



(12) 发明专利

(10) 授权公告号 CN 1639571 B

(45) 授权公告日 2010.05.26

(21) 申请号 03804559.1  
 (22) 申请日 2003.02.26  
 (30) 优先权数据  
 60/360,057 2002.02.26 US  
 (85) PCT申请进入国家阶段日  
 2004.08.25  
 (86) PCT申请的申请数据  
 PCT/US2003/005670 2003.02.26  
 (87) PCT申请的公布数据  
 W02003/073068 EN 2003.09.04  
 (73) 专利权人 金克克国际有限公司  
 地址 美国加利福尼亚州  
 (72) 发明人 F·A·哈丁  
 (74) 专利代理机构 中国国际贸易促进委员会专  
 利商标事务所 11038  
 代理人 程泳

(56) 对比文件  
 EP 0279994 A2, 1988.08.31, 说明书全文.  
 STICKLER M.M..CD4+ T-Cell epitope  
 determination using unexposed human donor  
 peripheral blood mononuclear cells. JOURNAL  
 OF IMMUNOTHERAPY23 6.2000, 23(6), 654-660.  
 MUCHA J.M..Enhanced immunogenicity  
 of a functional enzyme by T cell epitope  
 modification. BMC IMMUNOLOGY3.2002,

审查员 邢维玲

(51) Int. Cl.  
 G01N 33/53 (2006.01)  
 G01N 33/567 (2006.01)  
 C12N 5/06 (2006.01)

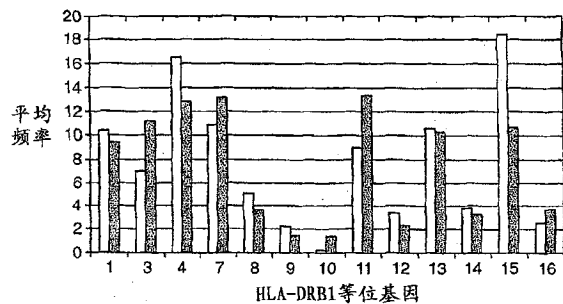
权利要求书 2 页 说明书 36 页 附图 13 页

(54) 发明名称

将蛋白质相对免疫原性进行分级的基于群体的评估方法和手段

(57) 摘要

本发明提供了评估群体的免疫反应特性的方法。特别地,本发明提供了定性地评估人类群体的免疫反应的方法,其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外,本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。



1. 用于将第一个蛋白质和至少一个附加蛋白质的相对免疫原性进行分级的方法,其包括下列步骤:

(a) 从第一个蛋白质中制备第一个肽组和从每个所述附加蛋白质中制备至少一个附加的肽组;

(b) 从单个人血源中获得含有单核细胞的溶液和原初 CD4<sup>+</sup>T- 细胞的溶液;

(c) 分化所述单核细胞以产生经分化的树突细胞的溶液;

(d) 将所述经分化的树突细胞的溶液和所述原初 CD4<sup>+</sup>T- 细胞与所述第一个肽组合并在一起;

(e) 将所述经分化的树突细胞的溶液和所述原初 CD4<sup>+</sup>T- 细胞与来自所述附加蛋白质的所述每个肽组合并在一起;

(f) 测量所述步骤 (d) 和 (e) 中所述 T- 细胞的增殖,以测定对于所述第一个和附加肽组中的每个肽的响应;

(g) 汇集在步骤 (f) 中所述 T- 细胞对于所述第一个蛋白质和所述附加蛋白质的响应;

(h) 使用公式  $\sum \left| f(i) - \frac{1}{p} \right|$  确定步骤 (g) 中所汇集的对于所述第一个蛋白质和所述附加蛋白质的响应的结构值,其中  $f(i)$  是对于单个肽的响应频率,  $p$  是肽组中肽的数量;和

(i) 将对于所述第一个蛋白质所获得的结构值与对于所述附加蛋白质的结构值进行比较,以确定所述第一个蛋白质和所述附加蛋白质的免疫原性分级。

2. 权利要求 1 的方法,其中将有最小结构值的蛋白质分级为比有更高结构值的蛋白质具有更弱的免疫原性。

3. 权利要求 1 的方法,所述方法是用于将两个蛋白质的相对免疫原性进行分级的方法,其中所述第二个附加蛋白质是所述第一个蛋白质的一种蛋白质变体,且与所述第一个蛋白质的区别在于一个或多个氨基酸。

4. 权利要求 3 的方法,其中将所述第二个蛋白质分级为比所述第一个蛋白质具有更弱的免疫原性。

5. 权利要求 3 的方法,其中将所述第一个蛋白质分级为比所述第二个蛋白质具有更弱的免疫原性。

6. 权利要求 3 的方法,其中所述第二个蛋白质比所述第一个蛋白质少含至少一个显著区域。

7. 权利要求 3 的方法,其中在步骤 (e) 中所述 T- 细胞的增殖处于背景水平。

8. 权利要求 7 的方法,所述方法用于将第一个蛋白质和至少一个变体蛋白质的相对免疫原性进行分级。

9. 权利要求 8 的方法,其中将至少一个所述变体蛋白质分级为比所述第一个蛋白质具有更弱的免疫原性。

10. 权利要求 8 的方法,其中将所述第一个蛋白质分级为比至少一个所述变体蛋白质具有更弱的免疫原性。

11. 权利要求 8 的方法,其中至少一个所述变体蛋白质比所述第一个蛋白质少含至少一个显著区域。

12. 权利要求 11 的方法,其中在步骤 (e) 中所述 T- 细胞对于至少一个变体蛋白质的增

殖处于背景水平。

13. 权利要求 1、3 或 8 的方法,其中所述蛋白质选自酶、激素、细胞因子、抗体、结构蛋白和结合蛋白。

14. 权利要求 1、3 或 8 的方法,其中抗所述第一个蛋白质的阳性响应具有 2.7-3.2 的刺激系数值。

15. 权利要求 1、3 或 8 的方法,其中抗至少一个所述变体蛋白质的阳性响应具有 2.7-3.2 的刺激系数值。

16. 权利要求 1、3 或 8 的方法,其中在步骤 (d) 中所述 T-细胞的所述增殖导致 2.95 或更高的刺激系数。

17. 权利要求 1、3 或 8 的方法,其中在步骤 (e) 中所述 T-细胞的所述增殖导致 2.95 或更高的刺激系数。

18. 权利要求 1、3 或 8 的方法,其中在步骤 (b) 中使用至少一个附加的人血源。

19. 权利要求 1、3 或 8 的方法,其中将对于每个所述人血源和所述蛋白质而获得的结构值进行比较。

20. 权利要求 19 的方法,其中将所述蛋白质的结构值和背景响应百分数用于分级所述蛋白质。

## 将蛋白质相对免疫原性进行分级的基于群体的评估方法和手段

### 发明领域

[0001] 本发明提供了评估群体的免疫反应特性的方法。特别地,本发明提供了定性地评估人类群体的免疫反应的方法,其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外,本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。

### [0002] 发明背景

[0003] 蛋白质具有潜在地引起威胁生命的免疫反应的能力。这一限制阻碍了蛋白质在消费者终端应用和产品中的广泛使用。实际上,这一引起免疫反应的潜在性已经引起了美国食品和药品管理局(FDA)的注意,结果导致在批准新的蛋白质治疗剂之前和之后都要求免疫原性测试。可是,虽然有大量的可用于评估免疫原性的动物模型,但是没有有效的在人中辨识相对免疫原性的方法。

[0004] 尽管有这些担忧,蛋白质的免疫原性在很长时间内已经是在酶生产工业中所关心的事情。已经证明,职业性的蛋白质接触可导致工业和实验室工作者的致敏。对于特定蛋白质的致敏通常通过测试来进行评估,例如皮肤针刺测试(skin-prick test),其可揭示出个体是否形成了对于该蛋白质的免疫反应。

[0005] 实际上,已经证明职业性的蛋白质接触可导致工业和实验室工作者的致敏。在大多数环境中,致敏可通过减少空气传播的蛋白质的水平来加以控制(见 Sarlo 和 Kirchner, *Curr. Opin. Allergy Clin. Immunol.*, 2 :97-101[2002]; 和 Schweigert 等人, *Clin. Exp. Allergy*30 :1511-1518[2000])。已经实施了职业暴露指南来控制空气传播的与蛋白质的接触。这些指南提供了与特定蛋白质接触的允许水平,其对于减少在指定的工业环境中发生的致敏事件的总数量是有用的。当要生产一种新的蛋白质时,对于该新的蛋白质建立职业暴露指南(OEG)是需要认真关注的问题。一般接受的用于确定这些指南的方法为豚鼠气管内测试法(GPIT)(见 Sarlo, *Fundam. Appl. Toxicol.*, 39 :44-52[1997])。在该测试中,通过气管内滴注大约 10-12 星期来将豚鼠与测试蛋白质进行接触。定时性地从这些动物取得血清样品,并通过本领域中已知的合适方法来测试其抗原特异性抗体的水平(例如,对于 IgG<sub>1</sub> 的被动皮肤测试(PCA),和对于沉淀 IgG 的微免疫扩散测试(MID))。将这些结果与从一组用对照蛋白质(例如 **ALCALASE**<sup>®</sup> 酶,从 Novo 商购可得)进行测试的豚鼠中获得的结果进行比较,该对照蛋白质具有已知和有效的暴露指南。考虑血清滴度的测定、MID 阳性和反应的时间,从而确定相对效价值。这一方法已经成功地用于设置对于许多工业酶的 OEG。

[0006] 可是,虽然 GPIT 测试是有用的,但是其消耗时间而且价格昂贵,还需要许多动物和多重测试周期。相对近来的一段时间,建立了基于小鼠的测试法,据报道这种方法可再现 GPIT 中获得的结果,其使用更便宜的和麻烦更少的动物模型。一些公司使用小鼠鼻内测试法(MINT;见 Robinson 等人, *Toxicol. Sci.* 43 :39-46[1998])来制定 OEG 指南。可是,还没有获得对于该模型的工业范围的认可(关于蛋白质过敏原性的预报性测试的综述,见 Robinson 等人,如上,以及 Kimber 等人, *Fundam. Appl. Toxicol.*, 33 :1-10[1996];和

Kimber 等人, *Toxicol. Sci.*, 48 :157-162[1999])。

[0007] 因此, 尽管动物模型是有用的, 但是它们是有限制的。在 GPIT 中使用部分远交的豚鼠使得需要使用大量的动物, 以便当比较小组之间的反应时获得统计上的显著性。另外, 对照动物反应的实验间变化是非常大的, 这使得基于单组对照反应的效价测定的可信性降低了。MINT 检测不会受到抗体反应中这么多变化的影响, 因为所使用的小鼠通常为 BDF1 小鼠, 其是两个高度近交小鼠品系之间的杂交品种。虽然这一附加的控制水平允许更有力的数据分析, 但是不同品系的小鼠对于相似的酶通常会得出非常不同的效价分级 (见 Blaikie, *Food Chem. Toxicol.*, 37 :897-904[1999]; 和 Blaikie 和 Basketter, *Food Chem. Toxicol.*, 37 :889-896[1999])。这很可能是由于小鼠品系中的免疫反应的特异性造成的, 这种小鼠品系经过近交而表达非常有限的 MHC 分子。另外, 虽然来自使用 MINT 检测的单个实验室的数据可能是有力的, 但是 MINT 检测也受到实验室间差异的困扰。

[0008] 值得注意的是, 所有的动物测试都没有能力合适地代表在人中对于指定蛋白质的免疫反应。小鼠的近交系呈现出具有其鼠科动物 MHC 分子所赋予的特异性的肽分子。人 HLA 分子虽然与小鼠 MHC 分子高度相关, 但是不具有相同的肽特异性。此外, 已经选择出能够表达单 I-A 和 / 或 I-E 分子的近交小鼠品系, 但是这种情况在高度远交的人类群体中是极少出现的。另外, 小鼠免疫系统具有许多未在人中发现的特性 (例如已经在小鼠中进行了描述的 Th1 对 Th2 的范例 (paradigm) 在人中却不清楚得多)。例如, 在人中, Th1 和 Th2 表型中具有可塑性, 其可以用 IFN- $\alpha$  基因中的基因不一致性来解释。与之相反, 在小鼠中 Th1 和 Th2 表型不是动态的, 这是由于在这些动物的 IFN- $\alpha$  基因中具有插入物 (见 Farrar, *Nat. Immunol.*, 1 :65-69[2000])。另外, 人在激活的 T- 细胞上表达 II 型 HLA 分子, 而小鼠则不表达。此外, 人供体通常带有内源性的病毒并常常具有亚临床感染, 而实验室小鼠通常保持在无特异性病原体 (SPF) 的环境中。另一个有关的事是, C57BI/6 小鼠品系 (形成转基因小鼠模型的常用背景材料) 具有确定的抗原加工缺陷, 这使得与人来源的数据进行比较的可信度存在疑问 (Kim 和 Jang, *Eur. J. Immunol.*, 22 :775-782[1992])。人 HLA 转基因小鼠已经可用于人免疫反应的机制研究 (见 Boyton 和 Altmann, *Clin. Exp. Immunol.*, 127 :4-11[2002]; Black 等人, *J. Immunol.*, 169 :5595-5600[2002]; Raju 等人, *Hum. Immunol.*, 63 :237-247[2002]; 和 Das 等人, *Rev. Immunogenet.*, 2 :105-114[2000])。可是, 这些动物的使用是有限制的, 因为 HLA 转基因小鼠受到种特异性免疫系统复合物的影响。另外, 由于 HLA 转基因的表达没有进行正确的调节, 至少一些用于构建这些小鼠的方法不能允许准确地分析肽特异性的反应。当表达单 HLA 分子时, HLA 转基因小鼠常常用于作图研究, 这一情况没有在人中发现。这一点对于 HLA-DQ 转基因小鼠来说是特别要注意的地方, 在这些转基因小鼠中显示不同 HLA-DQ 等位基因之间的杂交配对以形成新的肽呈递特异性 (见 Krcso 等人, *J. Immunol.*, 163 :1661-1665[1999])。因此, 尽管在蛋白质免疫原性的测定、评估和比较中取得了进展, 但是在本领域中仍需要简单的、可靠的和可重复的方法来进行这样的测定。

[0009] 同样地, 蛋白质在治疗、工业和营养应用上的使用受到引起或加剧有害免疫反应的潜在性的限制。这种潜在性对于重组人源蛋白质的使用来说是特别要关注的问题。实际上, 已经证明重组人源蛋白质可引起针对自身蛋白质的免疫反应, 结果导致自体免疫的形成 (Li 等人, *Blood* 98 :3241-3248[2001]; 和 Casadell 等人, *N. Eng. J. Med.*, 346 :469-475[2002])。在无意地引起对于工业或食物蛋白质的免疫反应之后所发生的免疫系

统的再次激活可以通过回避而最小化。可是,这不适合于使用人源治疗蛋白质的情况。因此,选择和 / 或创造免疫原性减弱的蛋白质变体对于改善所施用的蛋白质的安全性和有效性来说是必需的。挑选天然存在的免疫原性减退的蛋白质异构体是一种选择,其中存在几种具有相似活性的相关分子。不幸的是,这不适合于许多治疗蛋白质。因此,在本领域中有长期存在的对于生产免疫原性减退的蛋白质的方法的需要,这些蛋白质适合于作为治疗药物使用和用于其他应用。

#### [0010] 发明概述

[0011] 本发明提供了评估群体的免疫反应特性的方法。特别地,本发明提供了定性地评估人类群体的免疫反应的方法,其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外,本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。

[0012] 形成本发明是为了避免在除了人之外的动物的免疫原性分析中出现的问题。在本发明的优选实施方案中,提供了通过使用人外周血单核细胞 (PBMC) 作为测试“对象”来将蛋白质的免疫原性进行分级的方法。因为使用了大量人类样品,所以获得的信息可应用于一般的人类群体。重要的是,该数据不受在使用近交小鼠中出现的特异性问题的影响。在优选的实施方案中,本发明提供了基于蛋白质的总免疫原性来将蛋白质进行分级的方法。另外,通过将数据与先前存在的动物数据进行比较,本发明的方法提供了与蛋白质的相对效价有关的信息。例如,在本发明的形成期间,将四个经很好表征的工业过敏原以通过 GPIT 和 MINT 测试而确定的顺序进行放置,然后与用本发明的方法获得的结果进行比较,包括确定职业性暴露的工作者的致敏作用。

[0013] 在优选的实施方案中,本发明提供的方法包括使用作为抗原呈递细胞的树突细胞、偏移 (offset) 3 个氨基酸从而包含完整的目标蛋白质序列的 15 聚肽和从树突细胞供体获得的 CD4<sup>+</sup>T- 细胞。让 T- 细胞在肽 (每个肽分别进行测试) 和分化的树突细胞存在的情况下在样品中进行增殖。至于肽组 (pepset) 的制备和树突细胞的分化,不是想要将任何本发明的方法以任何特定的顺序来进行。例如,在一些实施方案中于树突细胞分化之前制备肽组,而在其他实施方案中于制备肽组之前进行树突细胞的分化,和在另外其他实施方案中树突细胞的分化和肽组的制备同时进行。因此,不是想要本发明限制于具有这些以任何特定顺序排列的步骤的方法。

[0014] 如果在对于肽的响应中的增殖导致 1.5-4.5 的刺激系数 (SI),则认为和记录该响应为“阳性”。对于一个供体组将每个肽的结果列成表格,虽然具有一些变化,但是其优选地反映出群体的一般 HLA 等位基因频率。测定基于确定离开直线性的差异而得到的“结构值”(structure value),该值用于将蛋白质的相对免疫原性进行分级。因此,本发明提供了在蛋白质的修饰中有用的信息,从而获得预测在人中为有效的减少了的响应率,而不需致敏志愿者。然后进行对于肽组的供体响应的分析,所述分析基于这些经设计而免疫原性减退的新蛋白质,从而计算出新蛋白质的结构值和确定它们的免疫原性与暴露潜在性。

[0015] 在一些优选的实施方案中,本发明提供了用于将蛋白质的相对免疫原性进行分级的检测系统 (即 **I-MUNE**<sup>®</sup>测)。在一个实施方案中,该方法包括在体外测量 CD4<sup>+</sup>T- 细胞在对于蛋白质的肽片段的响应中的增殖,汇集所测量的对于蛋白质的响应,确定所汇集的响应的结构值,和比较该蛋白质的结构值与第二个蛋白质的结构值,其中将具有最小结

构值的蛋白质分级为比具有更高结构值的蛋白质对于人类有着更弱的免疫原性。在另一个可选择的实施方案中,被测试的蛋白质是酶。在更进一步的实施方案中,酶为蛋白酶。在一个另外的实施方案中,被测试的蛋白质选自抗体、细胞因子和激素。在进一步的实施方案中,在并行的测试中测定每个肽片段和每个蛋白质的 T- 细胞增殖。在其他实施方案中,基于 2.7-3.2 的 SI 值来确定“阳性”响应。在特别优选的实施方案中,增殖的水平导致 2.95 或者更高的刺激系数。

[0016] 本发明也提供了方法来评估在人中变体蛋白质的减弱的免疫原性能力。在一些实施方案中,该方法包括减少一个或多个亲本蛋白质的显著区域 (prominent region) 至背景水平从而形成变体蛋白质,测定变体的结构值,和比较变体的结构值和亲本蛋白质的结构值,其中较低的结构值表明蛋白质具有减弱的免疫原性。在一些优选的实施方案中,蛋白质为酶。在一些可选择的实施方案中,蛋白质选自蛋白酶、细胞因子、激素、抗体、淀粉酶和其他酶,其他酶包括枯草杆菌蛋白酶、**ALCALASE**<sup>®</sup>酶、纤维素酶、脂肪酶、氧化酶、异构酶、激酶、磷酸酶、内酰胺酶和还原酶,但不局限于此。在进一步的实施方案中,减少至背景水平的显著区域的数目为 1-10,优选为 1-5。在另一个实施方案中,将在亲本蛋白质的显著区域中的一个或多个氨基酸残基加以改变以形成变体。

[0017] 本发明也提供了从一组相关蛋白质中选择出具有最弱免疫原性的蛋白质的方法。在一个实施方案中相关蛋白质为抗体,而在一个可选择的实施方案中它们为细胞因子,和在另一个实施方案中它们为激素。在进一步的实施方案中,相关蛋白质为结构蛋白。在另一个实施方案中,蛋白质为酶。在一些优选的实施方案中,酶选自蛋白酶、纤维素酶、脂肪酶、淀粉酶、氧化酶、异构酶、激酶、磷酸酶、内酰胺酶和还原酶。

[0018] 本发明进一步提供了通过使用相关蛋白质的相对分级来确定适合于(特别是在人中)减弱蛋白质免疫原性的 T- 细胞表位修饰的方法。本发明也提供了将蛋白质进行分类的方法,其是基于它们的背景响应百分数和它们的结构值来进行的。因此,在一些进一步的实施方案中,将经分析的蛋白质根据它们的背景响应百分数和结构值进行分类和/或分级。

[0019] 在一些实施方案中,本发明提供了用于将第一个蛋白质和至少一个附加蛋白质的相对免疫原性进行分级的方法,其包括下列步骤:(a) 从第一个蛋白质中制备第一个肽组和从每个附加蛋白质中制备至少一个附加的肽组,其中每个肽组;(b) 从单个人血源中获得含有树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞的溶液;(c) 分化树突细胞以产生经分化的树突细胞的溶液;(d) 将经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞与第一个肽组结合在一起;(e) 将经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞与来自附加蛋白质的每个肽组结合在一起;(f) 测量步骤 (d) 和 (e) 中 T- 细胞的增殖,以测定对于第一个和附加肽组中的每个肽的响应;(g) 汇集在步骤 (f) 中 T- 细胞对于第一个蛋白质和附加蛋白质的响应;(h) 确定步骤 (g) 中所汇集的对于第一个蛋白质和附加蛋白质的响应的结构值;和 (i) 将对于第一个蛋白质所获得的结构值与对于附加蛋白质的结构值进行比较,以确定第一个蛋白质和附加蛋白质的免疫原性分级。在一些优选的实施方案中肽组包含有长度为大约 15 个氨基酸的肽,而在一些特别优选的实施方案中每个肽与邻近肽重叠大约 3 个氨基酸。可是,这不是想要将肽组中的肽限制为任何特定的长度或者重叠,因为其他的肽长度和重叠数量在本发明中是有用的。

