

[19] 中华人民共和国国家知识产权局

[ 51 ] Int. Cl<sup>7</sup>  
C07K 14/50  
A61K 38/18 C07K 7/08



[12] 发明专利申请公开说明书

[21] 申请号 02804610.2

[43] 公开日 2004 年 4 月 21 日

[11] 公开号 CN 1491231A

[22] 申请日 2002.2.5 [21] 申请号 02804610.2  
[30] 优先权  
[32] 2001. 2. 6 [33] EP [31] 01102574.9  
[32] 2001. 2. 19 [33] EP [31] 01103954.2  
[86] 国际申请 PCT/EP02/01175 2002.2.5  
[87] 国际公布 WO02/062842 英 2002.8.15  
[85] 进入国家阶段日期 2003.8.6  
[71] 申请人 默克专利有限公司  
地址 德国达姆施塔特  
[72] 发明人 F·J·卡尔 G·卡特 T·琼斯  
S·威廉斯

[74] 专利代理机构 北京市中咨律师事务所  
代理人 黄革生 隗永良

权利要求书 4 页 说明书 27 页

[54] 发明名称 免疫原性减弱的经修饰角质形成细胞生长因子(KGF)

[57] 摘要

本发明涉及尤其是用于对人施用的、特别是用于治疗的多肽。所述的多肽是经修饰的多肽，其中，所述的修饰使得上述多肽在施用于人体时引起免疫应答的倾向减弱。本发明特别涉及对角质形成细胞生长因子(KGF)进行修饰以产生角质形成细胞生长因子(KGF)蛋白，该蛋白在体内使用时基本上无免疫原性，或比未经修饰的相应蛋白的免疫原性低。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种经修饰的分子，其具有角质形成细胞生长因子(KGF)的生物学活性，且当其在体内应用时基本上无免疫原性或免疫原性低于任何具有相同生物学活性但未经修饰的分子。

2. 如权利要求 1 所述的分子，其中，所述的免疫原性丧失是通过从原始的未经修饰的分子中去除一个或多个 T-细胞表位而实现的。

3. 如权利要求 1 或 2 所述的分子，其中，所述的免疫原性丧失是通过减少能与由所述分子衍生的肽相结合的 MHC 同种异型的数量来实现的。

4. 如权利要求 2 或 3 所述的分子，其中，去除了一个 T-细胞表位。

5. 如权利要求 2-4 中任意一项所述的分子，其中，原本存在的 T-细胞表位是 MHC II 类配体，或表现出经 II 类分子呈递作用后有刺激或结合 T-细胞的能力的肽序列。

6. 如权利要求 5 所述的分子，其中，所述的肽序列选自如表 1 所示的组。

7. 如权利要求 2-6 中任意一项所述的分子，其中，任何原本存在的 T-细胞表位中的 1-9 个氨基酸残基发生了改变。

8. 如权利要求 7 所述的分子，其中，有一个氨基酸残基发生了改变。

9. 如权利要求 7 或 8 所述的分子，其中，所述氨基酸残基的改变是用其他的氨基酸残基在特定的位置替代原本存在的氨基酸残基。

10. 如权利要求 9 所述的分子，其中，按表 2 所示进行一个或多个氨基酸残基的替代。

11. 如权利要求 10 所述的分子，其中，另外还按表 3 所示进行一个或多个氨基酸残基的替代以减少能与由所述分子衍生的肽相结合的 MHC 同种异型的数量。

12. 如权利要求 9 所述的分子，其中，按表 3 所示进行一个或多个氨基酸的替代。

13. 如权利要求 7 或 8 所述的分子, 其中, 所述氨基酸残基的改变是在特定的位置缺失原本存在的氨基酸残基。

14. 如权利要求 7 或 8 所述的分子, 其中, 所述氨基酸残基的改变是在特定的位置向原始序列中添加氨基酸残基。

15. 如权利要求 7-14 中任意一项所述的分子, 其中, 通过进一步的改变恢复所述分子的生物学活性。

16. 如权利要求 15 所述的分子, 其中, 进一步的改变是替代、添加或缺失特定的氨基酸。

17. 如权利要求 7-16 中任意一项所述的分子, 其中, 所述氨基酸的改变是参照同源蛋白序列进行的。

18. 如权利要求 7-16 中任意一项所述的分子, 其中, 所述氨基酸的改变是参照计算机模拟技术进行的。

19. 编码权利要求 1-18 中任意一项所述的经修饰角质形成细胞生长因子(KGF)的 DNA 序列。

20. 一种药物组合物, 其包含上述任意一项权利要求中所定义的具有角质形成细胞生长因子(KGF)生物学活性的经修饰分子, 并可任选地包含药学上可接受的载体、稀释剂或赋形剂。

21. 制造上述任意一项权利要求中所定义的具有角质形成细胞生长因子(KGF)生物学活性的经修饰分子的方法, 该方法包括下述步骤:

(i) 确定所述多肽或其中一部分的氨基酸序列;

(ii) 通过任意方法, 包括利用体外或计算机技术或生物学试验确定所述肽与 MHC 分子的结合, 由此鉴定所述蛋白的氨基酸序列中潜在的一个或多个 T-细胞表位;

(iii) 设计新的序列变异体, 其中, 经鉴定的潜在 T-细胞表位内有一个或多个氨基酸经过修饰, 由此基本上减弱或去除了所述 T-细胞表位的活性, 这一效果可由体外或计算机技术或生物学试验通过所述肽与 MHC 分子的结合来确定, 或通过肽-MHC 复合物与 T-细胞的结合来确定;

(iv) 通过重组 DNA 技术构建所述序列变异体, 并检测所述的变异体以便鉴定一个或多个具有所需性质的变异体; 和

(v) 任选地重复步骤(ii)-(iv)。

22. 如权利要求 21 所述的方法, 其中步骤(iii)是通过在任何原本存在的 T-细胞表位中替代、添加或缺失 1-9 个氨基酸残基来进行的。

23. 如权利要求 22 所述的方法, 其中, 所述的改变是参照同源蛋白序列和/或计算机模拟技术进行的。

24. 如权利要求 21-23 中任意一项所述的方法, 其中, 步骤(ii)是通过下述步骤进行的:

(a) 在所述的肽中选择一个具有已知氨基酸序列的区域;

(b) 然后由所选择的区域中顺序抽取预定统一大小且至少由 3 个氨基酸残基组成的重叠氨基酸残基片段;

(c) 通过对存在于抽样氨基酸残基片段中的每个疏水氨基酸残基侧链赋值求和, 计算每一抽样片段的 MHC II 类分子结合值;和

(d) 根据计算出的该片段的 MHC II 类分子结合分值鉴定至少一个适于修饰的片段, 以在基本上不减弱所述肽的治疗功效的前提下改变整体的 MHC II 类的结合分值。

25. 如权利要求 24 所述的方法, 其中, 步骤(c)通过下述步骤利用经改进而包含了 12-6 范德华配体-蛋白能量排斥项和配体构象能量项的 Böhm 评分函数进行, 所述步骤为:

(1) 提供 MHC II 类分子模型第一数据库;

(2) 提供所述 MHC II 类分子模型的容许肽主链的第二数据库;

(3) 从所述的第一数据库中筛选模型;

(4) 从所述的第二数据库中筛选容许肽主链;

(5) 鉴定在每个抽样片段中存在的氨基酸残基侧链;

(6) 确定存在于每个抽样片段中的所有侧链的结合亲和性值; 以及对每一模型和每一所述的主链重复步骤(1)到(5)。

26. 一种选自表 1 所示组的 13 肽 T-细胞表位肽，其具有潜在的 MHC II 类结合活性且由未经免疫遗传修饰的角质形成细胞生长因子(KGF)构建。

27. 一种肽序列，其由如权利要求 26 所述的 13 肽 T-细胞表位肽中的至少 9 个连续的氨基酸残基组成。

28. 如权利要求 26 所述的 13 肽 T-细胞表位肽用于生产在体内使用时基本上没有免疫原性或免疫原性低于具有相同生物学活性的任何未经修饰的分子的角质形成细胞生长因子(KGF)的用途。

29. 如权利要求 27 所述的肽序列用于生产在体内使用时基本上没有免疫原性或免疫原性低于具有相同生物学活性的任何未经修饰的分子的角质形成细胞生长因子(KGF)的用途。

## 免疫原性减弱的经修饰角质形成细胞生长因子(KGF)

### 发明领域

本发明涉及尤其是用于对人施用的、特别是用于治疗的多肽。所述的多肽是经修饰的多肽，其中，所述的修饰使得上述多肽在施用于人体时引起免疫应答的倾向减弱。本发明特别涉及对角质形成细胞生长因子(KGF)进行修饰以产生 KGF 蛋白变异体，该变异体在体内使用时基本上无免疫原性，或比未经修饰的相应蛋白的免疫原性低。本发明还涉及来自上述未经修饰蛋白的 T-细胞表位肽，由此可以构建免疫原性减弱的经修饰的角质形成细胞生长因子(KGF)变异体。

