcagc gcctgctggc cgagcacgac gccgtgcgca acgcgctggc gggagggcctg      600
aggccgcagg ccgtgcggcc gtctgcgcc ccgcggccgc cagggaccac cccggtegcc      660
gcctgccect ccgcgccttc ttctccccc ggccgggggg gcagctcgga gcccggtcc      720
ccgcgttcgc cctactcgtc ggacgacagc ggctgcgaag gcgcgctgag tctgcggag      780
cgcgagctac tcgacttctc cagctggtta ggggctaca ctagtggcca ccatccat      840
caccattaa                                     849

```

<210> 10

<211> 282

<212> PRT

<213> 流感病毒和人类

<400> 10

```

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

```

Ala Ala Ala Val Ala Arg Arg Asn Glu Arg Glu Arg Asn Arg Val Lys
 130 135 140
 Leu Val Asn Leu Gly Phe Gln Ala Leu Arg Gln His Val Pro His Gly
 145 150 155 160
 Gly Ala Ser Lys Lys Leu Ser Lys Val Glu Thr Leu Arg Ser Ala Val
 165 170 175
 Glu Tyr Ile Arg Ala Leu Gln Arg Leu Leu Ala Glu His Asp Ala Val
 180 185 190
 Arg Asn Ala Leu Ala Gly Gly Leu Arg Pro Gln Ala Val Arg Pro Ser
 195 200 205
 Ala Pro Arg Gly Pro Pro Gly Thr Thr Pro Val Ala Ala Ser Pro Ser
 210 215 220
 Arg Ala Ser Ser Ser Pro Gly Arg Gly Gly Ser Ser Glu Pro Gly Ser
 225 230 235 240
 Pro Arg Ser Ala Tyr Ser Ser Asp Asp Ser Gly Cys Glu Gly Ala Leu
 245 250 255
 Ser Pro Ala Glu Arg Glu Leu Leu Asp Phe Ser Ser Trp Leu Gly Gly
 260 265 270
 Tyr Thr Ser Gly His His His His His His
 275 280

<210> 11
 <211> 193
 <212> PRT
 <213> 人类

<400> 11
 Met Tyr Ser Thr Ala Glu Arg Ser Val Ser Thr Leu Leu Ser Phe Leu
 1 5 10 15