[0020] 在一些实施方案中,将有最小结构值的蛋白质分级为比有更高结构值的蛋白质具有更弱的免疫原性。在另外的实施方案中,至少两种蛋白质选自酶、激素、细胞因子、抗体、结构蛋白和结合蛋白。在更进一步的实施方案中,抗第一个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值。在另一些其他的实施方案中,抗附加蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值。在进一步的实施方案中,抗第一个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值,抗附加蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值。在一些实施方案中,步骤 (d) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数,而在另外的实施方案中,步骤 (e) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数。在更进一步的实施方案中,步骤 (d) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数,步骤 (e) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数。在一些特别优选的实施方案中,在步骤 (b) 中使用至少一个附加的人血源。在一些附加的特别优选的实施方案中,比较对于每一个人血源和蛋白质获得的结构值。本发明也提供了将蛋白质进行分类的方法,其是基于它们的背景响应百分数和它们的结构值来进行的。因此,在一些进一步的实施方案中,将经分析的蛋白质根据它们的背景响应百分数和结构值进行分类和 / 或分级。

[0021] 本发明也提供了用于将两个蛋白质的相对免疫原性进行分级的方法,其中第二个蛋白质是第一个蛋白质的一种蛋白质变体,其包括下列步骤:(a) 从第一个蛋白质中制备第一个肽组和从第二个蛋白质中制备第二个肽组;(b) 从单个人血源中获得含有树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞的溶液;(c) 分化树突细胞以产生经分化的树突细胞的溶液;(d) 将经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞与第一个肽组合并在一起;(e) 将经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞与第二个肽组合并在一起;(f) 测量步骤 (d) 和 (e) 中 T- 细胞的增殖,以测定对于第一个和第二个肽组中的每个肽的响应;(g) 汇集在步骤 (f) 中 T- 细胞对于第一个蛋白质和第二个蛋白质的响应;(h) 确定步骤 (g) 中所汇集的对于第一个蛋白质和第二个蛋白质的响应的结构值;(i) 将对于第一个蛋白质所获得的结构值与对于第二个蛋白质的结构值进行比较,以确定第一个蛋白质和第二个蛋白质的免疫原性分级。在一些实施方案中,将第二个蛋白质分级为比第一个蛋白质具有更弱的免疫原性,而在可选择的实施方案中,将第一个蛋白质分级为比第二个蛋白质具有更弱的免疫原性。在一些优选的实施方案中肽组包含有长度为大约 15 个氨基酸的肽,而在一些特别优选的实施方案中每个肽与邻近肽重叠大约 3 个氨基酸。可是,这不是想要将肽组中的肽限制为任何特定的长度或者重叠,因为其他的肽长度和重叠数量在本发明中是有用的。在另外的实施方案中,第一个和第二个蛋白质选自酶、激素、细胞因子、抗体、结构蛋白和结合蛋白。在更进一步的实施方案中,抗第一个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值,而在其他实施方案中,抗第二个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值。在另外的实施方案中,抗第一个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值,抗第二个蛋白质的阳性响应具有大约 2.7- 大约 3.2 的刺激系数值。在更进一步的实施方案中,步骤 (d) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数,步骤 (e) 中 T- 细胞的增殖导致大约 2.95 或者更高的刺激系数。在一些特别优选的实施方案中,在步骤 (b) 中使用至少一个附加的人血源。在一些附加的特别优选的实施方案中,比较对于每一个人血源和蛋白质获得的结构值。在一些实施方案中,第二个蛋白

质减少了至少一个第一个蛋白质中的显著区域。在进一步的实施方案中,步骤(e)中T-细胞的增殖为背景水平。在一些特别优选的实施方案中,比较对于每一个人血源和蛋白质获得的结构值。本发明也提供了将蛋白质进行分类的方法,其是基于它们的背景响应百分数和它们的结构值来进行的。因此,在一些进一步的实施方案中,将经分析的蛋白质根据它们的背景响应百分数和结构值进行分类和/或分级。

[0022] 本发明也提供了用于将第一个蛋白质和至少一个变体蛋白质的相对免疫原性进行分级的方法,其包括下列步骤:(a)从第一个蛋白质中制备第一个肽组和从每一个变体蛋白质中制备肽组;(b)从单个人血源中获得含有树突细胞的溶液和原初CD4<sup>+</sup>和/或CD8<sup>+</sup>T-细胞的溶液;(c)分化树突细胞以产生经分化的树突细胞的溶液;(d)将经分化的树突细胞的溶液和原初CD4<sup>+</sup>和/或CD8<sup>+</sup>T-细胞与第一个肽组合并在一起;(e)将经分化的树突细胞的溶液和原初CD4<sup>+</sup>和/或CD8<sup>+</sup>T-细胞与从每个变体蛋白质中制备得到的每个肽组合并在一起;(f)测量步骤(d)和(e)中T-细胞的增殖,以测定对于第一个和第二个肽组中的每个肽的响应;(g)汇集在步骤(f)中T-细胞对于第一个蛋白质和变体蛋白质的响应;(h)确定步骤(g)中所汇集的对于第一个蛋白质和变体蛋白质的响应的结构值;和(i)将对于第一个蛋白质所获得的结构值与对于变体蛋白质的结构值进行比较,以确定第一个蛋白质和变体蛋白质的免疫原性分级。在一些优选的实施方案中,肽组包含有长度为大约15个氨基酸的肽,而在一些特别优选的实施方案中,每个肽与邻近肽重叠大约3个氨基酸。可是,这不是想要将肽组中的肽限制为任何特定的长度或者重叠,因为其他的肽长度和重叠数量在本发明中是有用的。在一些优选的实施方案中,将至少一种变体蛋白质分级为比第一个蛋白质具有更弱的免疫原性,而在其他实施方案中,将第一个蛋白质分级为比至少一种变体蛋白质具有更弱的免疫原性。在另外的实施方案中,第一个和变体蛋白质选自酶、激素、细胞因子、抗体、结构蛋白和结合蛋白。在进一步的实施方案中,抗第一个蛋白质的阳性响应具有大约2.7-大约3.2的刺激系数值,而在其他实施方案中,抗变体蛋白质的阳性响应具有大约2.7-大约3.2的刺激系数值。在另外的实施方案中,抗第一个蛋白质的阳性响应具有大约2.7-大约3.2的刺激系数值,抗变体蛋白质的阳性响应具有大约2.7-大约3.2的刺激系数值。在更进一步的实施方案中,步骤(d)中T-细胞的增殖导致大约2.95或者更高的刺激系数,步骤(e)中T-细胞的增殖导致大约2.95或者更高的刺激系数。在一些特别优选的实施方案中,在步骤(b)中使用至少一个附加的人血源。在一些附加的特别优选的实施方案中,比较对于每一个人血源和蛋白质获得的结构值。在一些实施方案中,变体蛋白质减少了至少一个第一个蛋白质中的显著区域。在进一步的实施方案中,步骤(e)中T-细胞的增殖为背景水平。在一些优选的实施方案中,步骤(e)中T-细胞对于至少一种变体蛋白质的增殖为背景水平。在一些特别优选的实施方案中,比较对于每一个人血源和蛋白质获得的结构值。在进一步的实施方案中,在步骤(b)中使用至少一个附加的人血源。本发明也提供了将蛋白质进行分类的方法,其是基于它们的背景响应百分数和它们的结构值来进行的。因此,在一些进一步的实施方案中,将经分析的蛋白质根据它们的背景响应百分数和结构值进行分类和/或分级。

[0023] 本发明进一步提供了用于测定测试群体抗测试蛋白质的免疫反应的方法,其包括下列步骤:(a)从测试蛋白质中制备肽组;(b)获得多个含有人树突细胞的溶液和原初人CD4<sup>+</sup>和/或CD8<sup>+</sup>T-细胞的多个溶液,其中人树突细胞的溶液和原初人CD4<sup>+</sup>和/或CD8<sup>+</sup>T-

胞的溶液从测试群体中多个个体之中获得；(c) 分化树突细胞以产生多个含有经分化的树突细胞的溶液；(d) 将多个经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞的溶液与肽组合并在一起，其中将来自测试群体中的单个个体的每一个经分化的树突细胞的溶液和原初 CD4<sup>+</sup> 和 / 或 CD8<sup>+</sup>T- 细胞的溶液合并在一起；(e) 测量步骤 (d) 中 T- 细胞的增殖，以测定对于肽组中每个肽的响应；(g) 汇集在步骤 (e) 中 T- 细胞对于测试蛋白质的响应；(h) 确定步骤 (g) 中所汇集的对于测试蛋白质的响应的结构值；和 (i) 确定多个个体对于测试蛋白质的暴露水平。在一些优选的实施方案中肽组包含有长度为大约 15 个氨基酸的肽，而在一些特别优选的实施方案中每个肽与邻近肽重叠大约 3 个氨基酸。可是，这不是想要将肽组中的肽限制为任何特定的长度或者重叠，因为其他的肽长度和重叠数量在本发明中是有用的。在一些实施方案中，测试至少两种测试蛋白质。在一些优选的实施方案中，比较多个体对于测试蛋白质的暴露水平。在一些特别优选的实施方案中，将测试蛋白质进行修饰以产生在测试群体中表现出减弱的免疫反应的变体蛋白质。本发明也提供了将蛋白质进行分类的方法，其是基于它们的背景响应百分数和它们的结构值来进行的。因此，在一些进一步的实施方案中，将经分析的蛋白质根据它们的背景响应百分数和结构值进行分类和 / 或分级。

[0024] 附图简述

[0025] 图 1 显示了，与发表的“高加索人”HLA-DRB1 群体相比较，对于社区供体群体中 184 个随机个体的 HLA-DRB1 等位基因的平均频率。

[0026] 图 2 显示了具有 82 个随机个体的群体中的响应者百分数，这 82 个随机个体用来自地衣形芽孢杆菌 (*Bacillus licheniformis*)  $\alpha$ -淀粉酶的肽进行测试。偏移 3 个氨基酸的连续的 15 聚肽列于 x 轴，对于每一个肽有响应的供体的百分数显示于 y 轴。

[0027] 图 3 显示了具有 65 个随机个体的群体中的响应者百分数，这 65 个随机个体用来自迟缓芽孢杆菌 (*Bacillus lentus*) 枯草杆菌蛋白酶的肽进行测试。偏移 3 个氨基酸的连续的 15 聚肽列于 x 轴，对于每一个肽有响应的供体的百分数显示于 y 轴。

[0028] 图 4 显示了具有 113 个个体的群体中的响应者百分数，这 113 个个体用来自芽孢杆菌 (*Bacillus*) BPN' 枯草杆菌蛋白酶 Y217L 的两个肽组进行测试。偏移 3 个氨基酸的连续的 15 聚肽列于 x 轴，对于每一个肽有响应的供体的百分数显示于 y 轴。

[0029] 图 5 显示了具有 92 个个体的群体中的响应者百分数，这 92 个个体用来自 **ALCALASE**<sup>®</sup>酶的肽进行测试。偏移 3 个氨基酸的连续的 15 聚肽列于 x 轴，对于每一个肽有响应的供体的百分数显示于 y 轴。

[0030] 图 6 提供的图显示出计算所得的结构值随着增加每个肽的响应数目而减少。所显示的结构值为那些当积聚响应时对于  $\alpha$ -淀粉酶 (正方形) 和 BPN' Y217L (菱形) 所测定的值。

[0031] 图 7 中，板块 A 和 B 提供了对于四种工业酶的 GPIT (板块 A) 和 MINT (板块 B) 分级数据与结构系数值之间的比较。将在基于豚鼠 (GPIT) 和小鼠 (MINT) 的检测中测定的  $\alpha$ -淀粉酶、**ALCALASE**<sup>®</sup>酶、BPN' Y217L 和迟缓芽孢杆菌枯草杆菌蛋白酶的相对过敏原性与结构系数值 (y 轴) 进行比较。

[0032] 图 8 提供了显示出有限的数据组的图，该数据组表明用于计算 BPN' Y217L 变体的结构的变体肽响应。48 个社区供体用来自 BPN' Y217L 的序列的肽进行测试。偏移 3 个氨

氨基酸的连续的 15 聚肽列于 x 轴,对于每一个肽有响应的供体的百分数显示于 y 轴。最后两个肽代表了第 24 和 37 号肽的变体序列。

[0033] 图 9 提供的图显示出来自 30 个社区供体的 PBMC 对于 BPN' Y217L(空心三角形,结构值 = 0.53) 和未修饰的 BPN' Y217L 变体(实心正方形,结构值 = 0.40) 的最大增殖响应。每个供体的最大响应显示于 y 轴。2.0 的 SI 是对于“阳性”反应的阈值 (cut-off)。BPN' Y217L 和变体之间增殖响应的差异为  $p < 0.01$ 。

[0034] 图 10 提供的图显示出对于 25 种如实施例 5 中所述进行测试的蛋白质的相对结构值和响应背景百分数。

[0035] 图 11 提供的图显示出测试供体对于 11 种测试蛋白质中每种蛋白质的每个肽的平均响应百分数。

[0036] 图 12 提供的图显示出对于迟缓芽孢杆菌枯草杆菌蛋白酶的响应频率 ( $n = 65$  个社区供体)。板块 A 显示了对于描述枯草杆菌蛋白酶的线性肽的响应百分数。连续的肽显示于 x 轴。65 个供体中的响应百分数位于 y 轴。板块 B 显示了组内的响应频率。显示了对于在迟缓芽孢杆菌肽组中的肽的响应频率。

[0037] 图 13 提供的图显示出 7 个 SPT<sup>+</sup>(皮肤针刺测试阳性) 的供体对于迟缓芽孢杆菌肽的响应。将来自 7 个通过皮肤针刺测试而证实对于迟缓芽孢杆菌枯草杆菌蛋白酶致敏的供体的 PBMC 用于本发明的 I-MUNE<sup>®</sup> 检测之中,以测试它们对于迟缓芽孢杆菌枯草杆菌蛋白酶肽的响应。如果观察到 2.95 或者更高的 SI,则认为对于肽的响应是阳性的。对每一个肽有响应的供体的数目显示于 y 轴。连续的迟缓芽孢杆菌肽显示于 x 轴。

[0038] 图 14 提供了显示对于链激酶的 I-MUNE<sup>®</sup> 检测数据结果的图。板块 A 提供了每个肽 ( $n = 72$ ) 的响应者百分数。连续的链激酶肽显示于 x 轴。在数量为 72 的供体组中的响应者百分数显示于 y 轴。板块 B 显示了每个肽的响应频率。

[0039] 图 15 提供的表显示了用本发明的 I-MUNE<sup>®</sup> 检测系统获得的 I-MUNE<sup>®</sup> 检测结果与发表的链激酶表位之间的表位对比。

[0040] 图 16 提供了显示对于  $\beta$ -2 微球蛋白的 I-MUNE<sup>®</sup> 检测结果的图。板块 A 显示了每个肽 ( $n = 87$ ) 的响应者百分数。连续的人  $\beta$ -2 微球蛋白肽显示于 x 轴。数量为 87 的供体组中的响应百分数显示于 y 轴。板块 B 显示了每个肽的响应频率。

[0041] 图 17 提供的表显示了通过本发明的 I-MUNE<sup>®</sup> 检测系统鉴定的对于细菌蛋白酶中的表位肽的 IC<sub>50</sub> 结合值。低于 500nM 的值可认为是良好的结合物,其在表中以粗体加以强调。简并性指明了 18 个进行测试的全部等位基因中以低于 500nM 的 IC<sub>50</sub> 进行结合的 II 型 HLA 蛋白质的数目。

[0042] 发明描述

[0043] 本发明提供了评估群体的免疫反应特性的方法。特别地,本发明提供了定性地评估人类群体的免疫反应的方法,其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外,本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。

[0044] 本发明提供了用于在人类群体基础上鉴定 CD4<sup>+</sup>T- 细胞表位的离体技术。在对于目标蛋白质预先致敏的供体群体内,可以定义出所有记忆表位 (recall epitope)。对

于定义为对于目标蛋白质不致敏的供体群体, 鉴定基本的或者可交叉反应的表位。虽然不能正式地排除后者, 但是有许多事实支持这样的结论, 即所发现的表位为基本表位 (primary epitope)。首先, 在工业蛋白质中找到的表位主要为混杂的结合物, 其在体外结合检测中具有低的  $IC_{50}$  值。记忆响应的特征在于一段时间内的低的阈值, 而不是限定于最高的结合值 (见 Hesse 等人, *J. Immunol.*, 167 :1353-1361 [2001])。第二, 当使用可能未致敏的供体时, 总是发现全部记忆表位的亚组。这是基本的、免疫显性的表位的特征 (见 Muraro 等人, *J. Immunol.*, 164 :5474-5481 [2000]; Vanderlugt 等人, *Nat. Rev. Immunol.*, 2 :85-95 [2002]; Vanderlugt 等人, *J. Immunol.*, 164 :670-678 [2000]; 和 Yin 等人, *J. Immunol.*, 26 :2063-2068 [1998])。第三,  $\beta$ -2 微球蛋白以一组偏移 3 个氨基酸的 15 聚肽进行测试, 其代表了一组 52 个肽, 对于这些肽没有发现显著的表位响应。似乎不可能的是, 没有发现这些序列中的一个为在任何其他蛋白质中的交叉反应序列。第四, 当发现与在来自人病原体的蛋白质中找到的序列可交叉反应的表位时 (正如对于检查细菌酶蛋白质的情况), 对于表位肽的响应百分数是非常高的 (30%), 这比在其他 10 种如实施例 7 中所述进行测试的工业酶中所比较的任何响应要高得多。第五, 本发明的 **I-MUNE**<sup>®</sup> 检测系统通过使用富含  $CD4^+T^-$  细胞的响应者细胞和活化的源自单核细胞的树突细胞作为 APC 来进行。所见的增殖响应的程度是非常小的, 这与抗原特异性  $CD4^+T^-$  细胞的低前体频率相一致。所检测到的记忆增殖响应比在可能未致敏的群体中检测到的响应要强烈得多。最后, 用表位序列进行 BLAST 搜索。对于源自芽孢杆菌的蛋白质, 芽孢杆菌菌种含有在所鉴定的表位序列内有修饰的蛋白酶变体。可是, 不可能的是, 供体库会对于这些或者任何其他的芽孢杆菌丝氨酸蛋白酶变得敏感 (用上面所引用的明显的交叉反应例子)。有趣的是, 在 BPN'Y217L 中的 70-84 位氨基酸表位区域和推定的人源 ATP 依赖型 RNA 解旋酶中的区域之间具有一些同源性 (66% 同源性) (见 Imamura 等人, *Nucl. Acids Res.*, 26 :2063-2068 [1998])。可以期望与广泛表达的管家基因如该基因的同源性会引起耐受性, 而不是激起交叉反应的响应。

[0045] 在分析群体数据时, 背景率是重要的要考虑的事项。背景率是由于对于表位肽的聚积性阳性响应以及达到 2.95SI 阈值的随机事件而造成的。随机聚积性阳性响应的低水平反应了在人供体中  $CD4^+T^-$  细胞增殖状况的异质性 (见 Asquith 等人, *Trends Immunol.*, 23 :595-601 [2002])。虽然背景可以通过增加响应值阈值而人工地减少, 但是具有可测量的背景率使得能够确定响应频率是以非随机的方式在何处积聚的。例如, 对于 HPV16E6 的背景响应率比对于工业酶的背景响应率明显地要高, 这可能反映了在社区供体群体中 HPV16 感染的高度流行 (Lazcano-Ponce 等人, *Int' l J. Cancer* 91 :412-420 [2001]; 和 Stone 等人, *J. Infect. Dis.*, 186 :1396-1402 [2002])。对于链激酶可能出现同样的情况。

[0046] 尽管在 **I-MUNE**<sup>®</sup> 检测系统中包括有这么多可变情况, 但是对于表位响应频率的变异系数 (CV) 是非常良好的 (对于四种测试肽, 平均值为 20%)。有利的是, 这种重现性水平与在实验室内和供体间的主要 ELISPOT 数据重复测试中所报道的变异系数值进行比较, 这一重复测试是一种类似的离体检测 (Keilhoz 等人, *J. Immunother.*, 25 :97-138 [2002]; 和 Asai 等人, *Clin. Diag. Lab Immunol.*, 7 :145-154 [2000])。一般地, 当对于表位肽的反应百分数增加时, CV 值减小。另外, 具有减少的频率 (通常低于供体群体的 10%) 的非表位肽响应有着增加的 CV 值。例如, 在实施例 7 中, 总背景率为 3.15%, 其具有 1.6% 的标准偏差, 和 51% 的 CV。

[0047] 如果群体经证实对于目标蛋白质预致敏,那么用于确定表位肽的统计学方法是不同的。增加的背景响应可能是由于对于在记忆响应中所见的功能性激活的阈值减少了(见 Hesse 等人,如上)。对于功能性激活的阈值的减少导致本发明的**I-MUNE**<sup>®</sup>检测系统可检测出更多的表位。**I-MUNE**<sup>®</sup>检测系统的结果与来自致敏供体的数据之间的比较显示出,**I-MUNE**<sup>®</sup>检测数据中的显著表位响应与无性繁殖 CD4<sup>+</sup>T- 细胞系所限定的表位响应相一致。通过降低统计学方法的严格性水平,在**I-MUNE**<sup>®</sup>检测系统中表位肽的选择与发表的表位序列一致。在具有非常低的背景率的数据组如工业酶数据中表位状况的标明是更加严格的。当背景响应非常低时,如果使用严格性减少的测定,则许多肽积聚达到阈值的响应,但是总响应频率是非常低的,并且将会难以重现。通常地,当响应低于总群体的 10% 时,由于测试超过 100 个供体的技术困难它们将会变得难以重现。显著的表位响应可容易地从频率数据中推断出来,在这些频率数据中表位响应是超偏值。在未致敏的供体中的表位肽序列可能反映了能够引起再次增殖的紧密结合混杂表位 (Viola 和 Lanzavecchi, *Science* 273 :104-106[1996]; 和 Rachmilewitz 和 Lanzavecchia, *Trends Immunol.*, 23 : 592-595[2002])。这一点通过体外肽结合研究对于在两个工业酶中标明的表位肽而得到确认(见实施例 7)。

[0048] 本发明的**I-MUNE**<sup>®</sup>检测系统没有鉴定人  $\beta$ -2 微球蛋白中的任何表位。这一结果突出了本发明的**I-MUNE**<sup>®</sup>检测系统与基于算法的 II 型 HLA 结合预测方法之间的差异。肽结合算法可通过互联网自由地使用,而且对于本领域的人员来说是已知的,其预测在该序列中的 II 型结合表位。可是,正如此处给出的结果所例示的,与 II 型分子结合并不总是表明功能性表位的存在。对于确定 T- 细胞表位,与 II 型 HLA 结合是必需的,但是还不够。这是众所周知的预测方法的特性,因此这些方法常常用功能测试来加以补充。可是,本发明提供了来获得这一信息的更直接的方法。

