



(12) 发明专利申请

(10) 申请公布号 CN 115776987 A

(43) 申请公布日 2023.03.10

(21) 申请号 202180039282.4

(22) 申请日 2021.03.31

(30) 优先权数据

63/003,177 2020.03.31 US

(85) PCT国际申请进入国家阶段日

2022.11.29

(86) PCT国际申请的申请数据

PCT/US2021/025167 2021.03.31

(87) PCT国际申请的公布数据

WO2021/202727 EN 2021.10.07

(71) 申请人 瑞佩尔托利免疫医药股份有限公司

地址 美国马萨诸塞州

(72) 发明人 O·约斯弗森 T·马利亚

A·哈梅尔 J·斯怀恩

B·罗斯寇 M·拉什蒂安

(74) 专利代理机构 上海专利商标事务所有限公司 31100

专利代理师 陶家蓉

(51) Int.Cl.

G07K 14/705 (2006.01)

权利要求书9页 说明书102页 附图25页

(54) 发明名称

可条码化可交换肽-MHC多聚体文库

(57) 摘要

提供了MHC多聚体,其中加载肽的MHC单体通过单体和多聚结构域上的偶联部分共价连接到多聚结构域。多聚体还可以包括寡核苷酸条码。可以用多个pMHC多聚体进行肽交换以创建pMHC多聚体文库。还提供了制备和使用pMHC多聚体和文库的方法。提供了加载肽的MHC I类和MHC II类多聚体及其文库。

1. 一种生产主要组织相容性复合物 (MHC) 多聚体的方法, 该方法包括:
  - (a) 提供两个或多个MHC单体, 其中每个单体包括偶联部分;
  - (b) 提供多聚结构域, 其中所述多聚结构域的每个亚基包括偶联部分;
  - (c) 在足以使MHC单体和多聚结构域共价偶联的条件下, 将MHC单体和多聚化结构域组合以产生MHC多聚体。
2. 如权利要求1所述的方法, 其中MHC单体是MHC I类单体。
3. 如权利要求1所述的方法, 其中MHC单体是MHC II类单体。
4. 如权利要求1-3中任一项所述的方法, 其中所述MHC单体在与所述多聚结构域结合之前加载占位符肽。
5. 如权利要求1-4任一项所述的方法, 其中所述多聚结构域包括非共价结合位点, 进一步地, 其中用寡核苷酸条码标记所述MHC多聚体, 所述寡核苷酸条码结合所述多聚结构域的非共价结合位点。
6. 如权利要求1-5中任一项所述的方法, 其中所述多聚结构域是链霉亲和素。
7. 如权利要求5所述的方法, 其中所述多聚结构域是链霉亲和素, 并且所述寡核苷酸条码结合链霉亲和素的生物素结合位点。
8. 如权利要求1-7任一项所述的方法, 其中每个MHC单体的偶联部分包括X, 而多聚结构域的每个亚基的偶联部分包括Y, 其中
  - (i) X是末端炔烃, Y是叠氮化物;
  - (ii) X是叠氮化物, Y是末端炔烃;
  - (iii) X是应变炔烃, Y是叠氮化物;
  - (iv) X是叠氮化物, Y是应变炔烃;
  - (v) X是二烯, Y是亲二烯;
  - (vi) X是亲二烯, Y是二烯;
  - (vii) X是硫醇, Y是烯烃; 或
  - (viii) X是烯烃, Y是硫醇。
9. 如权利要求8所述的方法, 其中所述叠氮化物是铜整合叠氮化物。
10. 如权利要求9所述的方法, 其中所述铜整合叠氮化物是吡啶甲基叠氮化物。
11. 如权利要求1-10任一项所述的方法, 其中每个MHC单体的偶联部分和多聚结构域的每个亚基的偶联部分包含分选标签基序。
12. 如权利要求1-10任一项所述的方法, 其中每个MHC单体的偶联部分和多聚结构域的每个亚基的偶联部分包含内含蛋白序列。
13. 如权利要求1-12中任一项所述的方法, 其中所述多聚体是四聚体。
14. 如权利要求4所述的方法, 其进一步包括将占位符肽与结合MHC单体的拯救肽表位交换。
15. 一种条码标记的MHC多聚体, 包括:
  - (a) 两个或多个MHC单体;
  - (b) 多聚结构域, 其包含两个或多个亚基并具有至少一个非共价结合位点; 和
  - (c) 一种寡核苷酸条码;其中每个MHC单体通过共价连接与多聚结构域的亚基结合; 和

其中所述寡核苷酸条码通过与所述多聚结构域上的所述非共价结合位点的非共价键合而与所述多聚结构域结合。

16. 如权利要求15所述的MHC多聚体,其进一步包含加载到所述多聚体的每个MHC单体上的MHC结合肽。

17. 如权利要求15或16所述的MHC多聚体,其中所述MHC单体是MHC I类单体。

18. 如权利要求15或16所述的MHC多聚体,其中所述MHC单体是MHC II类单体。

19. 如权利要求15-18中任一项所述的MHC多聚体,其为四聚体。

20. 如权利要求19所述的MHC多聚体,其中所述多聚结构域是链霉亲和素。

21. 如权利要求20所述的MHC多聚体,其中所述寡核苷酸条码与链霉亲和素上的生物素结合位点非共价结合。

22. 如权利要求15-21任一项所述的方法,其中每个MHC单体包括偶联部分X,而多聚结构域的每个亚基包括偶联部分Y,其中

(i) X是末端炔烃,Y是叠氮化物;

(ii) X是叠氮化物,Y是末端炔烃;

(iii) X是应变炔烃,Y是叠氮化物;

(iv) X是叠氮化物,Y是应变炔烃;

(v) X是二烯,Y是亲二烯;

(vi) X是亲二烯,Y是二烯;

(vii) X是硫醇,Y是烯烃;或

(viii) X是烯烃,Y是硫醇。

23. 如权利要求22所述的MHC多聚体,其中所述叠氮化物是铜整合叠氮化物。

24. 如权利要求23所述的MHC多聚体,其中所述铜整合叠氮化物是吡啶甲基叠氮化物。

25. 如权利要求15-24中任一项所述的MHC多聚体,其中所述多聚结构域的每个MHC单体和每个亚基包括偶联部分,其中每个偶联部分包括分选标签基序。

26. 如权利要求15-24中任一项所述的MHC多聚体,其中所述多聚结构域的每个MHC单体和每个亚基包括偶联部分,其中每个偶联部分包括内含蛋白序列。

27. 一种产生加载肽的主要组织相容性复合物I类(pMHCI)多聚体的方法,该方法包括:

(a) 提供两个或多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

(b) 提供多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;

(c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将所述p\*MHCI单体和所述多聚结构域组合以产生p\*MHCI多聚体;和

(d) 用拯救肽表位替换结合在p\*MHCI多聚体中的每个p\*MHC I单体的肽结合沟中的占位符肽,以产生pMHCI多聚体。

28. 一种产生加载条码化肽的主要组织相容性复合物I类(pMHCI)多聚体的方法,该方法包括:

(a) 提供两个或多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个

MHCI单体的肽结合沟中的占位符肽；

(b) 提供多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分,并且所述多聚结构域包括至少一个非共价结合位点；

(c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将所述p\*MHCI单体和所述多聚结构域组合以产生p\*MHCI多聚体;和

(d) 用拯救肽表位替换结合在p\*MHCI多聚体中的每个p\*MHC I单体的肽结合沟中的占位符肽,以产生pMHCI多聚体;和

(e) 将寡核苷酸条码结合到多聚结构域上的非共价结合位点。

29. 一种产生加载肽的主要组织相容性复合物I类(pMHCI)多聚体的方法,该方法包括:

(a) 提供两个或多个加载占位符肽的MHCI (p\*MHCI) 单体,其各自包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 位于(i) 或(ii) C末端处的包含偶联部分的肽接头,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

(b) 提供多聚结构域,其在所述多聚结构域的每个亚基C末端处包含肽接头,所述肽接头包含偶联部分;

(c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将所述p\*MHCI单体和所述多聚结构域组合以产生p\*MHCI多聚体;和

(d) 用拯救肽表位替换结合在p\*MHCI多聚体中的每个p\*MHC I单体的肽结合沟中的占位符肽,以产生pMHCI多聚体。

30. 如权利要求27-29中任一项所述的方法,其中每个p\*MHC单体均为脊椎动物来源。

31. 如权利要求27-30中任一项所述的方法,其中每个p\*MHCI单体包括人MHCI重链多肽或其功能片段,以及人 $\beta$ 2-微球蛋白多肽或功能片段。

32. 如权利要求31所述的方法,其中其每个p\*MHCI单体是HLA-A、HLA-B或HLA-C。

33. 如权利要求32所述的方法,其中每个p\*MHCI单体是HLA-A。

34. 如权利要求27-33中任一项所述的方法,其中每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1结构域。

35. 如权利要求27-34中任一项所述的方法,其中每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1/ $\alpha$ 2异二聚体。

36. 如权利要求27-35中任一项所述的方法,其中每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1、 $\alpha$ 2和 $\alpha$ 3结构域。

37. 如权利要求27-36任一所述的方法,其中每个p\*MHCI单体的MHCI重链多肽或其功能片段包含与SEQ ID NO:28-159所示的任何氨基酸序列至少80%、85%、90%、95%或99%相同的 $\alpha$ 结构域。

38. 如权利要求27-37中任一项所述的方法,其中每个p\*MHCI单体是可溶的。

39. 如权利要求27-38中任一项所述的方法,其中每个p\*MHCI单体的 $\beta$ 2-微球蛋白多肽是野生型人 $\beta$ 2-微球蛋白。

40. 如权利要求27-39任一所述的方法,其中 $\beta$ 2-微球蛋白多肽包含与人 $\beta$ 2-微球蛋白的氨基酸序列(SEQ ID NO:2或160)至少80%、85%、90%、95%或99%相同的氨基酸序列。

41. 如权利要求27-40中任一项所述的方法,其中每个p\*MHC单体是融合蛋白,包括MHCI重链多肽或其功能片段,以及 $\beta$ 2-微球蛋白或其功能片段。



42. 如权利要求41所述的方法,其中p\*MHC融合蛋白包括MHCI重链多肽或其功能片段以及 $\beta$ 2-微球蛋白或其功能片段之间的肽接头。

43. 如权利要求27-42任一所述的方法,其中占位符肽是促进MHCI多肽折叠的肽或肽样化合物。

44. 如权利要求27-43中任一项所述的方法,其中所述占位符肽是2至25个氨基酸。

45. 如权利要求27-44中任一项所述的方法,其中所述占位符肽是8至11个氨基酸。

46. 如权利要求27-45中任一项所述的方法,其中所述占位符肽是9至10个氨基酸。

47. 如权利要求27-46任一所述的方法,其中与交换肽表位相比,占位符肽对MHCI肽结合沟的亲合力更低,并且其中步骤(d)包括在竞争测试中使p\*MHCI单体与过量肽表位接触。

48. 如权利要求47任一所述的方法,其中占位符肽的KD比交换肽表位低约10倍。

49. 如权利要求27-46中任一项所述的方法,其中所述占位符肽在约30-37°C的温度下是不稳定的,并且步骤(d)包括在存在肽表位的情况下将p\*MHCI单体暴露于约30-37°C。

50. 如权利要求27-46中任一项所述的方法,其中所述占位符肽在约pH2.5-5.5之间的酸性pH下是不稳定的,并且步骤(d)包括在肽表位存在下将p\*MHCI单体暴露于约pH 2.5-5之间的pH。

51. 如权利要求27-46中任一项所述的方法,其中所述占位符肽在约pH 9-11之间的碱性pH下是不稳定的,并且步骤(d)包括在肽表位存在下将p\*MHCI单体暴露于约pH 9-11之间的pH。

52. 如权利要求27-51中任一项所述的方法,其中所述占位符肽包括GILGFVJL (SEQ ID NO:7)。

53. 如权利要求27-51中任一项所述的方法,其中所述占位符肽由GILGFVJL (SEQ ID NO:7) 组成。

54. 如权利要求27-46中任一项所述的方法,其中所述占位符肽包括可切割部分,并且步骤(d)包括在足以切割所述占位符肽的条件下使所述p\*MHCI单体与肽表位接触。

55. 如权利要求54所述的方法,其中可切割部分是可光切割氨基酸,并且步骤(d)包括在足以诱导占位符肽中的可光切割部分切割和肽表位与MHCI单体结合的条件下将p\*MHCI单体暴露于UV光。

56. 如权利要求55所述的方法,其中所述可光切割氨基酸包括2-硝基苯基侧链。

57. 如权利要求55所述的方法,其中所述可光切割氨基酸包括3-氨基-3-(2-硝基苯基)丙酸。

58. 根据权利要求55所述的方法,其中所述可光切割氨基酸是(2-硝基)苯基甘氨酸。

59. 如权利要求55所述的方法,其中可光切割的占位符肽和与其结合的对应该MHC分子选自:A\*02:01,KILGFVJL (SEQ ID NO:15) 或GILGFVJL (SEQ ID NO:7), A\*01:01, STAPGJLEY (SEQ ID NO:16); A\*02:03, SVRDJLARL (SEQ ID NO:271); A\*02:06, LTAJFLIFL (SEQ ID NO:272); A\*02:07, LLDSJERL (SEQ ID NO:273); A\*02:11, KMDIJVPLL (SEQ ID NO:274); A\*03:01, RIYRJGATR (SEQ ID NO:17); A\*11:01, RVFAJSFIK (SEQ ID NO:18); A\*24:02, VYGJVRACL (SEQ ID NO:11); A\*33:03, FYVJGAANR (SEQ ID NO:275); B\*07:02, AARGJTLAM (SEQ ID NO:14); B\*15:02, ILGPPGJVY (SEQ ID NO:276); B\*35:01, KPIVVLJGY (SEQ ID NO:19); B\*44:05, EEFGAAJSF (SEQ ID NO:277); B\*46:01, KMKEIAJAY (SEQ ID NO:278); B\*55:02, KPWDJIPMV

(SEQ ID NO:279);C\*03:04,FVYGJSKTSL (SEQ ID NO:20),B\*08:01,FLRGRAJGL (SEQ ID NO:21);C\*07:02,VRIJHLYIL (SEQ ID NO:22);C\*04:01,QYDJAVYKL (SEQ ID NO:23);B\*15:01,ILGPJGSVY (SEQ ID NO:24);B\*40:01,TEADVQJWL (SEQ ID NO:25);B\*58:01,ISARGQJLF (SEQ ID NO:26);和C\*08:01,KAAJDLSHFL (SEQ ID NO:27),其中J是3-氨基-3-(2-硝基苯基)丙酸。

60.如权利要求54所述的方法,其中可切割部分包含化学选择部分,并且步骤(d)包括在足以切割化学选择部分的条件下使p\*MHCI单体与肽表位接触。

61.如权利要求60所述的方法,其中化学选择部分是连二亚硫酸钠敏感的偶氮苯接头,并且步骤(d)包括在连二亚硫酸钠存在下使p\*MHCI单体与肽表位接触。

62.如权利要求54所述的方法,其中可切割部分是高碘酸盐敏感氨基酸,且步骤(d)包括在足以切割占位符肽的条件下将p\*MHCI单体与肽表位接触。

63.如权利要求62所述的方法,其中所述高碘酸盐敏感氨基酸包括邻二醇部分或邻氨基醇。

64.如权利要求63所述的方法,其中高碘酸盐敏感氨基酸是 $\alpha, \gamma$ -二氨基- $\beta$ -羟基丁酸(DAHB)。

65.如权利要求54所述的方法,其中所述可切割部分是蛋白酶识别部分。

66.如权利要求65所述的方法,其中所述蛋白酶是氨基-肽酶。

67.如权利要求66所述的方法,其中所述蛋白酶是甲硫氨酸氨基-肽酶。

68.如权利要求27-46中任一项所述的方法,其中所述占位符肽是二肽。

69.如权利要求68所述的方法,其中所述二肽结合到所述MHCI结合沟的F口袋。

70.如权利要求68或69所述的方法,其中所述二肽的第二氨基酸是疏水性的。

71.如权利要求68-70任一所述的方法,其中二肽选自甘氨酸-亮氨酸(GL)、甘氨酸-缬氨酸(GV)、甘氨酸-甲硫氨酸(GM)、甘氨酸-环己基丙氨酸(GCha)、甘氨酸-高亮氨酸(GH1e)和甘氨酸-苯丙氨酸(GF)。

72.如权利要求27-71中任一项所述的方法,其中所述多聚结构域是二聚体、四聚体、六聚体、八聚体、十聚体或十二聚体。

73.如权利要求27-72中任一项所述的方法,其中所述多聚结构域是四聚体。

74.如权利要求27-73中任一项所述的方法,其中所述多聚结构域是同多聚体。

75.如权利要求27-73中任一项所述的方法,其中所述多聚结构域是异多聚体。

76.如权利要求27-75中任一项所述的方法,其中所述多聚结构域包括链霉亲和素或其衍生物。

77.如权利要求76所述的方法,其中所述多聚结构域是链霉亲和素或其衍生物的四聚体。

78.如权利要求76所述的方法,其中所述多聚结构域包括Strep-tag®或Strep-tactin®。

79.如权利要求27-78中任一项所述的方法,其中所述偶联部分连接到每个p\*MHCI单体的MHCI重链 $\alpha 1$ 结构域的C末端,并且步骤(c)包括将所述多聚化结构域共价偶联到所述MHCI  $\alpha 1$ 结构的C末端。

80.如权利要求27-78中任一项所述的方法,其中所述偶联部分连接到每个p\*MHCI单体

的MHCI重链 $\alpha 2$ 结构域的C末端,并且步骤(c)包括将所述多聚化结构域共价偶联到所述MHCI $\alpha 2$ 结构的C末端。

81.如权利要求27-78中任一项所述的方法,其中所述偶联部分连接到每个p\*MHCI单体的MHCI重链 $\alpha 3$ 结构域的C末端,并且步骤(c)包括将所述多聚化结构域共价偶联到所述MHCI $\alpha 3$ 结构的C末端。

82.如权利要求27-78中任一项所述的方法,其中所述偶联部分连接到每个p\*MHCI单体的每个p\*MHC单体的 $\beta 2$ -微球蛋白的C末端,并且步骤(c)包括将所述多聚结构域共价偶联到每个p\*MHC单体的 $\beta 2$ -微球蛋白的C末端。

83.如权利要求27-82中任一项所述的方法,其中步骤(c)包括所述两个或多个p\*MHC单体与所述多聚化结构域的化学偶联。

84.如权利要求83所述的方法,其中所述化学偶联由半胱氨酸生物偶联介导。

85.如权利要求83所述的方法,其中所述化学偶联由天然化学偶联介导。

86.如权利要求83所述的方法,其中所述化学偶联由点击化学介导。

87.如权利要求86所述的方法,其中每个p\*MHCI单体的偶联部分包括X,并且多聚化结构域的每个亚基的偶联部分包括Y,其中

(i) X是末端炔烃,Y是叠氮化物;

(ii) X是叠氮化物,Y是末端炔烃;

(iii) X是应变炔烃,Y是叠氮化物;

(iv) X是叠氮化物,Y是应变炔烃;

(v) X是二烯,Y是亲二烯;

(vi) X是亲二烯,Y是二烯;

(vii) X是硫醇,Y是烯炔;或

(viii) X是烯炔,Y是硫醇。

88.如权利要求87所述的方法,其中所述叠氮化物是铜螯合叠氮化物。

89.如权利要求88所述的方法,其中所述铜螯合叠氮化物是吡啶甲基叠氮化物。

90.如权利要求27-89中任一项所述的方法,其中每个p\*MHC结构域的偶联部分包括连接到N末端的肽接头,并且多聚结构域的每个亚基的偶联部分包括附接到多聚结构域的每个亚基的C末端的肽接头。

91.如权利要求90所述的方法,其中每个p\*MHC单体的C末端的肽接头包含 $(G)_n-X$ ,其中n为至少2,并且X是适合于化学偶联的部分,并且多聚结构域的每个亚基的C末端处的肽接头包括 $(G)_n-Y$ ,其中n为至少2,且Y是适合于与每个p\*MHC单体的X部分化学偶联的部分。

92.如权利要求27-78中任一项所述的方法,其中每个p\*MHCI单体与多聚结构域的共价偶联由半胱氨酸转肽酶介导。

93.如权利要求92所述的方法,其中所述半胱氨酸转肽酶是分选酶或其酶活性片段。

94.如权利要求93所述的方法,其中每个p\*MHC结构域的偶联部分包括C末端分选标签,并且多聚结构域的每个亚基的偶联部分包括N末端分选标签,和步骤(c)包括向p\*MHC I单体和多聚结构域的混合物中添加分选酶,并催化每个p\*MHC单体和多聚结构域之间肽键的形成,以制备p\*MHC-多聚体。

95.如权利要求92所述的方法,其中

步骤(a)中的两个或多个p\*MHCI单体是通过在包含适合于化学偶联的部分的肽接头存在下,使包含C末端分选标签的p\*MHC I单体与分选酶或其酶活性片段接触而产生的,其中分选酶或其酶活性片段介导肽接头与p\*MHC单体的偶联;

步骤(b)中的多聚结构域是通过在包含适合于化学偶联的部分的肽接头的存在下使包含C末端分选标签的多聚结构域与分选酶接触而产生的,其中所述分选酶或其酶活性片段介导肽接头与多聚结构域的每个亚基的N末端的偶联;和

步骤(c)包括在两个或多个p\*MHC单体的C末端的肽接头和多聚结构域的每个亚基的C末端的肽接头之间化学偶联,以产生p\*MHC多聚体。

96.如权利要求95所述的方法,其中每个p\*MHC单体的C末端的肽接头包含 $(G)_n-X$ ,其中n为至少2,并且X是适合于点击化学偶联的部分,并且多聚结构域的每个亚基的N末端处的肽接头包括 $(G)_n-Y$ ,其中n为至少2,且Y是适合于与每个p\*MHC单体的X部分点击化学偶联的部分。

97.如权利要求96所述的方法,其中

- (i) X是末端炔烃, Y是叠氮化物;
- (ii) X是叠氮化物, Y是末端炔烃;
- (iii) X是应变炔烃, Y是叠氮化物;
- (iv) X是叠氮化物, Y是应变炔烃;
- (v) X是二烯, Y是亲二烯;
- (vi) X是亲二烯, Y是二烯;
- (vii) X是硫醇, Y是烯炔;或
- (viii) X是烯炔, Y是硫醇。

98.如权利要求97所述的方法,其中所述叠氮化物是铜螯合叠氮化物。

99.如权利要求98所述的方法,其中所述铜螯合叠氮化物是吡啶甲基叠氮化物。

100.如权利要求93-99中任一项所述的方法,其中所述分选酶或其酶活性片段是Ca<sup>2+</sup>依赖性的。

101.如权利要求93-99中任一项所述的方法,其中所述分选酶或其酶活性片段是Ca<sup>2+</sup>非依赖性的。

102.如权利要求93-99中任一项所述的方法,其中所述分选酶是所述野生型分选酶的可溶性片段。

103.如权利要求93-99中任一项所述的方法,所述分选酶是金黄色葡萄球菌分选酶A的变体或同源物。

104.如权利要求93-99中任一项所述的方法,其中所述分选酶是所述修饰的分选酶A的可溶性片段。

105.如权利要求93-99中任一项所述的方法,其中所述分选酶是SrtAstaph突变体。

106.如权利要求105所述的方法,其中所述SrtAstaph突变体选自F40、SrtAstaph五突变体、2A-9和4S-9。

107.如权利要求27-78中任一项所述的方法,其中每个p\*MHCI单体与多聚结构域的共价偶联由内含蛋白介导。

108.如权利要求107所述的方法,其中所述内含蛋白选自MxeGyrA、SspDnaE、NpuDnaE、

AvaDnaE, Cfa (共有DnaE裂解内含蛋白)、gp41-1、gp41-8和NrdJ-1。

109. 如权利要求107所述的方法,其中,所述内含蛋白是裂解内含蛋白对。

110. 如权利要求27-78中任一项所述的方法,其中每个p\*MHCI单体与多聚结构域的共价偶联由内含蛋白肽标签介导。

111. 如权利要求27-78任一所述的方法,其中每个p\*MHCI单体在C末端包括N-内含蛋白片段,并且多聚结构域的每个亚基在N末端包括Npu-C-内含蛋白片段。

112. 如权利要求27-111中任一项所述的方法,其中所述多聚结构域包括标识符。

113. 如权利要求112所述的方法,其中,所述标识符是核酸标识符。

114. 如权利要求113所述的方法,其中所述核酸标识符编码所述肽。

115. 如权利要求113或114所述的方法,其中所述核酸标识符的长度为25个核苷酸至500个核苷酸。

116. 如权利要求113或114所述的方法,其中所述核酸标识符的长度为40个核苷酸至120个核苷酸。

117. 一种产生包含加载多种肽表位的MHCI (pMHCI) 多聚体的文库的方法,该方法包括:

(a) 提供多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含 (i) MHCI重链多肽或其功能片段, (ii)  $\beta$ 2-微球蛋白多肽或其功能片段, (iii) 偶联部分,和 (iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

(b) 提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;

(c) 在足以使两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,使p\*MHCI单体和多聚结构域接触,以产生p\*MHCI多聚体;和

(d) 用包含多个独特MHCI肽表位的肽文库替换多个p\*MHCI多聚体中的占位符肽,以产生多个加载肽的MHCI (pMHCI) 多聚体。

118. 一种产生包含多种条码化且加载肽的主要组织相容性复合物I类 (pMHCI) 多聚体的文库的方法,该方法包括:

(a) 提供多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含 (i) MHCI重链多肽或其功能片段, (ii)  $\beta$ 2-微球蛋白多肽或其功能片段, (iii) 偶联部分,和 (iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

(b) 提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分,并且其中所述多聚结构域包括至少一个非共价结合位点;

(c) 在足以使所述两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,将所述多个p\*MHCI单体和多个多聚结构域接触以产生多个p\*MHCI多聚体;

(d) 用多个独特拯救肽表位替换结合在p\*MHCI多聚体的肽结合沟中的占位符肽,以产生多个pMHCI多聚体;和

(e) 将寡核苷酸条码结合到多聚结构域上的非共价结合位点。

119. 一种产生包含多种条码化且加载肽的主要组织相容性复合物I类 (pMHCI) 多聚体的文库的方法,该方法包括:

(a) 提供多个加载占位符肽的MHCI (p\*MHCI) 单体,其各自包含 (i) MHCI重链多肽或其功能片段, (ii)  $\beta$ 2-微球蛋白多肽或其功能片段, (iii) 位于 (i) 或 (ii) C末端处的包含偶联部分的肽接头,和 (iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

(b) 提供多个多聚结构域,其在所述多聚结构域的每个亚基C末端处包含肽接头,所述肽接头包含偶联部分;

(c) 在足以使所述两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,将所述多个p\*MHCI单体和多个多聚结构域接触以产生多个p\*MHCI多聚体;和

(d) 用多个独特拯救肽表位替换结合在p\*MHCI多聚体的肽结合沟中的占位符肽,以产生多个pMHCI多聚体。

120. 如权利要求117-119中任一项所述的方法,其中所述MHCI肽表位文库是高多样性肽文库。

121. 如权利要求120所述的方法,其中肽文库包含约 $10^3$ 、约 $10^4$ 、约 $10^5$ 、约 $10^6$ 、约 $10^7$ 、约 $10^8$ 、约 $10^9$ 、约 $10^{10}$ 、约 $10^{11}$ 、约 $10^{12}$ 、约 $10^{13}$ 、约 $10^{14}$ 、约 $10^{15}$ 、约 $10^{16}$ 、约 $10^{17}$ 、约 $10^{18}$ 、约 $10^{19}$ 、约 $10^{20}$ 或更多不同的MHCI肽表位。

122. 如权利要求117-120中任一项所述的方法,其中所述MHCI肽表位衍生自单个抗原蛋白。

123. 如权利要求122所述的方法,其中所述MHCI肽表位包含抗原蛋白的重叠片段。

124. 如权利要求117-121中任一项所述的方法,其中所述多个独特肽表位由生物体的基因组、生物体的转录组、生物体的蛋白质组、生物体中的肽或蛋白质产生。

125. 如权利要求117-121中任一项所述的方法,其中所述多个独特肽表位由两个基因组之间的差异序列产生。

126. 如权利要求117-125中任一项所述的方法,其中所述pMHC多聚体中的每一个包括独特标识符部分。

127. 如权利要求126所述的方法,其中所述独特标识符部分是核酸。

128. 一种包含多个加载肽的MHCI (pMHCI) 多聚体的多肽文库,其中每个加载肽的pMHCI多聚体包含两个或多个与多聚结构域偶联的pMHC I单体。

129. 一种分离MHC-多聚体结合淋巴细胞的方法,其中所述方法包括:

(a) 使多个淋巴细胞与pMHCI多聚体文库接触;和

(b) 产生多个隔室,其中每个隔室包括与文库的pMHCI多聚体结合的淋巴细胞和捕获载体。

130. 根据权利要求129所述的方法,其中淋巴细胞是T细胞、B细胞或NK细胞。

131. 一种鉴定与pMHC多聚体结合的淋巴细胞的方法,该方法包括:

(a) 使多个淋巴细胞与pMHCI多聚体文库接触;

(b) 将与文库的pMHCI多聚体结合的多个淋巴细胞中的一个淋巴细胞划分为单个隔室,其中所述pMHCI多聚体包括独特标识符;和

(c) 确定与所述隔室化淋巴细胞结合的pMHCI的独特标识符。

## 可条码化可交换肽-MHC多聚体文库

### [0001] 相关申请

[0002] 本申请要求2020年3月31日提交的美国临时专利申请第63/003,177号的优先权,其完整内容通过引用纳入本文。

### 背景技术

[0003] 鉴定单个T细胞识别的肽对于理解和治疗免疫相关疾病以及预防疾病的疫苗开发非常重要。检测抗原反应性T细胞的技术利用了给定TCR与其肽-MHC (pMHC) 识别基序之间的相互作用。制备可溶性MHC分子的能力允许制备可溶性肽-MHC复合物,然后可将其制成多聚复合物。在各种研究和临床情况下,使用多聚化pMHC分子检测T细胞已成为检测抗原特异性T细胞的首选方法。

[0004] 自从Altman等人 (Science 274:94-961996) 表明,加载肽的MHC I类 (pMHCI) 分子的四聚化为T细胞受体 (TCR) -pMHC相互作用提供了足够的稳定性,从而允许使用流式细胞术检测荧光标记的MHC多聚体结合性T细胞以来,MHC多聚体已被用于检测抗原反应性T细胞。然而,由于当MHC I类分子不是与肽的复合物的一部分时,它们在很大程度上是不稳定的,因此基于pMHCI的技术最初受到了繁琐的分子生产的限制,其中每个肽都需要单独的折叠和纯化程序 (Bakker等人,Curr.Opin.Immunol.17:428-4332005)。

[0005] 最近,已经报道了具有共价连接肽的多种MHCI分子 (例如,由Goldberg等人,J.Cell.Mol.Med.15:1822-1832,2011综述)。还开发了几种类型的pMHCI微阵列系统,但大多数工作都集中在优化支持表面和修改结合和/或洗涤过程中应用的条件。这些系统的使用也受到限制,因为与现有的基于细胞术的分析相比,检测极限差且再现性低。例如,这种基于阵列的策略的一般限制是,给定T细胞存在追求给定阵列上显示的所有可能pMHCI相互作用的倾向。因此,细胞制剂中抗原反应性T细胞的频率通常需要>0.1%,以实现稳健的读取。

[0006] MHCI多聚体及其文库已使用生物素化肽-MHCI单体制备,然后与链霉亲和素上的生物素结合位点结合以形成四聚体 (参见例如Leisner等人,PLoS One 3(2):e16782008)。为了创建MHC I类文库,已经描述了寡核苷酸条码标签与链霉亲和素偶联的方法。然而,现有的策略涉及复杂和/或昂贵的方法,这些方法限制了大型文库的简易生产。例如,在一种方法中,在生物素化肽-HLA单体的四聚化之前,必须通过重叠延伸PCR对单个链霉亲和素前体进行单独条码化 (Zhang等人,Nature Biotech.2018;doi:10.1038.nbt.4282)。在另一种方法中,链霉亲和素偶联的葡聚糖是一种昂贵的试剂,用于产生一种右旋体 (dexramer),生物素化肽-HLA单体和生物素化条码寡核苷酸通过链霉亲和素偶联到葡聚糖骨架与之复合 (Bentzen等人,Nature Biotech.34:10:1037-10452016)。

[0007] 与采用pMHCI四聚体的方法类似,可溶性MHC II类分子也被用于制备pMHCII四聚体,其已用于研究CD4+T辅助细胞的抗原特异性 (综述于例如Nepom等人 (2002) Arthrit.Rheumat.46:5-12;Vollers和Stern (2008) Immunol.123:305-313;Cecconi等人 (2008) Cytometry 73A:1010-1018)。通常为了制备pMHCII多聚体,可溶性生物素化MHCII $\alpha$ /

$\beta$ 二聚体被重组表达,然后通过其生物素结合位点与链霉亲和素或亲和素结合而四聚。链霉亲和素或亲和素的荧光标记允许通过流式细胞术分离结合pMHCII多聚体的T细胞。关于MHCII分子的抗原肽加载,在一种方法中,肽共价连接到MHCII $\alpha/\beta$ 二聚体上。一些小组已经产生了加载有共价但可切割的“填充”肽的pMHCII,该肽可以在酸性条件下与感兴趣的肽交换(Day等人, *J Clin Invest*.2003;112(6):831-842)。

[0008] 在另一种方法中,制备“空”MHCII $\alpha/\beta$ 二聚体,然后加载可溶性MHCII结合肽(见例如Novak等人(1999) *J.Clin.Invest*.104:63-67;Nepom等人(2002) *Arthrit.Rheumat*.46:5-12;Macaubus等人(2006) *J.Immunol*.176:5069-5077)。虽然这种方法允许MHCII $\alpha/\beta$ 二聚体上加载肽的更多样性,但重组表达稳定的“空”MHCII $\alpha/\beta$ 二聚体的能力是有限的,因此再次阻碍了大规模pMHCII多聚体文库的制备。例如,已经报道了通过从大肠杆菌包涵体再折叠或通过昆虫细胞或哺乳动物细胞表达产生“空”MHCII $\alpha/\beta$ 二聚体,但产率太低,无法支持高通量方法(综述于Vollers和Stern(2008) *Immunology* 123:305-313)。

[0009] 因此,仍然需要产生肽-MHC文库(包括条码文库)的高效且成本有效的方法,其可用于多种方法,例如,筛选T细胞特异性以分析T细胞识别,例如,在全基因组水平,而不是仅限于选择模型抗原的分析。

## 发明内容

[0010] 本公开提供了生产条码化加载肽的MHC(pMHC)多聚体(例如四聚体)的方法,包括其文库。该方法使用有效的反应条件在短时间内提供pMHC多聚体的高蛋白产率,该反应条件允许多聚体易于肽交换和条码标记,从而允许大型pMHC多聚体文库的高效制备。因此,本文描述的组合物和方法适用于常规实验室研究以及大规模工业和临床应用,在pMHC多聚体有用的所有情况下。在一个实施方式中,pMHC多聚体是pMHC I类(pMHC I)多聚体,其可用于分析CD8+T细胞抗原识别。在另一个实施方式中,pMHC多聚体是pMHC II类(pMHC II)多聚体,其可用于分析CD4+T细胞抗原识别。本发明的MHC多聚体包括MHC单体和多聚结构域之间的共价连接,从而使得多聚结构域上的非共价结合位点易于用于条码标记。

[0011] 在一个方面,本发明提供了一种制备加载肽的MHC(pMHC)多聚体的方法,该多聚体包括两个或多个加载肽的MHC(mMHC)单体,其中每个pMHC单体共价连接到多聚结构域。特别地,pMHC单体通过不是生物素/链霉亲和素或生物素/亲和素相互作用的化学连接与多聚结构域连接,该连接在有效的整体化学反应中实现。通过使用pMHC单体上的偶联部分和多聚结构域实现这种化学连接,然后这些部分反应形成化学连接。然后,肽交换和寡核苷酸条码标记可以很容易地在pMHC多聚体上进行,从而实现高效的大规模pMHC文库生产。

[0012] 因此,在一个方面,本公开提供了一种生产主要组织相容性复合物(MHC)多聚体的方法,该方法包括:

[0013] (a) 提供两个或多个MHC单体,其中每个单体包括偶联部分;

[0014] (b) 提供多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;

[0015] (c) 在足以使MHC单体和多聚结构域共价偶联的条件下,将MHC单体和多聚化结构域组合以产生MHC多聚体。

[0016] 在一个实施方式中,MHC单体是MHC I类单体。在一个实施方式中,MHC单体是MHC II类单体。在某些实施方式中,MHC单体在与多聚结构域结合之前加载有占位符肽