 Leu Ala Pro Pro Cys Gly Thr Cys Cys Arg Ser Ala Trp Lys Pro Lys
 20 25 30
 Phe Thr Ser Phe Thr Arg Leu Arg Ser Arg Ser Leu Arg Arg Ala Thr
 35 40 45
 Ala Ala Ala Pro Pro Pro Val Ser Ala Val Ala Gly Arg Arg Arg Arg
 50 55 60
 Leu Gln Arg Asn Ser Ser Gly Asp Ala Gly Leu Arg Arg Ala Ala Gln
 65 70 75 80
 Pro Thr Gly Thr Gly Gly Ala Gly Gly Ala Asp Leu Gly Ser Val Pro
 85 90 95
 Pro Ser Ile Ala Pro Ala Ser Thr Arg Pro Leu Gln Val Pro Ala Arg
 100 105 110
 Arg Arg Lys Val Gln Ala Glu Glu Pro Glu Ala Thr Gly Lys Thr Val

115	120	125	
Ala Pro Gln Gly Gly Phe Trp His Gly Ala Ala Arg Gln Leu Pro Arg			
130	135	140	
Ala Arg Val Leu Gly Arg Leu Glu Pro Gly Asp Arg Arg Pro Ser Gly			
145	150	155	160
Gly Arg Pro Tyr Ala Pro Gly Ser Val Gly Arg Ser Cys Pro Ala Arg			
165	170	175	
Ala Ala Gly Leu Ser Gln Val Ser Ala Gly Ala Ala Gln Ala Ala Gly			
180	185	190	
Phe			

<210> 12

<211> 263

<212> PRT

<213> 小鼠

<400> 12

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

Thr Pro Pro Ser Asp Glu Cys Ala Gln Pro Ser Ala Ser Pro Ala Ser
 195 200 205
 Ala Ser Leu Ser Cys Ala Ser Thr Ser Pro Ser Pro Asp Arg Leu Gly
 210 215 220
 Cys Ser Glu Pro Thr Ser Pro Arg Ser Ala Tyr Ser Ser Glu Glu Ser
 225 230 235 240
 Ser Cys Glu Gly Glu Leu Ser Pro Met Glu Gln Glu Leu Leu Asp Phe
 245 250 255
 Ser Ser Trp Leu Gly Gly Tyr
 260

- <210> 13
- <211> 1051
- <212> DNA
- <213> 小鼠

<400> 13

```

gcccggagca tggaaacacg tcagctaggc catgaactgc acccgggagg ggtgggggtg      60