[0049] 重要的是要注意到此处描述的表位确定是在群体基础之上定义的。虽然显著表位常常显示出一定水平的 HLA 特异性,但是表位肽主要通过它们的混杂 HLA 结合能力来定义。因为这一点,如果要寻找免疫原性减退的蛋白质,那么这些表位可能是超型 (supertype) 结合物,因此它们代表了用于修饰的良好候选物。可是,要注意到,由于是基于群体的分析,用这些结果作为指导而形成的免疫原性减退的蛋白质并不总是在每个独立情况下都没有免疫原性。虽然如此,在群体基础上确定 T- 细胞表位在表征对感染源的免疫反应中是有用的(见 Novitsky 等人, *J. Virol.*, 76 :10155-10168[2002]; 和 Pathan 等人, *J. Immunol.*, 167 :5217-5225[2001])。这样的研究的一个目的是设计有效的疫苗,其中也保证包含有混杂超型结合物。有趣的是,当将在一个这些研究 (Pathan 等人,如上) 中给出的数据通过此处定义的暴露供体方法而进行的分析时,选择出了同样的一组优势表位响应(数据未显示)。

[0050] 除了其在传染病的确定以及蛋白质分析中的应用之外,本发明的方法还提供了用于在任何目标蛋白质中定位功能性 CD4<sup>+</sup>T- 细胞表位的方法。当认为供体群体没有接触目标蛋白质时,背景响应是低的,严格的统计学可以用于选择出 CD4<sup>+</sup> 表位序列。有趣的是,人蛋白质具有非常低的背景响应。高的背景水平是与供体接触目标蛋白质相应的,此时表位的确定依赖于较不严格的标准。表位的标明已经通过与经证实致敏的供体的数据进行比较

而得到确认。正如上面所指明的,在人  $\beta$ -2 微球蛋白中没有发现表位,这种情况也适用于普遍表达并且使免疫系统产生耐受性的蛋白质。因此,本 **I-MUNE**<sup>®</sup> 检测系统提供了用于预测基于群体的 CD4<sup>+</sup>T- 细胞表位的有价值的工具。这一技术的应用包括创造免疫原性减退的蛋白质变体,选择用于创造基于表位的疫苗的表位区域,和作为一种工具而包含于所有商用蛋白质的风险评估的评价系统中。

[0051] 实际上,本发明提供了减少 CD4<sup>+</sup>T- 细胞的致敏潜能的方法。这种方法特别可用于这样的目标群体中,即该目标群体以前未接触潜在的商用蛋白质或者任何其他想要用于人和其他动物的蛋白质。实际上,除了创造过敏原性 / 免疫原性减退的商用蛋白质变体之外, T- 细胞表位的鉴定是许多疫苗策略的基础 (Alexander 等人, *Immunol. Res.*, 18 : 79-2[1998]; 和 Berzofsky, *Ann. N. Y. Acad. Sci.*, 690 :256-264[1993])。由清除病原体的个体相对于没有清除病原体的个体所识别的 T- 细胞表位的鉴定有利于癌症和病毒疫苗的设计 (Manici 等人, *J. Exp. Med.*, 189 :871-87[1999]; Doolan 等人, *J. Immunol.*, 165 : 1123-1137; 和 Novitsky 等人, *J. Virol.*, 76 :10155-10168[2002])。过敏原性 / 免疫原性减退的蛋白质在个人护理、健康护理和家庭护理情况中的使用以及在商业应用中的使用也是清楚的。实际上,这样的过敏原性 / 免疫原性减退的蛋白质在无数的情况和应用中是有用的。

[0052] 对于创造 CD4<sup>+</sup>T- 细胞表位经修饰的蛋白质,第一个紧要的步骤是在蛋白质中定位功能性表位。由许多基于计算机的方法来预测与 II 型 HLA 分子结合的肽序列的定位 (Yu 等人, *Mol. Med.*, 8 :137-148[2002]; Rammensee 等人, *Immunogenet.*, 50 :213-219[1990]; Sturniolo 等人, *Nat. Biotechnol.*, 17 :555-561[1999]; 和 Altuvia 等人, *J. Mol. Biol.*, 249 :244-250[1995])。对于 CD4<sup>+</sup>T- 细胞的激活来说,与 HLA 的结合是必需的,但还不够。最佳的是必须进行体外和体内测试以确认功能性。对于基于计算机的方法在它们的能力上进行了改进,以便正确地鉴定紧密的 HLA 结合物,但是仍然缺乏对于结合非 II 型 HLA-DR 分子的预测,并且具有明显的错误阴性率。另外,功能的差异如耐受性的引起和通过激活的 T- 细胞而引起不同响应的表位不能用计算机建模来进行评估。

[0053] 因此,本发明提供了迄今无法获得的鉴定和确认功能性的方法以评估 CD4<sup>+</sup>T- 细胞表位经修饰的蛋白质。在一些实施方案中,本发明提供了在体外基于人细胞的方法,来定位来自任何目标蛋白质的免疫显性的、混杂的 II 型 HLA 表位。和其良好地应用于描绘基于群体的对于病原体来源的蛋白质的表位响应一样,该方法同样良好地应用于工业酶、食物过敏原和人治疗蛋白质以及任何其他目标蛋白质。在优选的实施方案中,大的供体组没有对于 HLA 类型进行预选择而进行测试。表位确定是基于将整个供体组对于所有来自蛋白质序列的肽的响应率进行统计学分析来完成的,因此代表了基于群体的表位。正如此处所指明的,本发明的方法能够区分供体群体已经接触的蛋白质和供体群体以前未遇到的或者还未对其变得敏感的蛋白质。在本发明的形成期间,将两个类型的分析与来自经证实抗原致敏的供体的增殖结果进行比较。另外,测试人  $\beta$ -2 微球蛋白,并将其确定为阴性对照。

[0054] 正如此处所提到的,表位肽通过与背景响应率的差异来标明。当对于多个随机供体组进行测试时,表位肽响应是可重现的,其具有 21% 这一变异系数中值。另外,如在此更详细地讨论的,本发明的 **I-MUNE**<sup>®</sup> 检测系统鉴定了对于蛋白质链激酶的免疫显性的、混杂的表位 (其代表了全部记忆表位的一个亚组)。此外,

**I-MUNE**<sup>®</sup>检测系统在阴性对照（即人  $\beta$ -2 微球蛋白）中没有发现表位。重要的是，本发明提供了方法以在没有对于 II 型 HLA 的类型进行预选择的情况下在任何蛋白质中鉴定功能性 CD4<sup>+</sup>T- 细胞表位，从而可表明供体群体是否预先接触了目标蛋白质，并且本发明对于体外测试不需要致敏的供体。

[0055] 在本发明的形成期间，在大的人供体库中使用肽特异性响应的统计分析提供了一种度量，这种度量以小鼠和豚鼠暴露模型所确定的顺序将四种工业酶进行分级。该分级方法也有利地与在职业性暴露的工作者中的人致敏率进行了比较。也进行本发明方法的附加的确认，这是基于已知在人中引起致敏的蛋白质的结构值来进行的。这些结果的比较表明了，致敏水平比对于人  $\beta$ -2 微球蛋白所测定的值高。在优选的实施方案中，本发明提供了比较方法来预测在人中多种相关和不相关蛋白质的免疫原性。因此，本发明提供的信息可用于蛋白质疗法的早期开发和其他基于蛋白质的应用以选择或创造出免疫原性减弱的变体。

[0056] 定义

[0057] 在此处除非另外定义，此处使用的所有技术和科学术语具有和与本发明相关的本领域中普通熟练技术人员的通常理解相同的含义。例如，微生物学和分子生物学词典 (Dictionary of Microbiology and Molecular Biology, 第二版, John Wiley & Sons 出版社, NY(1994))；和 Harper Collins 生物学词典 (The Harper Collins Dictionary of Biology, 第二版, Hale 和 Marham, Harper Perennial 出版社, NY(1991)) 为本领域的熟练技术人员提供了许多此处使用的术语的一般性词典。虽然任何相似于或者相当于此处所描述的那些的方法和材料可在本发明的实施中使用，但此处描述了优选的方法和材料。因此，在下面定义的术语通过参照作为一个整体的说明书而进行更完整的描述。

[0058] 在此处使用时，术语“群体”是指在给定的区域内相互联系的和 / 或居住的个体。在一些实施方案中，该术语用于大量的享有共同特征的个体（例如具有特殊的 HLA 类型的群体，等等）。虽然在优选的实施方案中该术语用于人类群体，但是这不是想要将该术语限制于人，因为其也可用于其他动物和生物。在一些实施方案中，该术语用于全部的条目 (item)、特征、从中可取出样本的个体等等。

[0059] 在此处使用时，术语“基于群体的免疫反应”是指群体成员的免疫反应特性（即特征）。

[0060] 在此处使用时，术语“免疫反应”是指由于生物体（例如人或者其他动物）抗免疫原而引起的免疫学反应。意图是，该术语包括所有类型的免疫反应，包括体液的（即抗体介导的）、细胞的和非特异性的免疫反应，但并不局限于此。在一些实施方案中，该术语反映了群体的免疫水平（即对于特定抗原“有免疫性”的人的数量，和 / 或对于特定抗原“没有免疫性”的人的数量）。

[0061] 在此处使用时，术语“减弱的免疫原性”是指，与原始的野生型（例如亲本或源）蛋白质相比，用变体（例如衍生物）蛋白质所观察到的免疫反应的减小。在本发明的优选实施方案中，提供了变体蛋白质，其与源蛋白质相比在体外和 / 或体内激起强度更小的免疫反应。可期望的是，这些具有减弱的免疫原性的蛋白质将会在多种应用中使用，包括生物产品、蛋白质治疗剂、食品和饲料、个人护理、去污剂和其他消费者相关的产品，以及在其他治疗方法、诊断等等中使用，但并不局限于此。

[0062] 在此处使用时,术语“增强的免疫原性”是指,与原始的野生型(例如亲本或源)蛋白质相比,用变体(例如衍生物)蛋白质所观察到的免疫反应的增加。在本发明的优选实施方案中,提供了变体蛋白质,其与源蛋白质相比在体外和/或体内激起强度更大的免疫反应。可期望的是,这些具有增强的免疫原性的蛋白质将会在多种应用中使用,包括生物产品、蛋白质治疗剂、食品和饲料添加剂,以及在其他治疗方法、诊断等等中使用,但并不局限于此。

[0063] 在此处使用时,术语“过敏性食物蛋白质”是指任何能够在人和其他动物中引起过敏反应的食物蛋白质。“推定的过敏性食物蛋白质”是可能为过敏原的食物蛋白质。“具有减弱的过敏性食物蛋白质”是经过修饰从而具有比原始的、未修饰的蛋白质更弱的过敏性的(即“过敏性减退的”)食物蛋白质。意图是,这些术语包括天然存在的食物蛋白质以及通过合成和/或使用重组技术生产出的食物蛋白质。

[0064] 在此处使用时,术语“改变的免疫反应”是指增强的或减弱的免疫反应。当蛋白质和肽激起的 T-细胞和/或 B-细胞响应比亲本(例如前体)蛋白质或肽(例如目标蛋白质)激起的 T-细胞和/或 B-细胞响应更强烈时,这些蛋白质和肽显示出“增强的免疫反应”。这一更强的响应的最终结果是增强的针对变体蛋白质或肽的抗体反应。当蛋白质和肽激起的 T-细胞和/或 B-细胞响应比亲本(例如前体)蛋白质或肽激起的 T-细胞和/或 B-细胞响应更弱时,这些蛋白质和肽显示出“减弱的免疫反应”。这一更弱的响应的最终结果是减弱的针对变体蛋白质或肽的抗体反应。在一些优选的实施方案中,亲本蛋白质是野生型蛋白质或肽。

[0065] 在此处使用时,术语“刺激系数”是指,与对照相比,肽的 T-细胞增殖响应的度量。SI 是这样计算出来的,即将在测试含有肽的 CD4<sup>+</sup>T-细胞和树突细胞培养物中获得的平均 CPM(每分钟计数)除以含有树突细胞和 CD4<sup>+</sup>T-细胞但没有肽的对照培养物的平均 CPM。该值对于每个供体和每个肽进行计算。虽然在一些实施方案中,大约 1.5-4.5 的 SI 值用于表明阳性响应,但用于表明阳性响应的优选的 SI 值为 2.5-3.5(包括端值),优选地为 2.7-3.2(包括端值),和更优选地为 2.9-3.1(包括端值)。此处描述的最优选的实施方案使用 2.95 的 SI 值。

[0066] 在此处使用时,术语“数据组”是指对于一组肽和一组对于它们抗每个测试蛋白质(即目标蛋白质)的响应进行测试的供体所汇集的数据。

[0067] 在此处使用时,术语“肽组”是指对于每个测试蛋白质(即目标蛋白质)所产生出的一组肽。肽组中的这些肽用来自每个供体的细胞进行测试。

[0068] 在此处使用时,术语“结构”和“结构值”是指用于将蛋白质的相对免疫原性进行分级的值。结构值根据下面的“偏离均值的总变化距离”公式进行测定:

$$[0069] \quad \sum \left| f(i) - \frac{1}{p} \right|$$

[0070] 其中:

[0071]  $\Sigma$  (大写字母西格马)为对于每个肽的响应频率减去该肽在组中的频率的绝对值之和;  $f(i)$  定义为对于单个肽的响应频率;和  $p$  为肽组中肽的数量。在本发明的优选实施方案中,对于每个进行测试的蛋白质测定结构值。基于所获得的结构值,从一系列进行测试的蛋白质中按最低值到最高值将测试蛋白质进行分级。在该进行分级的系列中,最低值表

明免疫原性最弱的蛋白质,而最高值表明免疫原性最强的蛋白质。

[0072] 结构值依赖于进行测试的供体的数量(即从不同的个体获得的血液样品的数量)。一般地,整个数据组的零响应提供了 1.0 的结构值。对于每个肽的响应数目相同就会得到结构值为 0。因此,在优选的实施方案中,应当对肽组进行测试直至数据组的大部分都有响应,以便数据能够准确地反映对于特定的肽和肽区域的响应。在特别优选的实施方案中,对于数据组中每个肽都有响应。可是,由于多种因素(例如不溶问题)一些数据组没有显示出对于数据组中每个肽的响应。

[0073] 虽然上面的公式是用于测定结构值的优选的公式,但是其他同等的公式在本发明中也是可用的。例如,“分布熵”(entropy of the distribution)公式在本发明中是可行的,以及多种本领域中的人员已知的其他公式。

[0074] 在一些实施方案中,肽组用至少与下列情况一样多的供体进行测试,即应当对于每个给出的肽产生响应,其具有 3% 非特异性响应的总比率。例如,在优选的实施方案中,88 个肽的肽组用最少 30 个供体进行测试。因此,在肽组包括更多的肽的实施方案中,供体的数量相应地进行调节。尽管如此,30 个供体还是优选的最小数量。当然,可以用本发明的方法测试更多的供体,甚至当在肽组中存在更少的肽时。在一些优选的实施方案中,数据组包括至少 50 个供体,以便提供良好的 HLA 等位基因的代表。

[0075] 在此处使用时,“显著响应”是指一个肽在数据组中产生体外 T- 细胞响应率,该响应率比背景响应率高大约 2.0 倍。在进一步的实施方案中,该响应比背景响应率高大约 2.0 倍 - 大约 5.0 倍。这样的响应也包括在该术语之内,即该响应代表了在背景响应率之上有大约 2.5-3.5 倍的增加、大约 2.8-3.2 倍的增加和 2.9-3.1 倍的增加。例如,在本发明的形成期间,对于一些肽记录到了显著响应。

[0076] 在此处使用时,“显著区域”是指用特定的肽组获得的 **I-MUNE**<sup>®</sup> 检测响应,其比背景响应率高大约 2.0 倍。在本发明的一个实施方案中,减少蛋白质的所有显著区域,从而降低它们在本发明的 **I-MUNE**<sup>®</sup> 检测系统中的响应。在进一步的实施方案中,显著区域的数目减少 1、2、3、4、5、6、7、8、9、10 或更多,并且优选地在相关蛋白质中减少 1-5 个显著区域。在一些实施方案中,显著区域也符合 T- 细胞表位的要求。

[0077] 在此处使用时,术语“样本”是以其最宽泛的含义进行使用的。可是,在优选的实施方案中,该术语用于指这样的样品(例如等分试样),该样品含有正进行分析、鉴定、修饰和/或与其他肽比较的肽(例如肽组中的肽,其含有目标蛋白质的序列)。因此,在多数情况下,该术语用于含有目标蛋白质或目标肽的材料。

[0078] 在此处使用时,“背景水平”和“背景响应”是指对于在任何测试蛋白质的数据组中任何给出的肽的平均响应者百分数。该值是通过将对于组中的所有肽的响应者百分数求平均值来测定的,这是通过对所有的测试供体进行汇集得到的。作为一个例子,3% 的背景响应表明,平均起来,当对 100 个供体进行测试时,对于数据组中的任何肽会有三个阳性(SI 大于 2.95) 反应。

[0079] 在此处使用时,“抗原呈递细胞”(“APC”)是指在其表面呈递抗原的免疫系统细胞,从而抗原能够被 T- 细胞表面上的受体识别。抗原呈递细胞包括树突细胞、交错突细胞、激活的 B- 细胞和巨噬细胞,但并不局限于此。

[0080] 在此处使用时,术语“T 淋巴细胞”和“T- 细胞”包括在 T 淋巴细胞谱系中的任何

细胞,从 T- 细胞前体(包括没有重排 T- 细胞受体基因的 Thy1 阳性细胞)至成熟的 T- 细胞(即对于 CD4 或 CD8 单阳性的、表面 TCR 阳性的细胞)。

[0081] 在此处使用时,术语“B 淋巴细胞”和“B- 细胞”包括在 B- 细胞谱系中的任何细胞,从 B- 细胞前体如前 B- 细胞(已经开始重排 Ig 重链基因的 B220<sup>+</sup> 细胞)至成熟的 B- 细胞和浆细胞。

[0082] 在此处使用时,“CD4<sup>+</sup>T- 细胞”和“CD4T- 细胞”是指辅助 T- 细胞,而“CD8<sup>+</sup>T- 细胞”和“CD8T- 细胞”是指细胞毒性 T- 细胞。

[0083] 在此处使用时,“B- 细胞增殖”是指在有或没有抗原存在的情况下将 B- 细胞与抗原呈递细胞一起培养期间所产生的 B- 细胞的数量。

[0084] 在此处使用时,“基本 B- 细胞增殖”是指,在不存在肽或蛋白质抗原的情况下,对于暴露于抗原呈递细胞作出响应而通常在个体中见到的 B- 细胞增殖的程度。对于此处的目的,基本 B- 细胞增殖水平是这样测定的,即在每个样品的基础之上对于每个个体测定在不存在抗原的情况下的 B- 细胞的增殖。

[0085] 在此处使用时,“B- 细胞表位”是指肽或蛋白质的特征,这些肽或蛋白质在对于含有那个抗原(即免疫原)的肽的免疫反应中被 B- 细胞受体所识别。

[0086] 在此处使用时,“改变的 B- 细胞表位”是指与前体肽或目标肽不同的表位氨基酸序列,因此目标肽变体在人或另一种动物中产生不同的(即改变的)免疫反应。期望的是,改变的免疫反应包括改变的免疫原性和/或过敏原性(即增强的或减弱的总免疫反应)。在一些实施方案中,改变的 B- 细胞表位包括替换和/或删除选自位于鉴定出的表位中的那些残基的氨基酸。在另一些可选择的实施方案中,改变的 B- 细胞表位包括在表位内添加一个或多个残基。

[0087] 在此处使用时,“T- 细胞增殖”是指在有或没有抗原存在的情况下将 T- 细胞与抗原呈递细胞一起培养期间所产生的 T- 细胞的数量。

[0088] 在此处使用时,“基本 T- 细胞增殖”是指,在不存在肽或蛋白质抗原的情况下,对于暴露于抗原呈递细胞作出响应而通常在个体中见到的 T- 细胞增殖的程度。对于此处的目的,基本 T- 细胞增殖水平是这样测定的,即在每个样品的基础之上对于每个个体测定在不存在抗原的情况下 T- 细胞对于抗原呈递细胞作出响应而产生的增殖。

[0089] 在此处使用时,“T- 细胞表位”是指肽或蛋白质的特征,这些肽或蛋白质在对于含有那个抗原(即免疫原)的肽的免疫反应的起始阶段中被 T- 细胞受体所识别。尽管不想将本发明限制于任何特定的机制,但一般相信 T- 细胞表位被 T- 细胞的识别是通过这样的机制来进行的,即在该机制中 T- 细胞识别与在抗原呈递细胞上表达的 I 型或 II 型 MHC(即 HLA) 分子结合的抗原的肽片段(见例如,Moeller, Immunol. Rev., 98 :187[1987])。

[0090] 在此处使用时,“改变的 T- 细胞表位”是指与前体肽或目标肽不同的表位氨基酸序列,因此目标肽变体在人或另一种动物中产生不同的免疫反应。期望的是,改变的免疫反应包括改变的免疫原性和/或过敏原性(即增强的或减弱的总免疫反应)。在一些实施方案中,改变的 T- 细胞表位包括替换和/或删除选自位于鉴定出的表位中的那些残基的氨基酸。在另一些可选择的实施方案中,改变的 T- 细胞表位包括在表位内添加一个或多个残基。

[0091] 在此处使用时,“目标蛋白质”是指待进行分析、鉴定和/或修饰的蛋白质(例如

蛋白酶)。天然存在的蛋白质以及重组蛋白质在本发明中是可用的。实际上,本发明可用于任何这样的蛋白质,即希望通过抗这些蛋白质来表征和/或调节人(或者其他动物)的免疫反应。在一些实施方案中,包括激素、细胞因子、抗体、酶、结构蛋白和结合蛋白在内的蛋白质可在本发明中使用。在一些实施方案中,包括胰岛素、红细胞生成素(EPO)、血小板生成素(TPO)和黄体生成素(LH)在内的激素可在本发明中使用,但并不局限于此。在进一步的实施方案中,包括干扰素(例如 IFN- $\alpha$  和 IFN- $\beta$ )、白细胞介素(例如 IL-1 至 IL-15)、肿瘤坏死因子(例如 TNF- $\alpha$  和 TNF- $\beta$ )和 GM-CSF 在内的细胞因子可在本发明中使用,但并不局限于此。在另一些其他的实施方案中,包括人类抗体和人源化抗体、任何类型的源自抗体的片段(例如单链抗体)在内的抗体(即免疫球蛋白)可在本发明中使用,但并不局限于此。在另一些其他实施方案中,包括食物过敏原(例如 Bere1[巴西坚果过敏原]和 Ara H1[花生过敏原])在内的结构蛋白可在本发明中使用,但并不局限于此。在另外的一些实施方案中,蛋白质为工业和/或医用酶。在一些实施方案中,优选类型的酶包括蛋白酶、纤维素酶、脂肪酶、酯酶、淀粉酶、酚氧化酶、氧化酶、通透酶、支链淀粉酶、异构酶、激酶、磷酸酶、内酰胺酶和还原酶,但并不局限于此。