(placeholder peptide)。

[0017] 在一个实施方式中,所述多聚结构域包括非共价结合位点,其中所述方法进一步包括用寡核苷酸条码标记所述MHC多聚体,所述寡核苷酸条码结合所述多聚结构域的非共价结合位点。

[0018] 在一个实施方式中,多聚体是四聚体。在一个实施方式中,多聚结构域是链霉亲和素。在一个实施方式中,多聚结构域是链霉亲和素,寡核苷酸条码结合链霉亲和素的生物素结合位点。

[0019] 关于MHC单体和多聚结构域之间的共价连接,在一个实施方式中,每个MHC单体的偶联部分包括X,而多聚结构域的每个亚基的偶联部分包括Y,其中

[0020] (i) X是末端炔烃,Y是叠氮化物;

[0021] (ii) X是叠氮化物,Y是末端炔烃;

[0022] (iii) X是应变炔烃,Y是叠氮化物;

[0023] (iv) X是叠氮化物,Y是应变炔烃;

[0024] (v) X是二烯,Y是亲二烯;

[0025] (vi) X是亲二烯,Y是二烯;

[0026] (vii) X是硫醇,Y是烯烃;或

[0027] (viii) X是烯烃,Y是硫醇。

[0028] 在一个实施方式中,叠氮化物是铜螯合叠氮化物。在一个实施方式中,铜螯合叠氮化物是吡啶甲基叠氮化物。

[0029] 在一个实施方式中,每个MHC单体的偶联部分和多聚结构域的每个亚基的偶联部分包括分选标签(sortag)基序。

[0030] 在一个实施方式中,每个MHC单体的偶联部分和多聚结构域的每个亚基的偶联部分包括内含蛋白序列。

[0031] 在一个实施方式中,该方法还包括将占位符肽与结合MHC单体的拯救肽表位交换。

[0032] 在另一方面,本发明涉及一种条码标记的MHC多聚体,包括:

[0033] (a) 两个或多个MHC单体;

[0034] (b) 多聚结构域,其包含两个或多个亚基并具有至少一个非共价结合位点;和

[0035] (c) 寡核苷酸条码;

[0036] 其中每个MHC单体通过共价连接与多聚结构域的亚基结合;和

[0037] 其中所述寡核苷酸条码通过与所述多聚结构域上的所述非共价结合位点的非共价键合而与所述多聚结构域结合。

[0038] 在一个实施方式中,MHC多聚体还包括加载到多聚体的每个MHC单体上的MHC结合肽。在一个实施方式中,MHC单体是MHC I类单体。在一个实施方式中,MHC单体是MHC II类单体。

[0039] 在一个实施方式中,MHC多聚体是四聚体。在一个实施方式中,多聚结构域是链霉亲和素。在一个实施方式中,寡核苷酸条码与链霉亲和素上的生物素结合位点非共价结合。

[0040] 在MHC多聚体的一个实施方式中,每个MHC单体包括偶联部分X,并且多聚结构域的每个亚基包括偶联部分Y,其中

[0041] (i) X是末端炔烃,Y是叠氮化物;

- [0042] (ii) X是叠氮化物,Y是末端炔烃;
- [0043] (iii) X是应变炔烃,Y是叠氮化物;
- [0044] (iv) X是叠氮化物,Y是应变炔烃;
- [0045] (v) X是二烯,Y是亲二烯;
- [0046] (vi) X是亲二烯,Y是二烯;
- [0047] (vii) X是硫醇,Y是烯烃;或
- [0048] (viii) X是烯烃,Y是硫醇。
- [0049] 在一个实施方式中,叠氮化物是铜螯合叠氮化物。在一个实施方式中,铜螯合叠氮化物是吡啶甲基叠氮化物。
- [0050] 在MHC多聚体的一个实施方式中,每个MHC单体和多聚结构域的每个亚基包括偶联部分,其中每个偶联部分包括分选标签基序。
- [0051] 在MHC多聚体的一个实施方式中,每个MHC单体和多聚结构域的每个亚基包括偶联部分,其中每个偶联部分包括内含蛋白序列。
- [0052] 在另一方面,本公开涉及制备MHC I类多聚体的方法。在一个实施方式中,提供了一种生产pMHCI多聚体的方法,该方法包括:(a) 提供两个或多个加载有占位符肽的MHCI (p\*MHCI) 单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;(b) 提供多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;(c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将所述p\*MHC I单体和所述多聚结构域组合以产生p\*MHCI多聚体;和(d) 用拯救肽表位替换p\*MHCI多聚体中每个p\*MHC I单体的肽结合沟中结合的占位符肽,以产生pMHCI多聚体。该方法还可以包括通过使多聚体与条码寡核苷酸反应,用寡核苷酸条码标记pMHCI多聚体,所述条码寡核苷酸包含结合pMHCI多聚体的结合部分,例如结合pMHCI多聚体的多聚结构域。
- [0053] 在另一方面,提供了一种生产条码化肽pMHCI多聚体的方法,该方法包括:
- [0054] (a) 提供两个或多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;
- [0055] (b) 提供多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分,并且所述多聚结构域包括至少一个非共价结合位点;
- [0056] (c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将所述p\*MHC I单体和所述多聚结构域组合以产生p\*MHCI多聚体;
- [0057] (d) 用拯救肽表位替换结合在p\*MHCI多聚体中的每个p\*MHC I单体的肽结合沟中的占位符肽,以产生pMHCI多聚体;和
- [0058] (e) 将寡核苷酸条码结合到多聚结构域上的非共价结合位点。
- [0059] 在另一个方面,提供了一种生产pMHCI多聚体的方法,该方法包括:
- [0060] (a) 提供两个或多个加载占位符肽的MHCI (p\*MHCI) 单体,其各自包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 位于(i) 或(ii) C末端处的包含偶联部分的肽接头,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;
- [0061] (b) 提供多聚结构域,其在所述多聚结构域的每个亚基C末端处包含肽接头,所述

肽接头包含偶联部分；

[0062] (c) 在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下，将所述p\*MHCI单体和所述多聚结构域组合以产生p\*MHCI多聚体；和

[0063] (d) 用拯救肽表位替换结合在p\*MHCI多聚体中的每个p\*MHCI单体的肽结合沟中的占位符肽，以产生p\*MHCI多聚体。

[0064] 任何合适的p\*MHC单体均可用于本文所述的方法中。在一个实施方式中，p\*MHC单体来源于脊椎动物。在另一个实施方式中，p\*MHCI单体包括人MHCI重链多肽或其功能片段，以及人 $\beta$ 2-微球蛋白多肽或其功能片段。在另一个实施方式中，其每个p\*MHCI单体是HLA-A、HLA-B或HLA-C。在另一个实施方式中，每个p\*MHCI单体是HLA-A。在另一个实施方式中，每个p\*MHCI单体是可溶的。

[0065] 在另一个实施方式中，每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1结构域。在另一个实施方式中，每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1/ $\alpha$ 2异二聚体。在另一个实施方式中，每个p\*MHCI单体的MHCI重链多肽或其功能片段包含MHCI $\alpha$ 1、 $\alpha$ 2和 $\alpha$ 3结构域。在另一个实施方式中，每个p\*MHCI单体的MHCI重链多肽或其功能片段包含与SEQ ID NO:28-159所示的任何氨基酸序列至少80%、85%、90%、95%或99%相同的 $\alpha$ 结构域。

[0066] 在另一个实施方式中，每个p\*MHCI单体包括 $\beta$ 2-微球蛋白结构域。在一个实施方式中，每个p\*MHCI单体的 $\beta$ 2-微球蛋白多肽是野生型人 $\beta$ 2-微球蛋白。在另一个实施方式中， $\beta$ 2-微球蛋白多肽包含与人 $\beta$ 2-微球蛋白的氨基酸序列(如SEQ ID NO:2或160中所示的氨基酸序列)至少80%、85%、90%、95%或99%相同的氨基酸序列。

[0067] 在另一个实施方式中，每个p\*MHC单体是融合蛋白。例如，在一个实施方式中，每个p\*MHC单体是包含MHCI重链或其功能片段和 $\beta$ 2-微球蛋白或其功能性片段的融合蛋白。在另一个实施方式中，p\*MHC融合蛋白包括MHCI重链或其功能片段与 $\beta$ 2-微球蛋白多肽或其功能性片段之间的肽接头。

[0068] 任何合适的占位符肽均可用于本文所述的方法中。在一个实施方式中，占位符肽是促进MHCI多肽折叠的肽或肽样化合物。在一个实施方式中，占位符肽是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个氨基酸。在另一个实施方式中，占位符肽是2至25个氨基酸。在另一个实施方式中，占位符肽是8至11个氨基酸。在另一个实施方式中，占位符肽是9个氨基酸。在另一个实施方式中，占位符肽是10个氨基酸。在另一个实施方式中，占位符肽包括GILGFVFLJL (SEQ ID NO:7)。在另一个实施方式中，占位符肽由GILGFVFLJL (SEQ ID NO:7) 组成。在其他实施方式中，占位符肽具有SEQ ID NO:8-27或271-279中任一所示的序列。

[0069] 在另一个实施方式中，与交换肽表位相比，占位符肽对MHCI肽结合沟的亲和力更低，并且其中步骤(d)包括在竞争测试中使p\*MHCI单体与过量肽表位接触。在另一个实施方式中，占位符肽的KD比交换肽表位低约4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍或15倍。在另一个实施方式中，占位符肽的KD比交换肽表位低约10倍。在另一个实施方式中，占位符肽比交换肽表位对MHCI结合沟具有更高的亲和力。

[0070] 在另一个实施方式中，占位符肽在约30-37°C的温度下是不稳定的，且步骤(d)包括在肽表位存在下将p\*MHCI单体暴露于约30-37°C(例如，30°C、31°C、32°C、33°C、34°C、35

℃、36℃、或37℃)的温度下。在另一个实施方式中,占位符肽在约pH 2.0-5.5(例如,pH 2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、5.0、5.1、5.2、5.3、5.4或5.5)的酸性pH下不稳定。在另一个实施方式中,在肽表位存在下,p\*MHCI单体暴露于约pH 2.0-5.5(例如,pH 2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、5.0、5.1、5.2、5.3、5.4或5.5)的pH。在另一个实施方式中,占位符肽在约pH 2.0-5.5(例如,pH 2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、5.0、5.1、5.2、5.3、5.4或5.5)的酸性pH下不稳定,且步骤(d)包括在肽表位存在下将p\*MHCI单体暴露于约pH 2.0-5.5(例如,pH 2.0、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3.0、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4.0、4.1、4.2、4.3、4.4、4.5、4.6、4.7、4.8、4.9、5.0、5.1、5.2、5.3、5.4或5.5)的pH。

[0071] 在另一个实施方式中,占位符肽在约pH 9-11(例如,9、9.1、9.2、9.3、9.4、9.5、9.6、9.7、9.8、9.9、10、10.1、10.2、10.3、10.4、10.5、10.6、10.7、10.8、10.9或11)的碱性pH下不稳定。在另一实施方式中,所述占位符肽在约pH 9-11(例如,9、9.1、9.2、9.3、9.4、9.5、9.6、9.7、9.8、9.9、10、10.1、10.2、10.3、10.4、10.5、10.6、10.7、10.8、10.9或11)之间的碱性pH下是不稳定的,并且所述p\*MHCI单体在肽表位存在下暴露于约pH 9-11(例如9、9.1、9.2、9.2、9.3、9.4、9.5、9.6、9.7、9.8、10、10.1、10.2、10.3、10.4、10.5、10.6、10.7、10.8、10.9或10)之间的pH。在另一实施方式中,所述占位符肽在约pH 9-11(例如,9、9.1、9.2、9.3、9.4、9.5、9.6、9.7、9.8、9.9、10、10.1、10.2、10.3、10.4、10.5、10.6、10.7、10.8、10.9或11)之间的碱性pH下是不稳定的,并且步骤(d)包括使p\*MHCI单体在肽表位存在下暴露于约pH 9-11(例如9、9.1、9.2、9.2、9.3、9.4、9.5、9.6、9.7、9.8、10、10.1、10.2、10.3、10.4、10.5、10.6、10.7、10.8、10.9或10)之间的pH。

[0072] 在一些实施方式中,占位符肽包括GILGFVFL(SAQ ID NO:7)。在一些实施方式中,占位符肽由GILGFVFL(SAQ ID NO:7)组成。在其他实施方式中,占位符肽具有SAQ ID NO:8-27或271-279中任一所示的序列。

[0073] 在一个实施方式中,占位符肽包括可切割部分。在一个实施方式中,该方法包括在足以切割占位符肽的条件下使p\*MHCI单体与肽表位接触。可使用任何合适的可切割部分。在一个实施方式中,可切割部分是可光切割氨基酸,并且该方法(例如,步骤(d))包括在足以诱导占位符肽中的可光切割部分切割和肽表位与MHCI单体结合的条件下将p\*MHCI单体暴露于UV光。在一个实施方式中,可光切割氨基酸包括(2-硝基)苯基侧链。在另一个实施方式中,可光切割氨基酸包括3-氨基-3-(2-硝基苯基)丙酸。在另一个实施方式中,可光切割氨基酸是(2-硝基)苯基甘氨酸。

[0074] 在其它实施方式中,可光切割的占位符肽和与其结合的对应该MHC分子选自:A\*02:01,KILGFVFL(SAQ ID NO:15)或GILGFVFL(SAQ ID NO:7),A\*01:01,STAPGJLEY(SAQ ID NO:16);A\*02:03,SVRDJLRL(SAQ ID NO:271);A\*02:06,LTAJFLIFL(SAQ ID NO:272);A\*02:07,LLDSDJERL(SAQ ID NO:273);A\*02:11,KMDIJVPLL(SAQ ID NO:274);A\*03:01,RIYRJGATR(SAQ ID NO:17);A\*11:01,RVFAJSFIK(SAQ ID NO:18);A\*24:02,VYGJVRACL(SAQ ID NO:11);A\*33:03,FYVJGAANR(SAQ ID NO:275);B\*07:02,AARGJTLAM(SAQ ID NO:14);B\*

15:02, ILGPPGJVY (SEQ ID NO:276); B\*35:01, KPIVVVLJGY (SEQ ID NO:19); B\*44:05, EEFGAAJSF (SEQ ID NO:277); B\*46:01, KMKEIAJAY (SEQ ID NO:278); B\*55:02, KPWDJIPMV (SEQ ID NO:279); C\*03:04, FVYGJSKTSL (SEQ ID NO:20), B\*08:01, FLRGRAJGL (SEQ ID NO:21); C\*07:02, VRIJHLYIL (SEQ ID NO:22); C\*04:01, QYDJAVYKL (SEQ ID NO:23); B\*15:01, ILGPJGSVY (SEQ ID NO:24); B\*40:01, TEADVQJWL (SEQ ID NO:25); B\*58:01, ISARGQJLF (SEQ ID NO:26); 和 C\*08:01, KAAJDLSHFL (SEQ ID NO:27), 其中 J 是 3-氨基-3-(2-硝基苯基)丙酸。

[0075] 在另一个实施方式中,可切割部分是包含化学选择性部分的氨基酸。在另一个实施方式中,该方法(例如,步骤(d))包括在足以切割化学选择性部分的条件下使 p\*MHCI 单体与肽表位接触。在另一个实施方式中,化学选择性部分是连二亚硫酸钠敏感的偶氮苯接头。在另一个实施方式中,该方法(例如,步骤(d))包括在连二亚硫酸钠存在下使 p\*MHCI 单体与肽表位接触。

[0076] 在另一个实施方式中,可切割部分是高碘酸盐敏感氨基酸。在另一个实施方式中,该方法(例如,步骤(d))包括在高碘酸盐存在下,在足以切割占位符肽的条件下,使 p\*MHCI 单体与肽表位接触。在另一个实施方式中,高碘酸盐敏感氨基酸包括邻二醇部分。在另一个实施方式中,高碘酸盐敏感氨基酸包括邻氨基醇部分。在另一个实施方式中,高碘酸盐敏感氨基酸是  $\alpha, \gamma$ -二氨基- $\beta$ -羟基丁酸(DAHB)。

[0077] 在另一个实施方式中,可切割部分是蛋白酶识别部分。在一个实施方式中,蛋白酶是氨基肽酶。在另一个实施方式中,蛋白酶是甲硫氨酸氨基肽酶。在另一些实施方式中,蛋白酶选自 FXa、凝血酶、TEV、HRV3C 和弗林蛋白酶。

[0078] 在一个实施方式中,占位符肽是二肽。在另一个实施方式中,二肽结合到 MHC I 结合沟的 F 口袋。在另一个实施方式中,二肽的第二氨基酸是疏水性的。在另一个实施方式中,二肽选自甘氨酸-亮氨酸(GL)、甘氨酸-缬氨酸(GV)、甘氨酸-甲硫氨酸(GM)、甘氨酸-环己基丙氨酸(GCha)、甘氨酸-高亮氨酸(GH1e)和甘氨酸-苯丙氨酸(GF)。

[0079] 可以使用任何合适的多聚结构域。在一个实施方式中,多聚结构域的每个亚基包括偶联部分。在另一个实施方式中,多聚结构域在多聚结构域的每个亚基的 N 末端处包含肽接头,该肽接头包含偶联部分。在另一个实施方式中,多聚结构域在多聚结构域的每个亚基的 C 末端处包含肽接头,该肽接头包含偶联部分。在一个实施方式中,多聚结构域是二聚体、四聚体、六聚体、八聚体、十聚体或十二聚体。在另一个实施方式中,多聚结构域是同多聚体。在另一个实施方式中,多聚结构域是异多聚体。在另一个实施方式中,多聚结构域包括链霉亲和素或其衍生物。在另一个实施方式中,多聚结构域是链霉亲和素或其衍生物的四聚体。在另一个实施方式中,多聚结构域包括 Strep-tag® 或 Strep-tactin®。

[0080] 在一个实施方式中,偶联部分连接到每个 p\*MHCI 单体的 MHC I 重链  $\alpha 1$  结构域的 C 末端。在另一个实施方式中,多聚结构域与 MHC I  $\alpha 1$  结构域的 C 末端共价偶联。在另一个实施方式中,偶联部分连接到每个 p\*MHCI 单体的 MHC I 重链  $\alpha 2$  结构域的 C 末端。在另一个实施方式中,多聚结构域与 MHC I  $\alpha 2$  结构域的 C 末端共价偶联。在另一个实施方式中,偶联部分连接到每个 p\*MHCI 单体的 MHC I 重链  $\alpha 3$  结构域的 C 末端。在另一个实施方式中,多聚结构域与 MHC I  $\alpha 3$  结构域的 C 末端共价偶联。在另一个实施方式中,偶联部分连接到每个 p\*MHCI 单体的每个 p\*MHC 单体的  $\beta 2$ -微球蛋白的 C 末端。在另一个实施方式中,多聚结构域共价偶联到每个 p\*MHC

单体的 $\beta 2$ -微球蛋白的C末端。在另一个实施方式中,每个p\*MHCI单体与多聚结构域的共价偶联由半胱氨酸转肽酶(例如分选酶或其酶活性片段)介导。

[0081] 在另一个实施方式中,两个或多个p\*MHC单体与多聚结构域化学偶联。在另一个实施方式中,化学偶联由半胱氨酸生物偶联介导。在另一个实施方式中,化学偶联由天然化学偶联介导。在另一个实施方式中,化学偶联由点击化学介导。

[0082] 在另一个实施方式中,每个p\*MHCI单体的偶联部分包括X,并且多聚结构域的每个亚基的偶联部分包括Y。例如,在一个实施方式中,X是末端炔烃,Y是叠氮化物。在另一个实施方式中,X是叠氮化物,Y是末端炔烃。在另一个实施方式中,X是应变炔烃,Y是叠氮化物。在另一个实施方式中,X是叠氮化物,Y是应变炔烃。在某些实施方式中,叠氮化物是铜整合叠氮化物。在一个实施方式中,铜整合叠氮化物是吡啶甲基叠氮化物。在另一个实施方式中,X是二烯,Y是亲二烯。在另一个实施方式中,X是二烯,Y是亲二烯。在另一个实施方式中,X是硫醇,Y是烯烃。在另一个实施方式中,X是烯烃,Y是硫醇。

[0083] 在另一个实施方式中,每个p\*MHC结构域的偶联部分包括连接到C末端的肽接头,并且多聚化结构域的每个亚基的偶联部包括连接到多聚结构域的每个亚基的C末端的肽接头。在另一个实施方式中,每个p\*MHC单体的C末端的肽接头包括 $(G)_n-X$ ,其中n至少为2,X是适合化学偶联的部分,并且多聚结构域的每个亚基的C末端的肽接头包含 $(G)_n-Y$ ,其中n为至少2,Y是适合于与每个p\*MHC单体的X部分化学偶联的部分。

[0084] 在另一个实施方式中,每个p\*MHCI单体的偶联部分包括C末端分选标签(sortag),并且多聚结构域的每个亚基的偶联部分包含N末端分选标签。在另一个实施方式中,每个p\*MHCI单体的偶联部分包括N末端分选标签(sortag),并且多聚结构域的每个亚基的偶联部分包含C末端分选标签。在另一个实施方式中,该方法(例如,上述制备MHC多聚体的方法的步骤(c))包括向p\*MHC I单体和多聚结构域的混合物中添加分选酶,并催化每个p\*MHC单体和多聚结构域之间肽键的形成,以产生p\*MHC-多聚体。

[0085] 在另一个实施方式中,两个或多个p\*MHCI单体(例如,在上述制备MHC多聚体的方法的步骤(a)中)是通过在包含适合于化学偶联的部分的肽接头存在下,使包含C末端分选标签的p\*MHC I单体与分选酶或其酶活性片段接触而产生的,其中分选酶或其酶活性片段介导肽接头与p\*MHC单体的偶联;多聚结构域(例如,在上述方法中的步骤(b)中)是通过在包含适合于化学偶联的部分的肽接头的存在下使包含N末端分选标签的多聚结构域与分选酶接触而产生的,其中所述分选酶或其酶活性片段介导肽接头与多聚结构域的每个亚基的N末端的偶联;和步骤(c)包括在两个或多个p\*MHC单体的C末端的肽接头和多聚结构域的每个亚基的N末端的肽接头之间化学偶联,以产生p\*MHC-多聚体。

[0086] 在一个实施方式中,每个p\*MHC单体的C末端的肽接头包括 $(G)_n-X$ ,其中n至少为2,X是适合点击化学偶联的部分,并且多聚结构域的每个亚基的N末端的肽接头包含 $Y-(G)_n$ ,其中n为至少2,Y是适合于与每个p\*MHC单体的X部分点击化学偶联的部分。在另一个实施方式中,X是末端炔,Y是叠氮化物。在另一个实施方式中,X是叠氮化物,Y是末端炔烃。在另一个实施方式中,X是应变炔烃,Y是叠氮化物。在另一个实施方式中,X是叠氮化物,Y是应变炔烃。在某些实施方式中,叠氮化物是铜整合叠氮化物。在一个实施方式中,铜整合叠氮化物是吡啶甲基叠氮化物。在另一个实施方式中,X是二烯,Y是亲二烯。在另一个实施方式中,X是二烯,Y是亲二烯。在另一个实施方式中,X是硫醇,Y是烯烃。在另一个实施方式中,X是烯

烃,Y是硫醇。

[0087] 在一个实施方式中,分选酶或其酶活性片段是Ca<sup>2+</sup>依赖性的。在另一个实施方式中,分选酶或其酶活性片段是Ca<sup>2+</sup>非依赖性的。在另一个实施方式中,分选酶或其酶活性片段是野生型分选酶的可溶性片段。在另一个实施方式中,分选酶或其酶活性片段是金黄色葡萄球菌(*S.aureus*)分选酶A的变体或同源物。在另一个实施方式中,分选酶或其酶活性片段是修饰的分选酶A。在另一个实施方式中,分选酶或其酶活性片段是SrtAstaph突变体。在另一个实施方式中,SrtStaph突变体选自F40、SrtAstaph五突变体、2A-9和4S-9。

[0088] 在一个实施方式中,每个p\*MHCI单体与多聚结构域的共价偶联是由内含蛋白介导的。在一个实施方式中,所述内含蛋白选自MxeGyrA、SspDnaE、NpuDnaE、AvaDnaE、Cfa(共有DnaE裂解内含蛋白)、gp41-1、gp41-8和NrdJ-1。在另一个实施方式中,内含蛋白是裂解内含蛋白对。在另一个实施方式中,每个p\*MHCI单体通过内含蛋白肽标签与多聚结构域偶联。在另一个实施方式中,每个p\*MHCI单体在C末端包括N-内含蛋白片段,并且多聚结构域的每个亚基在N末端包括Npu-C-内含蛋白片段。

[0089] 在一个实施方式中,拯救肽表位包括标识符。在一个实施方式中,标识符是核酸标识符。在一个实施方式中,标识符是核酸标识符。在另一个实施方式中,核酸标识符编码肽。在另一个实施方式中,核酸标识符的长度为25个核苷酸至500个核苷酸。在另一个实施方式中,核酸标识符的长度为80个核苷酸至120个核苷酸。

[0090] 在一个方面,本发明提供了一种生产多种pMHCI多聚体文库的方法及其生产方法,包括高通量方法。在一些实施方式中,pMHC多聚体进一步包含核酸标识符,允许如本文他处所述方便地检测和量化结合。

[0091] 在一个实施方式中,提供了一种产生包含加载多种肽表位的MHCI(pMHCI)多聚体的文库的方法,该方法包括:

[0092] (a) 提供多个加载占位符肽的MHCI(p\*MHCI)单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

[0093] (b) 提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;

[0094] (c) 在足以使所述两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,将所述p\*MHCI单体和所述多聚结构域组合以产生p\*MHCI多聚体;和

[0095] (d) 用包含多个独特MHCI肽表位的肽文库替换多个p\*MHCI多聚体中的占位符肽,以产生多个加载肽的MHCI(pMHCI)多聚体。

[0096] 在另一方面,提供了一种产生包含多种条码化且加载肽的主要组织相容性复合物I类(pMHCI)多聚体的文库的方法,该方法包括:

[0097] (a) 提供多个加载占位符肽的MHCI(p\*MHCI)单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;

[0098] (b) 提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分,并且所述多聚结构域包括至少一个非共价结合位点;

[0099] (c) 在足以使所述两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,将所述多个p\*MHCI单体和多个多聚结构域组合以产生多个p\*MHCI多聚体;

[0100] (d) 用多个独特拯救肽表位替换结合在p\*MHCI多聚体的肽结合沟中的占位符肽,以产生多个pMHC I多聚体;和

[0101] (e) 将寡核苷酸条码结合到多聚结构域上的非共价结合位点。

[0102] 在另一方面,提供了一种产生包含多种条码化且加载肽的主要组织相容性复合物 I类 (pMHC I) 多聚体的文库的方法,该方法包括:

[0103] (a) 提供多个加载占位符肽的MHC I (p\*MHC I) 单体,其各自包含 (i) MHC I重链多肽或其功能片段, (ii)  $\beta$ 2-微球蛋白多肽或其功能片段, (iii) 位于 (i) 或 (ii) C末端处的包含偶联部分的肽接头,和 (iv) 结合在每个MHC I单体的肽结合沟中的占位符肽;

[0104] (b) 提供多个多聚结构域,其在多聚结构域的每个亚基C末端处包含肽接头,所述肽接头包含偶联部分;

[0105] (c) 在足以使所述两个或多个p\*MHC I单体与多聚结构域共价偶联的条件下,将所述多个p\*MHC I单体和多个多聚结构域组合以产生多个p\*MHC I多聚体;和

[0106] (d) 用多个独特拯救肽表位替换结合在p\*MHC I多聚体的肽结合沟中的占位符肽,以产生多个pMHC I多聚体;和

[0107] (e) 将寡核苷酸条码结合到多聚结构域上的非共价结合位点。

[0108] 在另一方面,本公开提供了一种加载肽的MHC I类 (pMHC I) 多聚体文库,其中文库中的每个pMHC I多聚体包括两个或多个加载独特肽表位的pMHC I单体,并且其中每个pMHC I单体共价连接到多聚结构域的亚基。

[0109] 在一个实施方式中,MHC I肽表位文库是高多样性肽文库。在另一个实施方式中,肽文库包含约 $10^3$ 至 $10^{20}$ 个不同的MHC I肽表位。在另一个实施方式中,肽文库包含约 $10^3$ 、约 $10^4$ 、约 $10^5$ 、约 $10^6$ 、约 $10^7$ 、约 $10^8$ 、约 $10^9$ 、约 $10^{10}$ 、约 $10^{11}$ 、约 $10^{12}$ 、约 $10^{13}$ 、约 $10^{14}$ 、约 $10^{15}$ 、约 $10^{16}$ 、约 $10^{17}$ 、约 $10^{18}$ 、约 $10^{19}$ 、约 $10^{20}$ 或更多不同的MHC I肽表位。

[0110] 在一个实施方式中,MHC I肽表位衍生自单个抗原蛋白。在另一个实施方式中,MHC I肽表位包括抗原蛋白的重叠片段。在另一个实施方式中,多个独特的肽表位由生物体的基因组、生物体的转录组、生物体的蛋白质组或生物体的肽或蛋白质产生。在另一个实施方式中,多个独特的肽表位由两个基因组之间的差异序列产生。在另一个实施方式中,MHC I肽表位可以是感兴趣的特定肽表位的改变的肽配体 (APL)。

[0111] 在另一个实施方式中,每个pMHC I多聚体包括独特标识符部分。在一个实施方式中,独特标识符部分是核酸。

[0112] 在另一方面,提供了包含多个加载肽的MHC I (pMHC I) 多聚体的多肽文库,其中每个加载肽的pMHC I多聚体包含两个或多个与多聚结构域偶联的pMHC I单体。

[0113] 在另一方面,提供了一种分离MHC I-多聚体结合淋巴细胞的方法,其中所述方法包括:

[0114] (a) 使多个淋巴细胞与pMHC I多聚体文库接触;和

[0115] (b) 产生多个隔室,其中每个隔室包括与文库的pMHC I多聚体结合的淋巴细胞和捕获载体。在一个实施方式中,淋巴细胞是T细胞、B细胞或NK细胞。

[0116] 在另一个实施方式中,提供了一种鉴定与pMHC I多聚体结合的淋巴细胞的方法,该方法包括:

[0117] (a) 使多个淋巴细胞与pMHC I多聚体文库接触;



[0118] (b) 将与文库的pMHCI多聚体结合的多个淋巴细胞中的一个淋巴细胞划分为单个隔室,其中所述pMHCI多聚体包括独特标识符;和

[0119] (c) 确定与所述隔室化淋巴细胞结合的pMHCI的独特标识符。

[0120] 为了更好地理解本发明的本质和优势,参考以下与附图结合的详细说明。本发明能够在不偏离本发明的情况下进行各个方面的修改。因此,这些实施方式的附图和描述不是限制性的。

[0121] 附图简要说明

[0122] 所附权利要求书中具体说明了本发明的新特征。可参考以下详述更好地理解本发明的特征和优点,这些详述阐述了利用本发明原理的说明性实施方式和附图,其中:

[0123] 图1例举了各种点击化学处理和反应。

[0124] 图2说明了各种肽交换方法。

[0125] 图3A-3E显示了偶联反应的SDS-PAGE或Western印迹分析。草图描绘了与一个、两个、三个或四个HLA分子相连的SAv四聚体。箭头表示不需要的副产品。图3A:SAv-偶联反应的抗His-Western印迹分析。表中显示了每条泳道的描述。基于与参考SA蛋白的比较,反应程度约为94-97%。图3B:HLA-A2-DBCO-SA<sub>v</sub>-Az的SDS-PAGE图像。泳道1:SeeBlue Plus蛋白质标准品,泳道2:SA-Az(未煮沸),泳道3:SA-Az(煮沸),泳道4:HLA-A2-DBCO-SA<sub>v</sub>-Az(未煮沸,未还原),泳道5:HLA-A2-DBCO-SA<sub>v</sub>-Az(煮沸,还原)。图3C:HLA-A2-Az-SA<sub>v</sub>-DBCO的SDS-PAGE图像。泳道1:SeeBlue Plus蛋白质标准品,泳道2:HLA-A2-Az(未煮沸),泳道3:HLA-A2-Az-SA<sub>v</sub>-DBCO,(未煮沸),泳道4-7:HLA-A2-Az-SA<sub>v</sub>-DBCO反应(未煮沸)。图3D:HLA-A2-A1k-SA<sub>v</sub>-Az的SDS-PAGE图像。泳道1:SeeBlue Plus蛋白质标准品,泳道3:HLA-A2-A1k-SA<sub>v</sub>-Az(未煮沸,未还原),泳道5:HLA-A2-烯炔-SA<sub>v</sub>-Az(煮沸,还原)。图3E:偶联四聚体形式的HLA-A\*01:01、HLA-A\*03:01和HLA-A\*24:02的SDS-PAGE图像。样品为未煮沸/未还原(NB/NR)或煮沸/还原(煮沸/R)。

[0126] 图4.HLA-A2-N-内含蛋白/β2m/肽复合物与SA<sub>v</sub>-C-内含蛋白之间的内含蛋白剪接反应的SDS-PAGE分析。

[0127] 图5A和5B示出了通过差示扫描荧光法监测的UV交换。图5A显示了实施例1中产生的含有占位符GILGFVFL肽(SEQ ID NO:7)的HLA-A\*02:01-A1k-SA<sub>v</sub>-Az偶联四聚体,或在过量NLVPMVATV肽(SEQ ID NO:8)存在下进行UV交换后的差示扫描荧光分析(DSF),显示了20℃的稳定性增加,表明交换成更高亲和力的肽。图5B是HLA-A\*02生物素介导的四聚体的DSF,其通过在单体上进行UV交换,然后进行四聚,或通过四聚体本身上进行UV交换而产生,并证实了多聚体状态对UV交换的效率没有影响,并且本发明的多聚体具有与工业标准pMHC相同的稳定性。

[0128] 图6A-6F描绘了在生物素化HLA-A\*02单体和四聚体上进行肽交换后的流式细胞分析。用NLVPMVGTV肽(SEQ ID NO:9)扩增的供体PBMC用以下染色:仅抗CD8-BV785和抗-Flag-APC二抗(图6A),加载占位符肽GILGFVFL(SEQ ID NO:7)的50nM HLA-A\*02生物素介导的四聚体(图6B),与NLVPMVATV肽(SEQ ID NO:8)重折叠的50nM HLA-A\*02生物素介导的四聚体(图6C),通过UV交换在单体形式上加载NLVPMVATV肽(SEQ ID NO:8),然后用链霉亲和素四聚化的50nM HLA-A\*02生物素介导的四聚体(图6D),通过UV交换在四聚体形式本身上加载NLVPMVATV肽(SEQ ID NO:8)的50nM HLA-A\*02生物素介导的四聚体(图6E)和通过二肽交换

在四聚体形式本身上加载NLVPMVATV肽 (SEQ ID NO:8) 的50nM HLA-A\*02生物素介导的四聚体 (图6F)。

[0129] 图7A-7B描绘了在HLA-A\*02:01-A1k-SAv-Az偶联四聚体上进行UV交换后的流式细胞术分析。用NLVPMVATV肽 (SEQ ID NO:8) 扩增的供体PBMC用以下染色: 仅抗-链霉亲和素-PE和抗-Flag-APC二抗 (图7A) 或通过UV交换直接在四聚体形式上加载NLVPMVATV肽 (SEQ ID NO:8) 的1nM HLA-A\*02:01-A1k-SAv-Az偶联四聚体 (图7B)。

[0130] 图8A-8C描述了ELISA和DSF的比较, 作为UV交换的HLA-a\*02四聚体的稳定性测试。具体地, 图8A描述了与代表NLVPMVATV肽 (SEQ ID NO:8) 的改变的肽配体 (APL) 的192成员肽组UV交换的HLA-A\*02:01-A1k-SAv-Az偶联四聚体的ELISA分析。ELISA OD对每种肽的净MHC预测IC<sub>50</sub>作图。不同的肽跨越一系列ELISA信号。图8B显示了NLVPMVATV (SEQ ID NO:8) APL肽子集UV交换入生物素介导的四聚体的DSF曲线, 显示了一系列稳定性。图8C显示了NLVPMVATV (SEQ ID NO:8) APL肽子集UV交换入生物素介导的四聚体的DSF/ELISA相关性。