gaagcgcacg gtgtcagctt tgcagaatgt gtacaccaag gggagggcga ggcaaggaa      120
ggagggcgta agaaaggagg cgggtggcggg gcggaggaga ttatctatac ttttaaaaa      180
aaaggagcct cttagccgcg taaaggagac ttggggagcg cctgacagca cgcgcgggac      240
acgagagtac cacgcttccc tactcttttc agaccttgac tggtaggggg tcccaggact      300
gcaggaggcc agcgcgcgct gcacctaggga gtcctgcagc agtgccctgc ctgaggcccg      360
tgaaggtgca aacgtccact tcccaccgca cccggttccct cgcgagcaact tttcctgtgc      420
cgcaccagaa ctctagcag gggcccaggg gctgaatgca agcttgatgg acgcggcgc      480
gctgccaga ctcatgcca cctcgtctgg agtcgctgga goctgcgctg ctcgccggag      540
acaagcgtct ccggaattgc tgcgctgcag ccggcggcgg ogatctggag caaccgaggc      600
cagcagcagc tggcgctccg tggcacgccg caatgagcgc gagcgcgaacc gcgtaaagct      660
ggtaaacttg ggcttccagg cgtctgcgca gcacgtgccg cacggcggcg ccaacaagaa      720
gctgagtaag gtggagacgc tgcgctccgc ggtagagtac attcgtgcgc tgcagcggct      780
gctcgcagag cagcacagcg tgcggccggn gctcgtcggg gggctgttaa caccgcctac      840
tccgcgctcc gatgagtgca cgcagccctc tgcctcccct gccagcgggt ctctgtctg      900
cgctctacg tctccgtccc ggacctggg ctgctctgag cctacctccc cgcgctccgc      960
ctactcgtcg gaggaaagca gctgcgaggg agagctaagc ccgatggagc aggagctgct     1020
tgacttttcc agttggttag ggggetactg a                                     1051
    
```

- <210> 14
- <211> 260
- <212> PRT
- <213> 大鼠

<400> 14

Met Glu Ser His Phe Asn Trp Tyr Gly Val Pro Arg Leu Gln Lys Ala
 1 5 10 15