[0092] 在此处使用时,“蛋白质”是指任何由氨基酸组成的和被本领域熟练技术人员认为是蛋白质的化合物。术语“蛋白质”、“肽”和“多肽”在此处可互换使用。其中,肽是蛋白质的一部分,本领域的熟练技术人员可根据上下文理解术语的使用。术语“蛋白质”包括成熟形式的蛋白质,以及相关蛋白质的前体形式(pro-form)和前前体形式(prepro-form)。蛋白质的前前体形式包括这样的蛋白质的成熟形式,即该蛋白质具有可操作地连接到蛋白质氨基端的前序列,和可操作地连接到前序列氨基端的“前”或“信号”序列。

[0093] 在此处使用时,“野生型”和“天然”蛋白质是指那些在自然中找到的蛋白质。术语“野生型序列”和“野生型基因”在此处可互换使用,其是指在宿主细胞中天然的或天然存在的序列。在一些实施方案中,野生型序列是指作为蛋白质工程计划的起始点的目标序列。

[0094] 在此处使用时,“蛋白酶”是指天然存在的蛋白酶以及重组蛋白酶。蛋白酶为羧基水解酶,其通常可以切割蛋白质或肽的肽键。天然存在的蛋白酶包括  $\alpha$ -氨基肽水解酶、肽基氨基酸水解酶、酰氨基水解酶、丝氨酸羧肽酶、金属羧肽酶、巯基蛋白酶、羧基蛋白酶和金属蛋白酶这样的例子,但并不局限于此。包括丝氨酸、金属、巯基和酸蛋白酶,以及内切和外切蛋白酶。实际上,在一些优选的实施方案中,使用丝氨酸蛋白酶,例如胰凝乳蛋白酶和枯草杆菌蛋白酶。这两种丝氨酸蛋白酶具有含有天冬氨酸、组氨酸和丝氨酸的催化三联体。在枯草杆菌蛋白酶中,这些氨基酸的相对顺序从羧基端开始为天冬氨酸-组氨酸-丝氨酸,而在胰凝乳蛋白酶中,这些氨基酸的相对顺序从羧基端开始为组氨酸-天冬氨酸-丝氨酸。虽然枯草杆菌蛋白酶通常是从细菌、真菌或酵母来源中获得的,但是在此处使用时,“枯草杆菌蛋白酶”是指具有上面定义的枯草杆菌蛋白酶催化三联体的丝氨酸蛋白酶。另外,人枯草杆菌蛋白酶为具有枯草杆菌蛋白酶催化活性的人源蛋白质,例如人源蛋白酶的 kexin 家族。枯草杆菌蛋白酶对于本领域的熟练技术人员来说是熟知的,例如解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*) 枯草杆菌蛋白酶(BPN')、迟缓芽孢杆菌枯草杆菌蛋白酶、枯草芽孢杆菌(*Bacillus subtilis*) 枯草杆菌蛋白酶、地衣形芽孢杆菌枯草杆菌蛋白酶(见例如, U. S. 专利 4,760,025(RE 34,606)、U. S. 专利 5,204,015、U. S. 专利 5,185,258、EP 0328299 和 WO 89/06279)。

[0095] 在此处使用时,功能上相似的蛋白质被认为是“相关蛋白质”。在一些实施方案中,这些蛋白质来自不同的属和/或种(例如枯草芽孢杆菌枯草杆菌蛋白酶和迟缓芽孢杆菌枯草杆菌蛋白酶),包括了生物类型之间的差异(例如细菌枯草杆菌蛋白酶和真菌枯草杆菌蛋白酶)。在另外的实施方案中,相关蛋白质从同一物种中提供。实际上,不是想要将本发明限定于来自任何来源的相关蛋白质。

[0096] 在此处使用时,“衍生物”是指源自前体蛋白质(例如天然蛋白酶)并通过下列修饰的蛋白质(例如蛋白酶),即在 C-末端和 N-末端中的一个或两者上添加一个或多个氨基酸,在氨基酸序列中一个或许多不同的位点替换一个或多个氨基酸,和/或在蛋白质末端中的一个或两者上或者在氨基酸序列中的一个或多个位点上删除一个或多个氨基酸,和/或在氨基酸序列中的一个或多个位点上插入一个或多个氨基酸。蛋白质衍生物的制备优选地通过下列步骤来完成,即修饰编码天然蛋白质的 DNA 序列,将该 DNA 序列转化入合适的宿主中,然后表达经修饰的 DNA 序列以形成衍生蛋白酶。

[0097] 相关(和衍生)蛋白质的一种类型为“变体蛋白质”。在优选的实施方案中,变体蛋白质不同于亲本蛋白质,它们彼此之间相差少量的氨基酸残基。差异氨基酸残基的数目可以是一个或多个,优选为 1、2、3、4、5、10、15、20、30、40、50 或更多的氨基酸残基。在一个优选的实施方案中,变体之间差异氨基酸的数目为 1-10。在特别优选的实施方案中,相关蛋白质和特定变体蛋白质具有至少 50%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98% 或 99% 的氨基酸序列同一性。另外,在此处使用时,相关蛋白质或变体蛋白质是指在显著区域的数量上不同于另一个相关蛋白质或亲本蛋白质的蛋白质。例如,在一些实施方案中,变体蛋白质具有 1、2、3、4、5 或 10 个不同于亲本蛋白质的相应的显著区域。在一个实施方案中,变体的显著相应区域只产生背景水平的免疫反应。

[0098] 在此处使用时,“相应于”是指在蛋白质或肽中的计数位置上的残基,或者与另一个蛋白质或肽中的计数位置上的残基相似、同源或相当的残基。

[0099] 在此处使用时,“相应区域”一般是指在相关蛋白质或亲本蛋白质内相似的位置。

[0100] 在此处使用时,术语“相似序列”是指与目标蛋白质具有类似的功能、三级结构和/或保守残基的蛋白质之内的序列。在特别优选的实施方案中,相似序列包括位于或邻近表位的序列。例如,在含有  $\alpha$  螺旋或  $\beta$  折叠结构的表位区域内,在相似序列内氨基酸的置换优选地保持了同样的特异性结构。

[0101] 在此处使用时,“同源蛋白质”是指具有与目标蛋白质(例如蛋白酶)类似的催化活性、结构、抗原性和/或免疫反应的蛋白质(例如蛋白酶)。不是想要让同源物和目标蛋白质(例如蛋白酶)在进化上必需相关。因此,想要让该术语包括从不同物种中获得的具有同样功能的蛋白质。在一些优选的实施方案中,希望鉴定出具有与目标蛋白质类似的三级和/或一级结构的同源物,因为在目标蛋白质内用来自同源物的相似片段对于表位进行置换将会减少改变所带来的破坏。因此,在大多数情况下,同源性高的蛋白质提供了最想要的表位替换的来源。另一个原因是,其有利于寻找对于指定蛋白质的人相似物。

[0102] 在此处使用时,“同源基因”是指至少一对来自不同的但通常相关的物种的基因,它们互相相关和它们互相是同一的或非常类似的。该术语包括通过物种形成(即新物种的形成)而分离的基因(例如正向进化同源基因),以及通过基因复制而分离的基因(例如平行进化同源基因)。

[0103] 在此处使用时,“正向进化同源物”和“正向进化同源基因”是指在不同物种中通过物种形成而从共同的祖先基因(即同源基因)进化来的基因。通常地,正向进化同源物在进化过程期间保持了同样的功能。正向进化同源物的鉴定可用于在新近测序的基因组中基因功能的可靠预测。

[0104] 在此处使用时,“平行进化同源物”和“平行进化同源基因”是指在基因组内通过复制而相关的基因。虽然正向进化同源物在进化过程中保持了同样的功能,但是平行进化同源物却进化出了新功能,即使一些功能常常与原始的基因相关。平行进化同源基因的例子包括编码胰蛋白酶、胰凝乳蛋白酶、弹性蛋白酶和凝血酶的基因,但并不局限于此,它们都是丝氨酸蛋白酶,并在同一物种中一起出现。

[0105] 序列之间的同源性程度可以用本领域中已知的任何合适的方法来测定(见例如,Smith 和 Waterman, *Adv. Appl. Math.*, 2:482[1981]; Needleman 和 Wunsch, *J. Mol. Biol.*, 48:443[1970]; Pearson 和 Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444[1988]; Wisconsin 遗传学软件包中的程序如 GAP、BESTFIT、FASTA 和 TFASTA (Genetics Computer Group, Madison, WI); 和 Devereux 等人, *Nucl. Acid Res.*, 12:387-395[1984])。

[0106] 例如, PILEUP 是对于测定序列同源性水平有用的程序。PILEUP 通过使用渐进的、成对的比对而从一组相关序列中进行多序列对比。其也可以绘制显示出用于进行比对的聚类关系的树图。PILEUP 使用了 Feng 和 Doolittle 的渐进式比对方法的简化形式 (Feng 和 Doolittle, *J. Mol. Evol.*, 35:351-360[1987])。该方法类似于 Higgins 和 Sharp 所描述的方法 (Higgins 和 Sharp, *CABIOS*, 5:151-153[1989])。有用的 PILEUP 参数包括 3.00 的默认缺口权重、0.10 的默认缺口长度权重和加权的末端缺口。有用的算法的另一个例子为 Altschul 等人所描述的 BLAST 算法 (Altschul 等人, *J. Mol. Biol.*, 215:403-410[1990]; 和 Karlin 等人, *Proc. Natl. Acad. Sci. USA* 90:5873-5787[1993])。一个特别有用的 BLAST 程序为 WU-BLAST-2 程序(见 Altschul 等人, *Meth. Enzymol.*, 266:460-480[1996])。参数“W”、“T”和“X”决定了对比的灵敏度和速度。BLAST 程序使用的默认值有 11 的词长度 (W)、50 的 BLOSUM62 评分矩阵(见 Henikoff 和 Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915[1989]) 排列 (B)、10 的期望值 (E)、M' 5、N' -4 和两条链的比较。

[0107] 在此处使用时,“核酸序列同一性百分数(%)”定义为在候选序列中与序列的核苷酸残基一样的核苷酸残基的百分数。

[0108] 在此处使用时,术语“杂交”是指一种过程,通过这一过程核酸的一条链通过碱基配对与互补链结合在一起,这在本领域中是已知的。

[0109] 在此处使用时,“最大严格性”是指通常于大约  $T_m - 5^\circ\text{C}$  (比探针的  $T_m$  低  $5^\circ\text{C}$ ) 出现的杂交水平;“高严格性”是指于比  $T_m$  低大约  $5^\circ\text{C} - 10^\circ\text{C}$  出现的杂交水平;“中等严格性”是指于比  $T_m$  低大约  $10^\circ\text{C} - 20^\circ\text{C}$  出现的杂交水平;和“低严格性”是指于比  $T_m$  低大约  $20^\circ\text{C} - 25^\circ\text{C}$  出现的杂交水平。正如本领域中熟练技术人员将会理解的,最大严格性的杂交可以用于鉴定或检测相同的多核苷酸序列,而中等或低严格性的杂交可以用于鉴定或检测多核苷酸序列同源物。

[0110] 在一些实施方案中,“等效残基”是通过在前体蛋白质(即目标蛋白质)的三级结构水平上进行测定同源性来定义的,其中前体蛋白质的三级结构已经通过 x 射线晶体学进行了测定。等效残基定义为这样的残基,即对于这些残基,前体蛋白质和另一个蛋白质的特

定氨基酸残基的两个或更多个主链原子的原子坐标在排列比对之后位于 0.13nm 之内,优选地于 0.1nm 之内。在将最好的模型进行定向和定位以便给出蛋白质的非氢蛋白质原子的原子坐标的最大重叠之后来完成对比。在大多数实施方案中,最好的模型是这样的晶体学模型,即其给出对于在最高可用解析度的实验衍射数据的最低 R 因子。

[0111] 在一些实施方案中,优选地对于编码前体酶的氨基酸序列的“前体 DNA 序列”进行修饰,但在可选择的一些实施方案中,修饰通过操作前体蛋白质来进行。在残基不保守的情况下,将一个或多个氨基酸的置换限制于这样的替换,即该替换产生一种具有与自然中发现的氨基酸序列不一致的氨基酸序列的变体。在保守残基的情况下,这种置换不应当产生天然存在的序列。本发明提供的衍生物进一步包括改变蛋白酶特性的化学修饰。

[0112] 在一些优选的实施方案中,将蛋白质基因连接入合适的表达质粒之中。然后使用克隆的蛋白质基因转化或转染宿主细胞,以便表达蛋白质基因。在其含有熟知的对于质粒复制必需的元件的情况下该质粒可以在宿主中复制,或者可以将质粒设计为插入宿主染色体之中。提供必需的元件以用于有效的基因表达(例如可操作地连接到目标基因上的启动子)。在一些实施方案中,这些必需的元件以下列形式进行提供,即基因本身同源的启动子,如果其可被识别(即由宿主进行转录),和转录终止子(对于真核宿主细胞来说为聚腺苷酸化区域),其是外源的或者由蛋白质基因的内源终止子区域所提供。在一些实施方案中,也包括选择基因,例如抗生素抗性基因,其使得经质粒感染的宿主细胞能够通过含有抗微生物剂的培养基中生长而持续地进行培养。

[0113] 在涉及蛋白酶的实施方案中,通过检测蛋白酶与多种商用底物之间的相互作用而测定变体蛋白酶活性,并将其与目标蛋白酶进行比较,商用底物包括酪蛋白、角蛋白、弹性蛋白和胶原,但并不局限于此。实际上,可以想到的是蛋白酶活性可用任何本领域中已知的合适方法来测定。典型的用于测定蛋白酶活性的检测方法包括琥珀酰-Ala-Ala-Pro-Phe-对硝基苯胺(SAAPFpNA)(引用)检测法和 2,4,6-三硝基苯磺酸钠(TNBS)检测法,但并不局限于此。在 SAAPFpNA 检测法中,蛋白酶切割肽和对硝基苯胺之间的键从而给出在 405nm 吸收的可见黄色。在 TNBS 颜色反应方法中,该检测法测量酶催化水解底物形成含有游离氨基的多肽。这些氨基与 TNBS 相互作用而形成黄色复合物。因此,反应的颜色越深,所测得的活性越高。黄色可以用多种本领域中已知的分析器或分光光度计进行测定。

[0114] 变体蛋白酶的其他特性可以用本领域熟练技术人员已知的方法来测定。典型的特性包括热稳定性、碱稳定性和在多种底物或缓冲溶液或产品配剂中特定蛋白酶的稳定性,但并不局限于此。

[0115] 当联合此处公开的酶稳定性检测程序时,可以鉴定出通过随机诱变获得的突变体,其显示出增加或减少的碱或热稳定性,而同时保持酶活性。

[0116] 碱稳定性可以用已知的程序或者用此处描述的方法来测量。当与前体蛋白质相比较时,突变体的酶活性的半寿期至少有大约 5% 或更大的增加或减少(在大多数实施方案中,其优选为增加)可证明碱稳定性有实质性的改变。

[0117] 热稳定性可以用已知的程序或者用此处描述的方法来测量。当与前体蛋白质相比较而暴露于相对高的温度和中性 pH 时,突变体的催化活性的半寿期至少有大约 5% 或更大的增加或减少(在大多数实施方案中,其优选为增加)可证明热稳定性有实质性的改变。

[0118] 许多本发明的蛋白质变体在配制多种用于许多应用的组合物中使用,从个人护理至工业生产。例如,许多已知的化合物是在去污剂组合物中有用的合适表面活性剂,这些去污剂组合物含有本发明的蛋白质变体。这些去污剂包括非离子、阴离子、阳离子、阴离子或两性离子去污剂(见例如,US 专利 No. 4, 404, 128、US 专利 No. 4, 261, 868 和 US 专利 No. 5, 204, 015)。因此,可以预期到,如此处的描述而进行表征和修饰的蛋白质将会在多种去污剂应用中使用。本领域的技术人员熟悉不同的可用作清洁组合物的配剂。除了通常的清洁组合物之外,可以容易地理解,本发明的蛋白质变体可在天然或野生型蛋白质所用于的任何目的中使用。因此,这些变体例如可以用于块状肥皂或液体肥皂应用、盘碟护理配剂、表面清洁应用、接触式镜片清洁溶液和 / 或产品、肽水解、废水处理、纺织品应用,和在蛋白质生产中作为融合-切割酶等等。例如,除了减弱的过敏原性之外,本发明的变体可以在去污剂组合物中具有增强的性能(与前体比较)。实际上,不是想要将本发明的变体限制于任何特定的使用。在此处使用时,“去污剂中的增强性能”定义为,当在标准的洗涤循环之后用通常的评估方法进行测定时,测定到对于某些酶敏感的污点(例如草汁或血液)的清洁性能增加了。

[0119] 在一些实施方案中,本发明的方法所提供的蛋白质(特别是酶)可以大约 0.01% - 大约 5% (优选为 0.1% -0.5%) 的重量比的水平配制进已知的 pH 为 6.5-12.0 的粉末和液体去污剂之中。在一些实施方案中,这些去污清洁组合物进一步含有其他的酶,例如蛋白酶、淀粉酶、纤维素酶、脂肪酶或内切糖苷酶,以及助洗剂和稳定剂。

[0120] 向常规的清洁组合物中添加蛋白质不会造成任何特定的使用限制。也就是说,任何对于去污剂合适的温度和 pH 也适合于本组合物,只要 pH 位于上述范围和温度低于所述蛋白质的变性温度。另外,本发明的蛋白质还可用于无去污剂的清洁组合物,其单独或者与助洗剂和稳定剂联合使用。

[0121] 在一个实施方案中,本发明提供了用于纺织品处理的组合物,其含有本发明的变体蛋白质。该组合物可以用于处理例如丝绸或羊毛(见例如,RE 216, 034、EP 134, 267、US 4, 533, 359 和 EP 344, 259)。在一些实施方案中,根据本领域中所熟知的方法,根据蛋白水解活性来筛选这些变体。

[0122] 正如上面所表明的,在优选的实施方案中,当与它们的前体 DNA 所编码的天然蛋白质进行比较时,本发明的蛋白质显示出经修饰的免疫反应(例如抗原性和 / 或免疫原性)。在一些优选的实施方案中,蛋白质(例如蛋白酶)显示出减弱的过敏原性。本领域的熟练技术人员容易认识到,本发明的蛋白酶的使用将会在很大程度上取决于蛋白质的免疫学特性。例如,显示出减弱的免疫反应的蛋白酶可以用于清洁组合物。有效量的一种或多种此处描述的蛋白酶变体可在用于清洁多种需要去除蛋白质污点的表面的组合物中使用。这样的清洁组合物包括用于清洁坚硬表面的去污剂组合物、用于清洁织物的去污剂组合物、洗盘碟组合物、口腔清洁组合物和假牙清洁组合物。

[0123] 有效量的一种或多种具有减弱的过敏原性 / 免疫原性(根据本发明的方法进行分级)的相关和 / 或变体蛋白质可用于多种组合物,这些组合物为应用于角蛋白材料如甲和头发的组合物,包括那些用作头发喷雾组合物、头发香波和 / 或调理组合物的组合物,但并不局限于此,还有应用于头发生长调理目的的组合物,和对于治疗皮脂溢出、皮炎和 / 或落头皮屑的目的而应用于头发和头皮的组合物。

[0124] 在另一些实施方案中,有效量的一种或多种此处描述的蛋白酶变体可用于适合于对皮肤或毛发进行局部施用的组合物。这些组合物可以霜剂、洗液、凝胶等等形式进行配制,也可以配制为含水的组合物,或者可以配制为在含水连续相中具有一种或多种油相的乳浊液。

[0125] 另外,具有减弱的过敏原性/免疫原性的相关和/或变体蛋白质可用于其他应用,包括药物应用、药物递送应用和其他健康护理应用。

[0126] 发明详述

[0127] 本发明提供了评估群体的免疫反应特性的方法。特别地,本发明提供了定性地评估人类群体的免疫反应的方法,其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外,本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。

[0128] 本发明提供了通过分析个体供体对于一组描述了目标蛋白质的肽的响应率来评估任何蛋白质的总免疫原性潜能的方法。这些方法可用于选择免疫原性最弱的相关蛋白质的异构体。另外,这些方法可用于指导具有减弱的免疫原性的变体蛋白质的开发。

[0129] 在一些优选的实施方案中,基于群体的免疫反应特性可用于这些开发具有减弱的免疫原性的蛋白质的方法中。另外,本发明提供了确定一个特定的群体是否已经接触了目标蛋白质的方法,以及确定在该群体的个体中免疫反应的水平。该测定提供了信息以用于开发具有改变的免疫原性特征的蛋白质,这些特征在一些应用中是希望的,例如在生物产品、食品和饲料、蛋白质治疗剂、个人护理、健康护理产品、去污剂和其他消费者相关的产品中。

[0130] 本发明提供了用于研究群体的免疫反应的新方法。正如此处所表明的,对于涉及施用于人的蛋白质的应用的潜能测定目前使用非人的动物模型。另外,基于算法的 T-细胞表位的测定没有提供本发明的应用所提供的所需的信息。实际上,本发明提供了用于评估个体以及群体的免疫反应特性的方法,这提供了用于含蛋白质产品的合理设计和开发的重要信息。

[0131] 通过分析蛋白质的背景响应和结构值,可以在群体基础上确定任何目标蛋白质的免疫“历史”。高的背景响应表明群体的预暴露(即超过大约 4%的群体显示出对于测试蛋白质的免疫反应)。高的结构值对于具有低背景值的蛋白质表明潜在的免疫原,和当蛋白质具有高的背景值时表明新近的、频繁的和“高质量的”免疫反应。在一些实施方案中,观察到“高质量的”免疫反应,这是由于高水平的免疫原、抗免疫原的强免疫反应和/或强佐剂加强的反应。

[0132] 在一些实施方案中,具有高背景的低结构值表示减弱的免疫记忆响应、群体中不常见的响应、外源抗原引起的耐受性和/或对于高度不同的蛋白质的响应(即它们也可以是“减弱”记忆响应的产品)。可以想到,在这一类型的响应特性中代表了普通的、非过敏原的食物蛋白质。另外,具有低结构值和低背景的蛋白质表示在群体中无记忆响应的非免疫原性蛋白质,和/或人类群体对其耐受的蛋白质。在一些优选的实施方案中,修饰具有低的暴露背景水平的蛋白质,从而将其制成“过敏原性减退的”蛋白质(即当将它们暴露于人或其他动物时,它们不引起免疫反应或者引起较弱的反应)。