[0131] 图9A-9D描述了HLA-A\*01:01-A1k-SAv-Az偶联四聚体的质量控制分析。具体而言, 图9A描绘了具有低聚集性的HLA-A\*01:01四聚体的分析SEC色谱图。图9B描绘了未煮沸/未还原 (NB/NR) 或煮沸/还原 (煮沸/R) 的HLA-A\*01:01-A1k-SAv-Az偶联四聚体的SDS-PAGE。图9C描绘了加载占位符肽STAPGJLEY (SEQ ID NO:16) (无UV) 的HLA-A\*01:01-A1k-SAv-Az偶联四聚体的DSF, 或在不存在 (UV无肽) 或存在 (UV+VTEHDTLLY (SEQ ID NO:10)) 拯救肽的情况下进行UV交换后的DSF。图9D描述了用VTEHDTLLY肽 (SEQ ID NO:10) 扩增的PBMC的流式细胞术数据, 并用以下染色: 20nM通过重折叠加载VTEHDTLLY肽 (SEQ ID NO:10) 的HLA-A\*01:01生物素介导的四聚体 (重折叠VTE), 加载STAPGJLEY (SEQ ID NO:16) 的HLA-A\*01:01-A1k-SAv-Az偶联四聚体 (无UV), 或在拯救肽VTEHDTLLY (SEQ ID NO:10) 存在下进行UV交换后的HLA-A\*01:01-A1k-SAv-Az偶联四聚体 (UV+VTE)。描述了四聚体阳性细胞的分數 (%四聚体+) 和平均荧光强度 (MFI)。

[0132] 图10A-10D描述了HLA-A\*24:02-A1k-SAv-Az偶联四聚体的质量控制分析。具体而言, 图10A描绘了具有低聚集性的HLA-A\*24:02四聚体的分析SEC色谱图。图10B描绘了未煮沸/未还原 (NB/NR) 或煮沸/还原 (煮沸/R) 的HLA-A\*24:02-A1k-SAv-Az偶联四聚体的SDS-PAGE。图10C描绘了加载占位符肽VYGJVRACL (SEQ ID NO:11) (无UV) 的HLA-A\*24:02-A1k-SAv-Az偶联四聚体的DSF, 或在不存在 (UV无肽) 或存在 (UV+QYDPVAALF (SEQ ID NO:12)) 拯救肽的情况下进行UV交换后的DSF。图10D描述了用QYDPVAALF肽 (SEQ ID NO:12) 扩增的PBMC的流式细胞术数据, 并通过以下染色: 仅二抗, 20nM通过重折叠加载QYDPVAALF肽 (SEQ ID NO:12) 的HLA-A\*24:02生物素介导的四聚体 (重折叠QYD), 加载VYGJVRACL (SEQ ID NO:11) 的20nM HLA-A\*24:02-A1k-SAv-Az偶联四聚体 (无UV), 或在拯救肽QYDPVAALF (SEQ ID NO:12) 存在下进行UV交换后的20nM HLA-A\*24:02-A1k-SAv-Az偶联四聚体 (UV+QYD)。描述了四聚体阳性细胞的分數 (%四聚体+) 和平均荧光强度 (MFI)。

[0133] 图11A-11C描述了HLA-B\*07:02-A1k-SAv-Az偶联四聚体的质量控制分析。具体而言, 图11A描绘了具有低聚集性的HLA-B\*07:02四聚体的分析SEC色谱图。图11B描绘了未煮沸/未还原 (NB/NR) 的HLA-B\*07:02-A1k-SAv-Az偶联四聚体的SDS-PAGE。图11C描述了用RPHERNGFTVL肽 (SEQ ID NO:13) 扩增的PBMC的流式细胞术数据, 并通过以下染色: 仅二抗, 20nM通过重折叠加载RPHERNGFTVL肽 (SEQ ID NO:13) 的HLA-B\*07:02生物素介导的四聚体

(重折叠RPH),加载AARGJTLAM(SEQ ID NO:14)的20nM HLA-B\*07:02-A1k-SAv-Az偶联四聚体(无UV),或在拯救肽RPHENGFTVL(SEQ ID NO:13)存在下进行UV交换后的20nM HLA-B\*07:02-A1k-SAv-Az偶联四聚体(UV+RPH)。描述了四聚体阳性细胞的分數(%四聚体+)和平均荧光强度(MFI)。

[0134] 图12描述了用鉴定性寡核苷酸标签标记HLA-A\*02:01-A1k-SAv-Az偶联四聚体。将如实施例1所述制备的HLA-A\*02:01-A1k-SAv-Az偶联四聚体与5'生物素化寡核苷酸孵育,并用Western分离,用抗Flag抗体检测。寡聚物加成后的漂移条带表示四聚体标记。

[0135] 图13显示了条码化HLA-A\*02:01-A1k-SAv-Az APL文库的单细胞测序。通过单细胞测序鉴定的pMHC与单个T细胞结合的热图。代表2008个单个细胞的列按TCR克隆型聚类,行代表NLVPMATV(SEQ ID NO:8)的192个APL变体中的每一个。暖色表示通过鉴定性寡核苷酸标签读取的强pMHC-TCR相互作用。

[0136] 图14描述了在单模板条件下将肽编码模板PCR扩增到水凝胶上。在单模板条件下,对水凝胶珠批量或在液滴中包封后进行PCR。PCR后,打破液滴后释放的上清液与通过XbaI或模拟消化从珠中释放的产物并列运行。

[0137] 图15显示了液滴中单模板扩增的验证。用链霉亲和素-PE对在单模板条件下批量或液滴中的模板PCR扩增后的水凝胶进行染色。荧光水凝胶相对于总水凝胶进行定量,以确认单模板条件。

[0138] 图16A-16B描述了将HLA-A\*02:01-A1k-SAv-Az偶联四聚体加载到PCR扩增的水凝胶上。在加载偶联四聚体或随后用Benzonase或SmaI释放后,用抗-Flag-APC或抗-β2M-Alexa488染色的水凝胶的信噪比(图16A)。ELISA测定了在水凝胶加载步骤后或通过Benzonase或SmaI从加载的水凝胶中释放后留在上清液中的HLA-A\*02:01-A1k-SAv-Az偶联四聚体的浓度(图16B)。

[0139] 图17A-17B描绘了IVTT肽生产,以产生功能性UV交换的四聚体。用抗-SUMO结构域抗体检测Western:由编码SUMO-NLVPMVATV(SEQ ID NO:8)肽融合的PCR扩增子模板驱动的IVTT反应产物(+/-Ulp1蛋白酶)在泳道10-11中运行(图17A)。泳道2-9包含含SUMO结构域的标准品的稀释系列,该标准品用于将SUMO域的产率定量至~1μM(图17A)。由IVTT产生的肽通过UV交换产生的四聚体的流式分析(图17B)。四聚体在等摩尔合成NLVPMVATV(SEQ ID NO:8)肽(UV ex 1:1NLV-合成)或由SUMO-NLVPMVATV(SEQ ID NO:8)肽模板驱动的IVTT反应(+Ulp1)(UV ex NLV-IVTT)存在下进行UV交换,并在1nM下在NLVPMVATV(SEQ ID NO:9)扩增的PBMC上染色(图17B)。用NLVPMVATV(SEQ ID NO:8)或GILGFVFL(SEQ ID NO:7)肽重折叠的阳性和阴性对照四聚体也在1nM下染色,如图17B所示。

[0140] 图18显示了利用滴入法从水凝胶中产生和释放的pMHC四聚体的流式细胞术结果,其在抗原特异性CD8+T细胞上染色。

[0141] 图19是显示使用可交换可条码化四聚体的高通量条码化抗原文库生产的示意图。

[0142] 图20是一示意图,显示了分选标签(sortag)和点击化学用于p\*MHCII与SAv的偶联、占位符肽内肽接头的切割、占位符肽与拯救肽的交换以及与TCR的结合。

[0143] 图图21A-21E描述了p\*MHCII多聚体的产生。图21A:GGG-炔烃与单体p\*MHCIIα链偶联的抗-Myc Western印迹分析。图21B:p\*MHCII-A1k和SAv-Az点击反应后的SDS-PAGE分析。图21C:点击反应样品的HiLoad 26/600Superdex200SEC洗脱色谱图。图21D:从SEC获得的主

峰的抗-FLAG Western印迹分析。泳道1:Chameleon Duo预-染色蛋白梯度(Licor),泳道2:将样品加载至SEC柱之前的点击反应,泳道3和4:峰I的SEC样品,泳道5和6:峰II的SEC样品,泳道7:游离Sav。泳道编号对应于未煮沸的样品,而标有星号的泳道编号则对应于煮沸样品。图21E:从SEC获得的主峰的抗-His Western印迹分析。泳道编号与图21D所述相同。

[0144] 图22A-22C示出了pMHCII的消化、交换和TCR结合。图22A:因子Xa切割前后煮沸和未煮沸样品的SDS-PAGE分析。图22B:一种检测生物素化交换肽对pMHCII多聚体的结合能力的ELISA分析。图22C:BLI测定,测量HA特异性TCR与pMHCII多聚体之间的相互作用,其被交换以显示同源HA肽。黑色、浅灰色和深灰色曲线分别对应于将加载TCR的生物传感器移动到含有交换的pMHCII、未交换pMHCII和BLI缓冲液的孔中获得的信号。虚线定义了将生物传感器转移到没有分析物(离解)的孔。

[0145] 图23A-B示出了抗原特异性T细胞上pMHCII四聚体文库的染色。图23A:表达DRB1\*01:01等位基因并用流感血凝素表位PKYVKQNTLKLAT (SEQ ID NO:281)扩增的供体CD4+PBMC用具有抗-链霉亲和素-PE和抗CD4-BV510二抗的10成员DRB1\*01:01文库染色。将四聚体阳性门中的细胞分选,与用加载在A\*02:01四聚体上的ELAGIGILTV (SEQ ID NO:282)变体的6成员文库染色的MART1-抗原特异性CD8+T细胞混合,并对所得合并物进行单细胞测序,得到图23B所示的热图。分布在y轴上的最普遍的TCR克隆型特异性结合加载HA肽的DRB1\*01:01四聚体。

## 具体实施方式

### [0146] 定义

[0147] 除非另外定义,本文使用的所有技术和科学术语旨在具有本发明所属领域技术人员通常所理解的含义。本文所用技术意指本领域中通常理解的技术,包括本领域技术人员显而易见的那些技术的变化或等效技术的替代。虽然以下术语据信是本领域普通技术人员所熟知的,仍给予以下定义以助理解在此公开的主题。

[0148] 如本文所用,“约”将为本领域普通技术人员所理解,并且在某种程度上取决于其使用的上下文而变化。如果在使用该术语的文中该术语对于本领域普通技术人员来说是不明确的,则“约”将意味着高达特定值的正负10%。

[0149] 如本文所用,“改变的肽配体”或“APL”是指肽配体如MHC结合肽的改变或突变形式。与衍生出改变或突变的肽配体相比,肽配体经改变或突变的形式包含至少一种结构修饰(例如氨基酸取代)。例如,可以通过对已知MHC结合肽的系统性或随机突变来制备APL组,从而产生APL文库,该APL文库可用作MHC结合多肽的文库,用于加载到本文所述的MHC偶联多聚体上。

[0150] 本文中使用的术语“和/或”当用在多个事项的列表的情形下,表示单独或以任何可能组合或亚组合存在的事项。

[0151] 术语“抗原决定簇”或“表位”是T细胞受体、MHC分子或抗体的可变结构域与其特异性结合的抗原上的位点。表位既可以由连续氨基酸形成,也可以由蛋白质三级折叠并列的非连续氨基酸形成。由连续氨基酸形成的表位通常在暴露于变性溶剂时保留,而由三级折叠形成的表位通常在用变性溶剂处理时丢失。表位通常包括具有独特空间构型的至少3、4、5、6、7、8、9、10、11、12、13、14或15个氨基酸。用于确定哪些表位与给定TCR或抗体结合的方

法(即表位作图)在本领域是公知的,并且包括例如免疫印迹和免疫沉淀测定,其中测试来自抗原的重叠或连续肽与给定的TCR或免疫球蛋白的反应性。确定表位空间构象的方法包括本领域和本文所述的技术,例如,x射线晶体学核磁共振、低温电子显微镜(低温EM)、氘氘交换质谱(HDX-MS)和定点诱变(参见,例如,Epitope Mapping Protocols in Methods in Molecular Biology,卷66,G.E.Morris编(1996))。

[0152] 本文中使用的术语“亲合力”是指作为多价分子(例如,可溶性多聚pMHC-免疫球蛋白蛋白)的多个结合位点与靶分子的协同相互作用的函数的结合强度。存在许多技术来表征分子相互作用的亲合力,包括switchSENSE和表面等离子体共振(Gjelstrup等人, J. Immunol. 188:1292-1306, 2012); Vorup-Jensen, Adv. Drug. Deliv. Rev. 64:1759-1781, 2012)。

[0153] 如本文所用,“条码”也称寡核苷酸条码,是短核苷酸序列(例如,约1、2、3、4、5、6、7、8、9、10、11或12个核苷酸长),其鉴别与其偶联的分子。条码可用于例如识别反应混合物中的分子。条码独特识别与之偶联的分子,例如,通过使用每个都包含“独特分子标识符”条码的引物进行逆转录。在其他实施方式中,可以使用包含每个分子独特“分子条码”的引物。用条码标记分子的过程在本文中称为“条码化”“DNA条码”是在DNA测序过程中用于鉴定靶分子的DNA序列。在一些实施方式中,DNA条码文库是随机生成的,例如,通过在库中装配寡核苷酸。在其他实施方式中,DNA条码文库在计算机中合理设计,然后被制造。

[0154] “结合亲和力”通常指分子(例如TCR、pMHC)的单个结合位点与其结合伴侣之间的非共价相互作用的总和的强度。除非另有说明,如本文所用,“结合亲和力”是指反映结合对的成员(例如,TCR和抗原)之间1:1相互作用的内在结合亲和力。分子X对其伴侣Y的亲和力可以由解离常数(Kd)表示。例如,Kd可以是约200nM、150nM、100nM、60nM、50nM、40nM、30nM、20nM、10nM、8nM、6nM、4nM、2nM、1nM或更强,包括高达1 $\mu$ M。亲和力可以通过本领域已知的常用方法来测量,包括本文所述的那些方法。低亲和力TCR通常缓慢结合抗原并倾向于容易地解离,而高亲和力TCR通常更快地结合抗原并倾向于保持更长的结合。本领域已知测量结合亲和力的多种方法,其中任何一种都可用于本公开的目的。

[0155] 术语“生物正交化学”是指在不干扰天然生物化学过程的情况下,在活系统内部发生的任何化学反应。该术语包括化学反应,该化学反应是在水的存在下或生理pH下在体外发生的化学反应。考虑到生物正交性,反应是选择性的,避免与起始化合物中发现的其他官能团发生副反应。此外,反应伴侣之间产生的共价键应是强的,对生物反应具有化学惰性,并且不应影响所需分子的生物活性。

[0156] 如本文所用,术语“运载体”和“药学上可接受的运载体”包括生理上相容的任何及所有溶剂、分散介质、包衣、抗菌剂和抗真菌剂、等渗剂和吸收延迟剂等。

[0157] 本文中使用的术语“螯合配体”是指将放射性标记的人造基团共价连接到生物活性靶向分子(例如肽或蛋白质)的双官能偶联部分。双官能偶联部分利用官能团,如羧酸或活化酯用于酰胺偶联,异硫氰酸酯用于硫脲偶联,马来酰亚胺用于硫醇偶联。

[0158] 如本文所用,术语“可切割部分”指的是可切割的基序或序列。在一些实施方式中,可切割部分包含蛋白质,例如,酶促、可切割位点。在一些实施方式中,可切割部分包含化学可切割位点,例如,通过暴露于氧化/还原条件,光/声、温度、pH、压力等。

[0159] 术语“点击化学”是指一组可靠的选择性生物正交反应,用于快速合成新化合物和

组合文库。点击反应的特性包括模块性、范围广、产率高、立体特异性和简单产物分离(通过非色谱方法从惰性副产物中分离),以生产在生理条件下稳定的化合物。在放射化学和放射药理学中,点击化学是一组标记反应的总称,这些标记反应利用选择性和模块化构建块,并在没有催化剂的情况下实现化学选择性连接,以放射性标记生物相关化合物。“点击反应”可以用铜,也可以是无铜点击反应。点击化学处理和反应的非限制性示例显示在图1中。

[0160] 如本文所用,术语“足以进行共价偶联的条件”是指反应条件,包括但不限于温度、pH和反应组分的浓度,其适合于发生所需的共价偶联化学反应。

[0161] 如本文所用,术语“偶联多聚体”,也称为pMHC偶联多聚体,是指由包含偶联部分的pMHC单体与包含偶联部分的多聚结构域反应获得的反应产物,其中两个偶联部分相互反应以在pMHC单体和多聚结构域之间形成共价连接,从而形成偶联多聚体。在一个实施方式中,偶联多聚体是偶联四聚体,其中四个pMHC单体通过其偶联部分与多聚结构域反应,从而形成四聚体。在一个实施方式中,偶联多聚体是pMHCI偶联多聚体(例如四聚体),其中pMHC I类单体是多聚的。在一个实施方式中,偶联多聚体是pMHCII偶联多聚体(例如四聚体),其中pMHCII类单体是多聚的。

[0162] 如本文所用,术语“交联单元”可指与另一(相同或不同)分子连接的分子。在一些实施方式中,交联单元是单体。在一些实施方式中,交联是化学键。在一些实施方式中,交联是共价键。在一些实施方式中,交联是离子键。在一些实施方式中,交联改变了连接分子的至少一种物理性质,例如聚合物的物理性质。

[0163] 如本文所用,术语“内切蛋白酶”指的是切割非末端氨基酸的肽键的蛋白酶。

[0164] 如本文所用,术语“表位”(如“肽表位”)是指与免疫受体结合(相互作用或被其识别)的抗原(例如抗原蛋白)的一部分。因此,T细胞受体识别并结合与肽表位复合(加载)的MHC分子。

[0165] 本文中可互换使用的术语“可交换pMHC多肽”、“可交换pMHC多聚体”和“加载占位符肽的MHC多肽”指的是MHC单体和MHC多聚体,在MHC多肽的结合沟中包含占位符肽,也被称为“p\*MHC”单体或多聚体。“可交换”是指p\*MHC单体或p\*MHC-多聚体允许占位符肽与抗原肽交换的性质。在一个实施方式中,可交换pMHC或p\*MHC多肽包括MHC I类分子,在MHC I类分子的结合沟中具有MHC I型结合肽。在另一个实施方式中,可交换pMHC或p\*MHC多肽包括MHC II类分子,在MHC II类分子的结合沟中具有MHC II型结合肽。

[0166] 本文可互换使用的“融合蛋白”或“融合多肽”是指通过将两种多肽连接或融合成单个蛋白分子而制备的重组蛋白。

[0167] 本文中应用于MHC单体的术语“分离的”是指MHC糖蛋白,其处于非天然状态,例如,不与正常表达MHC的细胞的细胞膜结合。这个术语包括全长亚基链,以及MHC单体的功能性片段。功能性片段是包含抗原结合位点和被适当的T细胞受体识别所必需序列的片段。它通常包含全长链序列的至少约60-80%,通常90-95%。“分离的”MHC亚基成分可以从适当的细胞来源重组产生或溶解。在一个实施方式中,“分离的”MHC单体是MHC I类单体,例如MHC I类重链的可溶性形式( $\alpha$ 链),其与 $\beta 2$ -微球蛋白结合。在另一个实施方式中,“分离的”MHC单体是MHC II类单体,例如MHC II $\alpha/\beta$ 链的可溶性形式。

[0168] 如本文所使用的,术语“标识符”是指提供与标识符相对应的信息(例如特性)的数据的可读表示。

[0169] 如本文所用,术语“连接的”、“偶联的”、“融合的”或“融合”在指通过任何手段(包括重组或化学手段)将两种以上元件或组分或结构域连接在一起时可互换使用。

[0170] 术语“主要组织相容性复合体”或“MHC”是指含有一组基因的基因组位点,其编码多态性细胞膜结合糖蛋白,称为MHC经典I类和II类分子,通过分别向循环细胞毒性和辅助性T淋巴细胞递呈片段化蛋白的肽来调节免疫反应。在人类中,这组基因也称为“人白细胞抗原”或“HLA”系统。人MHC I类基因编码例如HLA-A、HL-B和HLA-C分子。HLA-A是人MHC I类细胞表面受体的三种主要类型之一。其他为HLA-B和HLA-C。HLA-A蛋白是一种异二聚体,由重 $\alpha$ 链和小 $\beta$ 链组成。 $\alpha$ 链由变体HLA-A基因编码, $\beta$ 链是非变体 $\beta$ 2微球蛋白( $\beta$ 2m)多肽。 $\beta$ 2微球蛋白多肽由人基因组的一个单独区域编码。HLA-A\*02(A\*02)是HLA-A血清型组内的人白细胞抗原血清型。血清型由HLA-A $\alpha$ 链 $\alpha$ 2结构域的抗体识别决定。对于A\*02, $\alpha$ 链由HLA-A\*02基因编码, $\beta$ 链由B2M基因座编码。人MHC II类基因编码例如HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-D1QB1、HLA-DRA和HLA-DRB1。人主要组织相容性复合体的完整核苷酸序列和基因图谱是公开的(例如,MHC测序联盟,Nature 401:921-923,1999)。

[0171] 如本文所用,术语“MHC分子”和“MHC蛋白”在本文中用于指由MHC I类和MHC II类基因编码的多态性糖蛋白,其参与肽表位向T细胞的递呈。术语“MHC I类”或“MHC I”可互换地用于指包含由三个结构域( $\alpha$ 1、 $\alpha$ 2和 $\alpha$ 3)组成的 $\alpha$ 链和第二个非变体 $\beta$ 2-微球蛋白的蛋白质分子。 $\alpha$ 3结构域是跨膜的,锚定MHC I型分子至细胞膜。抗原衍生的肽表位,位于 $\alpha$ 1/ $\alpha$ 2异二聚体的中心区域的肽结合沟中。MHC I类分子如HLA-A是向免疫系统递呈短多肽的过程的一部分。这些多肽的长度通常为9-11个氨基酸,并且来源于细胞表达的蛋白质。MHC I类分子向CD8+细胞毒性T细胞递呈抗原。术语“MHC II类”和“MHC II”可互换地用于指含有具有两个结构域( $\alpha$ 1和 $\alpha$ 2)的 $\alpha$ 链和具有两个结构域( $\beta$ 1和 $\beta$ 2)的 $\beta$ 链的蛋白质分子。肽结合沟由 $\alpha$ 1/ $\beta$ 1异二聚体形成。MHC II类分子向特异性CD4+T细胞递呈抗原。内源性递送至APC的抗原主要加工用于与MHC I类结合。外源性递送至APC的抗原主要加工用于与MHC II类结合。

[0172] 如本文所用,MHC蛋白(MHC I类或II类蛋白)还包括MHC变体,其含有氨基酸取代、缺失或插入,但仍结合MHC肽表位(MHCI类或MHC II类肽表位)。该术语还包括所有这些蛋白质的片段,例如保留肽结合的胞外结构域。

[0173] 术语“MHC蛋白”还包括非人脊椎动物物种的MHC蛋白。非人脊椎动物物种的MHC蛋白在这些脊椎动物物种疾病的检查和治疗中发挥作用,例如在兽医学和动物试验中在动物模型上检查人类疾病,例如小鼠(*mus musculus*)的EAE(实验性自身免疫性脑脊髓炎),其是人类疾病多发性硬化症的动物模型。例如,非人脊椎动物物种更具体为小鼠(*mus musculus*)、大鼠(褐家鼠(*rattus norvegicus*))、牛(*bos taurus*)、马(*equus equus*)和绿猴(*macaca mulatta*)。例如,小鼠的MHC蛋白被称为H-2蛋白,其中MHC I类蛋白由基因座H2K、H2L和H2D编码,MHC II类蛋白由基因座H2I编码。

[0174] 本文所用的“无肽MHC多肽”或“无肽MHC多聚体”是指MHC单体或MHC多聚体,其在MHC多肽的结合沟中不含肽。无肽MHC单体和多聚体也被称为“空”。在一个实施方式中,无肽MHC多肽或多聚体是MHC I类多肽或多聚体。在另一个实施方式中,无肽MHC多肽或多聚体是MHC II类多肽或多聚体。

[0175] 如本文所用,术语“多聚体”指的是多个单元。一些实施方式中,多聚体包含一种或多种不同单元。在一些实施方式中,多聚体中的单元是相同的。在一些实施方式中,多聚体

中的单元是不同的。在一些实施方式中,多聚体包含相同和不同单元的混合物。

[0176] 术语“肽表位”、“MHC肽表位”、“MHC肽抗原”和“MHC配体”在本文中可互换使用,并指可结合在MHC分子的肽结合沟中的MHC配体。肽表位通常可以由MHC分子递呈。肽表位通常具有8至25个氨基酸,这些氨基酸通过肽键连接。肽可以包含修饰,例如但不限于氨基酸残基的侧链、标记或标签的存在、合成氨基酸的存在、氨基酸的功能等效物等。典型的修饰包括由细胞机制产生的修饰,如聚糖加成和磷酸化。然而,其他类型的修改也在本公开的范围

内。  
[0177] 如本文所用,术语“肽交换”指的是竞争实验,其中占位符肽被除去并被“交换肽”(或“交换肽表位”)取代,本文也称为“拯救肽”(或者“拯救肽表位”)或“竞争肽”(或“竞争肽表位”)。通常,肽交换发生在通过肽切割释放占位符肽的条件下,或在允许拯救肽竞争与MHC单体或多聚体的结合口袋结合的合适条件下。例如,肽交换可以通过UV诱导交换、二肽诱导交换、温度诱导交换或本领域已知并在此公开的其他交换方法来实现。肽交换的示范性方法如图2所示。

[0178] 如本文所用,术语“肽文库”指的是多种肽。在一些实施方式中,文库包含具有独特序列的一种或多种肽。在一些实施方式中,文库中的每种肽具有不同的序列。在一些实施方式中,文库包含具有相同和不同序列的肽的混合物。

[0179] 如本文所用,术语“高多样性肽文库”指的是具有高度肽多变性的肽文库。例如,高多样性肽文库包含约 $10^3$ 、约 $10^4$ 、约 $10^5$ 、约 $10^6$ 、约 $10^7$ 、约 $10^8$ 、约 $10^9$ 、约 $10^{10}$ 、约 $10^{11}$ 、约 $10^{12}$ 、约 $10^{13}$ 、约 $10^{14}$ 、约 $10^{15}$ 、约 $10^{16}$ 、约 $10^{17}$ 、约 $10^{18}$ 、约 $10^{19}$ 、约 $10^{20}$ 或更多不同肽。

[0180] 如本文所用,术语“文库肽”指的是文库内的单个肽。

[0181] 如本文所用,术语“占位符肽”或“可交换肽”可互换地用于指与MHC蛋白(例如,MHC I或MHC II蛋白)具有足够亲和力结合并导致或促进MHC蛋白从未折叠状态适当折叠或折叠的MHC蛋白稳定化的肽或肽样化合物。占位符肽随后可以与不同的感兴趣肽(称为交换肽或拯救肽)交换。这种交换可以通过UV诱导交换、二肽诱导交换、温度诱导交换或本领域已知的其他交换方法来实现。

[0182] 术语“多肽”、“肽”和“蛋白质”在本文中互换使用,指氨基酸残基的聚合物。该术语适用于其中一个或多个氨基酸残基是相应天然产生氨基酸的人造化学模拟物的氨基酸聚合物,以及天然产生的氨基酸聚合物和非天然产生的氨基酸聚合物。术语“分离的蛋白质”和“分离的多肽”可互换地用于指已从其他组分(例如蛋白质、细胞材料)和/或化学品中分离或纯化的蛋白质(例如可溶性多聚蛋白)。通常,当多肽占样品中总蛋白质的至少60(例如,至少65、70、75、80、85、90、92、95、97或99)重量%时,多肽被纯化。

[0183] 如本文所用,术语“蛋白质折叠”指的是肽的空间结构。在一些实施方式中,氨基酸序列影响肽的空间结构或折叠。在一些实施方式中,肽可能以功能性构象折叠。在一些实施方式中,折叠的肽具有一种或多种生物学功能。在一些实施方式中,折叠的肽获得三维结构。

[0184] 如本文所用,术语“N末端氨基酸残基”指的是位于多肽的N末端的一个或多个氨基酸。

[0185] 如本文所用,术语“小泛素样修饰部分”或“SUMO结构域”或“SUMO部分”可互换使用,指的是特定蛋白酶识别部分。



[0186] 如本文所用,“标签”是指寡核苷酸组分,通常是DNA,其提供定位(address)与其连接的靶分子(例如偶联多聚体)的手段。例如,在一些实施方式中,标签包含允许对标签所附连的分子进行鉴定、识别和/或分子或生物化学操作的核苷酸序列(例如,通过提供用于退火寡核苷酸的位点和/或独特序列,所述寡核苷酸例如为用于通过DNA聚合酶延伸的引物,或用于捕获或用于连接反应的寡核苷酸)。将标签连接到靶分子的过程在本文中有时称为“标签化”,并且经历标签化或含有标签的靶分子被称为“带标签的”(例如,“带标签的偶联多聚体”)。标签可以是条码,衔接体序列,引物杂交位点或其组合。

[0187] 术语“T细胞”指的是一种白细胞,可以通过细胞表面存在T细胞受体而与其他白细胞区分开来。存在几种T细胞亚群,包括但不限于辅助T细胞(也称为 $T_H$ 细胞或 $CD4^+$ T细胞)和亚型,包括 $T_H1$ 、 $T_H2$ 、 $T_H3$ 、 $T_H17$ 、 $T_H9$ 和 $T_{FH}$ 细胞,细胞毒性T细胞(也称为 $T_C$ 细胞、 $CD8^+$ T细胞、细胞毒性T淋巴细胞、T杀伤细胞、杀伤性T细胞),记忆T细胞和亚型(包括中央记忆T细胞( $T_{CM}$ 细胞)、效应记忆T细胞( $T_{EM}$ 和 $T_{EMRA}$ 细胞)、常驻记忆T细胞,调节性T细胞(也称为 $T_{reg}$ 细胞或抑制性T细胞)和亚型,包括 $CD4^+FOXP3^+$  $T_{reg}$ 、 $CD4^+FOXP3^-T_{reg}$ 、Tr1细胞、Th3细胞和 $T_{reg}17$ 细胞,自然杀伤T细胞(又称为NKT细胞)、粘膜相关不变T细胞(MAIT)和 $\gamma\delta$ T细胞( $\gamma\delta$ T),包括 $V\gamma9/V\delta2$ T细胞。术语“T细胞毒性”包括由 $CD8^+$ T细胞活化介导的任何免疫反应。

[0188] 如本文所用的词语“T细胞受体”和术语“TCR”指一种T细胞表面蛋白,其使T细胞识别通常与一种或多种主要组织相容性复合物(MHC)分子结合的抗原和/或其表位。TCR的功能是识别抗原决定簇并引发免疫反应。通常,TCR是包含两条不同蛋白链的异二聚体。在绝大多数T细胞中,TCR包含 $\alpha$ 链和 $\beta$ 链。每条链包括两个胞外结构域:可变(V)区和恒定(C)区,后者在膜近端。 $\alpha$ 链和 $\beta$ 链的可变结构域由三个高可变区组成,这些高可变区也称为互补决定区(CDR)。CDR,特别是CDR3,主要负责接触抗原,从而定义了TCR的特异性,尽管 $\alpha$ 链的CDR1可以与抗原的N末端部分相互作用,而 $\beta$ 链的CDR1可以与抗原C末端部分相互作用。大约5%的T细胞具有由 $\gamma$ 和 $\delta$ ( $\gamma/\delta$ )链组成的TCR。氨基酸序列的所有编号以及TCR的蛋白质环和片层的命名均根据IMGT编号方案(IMGT,国际ImMunoGeneTics信息system@imgt.cines.fr; <http://imgt.cines.fr>;Lefranc等人,(2003)Dev Comp Immunol 27:55-77;Lefranc等人(2005)Dev Comp Immunol 29:185-203)。

[0189] 如本文所用,术语“可溶性T细胞受体”和“sTCR”是指TCR的异二聚体截短变体,其包括TCR $\alpha$ 链和 $\beta$ 链的胞外部分(例如,通过二硫键连接),但缺乏全长蛋白的跨膜和胞质结构域。与相应的天然TCR序列相比,可溶性TCR $\alpha$ 链和 $\beta$ 链的序列(氨基酸或核酸)可以与天然TCR中的相应序列相同,或者可以包括变体的可溶性TCR $\alpha$ 链和 $\beta$ 链序列。本文使用的术语“可溶性T细胞受体”包括具有变异或非变异可溶性TCR $\alpha$ 链和 $\beta$ 链序列的可溶性TCR。这些变异可能在可溶性TCR $\alpha$ 链和 $\beta$ 链序列的可变或恒定区,并且可以包括但不限于氨基酸缺失、插入、取代突变以及核酸序列的变化,这些变化不会改变氨基酸序列。变体保留了其母分子的结合功能。

[0190] 如本文所用,“TCR/pMHC复合物”是指通过T细胞受体(TCR)或其可溶性部分与加载肽的MHC分子之间结合形成的蛋白质复合物。因此,“TCR/pMHC复合物的成分”是指TCR的一个或多个亚基(例如 $V\alpha$ 、 $V\beta$ 、 $C\alpha$ 、 $C\beta$ ),或MHC或pMHC I类或II类分子的一个或者多个亚基。

[0191] 如本文所使用的,术语“无偏”指缺少一个或多个选择标准。

[0192] 综述

[0193] 本公开提供了高通量产生文库的方法,所述文库包含加载肽的MHC (pMHC) 多聚体,其在MHC结合沟中包含多个独特肽,且具有寡核苷酸条码标记,以实现文库成员的鉴定。在本文提供的方法中,用于生成pMHC多聚体的所有具有挑战性和潜在低效的化学步骤都是在单个批次反应中完成的,包括色谱净化和纯化,然后是高效的肽交换和寡核苷酸条码化。特别是,通过使用单体和多聚结构域上的偶联部分将pMHC单体连接到多聚结构域,这些偶联部分反应以在单体和多聚结构域之间形成稳定的化学连接(即共价键),从而形成pMHC偶联多聚体,如pMHC偶联四聚体。如本文所述,各种偶联部分和反应适合用于形成偶联多聚体,包括使用生物正交化学,如点击化学,其使得反应容易而高效。此外,当多聚结构域是链霉亲和素时,由于生物素结合位点不用于连接pMHC单体,因此该生物素结合部位可用于方便地连接生物素化寡核苷酸条码,从而容易且有效地标记多聚体。

[0194] 本文提供的pMHC多聚体文库在一系列治疗、诊断和研究应用中是有用的,基本上在pMHC多聚体有用的任何情况下都是有用的。例如,本文所述的pMHC多聚体可用于多种方法,例如在广泛的应用中鉴定和分离特异性T细胞。在一个实施方式中,pMHC多聚体是pMHC I类多聚体,其可用于确定CD8<sup>+</sup>T细胞(例如细胞毒性T细胞)的抗原特异性。在另一个实施方式中,pMHC多聚体是pMHC II类多聚体,其可用于确定CD4<sup>+</sup>T细胞(例如辅助T细胞)的抗原特异性。

#### [0195] I. MHC多肽

##### [0196] A. MHC I类多肽

[0197] I类组织相容性三元复合物由非共价键结合的三个部分组成。MHCI重链是一种约45kDa的多态性跨膜糖蛋白,由三个胞外结构域组成,每个结构域包含约90个氨基酸(N末端的 $\alpha 1$ 、 $\alpha 2$ 和 $\alpha 3$ )、约40个氨基酸的跨膜结构域和约30个氨基酸的胞质尾。MHCI重链的 $\alpha 1$ 和 $\alpha 2$ 结构域包含两个 $\alpha$ 螺旋区段,其形成肽结合沟或裂隙。一个约8-10个氨基酸的短肽非共价结合(“匹配”)到两个 $\alpha$ 螺旋之间的沟中。MHCI重链的 $\alpha 3$ 结构域靠近质膜。MHCI重链与 $\beta 2m$ 多肽非共价结合,形成三元复合物。在MHCI中,结合沟在两端被保守的酪氨酸残基封闭,导致结合肽的大小限制为通常8-10个残基,其C末端停靠在F口袋中。

[0198] 本公开提供了一种包含两种或多种MHCI或MHCI样多肽的多聚蛋白。MHCI分子可以适当地是脊椎动物MHC分子,例如人、小鼠、大鼠、猪、牛或禽类MHC分子。

[0199] 在一些实施方式中,本文所述的多聚MHCI多聚体,MHC分子是人MHC I类蛋白:HLA-A、HLA-C的HLA-B。在一些实施方式中,多聚体包括MHC I类分子(包括非经典MHC I类分子),包括但不限于CD1d、HLA E、HLA G、HLA F、HLA H、MIC A、MIC B、ULBP-1、ULBP-2和ULBP-3。来自各种脊椎动物物种的MHCI重链、 $\beta 2m$ 多肽和MHC I类分子的氨基酸序列在本领域是已知的,并且是公众可得的。