### 发明背景

有许多例子表明治疗蛋白的效率受限于针对所述治疗蛋白的干扰性免疫反应。已有若干种小鼠单克隆抗体表现出治疗多种人类疾病的前景，但在某些情况下由于诱导出相当程度的人抗小鼠抗体(HAMA)应答而未能成功应用 [Schroff, R. W.等(1985) *Cancer Res.* 45: 879-885; Shawler, D. L. 等 (1985) *J. Immunol.* 135: 1530-1535]。对于单克隆抗体，已发展了多种技术以试图减弱 HAMA 应答 [WO 89/09622; EP 0239400; EP 0438310; WO 91/06667]。这些重组 DNA 方法通常是减少最终的抗体构建体中小鼠的遗传信息，同时增加最终构建体中人的遗传信息。尽管如此，在许多个体中，所得的“人源化”抗体仍然引起患者的免疫应答 [Issacs J. D. (1990) *Sem. Immunol.* 2 : 449,456; Rebello, P. R. 等 (1999) *Transplantation* 68: 1417-1420]。

抗体并不是唯一一类作为治疗剂施用并可引起免疫应答的多肽分子。即使是来源于人且具有与人体内蛋白相同氨基酸序列的蛋白也可以在人体内诱导免疫应答。值得注意的例子包括粒细胞-巨噬细胞集落刺激

因子[Wadhwa, M. 等 (1999) Clin. Cancer Res. 5: 1353-1361]和干扰素  $\alpha 2$  [Russo, D. 等 (1996) Bri. J. Haem. 94: 300-305; Stein, R. 等 (1988) New Engl. J. Med. 318: 1409-1413]等的治疗性应用。

诱导免疫应答的主要因素是在蛋白中存在可经由 MHC II 类分子的呈递作用激活 T-细胞活性的肽(即所谓的 T-细胞表位)。这种潜在的 T-细胞表位通常定义为任何具有与 MHC II 类分子结合能力的氨基酸序列。可测定此种 T-细胞表位以建立 MHC 结合。隐含地, "T-细胞表位"是指当其与 MHC 分子结合时可被 T-细胞受体(TCR)识别的表位,并且至少原则上,这种表位可以通过使 TCR 促进 T-细胞应答而激活这些 T-细胞。但可以理解,某些可与 MHC II 类分子结合的肽可能存留在蛋白序列中,因为这种肽在施用此最终蛋白的生物中被识别为“自身的”。

已知这些 T-细胞表位肽中的某些可在肽、多肽或蛋白在细胞中降解的过程中释放出来,随后由主要组织相容性复合体(MHC)分子呈递以便引发 T-细胞激活作用。对于 MHC II 类分子呈递的肽,这种 T-细胞激活作用可,例如通过直接刺激 B 细胞产生抗体而引起抗体应答。

MHC II 类分子是一组高度多态的蛋白,其对辅助 T 细胞筛选和激活起重要作用。人白细胞抗原群 DR (HLA-DR)是这组蛋白的主要同种型也是本发明的焦点所在。鉴于同种型 HLA-DQ 和 HLA-DP 行使类似的功能,本发明对此两者均适用。MHC II 类 DR 分子由  $\alpha$  和  $\beta$  链组成,其 C-末端插入并穿越细胞膜。每一异源二聚体都具有可与长度在 9 到 20 个氨基酸范围内变化的肽相结合的配体结合结构域,但是结合沟中最多容纳 11 个氨基酸。配体结合结构域由  $\alpha$  链的 1-85 位氨基酸和  $\beta$  链的 1-94 位氨基酸组成。最近的研究表明 DQ 分子具有同源的结构,预期 DP 家族蛋白也非常类似。在人中已发现了 DR 同种型的约 70 种不同同种异型,对于 DQ 有 30 种不同的同种异型,对于 DP 有 47 种不同的同种异型。每一个体带有 2 到 4 个 DR 等位基因,两个 DQ 及两个 DP 等位基因。许多 DR 分子的结构已阐明,这种结构指向一种带有许多疏水口袋的开口肽结合沟,所述的疏水口袋与肽的疏水残基(口袋残基)相互

作用[Brown 等, Nature (1993) 364: 33; Stern 等, (1994) Nature 368: 215]。确立不同 II 类分子同种异型的多态性有助于肽结合沟中的不同肽结合表面的广泛多样性,并且在群体水平保证有关识别外来蛋白和引发针对病原生物免疫应答的最大灵活性。在配体结合结构域内存在相当数量的多态性,并且在不同地域和种族群体存在着显著不同的“家族”。这一多态性影响着肽结合结构域的结合特性,因此,不同的 DR 分子“家族”对不同序列属性的肽具有特异性,但也可以有一些重叠。这种特异性决定了最终负责驱动针对 B 细胞表位的抗体应答的 Th-细胞表位识别(II 类 T-细胞应答),所述的 B 细胞表位存在于衍生所述 Th-细胞表位的相同蛋白上。因此,个体中针对蛋白的免疫应答在很大程度上受到 T-细胞表位识别作用的影响,这种识别作用是个体中与 HLA-DR 同种异型的肽结合特异性的函数。因而,为对全球种群背景下的肽或蛋白中的 T-细胞表位进行鉴定,需要考虑到尽可能多样的 HLA-DR 同种异型组的结合特性,由此覆盖尽可能高比例的世界上的种群。

针对治疗性蛋白(如本发明的目的蛋白)的免疫应答通过 MHC II 类肽呈递途径进行。其间外来蛋白经吞噬和加工后与 DR、DQ 或 DP 型 MHC II 类分子结合以进行呈递。MHC II 类分子由专门抗原呈递细胞(APC)如巨噬细胞、树突状细胞等表达。通过在 T 细胞表面的关联性 T-细胞受体与 MHC II 类肽复合物相互作用,加之与某些其他共同受体,如 CD4 分子交联可诱导 T-细胞进入激活状态。上述激活作用可导致细胞因子释放,进一步激活其他淋巴细胞如 B 细胞,产生抗体或激活 T 杀伤细胞形成完整的细胞免疫应答。

肽与给定的 MHC II 类分子结合以备在 APC 表面呈递的能力依赖于多种因素,最主要的是肽的一级结构。这影响其蛋白酶剪切倾向及其在 MHC II 类分子的肽结合隙中的结合亲和性。在 APC 表面的 MHC II 类/肽复合物向能识别决定簇的特定 T-细胞受体(TCR)呈递一个结合面,其中所述决定簇由肽和 MHC II 类分子的暴露残基共同提供。

本领域中有鉴定能结合 MHC II 类分子的合成肽的方法(例如 WO98/52976 和 WO00/34317)。这种肽不是在所有情况下都行使 T 细胞表位的功能,特别是在体内会受加工途径和其他现象的影响。T-细胞表位鉴定是表位清除的第一步。鉴定及从蛋白中去除潜在 T-细胞表位先前已有公开。本领域中已有检测 T-细胞表位的方法,通常是通过计算机手段在经试验确定的 T-细胞表位中扫描公认的序列基元,或利用计算机技术预测 MHC II 类结合肽,特别是 DR-结合肽。

WO98/52976 和 WO00/34317 中公开了鉴定具有与人 MHC II 类 DR 同种异型亚群结合的潜在能力的多肽序列的计算机穿线方法 (computational threading approaches)。这些教导中,通过在人源或非人源治疗性抗体或非抗体蛋白一级序列中进行准确的氨基酸替换以去除预测的 T-细胞表位。

此外,利用重组 MHC 分子与合成肽的可溶性复合物(该复合物能与来自于人或受试实验动物外周血液样品中的 T-细胞克隆相结合)的技术也在本领域中有所应用[Kern, F. 等 (1998) *Nature Medicine* 4: 975-978; Kwok, W. W. 等 (2001) *TRENDS in Immunology* 22: 583-588], 这些技术也可以用于表位鉴定策略中。

根据上述描述,由此可能期望鉴定、去除或至少减少给定的具有重要的治疗价值但原本具有免疫原性的肽、多肽或蛋白中的 T-细胞表位。

“角质形成细胞生长因子(KGF)”是这些有治疗价值的分子之一。KGF 是成纤维细胞生长因子(FGF)/肝素结合生长因子蛋白家族的成员。它是一种主要在肺部表达的分泌糖蛋白,可通过刺激角质细胞和其他上皮细胞的生长促进伤口的愈合[Finch 等(1989), *Science* 24: 752-755; Rubin 等(1989), *Proc. Natl. Acad. Sci. U.S.A.* 86 : 802-806]。该糖蛋白的成熟(加工后的)形式包含 163 个氨基酸残基,可通过培养特定的细胞系,从条件培养基中分离得到[Rubin 等, (1989) 出处同上], 或利用重组技术生产[Ron 等(1993) *J. Biol. Chem.* 268: 2984-2988]。该蛋白在多种重要的疾病和损伤修复情况中对于刺激上皮细胞的生长具有治疗价值。