[0133] 为了建立一般供体群体未遇到的蛋白质的背景值,对于包括蛋白酶、淀粉酶、漆酶

和几丁质酶在内的 11 种工业酶进行 I-MUNE<sup>®</sup> 检测 (见 Mathies, Tenside Surf. Det., 34 :450-454[1997])。一种蛋白酶用以两种不同方式产生的肽 (PepSet 对从 Mimotope 纯化出的肽) 进行两次测试。每个肽组的测试供体的数目从 19-113 变化。每个肽组中肽的数目从 80-188 变化。当单个肽的刺激系数 (S. I. 或 SI) 为 2.95 或更高时, 将响应制成表格。计算出测试供体组中对于每个肽作出响应的供体百分数。计算出对于每个测试蛋白质的每个肽的平均响应百分数, 将其相对于测试供体的数量进行作图表示 (图 11)。相关系数为  $R^2 = 0.86$ 。相关性的斜率揭示了为 3.01% 的平均响应累积率。因此, 对于任何给定的用来自工业蛋白质的肽进行测试的供体, 平均 100 个肽中会有 3 个肽得出阳性 ( $SI \geq 2.95$ ) 响应。这一平均响应率包括表位肽 (见下) 和非表位肽。

[0134] 背景响应也可通过将完整的数据组中每个肽的响应百分数求平均值来计算。将 12 个测试的背景响应求平均值, 该值为 3.15+/-0.45 (平均值 +/- 标准误差), 其与通过相关趋势线的斜率确定的值是一致的。

[0135] 在本发明的形成期间, 一组蛋白质基于它们推测的在一般人类群体中的暴露而选择出来。这些蛋白质包括人乳头状瘤病毒 (HPV) 株系 16 和株系 18E6 蛋白、巴西坚果过敏原 Ber e 1 和链激酶。HPV 16 和 18 是最普遍形式的致癌性 HPV 病毒。通过年轻女性的抽样分析, 已经估计对于这些病毒的暴露水平为 5% 或更高 (Lazcano-Ponce 等人, Intl. J. Cancer 91 :412-420[2001]; Stone 等人, J. Infect. Dis., 186 :1396-1402[2002]; Goldsborough 等人, Mol. Cell Probes, 6 :451-457[1992]; 和 Lorincz 等人, Obstet. Gynecol., 79 :328-337[1992])。该比率随着地区和年龄而变化。巴西坚果过敏反应在 < 1% 的群体中发生, 但是接触食物中的巴西坚果却是普遍的 (Sicherer 和 Sampson, Curr. Opin. Pediatr., 12 :567-573[2000])。另外, 人外周血细胞培养物中链激酶特异性 T- 细胞响应率随着年龄而增加, 30% 的年轻供体有响应, 多于 70% 的超过 40 岁的供体有响应 (Warmerdam 等人, J. Immunol., 168 :155-161[2002])。将对于这四种蛋白质的肽组用来自地方社区血库的样品进行测试。对于所有这四种蛋白质的背景响应比在 11 种工业酶中发现的平均响应高。当与基于来自图 11 中 11 种工业酶的数据的期望值进行比较时, 这显示为更高的总背景响应百分数和更高的每个肽的响应频率。对于 HPV 16E6 和链激酶的背景响应明显地更高。这一结果与在供体库中推测的对于这些蛋白质更高的暴露率是一致的。对于 HPV 18E6 和 Ber e 1 的背景响应比工业蛋白质的平均值更高, 但并不是明显地不同。与工业蛋白质值相比背景值的增加是由于供体群体中 CD4<sup>+</sup> 记忆响应的作用, 其增加了对于给定蛋白质的总响应的幅度、数量和复杂性 (Kuhns 等人, Proc. Natl. Acad. Sci. USA 97 :12711-12716[2000]; Muraro 等人, J. Immunol., 164 :5474-5481[2000]; 和 Vanderlugt 和 Miller, Nat. Rev. Immunol., 2 :85-95[2002])。因此, 更高的背景率代表更高水平的对于测试蛋白质的致敏。可是, 不是想要将本发明限制于任何特定的抗这些蛋白质的总响应的机制。对于此处描述的蛋白质, 可以推断出, 我们的供体群体具有对于 HPV16E6 和链激酶的显著接触, 和具有对于 HPV 18E6 和 Ber e 1 的较少接触。对于 Ber e 1 和 HPV 18E6 的背景响应暗示了与这些蛋白质有接触, 但还没有达到 HPV 16E6 或链激酶的水平。

[0136] 除了这些蛋白质, 描述人蛋白质的肽组也在本发明形成期间进行测试。这些蛋白质包括干扰素 - $\beta$  (IFN- $\beta$ ) (在免疫反应期间广泛表达的细胞因子)、血小板生成素 (TPO) (其表达限制于骨髓的细胞因子) 和可溶的重组细胞因子受体分子 (肿瘤坏死因子受

体-1 ;TNF-R1)。对于这所有四种蛋白质的背景响应类似于工业酶的背景数据,这暗示供体正在对于这些组中的肽作出响应,好像它们未接触过这些蛋白质,或者说对于这些蛋白质是“天真的”(naive)。这些数据与对于这些特定蛋白质的外周耐受性的未知机制是一致的。

[0137] 在另外的实施方案中,进行与测试蛋白质相关的 T- 细胞和 / 或 B- 细胞表位的评估。在进一步的实施方案中,这一评估用于在这样的表位中进行合理的改变,从而减弱测试蛋白质的免疫原性 / 过敏原性 (即产生具有减弱的免疫原性的变体蛋白质)。然后这些变体蛋白质可用于多种应用,包括生物产品、蛋白质治疗剂、食品和饲料、个人护理、去污剂和其他消费者相关的产品,以及在其他治疗方法、诊断等等中使用,但并不局限于此。

[0138] 在优选的实施方案中,该方法使用作为抗原呈递细胞的树突细胞、偏移 3 个氨基酸从而包含完整的蛋白质序列的 15 聚肽和从树突细胞供体获得的 CD4<sup>+</sup>T- 细胞。如果对于特定的肽结合氘化胸苷的平均 CPM 高于或者等于 2.95 倍的背景 CPM,则记录下该“阳性”反应。对于一个大供体组将对于每个肽的结果制成表格,这样的大供体组应当能够反映一般的 HLA 等位基因频率 (具有一些变化)。基于“线性差异”(difference from linearity) 的确定而进行统计学计算,并将该结构值用于将这些蛋白质的相对免疫原性进行分级。正如此处所表明的,用本发明的方法获得的分级结果接近地反映了当这些蛋白质在职业性暴露的工作者中测定 (见 Sarlo, 如上) 或者在 GPIT 或 MINT 检测系统中测定 (见 Robinson, [1998], 如上) 时,这些作为呼吸性过敏原的蛋白质的免疫原性测定结果 (即通过 Sarlo 的 MID 检测法, [1997], 如上) 和过敏原性。

[0139] 在本发明的形成期间,发现一组包括三种已知免疫原的蛋白质的结构值是比较高的,这表明这些蛋白质可能能够在较大数量的经暴露的人群中引起免疫反应。相反地,小鼠 VH 36-60 基因家族成员的结构值是低的,这与其预测的免疫原性是一致的 (见 Olsson, J. Theor. Biol., 151 :111-122 [1991])。最后对于  $\beta$ -2 微球蛋白所测定的结构值是低的,这正如预测的一样,其中该预测假定该分子受到外周和中央耐受性机制的影响 (见 Guery 等人, J. Immunol., 154 :545-554 [1995])。

[0140] 正如此处所描述的,在另外的实验中,测试 25 种不同的蛋白质。这些数据提供了用于证实本发明的框架;不是想要将本发明限制于这 25 种蛋白质。实际上,本发明可用于任何合适的目标群体中对任何合适的目标蛋白质进行分析。与进行上述的初始实验一样,将蛋白质在此处描述的 I-MUNE<sup>®</sup> 检测系统中进行测试,并测定结构值。对于这 25 种蛋白质,结构值和背景响应描绘了在测试群体中具有变化的重要性属性的四个蛋白质亚组。此处描述的分级方法在那些具有低背景响应的蛋白质上得到了证实。此外,将所有的测试蛋白质与那些具有高背景响应的蛋白质进行比较。除了将蛋白质的潜在免疫原性进行分级之外,这些实施方案提供了关于一般群体出现的抗测试蛋白质的免疫反应类型的信息。

[0141] 在本发明的 I-MUNE<sup>®</sup> 检测系统中进行测试的蛋白质的相对免疫原性假定,蛋白质将会在体内以同样的剂量、在同样的配剂中、在一组相匹配的供体中和经过同样的给药过程而进行比较。该分析也排除了任何蛋白质的加工和 / 或呈递差异,以及一般的物理和结构特性 (即稳定性和活性)。

[0142] 本发明提供了易化在任何目标蛋白质中进行 T- 细胞表位定位的方法。例如,在一些优选的实施方案中,在不存在对于测试蛋白质致敏的个体的情况下测定 CD4<sup>+</sup>T- 细胞表

位。因此,对肽表位进行修饰,从而可获得预测在人中有效的减少的响应率,而不需要致敏自愿者。在一些实施方案中,分析对于经修饰的肽变体的供体响应,这一分析用于计算新蛋白质的结构值。例如,如图9中所示,构建成具有减少的结构值的蛋白酶变体在体外引起比亲本蛋白质显著更少的增殖。

[0143] 本发明在测定蛋白质的免疫原性中提供了明显的优点。与本发明相反,依靠用大量人供体复制品在体外激起更弱的响应来测试设计为具有更弱的免疫原性的蛋白质变体的这一测试方法不能合理地在豚鼠或小鼠中进行。转基因小鼠在它们的使用中是有限制的,这是由于这样的事实,即它们通常不能表达超过一个的 HLA 等位基因,甚至该基因常常不能以正确的环境(context)中进行表达。

[0144] 虽然蛋白质的分级并不意味着任何倍数的效价差异,但是豚鼠和小鼠模型中的效价差异众所周知是不准确的,其易受实验室间以及实验间可变性的影响,而且在小鼠中是品系依赖的。实际上,动物特别是豚鼠中效价的测定最多只是主观科学。目前,没有可靠的方法来测定效价。可是,本发明提供了进行效价测定的方法,其是基于将用本发明方法测定的数据与从动物实验中获得的数据进行对比而推断数据。这些效价值受到与用于使结构值结果标准化的动物数据一样的内在不准确性的影响,尽管有这样的事实,本发明的方法提供了经大大改进的方法来评估免疫原性(特别是在人中),和确定如何最好地减少蛋白质的免疫原性。

[0145] 此外,本发明提供了在人受试者(或者其他动物)中测定蛋白质的相对免疫原性的方法,其不需将受试者暴露于目标蛋白质。因此,为了进行测定,不会有使得个体对于具有潜在过敏原性/免疫原性的物质致敏的风险。重要的是,本发明提供了方法来对蛋白质相互之间的免疫原性进行分级,以及评估群体的免疫反应特性。实际上,本发明提供了方法来选择和/或开发具有减弱的免疫原性的蛋白质,和指导蛋白质的合理修饰,以创造和测试免疫原性减退的变体,这些变体适合于在人和任何其他动物中使用,特别是人。

#### [0146] 实验

[0147] 下面的实施例用于举例说明本发明的某些优选实施方案和方面,而不要将其解释为限制本发明的范围。

[0148] 在下面的实验相关内容中,使用下列缩写:eq(当量);M(摩尔/升); $\mu$ M(微摩尔/升);N(常态);mol(摩尔);mmol(毫摩尔); $\mu$ mol(微摩尔);nmol(纳摩尔);g(克);mg(毫克);kg(千克); $\mu$ g(微克);L(升);ml(毫升); $\mu$ l(微升);cm(厘米);mm(毫米); $\mu$ m(微米);nm(纳米); $^{\circ}$ C(摄氏度);h(小时);min(分钟);sec(秒);msec(毫秒); $\times$ g(倍重力);Ci(居里);OD(光密度);Dulbecco's 磷酸缓冲溶液(DPBS);HEPES(N-[2-羟乙基]哌嗪-N-[2-乙基磺酸]);HBS(HEPES 缓冲盐水);SDS(十二烷基硫酸钠);Tris-HCl(三[羟甲基]氨基甲烷-盐酸);Klenow(DNA 聚合酶 I 大(Klenow)片段);rpm(每分钟转数);EGTA(乙二醇双( $\beta$ -氨基乙醚)-N,N,N',N'-四乙酸);EDTA(乙二胺四乙酸);SPT<sup>+</sup>(皮肤针刺测试阳性);SPT<sup>-</sup>(皮肤针刺测试阴性);ATCC(美国典型培养物保藏中心, Rockville, MD);Cedar Lane(Cedar Lane 实验室, Ontario, 加拿大);Gibco/Life Technologies(Gibco/Life 技术公司, Grand Island, NY);Sigma(Sigma 化学制剂公司, St. Louis, MO);Pharmacia(Pharmacia 生物技术公司, Piscataway, NJ);Procter & Gamble(Procter 和 Gamble 公司, Cincinnati, OH);Genencor(Genencor 国际

公司, Palo Alto, CA); Endogen (Endogen 公司, Woburn, MA); Cedarlane (Cedarlane 公司, Toronto, 加拿大); Dynal (Dynal 公司, 挪威); Novo (Novo 工业 A/S 公司, Copenhagen, 丹麦); Biosynthesis (Biosynthesis 公司, Louisville, TX); TriLux Beta (TriLuxBeta 公司, Wallac, 芬兰); DuPont/NEN (DuPont/NEN 研究产品公司, Boston, MA); TomTec (Hamden, CT); 和 Stratagene (Stratagene 公司, La Jolla, CA)。

#### [0149] 肽

[0150] 所有的肽从商业来源 (Mimotopes, San Diego, CA) 获得。对于此处描述的 **I-MUNE**<sup>®</sup> 检测系统, 以多针方式 (multipin format) 来合成偏移 3 个氨基酸从而描述了完整的目标蛋白质序列的 15 聚肽 (见 Maeji 等人, J. Immunol. Meth., 134:23-33[1990])。将肽以大约 1-2mg/ml 重新悬浮在 DMSO 中, 并在使用之前贮存于 -70°C。每个肽至少分成两份进行测试, 但是对于小的肽组 (例如 Ber e 1), 这些肽通常分成三份进行测试。将每个肽的结果求平均值, 并对于每个肽计算出刺激系数 (SI)。

#### [0151] 蛋白质序列

[0152] 用本发明的方法测试来自下面经良好表征的工业酶的氨基酸序列, 并进行分级排序。这些蛋白质的序列可以公开地从数据库中获得, 例如 Medline。此处描述得最详细的蛋白质包括迟缓芽孢杆菌枯草杆菌蛋白酶 (Swissprot 收录号 P29600)、BPN'Y217L (Swissprot 收录号 P00782)、**ALCALASE**<sup>®</sup> 酶 (Swissprot 收录号 P00780) 和  $\alpha$ -淀粉酶 (Swissprot 收录号 P06278)。

#### [0153] 人供体血液样品

[0154] 从两个商业来源获得志愿者供体的人血液棕黄层样品 (Stanford 血液中心, Palo Alto, CA 和 Sacramento 医学基金会, Sacramento, CA)。通过密度分离而进一步纯化棕黄层样品。将每个样品用商购得到的基于 PCR 的试剂盒 (Bio-Synthesis) 进行 HLA 分型为 HLA-DR 和 HLA-DQ。确定供体库中 HLA DR 和 DQ 的表达与北美参考标准 (Mori 等人, Transplant., 64:1017-1027[1997]) 没有显著的差异。可是, 供体库显示出对于旧金山海湾地区常见种族划分所出现的轻微富集的证据。

#### [0155] 树突细胞和 CD4<sup>+</sup>T- 细胞的制备

[0156] 通过粘附到 AIM V 培养基 (Gibco/Life Technologies) 中的塑料上而纯化出单核细胞。将粘附细胞在含有 500 单位 /ml 重组人 IL-4 (Endogen) 和 800 单位 /ml 重组人 GM-CSF (Endogen) 的 AIM V 培养基中培养 5 天。在第 5 天, 分别以 50 单位 /ml 和 0.2 单位 /ml 加入重组人 IL-1 $\alpha$  (Endogen) 和重组人 TNF- $\alpha$  (Endogen)。在第 7 天, 完全成熟的树突细胞于 37°C 用 50  $\mu$ g/ml 丝裂霉素 C (Sigma) 处理 1 小时。经处理的树突细胞用 50mM EDTA 移入 PBS 中, 然后在 AIM V 培养基中进行洗涤, 并计数, 再以  $2 \times 10^5$  细胞 /ml 重新悬浮在 AIM V 培养基中。

[0157] 通过使用 Collect CD4 柱 (Cedarlane) 从人外周血单核细胞 (PBMC) 的冷冻等试样中进行负选择而纯化出 CD4<sup>+</sup>T- 细胞。用锥虫蓝 (Sigma) 排除法进行判断, CD4<sup>+</sup>T- 细胞群体一般 > 80% 为纯的, 和 > 95% 为有活力的。将 CD4<sup>+</sup>T- 细胞以  $2 \times 10^5$  细胞 /ml 重新悬浮在 AIM V 培养基中。

#### [0158] **I-MUNE**<sup>®</sup> 检测条件

[0159] 将 CD4<sup>+</sup>T- 细胞和树突细胞置于圆底 96 孔形式的平板中, 其数量为每孔 100  $\mu$ l 每

种细胞混合物。加入肽直至 0.25-0.5% DMSO 中大约 5  $\mu$ g/ml 的最终浓度。对照细胞含有 0.5% DMSO, 而没有加入肽。每个肽分成两份进行测试。将培养物于 37°C 在 5% CO<sub>2</sub> 中培养 5 天。在第 5 天, 向每个孔中加入 0.5  $\mu$  Ci 的氟化胸苷 (NEN DuPont)。在第 6 天, 用 TomTec 手动收获器 (TomTec) 在玻璃纤维垫上收获培养物, 然后进行处理以用于闪烁计数。通过测定每一组两个一样的孔 (TriLuxBeta) 的平均每分钟计数 (CPM) 值而对增殖进行评估。该方法也在 U. S. 专利 No. 6, 218, 165 和 Stickler 等人, J. Immunother. 23 :654-660 [2000] 中进行了描述, 这两个文献在此处引用作为参考。

#### [0160] 数据分析

[0161] 对于每个单个棕黄层样品, 分析所有肽的平均 CPM 值。每个肽的 CPM 值除以对照 (只有 DMSO) 孔的平均 CPM 值, 以确定“刺激系数”(SI)。用每个肽组测试供体, 直至汇集平均每个肽至少两个响应。将每个蛋白质的数据绘制成图, 以显示出在该组内对于每个肽的响应者百分数。如果 SI 值等于或大于 2.95, 则比较阳性响应。当其接近正常群体分布中三个标准偏差的差异时, 选择出该值。对于每个进行评估的蛋白质, 汇集单个供体对于单个肽的阳性响应。为了测定对于给定蛋白质的背景响应, 将对于该组中每个肽的响应者百分数求平均值, 并计算出标准偏差。汇集每个供体对于每个肽组的 SI 值, 并报告出响应者百分数。通过将对于该组中所有肽的响应百分数求平均值而计算出对于每个肽组的平均背景响应率。用 Poisson 统计学计算出对于数据组内每个肽的响应者数目的统计学显著性。如此处的描述, 使用不同的统计学方法。如果对于肽作出响应的供体数目与数据组所定义的 Poisson 分布具有  $p < 0.05$  的差异, 则认为对于肽的响应是显著的。

#### [0162] 肽结合分析

[0163] 除了上述的 **I-MUNE**<sup>®</sup> 检测, 也进行肽结合检测。在本发明的形成期间使用的肽结合检测在本领域中是已知的 (Southwood 等人, J. Immunol. 160 :3363-3373 [1998])。简要说, 从一组用 EBV 转化的细胞系中纯化出 HLA-DR 和 HLA-DQ 分子。用经表征的标准肽和未知的肽进行竞争检测。然后测定为了竞争 50% 的标准肽结合所需的未知肽的数量 (标记为 IC<sub>50</sub>)。

#### [0164] 统计学方法 \*

[0165] 基于 Poisson 统计学计算出肽响应的统计学显著性。基于响应的总数量和组中肽的数量, 响应者平均频率用于计算 Poisson 分布。如果  $p < 0.05$ , 则认为响应是显著的。另外, 进行具有不同变量的 two-tailed Student's t 检验。对于使用具有低背景响应率的数据

的表位测定, 使用基于保守 Poisson 的公式: 
$$= 1 - e \left( -n \left( 1 - \sum \frac{\lambda^x e^{-\lambda}}{x!} \right) \right)$$
, 其中 n = 组中肽

的数目, x = 对于目标肽的响应频率, 和  $\lambda$  = 数据组内的响应频率的中值。对于基于具有高背景响应率的数据的表位测定, 使用基于严格性更小的 Poisson 的测定  $1 - \left( \sum_{i=0}^x \frac{\lambda^x e^{-\lambda}}{x!} \right)$ , 其

中  $\lambda$  = 数据组内的响应频率的中值, 和 x = 对于目标肽的响应频率。

[0166] 在另外的实施方案中, 基于下面的公式计算出结构确定值:

$$[0167] \quad \sum \left| f(i) - \frac{1}{p} \right|$$

[0168] 其中  $\Sigma$  (大写字母西格马) 为对于每个肽的响应频率减去该肽在组中的频率的绝对值之和;  $f(i)$  定义为对于单个肽的响应频率; 和  $p$  为肽组中肽的数量。

[0169] 该等式得出了 0-2 的值, 其等于“结构值”。0 的值表明结果是完全没有结构, 2.0 的值表明所有的结构是围绕单个的区域高度结构化的。越接近 2.0 的值, 蛋白质就具有越强的免疫原性。因此, 低的值表明具有较弱的免疫原性的蛋白质。

[0170] 供体库中的 HLA 类型

[0171] 分析 HLA-DR 和 HLA-DQ 类型以确定它们与对于指定的表位肽的响应之间的联系。以一个自由度进行  $\chi^2$  分析, 以便用于确定显著性。于等位基因在响应者和非响应者库两者中都存在的情况下, 计算出相对风险。

[0172] 对于大约 185 个随机个体测定 HLA-DRB1 等位基因的表达。用低严格性 PCR 测定法进行 HLA 分型。按照制造商 (Bio-Synthesis) 的指导进行 PCR 反应。将汇集的对于 Stanford 和 Sacramento 样品的数据与发表的“高加索人”HLA-DRB1 频率 (见 Marsh 等人, HLA FactsBook, The Academic 出版社, San Diego, CA[2000], 第 398 页, 图 1) 进行比较。在这些社区中的供体群体富含 HLA-DR4 和 HLA-DR15。可是, 这些群体中这些等位基因的频率很好地位于报道的这两个等位基因的范围 (对于 HLA-DR4 为 5.2-24.8%, 对于 HLA-DR15 为 5.7-25.6%,) 之内。类似地, 对于 HLA-DR3、HLA-DR7 和 HLA-DR11, 它们的频率低于平均的高加索人频率, 但是在报道的这些等位基因的范围之内。还要注意的, 发现 HLA-DR15 在大量出现于旧金山海湾地区的种族群体中具有更高的频率。