[0200] 在一些实施方式中,MHCI重链 $\alpha$ 结构域是人的,并且包括例如来自人MHC I类分子的MHC I重链 $\alpha$ 结构域,选自:HLA-A\*01:01、HLA-A\*03:01、HLA-A\*11:01、HLA-A\*24:02、HLA-B\*07:02、HLA-C\*04:01、HLA-C\*07:02、HLA-B\*08:01、HLA-B\*35:01、HLA-B\*57:01、HLA-B\*57:03、HLA-E、HLA-C\*16:01、HLA-C\*08:02、HLA-C\*07:01、HLA-C\*05:01、HLA-B\*44:02、HLA-A\*29:02、HLA-B\*44:03、HLA-C\*03:04、HLA-B\*40:01、HLA-C\*06:02、HLA-B\*15:01、HLA-C\*03:03、HLA-A\*30:01、HLA-B\*13:02、HLA-C\*12:03、HLA-A\*26:01、HLA-B\*38:01、HLA-B\*14:02、HLA-A\*33:01、HLA-A\*23:01、HLA-A\*25:01、HLA-B\*18:01、HLA-B\*37:01、HLA-B\*51:01、HLA-

C\*14:02、HLA-C\*15:02、HLA-C\*02:02、HLA-B\*27:05、HLA-A\*31:01、HLA-A\*30:02、HLA-B\*42:01、HLA-C\*17:01、HLA-B\*35:02、HLA-B\*39:06、HLA-C\*03:02、HLA-B\*58:01、HLA-A\*33:03、HLA-A\*68:02、HLA-C\*01:02、HLA-C\*07:04、HLA-A\*68:01、HLA-A\*32:01、HLA-B\*49:01、HLA-B\*53:01、HLA-B\*50:01、HLA-A\*02:05、HLA-B\*55:01、HLA-B\*45:01、HLA-B\*52:01、HLA-C\*12:02、HLA-B\*35:03、HLA-B\*40:02、HLA-B\*15:03和/或HLA-A\*74:01。这些MHC I分子的全长氨基酸序列(包括信号序列和跨膜结构域)分别显示在SEQ ID NO:28-93中。这些MHC I分子可溶性形式的氨基酸序列(缺乏信号序列和跨膜结构域)分别显示在SEQ ID NO:94-159中。

[0201] 在一些实施方式中,本文所述的pMHC I多聚体包括MHC I重链的 $\alpha 1$ 和 $\alpha 2$ 结构域。在一些实施方式中,本文描述的化合物包括MHC I重链的 $\alpha 1$ 、 $\alpha 2$ 和 $\alpha 3$ 结构域。

[0202] 在一些实施方式中,多聚体中的两种或多种pMHC I或pMHC I样多肽包括 $\beta 2$ -微球蛋白多肽,例如人 $\beta 2$ -微球蛋白。在一些实施方式中, $\beta 2$ -微球蛋白是野生型人 $\beta 2$ -微球蛋白。在一些实施方式中, $\beta 2$ -微球蛋白包含与人 $\beta 2$ -微球蛋白质的氨基酸序列至少80%、85%、90%、95%或99%相同的氨基酸序列,其全长序列如SEQ ID NO:160(UniProt ID.NO.P61769)所示。或者,pMHC I多聚体中使用的人 $\beta 2$ -微球蛋白多肽可以包含或由SEQ ID NO:2所示的氨基酸序列组成。

[0203] 在一些实施方式中,多聚蛋白包括可溶性MHC I多肽。在一些实施方式中,MHC-多聚蛋白包括可溶性MHC I $\alpha$ 结构域和 $\beta 2$ -微球蛋白多肽。在一些实施方式中,可溶性MHC I蛋白包括MHC I重链 $\alpha 1$ 结构域和MHC I重链 $\alpha 2$ 结构域。

[0204] 或者,在一些实施方式中,MHC I单体是包含与MHC I重链或其功能片段共价连接的 $\beta 2m$ 多肽或其功能性片段的融合蛋白。在一些实施方式中, $\beta 2m$ 的羧基(-COOH)末端与MHC I重链的氨基(-NH<sub>2</sub>)末端共价连接。

[0205] 在一些实施方式中,MHC I单体包括MHC I单体的各个组分之间的一个或多个接头。在一些实施方式中,MHC I单体包括通过接头与 $\beta 2m$ 融合的重链。在一些实施方式中,重链和 $\beta 2m$ 之间的接头是柔性接头,例如由甘氨酸和丝氨酸制备。在一些实施方式中,重链和 $\beta 2m$ 之间的柔性接头的长度在5-20个残基之间。在其他实施方式中,重链和 $\beta 2m$ 之间的接头是刚性的,具有确定结构,例如由氨基酸如谷氨酸、丙氨酸、赖氨酸和亮氨酸制成。在一个实施方式中,接头是(G<sub>4</sub>S)<sub>4</sub>接头(SEQ ID NO:181,其中n=4)。

[0206] 许多MHC I类蛋白的氨基酸序列是已知的,基因已被克隆,因此,重链单体可以用重组方法表达。已经广泛描述了MHC I分子的表达和纯化方法(例如,Altman等人,Curr.Protoc.Enz.17.3.1-17.2-44,2016)。例如,MHC I重链和 $\beta 2$ -微球蛋白可以在单独的细胞中表达,通过纯化分离,然后在体外重折叠。例如,MHC多肽链可以在大肠杆菌中表达,其中MHC多肽链作为不溶性包涵体在细菌细胞中积聚。体外重折叠发生在重折叠缓冲液中,其中通过例如透析或稀释添加多肽。重折叠缓冲液可以是其中允许MHC多肽链和肽重建天然三聚体折叠的任何缓冲液。缓冲液可含有氧化剂和/或还原剂,从而产生帮助MHC蛋白建立正确折叠的氧化还原缓冲系统。合适的重折叠缓冲液的实例包括但不限于Tris缓冲液、CAPS缓冲液、TAP缓冲液、PBS缓冲液,其他磷酸盐缓冲液、碳酸盐缓冲液和Ches缓冲液。还可以添加伴侣分子或其他改善正确蛋白质折叠的分子,并且同样可以将增加溶解度和防止聚集体形成的试剂添加到缓冲液中。此类分子的实例包括但不限于精氨酸、GroE、HSP70、HSP90、小有机化合物、DnaK、ClpB、脯氨酸、甘氨酸、甜菜碱、甘油、吐温、盐、PLURONIC™。

[0207] 一旦表达, MHC I 复合物可以直接纯化为来自表达 MHC I 的细胞的整个 MHC I 或 MHC I 肽单体。MHC I 单体可以在细胞表面表达, 然后通过使用例如去污剂破坏细胞膜, 然后纯化 MHC I 来分离。在一些实施方式中, MHC 单体被表达至周质中, 表达细胞被裂解并释放纯化的 MHC I 单体。或者, MHC 单体可以从将表达蛋白分泌到培养上清液中的细胞的上清液中纯化。纯化 MHC I 单体的方法在本领域中是公知的, 例如, 通过使用亲和和标签与亲和层析、涂覆抗-标签的珠和/或涉及将 MHC I 蛋白固定到亲和基质的其他技术; 使用例如凝胶过滤的尺寸排阻色谱法, 离子交换或其他能够从细胞和/或细胞裂解物中分离 MHC 分子的方法。

[0208] 在一些实施方式中, MHC I 多肽的重组表达允许对 MHC 单体进行多种修饰。例如, 重组技术提供了删除疏水性跨膜结构域的羧基末端截短方法。羧基末端也可以任意选择以促进配体或标记物的偶联, 例如通过将半胱氨酸和/或赖氨酸残基引入分子中。合成基因通常包括限制性位点, 以帮助插入表达载体和操纵基因序列。然后将编码合适单体的基因插入表达载体中, 在合适的宿主如大肠杆菌、酵母、昆虫或其他合适的细胞中表达, 并获得重组蛋白。例如, MHC I 类多肽的产生包括 MHC I 轻链  $\beta 2m$ -微球蛋白 ( $\beta 2m$ ) 的细菌表达和折叠, 以及由 MHC I 重链  $\beta 2m$  和占位符肽组成的复合物的形成。

[0209] 在一些实施方式中, MHC I 单体在其重链或  $\beta 2m$  上生物素化。在一些实施方式中, MHC I 单体在加载肽之前通过重折叠或肽交换生物素化。MHC 单体的生物素化可以如本领域已知的那样实现, 例如通过将生物素连接到作为生物素化酶识别位点的特定连接位点。在一些实施方式中, 生物素化酶是 BirA。在一些实施方式中, 生物素化在体内在所需的蛋白质链上进行, 作为蛋白质表达期间的翻译后修饰。

#### [0210] B. MHC II 类多肽

[0211] MHC II 类分子是由  $\alpha$  链和  $\beta$  链组成的异二聚体, 两者都由 MHC 编码。 $\alpha$  链包括  $\alpha 1$  和  $\alpha 2$  结构域。 $\beta$  链包括  $\beta 1$  和  $\beta 2$  结构域。链的  $\alpha 1$  和  $\beta 1$  结构域链非共价相互作用, 形成膜远端肽结合结构域, 而  $\alpha 2$  和  $\beta 2$  结构域形成膜近端免疫球蛋白样结构域。肽表位结合的抗原结合沟由两个  $\alpha$  螺旋和  $\beta$ -片层组成。由于 MHC II 类分子的抗原结合沟两端都是开放的, 因此该沟可以容纳比 MHC I 类分子更长的肽表位。MHC II 类分子递呈的肽表位长度通常约为 15-24 个氨基酸残基。

[0212] 本公开提供了一种包含两种或多种 MHC II 或 MHC II 样多肽的多聚蛋白。MHC II 分子可以适当地是脊椎动物 MHC II 分子, 例如人、小鼠、大鼠、猪、牛或禽类 MHC II 分子。

[0213] 在一些实施方式中, 本文所述的多聚 MHC II 多聚体, MHC 分子是人 MHC I 类蛋白: HLA-DR, HLA-DQ, HLA-DX, HLA-DO, HLA-DZ, 和 HLA-DP。来自包括人类在内的多种脊椎动物物种的 MHC II  $\alpha$  和  $\beta$  链的氨基酸序列在本领域是已知的, 并且可以公开获得。

[0214] 在一些实施方式中, 人 MHC II 分子是选自以下的同种型: DRB1\*0101 (见例如 Cameron 等人, (2002) J. Immunol. Methods, 268:51-69; Cunliffe 等人, (2002) Eur. J. Immunol., 32:3366-3375; Danke 等人, (2003) J. Immunol., 171:3163-3169), DRB1\*1501 (见例如 Day 等人, (2003) J. Clin. Invest, 112:831-842), DRB5\*0101 (见例如 Day 等人, *ibid*), DRB1\*0301 (见例如 Bronke 等人, (2005) Hum. Immunol., 66:950-961), DRB1\*0401 (见例如 Meyer 等人, (2000) PNAS, 97:11433-11438; Novak 等人, (1999) J. Clin. Invest, 104:R63-R67; Kotzin 等人, (2000) PNAS, 97:291-296), DRB1\*0402 (见例如 Veldman 等人, (2007) Clin. Immunol., 122:330-337), DRB1\*0404 (见例如 Gebe 等人, (2001) J. Immunol. 167:3250-

3256), DRB1\*1101 (见例如Cunliffe, *ibid*; Moro等人, (2005) *BMC Immunol.*, 6:24), DRB1\*1302 (见例如Laughlin等人, (2007) *Infect. Immunol.* 75:1852-1860), DRB1\*0701 (见例如Danke, *ibid*), DQA1\*0102 (见例如Kwok等人, (2000) *J. Immunol.*, 164:4244-4249), DQB1\*0602 (见例如Kwok, *ibid*), DQA1\*0501 (见例如Quarsten等人, (2001) *J. Immunol.*, 167:4861-4868), DQB1\*0201 (见例如Quarsten, *ibid*), DPA1\*0103 (见例如Zhang等人, (2005) *Eur. J. Immunol.*, 35:1066-1075; Yang等人, (2005) *J. Clin. Immunol.*, 25:428-436), 和DPB1\*0401 (见例如Zhang, *ibid*; Yang, *ibid*)。

[0215] 在一些实施方式中, MHCII分子是人的, 并包含例如MHC II $\alpha$ 和 $\beta$ 链, 选自HLA-DRA\*01:01, HLA-DRB1\*01:01, HLA-DRB1\*01:02, HLA-DRB1\*03:01, HLA-DRB1\*04:01, HLA-DRB1\*04:04, HLA-DRB1\*07:01, HLA-DRB1\*08:01, HLA-DRB1\*10:01, HLA-DRB1\*11:01, HLA-DRB1\*11:04, HLA-DRB1\*13:01, HLA-DRB1\*13:02, HLA-DRB1\*14:01, HLA-DRB1\*15:01, HLA-DRB1\*15:03, HLA-DQA1\*01:01, HLA-DQB1\*05:01, HLA-DQA1\*01:02, HLA-DQB1\*06:02, HLA-DQA1\*03:01, HLA-DQB1\*03:02, HLA-DQA1\*05:01, HLA-DQB1\*02:01, HLA-DQB1\*03:01, HLA-DQB1\*03:03, HLA-DQB1\*04:02, HLA-DQB1\*05:03, HLA-DQB1\*06:03和HLA-DQB1\*06:04。这些MHCII链的全长氨基酸序列(包括信号序列和跨膜结构域)分别显示在SEQ ID NO:194-223中。这些MHC II链可溶性形式的氨基酸序列(缺乏信号序列和跨膜结构域)分别显示在SEQ ID NO:224-253中。

[0216] 在某些实施方式中, 可以将额外的氨基酸序列附加到MHCII分子的 $\alpha$ 链或 $\beta$ 链的C末端序列, 例如用于标记和/或用于将介导附着(例如偶联)的部分附着到多聚结构域。例如, 可以附加avitag(通过Sav的生物素结合位点介导结合), 例如具有Myc标签和His标签的avitag(SEQ ID NO:254)或具有Myc标签的avitag(SEQ ID NO:255)。在另一个实施方式中, 可以附加分选标签(sortag)(如本文所述, 可以通过分选酶介导点击化学部分的偶联), 如SEQ ID NO:257所示的分选标签或SEQ ID NO:256所示的带有His标签的分选标签。在另一个实施方式中, V5标签(SEQ ID NO:258)被附加到C末端。

[0217] 在某些实施方式中, 异源二聚配对可以附加到MHCII分子的 $\alpha$ 链和/或 $\beta$ 链的C末端序列。这种异二聚配对序列的非限制性实例包括Fos和Jun(例如, 分别具有SEQ ID NO:259和260中所示的氨基酸序列)、酸性和碱性亮氨酸拉链(例如, 具有SEQ ID NO:261和262中所示氨基酸序列), 用于杵臼技术的杵臼序列(例如, 分别具有SEQ ID NO:263和264中所示的氨基酸序列)或spytap和spycatcher序列(例如, 分别具有SEQ ID NO:265和266中所示氨基酸序列)。

[0218] 在某些实施方式中, MHC II结合性占位符肽包含在MHCII链之一(优选 $\beta$ 链)的表达构建体中, 使得占位符肽类和可消化接头编码在(N末端)上游的构建体中并与MHCII链的编码序列操作性连接。例如, 表达构建体可以编码(从N末端到C末端): 占位符肽、可消化接头、MHCII链(例如 $\beta$ 链)和C末端标签(例如, 编码SEQ ID NO:192所示的氨基酸序列)。在某些实施方式中, 还将N末端标签附加在占位符肽上游, 这允许在肽交换后去除未交换的肽物质。这种N末端标签的非限制性实例包括FLAG标签(例如, 具有SEQ ID NO:267中所示的氨基酸序列)、Strep标签(例如, 具有SEQ ID NO:268中所示氨基酸序列)和蛋白C标签(例如: 具有SEQ ID NO:269中所示氨基酸序列)。

[0219] 在一些实施方式中, 本文所述的pMHCII多聚体包括MHCII $\alpha$ 链的 $\alpha 1$ 和 $\alpha 2$ 结构域和

MHCII $\beta$ 链的 $\beta$ 1和 $\beta$ 2结构域。在一些实施方式中,本文描述的多聚体仅包括MHCII重链的 $\alpha$ 1和 $\beta$ 1结构域。在其他实施方式中,pMHCII多聚体包括与肽结合的 $\alpha$ 链和 $\beta$ 链。其他实施方式包括仅由 $\alpha$ 链和 $\beta$ 链组成的MHCII分子(所谓的无加载肽的“空”MHC II)、截短的 $\alpha$ 链(例如 $\alpha$ 1结构域)与全长 $\beta$ 链(空的或加载有肽)的组合、截短的 $\beta$ 链(例如 $\beta$ 1结构域)与全长 $\alpha$ 链(空的或加载有肽)的组合,或截短 $\alpha$ 链与截短的 $\beta$ 链(例如 $\alpha$ 1和 $\beta$ 1结构域,空的或加载有肽)的组合。

[0220] 在一些实施方式中,多聚蛋白包括可溶性MHC II多肽。在一些实施方式中,MHC多聚蛋白包括缺乏跨膜和胞内结构域的可溶性MHCII。

[0221] 许多MHC II类蛋白(包括人MHCII)的氨基酸序列在本领域是已知的,这些基因已被克隆。因此, $\alpha$ 和 $\beta$ 链单体可以使用重组方法表达。表达和纯化MHC II分子的方法已被广泛描述(例如Crawford等人(1998) *Immunity*, 8:675-682;Novak等人(1999) *J.Clin.Invest.*, 104:R63-R67;Nepom等人(2002) *Arthrit.Rheum.*, 46:5-12;Day等人(2003) *J.Clin.Invest.*, 112:831-842;Vollers and Stern(2008) *Immunol.*, 123:305-313;Cecconi等人(2008) *Cytometry*, 73A:1010-1018其全部内容通过引用并入本文)。

[0222] 对于MHC II分子, $\alpha$ 链和 $\beta$ 链可以作为单个多肽在单独的细胞中表达,或者作为融合蛋白在同一细胞中表达。MHC II肽复合物的肽可以单独生产,并在纯化完整MHC复合物后添加,或在体外重折叠期间添加,或与通过接头连接到 $\alpha$ 链和/或 $\beta$ 链任一链一起表达。遗传物质可以编码MHC II类 $\alpha$ 链和 $\beta$ 链的全部或仅一个片段。遗传物质可以与编码其他蛋白质的基因融合,包括用于纯化表达的多肽链的蛋白质(例如,纯化标签)、用于增加/减少多肽溶解度的蛋白质、用于检测多肽的蛋白质、参与MHC复合物与多聚结构域偶联和/或标记物与MHC复合物和/或MHC多聚体偶联的蛋白质。

[0223] 与MHC I复合物相比,MHC II复合物在体外变性后不易重折叠。只有一些MHC II等位基因可以在大肠杆菌中表达并在体外重折叠。因此,用于生产MHC II分子的优选表达系统是真核系统,其中蛋白质表达后不需要重折叠。优选的表达系统包括哺乳动物表达系统,例如CHO细胞、HEK细胞或其他适合表达人类蛋白质的哺乳动物细胞系。其他表达系统包括稳定的果蝇细胞转染子、杆状病毒感染的昆虫细胞或其他适合表达蛋白质的哺乳动物细胞系。

[0224] 可溶性MHC II复合物的稳定甚至比MHC I分子更重要,因为 $\alpha$ 链和 $\beta$ 链都参与了肽结合沟的形成,并且在未嵌入细胞膜时趋于解离。因此,在一个实施方式中,制备MHCII单体,其中肽与MHCII分子共价连接。例如,一种方法是单链MHC II类链-肽复合物的共价合成,通过工程改造与 $\beta$ 链cDNA接近的肽特异性互补DNA(cDNA)序列来指导(如Crawford等人(1999) *Immunity*, 8:675-682所述)。在这种策略中,所得多肽与从II类分子的氨基末端延伸的肽序列重折叠。肽中的系接头序列允许肽具有足够的灵活性,以占据成熟II类分子中的肽结合沟。可切割接头可用于肽和MHCII分子之间的共价连接的切割(例如,如Day等人(2003) *J.Clin.Invest.*, 112:831-842所述),从而允许肽交换和MHCII分子与其他肽(例如不同肽的文库)的加载。

[0225] 一旦表达,MHC II复合物可以直接纯化为来自表达MHCII的细胞的整个MHC II或MHC II肽单体。MHC II单体可以在细胞表面表达,然后通过使用例如去污剂破坏细胞膜,然后纯化MHC II来分离。在一些实施方式中,MHC单体被表达至周质中,表达细胞被裂解并释放纯化的MHC II单体。或者,MHC单体可以从将表达蛋白分泌到培养上清液中的细胞的上清

液中纯化。纯化MHC II单体的方法在本领域中是公知的,例如,通过使用亲和标签与亲和层析、涂覆抗-标签的珠和/或涉及将MHC II蛋白固定到亲和基质的其他技术;使用例如凝胶过滤的尺寸排阻色谱法、离子交换或其他能够从细胞和/或细胞裂解物中分离MHC分子的方法。

[0226] 在一些实施方式中,MHC II多肽的重组表达允许对MHC单体进行多种修饰。例如,重组技术提供了删除疏水性跨膜结构域的羧基末端截短方法。羧基末端也可以任意选择以促进配体或标记物的偶联,例如通过将半胱氨酸和/或赖氨酸残基引入分子中。合成基因通常包括限制性位点,以帮助插入表达载体和操纵基因序列。然后将编码合适单体的基因插入表达载体中,在合适的宿主如大肠杆菌、酵母、昆虫或其他合适的细胞中表达,并获得重组蛋白。

[0227] 在一些实施方式中,MHCII单体在其 $\alpha$ 链或 $\beta$ 链上被生物素化。在一些实施方式中,MHC II单体在加载肽之前通过重折叠或肽交换生物素化。MHC单体的生物素化可以如本领域已知的那样实现,例如通过将生物素连接到作为生物素化酶识别位点的特定连接位点。在一些实施方式中,生物素化酶是BirA。在一些实施方式中,生物素化在体内在所需的蛋白质链上进行,作为蛋白质表达期间的翻译后修饰。

[0228] II. 占位符肽

[0229] A. MHC I类占位符肽

[0230] 在本文提供的方法中,MHCI单体加载占位符肽,以促进MHCI单体能正确折叠,从而在多聚化之前产生加载占位符-肽的MHCI ( $p^*$ MHCI)。本领域已经描述了占位符肽的实例以及在占位符肽存在下体外诱导MHCI重链和 $\beta$ 2-微球蛋白折叠的方法(例如,Bakker等人,PNAS 105:3825-3830,2008;Rodenko等人,Nat. Prot. 1:1120-1132,2006)。

[0231] 在一些实施方式中,占位符肽是HLA-A、HLA-B或HLA-C肽。在一些实施方式中,占位符肽是HLA-A1肽(例如,A\*1:01结合肽)。在一些实施方式中,占位符肽是HLA-A2肽(例如,A\*02:01结合肽、A\*02:12结合肽、A\*02:06结合肽)。在其它实施方式中,占位符肽是HLA-A3肽(例如A\*3:01结合肽)、HLA-A11肽(例如A\*11:01结合肽)、HLA-A23肽(例如A\*23:01结合肽)、HLA-A24肽(例如A\*24:02结合肽)、HLA-A26肽(例如A\*26:01结合肽)、HLA-A29肽(例如A\*29:02结合肽)、HLA-A30肽(例如A\*30:01结合肽;A\*30:02结合肽)、HLA-A31肽(例如A\*31:01结合肽)、HLA-A32肽(例如A\*32:01结合肽)、HLA-A33肽(例如A\*33:01结合肽;A\*33:03结合肽)、HLA-A68肽(例如A\*68:02结合肽)、HLA-B7肽(例如B\*07:02结合肽)、HLA-B8肽(例如B\*08:01结合肽)、HLA-B15肽(例如B\*15:01结合肽;B\*15:03结合肽)、HLA-B18肽(例如B\*18:01结合肽)、HLA-B35肽(例如B\*35:01结合肽)、HLA-B38肽(例如B\*38:01结合肽)、HLA-B40肽(例如B\*40:01结合肽;B\*40:02结合肽)、HLA-B45肽(例如B\*45:01结合肽)、HLA-B51肽(例如B\*51:01结合肽)、HLA-B53肽(例如B\*53:01结合肽)、HLA-B58肽(例如B\*58:01结合肽)、HLA-C3肽(例如C\*03:03结合肽;C\*03:04结合肽)、HLA-C4肽(例如C\*04:01结合肽)、HLA-C7肽(例如C\*07:01结合肽;C\*07:02结合肽)或HLA-C8肽(例如C\*08:01结合肽)。在一些实施方式中,占位符肽是合成肽。

[0232] 在一些实施方式中,占位符肽对MHCI结合沟的亲和力低于拯救肽。在一些实施方式中,占位符肽对MHCI结合沟的亲和力比拯救肽低10倍。在一些实施方式中,占位符肽对MHCI结合沟的亲和力高于拯救肽;然而,通过使用过量浓度的拯救肽,占位符肽仍然可以被

拯救肽替换。

[0233] 在一些实施方式中,占位符肽是热不稳定的。在一些实施方式中,占位符肽在约30-37°C之间热不稳定。在一些实施方式中,占位符肽在30°C或以上、32°C或以上、34°C或以上、35°C或以上、36°C或以上、或约37°C或以上的温度下是不稳定的。已经描述了热不稳定占位符肽和鉴定和生产热不稳定占位符肽的方法(例如,WO 93/10220;WO 2005/047902;US 2008/0206789;Luimstra等人,Curr.Protoc.Imunol.126(1):e852019;Luimstra等人,J.Exp.Med.215(5):1493-1504,2018)。

[0234] 在一些实施方式中,占位符肽在酸性pH下是不稳定的。在一些实施方式中,占位符肽在约pH 2.5和6.5之间是不稳定的。在一些实施方式中,占位符肽在约2.5-6.0、3.0-6.0、3.0-6.5、3.5-6.0、3.5-6.5、4.0-6.0、4.0-6.5、4.5-6.0、4.5-6.5、5.0-6.0、5.0-6.5、5.0、5.5、6.0或6.5的pH下不稳定。在一些实施方式中,占位符肽在碱性pH下是不稳定的。在一些实施方式中,占位符肽在约pH 9-11之间是不稳定的。在一些实施方式中,占位符肽在pH 9或以上、pH 9.5或以上、在pH 10或约pH 10、在pH 10.5或约pH10.5、在pH 11或约pH11不稳定。产生和使用pH敏感占位符肽的方法是公开的,例如,如WO 93/10220中所述;US 2008/0206789;和Cameron等人,J.Immunol.Meth.268:51-59。

[0235] 在一些实施方式中,占位符肽包括可切割部分。本领域已知各种类型的可切割部分,包括例如通过光照射、酶、亲核或亲电试剂、还原和氧化试剂被切割的部分(例如,在Leriche等人,Biorg.Med.Chem.20(2):571-5822012中综述)。

[0236] 在一些实施方式中,可切割占位符肽包括一种或多种可光切割的非天然β-氨基酸。在一些实施方式中,占位符肽包括3-氨基-3-(2-硝基-苯基)-丙酸。在一些实施方式中,占位符肽包括(2-硝基)苯基甘氨酸。在一些实施方式中,占位符肽包括偶氮苯基团。在一些实施方式中,HLA-A2占位符肽是A\*02:01、KILGFVJV (SEQ ID NO:15) 或GILGFVJL (SEQ ID NO:7),其中J是3-氨基-3-(2-硝基)苯基丙酸。在一些实施方式中,占位符肽选自以下组:A\*01:01,STAPGJLEY (SEQ ID NO:16);A\*03:01,RIYRJGATR (SEQ ID NO:17);A\*11:01,RVFAJSFIK (SEQ ID NO:18);A\*24:02,VYGJVRACL (SEQ ID NO:11);B\*07:02,AARGJTLAM (SEQ ID NO:14);B\*35:01,KPIVVLJGY (SEQ ID NO:19);C\*03:04,FVYGJSKTS (SEQ ID NO:20),B\*08:01,FLRGRAJGL (SEQ ID NO:21);C\*07:02,VRIJHLYIL (SEQ ID NO:22);C\*04:01,QYDJAVYKL (SEQ ID NO:23);B\*15:01,ILGPJGSVY (SEQ ID NO:24);B\*40:01,TEADVQJWL (SEQ ID NO:25);B\*58:01,ISARGQLF (SEQ ID NO:26);和C\*08:01,KAAJDLSHFL (SEQ ID NO:27),其中J是3-氨基-3-(2-硝基)苯基丙酸。在另一个实施方式中,占位符肽包含SEQ ID NO:7-27或271-279中任一所示的序列。在另一个实施方式中,占位符肽由SEQ ID NO:7-27或271-279中任一所示的序列组成。

[0237] 产生含有可光切割氨基酸的占位符肽的方法在本领域是已知的,并且已经在前面描述过(例如,Toebe等人,Curr.Protoc.Imunol.87:18.16.1-18.16.20,2009;Bakker等人,上文,Rodenko等人,上文)。在各种实施方式中,使用先前描述的方法(例如,Toebe等人,Nat Med.2006Feb;12(2):246-51;Bakker等人,Proc Natl Acad Sci U S A.,2008年3月11日;105(10):3825-30;Rodenko等人,Nat Protoc.2006;1(3):1120-32;Frøsig等人,Cytometry A,2015年10月;87(10):967-75)在暴露于紫外光后切割可光切割占位符肽。

[0238] 在一些实施方式中,占位符肽包括化学选择性部分。在一些实施方式中,化学选择



性部分包括连二亚硫酸钠敏感的偶氮苯接头,其中偶氮苯包括至少一个包含电子供体基团的芳族基团,并且位于两个氨基酸残基之间。偶氮苯接头和化学选择性肽交换的方法在本领域是已知的,例如,如美国专利10,400,024所述。

[0239] 在一些实施方式中,占位符肽包括在暴露于氨肽酶时被切割的可切割部分。在一些实施方式中,氨基酸残基的切割通过使用甲硫氨酸氨肽酶发生。当第二位的氨基酸残基为例如甘氨酸、丙氨酸、丝氨酸、半胱氨酸或脯氨酸时,甲硫氨酸氨肽酶可从肽中切割甲硫氨酸。在一些实施方式中,可切割部分包括凝血酶切割结构域。

[0240] 在一些实施方式中,占位符肽包括对化学触发物敏感的可切割部分。在一些实施方式中,占位符肽包括高碘酸盐敏感氨基酸。在其它实施方式中,高碘酸盐敏感氨基酸包括邻二醇部分。在一些实施方式中,高碘酸盐敏感氨基酸包括邻氨基醇部分。在一些实施方式中,高碘酸盐敏感氨基酸是含1,2-氨基-醇的氨基酸。在一些实施方式中,高碘酸盐敏感氨基酸是 $\alpha$ ,  $\gamma$ -二氨基- $\beta$ -羟基丁酸(DAHB)。生产和使用含有高碘酸盐敏感氨基酸的肽的方法是公开的,例如,如Rodenko等人(J. Am. Chem. Soc. 131:12605-12313, 2009)和Amore等人(ChemBioChem 14:123-131, 2013)所述。

[0241] 在一些实施方式中,占位符肽是二肽。在一些实施方式中,二肽结合到MHC I结合沟的F口袋。在一些实施方式中,二肽的第二氨基酸是疏水性的。在一些实施方式中,二肽选自甘氨酸-亮氨酸(GL)、甘氨酸-缬氨酸(GV)、甘氨酸-甲硫氨酸(GM)、甘氨酸-环己基丙氨酸(GCha)、甘氨酸-高亮氨酸(GH1e)和甘氨酸-苯丙氨酸(GF)。生产和使用作为占位符肽的二肽的方法是公开的,例如,如Saini等人(PNAS 112:202-207, 2015)所述。

[0242] 在一些实施方式中,占位符肽包括GILGFVFL(SAQ ID NO:7)。在一些实施方式中,占位符肽由GILGFVFL(SAQ ID NO:7)组成。在其它实施方式中,占位符肽包含SAQ ID NO:8-27或271-279中任一所示的序列。在其它实施方式中,占位符肽由SAQ ID NO:8-27或271-279中任一所示的序列组成。

[0243] 在一些实施方式中,占位符肽还包括荧光标记。在一些实施方式中,荧光标记附着在占位符肽中的半胱氨酸残基上。

[0244] 在一些实施方式中,p\*MHCI分子被纯化,并被储存以用作储备分子的来源,当暴露于本文所述的肽交换条件时,所述储备分子可与感兴趣的肽表位交换。

#### [0245] B. MHC II类占位符肽

[0246] 在本文提供的方法中,MHC II单体加载占位符肽,以促进MHC II单体能正确折叠,从而在多聚化之前产生加载占位符-肽的MHC II(p\*MHCII)。在各种实施方式中,占位符肽是结合HLA-DR、HLA-DQ、HLA-DX、HLA-DO、HLA-DZ或HLA-DP的肽。在一些实施方式中,占位符肽是合成肽。

[0247] 在一些实施方式中,占位符肽对MHC II结合沟的亲合力低于拯救肽。在一些实施方式中,占位符肽对MHC II结合沟的亲合力比拯救肽低10倍。

[0248] 在一些实施方式中,占位符肽是热不稳定的。在一些实施方式中,占位符肽在约30-37°C之间热不稳定。在一些实施方式中,占位符肽在30°C或以上、32°C或以上、34°C或以上、35°C或以上、36°C或以上、或约37°C或以上的温度下是不稳定的。已经描述了热不稳定占位符肽和鉴定和生产热不稳定占位符肽的方法(例如,WO 93/10220;WO 2005/047902;US 2008/0206789;Luimstra等人,Curr. Protoc. Immunol. 126(1):e852019;Luimstra等人,

J. Exp. Med. 215 (5) :1493-1504, 2018)。

[0249] 在一些实施方式中,占位符肽在酸性pH下是不稳定的。在一些实施方式中,占位符肽在约pH 2.5和6.5之间是不稳定的。在一些实施方式中,占位符肽在约2.5-6.0、3.0-6.0、3.0-6.5、3.5-6.0、3.5-6.5、4.0-6.0、4.0-6.5、4.5-6.0、4.5-6.5、5.0-6.0、5.0-6.5、5.0、5.5、6.0或6.5的pH下不稳定。在一些实施方式中,占位符肽在碱性pH下是不稳定的。在一些实施方式中,占位符肽在约pH 9-11之间是不稳定的。在一些实施方式中,占位符肽在pH 9或以上、pH 9.5或以上、在pH 10或约pH 10、在pH 10.5或约pH10.5、在pH 11或约pH11不稳定。产生和使用pH敏感占位符肽的方法是公开的,例如,如WO 93/10220中所述;US 2008/0206789;和Cameron等人,J. Immunol. Meth. 268:51-59。

[0250] 在一些实施方式中,占位符肽包括可切割部分。本领域已知各种类型的可切割部分,包括例如通过光照射、酶、亲核或亲电试剂、还原和氧化试剂被切割的部分(例如,在Leriche等人,Biorg. Med. Chem. 20 (2) :571-5822012中综述)。

[0251] 在一个实施方式中,将占位符肽融合到降解标签上,并且通过在相应蛋白酶(消化降解标签)的存在下以及拯救肽的存在下的蛋白水解促进肽交换。

[0252] 在一些实施方式中,可切割占位符肽是可光切割肽,例如在暴露于UV光时被切割。例如,占位符肽可以包括一种或多种可光切割的非天然氨基酸。已经描述了结合MHCII的可光切割肽,例如,纳入UV敏感氨基酸类似物3-氨基-3-(2-硝基苯基)-丙酸盐(参见例如,Negrone和Stern (2018) PLoS One, 13 (7) :e0199704)。

[0253] 在一个实施方式中,MHCII占位符肽是CLIP肽,例如具有氨基酸序列KPVSKMRMATPLLQALPMGA (SEQ ID NO:189) 或ATPLLQALPMGA (SEQ ID NO:280)。在一个实施方式中,CLIP肽是可切割的。在一个实施方式中,MHCII单体与共价附着的可切割CLIP肽合成,例如通过合成单链MHC II类链-肽复合物,通过工程改造与β链cDNA接近的肽特异性互补DNA (cDNA) 序列来指导(参见例如Day等人 (2003) J. Clin. Invest., 112:831-842)。CLIP肽(作为占位符肽)和MHCII之间的共价连接的切割因此允许与其他MHCII结合肽进行肽交换。

[0254] 基于MHCII分子与其已知的MHCII结合肽的适当配对,本领域已经描述了可以用作占位符肽的其他MHCII结合肽。已知MHCII分子/MHCII结合肽对的非限制性实例包括:DRA1\*0101/DRB1\*0401和血凝素的免疫显性肽,HA<sub>307-319</sub>(见Novak等人 (1999) J. Clin. Invest., 104:R63-R67) 和HLA-DR\*1101和破伤风类毒素(TT)-衍生的p2肽(TT<sub>830-844</sub>),具有氨基酸序列QIYKANSKFIGITEL (SEQ ID NO:190) (见Cecconi等人 (2008) Cytometry, 73A:1010-1018)。