 Ser Asp Ala Cys Pro Arg Glu Ser Cys Ser Ser Ala Leu Pro Glu Ala
 20 25 30
 Arg Glu Gly Ala Asn Val His Phe Pro Pro His Pro Val Pro Arg Glu
 35 40 45
 His Phe Ser Cys Gly Ala Pro Lys Pro Val Ala Gly Ala Pro Ala Leu
 50 55 60
 Asn Ala Ser Leu Met Asp Gly Gly Ala Leu Pro Arg Leu Val Pro Thr
 65 70 75 80
 Ser Ser Gly Val Ala Gly Ala Cys Thr Ala Arg Arg Arg Pro Pro Ser
 85 90 95
 Pro Glu Leu Leu Arg Cys Ser Arg Arg Arg Arg Ser Gly Ala Thr Glu
 100 105 110
 Ala Ser Ser Ser Ser Ala Ala Val Ala Arg Arg Asn Glu Arg Glu Arg
 115 120 125
 Asn Arg Val Lys Leu Val Asn Leu Gly Phe Gln Ala Leu Arg Gln His
 130 135 140
 Val Pro His Gly Gly Ala Asn Lys Lys Leu Ser Lys Val Glu Thr Leu
 145 150 155 160
 Arg Ser Ala Val Glu Tyr Ile Arg Ala Leu Gln Arg Leu Leu Ala Glu
 165 170 175
 His Asp Ala Val Arg Ala Ala Leu Ser Gly Gly Leu Leu Thr Pro Ala
 180 185 190
 Thr Arg Pro Ser Asp Val Cys Thr Gln Pro Ser Ala Ser Pro Ala Ser
 195 200 205
 Ala Ser Leu Ser Cys Thr Ser Thr Ser Pro Asp Arg Leu Gly Cys Ser
 210 215 220
 Glu Pro Ala Ser Pro Arg Ser Ala Tyr Ser Ser Glu Asp Ser Ser Cys
 225 230 235 240
 Glu Gly Glu Thr Tyr Pro Met Gly Gln Met Phe Asp Phe Ser Asn Trp
 245 250 255
 Leu Gly Gly Tyr
 260

<210> 15
 <211> 1526
 <212> DNA
 <213> 大鼠

<400> 15

ttcaccggc tgcaagcgct aggtgtacgg agacctggca gctcttgggg ctttaaggact

60

gagcrccaga gccggtggag gttcctgtgg agtacattog gaccctctoa cagccccoga 120
 gagtgcggga cgtgcggagc gcagttcggg atctgcactc gaggacttgt cgaggaogca 180