[0173] 实施例 1

[0174] 对于四种已知的呼吸性过敏原所汇集的结果

[0175] 在该实施例中, 描述了用上述本发明的 **I-MUNE<sup>®</sup>** 检测和分析方法获得的并用于测试四种已知的呼吸性过敏原的结果。

[0176] A.  $\alpha$ -淀粉酶

[0177] 在这些实验中, 用来自  $\alpha$ -淀粉酶序列的肽测试 82 个个体。对于该组中肽的背景响应为 2.80+/-3.69%, 其很好地位于 3.16+/-1.57 这一在用 11 种工业酶的测试中获得的总平均值 (数据未显示) 之内。 $\alpha$ -淀粉酶的 34-48、160-174 和 442-456 位氨基酸具有显著响应 (见图 2)。所有这三个响应相对于背景响应是高度显著的 ( $p < 0.0001$ )。

[0178] B. 迟缓芽孢杆菌枯草杆菌蛋白酶

[0179] 在这些实验中, 用该蛋白质的两个重复的肽组测试 65 个个体, 并汇集结果。发现对于该肽组的背景为 3.45+/-2.90%, 但是其位于设定的范围之内。160-174 位氨基酸具有显著响应 ( $p = 0.0003$ ) (见图 3)。

[0180] C. BPN' Y217L

[0181] 在这些实验中, 用两个肽组测试 113 个个体。对于该数据组所汇集的平均值为 3.62%。70-84 和 109-123 位氨基酸具有显著响应 (见图 4)。响应区域也存在于 154 位氨基酸周围。

[0182] D. **ALCALASE<sup>®</sup>** 酶

[0183] 在这些实验中, 用来自该酶的肽测试 92 个个体。发现对于该蛋白质的背景响应是低的 (2.35%)。同样的肽组在两个短暂间隔的分析中进行测试, 并汇集数据。另外, 在对于该蛋白质的组中, 有明显更多的肽没有产生响应。19-33 位氨基酸具有显著响应 ( $p$

< 0.0001) (见图 5)。

[0184] 实施例 2

[0185] 结构计算

[0186] 该实施例描述了对于进行测试的四种酶所获得的结构值。结构值依赖于测试个体的数量。数据组的大部分的零响应率导致~ 1.0 的结构值。对于每个肽的同样响应数量产生 0 的结构值。因此,重要的是对肽组进行测试直至数据组的大部分的响应得到累积,以便数据能够准确地反映对于特定的肽和肽区域的响应率。结构值随着测试供体数目的增加而减少,直至达到通常为每个肽 2-3 个响应的稳定水平(见图 6)。必须将稳定结构值用于比较结构值。

[0187] 对于每一个进行测试的酶,将在数据组中汇集的响应用于计算结构。结构值为:对于淀粉酶为 0.81、对于 **ALCALASE**<sup>®</sup> 酶为 0.72、对于迟缓芽孢杆菌枯草杆菌蛋白酶为 0.64 和对于 BPN' Y217L 为 0.53,这些显示在表 1 中。

[0188] 表 1. 四种呼吸性过敏原的结构测定

[0189]

酶	肽	n	每个肽的 响应	表位区域 的数目	结构值
淀粉酶	157	82	2.29	3	0.81
迟缓芽孢杆菌枯草 杆菌蛋白酶	86	65	2.24	1	0.64
<b>ALCALASE</b> <sup>®</sup>	88	92	2.16	1	0.72
<b>BPN' Y217L</b>	88	113	3.65	2	0.53

[0190] 这些结果表明,当通过导致 2.95 或更高的 SI 的增殖水平来测量 CD4<sup>+</sup>T- 细胞的活化时,与用其他肽组测量的活性相比,淀粉酶肽组引起更高的活性。对于 BPN' Y217L 的结果表明来自该蛋白质序列的肽组具有最低的活性,其具有最低的结构数量。这四种测试蛋白质的结构值分级顺序为:

[0191] 淀粉酶 > **ALCALASE**<sup>®</sup> 酶 > 迟缓芽孢杆菌枯草杆菌蛋白酶 > BPN' Y217L

[0192] 实施例 3

[0193] 与动物模型的比较

[0194] 正如上面所指明的,两种动物模型已经用于预测工业蛋白质的过敏原性和免疫原性。因此,在该实施例中,描述了在这两种动物模型和本发明的方法之间所作的比较。豚鼠 (GPIT) 和 BDF1 小鼠 (MINT) 模型这两者都将蛋白质分级为下列顺序:淀粉酶 > **ALCALASE**<sup>®</sup> 酶 > 迟缓芽孢杆菌枯草杆菌蛋白酶 > BPN' Y217L。但是相对值不同。图 7 显示了对于 GPIT (板块 A) 和 MINT (板块 B) 效价值进行作图的结构值。通过使用本发明的方法而获得的基于人细胞的结构数据表明了两种方法之间的相关性 (R<sup>2</sup> 值分别为 0.86 和 0.84)。

[0195] 实施例 4

[0196] 附加蛋白质的结构值

[0197] 在该实施例中,描述了对于附加蛋白质所获得的结构值。例如,对于 Ber e 1(即巴西坚果中发现的主要过敏原)、人干扰素- $\beta$  (IFN- $\beta$ )、人血小板生成素 (TPO)、小鼠 VH 36-60 家族成员和人  $\beta$ -2 微球蛋白计算出结构值(见表 2)。

[0198] 表 2. 选定的附加蛋白质的结构值

[0199]

	肽	n	平均背景	每个肽 的响应	表位区域 的数目	结构值
hTpo	52	99	2.56	2.54	1	0.65
hIFN- $\beta$	52	88	3.17	2.79	1	0.75
Bere1	27	92	4.27	3.92	2	0.66
小鼠 VH36-60 家族	35	74	7.0	5.23	0	0.38
B-2 微球蛋白	36	87	3.9	3.39	0	0.39

[0200] 人 IFN- $\beta$ 、TPO 和 Ber e 1 都已知可在人中引起免疫反应(见 Scagnolari 等人, J. Interferon Cytokine Res., 22 :207-213[2002]; 和 Sicherer 和 Sampson, Curr. Opin. Pediatr., 12 :567-573[2000]; 和 Li 等人, Blood 98 :3241-3248[2001])。IFN- $\beta$ 、TPO 和 Ber e 1 的结构值都是比较高的。小鼠 VH 区域的值是比较低的,这暗示了该蛋白质比较没有免疫原性。该结果与小鼠重链家族的潜在免疫原性的结构分析(见 Olsson 等人, [1991], 如上)是一致的。另外,  $\beta$ -2 微球蛋白的结果是低的,这与对于该遍在表达的蛋白质引起的耐受性是一致的(Guery 等人, [1995], 如上)。

[0201] 实施例 5

[0202] 基于群体的免疫反应

[0203] 在该实施例中,描述了用于评估基于群体的群体免疫反应的实验。如上面所指明的,从 Stanford 和 Sacramento 获得供体血液,因为该群体具有与一般的美国“高加索人”群体没有显著差异的分布。将来自这些供体血液的样品在上述的 **I-MUNE**<sup>®</sup> 检测系统中进行测试。计算出结构值,并对于在 **I-MUNE**<sup>®</sup> 检测中测试的每一个蛋白质进行比较,对于该检测每个肽有超过两个的响应。进行测试的蛋白质为 Ber e 1(巴西坚果过敏原)、scFv(抗体的单链 V 区域;VH 和 VL 片段)、BLA( $\beta$ -内酰胺酶)、IFN- $\beta$ (干扰素- $\beta$ )、FNA(枯草杆菌蛋白酶-BPN'Y217L)、 $\alpha$ -淀粉酶、E6(在 HPV 株系 16、18、31 和 33 中的人乳头状瘤病毒 E6 蛋白质)、E7(在 HPV 株系 16、18、31、33、45 和 52 中的 HPV E7 蛋白质)、eglin(水蛭蛋白酶抑制剂;GenBank 收录号 CAA25380)、RECK(人蛋白酶抑制剂;实际上是测试在 971 个氨基酸的 RECK 蛋白质 [GenBank 收录号 NP\_066934] 中的一个小结构域)、链激酶、TPO(人血小板生成素)、ecotin(来自大肠杆菌 (E. coli) K12 的丝氨酸蛋白酶抑制剂;GenBank 收

录号 NP\_416713)、**ALCALASE**<sup>®</sup>酶、洒维奈斯、人  $\beta$ -2 微球蛋白、sTNFR1 (可溶的肿瘤坏死因子受体 1)。这些实验的结果显示在表 3 中。在该表中,数据表示有多少供体对于肽组中的每个肽作出响应 (即显示出具有  $SI > 2.95$  的增殖响应)。

[0204] 表 3. 结果

[0205]

测试蛋白质	结构值	响应 / 肽	背景%
Bere1	0.66	3.93	4.26
scFv	0.39	3.96	4.9
BLA	0.56	2.62	3.27
IFN-B	0.75	2.79	3.17
FNA	0.65	3.61	3.65
淀粉酶	0.81	2.29	2.79
E616	0.72	3.92	7.12
18	0.79	2.32	4.23
31	0.53	3.26	4.66
33	0.68	1.97	2.83
E716	0.66	3.9	4.33
18	0.44	3.19	3.55
31	0.78	3.1	5.26
33	0.54	2.55	4.32
45	0.76	2.44	3.75
52	0.59	3.69	5.68
Eglin	0.43	4.9	5.57

测试蛋白质	结构值	响应 / 肽	背景%
RECK	0.39	4.1	4.64
链激酶	0.44	4.48	6.22
TPO	0.65	2.24	2.53
Ecotin	0.64	3.98	5.69
Alcalase	0.72	2.16	2.35
GG36	0.65	2.24	3.45
$\beta$ -2 微球蛋白	0.39	3.38	3.9
sTNFR1	0.47	2.9	4.2

[0206] 确定暴露的四个区域和免疫反应水平。图 10 提供了显示出对于这些蛋白质的相对结构值和背景百分数的图。象限“1”反映了群体中未接触测试蛋白质的个体的数量（用 1 数据点 [◆] 来代表），而象限“2”反映了群体中已经接触了测试蛋白质的个体的数量，其中接触是更新近的、频繁的和 / 或在性质上更具有免疫原性的，象限“3”反映了群体中对于测试蛋白质具有耐受性和 / 或测试蛋白质没有免疫原性的个体的数量，和象限“4”反映了群体中已经接触了测试蛋白质（但是这一接触是很久以前发生的或者是不频繁地发生的）和 / 或测试蛋白质在性质上具有较弱免疫原性的个体的数量。

[0207] 象限“1”中的蛋白质为 BLA、IFN- $\beta$ 、FNA、淀粉酶、HPV33E6、HPV45E7、TPO、**ALCALASE**<sup>®</sup>酶和 GG36；而象限“2”中的蛋白质为 Ber e 1、HPV16E6、HPV18E6、HPV31E6、HPV16E7、HPV31E7、HPV33E7、HPV45E7、HPV52E7 和 ecotin；象限“3”中的蛋白质为 HPV18E7 和  $\beta$ -2 微球蛋白；和象限“4”中的蛋白质为 scFv、eglin、RECK、链激酶和 sTNFR1。因此，清楚的是，本发明提供了用于评估任何蛋白质的基于群体的免疫反应的方法。

[0208] 实施例 6

[0209] 形成具有减少的结构值的变体

[0210] 在该实施例中，提供了用于形成具有减少的结构值的变体的方法。如何将结构分析用于计算设计为减弱了在人中的免疫原性的变体蛋白质的总免疫原性，作为这样的例子，对于变体计算出结构值，该变体中对于 BPN' Y217L 中的 70-84 和 109-123 位氨基酸的显著响应减弱为背景水平的响应。用 BPN' Y217L 的 70-84 和 109-123 区域的肽变体测试 48 个个体的有限数据组。发现对于变体的响应位于背景水平。通过将对于 70-84 和 109-123 的响应减弱为背景水平而对 113 个个体的完整数据组进行修饰，以用于结构计算。以这种方式计算出结构，以便预测如果 113 个个体用亲本分子进行测试，那么结构值将会是多少。因为从计算中去除了响应，所以将相同数量的响应随机分散于数据组中，以保持同样的总

响应率。计算出经修饰的蛋白质变体的结构值为 0.40 (见表 4)。

[0211] 表 4. 潜在蛋白酶变体的结构计算值

[0212]

蛋白酶	显著表位	结构值
BPN' Y217L	2	0.53
BPN' 变体	0	0.40

[0213] 另外,体外数据表明具有较低结构值的蛋白酶变体引起较少的增殖。在这些实验中,来自 30 个社区供体的 PMBC 用完整的蛋白质亲本酶 (BPN' Y217L) 或者变体蛋白酶以不同的参数进行测试。将酶失活,并在 5-40  $\mu$ g/ml 的范围内进行测试。对于每个蛋白质所达到的最大 SI 值显示在图 9 中。亲本蛋白酶具有 0.53 的结构值,变体具有 0.40 的结构值。在这 30 个供体上进行测试的两个蛋白质的最佳 SI 值之间的差异是显著的,其具有  $p < 0.01$  的 two-tailed 参量 t 检验值。这些结构表明,将结构值从 0.53 减少至 0.40 对于分子的体外抗原性具有很大的影响。

[0214] 在本发明的优选方法中,当将变体蛋白质与亲本蛋白质在体外或体内进行比较时,优选地将这些蛋白质以同样的剂量、在同样的配剂中、在一组相匹配的供体中和经过同样的给药曲线进行比较。变体蛋白质应当保持了亲本蛋白质的一般物理和结构特性,例如稳定性和活性。另外,结构分析排除了亲本蛋白质与其变体之间的加工差异。

[0215] 实施例 7

[0216] CD4<sup>+</sup>T- 细胞表位的标明

[0217] 在该实施例中,给出了来自未暴露和暴露的供体的数据。除了上面实施例中的数据之外还提供了这些数据。

[0218] 未暴露的供体

[0219] 65 个供体用一组经合成而覆盖迟缓芽孢杆菌枯草杆菌蛋白酶序列的 15 聚肽进行测试。该 65 个供体对于每个肽的响应百分数显示在图 12 的板块 A 中。相应于 160-174 位氨基酸的 #54 位置上的显著响应是明显的。在 23 和 31 肽位置 (67-81 和 91-105 位氨基酸) 上的另一个显著的区域也是明显的。对于组中的肽的响应频率显示在图 12 的板块 B 中。清楚的是,在 160-174 位氨基酸上对于肽的响应频率不同于对于该组中其他肽的响应频率。可是,必须测定在 67-81 和 91-105 位氨基酸上的响应显著性。显著性可通过下面的方法测定:建立频率数据的 Poisson 分布,然后确定这样的可能性,即含有由组中肽的数目代表的值数目的数据组将包括该值作为其最高的成员的可能性。对于 160-174 位氨基酸代表的肽,该可能性为  $p = 0.0004$ 。对于其他两个肽,该可能性为  $p = 0.50$ 。

[0220] 作为表位选择标准的试验,将经用皮肤针刺测试证实已经接触了迟缓芽孢杆菌枯草杆菌蛋白酶的一组 7 个供体也用此处描述的 I-MUNE<sup>®</sup>检测系统进行测试。显示了所有 7 个供体对于每个肽的响应数目 (见图 13)。发现只有一个肽引起超过两个响应。对于 163-177 位氨基酸肽的三个响应者包括两个 HLA-DR2(15) 阳性供体。以前注意到了对于该肽的响应与 HLA-DR2(15) 之间的联系 (Stickler 等人, J. Immunother., 23: 654-660 [2000])。有两个供体对于包括 67-81 区域在内的 6 个肽区域作出了响应。来自暴

露供体数据的其他肽在未暴露供体数据中是不显著的。67-81 区域与相关蛋白酶中已知的 CD4<sup>+</sup>T<sup>-</sup> 细胞表位具有高度的同源性 (14/15 个氨基酸的同一性), 这些供体的一半对于该第二蛋白酶也是 SPT<sup>+</sup> 的。因此, 作为保守的估计, 在未暴露的供体群体中发现一个经证实的表位, 而且发现该表位在一组被经证实暴露于蛋白质的供体所识别的表位中是显著的。

[0221] 对于来自解淀粉芽孢杆菌的另一种相关枯草杆菌蛋白酶也观察到类似的结果。描述了高度显著的两个显著表位区域, 也在一组经证实 SPT<sup>+</sup> 的供体中发现这两个表位 (数据未显示)。如上所述, 在来自经暴露的供体的汇集数据中见到了更显著的表位区域, 并且在未暴露的供体组中所限定的表位肽是这些的一个亚组。

[0222] 记忆响应

[0223] 对于来自链激酶序列的一组肽施行上述的 **I-MUNE**<sup>®</sup> 检测。选择链激酶用于这些实验是由于这样的事实, 即一般的群体在一段时间内积聚对于该蛋白质的特异性响应 (见 Warmerdam 等人, J. Immunol., 168:155-161 [2002])。用该蛋白质在本发明的 **I-MUNE**<sup>®</sup> 检测系统中测试一组 72 个社区供体。对于链激酶组中的肽的响应显示于图 14 的板块 A 中。在链激酶数据组中没有明显显著的响应。这一点清楚地显示在频率数据中 (见图 4, 板块 B), 其中不像迟缓芽孢杆菌枯草杆菌蛋白酶的数据, 没有单个肽积聚一定比率的响应, 其中该比率明显不同于对于其他肽的响应的分布。可是, 位于位置 5 (13-27 位氨基酸)、20 和 21 (58-75 位氨基酸)、29 (85-99 位氨基酸) 和 36 (106-120 位氨基酸) 的显著响应率是重要的。数据组显示出每个肽 4.48 个响应的这一平均响应值 (背景 = 6.22%; 见表 5, 下面)。如果将该值用于限定 Poisson 分布的中值, 那么保守性较小的分析表明由所有上面概括的显著肽所显示的响应频率是显著的 ( $p < 0.05$ )。该分析比用于确定在未暴露供体中发现的表位的显著性的分析具有小得多的保守性, 因为 Poisson 分布是由背景值的中值限定的, 并且来自该值的差异用于确定显著性。

[0224] 表 5. 推测为对供体进行预暴露的蛋白质的背景值

[0225]

	测试的供体	期望的响应 / 肽 e <sup>b</sup>	响应 / 肽 e 测得 <sup>c</sup>	背景 +/- sd <sup>d</sup>	t 检验 <sup>e</sup>
11 种工业酶	n. a. <sup>a</sup>	n. a.	n. a.	3.15 +/- 1.57	n. a.
HPV16E6	55	1.65	3.92	7.12 +/- 6.48	P = 0.0003
HPV18E6	55	1.65	2.32	4.23 +/- 4.25	P = 0.16
Bere1	92	2.77	3.92	4.26 +/- 4.05	P = 0.22
链激酶	72	2.17	4.48	6.22 +/- 3.47	P = 0.0001
IFN-β	88	2.65	2.79	3.17 +/- 3.28	n. d. <sup>f</sup>

	测试的供体	期望的响应 / 肽 e <sup>b</sup>	响应 / 肽 e 测得 <sup>c</sup>	背景 +/- sd <sup>d</sup>	t 检验 <sup>e</sup>
TPO	99	2.99	2.51	2.54+/-2.23	n. d.
TNF-R1	69	2.08	1.54	223+/-1.95	n. d.