[0255] III. p\*MHC多聚体的产生

[0256] 用于制备本文提供的pMHC多聚体的多聚结构域包括蛋白质、多肽或其他适合于两种或多种pMHC或p\*MHC单体共价偶联的多聚部分,其不干扰pMHC多肽与细胞的结合。在一些实施方式中,多聚结构域包括蛋白质亚基。在一些实施方式中,多聚结构域是蛋白质亚基的同多聚体。在一些实施方式中,多聚结构域是蛋白质亚基的异多聚体。在一些实施方式中,多聚体是二聚体、三聚体、四聚体、五聚体、六聚体、八聚体十聚体或十二聚体。在一个优选实施方式中,pMHC多聚体是四聚体。

[0257] 合适的结合实体的实例是链霉亲和素(SA)和亲和素及其衍生物、生物素、免疫球蛋白、抗体(单克隆、多克隆和重组)、抗体片段及其衍生物、AP-1的亮氨酸拉链结构域(jun和fos)、六-his(金属螯合物部分)、六-帽GST(谷胱甘肽S-转移酶)谷胱甘肽亲和力、钙调素

结合肽 (CBP)、Strep-tag®、纤维素结合域、麦芽糖结合蛋白、S-肽标签、甲壳素结合标签、免疫反应表位、表位标签、E2标签、HA表位标签、Myc表位、FLAG表位、AU1和AU5表位、Glu-Glu表位、KT3表位、IRS表位、B标签表位、蛋白激酶C表位、VSV表位、介导与多种化合物(包括碳水化合物)结合的凝集素,脂质和蛋白质,例如Con A(白刀豆(*Canavalia ensiformis*))或WGA(小麦胚芽凝集素)和四联蛋白或蛋白A或G(抗体亲和力)或卷曲螺旋多肽,如亮氨酸拉链。还包括此类结合实体的组合。

[0258] 在一些实施方式中,多聚结构域是链霉亲和素(SA或SAv)或其衍生物的四聚体。在一些实施方式中,多聚结构域是四聚链霉亲和素。在一些实施方式中,四聚体包括Strep-tag®或Strep-tactin®。Strep-tag®或Strep-tactin®分别在美国专利号5,506,121和美国专利号6,103,493中描述,并可从多个来源商业获得。为了通过SAv的生物素结合位点将MHC单体非共价附着到链霉亲和素上,可以将avitag(例如具有SEQ ID NO:161中所示的氨基酸序列,其包括6xHis标签和FLAG标签)纳入MHC单体内,例如在C末端(参见例如实施例3)。

[0259] 在本文提供的方法中,pMHC多聚体是通过每个p\*MHC单体与多聚结构域的每个亚基的N或C末端共价偶联而产生的,从而产生本文中称为偶联多聚体的反应产物。在一个实施方式中,偶联多聚体是pMHC I类(pMHCI)偶联多聚体。在另一个实施方式中,偶联多聚体是pMHC II类(pMHCII)偶联多聚体。

[0260] 在一些实施方式中,pMHCI多聚体通过多聚结构域与MHCI $\alpha$ 1结构域的C末端共价偶联而产生。在一些实施方式中,pMHCI多聚体通过多聚结构域与MHCI $\alpha$ 2结构域的C末端共价偶联而产生。在一些实施方式中,pMHCI多聚体通过多聚结构域与MHCI $\alpha$ 3结构域的C末端共价偶联而产生。在一些实施方式中,pMHCI多聚体通过将多聚结构域共价偶联到每个p\*MHC单体的 $\beta$ 2-微球蛋白的C末端而产生。

[0261] 在一个优选实施方式中,pMHCII多聚体通过多聚结构域与MHCII $\alpha$ 链的共价偶联而产生。在另一个实施方式中,pMHCII多聚体通过多聚结构域与MHCII $\beta$ 链的共价偶联而产生。在某些实施方式中,pMHCII多聚体通过多聚结构域与MHCII $\alpha$ 1结构域的C末端共价偶联而产生。在某些实施方式中,pMHC II多聚体通过多聚结构域与MHC II $\alpha$ 2结构域的C末端共价偶联而产生。在某些实施方式中,pMHC II多聚体通过多聚结构域与MHC II $\beta$ 1结构域的C末端共价偶联而产生。在某些实施方式中,pMHC II多聚体通过多聚结构域与MHC II $\beta$ 2结构域的C末端共价偶联而产生。

[0262] 本文提供了在每个MHC单体和多聚结构域之间形成共价键的多种合适方法。

#### [0263] A. 化学生物偶联

[0264] 在一些实施方式中,p\*MHC多聚体通过化学偶联产生。在一些实施方式中,化学偶联通过p\*MHC多肽与多聚结构域的半胱氨酸生物偶联介导。在一些实施方式中,半胱氨酸生物偶联通过半胱氨酸烷基化介导。在一些实施方式中,半胱氨酸生物偶联通过半胱氨酸氧化介导。在其他实施方式中,半胱氨酸生物偶联通过脱硫反应介导。在一些实施方式中,半胱氨酸生物偶联由碘乙酰胺介导。在一些实施方式中,半胱氨酸生物偶联由马来酰亚胺介导。已经描述了利用半胱氨酸介导的两个部分的连接的方法,这两个部分可以用于产生本文公开的pMHC多聚体,例如,参见Chalker等人,Chem Asian J.4(5):630-402009;Spicer等人,Nat Commun.5:4740,2015。

[0265] 在一些实施方式中, MHC多聚体通过化学修饰除半胱氨酸以外的氨基酸而产生, 包括但不限于赖氨酸、酪氨酸、精氨酸、谷氨酸、天冬氨酸、丝氨酸、苏氨酸、甲硫氨酸、组氨酸和色氨酸侧链, 以及N末端胺或C末端羧基, 如前所述 (Baslé等人, *M Chem Biol.* 17 (3) : 213-272010; Hu等人, *Chem Soc Rev.* 45 (6) : 1691-719, 2016; Lin等人, *Science* 355 (6325) : 597-602, 2017)。

#### [0266] B. 天然化学连接

[0267] 在一些实施方式中, pMHC多聚体通过天然化学连接 (NCL) 产生, 其中每个p\*MHC多肽包括C末端硫酯, 并且多聚结构域的每个亚基包括N末端半胱氨酸残基或其功能等效物, 其中半胱氨酸侧链和硫酯之间的反应不可逆地形成天然肽键, 从而将p\*MHC单体连接到多聚结构域。已经描述了NCL的方法 (Hejjaoui等人, *M Protein Sci.* 24 (7) : 1087-99, 2015) Mandal等人, *Proc Natl Acad Sci USA* 109 (37) : 14779-84, 2012; Torbeev等人, *Proc Natl Acad Sci USA* 110 (50) : 20051-6, 2013)。

[0268] 在一些实施方式中,  $\beta$ -和/或 $\gamma$ -硫代氨基酸被掺入p\*MHC单体中。在一些实施方式中,  $\beta$ -和/或 $\gamma$ -硫代氨基酸取代多聚结构域的每个亚基的N末端位置的半胱氨酸样残基, 例如, 为反式硫酯化提供反应性硫醇。然后, 脱硫方法可以产生所需的天然侧链。在一些实施方式中, NCL在丙氨酸残基处进行。在其它实施方式中, 在苯丙氨酸 (Crich&Banerjee, 2007)、缬氨酸 (Chen等人2008; Haase等人2008)、亮氨酸 (Harpaz等人2010; Tan等人2010)、苏氨酸 (Chen等人2010b)、赖氨酸 (El Oualid等人2010; Kumar等人2009; Yang等人2009)、脯氨酸 (Shang等人2011)、谷氨酸 (Siman等人2012)、精氨酸 (Malins等人2013)、色氨酸 (Malins等人2014)、天冬氨酸 (Thompson等人2013)、谷氨酰胺 (Cergol等人2014) 和天冬酰胺 (Sayers等人2015) 处进行NCL。已经描述了去除纯化步骤并提高连接产物产率的连接/脱硫方法 (Moyal等人, 2013; Thompson等人, 2014)。

#### [0269] C. 点击化学介导的生物正交偶联

[0270] 在一些实施方式中, 通过每个p\*MHC单体的C末端偶联部分和多聚结构域每个亚基的N末端偶联部分之间的生物正交偶联来产生p\*MHC-多聚体。在一些实施方式中, 生物正交偶联由“点击化学”介导。(见例如Kolb, Finn and Sharpless, *Angewandte Chemie International Edition* (2001) 40: 2004-2021)。适用于点击化学、反应条件和相关方法的偶联部分在本领域中可用 (例如, Kolb等人, *Angewandte Chemie International Edition* 40: 2004-2021; Evans, *澳大利亚化学杂志* 60: 384-395, 2007; Lahann, *生物技术和材料科学点击化学*, John Wiley&Sons Ltd, ISBN 978-0-470-69970-6, 2009)。在一些实施方式中, 点击化学部分可以包括或由末端炔烃、叠氮化物、应变炔烃、二烯、亲二烯、烷氧基胺、羰基、膦、酰肼、硫醇或烯炔部分组成。在某些实施方式中, 叠氮化物是铜整合叠氮化物。在一个实施方式中, 铜整合叠氮化物是吡啶甲基叠氮化物, 例如Gly-Gly-Gly-(PEG)4-吡啶甲基-叠氮化物。用于点击化学反应的试剂是市售的, 例如来自点击化学工具 (Scottsdale, AZ) 或GenScript (Piscataway, NJ)。

[0271] 对于通过点击化学将每个p\*MHC单体偶联到多聚结构域的亚基, 蛋白质的点击化学部分必须彼此反应, 例如, 其中每个p\*MHC单体的点击化学部分之一的反应基团与多聚结构域的亚基上的第二点击化学部分的反应性基团反应以形成共价键。这种反应性成对点击化学处理对本领域技术人员来说是公知的, 并且包括但不限于图1所示的那些。

[0272] 在一些实施方式中,每个p\*MHC偶联部分可以在点击化学反应条件下与多聚结构域的每个亚基的偶联部分共价偶联。在一些实施方式中,使用分选酶介导的偶联来在每个p\*MHC单体的C末端装配第一点击化学部分,并在多聚结构域的每个亚基上装配第二点击化学部分反应。在本文提供的方法中,在点击化学条件下,将两个或多个含有第一点击化学部分的p\*MHC单体在多聚结构域的每个亚基的C末端与第二点击化学部分偶联。例如在W02013/00355中描述了利用分选酶连接点击化学部分的方法,其全部内容通过引用并入本文。实施例1、5、6和7中详细描述了使用炔烃-叠氮点击化学与分选酶介导的偶联组合制备的pMHC多聚体的非限制性实例。

[0273] 在一些实施方式中,使用内含蛋白介导的偶联来在每个p\*MHC单体的C末端装配第一点击化学部分,并在多聚结构域的每个亚基上装配第二点击化学部分反应。本文进一步描述了利用内含蛋白介导的偶联物的方法。

[0274] 在一些实施方式中,本文提供的点击化学介导的p\*MHC单体与多聚结构域的共价偶联的方法包括C末端硫酯与 $\beta$ -氨基硫醇的天然化学连接(Xiao J, Tolbert TJ *Org Lett.* 2009 Sep 17; 11 (18) : 4144-7)。

[0275] 在一些实施方式中,用于生产p\*MHC多聚体的点击化学包括1,3-偶极环加成(例如,Cu(I)催化的逐步变体,通常简称为“点击反应”;参见例如Tornoe等人,*Journal of Organic Chemistry* (2002) 67: 3057-3064)。铜和钌是反应中常用的催化剂。使用铜作为催化剂导致形成1,4-区域异构体,而钌导致形成1,5-区域异构体。

[0276] 在一些实施方式中,MHC单体通过表达蛋白连接(EPL)连接到炔基化肽上,然后通过Cu(I)催化的末端叠氮-炔环加成(CuAAC)与叠氮标记的多聚结构域偶联。

[0277] 在一些实施方式中,点击化学偶联包括环加成反应,例如Diels-Alder反应。在一些实施方式中,MHCI和多聚结构域通过叠氮-炔1,3-偶极环加成(“点击化学”)偶联。在一些实施方式中,通过Cu(I)催化的环加成(CuAAC)的存在促进环加成。

[0278] 在一些实施方式中,点击化学偶联包括对小应变环如环氧化物和氮丙啶的亲核加成。在一些实施方式中,通过应变环辛炔系统促进环加成,例如,如Agard NJ, Prescher JA, Bertozzi CR *J Am Chem Soc.* 2004年11月24日; 126 (46) : 15046-7。

[0279] 在一些实施方式中,点击化学偶联包括对活化羰基的亲核加成。

[0280] 在一些实施方式中,pMHC单体和多聚结构域的偶联通过生物正交反应进行。在一些实施方式中,MHC和多聚结构域通过应变亲二烯和四嗪二烯之间的反电子需求Diels-Alder反应偶联,例如,如Blackman ML, Royzen M, Fox JM *J Am Chem Soc.* 2008年10月15日; 130 (41) : 13518-9; 和Devaraj NK, Weissleder R, Hilderbrand SA *Bioconjug Chem.* 2008年12月; 19 (12) : 2297-9)。在一些实施方式中,亲二烯是反式环辛烯。在一些实施方式中,亲二烯是降冰片烯。

#### [0281] D. 分选酶介导的偶联

[0282] 在一些实施方式中,p\*MHC单体和多聚结构域之间的偶联由半胱氨酸转肽酶介导。在一些实施方式中,半胱氨酸转肽酶是分选酶或其酶活性片段。已经描述了多种分选酶并可商购(例如,Antos等人,*Curr. Opin. Struct. Biol.* 38: 111-118, 2016)。分类酶识别并切割一个氨基酸基序,称为“分选标签”,以在两个多肽上的酰基供体和受体位点之间产生肽键,从而连接含有N或C末端分选标签的不同多肽。实施例1、5、6和7中详细描述了使用分选酶介

导的偶联(与炔炔-叠氮点击化学组合)制备的pMHC多聚体的非限制性实例。

[0283] 因此,在一些实施方式中,每个p\*MHC单体包括C末端分选标签,并且多聚结构域的每个亚基包括N末端分选标签。在其他实施方式中,每个p\*MHC单体包括N末端分选标签,并且多聚结构域的每个亚基包括C末端分选标签。在一些实施方式中,分选酶催化MHC多肽与多聚结构域的每个亚基之间肽键的形成。

[0284] 在一些实施方式中,识别基序被添加到每一个pMHC单体的C末端,寡-甘氨酸基序被加入到多聚结构域的每一个亚基的N末端。在将分选酶添加到MHC单体和多聚结构域的混合物中后,多肽通过天然肽键共价连接以产生pMHC多聚体。

[0285] 在一些实施方式中,MHC单体和/或多聚结构域与分选标签框内表达。在一些实施方式中,可以包括额外的标签,例如,6x His标签(Sinisi等人Bioconag.Chem 23:1119-1126,2012)、亲核荧光染料(Nair等人Immun.Inflamm.Dis.1:3-13,2013)和/或FLAG标签(Greineder等人Bioconag.Chem.29:56-66,2018)。

[0286] 在一些实施方式中,分选标签包含适合于MHC单体和多聚结构域之间化学偶联的修饰氨基酸。在一些实施方式中,分选标签包含C末端叠氮赖氨酸残基,以实现如本文所述的定向点击化学偶联。

[0287] 在一些实施方式中,MHC多肽和/或多聚结构域包括多肽和分选标签之间的接头。在一些实施方式中,每个MHC多肽和多聚结构域的每个亚基包括带有接头的分选标签。例如,在Greineder等人Bioconjug.Chem.29:56-66,2018中描述了合适的接头。在一些实施方式中,接头是半刚性接头。在一些实施方式中,接头包括(SSSSG)<sub>2</sub>SAA(SEQ ID NO:182)。在一些实施方式中,接头包括(G)<sub>5</sub>(SEQ ID NO:183)。

[0288] 在一些实施方式中,分选标签包含荧光团修饰的赖氨酸残基,以促进反应进展和效率的测量。

[0289] 在一些实施方式中,分选酶是Ca<sup>2+</sup>依赖性的。在一些实施方式中,分选酶是Ca<sup>2+</sup>非依赖性的。

[0290] 在一些实施方式中,分选标签标记的MHC分子是可溶性HLA-A2分子(HLA-A\*02:01),其具有C末端分选标签和6xHis标签,例如具有SEQ ID NO:1所示的氨基酸序列。在一些实施方式中,分选标签标记的多聚结构域是具有C末端分选标签和6xHis标签的链霉亲和素分子,例如具有SEQ ID NO:3所示的氨基酸序列。在一些实施方式中,具有6xHis标签的分选标签标记具有SEQ ID NO:162所示的氨基酸序列。本领域已知各种其他分选标签序列,适合用于制备本发明的偶联多聚体,其非限制性实例将在下文进一步描述。

[0291] 在一些实施方式中,分选标签包含氨基酸序列LPXTG(SEQ ID NO:163),其中X是任何氨基酸,并且分选酶在基序内的苏氨酸和甘氨酸骨架之间切割。

[0292] 在一些实施方式中,分选酶识别包含选自IPKTG(SEQ ID NO:164)、MPXTG(SEQ ID NO:165)、LAETG(SEQ ID NO:166)、LPXAG(SESEQ ID NO:167)、LPESG(SEQ ID NO:168)、LPELG(SEQ ID NO:169)或LPEVG(SEQ ID NO:170)的氨基酸序列的分选标签。

[0293] 在一些实施方式中,分选酶是SrtAstaph突变体。在一些实施方式中,SrtAstaph突变体是F40,识别基序是XPKTG(SEQ ID NO:171)(Piotukh等人,J.Am.Chem.Soc.2011 133:17536-17539)。在一些实施方式中,SrtAstaph突变体是F40,识别基序是APKTG(SEQ ID NO:172)、DPKTG(SEQ ID NO:173)或SPKTG(SEQ ID NO:174)。

[0294] 在一些实施方式中,SrtAstaph突变体是SrtAstaph五突变体,识别基序是LPXTG (SEQ ID NO:163),其中X是任何氨基酸、LPEXG (SEQ ID NO:175),其中X是任何氨基酸或LAETG (SEQ ID NO:166)。在一些实施方式中,突变体是SrtAstaph五突变体,识别基序是LPEAG (SEQ ID NO:176),LPECG (SEQ ID NO:177)或LPESG (SEQ ID NO:168)。在一些实施方式中,SrtAstaph突变体是2A-9,识别基序是LAETG (SEQ ID NO:166)。在一些实施方式中,SrtAstaph突变体是4S-9,识别基序是LPEXG (SEQ ID NO:178),其中X=A,C或S)。

[0295] 在一些实施方式中,分选酶是野生型分选酶的可溶性片段。在一些实施方式中,分选酶是修饰的分选酶A的可溶性片段 (Mao H,Hart SA,Schink a,Pollok BA,J Am Chem Soc.2004年3月10日;126 (9):2670-1A)。

[0296] 在一些实施方式中,分选酶是金黄色葡萄球菌分选酶A的变体或同源物 (Antos JM,Truttmann MC,Ploegh HL Curr Opin Struct Biol.2016年6月;38:111-8;Dorr BM,Ham HO,An C,Chaikof EL,Liu DR Proc Natl Acad Sci U S A.2014年9月16;111 (37):13343-8;Glasgow JE,Salit ML,Cochran JR J Am Chem Soc.2016年6月22;138 (24):7496-9)。

[0297] 还描述了将分选标签偶联到蛋白质中的方法。(Matsumoto T,Furuta K,Tanaka T,Kondo A ACS Synth Biol.2016年11月18;5 (11):1284-1289;Williams FP,Milbradt AG,Embrey KJ,Bobby R PLoS One.2016;11 (4):e0154607;和Witte MD,Cragolini JJ,Dougan SK,Yoder NC,Popp MW,Ploegh HL Proc Natl Acad Sci U S A.2012年7月24;109 (30):11993-8;Mao H,Hart SA,Schink A,Pollok BA J Am Chem Soc.2004年3月10;126 (9):2670-1;Guimaraes CP,Witte MD,Theile CS,Bozkurt G,Kundrat L,Blom AE,Ploegh HL Nat Protoc.2013年9月;8 (9):1787-99和Theile CS,Witte MD,Blom AE,Kundrat L,Ploegh HL,Guimaraes CP Nat Protoc.2013年9月;8 (9):1800-7。

[0298] 在一些实施方式中,例如在反应进行时,通过透析或离心除去由分选酶反应产生的氨基甘氨酸肽片段 (Freiburger L,Sonntag M,Hennig J,Li J,Zou P,Sattler M J Biomol NMR.2015年9月;63 (1):1-8)。在一些实施方式中,亲和固定化策略或基于流动的平台用于选择性去除反应组分 (Policarpo RL,Kang H,Liao X,Rabideau AE,Simon MD,Pentelute BL Angew Chem Int Ed Engl.2014年8月25;53 (35):9203-8)。

[0299] 在一些实施方式中,反应的平衡可以通过连接产物或副产物失活来控制。例如,在一些实施方式中,通过连接添加到供体和受体的WTWTW (SEQ ID NO:179) 基序来控制反应,如Yamamura Y,Hirakawa H,Yamaguchi S,Nagamune T Chem Commun (Camb) 2011年4月28日;47 (16):4742-4中所述。在其他实施方式中,通过酰基供体甘氨酸的化学修饰使副产物失活,例如在Liu F,Luo EY,Flora DB,Mezo AR J Org Chem.2014年1月17日;79 (2):487-92;和Williamson DJ,Webb ME,Turnbull WB Nat Protoc.2014年2月;9 (2):253-62) 中所述。

#### [0300] E. 内含蛋白介导的偶联

[0301] 内含蛋白是天然存在的自剪接蛋白亚结构域,能够从更大的蛋白质结构中切除自身的蛋白亚结构,同时将两个先前侧翼肽区 (“外显蛋白”) 连接在一起,形成成熟的宿主蛋白。已经开发了基于内含蛋白的蛋白质修饰和连接方法。内含蛋白是一种能够催化蛋白质剪接反应的内部蛋白质序列,该剪接反应从前体蛋白质中切除内含蛋白序列并用肽键连接

侧翼序列(N-和C-外显蛋白)。实施例2中详细描述了使用内含蛋白介导的偶联制备的pMHC多聚体的非限制性实例。

[0302] 如本文所用,术语“裂解内含蛋白”是指任何内含蛋白,其中N末端内含蛋白区段和C末端内含蛋白区段之间存在一个或多个肽键断裂,使得N末端和C末端的内含蛋白区段成为独立的分子,可以经典(Cannon)-共价重结合或重建为具有剪接或切割反应功能的内含蛋白。任何催化活性内含蛋白或其片段可用于衍生用于本文公开的系统和方法的裂解内含蛋白。例如,在一个方面,裂解内含蛋白可以来源于真核内含蛋白。在另一个方面,裂解内含蛋白可以来源于细菌内含蛋白。在另一个方面,裂解内含蛋白可以来源于古细菌内含蛋白。优选地,如此衍生的裂解内含蛋白将仅具有催化剪接反应所必需的氨基酸序列。

[0303] 如本文所使用的,“N末端内含蛋白区段”是指任何含有N末端氨基酸序列的内含蛋白序列,当与相应的C末端内含蛋白区段结合时,该N末端氨基酸序列具有剪接和/或切割反应的功能。因此,N末端内含蛋白区段还包括在发生剪接时剪接出的序列。N末端内含蛋白片段可以包括天然存在(天然)内含蛋白序列的N末端部分的修饰序列。例如,N末端内含蛋白区段可以包含额外的氨基酸残基和/或突变残基,只要包含这样的额外和/或突变残基不会使内含蛋白对剪接或切割没有功能。优选地,额外和/或突变残基的纳入改善或增强了内含蛋白的剪接活性和/或可控性。非内含蛋白残基也可以与内含蛋白区段基因融合,以提供额外的功能,例如亲和纯化或共价固定的能力。

[0304] 如本文所使用的,“C末端内含蛋白区段”是指任何含有C末端氨基酸序列的内含蛋白序列,当与相应的N末端内含蛋白区段结合时,该C末端氨基酸序列具有剪接和/或切割反应的功能。在一个方面,C末端内含蛋白区段包括在发生剪接时剪接出的序列。在另一个方面,从与其C末端融合的肽序列中切割C末端内含蛋白区段。C末端内含蛋白片段可以包括天然存在(天然)内含蛋白序列的C末端部分的修饰序列。例如,C末端内含蛋白区段可以包含额外的氨基酸残基和/或突变残基,只要包含这样的额外和/或突变残基不会使C末端内含蛋白区段对剪接或切割没有功能。

[0305] 表达蛋白连接(EPL)是指重组蛋白与C末端硫酯和第二试剂与N末端半胱氨酸之间的天然化学连接。通过使用由内含蛋白(介入蛋白)介导的自加工(也称为蛋白质剪接),C末端硫酯可以很容易地被引入任何重组蛋白(即靶向配体)上。内含蛋白是一种蛋白质,可以从更大的前体多肽链中切除自身,利用一种过程在两侧的外显蛋白(外部蛋白)片段之间形成天然肽键。当将自加工蛋白克隆到靶向配体下游时,可以使用硫醇(例如2-巯基乙磺酸,MESNA)诱导自加工蛋白的位点特异性切割,从而形成反应性硫酯。然后,硫酯将与任何具有N末端半胱氨酸的试剂反应。EPL以位点特异性方式操作,如果两个官能团都为高浓度,则已知反应非常有效。(在Elias等人Small 6:2460-2468中综述)。

[0306] 因此,在一些实施方式中,MHC单体通过表达蛋白连接(EPL)连接到炔基化肽上,然后通过Cu(I)催化的末端叠氮-炔环加成(CuAAC)与叠氮标记的多聚结构域偶联。

[0307] 在一些实施方式中,MHC单体通过内含蛋白肽标签与多聚结构域偶联。在一些实施方式中,MHC多肽包括C末端硫酯,多聚结构域包括融合到修饰内含蛋白(缺乏进行酯交换的能力)的N-外显蛋白,并且通过添加外源硫醇发生酯交换。

[0308] 现在已公开了许多内含蛋白,包括但不限于MxeGyrA(Frutos等人(2010); Southworth等人(1999); SspDnaE(Shah等人(2012); Wu等人(1998); NpuDnaE(Shah等人



(2012); Vila-Perello等人(2013); AvaDnaE (David等人(2015); Shah等人(2012); Cfa (共有 DnaE裂解内含蛋白) (Stevens等人(2016)); gp41-1和gp41-8 (Carvajal-Vallejos等人(2012)); NrdJ-1 (Carvajal-Vallejos等人(2012)); IMPDH-1 (Carvajal-Vallejos et al) 和AceL-TerL (Thiel等人(2014))。表1总结了这些内含蛋白的性质和用途。

[0309] 表1:

内含蛋白	温度 (°C)	$t_{1/2}$ *
<i>MxeGyrA</i>	25	10 小时
<i>SspDnaE</i>	37	76 分钟
[0310] <i>NpuDnaE</i>	37	19 秒
<i>AvaDnaE</i>	37	23 秒
Cfa (共有 DnaE 裂解内含蛋白)	30	20 秒
gp41-1	45	4 秒
gp41-8	37	15 秒
[0311] NrdJ-1	37	7 秒
IMPDH-1	37	8 秒
AceL-TerL	8	7.2 分钟

[0312] 在一些实施方式中,该内含蛋白是来自蟾分枝杆菌 (*Mycobacterium xenopi*) 的 198残基促旋酶A内含蛋白 (*Mxe GyrA*) (Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB *Biotechniques*. 1999年7月; 27 (1) : 110-41116118-20)。在一些实施方式中,内含蛋白来自蓝藻集胞藻种 (*Synechocystis sp.*) PCC6803株 (*Ssp*)。

[0313] 在一些实施方式中,内含蛋白是裂解内含蛋白对。在一些实施方式中,裂解内含蛋白对是正交裂解内含蛋白对 (Carvajal Vallejos P, Pallissé R, Mootz HD, Schmidt SR *J Biol Chem*. 2012年8月17日; 287 (34) : 28686-96; Shah NH, Vila Perelló M, Muir TW *Angew Chem Int Ed Engl*. 2011年7月11日; 50 (29) : 6511-5)。

[0314] 在一些实施方式中,裂解内含蛋白对是短至6或11个残基的人工裂解内含蛋白对 (Appleby JH, Zhou K, Volkmann G, Liu XQ *J Biol Chem*. 2009年3月6日; 284 (10) : 6194-9; Ludwig C, Pfeiff M, Linne U, Mootz HD *Angew Chem Int Ed Engl*. 2006年8月4日; 45

(31):5218-21)。

[0315] 在一些实施方式中,内含蛋白是DnaE内含蛋白。在一些实施方式中,DnaE内含蛋白来自点形念珠藻(*Nostoc punctiforme*) (Npu)。在一些实施方式中,内含蛋白是gp41-1内含蛋白。在一些实施方式中,内含蛋白是gp41-8内含蛋白。在一些实施方式中,内含蛋白是IMPDH-1内含蛋白。在一些实施方式中,内含蛋白是NrdJ内含蛋白。

[0316] 在一些实施方式中,裂解内含蛋白对是AceL-TerL(Thiel IV,Volkman G,Pietrokovski S,Mootz HD *Angew Chem Int Ed Engl.*2014年1月27日;53(5):1306-10)。

[0317] 在一些实施方式中,该内含蛋白包括共有裂解内含蛋白序列(Cfa)(Stevens AJ,Brown ZZ,Shah NH,Sekar G,Cowburn D,Muir TW.*Design of a split intein with exceptional protein splicing activity.**Journal of the American Chemical Society.*2016;138(7):2162-2165)。

[0318] 可获得许多用于内含蛋白介导偶联的方案,并且本文在实施例2中提供了示例性方法。本领域已经描述了用于蛋白质偶联的合适内含蛋白序列和方案,例如Stevens等人,*J. Am. Chem. Soc.*,138,2162-2165,2016;Shah等人,*J. Am. Chem. Soc.*,134,11338-11341,2012;和Vila Perello等人,*J. Am. Chem. Soc.*,135,286-292,2013;Batjargal S,Walters CR,Peterson EJ *J Am Chem Soc.*2015年2月11日;137(5):1734-7;以及Guan D,Ramirez M,Chen Z *Biotechnol Bioeng.*2013年9月;110(9):2471-81,其全部内容通过引用并入本文。

[0319] 在一些实施方式中,内含蛋白标记的MHC分子是具有N-内含蛋白标签的可溶性HLA-A2分子(HLA-A\*02:01),例如具有SEQ ID NO:4所示的氨基酸序列。在一些实施方式中,内含蛋白标记的多聚结构域是具有C末端内含蛋白标签和FLAG标签的链霉亲和素分子,例如具有SEQ ID NO:5所示的氨基酸序列。在一些实施方式中,包括FLAG标签的N-内含蛋白标签具有SEQ ID NO:180中所示的氨基酸序列。本领域已知各种其他N-内含蛋白和C-内含蛋白序列,适合用于制备本公开的偶联多聚体,其非限制性实例在上文引用的参考文献中描述。

#### [0320] F. 其它生物偶联方法

[0321] 在一些实施方式中,MHC和多聚结构域的偶联是酶介导的。在一些实施方式中,该酶是甲酰甘氨酸生成酶(FGE),其识别CXPXR氨基酸序列基序并将半胱氨酸残基转化为甲酰基甘氨酸,从而引入醛官能团(Wu P,Shui W,Carlson BL,Hu N,Rabuka D,Lee J,Bertozzi CR *Proc Natl Acad Sci U S A.*2009年3月3日;106(9):3000-5),其经历生物正交转化,例如肟化和肟基-Pictet-Spengler反应(Agarwal P,Kudirka R,Albers AE,Barfield RM,de Hart GW,Drake PM,Jones LC,Rabuka D *Bioconjug Chem.*2013年6月19日;24(6):846-51;Dirksen A,Dawson PE *Bioconjug Chem.*2008年12月;19(12):2543-8)。

[0322] 位点特异性生物结合策略为定向蛋白质修饰提供了许多可能性。在各种基于酶的偶联方案中,甲酰基甘氨酸生成酶允许将氨基酸C $\alpha$ -甲酰甘氨酸(FGly)翻译后引入重组蛋白,从不同共有基序内的半胱氨酸或丝氨酸残基开始。含醛FGly残基显示出与所有其他天然氨基酸的正交反应性,因此可用于蛋白质支架上的位点特异性标记反应。(在Kruger等人,*Biol Chem.*2019年2月25日;400(3) doi:10.1515/hsz-2018-0358综述)

[0323] 甲酰甘氨酸生成酶(FGE)识别五肽共有序列CxPxR,并将该序列中的半胱氨酸特异

性氧化为一种不寻常的含醛甲酰甘氨酸。FGE识别序列或醛标签可以插入原核或真核表达系统中产生的异源重组蛋白中。半胱氨酸向甲酰甘氨酸的转化是通过FGE的瞬时或稳定细胞系的共-过度表达实现的,所得醛可以选择性地与 $\alpha$ -亲核试剂反应,生成位点选择性修饰的生物偶联物(Rabuka等人,Nat Protoc.2012年5月10日;7(6):1052-1067)。

[0324] 在一些实施方式中,该酶是硫辛酸连接酶,一种修饰13个残基靶序列内赖氨酸侧链(Uttamapinant C,White KA,Baruah H,Thompson S,Fernández-Suárez M,Puthenveetil S,Ting AY Proc Natl Acad Sci U S A.2010年6月15日;107(24):10914-9)以引入生物正交基团的酶,包括叠氮化合物、芳族醛和肼、对-碘苯基衍生物,降冰片烯和反式环辛烯(综述于Debelouchina等人,Q.Rev Biophys.2017;50e7.doi:10.1017/S003358351700021)。

[0325] 在其他实施方式中,该酶是生物素连接酶、法尼基转移酶、转谷氨酰胺酶或N-肉豆蔻酰基转移酶(综述于Rashidian M,Dozier JK,Distefano MD Bioconjug Chem.2013年8月21日;24(8):1277-94)。

#### [0326] G. 肽接头

[0327] 在其他实施方式中,p\*MHC多聚体包括肽接头。术语“肽接头”表示天然和/或合成来源的线性氨基酸链。接头具有确保相互偶联的多肽能够通过允许多肽正确折叠和正确递呈而发挥其生物活性的功能。肽接头可包含重复的氨基酸序列或天然存在的多肽序列。在一些实施方式中,肽接头具有2至50个氨基酸的长度。在一些实施方式中,肽接头为3至30个氨基酸、5至25个氨基酸、5至20个氨基酸或10至20个氨基酸。

[0328] 在一些实施方式中,肽接头富含甘氨酸、谷氨酰胺和/或丝氨酸残基。这些残基例如以多达五个氨基酸的小重复单元排列。这个小重复单元可以重复一到五次。在多聚体单元的氨基和/或羧基端可添加至多六个额外的任意天然存在的氨基酸。其他合成肽接头由单个氨基酸组成,该氨基酸重复10至20次,并且在氨基和/或羧基端可包含多达六个额外的任意天然存在的氨基酸。所有肽接头都可以由核酸分子编码,因此可以重组表达。由于接头本身是肽,由接头连接的多肽通过两个氨基酸之间形成的肽键与接头连接。

[0329] 合适的肽接头在本领域中是公知的,并且公开于例如US2010/0210511US2010/0179094和US2012/0094909中,其通过引用整体并入本文。其它接头在以下文献中提供:美国专利号5,525,491;Alfthan等人,Protein Eng.,1995,8:725-731;Shan等人,J.Immunol.,1999,162:6589-6595;Newton等人,Biochemistry,1996,35:545-553;Megeed等人;Biomacromolecules,2006,7:999-1004;和Perisic等人,Structure,1994,12:1217-1226;其各自通过引用出于所有目的将其全部内容纳入本文。

[0330] 在一些实施方式中,多肽接头是合成的。如本文所用,关于多肽接头的术语“合成的”包括肽(或多肽),其包含氨基酸序列(其可以是或可以不是天然存在的),该氨基酸序列以氨基酸的线性序列连接到其在自然界中非天然连接的序列(其可能是或可能不是天然存在的)。例如,多肽接头可以包括非天然多肽,其是天然多肽的修饰形式(例如包括诸如添加、取代或缺失的突变)或包含第一氨基酸序列(其可以是或可以不是天然存在的)。例如,可以使用多肽接头以确保每个多聚融合多肽的结合部分(TCR或MHC)、多聚结构域和Igg-框架并列,以确保功能性多聚蛋白复合物正确折叠和形成。优选地,多肽接头将是相对非免疫原性的,并且不会抑制结合蛋白的单体亚基之间的任何非共价结合。