 ttaagctaag catctgctcg gagcatggaa tcgcacttta actggtagcg ggtcccaagg 240
 ctccagaagg ctagcgacgc gtgccctagg gaatcctgca gcagtgccct gcctgaggcc 300
 cgtgaaggtg cgaacgtoca cttcccaccg caccoggttc ctgcgagca cttttcctgt 360
 ggcgcaccga aaccogtagc gggggccccg gcgctgaatg caagcttgat ggacggcggc 420
 gcgctgcccga gactcgtgcc cacctcgtct ggagtcgctg gagcctgcae tgctcggcgg 480
 agaccccctg ccccggaact gcttcgctgc agcogaoggc ggcgatcggg agcaaccgag 540
 gccagcagca gctcggcggc cgtggcacgc cgcaatgagc gtgagcggaa ccgcgtaaag 600
 ctggtaaaact tgggcttcca ggcgctgagg cagcacgtgc cgcacggcgg cgccaacaag 660
 aagctgagta aggtggagac gctgcgctcc gcggtagagt acatccgtgc gctgcagcgg 720
 ctgctagcag agcagcagc ggtgcgtgct gcgctctctg ggggtctatt aacaccgct 780
 actcggccgt ccgatgtgtg cagcagccc tcgcctccc ctgccagcgc gtctctgtcc 840
 tgcaccteta catccccaga ccgcctaggc tgctccgagc ctgcctctcc gcgctccgcc 900
 tactcgtcgg aggacagcag ctgcgaggga gagacttacc cgatggggca gatgtttgac 960
 ttttccaatt ggttaggggg ctactgagca ccccacccc ctaagctgog tccctgggtg 1020
 tcccctggtg gacctacctg cgtttcttgc ccaggaaacc tgggcccctg ccttaccat 1080
 gctgtctagt gcagcctgac caaatgccc gtactgacct ctgctcggcc tccacgcgc 1140
 ggaatgacat cttccatctc ccagtccttg ccgaaccagg acttggaat ttctcaggag 1200
 aaagaatttt acaatgacaa tctgcttttt atcaattaac ttgaactgct ggaggactct 1260
 gctgaaaata tgaagaatta tttttataca aaggatcctt aagcttggag cacaataaag 1320
 atgacctctg tctctcacc ccactgtcta gaactttcca acctggccaa agtgtggacg 1380