[0226] 在该表中，“a”表示“不可应用的”；“b”表示基于来自显示在图 11 中的 11 种工业蛋白质的数据而对于测试供体的数目所期望的每个肽的响应数目；“c”表示通过实验测定的对于测试蛋白质来说每个肽的响应值；“d”表示对于测试蛋白质的背景响应值；“e”表示将 11 种工业酶的背景值与测试蛋白质的背景响应进行比较而应用的 two-tailed 不等变量 t 检验；和“f”表示“未测定”。

[0227] 将在 **I-MUNE**<sup>®</sup> 检测中鉴定的五个表位肽与发表的表位进行比较，这些发表的表位是通过使用来自具有对链激酶的抗原特异性响应的供体的无性繁殖 CD4<sup>+</sup>T- 细胞系而确定的（见图 15）。

[0228] 用来自 10 个供体的无性繁殖 T- 细胞确定的区域即 D1、F2、C3、和 D4 含有核心序列（大多数响应克隆之间的共有肽序列），所述核心序列分别相当于 **I-MUNE**<sup>®</sup> 检测鉴定的肽 5、20、21 和 36。**I-MUNE**<sup>®</sup> 检测在位置 29（85-99 位氨基酸）上鉴定了一个表位肽，其是用 CD4<sup>+</sup>T- 细胞克隆未曾检测到的。该肽与 HLA-DR5（11）的存在相联系。只有一个对于 CD4<sup>+</sup>T- 细胞克隆研究提供了克隆的供体带有该等位基因，因此其可能被遗漏了。还有一种可能，该肽可能不是从链激酶中加工出来的，因此该结果在 **I-MUNE**<sup>®</sup> 检测数据组中可能是假阳性。可是，据以前报道，在蛋白质的羧基端的区域 A5 被 T- 细胞克隆所识别（见 Warmerdam 等人，如上）。**I-MUNE**<sup>®</sup> 检测在该区域的一个亚组中定位了一个表位即肽 36，其相当于邻近的 D4 区域。总的来说，在用所描述的保守性较小的表位标明法所发现的表位与发表的表位之间进行对比，其结果是很好的。另外，所报道的 HLA 相关性在两个数据组之间是一致的（见图 15）。

[0229] 阴性对照

[0230] 作为阴性对照，用来自 87 个社区供体的样品也在 **I-MUNE**<sup>®</sup> 检测中测试人  $\beta$ -2 微球蛋白。选择该蛋白质作为阴性对照是因为其作为 I 型 HLA 分子的一部分而存在于所有体细胞的表面上。另外， $\beta$ -2 微球蛋白于 T- 细胞发育期间在胸腺中表达。中央和外周耐受性机制都应当会影响 T- 细胞库，从而去除任何与  $\beta$ -2 微球蛋白来源的肽有明显交叉反应的 CD4<sup>+</sup>T- 细胞（见 Guery 等人，J. Immunol., 154 :545-554 [1995]）。最后，在该分子中有最小的等位基因变化。在数据库搜索中发现一个等位基因变体（未显示）。结果显示于图 16 中。对于  $\beta$ -2 微球蛋白的平均背景响应为 3.90+/-1.82%。对于肽的响应百分数显示于图 16 的板块 A 中，响应频率显示于图 16 的板块 B 中。基于统计学方法，对于具有低背景响应率的未暴露供体群体来说，没有一个肽响应是显著的。

[0231] 响应率的可重现性

[0232] 通过表位肽的重复测试来确定表位肽响应的可重现性。肽至少合成两次，并在多

个分立的供体组上进行测试。对于每个测试所用的供体数量从 27 至 103 个供体变化。比较对于肽的平均响应百分数。结果显示在表 6 中。四种表位肽的平均变异系数 (CV) 为 20%，中值为 21%。CV 的范围为 9.3-27%。这些值可以有利地与其他基于人细胞的体外检测 (Keilholz 等人, *J. Immunother.*, 25 :97-138[2000] ;和 Asai 等人, *Clin. Diagn. Lab. Immunol.*, 7 :145-154[2000]) 进行比较。在表 6 中,“s. d.” 是标准偏差,“s. e.” 是标准误差,“s. d. / 平均 \* 100” 是 CV 百分数。显示了对于四种肽的平均值和中值。

[0233] 表 6. 表位肽响应的可重现性

[0234]

	测试数目	平均值	s. d.	s. e.	% CV
IFN-B	3	16.41	1.53	0.88	9.32
TPO	3	9.18	1.83	1.06	19.99
BPN' Y217L#24	4	11.69	2.71	1.35	23.18
BPN' Y217L#37	4	12.91	3.51	1.76	27.19
				对于所有 的平均值	19.92
				中值	21.59

[0235] 用结合研究确认的表位

[0236] 对于在两个相关工业细菌蛋白酶中定义的肽表位来测定对于 II 型 HLA 结合的  $IC_{50}$  (见图 17)。将肽在与 18 种不同的 HLA-DR 和 -DQ 蛋白质结合的竞争检测中进行测试。发现迟缓芽孢杆菌枯草杆菌蛋白酶中的显著表位以两个不同的框架 (160-174 和 157-171) 结合一定范围的 HLA-DR 和 -DQ 分子,这表明有着混杂的结合。发现与 HLA-DR2(15) 结合的肽是极好的,其具有 127nM 的  $IC_{50}$ 。只有 HLA-DR1 显示了较低的  $IC_{50}$  值。在由 **I-MUNE**<sup>®</sup> 检测于解淀粉芽孢杆菌枯草杆菌蛋白酶 BPN' Y217L 中所定义的两个表位之中,发现第二个表位 (109-123 位氨基酸) 在 HLA 分析和该实施例所描述的结合分析中都是混杂的。第一个表位 (70-84 位氨基酸) 也结合大多数进行测试的 II 型 HLA 分子,但是其以 0.69nM 的  $IC_{50}$  结合 HLA-DR6(13)。这可能解释了在数据库中所看到的对于该肽的响应与供体的联系 ( $p = 0.00015$ ; 相对风险 = 7.22,  $n = 113$  个测试供体)。认为那些具有小于 500nM 的值的结果为良好的结合物,并在图 17 中以粗体加以强调。还有,在该图中,简并性指明了 18 个进行测试的全部等位基因中 II 型 HLA 蛋白质的数目,它们以低于 500nM 的  $IC_{50}$  进行结合。

[0237] 序列表

[0238] <110>Genencor International, Inc.

[0239] Harding, Fiona

- [0240] <120> 将蛋白质相对免疫原性进行分级的基于群体的评估方法和手段
- [0241] <130>GC721-2-PCT
- [0242] <140>PCT/US03/05670
- [0243] <141>2003-02-26
- [0244] <150>US 60/260,057
- [0245] <151>2002-02-26
- [0246] <160>1
- [0247] <170>FastSEQ for Windows Version 4.0
- [0248] <210>1
- [0249] <211>4
- [0250] <212>PRT
- [0251] <213> 人工序列
- [0252] <220>
- [0253] <223> 肽检测序列
- [0254] <400>1
- [0255] Ala Ala Pro Phe
- [0256] 1