本发明特别涉及具有 163 个氨基酸残基的人 KGF 蛋白的成熟（经加工的）形式。

也有人公开了 KGF 分子[例如 US 6,008,328; WO 90/08771]，包括经修饰的 KGF [Ron 等(1993)出处同上; WO 95/01434]。但这些教导并没有认识到 T-细胞表位对所述蛋白免疫原性的重要性，也不能据此联想到按照本发明的方案以特异的、可控制的方式直接影响所述蛋白的性质。

角质形成细胞生长因子(KGF)的氨基酸序列（以单字母密码表示）如下：

MCNDMTPEQMATNVNCSSPERHTRSVDYMEGGDIRVRLFCRTQWYL  
RIDKRGKVKGTTQEMKNNYNIMEIRTVAVGIVAIAKGVSEFYLAMNKEGK  
LYAKKECNEDCNFKELILENHYNTYASAKWTHNGGEMFVALNQKGIPV  
RGKKTKKEQKTAHFLPMAIT

但是，一直存在对具有改进的性质的角质形成细胞生长因子(KGF)类似物的需要。所需要的改进包括用于表达和纯化所述治疗剂的可供选择的方案和形式，以及尤其是所述蛋白生物学性质的改善。特别需要改进的是在施用于人体时在体内的特性。在这方面，非常需要提供对人体具有减弱的或没有诱导免疫应答可能性的角质形成细胞生长因子(KGF)。

## 发明概述及内容

本发明提供了经修饰的“角质形成细胞生长因子(KGF)”，其中，通过减少或去除大量潜在的 T-细胞表位的方式对该因子的免疫学特性进行修饰。本发明提供去除了一个或多个 T 细胞表位的角质形成细胞生长因子(KGF)的修饰形式。

由其他哺乳动物或脊椎动物中鉴定的 KGF 蛋白与与本发明中公开的肽序列之间存在大量共有序列，且与列表中公开的肽序列间存在大量基本上相同的共有序列。这些蛋白序列也落入本发明的范围内。

本发明公开了在角质形成细胞生长因子一级序列内鉴定的序列，由于它们具有与 MHC II 类分子结合的可能性，所以是潜在的 T-细胞表位。本发明特别涉及具有 163 个氨基酸残基的人 KGF 蛋白。

本发明还公开了在基本上不影响生物学活性的前提下需要通过特定的氨基酸替代、添加或缺失进行改变的本发明分子的一级序列中的特定位点。在只有同时丧失生物学活性才能去掉免疫原性的情况下，可通过在所述蛋白的氨基酸序列中做进一步的改造以恢复所述的活性。

本发明还公开了生产这种经修饰的分子的方法，尤其是鉴定需要改变以减少或去除免疫原性位点的 T-细胞表位的方法。

本发明可应用于任何具有与本发明公开的一级氨基酸序列基本相同的氨基酸序列的 KGF 分子，因此可包括利用基因工程手段或其他方法获得的、可能不是包含 163 个氨基酸残基的 KGF 分子。

预期本发明的蛋白在人体中的循环时间会延长，因此对慢性或复发性疾病，如角质形成细胞生长因子(KGF)的多种适应症特别有益。本发明提供经修饰的 KGF 蛋白，预期其在体内将表现出改进的性质。这些经修饰的 KGF 分子可应用于药物组合物。

总之，本发明涉及下述内容：

- 一种经修饰的分子，其具有角质形成细胞生长因子(KGF)的生物学活性，且当其在体内应用时基本上无免疫原性或免疫原性低于任何具有相同生物学活性但未经修饰的分子；
- 如上所述的分子，其中，所述的免疫原性丧失是通过从原始的未经修饰的分子中去除一个或多个 T-细胞表位而实现的；
- 如上所述的分子，其中，所述的免疫原性丧失是通过减少能与由所述分子衍生的肽相结合的 MHC 同种异型的数量来实现的；
- 如上所述的分子，其中，去除了一个 T-细胞表位；
- 如上所述的分子，其中，原本存在的 T-细胞表位是 MHC II 类配体，或表现出经 II 类分子呈递作用后有刺激或结合 T-细胞的能力的肽序列；
- 如上所述的分子，其中，所述的肽序列选自如表 1 所示的组；

- 如上所述的分子，其中，任何原本存在的 T-细胞表位中的 1-9 个氨基酸残基，优选 1 个氨基酸残基发生了改变；
- 如上所述的分子，其中，所述氨基酸残基的改变是用其他的氨基酸残基在特定的位置替代、添加或缺失原本存在的氨基酸残基；
- 如上所述的分子，其中，按表 2 所示进行一个或多个氨基酸残基的替代；
- 如上所述的分子，其中，另外还按表 3 所示进行一个或多个氨基酸残基的替代以减少能与由所述分子衍生的肽相结合的 MHC 同种异型的数量；
- 如上所述的分子，其中，在需要时常通过替代、添加或缺失特定的氨基酸作进一步改变以恢复所述分子的生物学活性；
- 编码任何上述和下述经修饰的分子的 DNA 序列或分子；
- 药物组合物，其包含如上述和/或权利要求中定义的具有角质形成细胞生长因子(KGF)生物学活性的经修饰分子，并可任选地包含药学上可接受的载体、稀释剂或赋形剂；
- 制造如上引用的任何权利要求中所定义的具有角质形成细胞生长因子(KGF)生物学活性的经修饰分子的方法，该方法包括下述步骤：(i) 确定所述多肽或其中一部分的氨基酸序列；(ii)通过任意方法，包括利用体外或计算机(*in silico*)技术或生物学试验确定所述肽与 MHC 分子的结合，由此鉴定所述蛋白的氨基酸序列中潜在的一个或多个 T-细胞表位；(iii)设计新的序列变异体，其中，经鉴定的潜在 T-细胞表位内有一个或多个氨基酸经过修饰，由此基本上减弱或去除了所述 T-细胞表位的活性，这一效果可由体外或计算机技术或生物学试验通过所述肽与 MHC 分子的结合来确定；(iv)通过重组 DNA 技术构建所述序列变异体，并检测所述的变异体以便鉴定一个或多个具有所需性质的变异体；和(v)任选地重复步骤(ii)-(iv)；
- 如上所述的方法，其中步骤(iii)是通过在任何原本存在的 T-细胞表位中替代、添加或缺失 1-9 个氨基酸残基来进行的；
- 如上所述的方法，其中，所述的改变是参照同源蛋白质序列和/或计

计算机模拟技术进行的；

- 如上所述的方法，其中步骤(ii)是通过下述步骤进行的：(a)在所述的肽中选择一个具有已知氨基酸序列的区域；(b)然后由所选择的区域中顺序抽取预定统一大小且至少由 3 个氨基酸残基组成的重叠氨基酸残基片段；(c)通过对存在于抽样氨基酸残基片段中的每个疏水氨基酸残基侧链赋值求和，计算每一抽样片段的 MHC II 类分子结合分值；和(d)根据计算出的该片段的 MHC II 类分子结合分值鉴定至少一个适于修饰的片段，以在基本上不减弱所述肽的治疗功效的前提下改变整体的 MHC II 类结合分值；步骤(c)优选地通过下述步骤利用经改进而包含了 12-6 范德华配体-蛋白能量排斥项和配体构象能量项的 Böhm 评分函数 (scoring function) 进行，所述步骤为(1)提供 MHC II 类分子模型第一数据库；(2)提供所述 MHC II 类分子模型的容许肽主链(allowed peptide backbone)的第二数据库；(3)从第一数据库中筛选模型；(4)从第二数据库中筛选容许肽主链；(5)鉴定在每个抽样片段中存在的氨基酸残基侧链；(6)确定存在于每个抽样片段中的所有侧链的结合亲和性值；以及对每一模型和每一所述的主链重复步骤(1)到(5)；
- 选自表 1 的具有潜在的 MHC II 类结合活性且由未经免疫遗传修饰的角质形成细胞生长因子(KGF)构建的 13 肽 T-细胞表位肽，及其在制造在体内使用时基本上没有免疫原性或免疫原性低于具有相同生物学活性的任何未经修饰的分子的 KGF 中的用途；
- 由上述的 13 肽 T-细胞表位肽中的至少 9 个连续的氨基酸残基组成的肽序列，及其在制造在体内使用时基本上没有免疫原性或免疫原性低于具有相同生物学活性的任何未经修饰的分子的 KGF 中的用途。