[0331] 在一些实施方式中,接头是Gly-Ser多肽接头,即由甘氨酸和丝氨酸残基组成的肽。一种示例性Gly-Ser多肽接头包含氨基酸序列(Gly4Ser)<sub>n</sub>,其中n=1-6(SEQ ID NO:181)。在某些实施方式中,n=1。在某些实施方式中,n=2。在某些实施方式中,n=3。在某些实施方式中,n=4。在某些实施方式中,n=5。在某些实施方式中,n=6。其它示例性Gly-Ser多肽接头包含氨基酸序列Ser(Gly4Ser)<sub>n</sub>,其中n=1-10(SEQ ID NO:184)。在某些实施方式中,n=1。在某些实施方式中,n=2。在某些实施方式中,n=3,即Ser(Gly4Ser)<sub>3</sub>。在某些实施方式中,n=4,即Ser(Gly4Ser)<sub>4</sub>。在某些实施方式中,n=5。在某些实施方式中,n=6。在某些实施方式中,n=7。在某些实施方式中,n=8。在某些实施方式中,n=9。在某些实施方式中,n=10。

[0332] 其他示例性接头包括GS接头(即,(GS)<sub>n</sub>)、GGSG接头(即(GGSG)<sub>n</sub>) (SEQ ID NO:185)、GSAT接头(SEQ ID NO:186)、SEG接头和GGS接头(即(GGSGS)<sub>n</sub>) (SEQ ID NO:187),其中n是正整数(例如,1、2、3、4或5)。用于多聚融合蛋白的其他合适接头可以使用公开可用的数据库找到,例如接头数据库([ibi.vu.nl/programs/linkerdbwww](http://ibi.vu.nl/programs/linkerdbwww))。接头数据库是多功能酶中的结构域间接头的数据库,其用作新型多聚体融合蛋白中的潜在接头(参见例如,George等人,Protein Engineering 2002;15:871-9)。

[0333] 可以使用本领域已知的技术将多肽接头引入多肽序列中。修饰可以通过DNA序列分析来确认。质粒DNA可用于转化宿主细胞以稳定所产生的多肽的生产。

#### [0334] H. 其它肽接头和标签

[0335] 适用于本文提供的方法和组合物的其他标签包括亲和标签,包括但不限于酶、蛋白质结构域或小多肽,其以高特异性结合到一系列底物,例如碳水化合物、小生物分子、金属螯合物、抗体等,以允许蛋白质的快速和有效纯化。溶解度标签增强蛋白质的正确折叠和溶解度,并经常与亲和性标签串联使用。

[0336] 包括但不限于6x His、FLAG、Strep II和钙调素结合肽(CBP)标签的小尺寸标签具有使对MHC多肽的结构、活性和特性的影响最小化的益处。(Zhao等人J.Anal.Chem.2013 581093)

[0337] 在一些实施方式中,标签是FLAG标签。FLAG标签是一种亲水性八肽表位标签,其特异性结合具有不同识别和结合特性的数种特异性抗-FLAG单克隆抗体,如M1、M2和M5(Einhauer等人J.Biochem.Biophys.49:455-465,2001:Hopp等人Mol.Immunol.33:601-608,1996)。FLAG融合蛋白可以被具有钙依赖性(例如M2)或钙非依赖性方式的单克隆抗体识别。特别是,附着在融合蛋白N末端的标签对于用M1单克隆抗体进行免疫亲和纯化是必需的,而M2是位置不敏感的。

#### [0338] IV. MHC肽表位

##### [0339] A. 肽表位选择

[0340] 已经开发了各种方法来鉴定可能是T细胞表位的新的MHC结合肽,并且许多实验方法开始于从给定的蛋白质序列构建肽片段的重叠文库,通过合成恒定长度(n-聚体)氨基酸序列,这些氨基酸序列沿蛋白质序列彼此偏移固定数量的氨基酸。然后,可以在许多试验中评估每个序列的MHC结合特性和活化T细胞的潜力。

[0341] 已经用上述方法和其他方法鉴定的现有MHC结合肽(例如MHC结合沟构象和电荷分布的晶体学分析)导致对最常见的MHC等位基因定义了结合基序,为什么类型的推定MHC结

合肽实际上可以很好地结合给定等位基因的MHC分子制定了规则。这些基序已被转化为预测肽与MHC分子结合的预测计算机算法,例如SYFPEITHI算法(Rammensee H.-G.等人(1995),*Immunogenetics* 41:178-228)。

[0342] 通过使用SYFPEITHI (Rammensee等人*Immunogenetics* 50:213-2191999)和来自IEDB的人工神经网络(ANN)和稳定矩阵法(SMM)算法(Peters等人*PLoS Biol.*3:e912005),分析所需抗原的蛋白质序列以寻找可能的HLA特异性抗原。根据SYFPEITHY的预测结合值>21、ANN的预测结合价值<6000或SMM的预测结合数值<600来选择肽。合成所选择的肽。

[0343] 可以使用如前所述的荧光偏振(FP)分析进行结合测定(例如,Buchi等人,*Biochemistry* 43:14852-14863,2004;Sette等人,*Mol. Immunol.*31:813-822)。为了确定肽的结合能力,可以在FP竞争试验中用占位符肽确定相对于对照的抑制百分数。

[0344] 在一些实施方式中,与pMHC多聚体结合的肽来自无偏肽文库。在一些实施方式中,肽是9聚体。在一些实施方式中,与pMHCI多聚体结合的肽是9聚体,其包括在位置2和9处具有关键氨基酸的HLA-A2结合基序,其可包括异亮氨酸(I)、缬氨酸(V)或亮氨酸(L)。

[0345] 在一些实施方式中,文库包含通过转录和翻译任何感兴趣的多核苷酸序列而产生的所有k-聚肽,例如,在所有六个阅读框中计算机产生基因组或宏基因组正向和反向链的转录和翻译产物。

[0346] 在一些实施方式中,本公开的文库包括可从感兴趣外显子的计算机翻译衍生的所有k-聚肽。

[0347] 在一些实施方式中,本公开的文库包括可从感兴趣转录组的计算机翻译衍生的所有k-聚肽。

[0348] 在一些实施方式中,本发明的文库包含可从感兴趣蛋白质组中衍生的所有k-聚肽。

[0349] 在一些实施方式中,本公开的文库包括可从感兴趣ORF组的计算机翻译衍生的所有k-聚肽。

[0350] 在一些实施方式中,可以使用算法来选择肽文库中的肽。例如,可以使用算法来预测最有可能折叠或停靠在MHC/HLA结合口袋中的肽,并且可以选择高于特定阈值的肽包含在文库中。

[0351] 在一些实施方式中,本发明的文库包含可从一组基因组、蛋白质组、转录组、ORF组或其任何组合的计算机转录和翻译或翻译中衍生的所有肽。

[0352] 在一些实施方式中,肽来源于来自一组样品(例如来自患者群体的临床样品或一组病原体基因组)的多核苷酸序列的计算机转录和翻译或翻译。

[0353] 在一些实施方式中,所述肽衍生自差异基因组、蛋白质组、转录组、ORF组或其任何组合,其中比较两个或多个基因组、蛋白质组、转录组或ORF组或其组合以识别作为差异序列的序列(例如,它们之间不同的序列)。在一些实施方式中,通过比较感兴趣的组织来鉴定肽序列。在一些实施方式中,通过比较感兴趣的细胞来鉴定肽序列。在一些实施方式中,通过比较患病细胞或组织与健康细胞或组织来鉴定肽序列。在一些实施方式中,患病的细胞或组织是癌细胞或组织。在一些实施方式中,患病细胞来源于患有自身免疫病的个体。

[0354] 在一些实施方式中,肽衍生自基因组、蛋白质组、转录组、ORF组或其任何组合的同源序列,其中比较两个或多个基因组、蛋白质、转录组和ORF组或其组合以鉴定同源序列。

[0355] 在一些实施方式中,所述肽衍生自感兴趣序列中的突变,例如,可由编码抗原或表位的多核苷酸序列中单核苷酸突变产生的所有9-聚肽。

[0356] 在一些实施方式中,所述肽是重叠肽文库,其包括来自模板序列(例如,在计算机翻译的基因组中)的重叠肽,其中设定长度的重叠肽偏离限定数量的残基。

[0357] 在一些实施方式中,肽的选择包括基于特定HLA类型的预测结合亲和力对肽进行优先级排序。

[0358] 在一些实施方式中,为本发明文库选择肽基于群体(例如,人类人群)中的流行程度优先考虑HLA类型或等位基因。

[0359] 在一些实施方式中,文库包含通过转录和翻译任何感兴趣的多核苷酸序列而产生的所有k-聚肽,例如,在所有六个阅读框中计算机产生基因组或宏基因组正向和反向链的转录和翻译产物。在一些实施方式中,本发明的文库包含可从哺乳动物基因组(例如,小鼠基因组、人类基因组、患者基因组、自身免疫患者基因组或癌症基因组)计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从微生物基因组(例如,细菌基因组、病毒基因组、原生动物基因组、原生生物基因组、酵母基因组、古细菌基因组或噬菌体基因组)计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从病原体基因组,例如,细菌病原体基因组、病毒病原体基因组、真菌病原体基因组、机会致病病原体基因组、条件病原体基因组或真核寄生虫基因组计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库可来自植物基因组或真菌基因组。在一些实施方式中,本发明的文库包含衍生自基因组的计算机转录和翻译的k-聚肽,其中基因组在计算机转录和翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0360] 在一些实施方式中,本发明的文库包含可从感兴趣的外显子组(例如,哺乳动物外显子组、人类外显子组、小鼠外显子组、患者外显子组、自身免疫患者外显子组、癌症外显子组、病毒外显子组、原生动物外显子组、原生生物外显子组、酵母外显子组、病原体外显子组、真核寄生虫外显子组、植物外显子组或真菌外显子组)的计算机翻译中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含衍生自外显子组的计算机翻译的k-聚肽,其中外显子组在计算机翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0361] 在一些实施方式中,本发明的文库包含可从感兴趣的转录组(例如,哺乳动物转录组、人类转录组、小鼠转录组、患者转录组、自身免疫患者转录组、癌症转录组、微生物转录组、细菌转录组、病毒转录组、原生动物转录组、原生生物转录组、酵母转录组、古细菌转录组、噬菌体转录组、病原体转录组、真核寄生虫转录组、植物转录组、真菌转录组,来自RNA测序的转录组、微生物组转录组或来自宏基因组RNA测序的转录组)的计算机翻译中衍生的所有k-聚肽。在一些实施例中,本发明的文库包含衍生自转录组的计算机翻译的k-聚肽,其中转录组在计算机翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0362] 在一些实施方式中,本发明的文库包含可从感兴趣的蛋白质组(例如,哺乳动物蛋白质组、人类蛋白质组、小鼠蛋白质组、患者蛋白质组、自身免疫患者蛋白质组、癌症蛋白质组、微生物蛋白质组、细菌蛋白质组、病毒蛋白质组、原生动物蛋白质组、原生生物蛋白质

组、酵母蛋白质组、古细菌蛋白质组、噬菌体蛋白质组、病原体蛋白质组、真核寄生虫蛋白质组、植物蛋白质组或真菌蛋白质组) 衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含源自蛋白质组的k-聚肽, 其中k-聚肽从蛋白质组序列修饰而来, 例如, 包含突变(例如, 替换、插入、删除)的k-聚肽。

[0363] 在一些实施方式中, 本发明的文库包含可从感兴趣的ORF组(例如, 哺乳动物ORF组、人类ORF组、小鼠ORF组、患者ORF组、自身免疫患者ORF组、癌症ORF组、微生物ORF组、细菌ORF组、病毒ORF组、原生动物ORF组、原生生物ORF组、酵母ORF组、古细菌ORF组、噬菌体ORF组、病原体ORF组、真核寄生虫ORF组、植物ORF组、真菌ORF组, 来自RNA测序的ORF组、微生物组ORF组或来自宏基因组RNA测序的ORF组)的计算机翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含衍生自ORF组的计算机翻译的k-聚肽, 其中ORF组在计算机翻译期间被修饰, 例如, 计算机突变以产生包含突变(例如, 替换、插入、删除)的k-聚肽。

[0364] 在一些实施方式中, 本发明的文库包含可从一组基因组、蛋白质组、转录组、ORF组或其任何组合的计算机转录和翻译或翻译中衍生的所有k-聚肽。在一些实施例中, 本发明的文库包含可从一组样品(例如, 来自患者群体的临床样品或一组病原体基因组)的多核苷酸序列的计算机转录和翻译或翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含可从一组病毒基因组, 例如人病毒组的计算机转录和翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含衍生自一组基因组、蛋白质组、转录组、ORF组或其任意组合的计算机转录和翻译的所有k-聚肽, 其中来源序列在计算机翻译期间被修饰, 例如, 计算机突变以产生包含突变(例如, 替换、插入、删除)的k-聚肽。

[0365] 在一些实施方式中, 本发明的文库包含可从差异基因组、蛋白质组、转录组、ORF组或其任何组合衍生的所有k-聚肽, 其中比较两个或更多基因组、蛋白质组、转录组、ORF组或其组合以鉴定为差异序列(在其之间区分)的序列(例如, 核苷酸序列、氨基酸序列、核苷酸丰度或蛋白质丰度不同)。在一些实施方式中, 通过比较感兴趣的组织来生成基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较来自感兴趣细胞(例如, 健康细胞与癌细胞)的序列来产生基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较感兴趣生物体的序列来生成基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较感兴趣的对象(例如患病对健康对象)来生成基因组、蛋白质组、转录组或ORF组的差异序列。

[0366] 在一些实施方式中, 本发明的文库包含可从基因组、蛋白质组、转录组、ORF组或其任何组合的同源序列衍生的所有k-聚肽, 其中比较两个或更多基因组、蛋白质组、转录组、ORF组, 或其组合以鉴定同源序列(例如, 具有一定程度的同源性)的序列, 例如同源核苷酸序列、同源氨基酸序列、同源核苷酸丰度或同源蛋白质丰度。在一些实施方式中, 通过比较感兴趣的组织来生成基因组、蛋白质组、转录组或ORF组的同源序列。在一些实施方式中, 基因组、蛋白质组、转录组或ORF组的同源序列是通过比较来自感兴趣的细胞(例如, 健康细胞与涉及自身免疫的细胞(例如, 诱导自身免疫的细胞或在自身免疫期间被靶向的细胞)的序列来生成的。在一些实施方式中, 通过比较感兴趣的生物体的序列来生成基因组、蛋白质组、转录组或ORF组的同源序列。在一些实施方式中, 通过比较感兴趣的对象(例如患病对健康对象)来生成基因组、蛋白质组、转录组或ORF组的同源序列。

[0367] 在一些实施方式中, 本发明的文库包含可从感兴趣的多肽序列衍生的所有k-聚

肽,例如,覆盖病毒蛋白质的完整蛋白质序列的所有可能的9-聚肽。在一些实施方式中,本发明的文库包含可从感兴趣的多肽序列产生的k-聚肽,其中感兴趣的多肽序列被修饰,例如在计算机中突变以生成包含突变(例如,取代、插入、删除)的k-聚肽。

[0368] 在一些实施方式中,本发明的文库包含可从感兴趣序列中的突变衍生的所有k-聚肽,例如,可从编码抗原或表位的多核苷酸序列中的单核苷酸突变生成的所有9-聚肽。例如,本发明的文库包含可由编码抗原或表位的多核苷酸序列中的两个、三个、四个、五个、六个、七个、八个或九个核苷酸突变产生的所有9-聚肽。在一些实施方式中,本发明的文库包含可从丙氨酸取代衍生的所有k-聚肽,例如,本文所述序列中任何位置的丙氨酸取代(例如,蛋白质、一组蛋白质、蛋白质组、计算机转录和翻译的基因组)。在一些实施方式中,本发明的文库包括位置扫描文库,其中所选氨基酸残基依次被所有其他天然氨基酸取代。在一些实施方式中,本发明的文库包括组合位置扫描文库,其中所选氨基酸残基每次在两个或多个位置依次被所有其他天然氨基酸取代。在一些实施方式中,本发明的文库包含重叠肽文库,其包含来自模板序列(例如,在计算机中翻译的基因组)的重叠肽,其中设定长度的重叠肽被确定数量的残基抵消。在一些实施方式中,本发明的文库包含T细胞截短肽文库,其中文库的每个复制物包含在一末端截短的等摩尔肽混合物(例如,可从标称11-聚体的C末端截短衍生8-聚体、9-聚体、10-聚体和11-聚体)。在一些实施方式中,本发明的文库包含定制肽组,其中定制肽组以列表形式提供。

[0369] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是病毒基因组、外显子组、转录组、蛋白质组或ORF组。病毒的非限制性实例包括腺病毒、腺相关病毒、Aichi病毒、澳大利亚蝙蝠丽沙病毒、BK多瘤病毒、版纳(Banna)病毒、巴马森林病毒、布尼亚韦拉病毒、拉克罗斯本扬病毒(Bunyavirus La Crosse)、雪鞋兔布尼亚病毒(Bunyavirus snowshoe hare)、猕猴疱疹病毒(Cercopithecine herpesvirus)、金迪普拉病毒(Chandipura virus)、基孔肯亚病毒(Chikungunya virus)、柯萨病毒(Cosavirus) A、牛痘病毒、柯萨奇病毒(Coxsackievirus)、克里米亚-刚果出血热病毒、巨细胞病毒(CMV)、登革热病毒、多理病毒(Dhori virus)、杜比病毒(Dugbe Virus)、杜文黑基病毒(Duvenhage Virus)、东部马脑炎病毒、埃博拉病毒、埃博拉病毒、埃可病毒(Echovirus)、脑心肌炎病毒、爱泼斯坦-巴尔病毒(EBV)、欧洲蝙蝠丽沙病毒、GB病毒C/庚型肝炎病毒、汉他病毒(Hantaan virus)、亨德拉病毒(Hendra virus)、甲型肝炎病毒、乙型肝炎病毒、丙型肝炎病毒、戊型肝炎病毒、丁型肝炎病毒、马痘病毒、人腺病毒、人星状病毒、人冠状病毒、人巨细胞病毒、人内源性反转录病毒(HERV)、人肠病毒、人疱疹病毒(例如,HHV-1、HHV-2、HHV-6A、HHV-6B、HHV-7、HHV-8)、人免疫缺陷病毒(例如,HIV-1、HIV-2)、人乳头瘤病毒(如HPV-1、HPV-2、HPV-16、HPV-18)、人副流感病毒、人细小病毒B19、人呼吸道合胞病毒(RSV)、人鼻病毒、人SARS冠状病毒、人泡沫反转录病毒(spuma retrovirus)、人类嗜T淋巴细胞病毒(HTLV,如HTLV-1、HTLV-2、HTLV-3)、人 $\tau$ 病毒、甲型流感病毒、乙型流感病毒、丙型流感病毒、伊斯法罕病毒(Isfahan virus)、JC多瘤病毒、日本脑炎病毒、胡宁沙状病毒(Juin arenavirus)、KI多瘤病毒、昆津病毒、拉各斯-蝙蝠病毒、维多利亚湖-马尔堡病毒、兰加特病毒(Langat virus)、拉沙病毒(Lassa Virus)、Lordsdale病毒、跳跃病病毒(Louping ill virus)、淋巴细胞性脉络丛脑膜炎病毒(Lymphocytic choriomeningitis virus)、马丘波病毒(Machupo virus)、马亚罗病毒(Mayaro virus)、MERS冠状病毒、麻疹病毒、门戈脑心肌炎病毒(Mengo



encephalomyocarditis virus)、梅克尔细胞多瘤病毒(Merkel cell polyomavirus)、蒙古拉病毒(Mokola virus)、传染性软疣病毒(Molluscum contagiosum virus)、猴痘病毒、腮腺炎病毒、墨莱溪谷脑炎病毒(Murray valley encephalitis virus)、纽约病毒、尼帕病毒(Nipah virus)、诺如病毒(Norovirus)、诺沃克病毒、阿尼昂-尼昂病毒(O'nyong-nyong virus)、羊接触传染性化脓性口炎病毒(Orf virus)、奥罗普切病毒(Oropouche virus)、皮钦德病毒(Pichinde virus)、脊髓灰质炎病毒、蓬塔托罗白蛉病毒(Punta toro phlebovirus)、普马拉病毒(Puumala virus)、狂犬病病毒、裂谷热病毒、罗沙病毒A(Rosavirus A)、罗斯河病毒(Ross river virus)、轮状病毒(例如,轮状病毒A、轮状病毒B、轮状病毒C、轮状病毒X)、风疹病毒(Rubella virus)、鹭山病毒(Sagiyama virus)、赛利病毒(Salivirus)A、西西里白蛉热病毒(Sandfly fever sicilian virus)、札幌病毒、塞姆利基森林病毒(Semliki forest virus)、汉城病毒、猴泡沫病毒(Simian foamy virus)、猴病毒5、辛德比斯病毒(Sindbis virus)、南安普敦病毒、圣路易斯脑炎病毒、蜱传波瓦桑病毒(Tick-borne powassan virus)、细环病毒(Torque teno virus)、托斯卡纳病毒、乌库涅米病毒(Uukuniemi virus)、痘苗病毒、水痘-带状疱疹病毒(Varicella-zoster virus)、天花病毒(Variola virus)、委内瑞拉马脑炎病毒、水泡性口炎病毒(Vesicular stomatitis virus)、西部马脑炎病毒、WU多瘤病毒、西尼罗河病毒、亚巴猴肿瘤病毒(Yaba monkey tumor virus)、亚巴样病病毒、黄热病病毒、和寨卡病毒。

[0370] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是癌基因组、外显子组、转录组、蛋白质组或ORF组。在一些实施方式中,本发明的文库包含已知的癌症新表位。在一些实施方式中,本发明的文库包含可从已知癌症抗原蛋白中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从参与上皮间质转化的基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从癌症相关基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从突变性癌症驱动基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从原癌基因、癌基因或肿瘤抑制基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从原癌基因、癌基因或肿瘤抑制基因中衍生的所有k-聚肽,其中k-聚肽包含如本文所述的突变(例如,氨基酸取代、丙氨酸取代、位置扫描、组合位置扫描等)。

[0371] 癌症的非限制性示例包括急性淋巴细胞白血病(ALL)、急性髓细胞白血病(AML)、肾上腺皮质癌、艾滋病相关癌、艾滋病相关淋巴瘤、肛门癌、阑尾癌、星状细胞瘤、非典型畸胎瘤样/横纹肌样瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、大脑瘤、乳腺癌、支气管肿瘤、伯基特氏淋巴瘤、类癌肿瘤、原发部位不明转移癌(Carcinoma of Unknown Primary)、心脏肿瘤、中枢神经系统癌、宫颈癌、胆管癌(Cholangiocarcinoma)、脊索瘤、慢性淋巴细胞白血病(CLL)、慢性粒细胞白血病(CML)、慢性骨髓增生性肿瘤、结直肠癌、颅咽管瘤(Craniopharyngioma)、皮肤T细胞淋巴瘤、原位乳腺导管癌(Ductal Carcinoma In Situ)、胚芽肿瘤(Embryonal Tumor)、子宫内膜癌、上皮癌、室管膜瘤、食管癌、成感觉神经细胞瘤(Esthesioneuroblastoma)、尤因氏肉瘤、颅外生殖细胞瘤、性腺外生殖细胞瘤、眼癌、输卵管癌、骨纤维组织细胞瘤、胆囊癌、胃癌、胃肠道类癌肿瘤、胃肠道间质瘤(GIST)、生殖细胞瘤、妊娠滋养细胞疾病(Gestational Trophoblastic Disease)、毛细胞白血病、头颈癌、肝细胞癌、组织细胞增多病、霍奇金氏淋巴瘤、下咽癌、眼内黑色素瘤、胰岛细胞瘤、卡波济氏

肉瘤、肾(肾细胞)癌、朗格汉斯细胞组织细胞增生症、喉癌、白血病、唇和口腔癌、肝癌、肺癌(非小细胞和小细胞)、淋巴瘤、男性乳腺癌、骨的恶性纤维组织细胞瘤和骨肉瘤、黑色素瘤、默克尔细胞癌、间皮瘤、转移癌、原发性隐匿的转移性颈鳞状细胞癌、中线束癌、口癌、多发性内分泌腺瘤综合征、多发性骨髓瘤、蕈样肉芽肿、骨髓增生异常综合征、骨髓增生异常/骨髓增生性肿瘤、鼻腔癌、鼻咽癌、神经母细胞瘤、非霍奇金淋巴瘤、非小细胞肺癌、口腔癌、唇癌和口腔癌、口咽癌、骨肉瘤、卵巢癌、胰腺癌、胰腺神经内分泌肿瘤、乳头瘤、副神经节瘤、副鼻窦癌、甲状旁腺癌、阴茎癌、咽癌、嗜铬细胞瘤、垂体瘤、浆细胞瘤、胸膜肺母细胞瘤、原发性中枢神经系统(CNS)淋巴瘤、原发性腹膜癌、前列腺癌、直肠癌、复发癌、视网膜母细胞瘤、横纹肌肉瘤、唾液腺癌、肉瘤、Sézary综合征、皮肤癌、小细胞肺癌、小肠癌、软组织肉瘤、皮肤鳞状细胞癌、隐匿原发性颈部鳞状细胞癌、胃癌、T细胞淋巴瘤、睾丸癌、喉癌、胸腺瘤和胸腺瘤、甲状腺癌、移行细胞癌、输尿管和肾盂癌、尿道癌、子宫癌、子宫肉瘤、阴道癌、血管瘤、外阴癌和肾母细胞瘤。

[0372] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是炎症或自身免疫病基因组、外显子组、转录组、蛋白质组或ORF组。在一些实施方式中,本发明的文库包含已知的炎症或自身免疫性新表位或自身表位。在一些实施方式中,本发明的文库包含可从已知炎症或自身免疫性抗原蛋白中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从炎症或自身免疫相关基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从炎症或自身免疫相关驱动基因的突变中衍生的所有k-聚肽。

[0373] 炎症或自身免疫病或病症的非限制性例子包括急性播散性脑脊髓炎(ADEM);急性坏死性出血性白脑炎;艾迪生病;佐剂诱导性关节炎;无丙种球蛋白血症;斑秃;淀粉样变性;强直性脊柱炎;抗GBM/抗TBM肾炎;抗磷脂综合征;自身免疫性血管水肿;自身免疫性再生障碍性贫血;自身免疫失调;自身免疫性胃萎缩;自身免疫性溶血性贫血;自身免疫性肝炎;自身免疫性高脂血症;自身免疫性免疫缺陷;自身免疫性内耳病(AIED);自身免疫性心肌炎;自身免疫性卵巢炎;自身免疫性胰腺炎;自身免疫性视网膜病变;自身免疫性血小板减少性紫癜(ATP);自身免疫性甲状腺疾病;自身免疫性荨麻疹;轴突和神经元神经病;巴洛病;白塞病;大疱性类天疱疮;心肌病;卡斯特曼病;乳糜泻;恰加丝病;慢性炎症脱髓鞘性多发性神经病(CIDP);慢性复发性多灶性骨髓炎(CRMO);变应性肉芽肿性血管炎(Churg-Strauss syndrome);瘢痕性类天疱疮/良性粘膜类天疱疮;克罗恩病;Cogan综合征;胶原诱导性关节炎;冷凝集素病;先天性心脏传导阻滞;柯萨奇心肌炎;CREST病;原发性混合冷球蛋白血症;脱髓鞘性神经病;疱疹样皮炎;皮炎;Devic病(视神经脊髓炎);盘状狼疮;Dressler综合征;子宫内膜异位症;嗜酸性食管炎;嗜酸性筋膜炎;结节性红斑;实验性变态反应性脑脊髓炎;实验性自身免疫性脑脊髓炎;埃文斯综合征;纤维肌痛;纤维性肺泡炎;巨细胞动脉炎(颞动脉炎);巨细胞性心肌炎;肾小球肾炎;古德帕斯综合征;肉芽肿性多血管炎(GPA)(以前称为韦格纳肉芽肿);Graves病;Guillain-Barre综合征;桥本脑炎;桥本甲状腺炎;溶血性贫血;过敏性紫癜;妊娠疱疹;低丙种球蛋白血症;特发性血小板减少性紫癜(ITP);IgA肾病;IgG4相关硬化性疾病;免疫调节性脂蛋白;包涵体肌炎;间质性膀胱炎;炎症性肠病;青少年关节炎;青少年少关节炎;青少年糖尿病(1型糖尿病);青少年肌炎;川崎综合征;lambert-Eaton综合征;白细胞破碎性血管炎;扁平苔藓;硬化苔藓;木质结膜炎;线性IgA病(LAD);狼疮(SLE);莱姆病,慢性;梅尼埃病;显微镜下多血管炎;混合结缔组织病;

莫伦氏溃疡; Mucha-Habermann病; 多发性硬化症; 重症肌无力; 肌炎; 嗜睡症; 视神经脊髓炎 (Devic's); 中性粒细胞减少症; 非肥胖性糖尿病; 眼瘢痕性类天疱疮; 视神经炎; 回文风湿病; PANDAS (与链球菌相关的小儿自身免疫性神经精神疾病); 副肿瘤性小脑变性; 阵发性夜间血色素尿症 (PNH); Parry Romberg综合征; Parsonnage-Turner综合征; 睫状体扁平部炎 (外周葡萄膜炎); 天疱疮; 寻常天疱疮; 外周神经病; 静脉周围脑脊髓炎; 恶性贫血; POEMS综合征; 结节性多动脉炎; I型、II型和III型自身免疫性多腺体综合征; 风湿性多发性肌痛; 多发性肌炎; 心肌梗塞后综合征; 心包切开术后综合征; 孕激素皮炎; 原发性胆汁性肝硬化; 原发性硬化性胆管炎; 银屑病; 斑块型银屑病; 银屑病性关节炎; 特发性肺纤维化; 坏疽性脓皮病; 纯红细胞再生障碍; 雷诺现象; 反应性关节炎; 反射性交感神经营养不良; Reiter综合征; 复发性多软骨炎; 不安腿综合征; 腹膜后纤维化; 风湿热; 类风湿关节炎; 结节病; Schmidt综合征; 巩膜炎; 硬皮病; 硬化性胆管炎; 硬化性涎腺炎; 干燥综合征; 精子和睾丸自身免疫; 僵人综合征; 亚急性细菌性心内膜炎; Susac综合征; 交感性眼炎; 系统性红斑狼疮 (SLE); 系统性硬化; Takayasu动脉炎; 颞动脉炎/巨细胞动脉炎; 血小板减少性紫癜 (TTP); Tolosa-Hunt综合征; 横纹肌炎; 1型糖尿病; 溃疡性结肠炎; 未分化结缔组织病 (UCTD); 葡萄膜炎; 血管炎; 水疱性皮肤病; 白癜风; 韦格纳肉芽肿病 (现称为肉芽肿性多血管炎 (GPA))。炎症或自身免疫病或病况的非限制性例子包括感染, 例如慢性感染、潜伏感染、缓慢感染、持续病毒感染、细菌感染、真菌感染、支原体感染或寄生虫感染。

[0374] 如例如在美国临时申请第62/791,601号中所述, 该申请通过引用整体并入本文。

#### [0375] B. 肽生产

[0376] 适用于pMHC多聚体的肽根据本领域已知的方法产生, 或由商业供应商合成生产, 或根据制造商的说明使用肽合成仪合成。例如, 在一些实施方式中, 适用于pMHC多聚体的肽可以通过计算机生产方法制备。

[0377] 在其它实施方式中, 肽可以通过化学方法合成, 例如, 茶包 (tea bag) 合成、数字光刻、针式 (pin) 合成和SPOT合成。例如, 可以通过SPOT合成产生肽阵列, 其中氨基酸链通过添加氨基酸和切割侧链保护基团的重复循环构建在纤维素膜上。

[0378] 在其它实施方式中, 可使用重组DNA技术表达肽, 例如, 将表达构建体引入细菌细胞、昆虫细胞或哺乳动物细胞, 并从细胞提取物中纯化重组蛋白。

[0379] 在一些实施方式中, 可通过体外转录和翻译合成肽, 其中合成利用无细胞环境中转录和翻译的生物学原理, 例如, 通过提供核酸模板、相关构建模块 (例如RNA、氨基酸)、酶 (例如RNA聚合酶、核糖体) 和条件。

[0380] 在一些实施方式中, 体外转录和翻译可以包括无细胞蛋白质合成 (CFPS)。通过CFPS获得高产量需要使用细菌系统, 其中翻译序列的第一个氨基酸是N-甲酰甲硫氨酸 (fMet)。该残基与甲硫氨酸的不同之处在于包含中性甲酰基团 (HCO) 而不是带正电荷的氨基-末端 ( $\text{NH}_3^+$ )。工程改造构建体以包括编码酶切结构域和文库多肽的基因, 如美国临时申请第62/791,601号所述, 该申请通过引用全文并入本文。

[0381] 去除至少初始甲硫氨酸氨基酸可以成功折叠肽并将其装载到MHC蛋白上。此外, 去除至少初始甲硫氨酸氨基酸允许肽文库多样性的上限更大, 例如,  $20^x$ , 其中x是肽的长度, 而包含该残基会将文库多样性限制为  $20^{(x-1)}$ 。

[0382] 在一些实施方式中, 利用体外转录/翻译 (IVTT) 系统合成肽, 该系统可以将例如

DNA构建体转录成RNA,然后将RNA翻译成蛋白质。例如,本公开的方法包括用于进行体外转录/翻译 (IVTT) 以生产高多样性的肽文库并允许正确的蛋白质折叠的方法。IVTT可以允许在无细胞环境下直接从DNA或RNA模板生产蛋白质。

[0383] 可以使用例如PCR产物、线性DNA质粒、环状DNA质粒或具有核糖体结合位点 (RBS) 序列的mRNA模板进行本文使用的IVTT方法。合适的模板被分离后,可以将转录组分加入到模板中,其包括,例如,三磷酸核糖核苷酸,和RNA聚合酶。转录完成后,可以加入翻译组分,其可在,例如,兔网织红细胞裂解液或小麦胚芽提取物中找到。在一些方法中,转录和翻译可在单个步骤中发生,其中,例如,在兔网织红细胞裂解液或小麦胚芽提取物中找到的纯化的翻译组分与转录组分同时添加到核酸模板中。

[0384] 在一些实施方式中,编码肽的N末端甲硫氨酸残基和可切割部分的核苷酸序列可在DNA构建体或RNA构建体中被编码。可切割部分位于使得该肽的至少一个N末端氨基酸残基在可切割部分之前或之内。在一些实施方式中,该方法包括编码可切割部分,其位于使得肽的一个N末端氨基酸残基在可切割部分之前或之内。在一些实施方式中,一个N末端氨基酸残基是甲硫氨酸残基。可以使用酶,例如对可切割部分特异的蛋白酶切割可切割部分,该蛋白酶也可以从肽的剩余部分切掉可切割部分。

[0385] 如本文所述的DNA或RNA构建体中编码的可切割部分的实例包括任何可被酶切割的可切割部分。在一些实施方式中,可由蛋白酶切割可切割部分。可使用对切割部分特异性的酶从肽中切下可切割部分。例如,该酶可以是因子Xa、人鼻病毒3C蛋白酶、AcTEV™蛋白酶, WELQut蛋白酶, Genenase™, 小泛素样修饰物 (SUMO) 蛋白、Ulp1蛋白酶或肠激酶。通过识别三级结构而非氨基酸序列, Ulp1蛋白酶可以特定的方式切下可切割部分。肠激酶 (肠肽酶) 也可用于从候选肽中切割可切割部分。肠激酶可以在赖氨酸之后在以下切割位点切割: DDDDK (SEQ ID NO:188)。肠激酶还可以在其他碱性残基处切割,取决于蛋白质底物的序列和构型。

[0386] 在一些实施方式中,可切割部分可以是小泛素样修饰 (SUMO) 蛋白质。可以使用对SUMO特异的蛋白酶从肽中切割下SUMO结构域。在一些实施方式中,可切割部分可以是肠激酶切割位点: DDDDK (SEQ ID NO.:188)。蛋白酶可以是,例如, Ulp1蛋白酶或肠激酶。通过识别SUMO的三级结构而非氨基酸序列, Ulp1蛋白酶可以特定的方式切下SUMO。肠激酶 (肠肽酶) 也可用于在赖氨酸之后在以下切割位点切割: DDDDK (SEQ ID NO:188)。肠激酶还可以在其他碱性氨基酸处切割,取决于蛋白质底物的序列。

[0387] 在编码肽的构建体翻译期间或之后, N末端氨基酸残基 (例如SUMO结构域) 可被高效地切割以产生恰当折叠的肽。在一些实施方式中,至少一个N末端氨基酸残基被切割以产生肽。在一些实施方式中,切割一个、两个、三个、四个、五个六个、七个、八个、九个、十个或更多N末端氨基酸残基以产生肽。该N末端氨基酸可以是任何氨基酸残基。该N末端氨基酸残基可以是甲硫氨酸氨基酸残基。因此该恰当折叠的肽不局限于具有N末端甲硫氨酸,还可以是通过无细胞体外方法产生的高多样性肽文库的部分。