 ggtcgggccc tgagggcaag atgcctgget gcaccctct tctcttccg aagcctatec 1440
 tgacgctgat gtttgccag tgtgggaacc ctgctattgc aaagtgtact attctataaa 1500
 agttgttttt cattggaaag gaattc 1526

<210> 16
 <211> 10
 <212> PRT
 <213> 人类

<400> 16
 Lys Leu Val Asn Leu Gly Phe Gln Ala Leu
 1 5 10

<210> 17
 <211> 9
 <212> PRT
 <213> 人类

<400> 17
 Glu Leu Leu Asp Phe Ser Ser Trp Leu

1 5

<210> 18

<211> 9

<212> PRT

<213> 人类

<400> 18

Arg Leu Leu Ala Glu His Asp Ala Val

1 5

<210> 19

<211> 9

<212> PRT

<213> 人类

<400> 19

Lys Leu Val Asn Leu Gly Phe Gln Ala

1 5

<210> 20

<211> 9

<212> PRT

<213> 人类

<400> 20

Glu Tyr Ile Arg Ala Leu Gln Arg Leu

1 5

<210> 21

<211> 10

<212> PRT

<213> 人类

<400> 21

Glu Tyr Ile Arg Ala Leu Gln Arg Leu Leu

1 5 10

<210> 22

<211> 10

<212> PRT

<213> 人类

<400> 22

Ala Val Arg Asn Ala Leu Ala Gly Gly Leu
1 5 10

<210> 23

<211> 10

<212> PRT

<213> 人类

<400> 23

Ser Glu Pro Gly Ser Pro Arg Ser Ala Tyr
1 5 10

<210> 24

<211> 10

<212> PRT

<213> 人类

<400> 24

Val Glu Thr Leu Arg Ser Ala Val Glu Tyr
1 5 10

<210> 25

<211> 9

<212> PRT

<213> 人类

<400> 25

Ile Arg Ala Leu Gln Arg Leu Leu Ala
1 5

<210> 26

<211> 9

<212> PRT

<213> 人类

<400> 26

Leu Arg Pro Gln Ala Val Arg Pro Ser
1 5

<210> 27

<211> 9

<212> PRT

<213> 人类

<400> 27

Leu Arg Gln His Val Pro His Gly Gly
1 5

<210> 28

<211> 9

<212> PRT

<213> 人类

<400> 28

Leu Gly Phe Gln Ala Leu Arg Gln His
1 5

<210> 29

<211> 9

<212> PRT

<213> Human

<400> 29