根据对本发明的理解，术语“T-细胞表位”是指具有下述能力的氨基酸序列，即能结合 MCH II、能刺激 T-细胞和/或以与 MHC II 的复合物的形式结合（但不一定可测量地激活）T-细胞。此处及后附权利要求中所用的术语“多肽”是指包含两个或多个氨基酸的化合物。氨基酸之间通过肽键相连(定义见下)。肽的生物生产中涉及 20 种不同的天然氨基

酸，任意数量的所述氨基酸可按任意的顺序连接形成肽链或环。用于生物生产肽中的天然氨基酸全部具有 L-构型。可应用常规的合成方法利用 L-氨基酸、D-氨基酸或两种不同构型的氨基酸的各种组合制备合成肽。一些肽仅包含少量的氨基酸单元。例如含有不到 10 个氨基酸的短肽有时被称作“寡肽”。其他的包含大量氨基酸残基，例如达 100 个或更多个氨基酸残基的肽称作“多肽”。习惯上将含有 3 个或 3 个以上氨基酸的任何肽链多看作“多肽”，而将“寡肽”视为特定类型的短“多肽”。因而本文所提到的“多肽”也包括“寡肽”。而且，所用到的“肽”包括多肽、寡肽和蛋白。不同的氨基酸排列形式形成不同的多肽或蛋白。因此，多肽的数量以及可形成的蛋白的数量实际上是无限的。“ $\alpha$ 碳( $C\alpha$ )”是肽链中碳-氢(CH)组分中的碳原子。“侧链”是  $C\alpha$  的侧基，其可包含简单的或复杂的基团或部分，且具有与所述肽的大小相比可显著变化的外形大小。

本发明可应用于与此处公开的角质形成细胞生长因子(KGF)具有基本上相同的一级氨基酸序列的任何角质形成细胞生长因子(KGF)分子，因此包括利用基因工程手段或其他方法获得的、可能不是包含 163 个氨基酸残基的角质形成细胞生长因子(KGF)。

角质形成细胞生长因子(KGF)蛋白，例如从其他哺乳动物来源中鉴定出的相关蛋白与本发明中公开的肽序列之间存在大量共有序列，且与列表中公开的肽序列间存在大量基本上相同的共有序列。因此这样的蛋白序列也落入本发明的保护范围。

本发明是为了克服实际应用中存在的下述问题，即将可溶性蛋白引入自体生物中可引发免疫应答，产生可与所述可溶性蛋白相结合的宿主抗体。其中的一个例子是干扰素  $\alpha 2$ ，尽管这一蛋白是内源产生的，但许多病人都会产生针对它的抗体[Russo, D.等(1996)出处同上；Stein, R.等(1988)出处同上]。将角质形成细胞生长因子(KGF)用于治疗用途时也可能存在类似的问题，本发明试图通过提供在施用于人体时引发免疫应答的倾向发生改变的角质形成细胞生长因子(KGF)蛋白来解决这一问题。

本发明中形成经修饰的角质形成细胞生长因子(KGF)的总的方法包括下述步骤:

(a) 确定多肽或其中一部分的氨基酸序列;

(b) 通过任意方法,包括利用体外或计算机技术或生物学试验确定所述肽与 MHC 分子的结合,由此鉴定所述蛋白的氨基酸序列中潜在的一个或多个 T-细胞表位;

(c) 设计新的序列变异体,其中,经鉴定的潜在 T-细胞表位内有一个或多个氨基酸经过修饰,由此基本上减弱或去除了所述 T-细胞表位的活性,这一效果可由体外或计算机技术或生物学试验通过所述肽与 MHC 分子的结合来确定。构建此序列变异体以避免由所述的序列变异体产生新的潜在 T-细胞表位,否则所述的新的潜在 T 细胞表位又通过此种方式进行修饰以基本上减弱或消除 T-细胞表位活性;和

(d) 通过重组 DNA 技术构建所述序列变异体,并检测所述的变异体以便根据已知的重组技术鉴定一个或多个具有所需功能的变异体。

对于步骤(b)中对潜在 T-细胞表位的鉴定可依照本领域已公开的方法进行。在 WO 98/59244; WO 98/52976; WO 00/34317 中也公开了适当的方法,并优选用于鉴定角质形成细胞生长因子(KGF)-衍生的肽对 MHC II 类分子的结合倾向。

在实施例部分公开了另一种非常有效的通过计算鉴定 T-细胞表位的方法,它是本发明的优选实施方案。

实践中,将制备许多角质形成细胞生长因子(KGF)蛋白变异体并检测所需的免疫和功能特性。最优选通过重组 DNA 技术生产所述的变异体蛋白,同时也可以利用其他的方法,包括化学合成角质形成细胞生长因子(KGF)片段。

涉及 163 个氨基酸残基的角质形成细胞生长因子(KGF)蛋白序列的根据上述方案中步骤(b)的分析结果列于表 1。

**表 1: 具有潜在人 MHC II 类结合活性的角质形成细胞生长因子(KGF)的肽序列**

NDMTPEQMATNVN, DMTPEQMATNVNC, EQMATNVNCSSPE, TNVNCSSPERHTR,  
 RSYDYMEGGDIRV, YDYMEGGDIRVRR, DYMEGGDIRVRRL, GDIRVRRLFCRTQ,  
 IRVRRLFCRTQWY, RRLFCRTQWYLRI, RLFCRTQWYLRI, TQWYLRIIDKRGKV,  
 QWYLRIIDKRGKVK, WYLRIIDKRGKVG, LRIDKRGKVGKTQ, GKVKGKTQEMKNY,  
 QEMKNYNNIMEIR, NNYNNIMEIRTVAV, YNIMEIRTVAVGI, NIMEIRTVAVGIV,  
 MEIRTVAVGIVAI, RTVAVGIVAIKGV, VAVGIVAIKGVES, VGIVAIKGVESF,  
 VAIKGVESFYLA, KGVESFYLAMNK, SEFYLAMNKEGKL, EFLAMNKEGKLY,  
 FYLAMNKEGKLYA, LAMNKEGKLYAKK, GKLYAKKECNEDC, KLYAKKECNEDCN,  
 CNFKELILENHYN, KELILENHYNNTYA, ELILENHYNNTYAS, LILENHYNNTYASA,  
 NHYNNTYASAKWTH, NTYASAKWTHNGG, AKWTHNGGEMFVA, GEMFVALNQGKIP,  
 EMFVALNQGKIPV, FVALNQGKIPVRG, VALNQGKIPVRGK, KGIPVRGKKTKE,  
 IPVRGKKTKEQK, KTKKEQKTAHFLP

肽是 13 肽, 氨基酸用单字母表示。

涉及本发明的经修饰分子的根据上述方案中步骤(c)和(d)所得的设计结果和构建体列于表 2 和表 3。

**表 2: 导致角质形成细胞生长因子(KGF)的潜在 T-细胞表位消除的替代 (WT = 野生型)。**

| 残基# | WT 残基 | 替代                        |
|-----|-------|---------------------------|
| 5   | M     | A C D E G H K N P Q R S T |
| 10  | M     | A C D E G H K N P Q R S T |
| 14  | V     | A C D E G H K N P Q R S T |
| 26  | Y     | A C D E G H K N P Q R S T |
| 28  | Y     | A C D E G H K N P Q R S T |
| 29  | M     | A C D E G H K N P Q R S T |
| 34  | I     | A C D E G H K N P Q R S T |
| 36  | V     | A C D E G H K N P Q R S T |
| 39  | L     | A C D E G H K N P Q R S T |
| 40  | F     | A C D E G H K N P Q R S T |
| 45  | W     | A C D E G H K N P Q R S T |
| 46  | Y     | A C D E G H K N P Q R S T |

| 残基# | WT 残基 | 替代 |   |   |   |   |   |   |   |   |   |   |   |   |
|-----|-------|----|---|---|---|---|---|---|---|---|---|---|---|---|
| 47  | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 49  | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 55  | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 61  | M     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 65  | Y     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 67  | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 68  | M     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 70  | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 73  | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 75  | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 77  | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 78  | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 80  | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 83  | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 87  | F     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 88  | Y     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 89  | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 91  | M     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 97  | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 98  | Y     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 109 | F     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 112 | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 113 | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 114 | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 118 | Y     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 121 | Y     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 126 | W     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 133 | M     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 134 | F     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 135 | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 137 | L     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 142 | I     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |
| 144 | V     | A  | C | D | E | G | H | K | N | P | Q | R | S | T |