[0388] 在编码肽的构建体翻译之后, N末端氨基酸残基可被切割以产生用于高多样性肽文库的肽。在一些实施方式中,至少一个N末端氨基酸残基被切割以产生肽。在一些实施方式中,一个或多个N末端氨基酸被切割,例如2,3,4,5,6,7,8,9,10,11,12,13,14,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,105,110,115,120,125,130,140,

150,160,170,180,190,200,250或更多N末端氨基酸残基被切割,以产生肽。该N末端氨基酸可以是任何氨基酸残基。该N末端氨基酸残基可以是甲硫氨酸氨基酸残基。

[0389] 在一些实施方式中,DNA或RNA构建体包含嘌呤霉素。在一些实施方式中,DNA或RNA构建体包含缺少终止密码子的间隔子序列。在一些实施方式中,肽通过亲和和标签纯化(例如,使用FLAG标签)来纯化。在一些实施方式中,肽包含HaloTag酶序列。在一些实施方式中,肽包含亲和素或链霉亲和素。

[0390] 对于哺乳动物表达,在哺乳动物表达载体中设计了一种编码CMV肽的构建体,其具有带或不带C末端His标记的C末端Flag标记。根据厂商的建议,通过瞬时转染Expi293F或ExpiCHO-S细胞(生命技术(Life Technologies))表达肽。

[0391] 用抗Flag亲和层析(Genscript)或Ni亲和层析从细胞培养上清液中纯化肽。在亲水性树脂(GE生命科学)上进行尺寸排阻层析(SEC),该树脂在20mM HEPES,150mM NaCl,pH 7.2中预平衡。

[0392] 或者,使用23mM磷酸钠,500mM氯化钠,500mM咪唑(pH 7.4)的柱缓冲液,通过Ni亲和层析纯化肽,而无需SEC纯化。

[0393] 哺乳动物细胞中产生的肽通过280nm处的UV进行定量,而CFPS产生的肽通过夹心ELISA相对于标准蛋白质进行定量。

#### [0394] V. 肽交换

[0395] p\*MHC多聚体用于通过如本文所述的原位或体外肽交换反应产生加载有多种独特肽表位的pMHC多聚体的文库或微阵列。在一些实施方式中,肽交换反应以多孔形式并在天然条件下进行。结合由多种技术确定,如ELISA,其监测MHC结构的稳定性,或通过生物物理技术监测肽结合,如荧光偏振。实施例4中详细描述了通过二肽交换或UV介导的交换进行肽交换的非限制性实例。

[0396] 在一些实施方式中,为了测量占位符肽或肽片段的解离效率,在存在未标记交换肽的情况下,在交换反应中使用荧光标记的占位符肽。荧光标记的p\*MHC多聚体的等分试样要么未经处理,要么暴露于肽交换条件(例如,UV暴露)不同的时间段。通过荧光分析监测剩余含p\*MHC的占位符肽量,以监测p\*MHC-复合物的减少。

[0397] 在一些实施方式中,与交换肽表位相比,占位符肽对MHC肽结合沟的亲合力更低,并且其中步骤(d)包括在竞争测试中使p\*MHC单体与过量肽表位接触。在一些实施方式中,占位符肽的KD比交换肽表位低约10倍。

[0398] 与MHC分子的肽结合沟结合的肽可以是天然存在的肽,但也可以使用特定MHC分子或其所属的超型家族的B和F口袋的结合特异性的知识来合成。可以使用MHC复合物的可用3D结构和关于相应MHC分子的结合口袋特异性的知识来生成合适的配体。

[0399] MHC I多肽的肽结合特异性主要由B和F结合口袋的物理化学性质以耦合方式控制。B和F结合口袋通常与肽中的“锚定残基”结合,该“锚定残基”定义了MHC的肽结合沟中肽的结合。观察到的MHC分子肽结合沟氨基酸残基的多样性确定了单个MHC分子的肽结合和递呈谱(Chang等人,2011;Frontiers in Bioscience, Landmark Edition, Vol.16:3014-3035)。例如,如Sidney等人(BMC Immunology Vol.9:12008)所述,对于大量MHC分子,已经阐明了锚定残基的口袋的特异性。

[0400] 本公开进一步提供了一种生产p\*MHC多聚体的方法,包括:生产其中结合沟中的肽

是占位符肽的p\*MHC-多聚体;使p\*MHC多聚体与还原剂接触以除去占位符肽;以及在足以结合MHC肽结合沟中的肽表位的条件下使p\*MHC多聚体与MHC肽表位接触。

[0401] 两个接触步骤优选通过提供含有MHC肽表位和还原剂的MHC分子的样品来进行。优选在加入还原剂时存在MHC肽表位。在一些实施方式中,每次反应添加一个MHC肽表位。在一些实施方式中,将两个或多个肽表位添加到反应中。

[0402] 在一些实施方式中,通过将混合物的温度升高至约30°-37°C。在一些实施方式中,将混合物升高至31°,32°,33°,34°,35°,36°或37°C。

[0403] 在一些实施方式中,通过将混合物的pH降低至约pH 2.5-5.5来诱导肽交换。在一些实施方式中,通过将混合物的pH增加至约pH 9-11来诱导肽交换。

[0404] 在一些实施方式中,占位符肽包括可光切割部分以形成所述的pMHC复合物(例如Toebes等,Nat. Med. 12:246-251,2006;Bakker等人,PNAS 105:3825-383,2008;Frosig等人,Cytometry Part A,87A:967-975,2015;Chang等人,Eur. J. Immunol. 43:1109-112,2013)。在一些实施方式中,占位符肽包括含有(2-硝基)苯基侧链的非天然氨基酸。在一些实施方式中,氨基酸是包含3-氨基-3-(2-硝基)苯基丙酸的UV敏感β-氨基酸。在一些实施方式中,UV敏感氨基酸是(2-硝基)苯基甘氨酸。

[0405] 在一些实施方式中,占位符肽是HLA-A2肽。在一些实施方式中,HLA-A2占位符肽是p\*A2,KILGCVFJV (SEQ ID NO:15)或GILGFVJL (SEQ ID NO:7),其中J是3-氨基-3-(2-硝基)苯基丙酸。

[0406] 在一些实施方式中,占位符肽是含有可光切割部分的HLA-A1、-A3、A11或-B7肽。在一些实施方式中,占位符肽选自以下组:A\*01:01,STAPGJLEY (SEQ ID NO:16);A\*03:01,RIYRJGATR (SEQ ID NO:17);A\*11:01,RVFAJSFIK (SEQ ID NO:18);A\*24:02,VYGJVRACL (SEQ ID NO:11);B\*07:02,AARGJTLAM (SEQ ID NO:14);B\*35:01,KPIVVLJGY (SEQ ID NO:19);C\*03:04,FVYGJSKTSL (SEQ ID NO:20);B\*08:01,FLRGRAJGL (SEQ ID NO:21);C\*07:02,VRIJHLYIL (SEQ ID NO:22);C\*04:01,QYDJAVYKL (SEQ ID NO:23);B\*15:01,ILGPJGSVY (SEQ ID NO:24);B\*40:01,TEADVQJWL (SEQ ID NO:25);B\*58:01,ISARGQJLF (SEQ ID NO:26);和C\*08:01,KAJDLSHFL (SEQ ID NO:27),其中J是3-氨基-3-(2-硝基)苯基丙酸。在其他实施方式中,占位符肽具有SEQ ID NO:7-27或271-279中任一所示的序列。

[0407] 在一些实施方式中,占位符肽还包括荧光标记。在一些实施方式中,荧光标记附着在占位符肽中的半胱氨酸残基上。

[0408] 在长波长UV照射下,肽在一种或多种肽的存在下被切割并从MHC复合物解离,以促进稳定的pMHC单体或多聚体的形成。通常,MHC肽交换以多孔形式进行,用于如本文所述的肽配体的高通量筛选。只有能够有效结合和稳定肽受体MHC分子的候选肽才能阻止MHC复合物的解离。肽交换可以通过多种技术,例如ELISA或荧光偏振监测,如Rodenko等人(Nat. Protocol. 1:1120-1132,2006)中所述。

[0409] 随后通过凝胶过滤HPLC和MHC-ELISA分析所得的pMHC多聚体以确定三个参数:MHC重折叠效率、pMHC复合物在无UV暴露下的稳定性以及复合物的UV敏感性。

[0410] 某些二肽可以帮助MHC I类分子的折叠和肽交换。二肽特异性结合到MHC I类分子的F口袋以促进肽交换,迄今为止已在HLA-A\*02:01、HLA-B\*27:05和H-2Kb分子中描述并验证了肽交换(Saini等人,Proc Natl Acad Sci U S A.2013年9月17日;110(38):15383-8)。

[0411] 因此,在一些实施方式中,占位符肽与一个或多个感兴趣肽的肽交换由二肽催化,该二肽催化MHC I类分子上的快速肽交换(参见例如Saini等人,Proc Natl Acad Sci U S A.2015年1月6日;112(1):202)。合适的二肽是具有疏水性第二残基的二肽。在一些实施方式中,二肽是甘氨酸亮氨酸(GL)、甘氨酸缬氨酸(GV)、甘氨酸甲硫氨酸(GM)、甘氨酸环己基丙氨酸(GCha)、甘氨酸高亮氨酸(GHle)或甘氨酸苯丙氨酸(GF)。

[0412] 在另一个实施方式中,占位符肽与一个或多个感兴趣肽的肽交换是通过伴侣介导的肽交换来实现的,例如使用分子伴侣TAPBPR,如Overall等人(2020)Nature Comm.11:1909中所述。

#### [0413] VI.pMHC文库生产

[0414] 在一个方面,本文提供了制备包含多种加载的肽表位的pMHC多聚体文库的方法。制备肽交换的条码pMHC文库的各个步骤如图18所示。这些步骤使用本领域已知的用于制备条码化文库的标准方法,包括使用单细胞测序、使用多孔水凝胶、使用单模板PCR生成肽编码扩增子(条码)和使用液滴(in-drop)体外转录/翻译(IVTT)。

[0415] 实施例9中描述了用合并的条码化UV肽交换的MHC四聚体进行单细胞测序的非限制性示例。实施例10中详细描述了用于高通量生产条码化UV肽交换的MHC四聚体池的多孔水凝胶生产的非限制性实例。实施例11中详细描述了使用单模板PCR生成肽编码扩增子的非限制性示例。实施例12中描述了在水凝胶上加载可条码化、可交换的MHC四聚体的非限制性示例。实施例13中详细描述了肽和UV交换到加载的MHC四聚体中的液滴体外转录/翻译(IVTT)的非限制性示例。实施例14中详细描述了从水凝胶中释放UV肽交换的条码化pMHC四聚体的非限制性示例。

[0416] 在一些实施方式中,该方法包括:(a)提供多种加载占位符肽的MHCI (p\*MHCI) 单体,其各自包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分,和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;(b)提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分;(c)在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将p\*MHC I单体和所述多聚结构域组合以产生p\*MHCI多聚体;和(d)用包含多个独特MHCI肽表位的肽文库替换多个p\*MHCI多聚体中的占位符肽,以产生多个加载肽的MHCI (pMHC) 多聚体。

[0417] 在一些实施方式中,该方法包括:(a)提供多个加载占位符肽的MHCI (p\*MHCI) 单体,每个单体包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段,(iii) 偶联部分和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;(b)提供多个多聚结构域,其中所述多聚结构域的每个亚基包括偶联部分,所述多聚结构域包括至少一个非共价结合位点;(c)在足以使两个或多个p\*MHCI单体与多聚结构域共价偶联的条件下,将多个p\*MHC I单体和多个多聚结构域组合,以产生多个p\*MHCI多聚体;(d)用多个拯救肽表位替换p\*MHCI多聚体中肽结合沟中结合的占位符肽,以产生多个pMHC多聚体;和(e)将寡核苷酸条码结合到多聚结构域上的非共价结合位点。

[0418] 在一些实施方式中,该方法包括:(a)提供多种加载占位符肽的MHCI (p\*MHCI) 单体,其各自包含(i) MHCI重链多肽或其功能片段,(ii)  $\beta$ 2-微球蛋白多肽或其功能片段;(iii) 位于(i) 或(ii) C末端处的包含偶联部分的肽接头;和(iv) 结合在每个MHCI单体的肽结合沟中的占位符肽;(b)提供多个多聚结构域,其在多聚结构域的每个亚基N末端处包含

肽接头,所述肽接头包含偶联部分;(c)在足以使所述两个或多个p\*MHCI单体与所述多聚结构域共价偶联的条件下,将多个p\*MHC I单体和多个多聚结构域组合,以产生多个p\*MHCI多聚体;和(d)用多个独特MHC I肽表位替换p\*MHCI多聚体肽结合沟中结合的占位符肽,以产生多个pMHC I多聚体。

#### [0419] VII. 标记

[0420] pMHC多聚体可与荧光标记偶联,从而允许鉴定结合肽-MHC多聚体的T细胞,例如,通过流式细胞术或显微镜检。也可以基于荧光标记通过荧光活化细胞分选的方式选择T细胞。

[0421] 在一些实施方式中,一个或多个可检测标记物与接头偶联。根据本发明,“可检测标记”是允许检测生物或化学特性或系统变化的任何分子或官能团,例如样品中靶物质的存在。

[0422] 可使用的可检测标记的实例包括荧光团、发色团、电化学发光标记、生物发光标记、聚合物、聚合物颗粒、珠子或其他固体表面、金或其他金属颗粒或重原子、自旋标记、放射性同位素、酶底物、半抗原、抗原、量子点、氨基己基、苾、核酸或核酸类似物或蛋白质,例如受体、肽配体或底物、酶和抗体(包括抗体片段)。

[0423] 可使用的聚合物颗粒标签的实例包括聚苯乙烯、PMMA或二氧化硅的微粒、珠或胶乳颗粒,其可嵌入荧光染料,或含有染料、酶或底物的聚合物胶束或胶囊。可使用的金属颗粒的例子包括金颗粒和涂覆的金颗粒,其可被银渍转化。在一些实施方式中可偶联的半抗原实例为荧光团、myc、硝基酪氨酸、生物素、亲和素、链霉亲和素、2,4-二硝基苯基、地高辛、溴脱氧尿苷、磺酸盐、乙酰氨基苄、三硝基苯酚汞和雌二醇。

[0424] 可使用的酶的例子包括辣根过氧化物酶(HRP)、碱性磷酸酶(AP)、 $\beta$ -半乳糖苷酶(GAL)、葡糖-6-磷酸脱氢酶、 $\beta$ -N-乙酰氨基葡萄糖苷酶、 $\beta$ -葡萄糖醛酸酶、转化酶、黄嘌呤氧化酶、萤火虫荧光素酶和葡糖氧化酶(GO)。用于辣根过氧化物酶(HRP)的常用底物的例子包括3,3'-二氨基联苯胺(DAB)、具有镍增强作用的二氨基联苯胺、3-氨基-9-乙基咪唑(AEC)、二盐酸联苯胺(BDHC)、Hanker-Yates试剂(HYR)、吡啶芬蓝(IB)、四甲基联苯胺(TMB)、4-氯-1-萘酚(CN)、 $\alpha$ -萘酚派洛宁( $\alpha$ -NP)、邻联茴香胺(OD)、5-溴-4-氯-3-吡啶磷酸(BCIP)、硝基蓝四唑(NBT)、2-(对碘苯基)-3-对硝基苯基-5-苯基四唑氯化物(INT)、四硝基蓝四唑(TNBT)、 $\delta$ -溴-氯-S-吡啶基- $\beta$ -D-半乳糖苷/铁-亚铁氰化物(BCIG/FF)。碱性磷酸酶常用底物的实例包括萘酚-AS-B1-磷酸/快红TR(NABP/FR)、萘酚-AS-MX-磷酸/快红TR(NAMP/FR)、萘酚-AS-B1-磷酸/快红TR(NABP/FR)、萘酚-AS-MX-磷酸/快红TR(NAMP/FR)、萘酚-AS-B1-磷酸/新复红(NABP/NF)、溴氯吡啶基磷酸/硝基蓝四唑(BCIP/NBT)、 $\delta$ -溴-氯-S-吡啶基- $\beta$ -D-半乳糖吡喃糖苷(BCIG)。

[0425] 可使用的发光标记物的实例包括鲁米诺、异鲁米醇、吡啶酯、1,2-二氧杂环丁烷和吡啶并咪唑。电化学发光标记物的实例包括钌衍生物。可使用的放射性标记的实例包括碘、钴、硒、氢、碳、硫和磷的放射性同位素。

[0426] 一些“可检测标记”还包括“颜色标记”,其中系统中的生物变化或事件可以通过颜色的存在或颜色的变化来检测。“颜色标记”的例子有发色团、荧光团、化学发光化合物、电化学发光标记、生物发光标记和催化底物颜色变化的酶。

[0427] 本文所述的“荧光团”是在用一种或多种波长的电磁辐射激发时发出可检测电磁



辐射的分子。本领域已知多种荧光团，由化学家开发用作可检测的分子标记，并可与本文提供的pMHC多聚体偶联。示例包括FLUORESCHEIN.TM或其衍生物，如FLUORESCHEIN®-5-异硫氰酸酯(FITC)、5-(和6)-羧基FLUORESCHEIN®、5-或6-羧基FLUORESCHEIN®、6-(FLUORESCHEIN®)-5-(和6)-羧酰氨基己酸、FLUORESCHEIN®异硫氰酸酯、罗丹明或其衍生物，如四甲基罗丹明和四甲基罗丹明-5-(和-6)异硫氰酸盐(TRITC)。其他荧光团包括：香豆素染料，如(二乙氨基)香豆素或7-氨基-4-甲基香豆素-3-乙酸、琥珀酰亚胺酯(AMCA)；磺基罗丹明101磺酰氯(TexasRed®或TexasRed®磺酰氯)；5-(和-6)-羧基罗丹胺101，琥珀酰亚胺酯，也称为5-(和6)-羧基-X-罗丹明，琥珀酰亚胺酯(CXR)；丽丝胺或丽丝胺衍生物，如丽丝胺罗丹明B磺酰氯(LisR)；5-(和-6)-羧基FLUORESCHEIN®，琥珀酰亚胺酯(CFI)；FLUORESCHEIN®5-异硫氰酸酯(FITC)；7-二乙氨基香豆素-3-羧酸、琥珀酰亚胺酯(DECCA)；5-(和-6)-羧基四甲基罗丹明琥珀酰亚胺酯(CTMR)；7-羟基香豆素-3-羧酸，琥珀酰亚胺酯(HCCA)；6-FLUORESCHEIN®-5-(和-6)-羧酰胺基己酸(FCHA)；N-(4,4-二氟-5,7-二甲基-4-硼-3a,4a-二氮杂-3-二环戊二烯并苯丙酸，琥珀酰亚胺酯；也称为5,7-双甲基BODIPY®丙酸，琥珀酰亚胺酯(DMBP)；“活化FLUORESCHEIN®衍生物”(FAP)，可从Probes, Inc.购得；曙红-5-异硫氰酸酯(EITC)；赤藓红素-5-异硫氰酸酯(Er1TC)；和Cascade®蓝色乙酰叠氮化物(CBAA)(1-羟基-3,6,8-苾-三磺酸的0-乙酰叠氮衍生物)。用于本发明的其他可能荧光团包括荧光蛋白，如绿色荧光蛋白及其类似物或衍生物，荧光氨基酸，如酪氨酸和色氨酸及其类似物，荧光核苷，以及其他荧光分子，如Cy2、Cy3、Cy3.5、CY5.TM、CY5.TM.5、Cy7、IR染料、Dyomics染料、藻红蛋白、俄勒冈绿488、太平洋蓝、罗丹明绿和Alexa染料。荧光标记的其他实例包括R-藻红蛋白或异藻红蛋白(allo-phycoerythrin)的偶联物，无机荧光标记例如基于半导体材料的颗粒，如涂覆的CdSe纳米晶体。

[0428] 上述许多荧光团以及其他荧光团可从Probes, Inc. (俄勒冈州尤金)、Pierce Chemical Co. (伊利诺伊州罗克福德)或Sigma-Aldrich Co. (密苏里州圣路易斯)等公司购得。

[0429] 可检测标记可通过多种方法检测，包括例如对于光学标记通过反射率、透射率、光散射、旋光和荧光或其组合，或者对于放射性标记通过薄膜、闪烁计数或磷成像。见例如Larsson, 1988, Immunocytochemistry: Theory and Practice, (CRC Press, Boca Raton, Fla.); Methods in Molecular Biology, 80卷, 1998, John D. Pound (编) (Humana Press, Totowa, N.J.)。在一些实施方式中，采用了一种以上的可检测标记。

[0430] VIII. 标识符/条码化

[0431] 在某些实施例中，本发明的偶联多聚体包括标识符标签或标记物，例如寡核苷酸条码，其有助于识别偶联多聚体。通常，标识符标签，例如寡核苷酸条码，例如通过标识符标签(例如寡核苷酸条码)上的结合部分附着到偶联多聚体的多聚结构域，该结合部分结合到多聚结构域上的结合位点。例如，当多聚结构域是链霉亲和素或亲和素时，由于pMHCI单体在生物素结合位点以外的位点与多聚结合域偶联，可以使用标识符标签的生物素化形式(例如生物素化寡核苷酸条码)用标识符标签(例如寡核苷酸条码)标记偶联多聚体。然后，通过将偶联多聚体与生物素化标识符标签(例如生物素化寡核苷酸条码)孵育，可以容易地实现偶联多聚体的标记。实施例8详细描述了使用生物素化寡核苷酸对偶联多聚体进行条

码化的非限制性示例。

[0432] 在另一个实施方式中,在多聚体的肽部分中用标识符标签(例如寡核苷酸条码)标记偶联多聚体。也就是说,条码标记的MHC结合肽可用于本文所述的交换反应中,以加载带有条码标记肽的偶联多聚体。

[0433] 通常,寡核苷酸条码是范围为10至50个核苷酸以上的独特寡核苷酸序列。条码在3'和5'端具有共享的扩增序列,在中间为独特序列。该序列可以通过测序来揭示,并且可以作为给定分子的特定条码。

[0434] 在一个实施方式中,条码的核酸成分(通常为DNA)具有特殊结构。因此,在一个实施方式中,至少一个核酸分子由至少5'第一引物区、中心区(条码区)和3'第二引物区组成。以这种方式,中心区(条码区)可用引物组扩增。核酸分子的长度也可以变化。因此,在其他实施方式中,所述至少一种核酸分子的长度在20-100个核苷酸的范围内,例如30-100个核苷酸,例如30-80个核苷酸,如30-50个核苷酸。在一个实施方式中,核酸标识符的长度为40个核苷酸至120个核苷酸。寡核苷酸条码与偶联多聚体的耦合也可能不同。因此,在一个实施方式中,所述至少一个寡核苷酸条码通过生物素结合域与所述偶联多聚体内的链霉亲和素或亲和素相互作用而与所述偶联多聚体连接。也可以使用其他偶联部分,这取决于与偶联多聚体的适当结合位点(例如,在偶联多聚体的多聚结构域内)和可连接到寡核苷酸条码分子以促进连接的适当对应结合域的可实现性。

[0435] 在另一个实施方式中,所述至少寡核苷酸条码分子包括DNA、RNA和/或人工核苷酸(例如PLA或LNA)或由其组成。优选DNA,但可以包括其他核苷酸以例如增加稳定性。

[0436] 本领域熟知条码技术的应用,参见例如Shiroguchi等,Proc Natl Acad Sci U S A., 2012年1月24日;109(4):1347-52;和Smith等,Nucleic Acids Research, 2010年7月;38(13)11:e142。使用条码技术的其他方法和组合物包括U.S. 2016/0060621中描述的那些。还描述了使用条码技术特异性标记MHC多聚体,见例如Bentzen等人,Nature Biotech. 34:10:1037-1045, 2016;Bentzen and Hadrup,Cancer Immunol.Immunotherap.66:657-666, 2017。本领域已知制备条码寡核苷酸的标准方法,包括将其与可结合偶联多聚体的合适结合部分(例如生物素化)结合,并可应用于制备用于标记偶联多聚体的条码寡核苷酸。

[0437] 生成可定制DNA条码文库的方法是公开的。程序包括Generator和nxCODE,分别由96-587个条码组成,以及DNA条码软件包和TagD软件(报告生成由100,000个条码组成的文库)。

[0438] 本领域已经描述了各种大规模条码文库的制备,这些方法可以用于获得用于标记pMHC偶联多聚体文库的条码文库。例如,Xu等描述了一组240,000种独特的25聚寡核苷酸,其序列具有相似的扩增特性,同时维持其识别基序的最大多样性(Xu等人,PNAS 106:2289-2294, 2008)。Wang等人描述了使用粒子群优化构建条码集(Wang等人,IEEE/ACM Trans.Comput.Biol.Bioinform. 15:999-1002)。Lyons描述了由多达一百万成员组成的大规模DNA条码文库的产生。(Lyons,Sci.Reports 7:13899,2017)。

[0439] 在一些情况下,独特分子标识符条码由在靶核酸一端带标签的核苷酸的连续序列编码。在其它情况下,独特分子标识符(UMI)条码由不连续的序列编码。不连续的UMI可在靶核酸的第一端具有条码的部分,并且在靶核酸的第二端具有条码的部分。在一些情况下,UMI是不连续的条码,其在靶核酸的第一端包含可变长度条码序列,在第二端包含第二标识

符序列。在一些情况下,UMI是不连续的条码,其在靶核酸的第一端具有可变长度条码序列,在第二端具有第二标识符序列,其中所述第二标识符序列由转座酶片段化事件的位置确定,例如,转座酶片段化位点和转座子端插入事件。

[0440] 在一些情况下,条码是“可变长度条码”。如本文所用,可变长度条码是在长度方面与群体中的其它可变长度条码寡核苷酸不同的寡核苷酸,这可以通过条码中连续核苷酸的数量来鉴别。在一些情况下,除了可变长度之外,还可通过使用可变核苷酸序列来提供可变长度条码的附加条码复杂性,如以上段落中所述。

[0441] 在一个示例性的实施方式中,可变长度条码的长度可为0至不超过5个核苷酸。这样的可变长度条码可以用术语“[0-5]”表示。在这样的实施方式中,应理解,连接于这种可变长度条码的靶核酸的群体预期包括与具有至少1个核苷酸的可变长度条码相连的至少一个靶核酸(例如,连接于具有仅1个,仅2个,仅3个,仅4个或仅5个核苷酸的可变长度条码)。在这样的实施方式中,还应理解,连接于所述可变长度条码的靶核酸的群体可包括不包含可变长度条码(即,长度为0的可变长度条码)的至少一个靶核酸,和/或包含具有仅1个核苷酸的可变长度条码的至少一个靶核酸,和/或包含具有仅2个核苷酸的可变长度条码的至少一个靶核酸,和/或包含具有仅3个核苷酸的可变长度条码的至少一个靶核酸,和/或包含具有仅4个核苷酸的可变长度条码的至少一个靶核酸,和/或包含具有仅5个核苷酸的可变长度条码的至少一个靶核酸。在这样的实施方式中,[0-5]可变长度条码可以通过其自身而独特地识别(区分)具有相同序列的5个不同靶核酸分子。此外,在这样的实施方式中,对于每个不同的靶核酸序列,[0-5]可变长度条码可独特地识别(区分)第一序列的5个不同的靶核酸分子,第二序列的5个不同的靶核酸分子等。此外,条码标记的MHC多聚体可以与单细胞分选和TCR测序组合使用,其中TCR的特异性可以通过共附着的条码来确定。这将使我们能够从同一样品中平行鉴定可能的1000+种不同抗原反应性T细胞的TCR特异性,并将TCR序列与抗原特异性相匹配。该技术的未来潜力与基于TCR序列预测抗原反应性的能力有关。

[0442] 条码标记的MHC多聚体文库的复杂性将允许在给定个体中个性化选择相关TCR。

[0443] 条码与多聚体共附着,并作为特定肽-MHC复合物的特定标记物。以这种方式,可以混合至少1000至10,000或更多不同的肽-MHC多聚体,允许与来自血液或其他生物样品的T细胞特异性相互作用,洗出未结合的MHC-多聚体并确定DNA-条码的序列。当选择感兴趣的细胞群时,背景水平以上的条码序列将提供指纹,用于识别给定细胞群中存在的抗原反应性细胞。每个特定条码的序列读取次数将与特定T细胞的频率相关,并且可以通过将读取频率与T细胞的输入频率进行比较来估计频率。

[0444] DNA条码用作抗原特异性T细胞的特异性标签,并可用于在例如单细胞分选、功能分析或表型评估后确定T细胞的特异性。以这种方式,抗原特异性可以与T细胞受体序列(可以通过单细胞测序方法揭示)以及抗原特异性细胞的功能和表型特征两者相关联。

[0445] 条码标记的MHC多聚体文库可用于定量评估MHC多聚体与给定T细胞克隆或TCR转导/转染细胞的结合。由于条码标签的测序允许在同一细胞群上同时测定几个不同的标签,因此该策略可用于确定给定TCR相对于相关肽-MHC多聚体文库的亲和力。基于文库中每个不同肽MHC多聚体的TCR结合的定量贡献,确定不同DNA-条码序列在最终读出中的相对贡献。通过基于滴定的分析,可以确定TCR与肽-MHC多聚体的大型文库的定量结合特性,这些多聚体都合并到一个样品中。为此,MHC多聚体文库可以特异性地容纳相关肽序列或丙氨酸

取代肽文库。

[0446] 在一些实施方式中,可以对多个样品中的每个样品使用独特标识符。在一些实施方式中,可以在两个或多个样品之间共享标识符。在一些实施方式中,标识符可以包括在所有样品之间共享的一些序列,以及对一个样品独特的其他序列。在一些实施例中,标识符可以包括所有样品之间共享的序列和对一个样品独特的序列。在一些实施方式中,样品之间共有的序列可用于标识符扩增(例如,使用合适的引物进行PCR扩增)。在一些实施方式中,对一个样品独特的或在样品亚组之间共有的序列可用于经由qPCR进行检测或量化(例如,水解探针的序列,例如TaqMan探针)。在一些实施方式中,一个样品的独特序列或样品亚组之间的共有序列可用于通过测序进行检测或量化。

[0447] 在一些实施方式中,标识符可包括独特的计算机上生成的序列;每个标识符序列可分配给多个样品中的一个样品,并且标识符-样品的分配可存储在数据库中。在一些实施方式中,标识符可包含编码肽或蛋白质的全部或部分的核苷酸序列。在一些实施方式中,标识符可包含编码开放阅读框的核苷酸序列。在一些实施方式中,标识符可包含包括启动子序列的核苷酸序列。在一些实施方式中,标识符可包含核苷酸序列,其包括DNA结合蛋白(例如转录因子或聚合酶)的结合位点。在一些实施方式中,标识符可包含核酸酶(例如限制性内切酶)靶向的一个或多个序列。在一些实施方式中,标识符可包含序列的体外转录和翻译所需的全部序列元件。在一些实施方式中,标识符不包括序列的体外转录和翻译所需的全部序列元件。

[0448] 在一些实施方式中,标识符可包含生物素化核苷酸序列。在一些实施方式中,标识符可通过使用生物素化引物的PCR扩增进行生物素化。在一些实施方式中,标识符可通过使用Klenow DNA聚合酶、缺口翻译或混合引物标记RNA聚合酶(包括T7、T3和SP6 RNA聚合酶)酶促纳入生物素化标签(例如,生物素dUTP标签)来生物素化。在一些实施方式中,标识符可通过光生物素化进行生物素化,例如,可将光可活化生物素添加到样品中,且用紫外光照射样品。

[0449] 在一些实施方式中,例如通过模板DNA的PCR扩增,可从模板多核苷酸生成标识符。在一些实施方式中,模板多核苷酸可包含编码开放阅读框的核苷酸序列。在一些实施方式中,模板多核苷酸可包含包括启动子序列的核苷酸序列。在一些实施方式中,模板多核苷酸可包含核苷酸序列,其包括DNA结合蛋白(例如转录因子或聚合酶)的结合位点。在一些实施方式中,模板多核苷酸可包含核酸酶(例如限制性内切酶)靶向的一个或多个序列。在一些实施方式中,模板多核苷酸可包含序列的体外转录和翻译所需的全部序列元件。在一些实施方式中,模板多核苷酸不包括序列的体外转录和翻译所需的全部序列元件。

[0450] 带有附着标识符(例如寡核苷酸条码)的pMHC多聚体可以与多个T细胞一起孵育,然后将T细胞分选入单细胞隔室。裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并核酸并测序。标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR-抗原特异性谱。

[0451] 可鉴定出对肽文库的肽显示结合亲和力的多个TCR,并可鉴定出对特异性TCR显示结合亲和力的多种肽。

[0452] 可鉴定导致TCR结合亲和力增加的所鉴定TCR-抗原对的抗原中的表位突变。

[0453] 可鉴定出与疾病相关蛋白的控制相关的肽和TCR序列,并用于设计疫苗和细胞疗法。

[0454] 为了评估对治疗的反应,对于测序的每个肽标识符,鉴定相应的TCR序列。鉴定出对肽文库的一些肽显示结合亲和力的多个TCR,并鉴定出对一些TCR显示结合亲和力的多种肽。纵向跟踪对象,并比较分析结果,以确定与免疫治疗成功反应相关的肽和TCR序列。

[0455] IX. 载体和多核苷酸

[0456] 本公开还包括编码本文描述的任何蛋白质的核酸序列。如本领域技术人员所理解的,由于第三碱基简并,在编码核苷酸序列中,几乎每个氨基酸都可以由一个以上的三联体密码子表示。此外,微小的碱基对变化可能导致编码的氨基酸序列的保守取代,但预计不会显著改变基因产物的生物活性。因此,编码本文所述蛋白质的核酸序列可以在序列上稍加修饰,但仍编码其各自的基因产物。

[0457] 编码本文所述各种蛋白质或多肽中任一种的核酸可以化学合成。可以选择密码子的使用,以便改进细胞中的表达。这种密码子的使用将取决于所选的细胞类型。已经为大肠杆菌和其他细菌,以及哺乳动物细胞、植物细胞、酵母细胞和昆虫细胞开发了专门的密码子使用模式。见例如:Mayfield等人,Proc.Natl.Acad.Sci.USA,100(2):438-442(2003年1月21日);Sinclair等人,Protein Expr.Purif.,26(I):96-105(2002十月);Connell,N.D.,Curr.Opin.Biotechnol.,12(5):446-449(2001十月);Makrides等人,Microbiol.Rev.,60(3):512-538(1996年9月);和Sharp等人,Yeast,7(7):657-678(1991十月)。

[0458] 核酸操作的一般技术描述于例如Sambrook等人,Molecular Cloning:A Laboratory Manual,第二版,1-3卷,Cold Spring Harbor Laboratory Press(1989),或Ausubel,F.等人,Current Protocols in Molecular Biology,Green Publishing and Wiley-Interscience,New York(1987)和定期更新,通过引用并入本文。通常,编码多肽的DNA可操作地连接到源自哺乳动物、病毒或昆虫基因的合适的转录或翻译调控元件。这些控制序列包括转录启动子,控制转录的可选操纵子序列,编码合适的mRNA核糖体结合位点的序列,以及控制转录和翻译终止的序列。还纳入了通常由复制起源赋予的在宿主中复制的能力,以及促进转化子识别的选择基因。

[0459] 本文所述的蛋白不仅可以直接重组生产,而且可以作为与异源多肽(优选信号序列或其他在成熟蛋白质或多肽的N末端具有特定切割位点的多肽)的融合多肽。所选择的异源信号序列优选是被宿主细胞识别和处理(即被信号肽酶切割)的序列。

[0460] 对于不识别和处理天然信号序列的原核宿主细胞,信号序列被选自例如碱性磷酸酶、青霉素酶、1pp或热稳定肠毒素II前导物的原核信号序列取代。

[0461] 对于酵母分泌,天然信号序列可由例如酵母转化酶前导物、 $\alpha$ 因子前导物(包括酿酒酵母和克鲁维酵母 $\alpha$ 因子前导物)或酸性磷酸酶前导物、白色念珠菌(*C.albicans*)葡糖淀粉酶前导物或美国专利号5,631,144中所述的信号序列取代。在哺乳动物细胞表达中,可获得哺乳动物信号序列以及病毒分泌前导物,例如单纯疱疹gD信号。这种前体区域的DNA可以在阅读框中连接到编码蛋白质的DNA上。

[0462] 表达载体和克隆载体都含有使载体能够在一种或多种选定的宿主细胞中复制的核酸序列。在克隆载体中,该序列通常是使载体能够独立于宿主染色体DNA复制的序列,并且包括复制起点或自主复制序列。这些序列对于各种细菌,酵母和病毒是众所周知的。质粒

pBR322的复制起点适用于大多数革兰氏阴性细菌,2微米质粒起点适用于酵母,而各种病毒起点(如SV40、多瘤病毒、腺病毒、VSV或BPV)可用于在哺乳动物细胞中克隆载体。通常,哺乳动物表达载体不需要复制起始点组分(SV40起始点通常仅因其含有早期启动子而可用)。