Val Arg Asn Ala Leu Ala Gly Gly Leu
1 5

<210> 30

<211> 9

<212> PRT

<213> 人类

<400> 30

Tyr Ile Arg Ala Leu Gln Arg Leu Leu
1 5

<210> 31

<211> 9

<212> PRT

<213> 人类

<400> 31

Leu Val Asn Leu Gly Phe Gln Ala Leu

1 5

<210> 32

<211> 9

<212> PRT

<213> 人类

<400> 32

Val Glu Tyr Ile Arg Ala Leu Gln Arg

1 5

<210> 33

<211> 9

<212> PRT

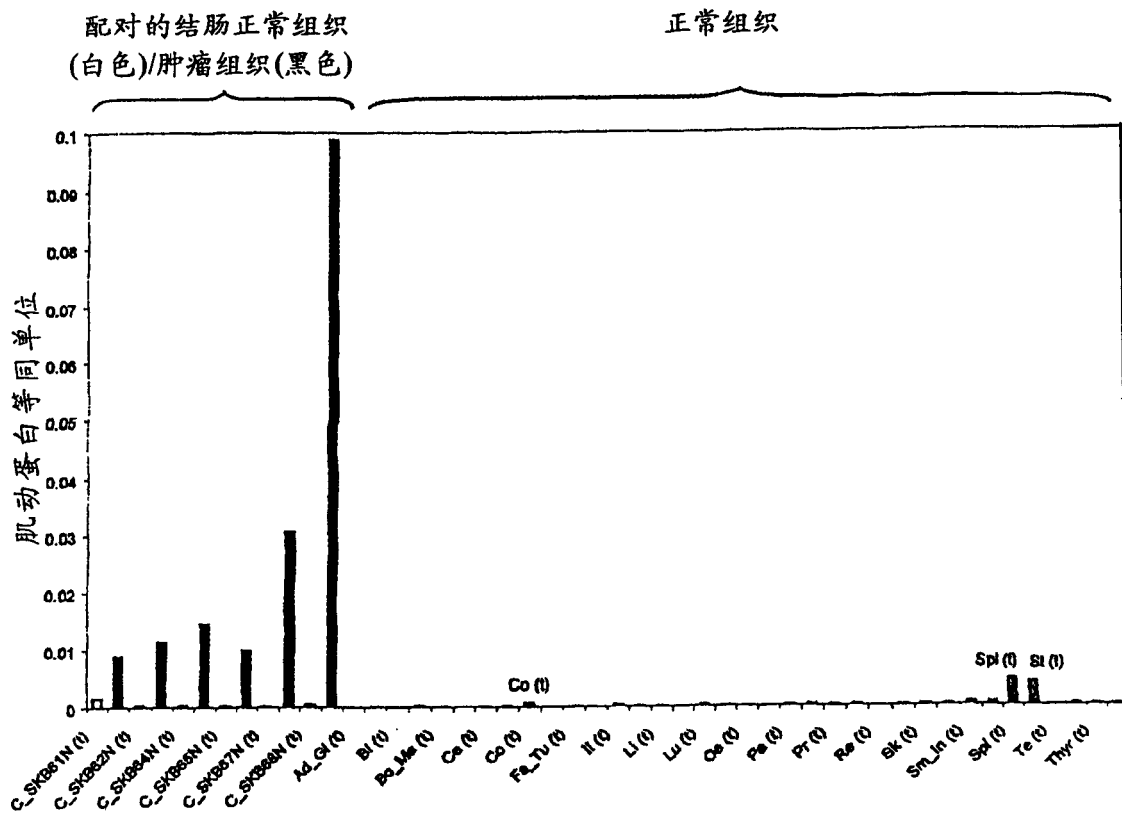
<213> 人类

<400> 33

Leu Leu Arg Cys Ser Arg Arg Arg Arg

1 5

图 1



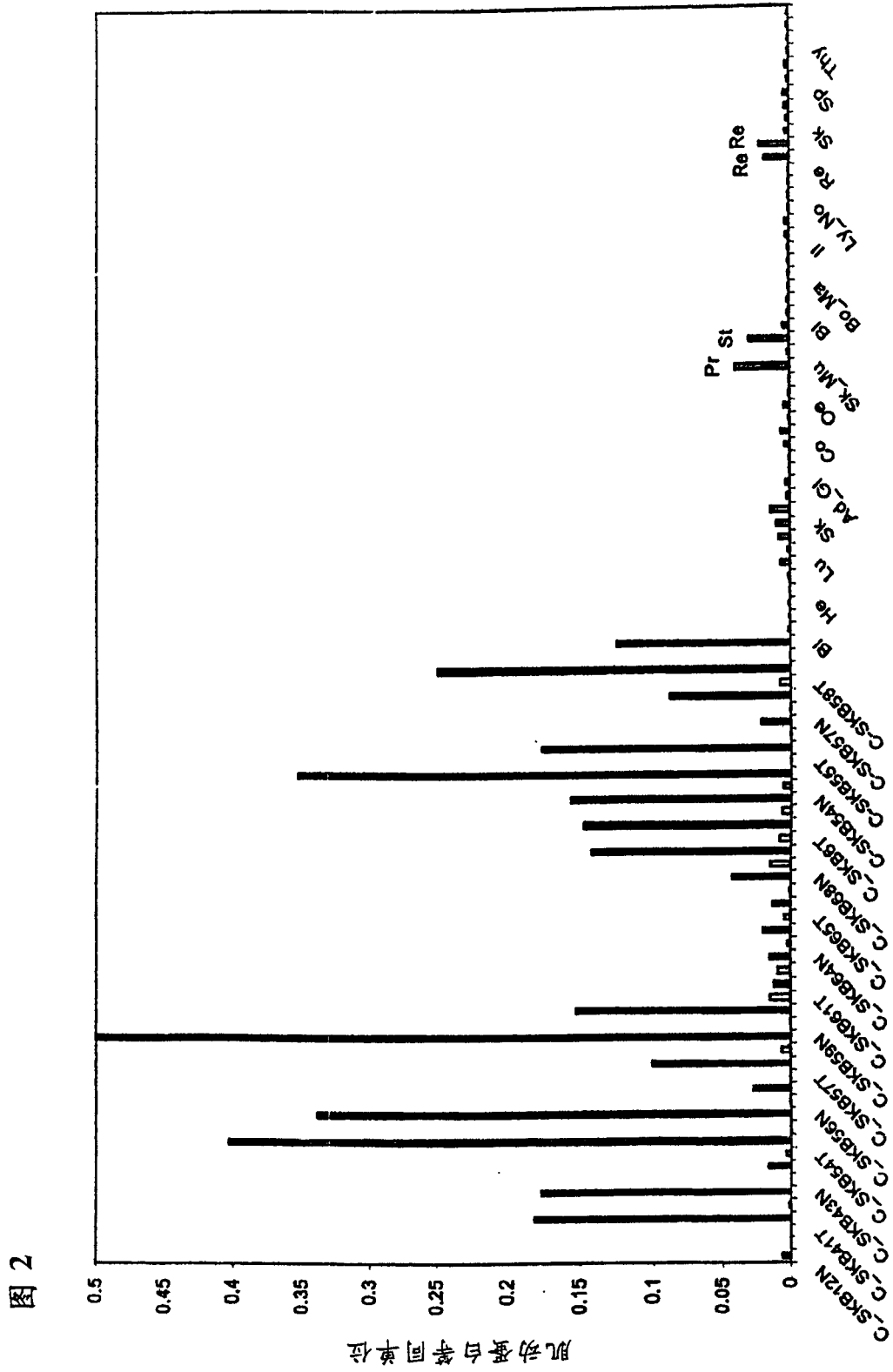


图 3

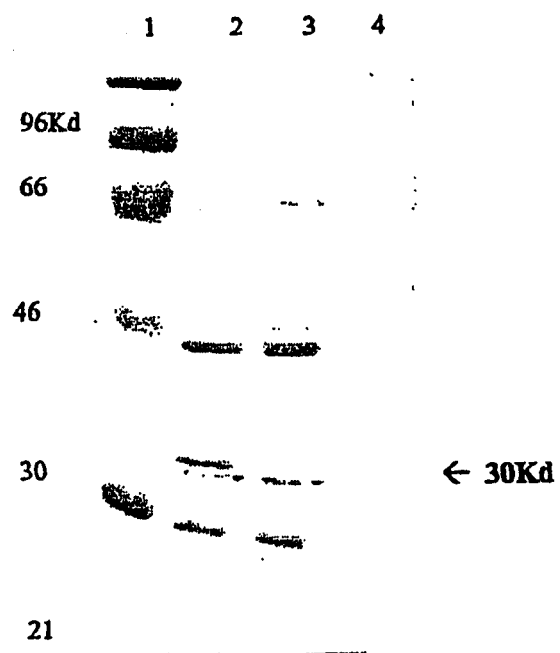


图 4

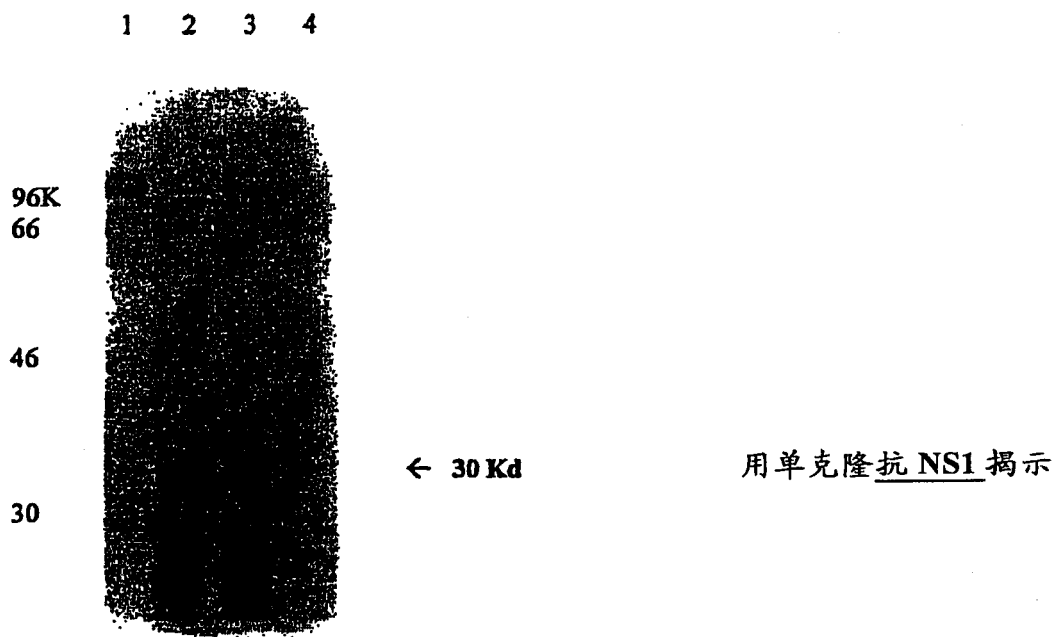


图 5

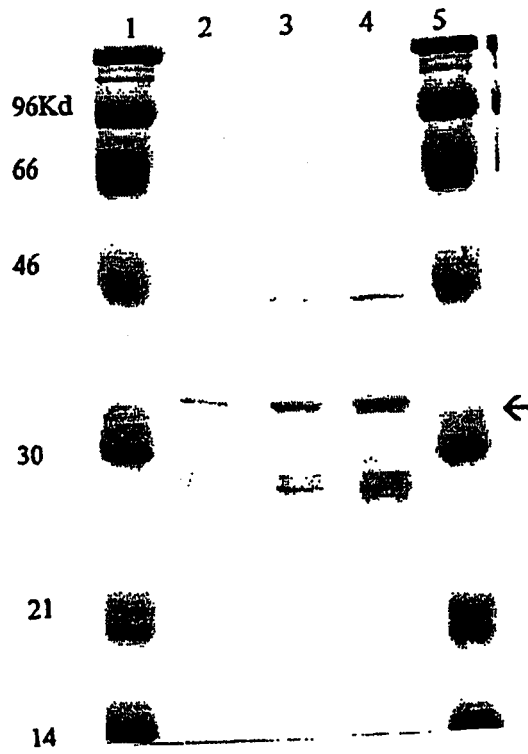
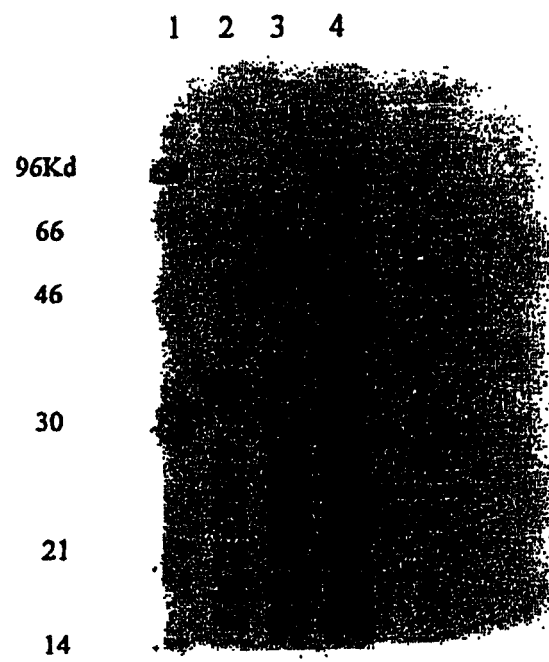


图 6



专利名称(译)	肿瘤特异性动物蛋白		
公开(公告)号	CN1840178A	公开(公告)日	2006-10-04
申请号	CN200610068150.1	申请日	2001-02-16
[标]申请(专利权)人(译)	史密斯克莱恩比彻姆生物有限公司		
申请(专利权)人(译)	史密斯克莱恩比彻姆生物有限公司		
当前申请(专利权)人(译)	史密斯克莱恩比彻姆生物有限公司		
[标]发明人	TEV卡贝宗 斯尔瓦 J P卡萨特 T科彻 SRJ T高利斯 C维纳尔斯 Y 德巴索尔斯		
发明人	T·E·V·卡贝宗 - 斯尔瓦 J· - P·卡萨特 T·科彻 S·R·J· - T·高利斯 C·维纳尔斯 Y 德巴索尔斯		
IPC分类号	A61K38/17 A61K31/711 C07K14/47 C07K17/02 C07K19/00 C07K16/18 C12N15/62 C12N15/12 C12N15/70 C12N15/85 C12N1/21 G01N33/50 G01N33/53 C12Q1/68 A61K C07K C12N C12Q G01N		
代理人(译)	梁谋		
优先权	2000021080 2000-08-25 GB 2000004269 2000-02-23 GB 2000009905 2000-04-20 GB		
其他公开文献	CN1840178B		
外部链接	Espacenet SIPO		

摘要(译)

本发明公开了CASB7439多肽和多核苷酸、包含它们的免疫原性组合物以及通过重组技术生产这样的多肽的方法。也公开了CASB7439多肽和多核苷酸在诊断学上的使用方法以及用于预防性和治疗性治疗癌症，尤其是结肠直肠癌、自身免疫病和相关病症的疫苗。