表3:导致相应于一个或多个MHC同种异型的潜在T-细胞表位去除的额外替代

| 残基# | WT 残基 | 替代                  |
|-----|-------|---------------------|
| 5   | M     | W Y                 |
| 8   | E     | P T                 |
| 10  | M     | L W Y               |
| 11  | A     | D E H K N P Q R S T |

| 残基# | WT 残基 | 替代                  |
|-----|-------|---------------------|
| 13  | N     | H P T               |
| 26  | Y     | M W                 |
| 29  | M     | I V W Y             |
| 30  | E     | A C G P             |
| 31  | G     | D H P               |
| 32  | G     | D E H K N P Q R S T |
| 33  | D     | A C G P             |
| 34  | I     | M W Y               |
| 35  | R     | P T                 |
| 36  | V     | M W Y               |
| 37  | R     | H P T               |
| 39  | L     | F I M V W Y         |
| 40  | F     | M W Y               |
| 41  | C     | H P                 |
| 42  | R     | A C G P T           |
| 43  | T     | P                   |
| 44  | Q     | A C G P T           |
| 46  | Y     | M                   |
| 47  | L     | V W Y               |
| 49  | I     | W Y                 |
| 50  | D     | A C G P             |
| 51  | K     | H P                 |
| 52  | R     | T                   |
| 54  | K     | A H P T             |
| 55  | V     | I Y                 |
| 57  | G     | P                   |
| 61  | M     | W Y                 |
| 62  | K     | A C G P             |
| 63  | N     | A C G P             |
| 64  | N     | P T                 |
| 66  | N     | H P                 |
| 67  | I     | M W Y               |
| 68  | M     | I V W Y             |
| 69  | E     | A C G H P S T       |
| 70  | I     | L M V W Y           |
| 71  | R     | A C G P T           |
| 72  | T     | A C G H P           |
| 73  | V     | I L M W Y           |
| 74  | A     | P T                 |
| 75  | V     | I L M W Y           |
| 76  | G     | D E H K N P Q R S T |
| 77  | I     | M W                 |
| 78  | V     | F I L W Y           |
| 79  | A     | P T                 |

| 残基# | WT 残基 | 替代                      |
|-----|-------|-------------------------|
| 80  | I     | F M V W                 |
| 81  | K     | A C G P T               |
| 82  | G     | D E H K N P Q S         |
| 83  | V     | F M W Y                 |
| 84  | E     | A C G P                 |
| 85  | S     | P T                     |
| 86  | E     | P T                     |
| 87  | F     | M W Y                   |
| 89  | L     | F I M V W Y             |
| 90  | A     | C D E G H K N P Q R S T |
| 91  | M     | W                       |
| 92  | N     | A C G H P T             |
| 93  | K     | A C G P T               |
| 94  | E     | A C G H P               |
| 95  | G     | C D E H K N P Q R S T   |
| 96  | K     | A C G H P               |
| 97  | L     | I M V W Y               |
| 98  | Y     | W                       |
| 99  | A     | D E H K N P Q R S T V   |
| 100 | K     | H N P Q S T             |
| 101 | K     | A C G P T               |
| 102 | E     | H P                     |
| 103 | C     | D E G H K N P Q R S T   |
| 104 | N     | A C G P T               |
| 105 | E     | H P T                   |
| 106 | D     | H P T                   |
| 108 | N     | A C G P                 |
| 109 | F     | M W                     |
| 110 | K     | A C G P                 |
| 111 | E     | A C G P                 |
| 112 | L     | F I M V W Y             |
| 113 | I     | M W                     |
| 115 | E     | A C G P T               |
| 116 | N     | A C G P T               |
| 117 | H     | P T                     |
| 118 | Y     | M                       |
| 119 | N     | A C G P T               |
| 120 | T     | P                       |
| 121 | Y     | I M V W                 |
| 123 | S     | A C G H P               |
| 124 | A     | C D E H K N P Q R S T   |
| 125 | K     | A C G P                 |
| 127 | T     | P                       |
| 128 | H     | A C G P                 |

| 残基# | WT 残基 | 替代            |
|-----|-------|---------------|
| 129 | N     | H P T         |
| 137 | L     | F I M V W Y   |
| 138 | N     | A C G P       |
| 139 | Q     | A C G P       |
| 140 | K     | H N P Q R S T |
| 142 | I     | M W Y         |
| 143 | P     | T             |
| 144 | V     | A             |
| 145 | R     | H K N P Q S T |
| 147 | K     | A C G P       |

本发明涉及角质形成细胞生长因子(KGF)类似物,其中,在可以导致所述蛋白中潜在的 T 细胞表位的活性明显减弱或从所述蛋白中去除一个或多个潜在的 T-细胞表位活性的位点替代了至少一个氨基酸残基。表 1 中鉴定的任何潜在的 MHC II 类配体中特定位点的一个或多个氨基酸的替代,可产生当作为治疗剂施用于人体时具有减弱的潜在免疫原性的角质形成细胞生长因子(KGF)分子。优选地,在预计可以实现基本上减弱或去除 T-细胞表位活性的肽序列中的适当位点进行氨基酸替代。实践中,合适的位点优选等同于在 MHC II 类结合沟内提供的疏水口袋之一中结合的氨基酸残基。

最优选是在所述肽中称作 P1 或 P1 锚的位置改变裂缝中第一口袋内的结合。公认地,肽的 P1 锚残基和 MHC II 类结合沟第一口袋之间的结合相互作用的质量是对整个肽的总结合亲和性的主要决定因素。在所述肽这一位置的适当替代应当是替换为不易容纳到所述口袋中的残基,例如替换为更亲水的残基。所述肽中相应与 MHC 结合裂缝内其他口袋区域结合的位置处的氨基酸残基也被认为是落入本发明的范围内。

可以理解,由给定的潜在 T-细胞表位内的单一氨基酸替代导致该表位去除的路线是最优选的。也可以进行单一表位内的组合替代,例如,这对于单独定义的表位间彼此重叠的情况特别适宜。此外,在给定的表位内一个一个地进行或在一个表位内组合进行的氨基酸替代也可以发生在非对应于 MHC II 类结合沟的"口袋残基"的位置,而是在所述肽内

的任意位点进行。替代可参照同源结构或由本领域已知的计算机技术产生的结构方法，也可以根据本发明分子的已知结构特征进行。所有此类替代均落入本发明的范围内。

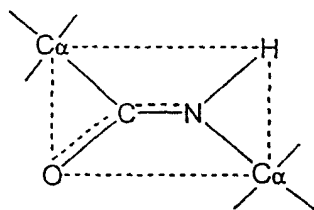
也可以考虑不在上面所鉴定的肽内进行氨基酸替代，特别是与在所列表肽内进行的替代相结合的情况下。例如可以考虑利用某种改变恢复变异体分子的结构或生物学功能。这种补偿性的改变和由角质形成细胞生长因子(KGF)多肽中缺失或添加特定的氨基酸残基得到具有所需活性的变异体的改变，以及任何本发明公开的肽中的改变均落入本发明的范围内。

本发明的范围涉及经修饰的角质形成细胞生长因子(KGF)，含有上述经修饰的角质形成细胞生长因子(KGF)蛋白或经修饰的角质形成细胞生长因子(KGF)蛋白片段的组合物，及其相关的组合物应认为均落入本发明的范围内。另一方面，本发明涉及编码经修饰的角质形成细胞生长因子(KGF)实体的核酸。另一方面本发明涉及利用经修饰的角质形成细胞生长因子(KGF)蛋白对人进行治疗的方法。

## 实施例

有多种因素对决定蛋白或多肽的总体结构起重要作用。首先是肽键，即将氨基酸连接在一起形成链的键，它是一种共价键。这种键是平面结构的，实质上是一种取代的酰胺。“酰胺”指含-CONH-基团的一组有机化合物中的任何一个化合物。

连接相邻氨基酸的  $C\alpha$  的平面肽键如下所示：



由于  $O=C$  和  $C-N$  原子位于一个相对刚性的平面中，所以不会发生沿这些轴的自由旋转。因此，图中虚线所示的平面有时被称作“酰胺”

平面或“肽平面”，肽主链中的氧(O)、碳(C)、氮(N)和氢(H)原子位于其中。 $C\alpha$ 原子位于酰胺平面中相对的角上。由于肽或酰胺平面中的  $O=C$  和  $C-N$  原子基本上不发生旋转，所以多肽链包含一系列连接  $C\alpha$  原子的平面肽键。