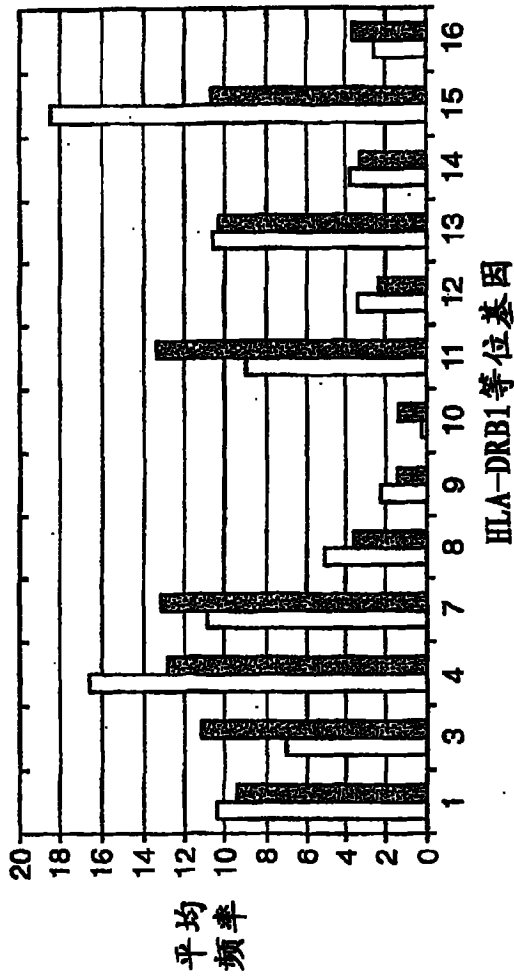


图 1

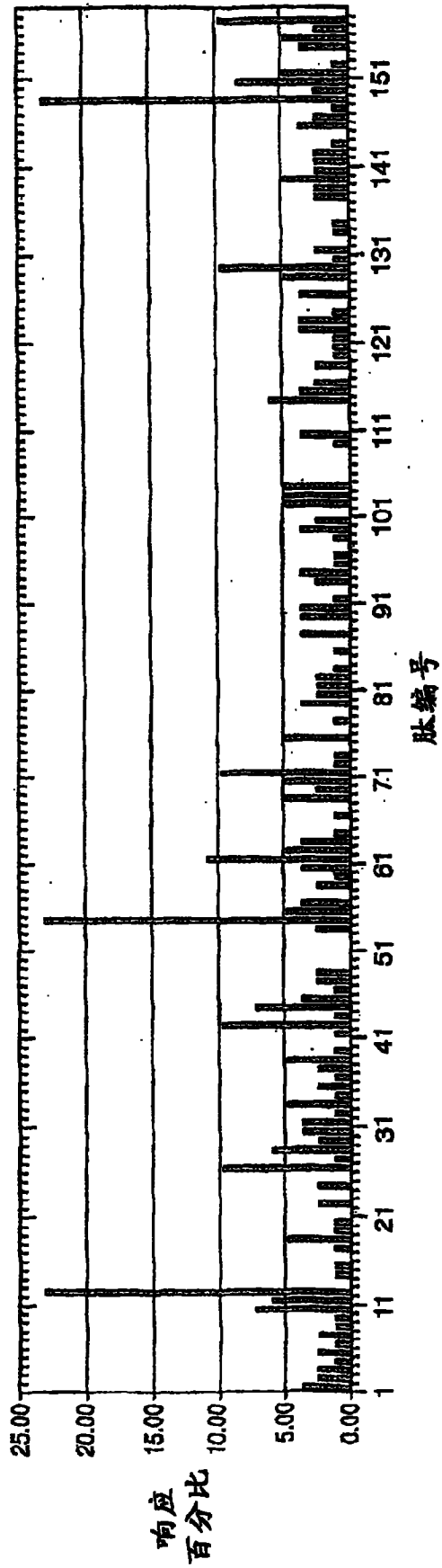


图 2



图 3

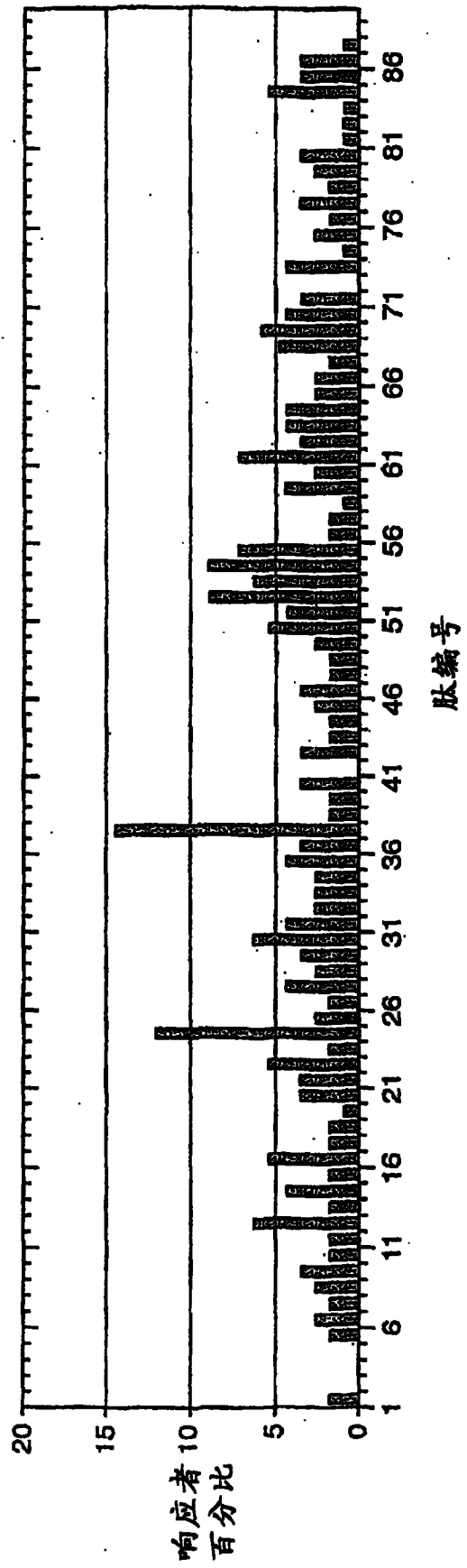


图 4

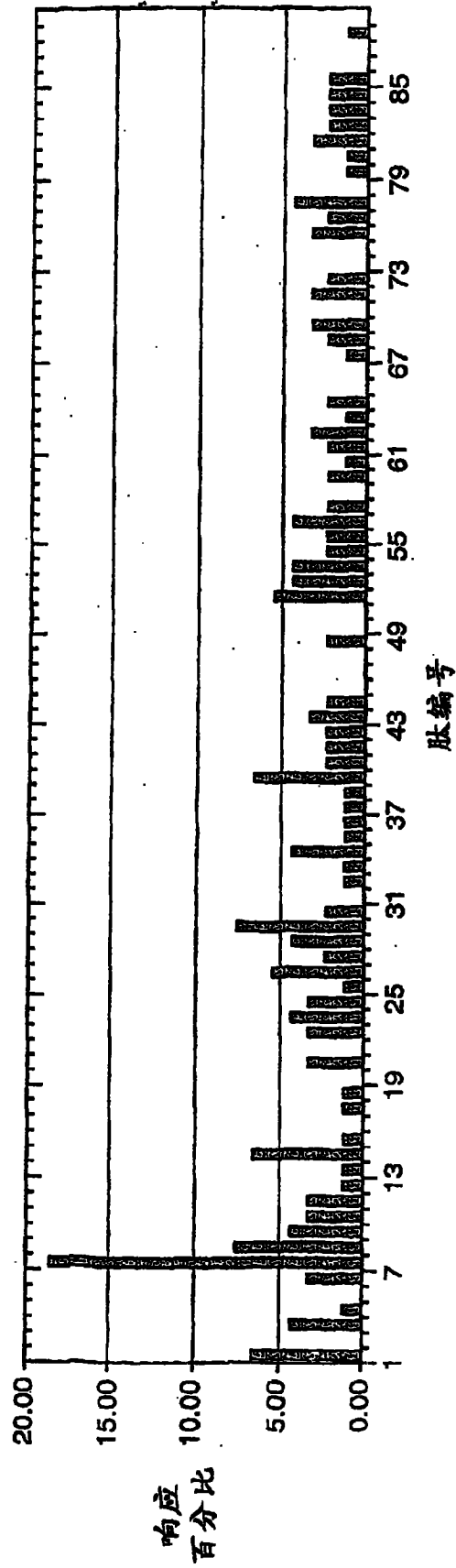


图 5

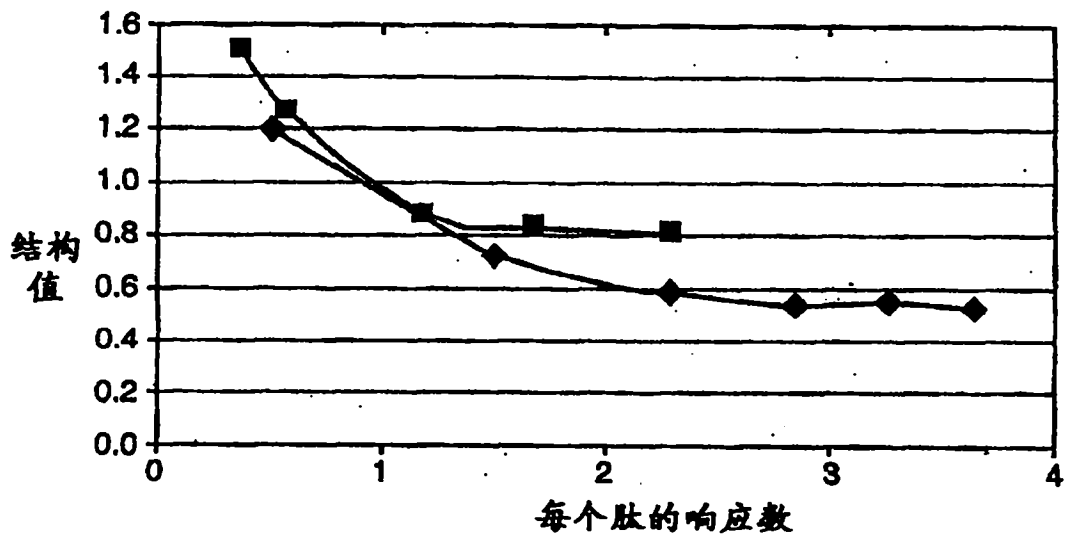


图 6

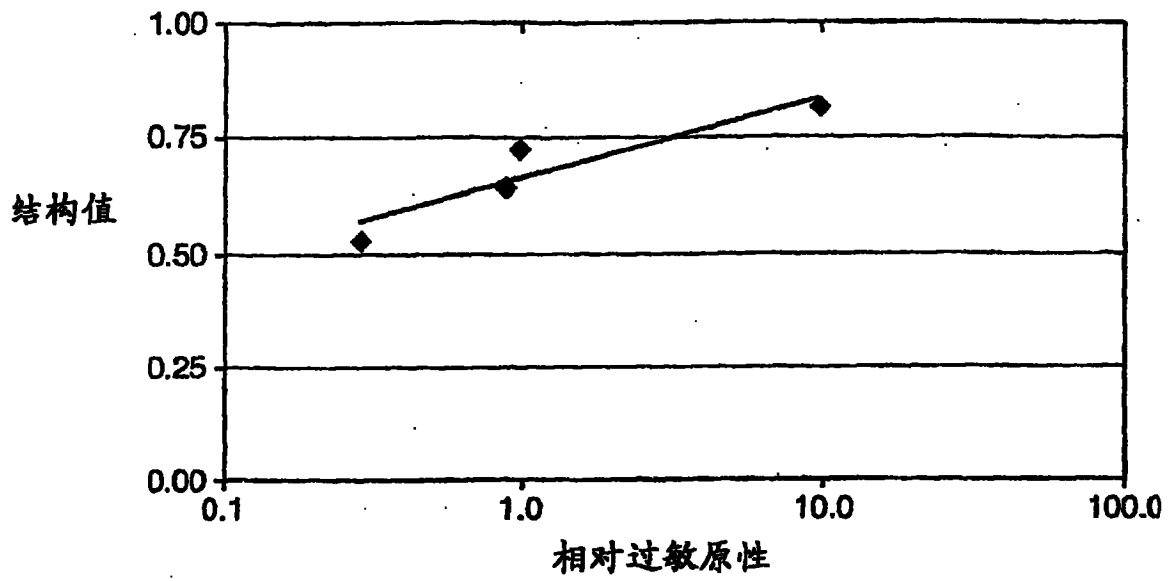


图 7A

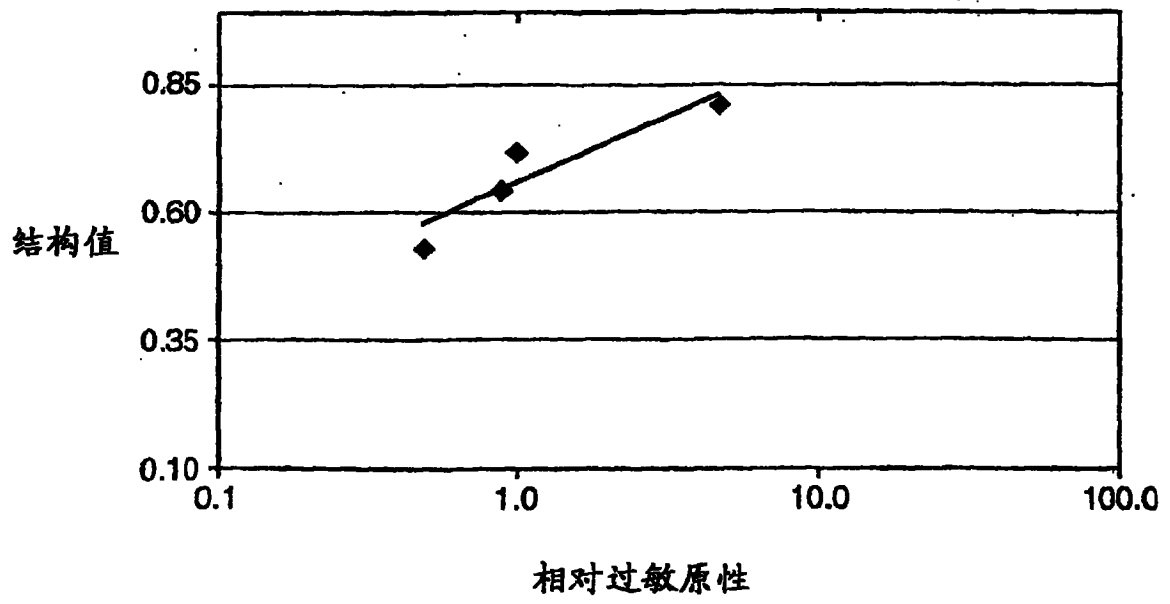


图 7B

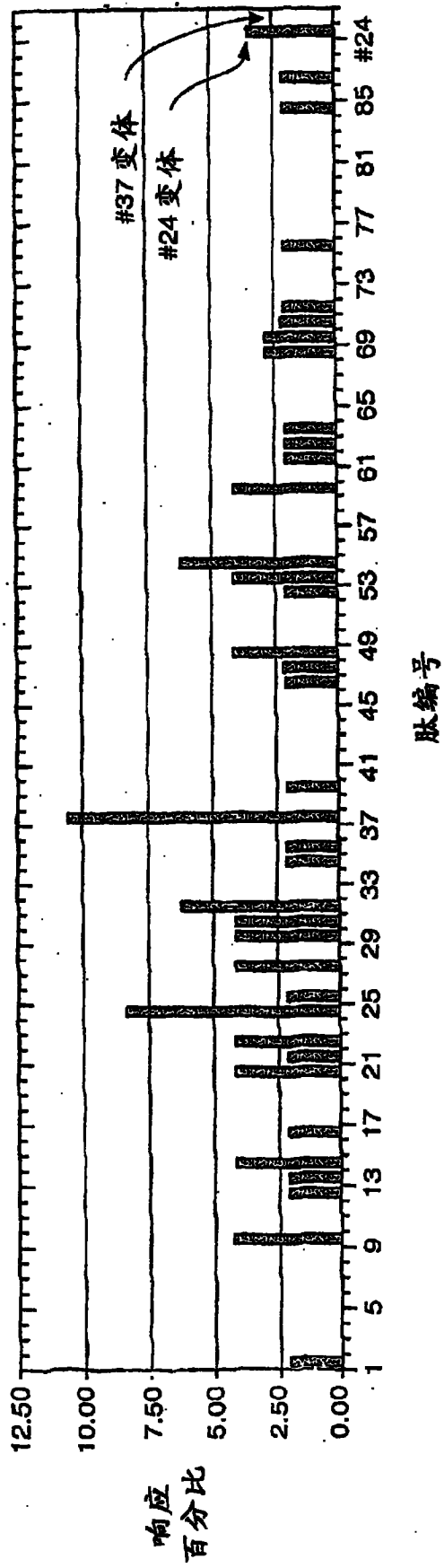


图 8

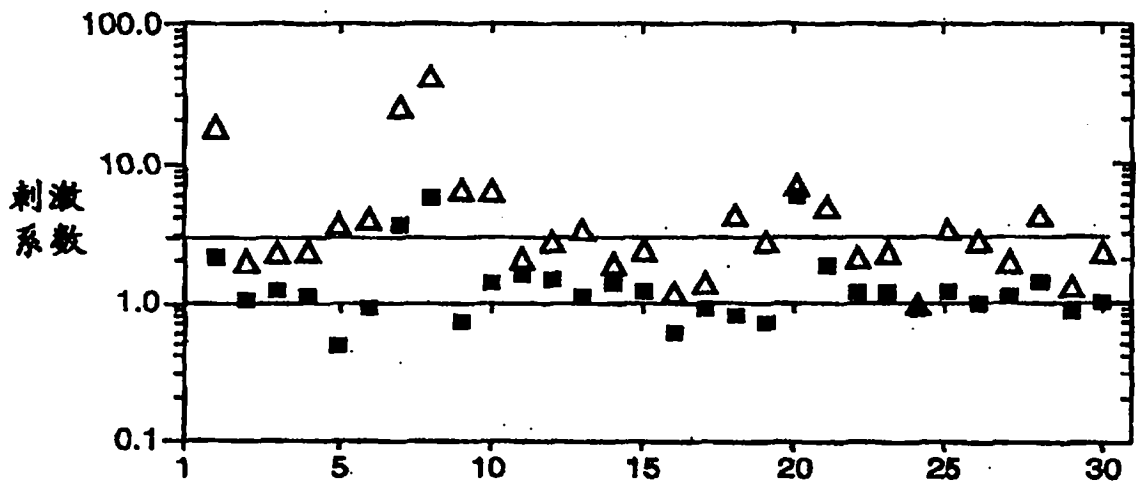


图 9

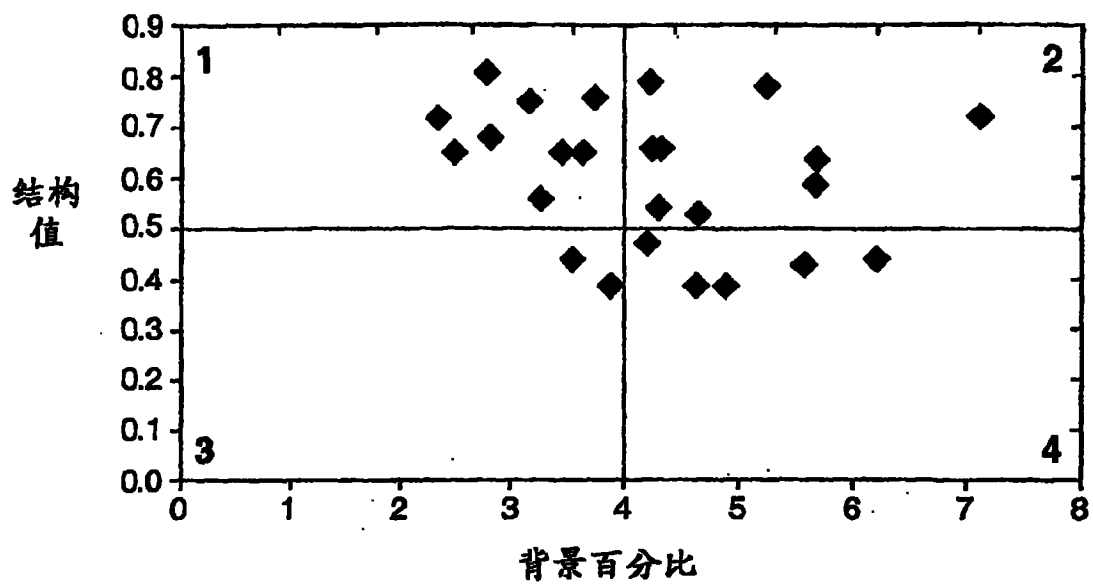


图 10

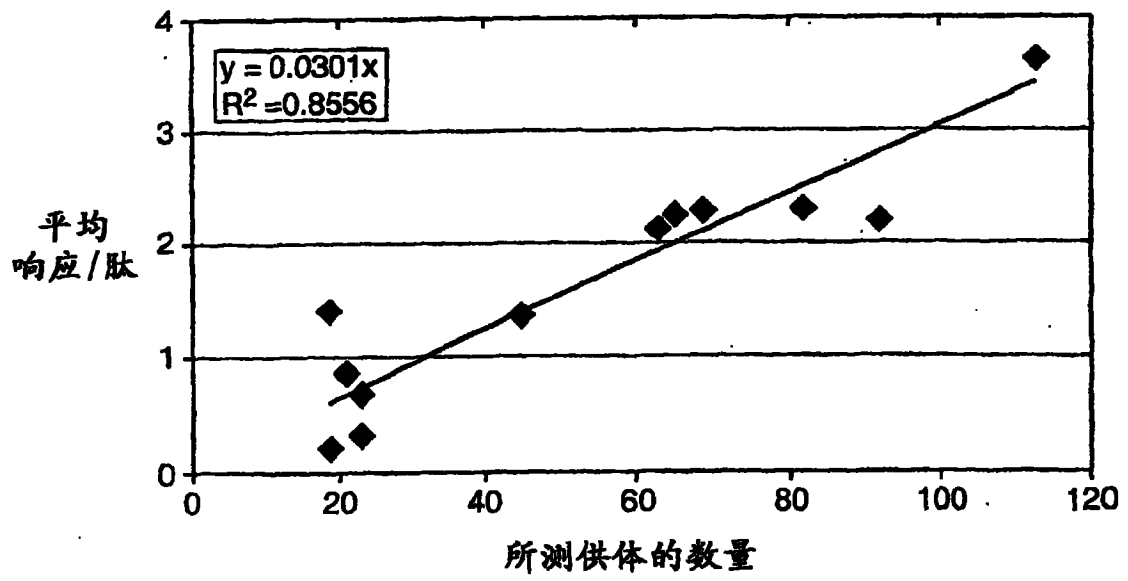


图 11

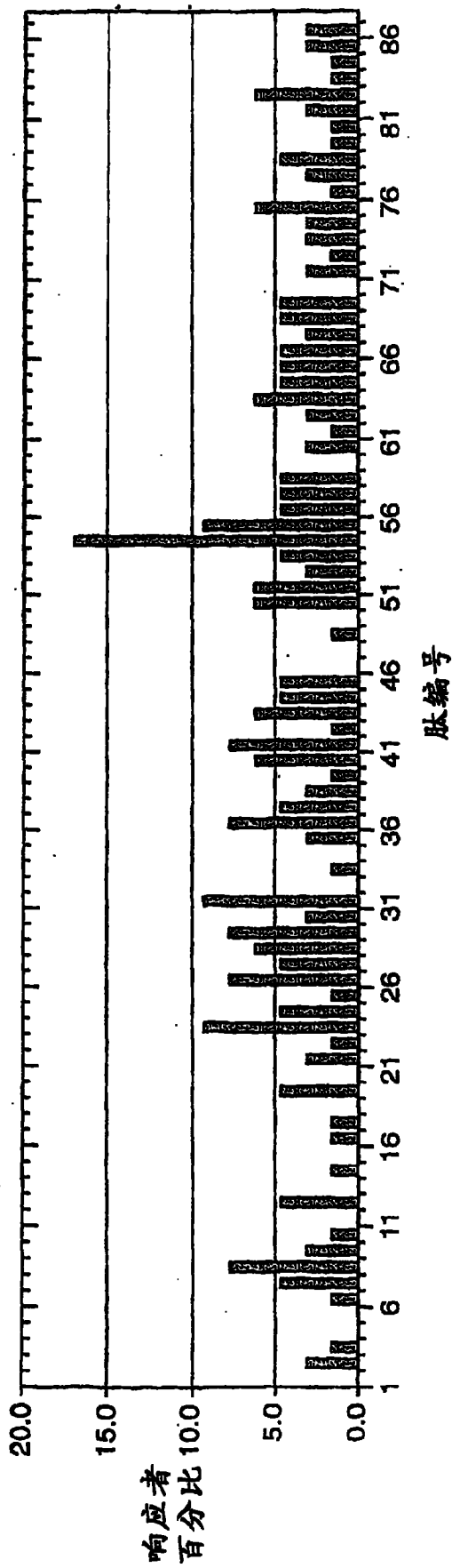
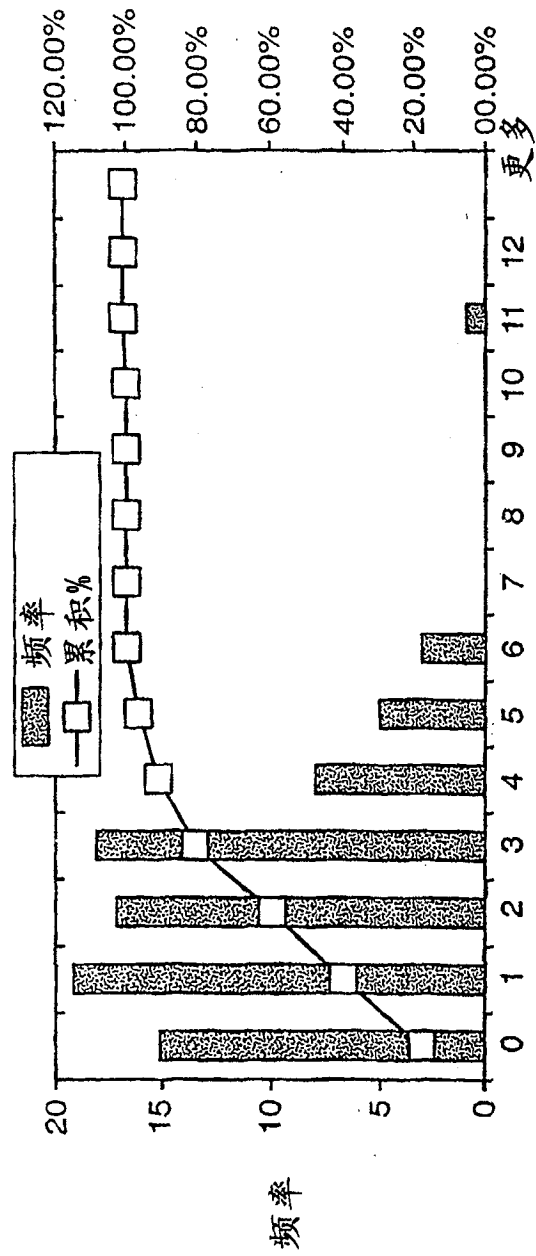


图 12A



对于每个迟缓芽孢杆菌枯草杆菌蛋白酶肽的响应

图 12B

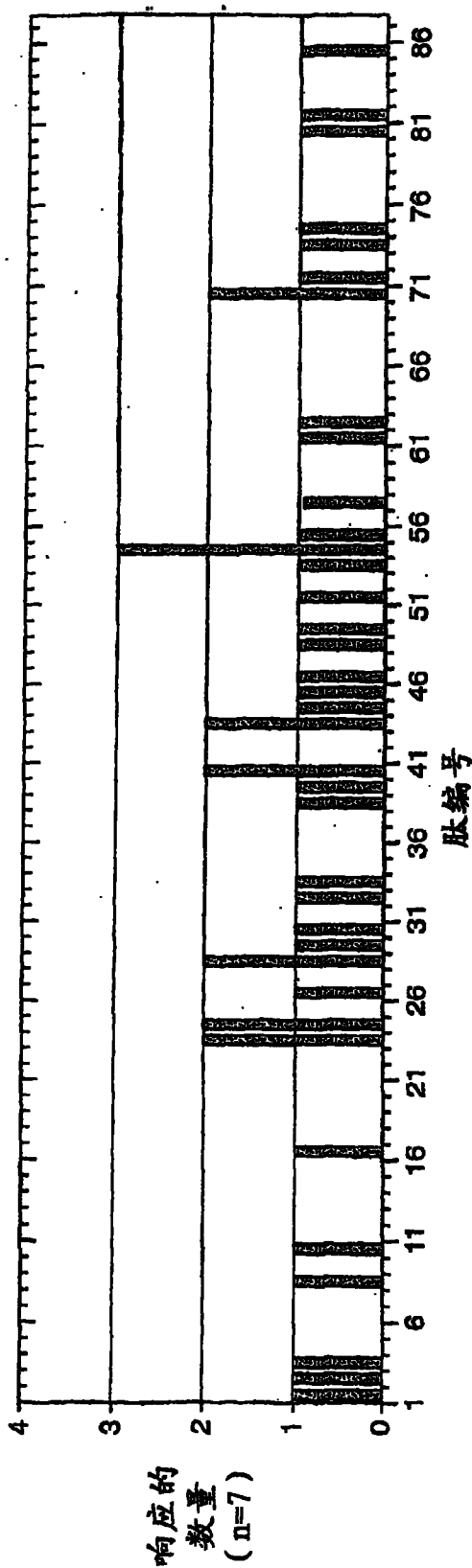


图 13

munc<sup>®</sup>HLA关联性  
链激酶肽

	HLA 关联性	p	HLA 增加的	HLA 降低的	Warmerdam, et al (2002)* 区域
5			DQ2		D1 DR3, 7, 2
20	DR2 (15)	p < 0.03	DQ6	DR7	F2 DR1, 4
21			DR6 (13)		C3 DR2, 3, 4, 6, 7, 8
29	DR5 (11)	p < 0.01			n.f.#
36			DR3		D4 DR3, 6
n.f.#					A5 DR2, 5, 6, 8

图 15

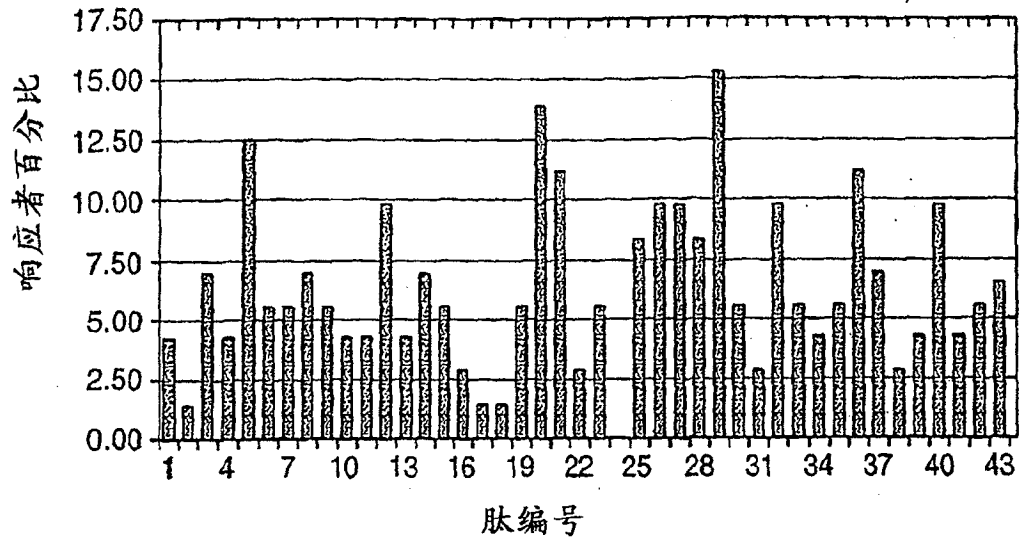


图 14A

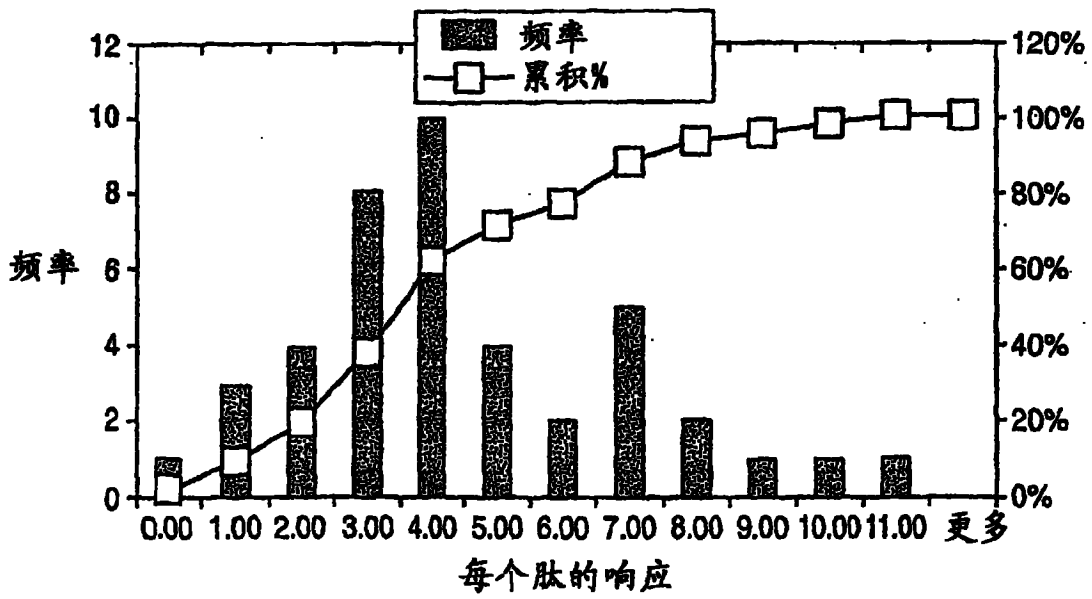


图 14B

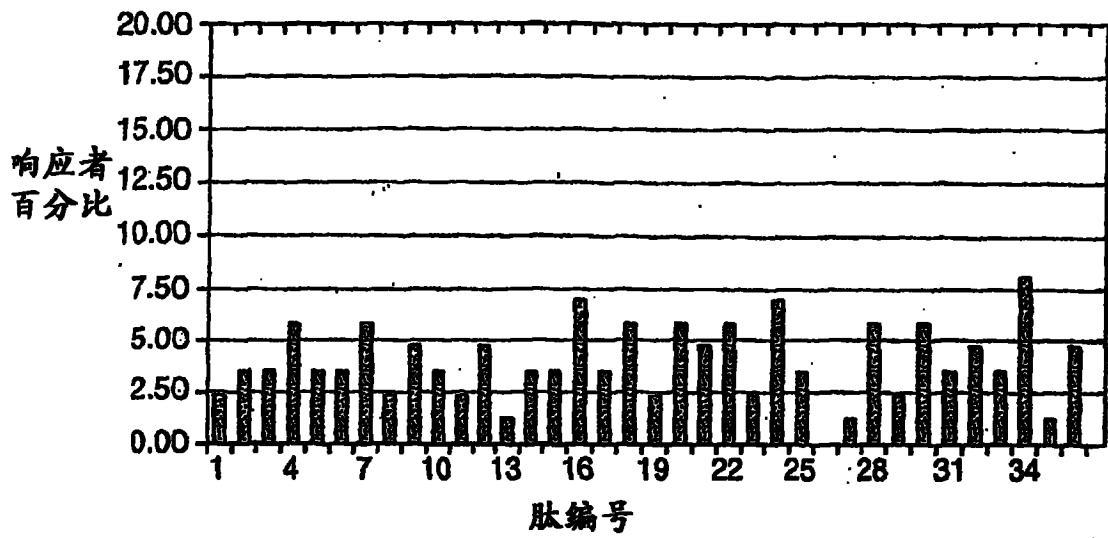


图 16A

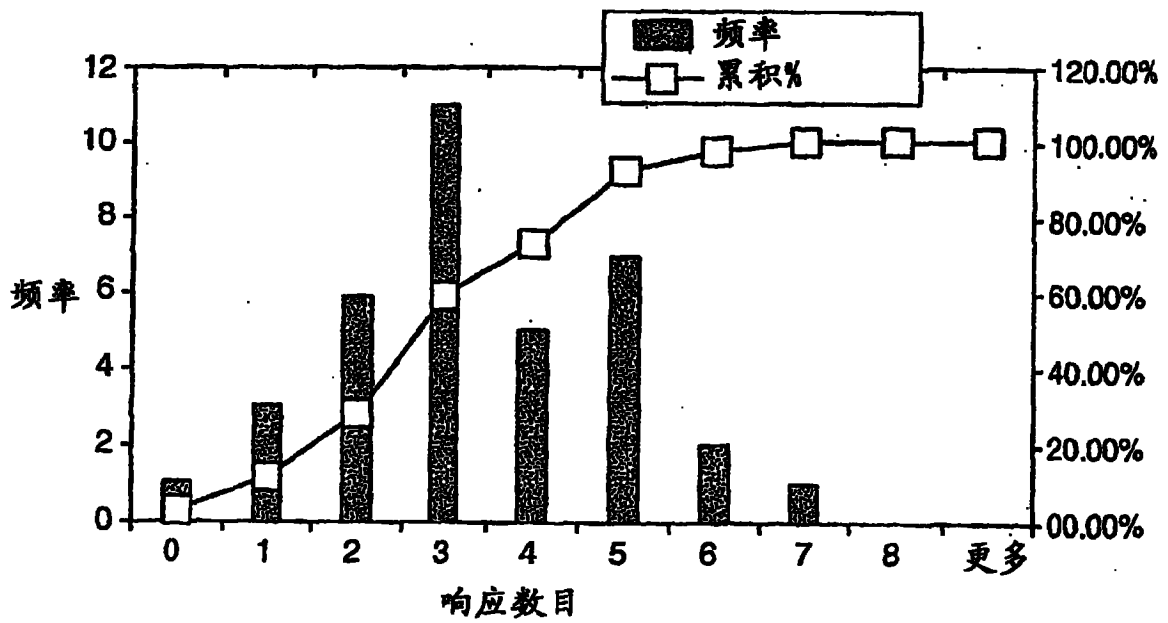


图 16B

与纯化的HLA结合的nM IC<sub>50</sub>

蛋白质	DRB1										多样性
	*0101 (DR1)	*0301 (DR3w17)	*0401 (DR4w4)	*0404 (DR4w14)	*0405 (DR4w15)	*0701 (DR7)	*0802 (DR8w2)	*0901 (DR9)	*1101 (DR5w11)	*1201 (DR5w12)	
BPN' Y217L.70	6.5	8737	33	5.7	166	154	1711	46	2382	80	
BPN' Y217L.109	8.8	-	30	166	37	58	2192	43	3019	1235	
远缘芽孢杆菌157	1065	16,433	4794	7575	6784	724	>16,333	1484	--	--	
远缘芽孢杆菌160	13	-	142	5542	1348	138	2033	164	5554	--	

DRB1	DRB3/4/5			DQ				多样性
	*1501 (DR2w2β1)	*DRB3*0101 (DR52a)	*DRB4*0101 (DRw53)	*DRB5*0101 (DR2w2β2)	DQA1*0501/ DQB1*0201 (DQ2)	DQA1*0301/ DQB1*0301 (DQ3.1)	DQA1*0301/ DQB1*0302 (DQ3.2)	
*1302 (DR6w19)	21	2010	31	15,689	670	440	2069	n/18
0.69	683	119	1071	1024	97	2182	80	12
9.8	865	>9434	>9667	--	6157	6009	5009	11
2009	127	6157	8257	1726	1296	63	1046	2
559								7

图 17

专利名称(译)	将蛋白质相对免疫原性进行分级的基于群体的评估方法和手段		
公开(公告)号	<a href="#">CN1639571B</a>	公开(公告)日	2010-05-26
申请号	CN03804559.1	申请日	2003-02-26
[标]申请(专利权)人(译)	金克克国际有限公司		
申请(专利权)人(译)	金克克国际有限公司		
当前申请(专利权)人(译)	金克克国际有限公司		
[标]发明人	FA哈丁		
发明人	F·A·哈丁		
IPC分类号	G01N33/53 G01N33/567 C12N5/06 C12N9/56 C12Q1/02 G01N33/50 G01N33/566		
CPC分类号	G01N2333/96433 G01N33/505 C12N9/54		
代理人(译)	程泳		
优先权	60/360057 2002-02-26 US		
其他公开文献	CN1639571A		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

本发明提供了评估群体的免疫反应特性的方法。特别地，本发明提供了定性评估人类群体的免疫反应的方法，其中对于针对任何目标蛋白质的免疫反应进行分析。本发明进一步提供了基于蛋白质的相对免疫原性来将蛋白质进行分级的方法。另外，本发明提供了方法来创造在多种应用中使用的具有减弱的免疫原性的蛋白质。