第二个对决定多肽或蛋白的整体结构或构象起重要作用的因素是绕共有  $C\alpha$  键的每一酰胺平面的转角。此后术语“转角”和“扭转角”是等同的术语。假定 O、C、N 和 H 原子保留在酰胺平面中(这通常是一种正确的假设，尽管在一些构象中这些原子会轻微的偏移平面)，这些转角确定了 N 和 R 多肽主链构象，即相邻残基之间的结构。这两个转角称为  $\phi$  和  $\psi$ 。因此，一套  $\phi_i$  和  $\psi_i$  角(其中，脚标  $i$  代表多肽链中的特定残基)有效地规定了多肽链的二级结构。在文献中定义了用于确定  $\phi$  和  $\psi$  角的惯例，即在给定的多肽中酰胺平面形成 0 度角的参考点，以及哪个角是  $\phi$  角，哪个角是  $\psi$  角的定义。参见 Ramachandran 等, *Adv. Prot. Chem.* 23: 283-437 (1968), 285-94 页，这些页中的内容在此引入作为参考。本发明的方法可应用于任何蛋白，并部分基于下述发现，即人 MHC II 类分子结合沟的初级口袋 1 锚定位点可以具有设计好的对特定氨基酸侧链的特异性。这一口袋的特异性由 MHC II 类分子  $\beta$  链第 86 位的氨基酸的身份来确定。这一位点位于口袋 1 的底部并决定可容纳于这一口袋中的氨基酸侧链的大小。Marshall, K. W., *J. Immunol.*, 152: 4946-4956 (1994)。如果这一残基是甘氨酸，则所有的疏水性脂肪族和芳香族氨基酸(疏水性脂肪族氨基酸是：缬氨酸、亮氨酸、异亮氨酸、甲硫氨酸，芳香族氨基酸是：苯丙氨酸、酪氨酸和色氨酸)均可容纳于所述的口袋中，优选芳香族侧链。如果这一口袋残基是缬氨酸，则该氨基酸的侧链伸到口袋中并限制了可容纳的肽侧链的大小，所以只有疏水性脂肪族侧链可容纳进去。因此在氨基酸残基序列中，无论哪里发现了带有疏水性脂肪族或芳香族侧链的氨基酸，即有存在 MHC II 类限制性 T-细胞表位的可能性。但是，如果所述的侧链是疏水性脂肪族侧链，其与 T-细胞表位结合的可能性约是芳香族侧链的两倍(假定 1 型口袋近似平均地分布于全球种群中)。

将本发明具体化的计算机方法描绘出肽区域包含 T-细胞表位的可能性,该方法如下:(1)扫描预定长度肽片段的一级序列,并鉴定存在的所有疏水性脂肪族和芳香族侧链。(2)对疏水性脂肪族侧链赋予比芳香族侧链高的值;优选两倍于赋予芳香族侧链的值,例如,给疏水性脂肪族侧链赋值为 2,给芳香族侧链赋值为 1。(3)将所述肽中的预定统一长度的每一重叠氨基酸残基片段(窗口)中确定存在的值总和起来,再将某一特定片段(窗口)的总值赋予该片段(窗口)中间位置的某个单个氨基酸残基,优选赋予处于抽样片段(窗口)中间点的氨基酸。将这一过程对每一抽样的重叠氨基酸残基片段(窗口)重复进行。因此,所述肽的每一氨基酸残基均被赋予了一个值,该值与 T-细胞表位存在于此特定片段(窗口)中的可能性相关。(4)用按照上述步骤 3 中的描述计算、赋予的值对被评估的整个氨基酸残基序列的氨基酸坐标作图。(5)序列中具有预定值(例如该值为 1)的所有部分均被认为可能包含 T 细胞表位,并且在需要时可进行修饰。在这一方面本发明提供了通用的方法,由此可描述可能包含 T-细胞表位的肽区域。在这些区域中对所述的肽进行修饰有可能改变 MHC II 类的结合特性。

依照本发明的另一方面,可利用更复杂的计算方法更精确地预测 T-细胞表位,该方法考虑了肽与 MHC II 等位基因模型之间的相互作用。根据这一方面,计算机预测存在于所述肽中的 T-细胞表位包括,根据所有已知的 MHC II 类分子结构构建至少 42 个 MHC II 类等位基因模型,将这些模型应用于 T-细胞表位计算机鉴定的方法,构建每一模型的肽主链文库以便允许在相关肽主链 $\alpha$ 碳( $C\alpha$ )位置具有已知的变异性,对于在肽和 MHC II 类分子相互作用的关键位置上的 20 种可供选择的氨基酸中的每一种,构建与每个模型对接的每一主链的氨基酸侧链构象文库,利用这些主链和侧链构象文库并结合评分函数选择对与特定 MHC II 类分子结合的特定肽而言最佳的主链和侧链构象,并从这一相互作用推导出结合分数。

MHC II 类分子模型可从 Brookhaven 蛋白数据库("PDB")中的许多类似的结构出发通过同源建模推导得出。它们可通过使用引入了模拟退

火算法的半自动同源建模软件(Modeller, Sali A. & Blundell TL., 1993. J. Mol Biol 234 : 779-815)并结合用于能量最小化的 CHARMM 力场(购自 Molecular Simulations Inc., San Diego, Ca.)来制备。也可以应用其他的建模方法。

本发明的方法与下述的其他计算方法有着显著的不同, 这些方法为: 利用从实验中得来的关于一小组 MHC II 类分子结合沟中每一位点的每一种氨基酸选项的结合数据文库的方法(Marshall, K. W.等, Biomed. Pept. Proteins Nucleic Acids, 1 (3): 157-162) (1995); 或利用类似的实验结合数据以定义所述的沟中特定结合口袋类型的结合特性(同样利用相对小的 MHC II 类分子亚组)然后将这一口袋文库中的口袋类型进行‘混合和匹配’以人工构建更“实际的”MHC II 类分子的方法(Sturniolo T.等, Nat. Biotech, 17 (6): 555-561 (1999)。这两种现有方法的主要缺陷在于实验的复杂性和需要合成大量的肽变异体造成仅有少量的 MHC II 类分子可通过实验扫描。因此第一种已知的方法仅能预测少量的 MHC II 类分子。第二种已知的方法还假设在不同 II 类等位基因的背景下在一个分子中衬有类似氨基酸的口袋将具有相同的结合特性, 并且其另外的缺陷在于, 仅仅可“实际地”地构建出那些包含口袋文库中所包含的口袋的 MHC II 类分子。利用本发明的建模方法可推导出任意数量和类型的 MHC II 类分子的结构, 因此可特异性地选择等位基因以代表全球种群的特征。此外, 扫描的 MHC II 类分子的数量可通过构建更多的模型而增加而无需通过复杂的实验获得额外的数据。利用主链文库使得被扫描的各种肽在与特定的 MHC II 类分子结合时其 C $\alpha$  原子位置处可进行变化。这也与上述现有技术中的计算机方法不同, 在那些方法中依赖于利用简化的肽主链来扫描结合在特定口袋中的氨基酸。这些简化的主链不可能代表在“真正的”肽中的主链文库构象, 导致对肽结合的预测不准确。本发明的主链文库是通过叠加蛋白数据库中所有与 MHC II 类分子结合的肽的主链, 并考虑到位于结合沟内的 11 个氨基酸的每个氨基酸的 C $\alpha$  原子之间的均方根(RMS)差而构建的。尽管该文库

可来自少量合适的可获得的小鼠和人的结构(当前为 13 种),为了允许存在甚至更大变异的可能性,将每一 C $\alpha$ - $\beta$ 位点的 RMS 数字提高 50%。然后确定每一氨基酸的平均 C $\alpha$ 位置,围绕这一点划一个球,其半径等于在该位置的 RMS 差加 50%。该球体代表所有可允许的 C $\alpha$ 位置。自具有最小 RMS 差的 C $\alpha$ (上述口袋 1 中氨基酸残基的 C $\alpha$ , 等同于结合沟中 11 个残基的位置 2)起运作,将所述的球三维网格化,网格内的每个顶点作为该氨基酸 C $\alpha$ 的可能位置。将后续的酰胺平面(相应于与后续氨基酸的肽键)移动到这些 C $\alpha$  的每一个上面,将  $\phi$  和  $\psi$  角以设定的间隔逐步地转动以便于安置后续的 C $\alpha$ 。如果后续的 C $\alpha$ 落入对这一 C $\alpha$ 而言‘可被允许的位置球’中,则此二肽的方向即可被接受,如果其落入所述球之外则所得的二肽不能被接受。对每一后续 C $\alpha$ 位置均重复这一过程,使肽从所述的口袋 1 C $\alpha$  ‘种子’开始生长,直到全部 9 个后续的 C $\alpha$  的位置均根据之前 C $\alpha$  的所有可能排列确定下来。

然后对口袋 1 前的单个 C $\alpha$ 重复上述步骤 1 次以上以构建定位于结合沟内的主链 C $\alpha$ 位置文库。生成的主链数目取决于几种因素:‘可被允许的位置球’的大小;对口袋 1 位点处‘最初的球’网格化的细度;用于定位后续 C $\alpha$  的  $\phi$  和  $\psi$  角逐步旋转的细度。利用这一程序可以构建大的主链文库。主链文库越大越可能发现对 MHC II 类分子结合沟内特定肽的最适主链。

鉴于和结合结构域的氨基酸可能存在冲突,所以不是所有的主链均适合于与所有 MHC II 类分子模型‘对接’(docking),故对每个等位基因建立包含适合于该等位基因的主链的亚文库。利用所述的主链文库并结合 MHC II 类分子模型可以构建出由与每一容许主链对接的每一 MHC II 类分子结合沟的每一位点中的每一氨基酸的容许侧链构象所组成的详尽数据库。可以利用简单的立体重叠函数构建这一数据组,其中,主链与 MHC II 类分子对接,氨基酸侧链在所需位置被嫁接到主链上。将侧链上可旋转的键以设定的间隔逐步旋转,记录下依赖于该键的原子的最终定位。将所述原子与结合沟侧链原子间的相互作用记录下来,根

据下述的标准确定是否接受这些位置:如此定位的所有原子的重叠总量不能超过预定值。因此,构象搜索的严谨度是在键的逐步旋转中所用间隔及对总重叠的预定限度的函数。如果已知特定的口袋是刚性的,则后一值可较小,但若已知口袋侧链位置相对灵活则严谨度可放松。这样便可以模拟结合沟口袋内灵活性的变化。针对与每一 MHC II 类分子对接后每一主链的所有位点上的所有氨基酸重复这种构象搜索以建立详尽的侧链构象数据库。

用适当的数学表达式评价 MHC II 类分子模型与肽配体构象的结合能量,所述的肽配体构象需通过扫描上述的主链/侧链大数据库根据经验获得。这样,通过对每一长度在 9-20 个氨基酸范围内变化(尽管对于每一次扫描长度是一定的)的可能肽进行下述计算,扫描蛋白以搜索潜在的 T-细胞表位:选择 MHC II 类分子及适合于该分子的肽主链,将相应于所需肽序列的侧链移植到其上。对于氨基酸的每一容许构象(由上述数据库获得),收集与主链上特定位点的特定侧链相关的原子身份和原子间距数据。对沿主链的每一侧链重复此过程,利用评分函数推导肽得分。保留该主链的最佳得分,对所选模型的每一容许主链重复该过程。比较所有容许主链的得分,最高的得分被认为是该 MHC II 类模型中所需肽的得分。对每一模型用从扫描的蛋白得到的所有可能肽重复上述过程,列出肽相对于模型的得分。

在本发明中,用于结合亲和力计算的每种配体都是选自上述肽或蛋白的氨基酸片段。因此所述配体为来自已知序列的肽、多肽或蛋白的长度为约 9 到 20 个氨基酸的选定氨基酸链。此后术语“氨基酸”和“残基”视为等同的术语。将移植到选自上述主链文库的主链上的待检测肽中的连续氨基酸形式的配体,通过肽主链上 C<sup>α</sup>-原子坐标定位到来自 MHC II 类分子模型库的 MHC II 类分子的结合裂缝中,并从所允许的构象数据选择每一侧链的允许构象。相关的原子身份和原子间距也来自这一数据库并用于计算肽结合分数。将对 MHC II 类结合口袋具有高亲和力的配体作为候选标记出来用于定点诱变。在标记的配体中(也由此

在目的蛋白中) 进行氨基酸替代, 然后用评分函数重新测定以确定使结合亲和力降低到预定的阈值以下的变化。这些变化即可引入到目的蛋白中以去除 T-细胞表位。肽配体与 MHC II 类分子结合沟的结合涉及非共价键相互作用, 其包括但不限于: 氢键、静电相互作用、疏水(亲脂)相互作用和范德华相互作用。它们包括在下面将详细描述肽评分函数中。应当理解, 氢键是非共价键, 其可在极性 or 带电的基团之间形成, 由被两个其他原子共享的氢原子构成。

氢供体中的氢带正电荷, 而氢受体带有部分负电荷。为肽/蛋白相互作用的目的, 氢键供体可以是连接氢的氮, 或连接在氧或氮上的氢。氢键受体原子可以是没有连接氢的氧、没有连接氢并具有一或两个连接的氮或仅有一个连接的硫。某些原子, 如连接了氢的氧或亚胺氮(如  $C=NH$ ), 既可以是氢受体也可以是氢供体。氢键的能量在 3-7 Kcal/mol, 大大强于范德华键, 但弱于共价键。氢键具有高度的方向性, 且当供体原子、氢原子和受体原子共线时最强。静电键是在带有相反电荷的离子对间形成的, 根据库仑定律这种相互作用的强度与原子间距离的平方成反比。离子对间的最佳距离是约  $2.8\text{\AA}$ 。在肽/蛋白相互作用中, 可在精氨酸、组氨酸或赖氨酸和天冬氨酸或谷氨酸之间形成静电键。该键的强度依赖于电离基团的  $pK_a$  和介质的介电常数, 尽管其与氢键的强度类似。亲脂相互作用是蛋白和肽配体之间发生的有利疏水-疏水相互作用。这种相互作用通常出现在埋于结合沟口袋中的肽的疏水性氨基酸侧链间, 以使它们不暴露在溶剂中。将疏水残基暴露于溶剂中是非常不利的, 因为周围的溶剂分子被迫在彼此间形成氢键而形成笼状结构。所致的熵值降低是非常不利的。亲脂性原子可以是既非极性又不是氢受体的硫和非极性的碳原子。范德华键是相距  $3-4\text{\AA}$  的原子间的非特异性的力。它比氢键和静电键弱、特异性低。原子周围的电荷分布随时间变化, 并且在任何瞬间电荷分布均是不对称的。这种瞬间的电荷不对称性诱导临近原子中的类似不对称性。在范德华接触距离所导致的原子之间的吸引力达到最大, 而在约  $1\text{\AA}$  到  $2\text{\AA}$  处迅速消失。相反, 当原子间隔的距离小

于此接触距离时，由于原子外部的电子云重叠使不断增强的斥力成为主导。尽管与静电和氢键相比，此吸引力相对较弱(约 0.6 Kcal/mol)，但所述斥力对于决定肽配体是否能与蛋白成功结合可能非常重要。

在一个实施方案中，利用 Böhm 评分函数(SCORE1 方法)评估结合常数(Böhm, H. J., J. Comput Aided Mol. Des., 8 (3): 243-256 (1994)，该文献在此全文引入作为参考)。在另一个实施方案中，用评分函数(SCORE2 方法)评估结合亲和力作为含有 T-细胞表位的配体的指示物(Böhm, H. J., J. Comput Aided Mol. Des., 12 (4): 309-323 (1998)，该文献在此全文引入作为参考)。但是上述文献中描述的 Böhm 评分函数是用于评估下述情况中配体对蛋白的结合亲和力的，即已知所述的配体可成功地与所述蛋白结合，且蛋白/配体复合物的结构已解析，这一结构已列于蛋白数据库("PDB")中。因此，利用已知的阳性结合数据对评分函数作了发展。为了区分阳性和阴性的结合体，需向方程中加入排斥项。此外，可通过以成对的方式计算亲脂相互作用，而非利用上述 Böhm 函数中基于面积的能量项来进行更理想的结合能量评估。因此，在一个优选实施方案中，用经修饰的 Böhm 评分函数评估结合能。在经修饰的 Böhm 评分函数中，在评估蛋白和配体之间的结合能( $\Delta G_{\text{bind}}$ )时考虑了下述参数：由于配体的平移和转动熵的整体损失造成的结合能减低( $\Delta G_0$ )；理想氢键的贡献( $\Delta G_{\text{hb}}$ )，其中至少一个配对物是中性的；无扰离子相互作用的贡献( $\Delta G_{\text{ionic}}$ )；亲脂配体原子和亲脂受体原子之间的亲脂相互作用( $\Delta G_{\text{lipo}}$ )；由于配体中内在自由度的冻结，即绕每一 C-C 键的旋转自由度降低造成的结合能损失( $\Delta G_{\text{rot}}$ )；蛋白和配体之间相互作用的能量( $E_{\text{vdw}}$ )。考虑到这些项给出等式 1：

$$(\Delta G_{\text{bind}}) = (\Delta G_0) + (\Delta G_{\text{hb}} \times N_{\text{hb}}) + (\Delta G_{\text{ionic}} \times N_{\text{ionic}}) + (\Delta G_{\text{lipo}} \times N_{\text{lipo}}) + (\Delta G_{\text{rot}} + N_{\text{rot}}) + (E_{\text{vdw}})$$

其中 N 是对于特定项限定的相互作用数目，在一个实施方案中， $\Delta G_0$ 、 $\Delta G_{\text{hb}}$ 、 $\Delta G_{\text{ionic}}$ 、 $\Delta G_{\text{lipo}}$  和  $\Delta G_{\text{rot}}$  是常数，其值分别为：5.4、-4.7、-4.7、-0.17 和 1.4。

$N_{hb}$  项依照等式 2 计算:

$$N_{hb} = \sum_{h-bond} f(\Delta R, \Delta \alpha) \times f(N_{neighb}) \times f_{pcs}$$

$f(\Delta R, \Delta \alpha)$  是罚函数, 其解决氢键自理想情况的巨大偏离, 其依照等式 3 计算:

$$f(\Delta R, \Delta \alpha) = f_1(\Delta R) \times f_2(\Delta \alpha)$$

其中:

如果  $\Delta R \leq TOL$  则  $f_1(\Delta R) = 1$ , 或者

如果  $\Delta R \leq 0.4 + TOL$  则  $f_1(\Delta R) = 1 - (\Delta R - TOL)/0.4$ , 或者

如果  $\Delta R > 0.4 + TOL$  则  $f_1(\Delta R) = 0$

并且:

如果  $\Delta \alpha < 30^\circ$  则  $f_2(\Delta \alpha) = 1$ , 或者

如果  $\Delta \alpha \leq 80^\circ$  则  $f_2(\Delta \alpha) = 1 - (\Delta \alpha - 30)/50$ , 或者

如果  $\Delta \alpha > 80^\circ$  则  $f_2(\Delta \alpha) = 0$

$TOL$  是氢键键长 =  $0.25 \text{ \AA}$  中所能允许的偏差

$\Delta R$  是 H-O/N 氢键键长与理想值 =  $1.9 \text{ \AA}$  的偏差

$\Delta \alpha$  是氢键键角  $\angle_{N/O-H..O/N}$  与  $180^\circ$  理想值的偏差

$f(N_{neighb})$  区分蛋白表面的凹凸部分, 并因此赋予口袋中而非蛋白表面的极性相互作用更高的权重。这一函数根据下述等式 4 计算:

$$f(N_{neighb}) = (N_{neighb}/N_{neighb,0})^\alpha, \text{ 其中 } \alpha = 0.5$$

$N_{neighb}$  为蛋白中与任意给定蛋白原子之间的距离小于  $5 \text{ \AA}$  的非氢原子的数量。

$N_{neighb,0}$  是常数 = 25

$f_{pcs}$  是用于估计每氢键的极性接触表面面积的函数, 由此区分强和弱的氢键, 其值由下述的标准确定:

当  $A_{polar}/N_{HB} < 10 \text{ \AA}^2$  时  $f_{pcs} = \beta$

当  $A_{polar}/N_{HB} > 10 \text{ \AA}^2$  时  $f_{pcs} = 1$

$A_{\text{polar}}$  是极性蛋白-配体接触面的大小

$N_{\text{HB}}$  是氢键的数目

$\beta$  是常数 = 1.2

由于假定了相同的几何相关性,在实施经修饰的 Böhmer 评分函数时,离子相互作用的贡献  $\Delta G_{\text{ionic}}$  用与上述有关氢键的类似方式计算。 $N_{\text{lipo}}$  项按下述的等式 5 计算:

$$N_{\text{lipo}} = \sum_{\text{IL}} f(r_{\text{IL}})$$

根据下述标准,对于所有亲脂配体原子 I 和所有亲脂蛋白原子 L,计算  $f(r_{\text{IL}})$ :

$$\text{当 } r_{\text{IL}} \leq R1 \text{ 时} \quad f(r_{\text{IL}}) = 1$$

$$\text{当 } R2 < r_{\text{IL}} < R1 \text{ 时} \quad f(r_{\text{IL}}) = (r_{\text{IL}} - R1)/(R2 - R1)$$

$$\text{当 } r_{\text{IL}} \geq R2 \text{ 时} \quad f(r_{\text{IL}}) = 0$$

其中:  $R1 = r_1^{\text{vdw}} + r_L^{\text{vdw}} + 0.5$

$$R2 = R1 + 3.0$$

$r_1^{\text{vdw}}$  是原子 I 的范德华半径

$r_L^{\text{vdw}}$  是原子 L 的范德华半径

$N_{\text{rot}}$  项是氨基酸侧链中可旋转的键的数目,其被视为无环的  $\text{sp}^3\text{-sp}^3$  及  $\text{sp}^3\text{-sp}^2$  键的数目。末端  $\text{-CH}_3$  或  $\text{-NH}_3$  的旋转未考虑进去。

最终,项  $E_{\text{vdw}}$  依照如下等式 6 计算:

$$E_{\text{vdw}} = \epsilon_1 \epsilon_2 ((r_1^{\text{vdw}} + r_2^{\text{vdw}})^{12}/r^{12} - (r_1^{\text{vdw}} + r_2^{\text{vdw}})^6/r^6),$$

其中:  $\epsilon_1$  和  $\epsilon_2$  是取决于原子身份的常数

$r_1^{\text{vdw}} + r_2^{\text{vdw}}$  是范德华原子半径

$r$  是原子对间的距离。

关于式 6,在一个实施方案中,  $\epsilon_1$  和  $\epsilon_2$  常数被赋予如下原子值,分别为: C: 0.245, N: 0.283, O: 0.316, S: 0.316 (即分别对于碳、氮、氧和硫原子)。对于式 5 和 6,给予范德华半径如下原子值,分别为 C: 1.85, N: 1.75, O: 1.60, S: 2.00 Å。

应当理解上述等式中所有预定的值和给定的常数都是在现有的对蛋白配体相互作用的理解决限内具体相对于此处所用的计算类型确定

的。因此，随着这种评分函数的进一步精练，这些值和常数也会因此而改变，任何能在蛋白和配体结合能的评估方面给出所需结果的适宜数值均可使用，而且，其也落入本发明的保护范围。

如上所述，所述的评分函数应用于由上述侧链构象、原子身份和原子间距数据库中提取的数据。为本说明书的目的，该数据库中包含的 MHC II 类分子数是 42 个模型加上 4 个已解析的结构。从上述描述中可清楚地了解到，本发明的计算机构建方法的模块性质意味着，可简单地添加新的模型，并利用肽主链文库和侧链构象搜索功能进行扫描以创建其它的可通过上述的肽评分函数处理的数据集。这使得经扫描的 MHC II 类分子库可以很容易地增加，或者如果可以获得相关数据，则可以替换结构和相关数据以创建现有等位基因的更精确的模型。

本发明的预测方法可以相对于包含大量已通过实验确定了其对不同 MHC II 类分子的亲和力的肽的数据集进行校准。将计算值与实验数据相比较，可确定一截断值，已知该值之上所有经实验确定的 T-细胞表位都得以正确的预测。

应当理解，尽管上述评分函数与现有的一些复杂方法相比相对简单，但计算进行得非常迅速。还应当理解的是，其目的并不在于计算出对接到所选择的 MHC II 类蛋白结合沟内的每种肽的真正结合能本身。根本的目的在于获得相对的结合能数据以助于根据所选蛋白的一级结构(即氨基酸序列)预测 T-细胞表位的定位。相对高的结合能或结合能高于选定的阈值意味着在配体中存在 T-细胞表位。然后将所述配体进行至少一轮氨基酸替代，并再次计算结合能。由于计算可迅速进行，对肽序列的这些操作可在现有成本划算的计算机硬件上于程序用户界面中互动进行。由此不需要对计算机硬件进行大量投资。本领域的技术人员应当了解，也可使用其他软件达到相同的目的。特别是可以使用能将配体对接入蛋白结合位点的更复杂的软件，并与能量最小化相结合。对接软件的例子包括：DOCK (Kuntz 等, J. Mol. Biol., 161: 269-288 (1982)), LUDI (Böhm, H. J., J. Comput Aided Mol. Des., 8: 623-632 (1994))和 FLEXX (Rarey M.等, ISMB, 3: 300-308 (1995))。分子建模和

操作软件的例子包括：AMBER (Tripos)和 CHARMM (Molecular Simulations Inc.)。使用这些计算机方法将严重限制本发明方法的信息吞吐量，这是由于进行必要的计算所需的处理时间导致的。但是可行的方式为，将这些方法作为‘二级筛选’以获得关于通过本发明的方法发现为‘阳性结合体’的肽的更精确的肽结合能计算值。用于复杂的分子机械或分子动力学计算的处理时间的限制是由进行所述计算的软件设计和目前计算机硬件技术限制共同确定的。可以预期在将来，随着编写更高效的代码和计算机处理器速度的不断提高，在更易控制的时间框架内进行上述计算是可行的。

有关用于大分子的能量函数的其他信息和有关在折叠蛋白结构内发生的多种相互作用的考虑可参考下述文献：Brooks, B. R., 等., J. Comput. Chem., 4 : 187-217 (1983), 有关蛋白-配体一般相互作用的信息参见：Dauber-Osguthorpe 等, Proteins 4 (1) : 31-47 (1988), 这些文献均全文引入作为参考。其他有用的背景资料也可参见 Fasman, G. D.编, Prediction of Protein Structure and the Principles of Protein Conformation, Plenum Press, New York, ISBN: 0-306 4313-9。