[0463] 表达和克隆载体可包含选择基因,也称为可选择标志物。典型的选择基因编码蛋白质,其将(a)赋予对抗生素或其他毒素例如氨基青霉素、新霉素、甲氨蝶呤或四环素的抗性;(b)弥补营养缺陷型缺陷;或(c)提供无法从复合培养基中获得的关键营养素,例如编码杆菌D-丙氨酸消旋酶的基因。

[0464] 表达和克隆载体通常包含启动子,其被宿主生物体识别且操作性地连接至编码本文所述蛋白质的核酸序列,例如基于纤连蛋白的支架蛋白。适用于原核生物宿主的启动子包括phoA启动子、 $\beta$ -内酰胺酶和乳糖启动子系统、碱性磷酸酶、色氨酸(trp)启动子系统和杂交启动子,如tan启动子。然而其它已知的细菌启动子也是合适的。用于细菌系统的启动子通常还含有与编码本文所述蛋白质的DNA可操作地连接的Shine-Dalgarno(S.D.)序列。启动子序列对于真核生物是已知的。几乎所有真核生物基因都有一个富含AT的区域,位于转录起始位点上游约25至30个碱基。在许多基因转录开始处上游70至80个碱基的另一个序列是CNCAAT区,其中N可以是任何核苷酸。在大多数真核生物基因的3'端是AATAAA序列,它可以是将poly A尾部添加到编码序列3'端的信号。所有这些序列都适合插入真核表达载体中。

[0465] 用于酵母宿主的合适启动子序列的例子包括3-磷酸甘油酸激酶或其他糖酵解酶,如烯醇酶,3-磷酸甘油醛脱氢酶,己糖激酶,丙酮酸脱羧酶,磷酸果糖激酶,葡萄糖-6磷酸异构酶,3-磷酸甘油酸突变酶,丙酮酸激酶,三糖磷酸异构酶,磷酸葡萄糖异构酶和葡萄糖激酶的启动子。

[0466] 哺乳动物宿主细胞中载体的转录可以被控制,由例如,从病毒基因组中获得的启动子,例如多瘤病毒、禽痘病毒、腺病毒(如腺病毒2)、牛乳头瘤病毒、禽肉瘤病毒、巨细胞病毒、逆转录病毒、乙肝病毒和最理想的猿猴病毒40(SV40),来自异源哺乳动物的启动子,例如,肌动蛋白启动子或免疫球蛋白启动子、来自热休克启动子,只要这些启动子与宿主细胞系统相容。

[0467] 高等真核生物对本文描述的DNA编码蛋白的转录通常通过将增强子序列插入载体中而增加。现在已经从哺乳动物基因(珠蛋白、弹性蛋白酶、白蛋白、甲胎蛋白和胰岛素)中得知了许多增强子序列。然而,通常人们会使用来自真核细胞病毒的增强子。例子包括复制起点后侧(bp100-270)的SV40增强子、巨细胞病毒早期启动子增强子、复制起点后侧的多瘤增强子和腺病毒增强子。另见Yaniv,Nature,297:17-18(1982)关于真核生物启动子活化的增强元件。增强子可以被剪接到表达载体中编码肽序列的5'或3'的位置,但优选是位于离启动子5'的位置。

[0468] 用于真核宿主细胞(例如酵母、真菌、昆虫、植物、动物、人或来自其它多细胞生物体的有核细胞)的表达载体还将包含用于终止转录和用于稳定化mRNA所必需的序列。这种序列通常可以从真核生物或病毒DNA或cDNA的5',偶尔3'非翻译区获得。这些区域包含在编码本文所述蛋白质的mRNA的非翻译部分中转录为聚腺苷化片段的核苷酸片段。一种有用的转录终止组分是牛生长激素聚腺苷酸化区。参见WO 94/11026和其中公开的表达载体。

[0469] 重组DNA还可以包括可用于纯化蛋白质的任何类型的蛋白质标签序列。蛋白质标

签的例子包括但不限于：组氨酸标签、FLAG标签、Myc标签、HA标签、或GST标签。用于细菌、真菌、酵母和哺乳动物细胞宿主的适当克隆和表达载体可在克隆载体：实验室手册(Elsevier, New York (1985))中找到,其相关公开内容通过引用并入本文。

[0470] 使用适用于宿主细胞的方法将表达构建体引入宿主细胞中,这对于本领域技术人员来说是显而易见的。本领域已知将核酸引入宿主细胞的多种方法,包括但不限于电穿孔;使用氯化钙、氯化铷、磷酸钙、DEAE-葡聚糖或其他物质进行转染;微弹轰击;脂肪感染;和感染(其中载体是传染源)。

[0471] 合适的宿主细胞包括原核生物、酵母、哺乳动物细胞或细菌细胞。合适的细菌包括革兰氏阴性或革兰氏阳性生物,例如大肠杆菌或芽孢杆菌。酵母,优选来自酵母物种,如酿酒酵母(*S. cerevisiae*),也可用于生产多肽。各种哺乳动物或昆虫细胞培养系统也可用于表达重组蛋白。Luckow等人(*Bio/Technology*, 6:47 (1988))综述了昆虫细胞中生产异源蛋白的杆状病毒系统。合适的哺乳动物宿主细胞系的实例包括内皮细胞、COS-7猴肾细胞、CV-1、L细胞、C127、3T3、中国仓鼠卵巢(CHO)、人胚胎肾细胞、HeLa、293、293T和BHK细胞系。通过培养合适的宿主/载体系统以表达重组蛋白来制备纯化的多肽。对于许多应用,本文描述的许多多肽的小尺寸将使大肠杆菌中的表达成为优选的表达方法。然后从培养基或细胞提取物中纯化蛋白质。

[0472] 用于生产本发明蛋白质的宿主细胞可以在多种培养基中培养。市售可得培养基如汉姆氏F10(Ham's F10)(西格玛(Sigma))、最小必需培养基(MEM,西格玛)、RPMI-1640(西格玛)和达氏改良的伊氏培养基(DMEM,西格玛)都适合培养宿主细胞。此外,Ham等人(*Meth. Enzymol.*, 58:44 (1979)), Barites等人(*Anal. Biochem.*, 102:255 (1980)), 美国专利号4,767,704、4,657,866、4,927,762、4,560,655、5,122,469、6,048,728、5,672,502或美国专利号RE 30,985中所述的许多培养基可用作宿主细胞的培养基。任何这些培养基可根据需要补充激素和/或其他生长因子(例如胰岛素、转铁蛋白或表皮生长因子)、盐类(如氯化钠、钙、镁和磷酸盐)、缓冲剂(如HEPES)、核苷酸(如腺苷和胸苷)、抗生素(例如庆大霉素药物)、微量元素(定义为无机化合物,其最终浓度通常在微摩尔范围内)和葡萄糖或等同的能量来源。任何其他必需的补充剂也可以以本领域技术人员已知的适当浓度包含在内。培养条件,如温度、pH值等,是以前用于选择表达的宿主细胞的条件,对普通技术人员来说是显而易见的。

[0473] 本文描述的蛋白质也可以使用无细胞翻译系统生产。为此目的,编码多肽的核酸必须被修饰以允许体外转录产生mRNA,并允许mRNA在所使用的特定无细胞系统(真核生物如哺乳动物或酵母无细胞翻译系统或原核生物如细菌无细胞翻译系统)中进行无细胞翻译。

[0474] 本文描述的蛋白质也可以通过化学合成(例如,通过Solid Phase Peptide Synthesis, 2nd Edition, The Pierce Chemical Co., Rockford, Ill. (1984)中描述的方法)来生产。蛋白质的修饰也可以通过化学合成产生。

[0475] 本发明的蛋白质可以通过蛋白质化学领域中通常已知的蛋白质的分离/纯化方法来纯化。非限制性例子包括抽提、重结晶、盐析(例如,用硫酸铵或硫酸钠)、离心、透析、超滤、吸附色谱、离子交换色谱、疏水色谱、正相色谱、反相色谱、凝胶过滤、凝胶渗透色谱、亲和色谱、电泳、逆流分布或这些的任何组合。纯化后,多肽可通过本领域已知的多种方法中



的任一种交换到不同的缓冲液中和/或浓缩,包括但不限于过滤和透析。

[0476] 纯化的多肽优选纯度至少为85%,优选至少为95%,和最优选为至少为98%。无论纯度的确切数值如何,多肽对于其预期用途而言都足够纯净。

[0477] X. 使用方法

[0478] 本发明的另一方面涉及用于检测例如样品中抗原反应性T细胞的方法。通常,方法包括提供本方面的多个pMHC偶联多聚体;使偶联多聚体与所述样品接触;以及检测所述偶联多聚体与所述样品内的抗原反应性T细胞的结合,从而检测对所述多个偶联多聚体中存在的抗原肽反应的T细胞。在一个实施方式中,通过扩增连接到偶联多聚体的寡核苷酸条码的条码区来检测结合。通常,对于pMHCI偶联多聚体,抗原反应性T细胞是CD8<sup>+</sup>T细胞,其TCR识别肽结合的MHC I类分子,而对于pMHCII偶联多聚体,抗原反应性T细胞为CD4<sup>+</sup>T细胞,其TCR识别肽结合的MHC II类分子。

[0479] 这种偶联多聚体技术允许在单个样品中检测多个(可能>1000个)不同的抗原特异性T细胞。例如,该技术可用于T细胞表位作图、免疫识别发现、诊断测试和测量疫苗接种或免疫相关疗法后的免疫反应性。对于治疗用途,pMHC偶联多聚体允许鉴定和选择用于治疗抗原特异性T细胞,例如用于过继性T细胞转移治疗。

[0480] A. 测定

[0481] 在本发明的一个实施方式中,MHC多聚体可用于使用流式细胞术或流式细胞术样分析检测流体样品中的单个T细胞。

[0482] 液体细胞样品可以使用流式细胞仪进行分析,流式细胞分析仪能够检测和计数通过激光束的流中的单个细胞。为了使用MHC多聚体鉴定特定的T细胞,通过用MHC多聚体孵育细胞,用荧光标记的MHC多聚体对细胞进行染色,然后用大量液体迫使细胞通过喷嘴,形成间隔细胞流。每个细胞通过激光束,与细胞结合的任何荧光色素都被激发,从而发出荧光。灵敏的光电倍增管检测发射的荧光,提供有关与细胞结合的MHC多聚体的数量信息。通过该方法,MHC多聚体可用于鉴定液体样品中的单个T细胞和/或特定T细胞群。

[0483] 能够在流式细胞术分析中通过MHC多聚体分析的细胞样品包括但不限于血液样品或其部分、T细胞系(杂交瘤、转染细胞)和匀浆组织,如脾脏、淋巴结、肿瘤、脑或任何其他包含T细胞的组织。

[0484] 分析血样时,在流式细胞仪分析之前,可以使用红细胞经或不经裂解的全血。裂解试剂可以在用MHC多聚体染色之前或之后添加。当在不裂解红细胞的情况下分析血液样品时,可以包括一种或多种门控试剂以区分淋巴细胞和红细胞。优选的门控试剂是对红细胞表面蛋白特异的标志物分子,能够从样品的剩余细胞中扣除该细胞群。例如,荧光标记的CD45特异性标志物分子(例如抗体)可用于设置触发区分因子,以允许流式细胞仪区分红细胞和染色的白细胞。

[0485] 作为全血分析的替代方案,可在流式细胞术分析之前纯化淋巴细胞,例如使用标准程序,如aFICOLL®-Hypaque梯度。另一种可能性是从血液样品中分离T细胞,例如,通过将样品添加到固定在固相载体上的抗体或其他T细胞特异性标记物中。然后标志物特异性T细胞将附着到固相载体上,洗涤后可洗脱特异性T淋巴细胞。然后,该纯化的T细胞群可以与MHC多聚体一起用于流式细胞术分析。

[0486] T细胞也可以通过玫瑰花结(rosetting)从其他淋巴细胞或血细胞中纯化。人T细



胞与绵羊红细胞形成自发的玫瑰花结,也称为E-玫瑰花结形成。可通过将淋巴细胞与绵羊红细胞孵育,然后在密度梯度(如FICOLL®Hypaque梯度)上纯化,形成E-玫瑰花结。

[0487] 可以在分析之前去除B细胞、NK细胞或其他细胞群等不需要的细胞,而不是主动分离T细胞。去除不需要细胞的优选方法是将样品与固定在固相载体上对不需要细胞上的一种或多种表面蛋白特异性的标志物分子一起孵育。一个实例包括使用包被了抗体或对不需要的细胞上的表面受体特异性的其他标志物分子的珠,例如针对CD19、CD56、CD14、CD15或其他标志物。简单说,将涂有特定表面标志物的珠添加到细胞样品中。不同于具有适当表面受体的所需T细胞的细胞将与珠结合。通过例如离心或磁提取(当使用磁珠时)去除珠,并富集剩余细胞中的T细胞。

[0488] 另一个例子是使用柱亲和层析,柱材料上涂覆有抗体或其他对不需要的细胞特异性的标志物。

[0489] 或者,可以将特异性抗体或标志物与补体一起添加到血液样品中,从而杀死抗体或标志物识别的细胞。

[0490] 分析中可以包括各种门控试剂。本文的门控试剂是指通过结合独特的表面蛋白或胞内组分或胞内分泌组分来识别细胞亚群的标记的抗体或其他标记的标志物分子。当使用MHC多聚体时,优选的门控试剂是针对CD2、CD3、CD4和CD8的抗体和标志物分子,用于识别T细胞的主要亚群。其他优选的门控试剂是针对CD11a、CD14、CD15、CD19、CD25、CD30、CD37、CD49a、CD49e、CD56、CD27、CD28、CD45、CD45RA、CD45RO、CD45RB、CCR7、CCR5、CD62L、CD75、CD94、CD99、CD107b、CD109、CD152、CD153、CD154、CD160、CD161、CD178、CDw 197、CDw 217、Cd229、CD245、CD247、Foxp3的抗体和标志物,或识别不同淋巴细胞、淋巴细胞群或其他细胞群独特的特定蛋白质的其他抗体或标志物分子。还包括针对白介素的抗体和标志物,例如IL-2、IL-4、IL-6、IL-10、IL-12、IL-21;干扰素,如INF  $\gamma$ 、TNF $\alpha$ 、TNF $\beta$ ,或其他细胞因子或趋化因子。

[0491] 可以在向样品中添加MHC多聚体之前、之后或同时添加门控试剂。在用MHC多聚体标记之后和在流式细胞仪上分析之前,可以用固定试剂(例如甲醛、乙醇或甲醇)处理染色的细胞,以将结合的MHC多聚体交联到细胞表面。还可直接分析染色细胞而无需固定。

[0492] 在一个实施方式中,流式细胞仪可以配置为分离和收集特定类型的细胞。这叫做细胞分选。MHC多聚体结合流式细胞仪分选可用于分离抗原特异性T细胞群。如上所述的门控试剂可以包括进一步指定待分离的T细胞群。分离和收集的特异性T细胞群然后可以如本文他处所述进一步操作,例如在体外扩增。

[0493] 样品中MHC肽特异性T细胞浓度的直接测定可以通过用MHC多聚体和相关门控试剂对血细胞或其他细胞样品染色,然后加入精确量的已知浓度的计数珠来获得。一般来说,计数珠是具有散射特性的微粒,当通过流式细胞仪进行记录时,这些微粒将其置于感兴趣细胞的环境中。它们可以用抗体、荧光色素或其他标志物分子标记,也可以不标记。在一些实施方式中,珠是聚苯乙烯珠,其中分子嵌入聚合物,在流式细胞仪的大多数通道中是荧光的。在此,术语“计数珠”和“微粒”可以互换使用。

[0494] 适用的珠或微粒包括用于凝胶色谱的珠,例如凝胶过滤介质,如SEPHADEX®。此类合适的微珠包括但不限于珠粒径为40-120 $\mu$ m的SEPHADEX®G-10(SigmaAldrich目录号27,103-9)、珠粒径为40-120 $\mu$ m的SEPHADEX®G-15(Sigma-Aldrich目录号27,104-

7)、珠粒径为20-50 $\mu\text{m}$ 的SEPHADEX®G-25 (Sigma-Aldrich目录号27,106-3)、珠粒径为20-80 $\mu\text{m}$ 的SEPHADEX®G-25 (Sigma-Aldrich目录号27,107-1)、珠粒径为50-150 $\mu\text{m}$ 的SEPHADEX®G-25 (Sigma-Aldrich目录号27,109-8)、珠粒径100-300 $\mu\text{m}$ 的SEPHADEX®G-25 (Sigma-Aldrich产品目录号27,110-1)、珠粒径20-50 $\mu\text{m}$ 的SEPHADEX®G-50 (Sigma Aldrich目录号27,112-8)、珠粒径20-80 $\mu\text{m}$ 的SEPHADEX®G-50 (Sigma Aldrich目录号27,113-6)、珠粒径为50-150 $\mu\text{m}$ 的SEPHADEX®G-50 (Sigma-Aldrich产品目录号27,114-4)、珠粒径为100-300 $\mu\text{m}$ 的SEPHADEX®G-50 (Sigma-Aldrich目录号27,115-2)、珠粒径为20-50 $\mu\text{m}$ 的SEPHADEX®G-75 (Sigma-Aldrich目录号27,116-0)、珠粒径为40-120 $\mu\text{m}$ 的SEPHADEX®G-75 (Sigma-Aldrich目录号27,117-9)、珠粒径为20-50 $\mu\text{m}$ 的SEPHADEX®G-100 (Sigma-Aldrich产品目录号27,118-7)、珠粒径为40-120 $\mu\text{m}$ 的SEPHADEX®G-100 (Sigma Aldrich商品目录号27,119-5)、珠粒径为40-120 $\mu\text{m}$ 的SEPHADEX®G-150 (Sigma Aldrich商品目录号27,121-7)、和珠粒径为40-120 $\mu\text{m}$ 的SEPHADEX®G-200 (SigmaAldrich目录号27,123-3)。

[0495] 用于本文所述方法和组合物的其它优选颗粒包括塑料微珠。虽然塑料微珠通常是固体,但它们内部也可能是中空的,可以是囊泡和其他微载体。它们不必是完美的球体才能在这里描述的方法中发挥作用。塑料材料如聚苯乙烯、聚丙烯酰胺和其他乳胶材料可用于制造珠,但也可使用其他塑料材料如聚氯乙烯、聚丙烯等。

[0496] 计数珠用作参考群体,以测量分析样品的准确体积。在流式细胞仪上分析样品,并使用例如预定义的门控策略确定MHC特异性T细胞的数量,然后将该数量与相同样品中已计数的计数珠的数量相关联。

[0497] 还可以使用对上调或下调的表面暴露受体具有特异性的标记分子以及MHC多聚体来进行样品中特异性T细胞的检测以及同时检测T细胞的活化状态。标志物分子和MHC多聚体可以用相同的标记分子或不同标记分子标记,同时或依次或单独添加到样品中。

[0498] 1. 用显微镜检测液体样品中的单个T细胞

[0499] 检测流体样品中单个T细胞的另一优选方法是使用显微镜。显微镜包括任何类型的显微镜,包括光学、电子和扫描探针显微镜、亮场显微镜、暗场显微镜、相差显微镜、差分干涉对比显微镜、荧光显微镜、共焦激光扫描显微镜、X射线显微镜、透射电子显微镜、扫描电子显微镜、原子力显微镜、扫描隧道显微镜和光子力显微镜。这可如下进行:将T细胞悬液添加到MHC多聚体中,洗涤样品,然后测量与每个细胞结合的MHC多聚体的量。结合的MHC多聚体可以直接标记或通过添加标记的标志物分子来测量。然后将样品铺展在载玻片或类似能够区分单个细胞的薄层中,使用显微镜鉴定标记的细胞。根据标记的类型,可以使用不同类型的显微镜,例如,如果使用荧光标记,则使用荧光显微镜进行分析。例如,MHC多聚体可以用荧光色素标记或用荧光抗体检测结合的MHC多聚体。然后可以使用例如免疫荧光显微镜或共焦荧光显微镜观察具有结合的荧光MHC多聚体的细胞。

[0500] 2. 免疫组织化学(IHC)

[0501] IHC是一种可以使用MHC多聚体直接检测特定T细胞的方法,例如在固体组织切片中。在一些实施方式中,固定或冷冻组织样品的切片与MHC多聚体一起孵育,使得MHC多聚体结合组织中的特异性T细胞。MHC多聚体可以用荧光色素、发色团或任何其他可检测的标记

分子标记。MHC多聚体的标记可以直接或通过第二标志物分子进行。例如，MHC多聚体可以用可被例如二抗识别的标签标记，该二抗任选地用HRP或另一标记标记。然后通过其荧光或吸光度（对于荧光团或发色团），或通过添加针对该标签的酶标记抗体或MHC多聚体的另一组分（例如，一条蛋白链，一个或多个多聚结构域上的标记）来检测结合的MHC多聚体。该酶例如可以是辣根过氧化物酶（HRP）或碱性磷酸酶（AP），这两者都将无色底物原位转化为有色反应产物。该有色沉积物鉴定MHC多聚体的结合位点，并可在例如光显微镜下观察。MHC多聚体也可以用例如HRP或AP直接标记，并在不需要额外抗体的情况下用于IHC。

[0502] 在一些实施方式中，检测固体组织中的T细胞包括使用石蜡包埋的组织，从其制作组织切片并在染色前固定在福尔马林中。抗体是用于福尔马林固定组织切片染色的标准试剂；这些抗体通常识别线性表位。相比之下，大多数MHC多聚体有望识别TCR上的构象表位。在这种情况下，TCR的天然结构需要至少部分保存在固定组织中。

[0503] 在其他实施方式中，染色从冷冻组织块获得的组织切片。在这种类型的染色中，在MHC多聚体染色后进行固定。

#### [0504] 3. 免疫荧光显微镜检

[0505] 在一些实施方式中，MHC多聚体可用于鉴定固体组织切片中的特异性T细胞。除了通过酶促反应显示结合的MHC多聚体，用荧光色素标记MHC多聚体或用荧光抗体检测结合的MHC多聚体。可在免疫荧光显微镜或共焦荧光显微镜中观察具有结合的荧光MHC多聚体的细胞。该方法还可用于使用本文别处描述的用于检测流体样品中的T细胞的原理检测流体样品的T细胞。

#### [0506] 4. 体内固态组织中的T细胞检测

[0507] MHC多聚体也可用于体内检测固体组织中的T细胞。为了体内检测T细胞，将标记的MHC多聚体注射到待研究个体体内。MHC多聚体可以用例如顺磁同位素标记。使用磁共振成像（MRI）扫描仪或电子自旋共振（ESR）扫描仪，可以测量和定位结合T细胞的MHC多聚体。通常，可以使用任何用于诊断成像可视化的常规方法。通常伽马和正电子发射放射同位素用于摄像，顺磁同位素用于MRI。

#### [0508] 5. 检测固定在固相载体上的T细胞。

[0509] 在许多应用中，将T细胞固定在固相或半固相载体上可能是有利的。这种载体可以是任何适合于固定化、分离等的载体。非限制性实例包括颗粒、珠、可生物降解颗粒、片材、凝胶、滤器、膜（例如尼龙膜）、纤维、毛细管、针、微量滴定条、管、平板或孔、梳、移液管尖端、微阵列、芯片、载玻片或任何固体表面材料。如果需要，可以标记固相或半固相载体。载体还可以具有散射特性或尺寸，这使得能够在相同性质的载体之间进行区分，例如不同尺寸或散射特性、颜色或强度的颗粒。

[0510] MHC多聚体可用于检测固定化T细胞的方法的一个例子是ELISA（酶联免疫吸附试验）。ELISA是最初用于检测抗体-抗原相互作用的结合试验。检测基于酶反应，常用的酶是例如HRP和AP。MHC多聚体可用于基于ELISA的测定，用于分析固定在微量滴定板孔中的纯化TCR和T细胞。结合的MHC多聚体可以通过例如HRP或AP与MHC多聚体（例如一个或多个多聚结构域或MHC蛋白）的直接化学偶联，或者例如通过HRP或AP-偶联抗体或与MHC多聚体结合的其他标志物分子来标记。然后通过添加被HRP或AP酶转化为可检测产物（例如有色）的底物（例如无色的）来检测酶标记。

[0511] 固相载体可以由例如玻璃、二氧化硅、乳胶、塑料或任何聚合物材料制成。载体也可以由生物降解材料制成。一般来说,载体的性质并不重要,可以使用多种材料。载体的表面可以是疏水的或亲水的。非磁性聚合物珠也可适用。这类产品可从各种制造商获得,例如 Dynal Particles AS、Qiagen、Amersham Biosciences、Serotec、Seradyne、Merck、Nippon Paint、Chemagen、Promega、Prolabo、Polysciences、Agowa和Bangs Laboratories。

[0512] 合适载体的另一个例子是磁珠或颗粒。本文中处处使用的术语“磁性”意指载体在置于磁场中时能够具有施加给它的磁矩,因此在磁场的作用下可移动。换言之,包含磁珠或颗粒的载体可以通过磁性聚集容易地去除,这提供了从溶液中分离出磁珠或颗粒的快速、简单和有效的方法。磁珠和颗粒可以适当地是顺磁性或超顺磁性的。超顺磁性珠和颗粒例如描述于EP 0 106 873中。磁珠和颗粒可从几家厂商获得,例如Dynal Biotech ASA(挪威奥斯陆,以前为Dynal AS,例如DYNABEADS.RTM)。

[0513] 6. 微芯片MHC多聚体技术

[0514] 通过将不同的MHC多聚体固定在固体载体上,可以形成MHC多聚体的微阵列,以形成空间阵列,其中位置指定了固定在该位置的MHC-肽复合物或特定空MHC的身份。当标记的细胞通过微阵列(例如血细胞)时,微阵列中携带MHC多聚体特异性TCR的细胞将被固定。因此,标记将位于微阵列的特定区域,这将允许识别结合细胞的MHC多聚体,从而允许识别例如对固定的MHC多聚体具有识别特异性的T细胞。或者,可以在细胞与MHC多聚体结合后对其进行标记。该标记可以是预期结合MHC多聚体的细胞类型的特异性标记,或者该标记可以对细胞进行一般染色(例如,结合DNA的标记)。或者,细胞因子捕获抗体可以与MHC一起在固体支持物上共同形成斑点(co-spotted),分析结合的抗原特异性T细胞的细胞因子分泌。这是可能的,因为当识别和结合特定MHC-肽复合物时,T细胞被刺激分泌细胞因子。

[0515] 7. 用pMHC多聚体间接检测T细胞

[0516] 也可以使用MHC多聚体间接检测样品中的T细胞。在间接检测中,通过检测TCR-MHC-肽复合物相互作用的结果来测量T细胞的数量或活性。MHC多聚体和T细胞之间的相互作用可能刺激T细胞,导致T细胞的活化,T细胞群的细胞分裂和增殖,或者导致T细胞失活。所有这些机制都可以使用能够检测这些事件的检测方法来测量。

[0517] 活化的实例测量包括特异性可溶性因子例如细胞因子分泌的测量,例如可通过流式细胞术章节所述的流式细胞术,通过例如流式细胞术和/或ELISA-样方法测量活化标志物表达,如CD27和CD28和/或其他受体的表达,以及测量T细胞效应功能,例如CD8 T细胞的细胞杀伤力,其可用细胞杀伤试验如铬释放试验测量,这是本领域技术人员已知的。

[0518] 增殖的示例性测量包括但不限于mRNA的测量、胸苷的掺入或其他分子如溴-2'-脱氧尿苷(BrdU)掺入的测量。

[0519] T细胞失活的示例性测量包括但不限于特异性TCR阻断作用的测量和细胞凋亡的测量。

[0520] 当与不同的T细胞群接触时,如对象外周血淋巴细胞(PBL)样品中所含的T细胞,样品中T细胞识别的那些含有pMHC的四聚体将与匹配的T细胞结合。使用荧光流式细胞术分析反应内容物,以确定、定量和/或分离那些结合有MHC四聚体的T细胞。

[0521] B. 筛选

[0522] 本发明的偶联多聚体可用于各种不同的筛选测定。例如,在一个实施方式中,在诱

导占位符肽释放和抗原衍生肽结合的条件下,将衍生自一种或多种抗原的荧光标记肽文库施用于包含占位符肽的pMHC多聚体。通过荧光偏振分析监测肽交换。占位符肽的使用允许在生理条件下产生空的肽受体MHC多聚体。这种筛选方法可用于鉴定与MHC分子结合的肽配体。肽交换反应可以在天然条件下以多孔形式进行。结合可由多种技术确定,如ELISA,其监测MHC结构的稳定性,或通过生物物理技术监测肽结合,如荧光偏振。这种筛选方法还可用于扫描肽组(例如源自病原体基因组、肿瘤相关抗原或自身免疫抗原的肽组)以寻找MHC配体。

[0523] 本文公开的pMHC偶联多聚体及其文库可用于多种筛选方法中,这些方法允许方便地检测和定量与免疫细胞受体的抗原特异性结合。例如,偶联多聚体文库可允许检测对特定抗原特异性的T细胞,多重检测给定样品中的T细胞特异性,将TCR序列与特异性匹配(例如,通过单细胞测序),比较性TCR亲和力测定,确定给定TCR的共有特异性序列,或针对感兴趣的序列绘制T细胞的抗原反应性。偶联多聚体还可用于检测携带特定MHC I多肽特异性受体的自然杀伤(NK)细胞。

[0524] 所得的pMHC偶联多聚体文库可用于T细胞筛选以确定抗原反应性T细胞,如Simon等人,Cancer Immunol Res,2014,2(12):1230-1244中所述。

[0525] 在一些实施方式中,本公开提供了一种用于分离表达TCR的细胞-pMHC对的方法,包括使多个表达TCR的细胞与本文所述的pMHC多聚体文库接触;产生多个隔室,其中所述多个隔室内的一个隔室包括与所述文库的pMHC结合的多个表达TCR的细胞中的一个表达TCR的细胞,从而在所述隔室中分离所述表达TCR的细胞-pMHC对。在一些实施方式中,表达TCR的细胞是T细胞,例如,当使用pMHCI多聚体文库时是CD8+T细胞,或者当使用pMHC II多聚体文库时,是CD4+T细胞。在一些实施方式中,可转染或转导细胞以表达TCR。在一些实施方式中,可转染或转导非淋巴细胞以表达TCR。

#### [0526] C. 鉴定方法

[0527] 本发明的pMHC偶联多聚体可用于鉴定感兴趣的抗原特异性T细胞,例如通过用pMHCI偶联多聚体文库筛选多个T细胞。在各种实施方式中,所述文库包括pMHC偶联多聚体,其加载了超过10、超过100、超过500、1000、超过2000、超过5000、超过10,000、超过 $10^6$ 、超过 $10^7$ 、超过 $10^8$ 、超过 $10^9$ 或超过 $10^{10}$ 种独特肽的多样性。该鉴定方法可以包括在单个隔室中将结合到文库的pMHC偶联多聚体的多个细胞中的一个细胞隔室化,其中pMHC偶联多聚体包括独特标识符;以及确定结合到隔室化细胞的每个pMHC偶联多聚体的独特标识符。隔室可以是单独的空间,例如孔、平板、分离的边界、相漂移、容器、囊泡、细胞等。

[0528] 在一些实施方式中,本文公开的组合物和方法可用于鉴定结合到TCR的多种肽。在一些实施方式中,本文公开的组合物和方法可用于鉴定结合到pMHC的多个TCR。在一些实施方式中,本文公开的组合物和方法可用于鉴定结合多个pMHC的多个TCR(例如,结合衍生自病原体文库、癌症文库或自身免疫文库的pMHC多聚体的多个TCR)。

[0529] 在一些实施方式中,本文公开的组合物和方法用于鉴定TCR-抗原特异性。

[0530] 在一些实施方式中,所选T细胞上TCR的身份通过测序确定(例如,对TCR的可变区、高变区或互补决定区(CDR)进行测序)。在一些实施方式中,结合的pMHC的肽(与TCR结合)的身份通过测序(例如,使用本文所公开的标识符)确定。

[0531] 在一个实施方式中,本发明的pMHC偶联多聚体可用于通过流式细胞术检测抗原特

异性T细胞,或可用于T细胞纯化。本发明的组合物和方法允许产生非常大的加载肽的MHC多聚体集合,当使用pMHCI多聚体时,该多聚体非常适合于快速鉴定细胞毒性T细胞(即CD8+T细胞)抗原,当使用MHCII多聚体的时候,该多聚体非常适合快速鉴定辅助T细胞(如CD4+T细胞)。

[0532] 在一个实施方式中,附着于固相表面的pMHC偶联多聚体可用于检测T细胞功能。固定在固体表面的肽-MHC抗原复合物可以通过TCR刺激T细胞活性,从而允许研究TCR刺激后的下游T细胞功能。

[0533] 在一些实施方式中,本文公开的组合物和方法用于确定经鉴定的MHC结合肽中的突变如何影响TCR结合。在一些实施方式中,本文公开的组合物和方法用于识别经鉴定的MHC结合肽中导致增强或降低TCR结合亲和力的突变。在一些实施方式中,本文公开的组合物和方法用于识别经鉴定的MHC-结合肽中保持TCR结合亲和力的突变。在一些实施方式中,本文公开的组合物和方法用于识别经鉴定的MHC结合肽中导致TCR结合亲和力丧失的突变。

[0534] 在一些实施方式中,本文公开的组合物和方法用于确定使用本文描述的方法鉴定的TCR中的突变如何改变肽表位的结合。在一些实施方式中,本文公开的组合物和方法用于鉴定TCR中导致肽表位结合亲和力降低或增加的突变。在一些实施方式中,本文公开的组合物和方法可用于鉴定TCR中保留肽表位结合的突变。在一些实施方式中,本文公开的组合物和方法可用于鉴定TCR中导致肽表位结合丧失的突变。

[0535] 在一些实施方式中,本文公开的方法在来自多个对象的T细胞上进行。在一些实施方式中,分析来自多个对象的数据允许鉴定由多个对象识别的MHC-结合肽表位。在一些实施方式中,分析来自多个对象的数据允许鉴定由多个TCR克隆型识别的MHC-结合肽表位。在一些实施方式中,分析来自多个对象的数据允许鉴定由多个患者识别的MHC结合肽表位,例如,多个癌症患者、多个具有自身免疫状况的患者或具有针对病原体的保护性免疫的多个患者。在一些实施方式中,分析来自多个对象的数据允许鉴定包括不同HLA型或等位基因的对象中识别的MHC结合肽表位。在一些实施方式中,分析来自多个对象的数据允许鉴定显示趋集型抗原结合的TCR的不同高变或互补决定区序列。

[0536] 在一些实施方式中,本文公开的方法使用多个文库来执行。在一些实施方式中,分析来自多个文库的数据允许鉴定文库之间共有的反应性MHC结合肽表位,例如,呈现TCR亲和力的抗原,其存在于多个病原体菌株、多个癌症类型、多个癌症患者、多个自身免疫疾病或多个自身免疫病况中。在一些实施方式中,分析来自多个文库的数据允许在文库中鉴定不同的反应性MHCI结合肽表位,例如存在于病原体菌株、癌症、病况或患者的子集中的抗原。

[0537] 在一些实施方式中,使用本发明的pMHC偶联多聚体文库鉴定的T细胞进行基因表达分析(例如,RNA-seq、qPCR)。在一些实施方式中,对在本发明文库中被鉴定为具有对肽表现出特异性的受体的细胞进行基因表达分析。例如,确定表达与病原体文库、癌症文库或自身免疫文库衍生的pMHC偶联多聚体结合的TCR的细胞,进行基因表达分析。基因表达分析可以是全局性的,也可以是对性的。分析表达的基因包括但不限于具有已知功能的基因、编码免疫效应分子(例如穿孔素、颗粒酶、细胞因子、趋化因子)、免疫检查点分子、促炎分子、抗炎分子、谱系标志物、整合素、选择素、淋巴细胞记忆标志物、死亡受体、胱冬酶、细胞周期检查点分子、酶、磷酸酶、激酶、脂肪酶的基因,和代谢基因。

[0538] 在一些实施方式中,基因表达分析可与pMHC偶联多聚体文库筛选同时进行。在一些实施方式中,基因表达分析可在pMHC偶联多聚体文库筛选结果分析之后进行。在一些实施方式中,基因表达分析可在pMHC偶联多聚体文库筛选结果分析之前进行。在一些实施方式中,基因表达分析允许对使用本文所述方法产生的pMHC-T细胞受体配对中鉴定出的感兴趣的细胞进行免疫分型。

[0539] 本文所述的方法和组合物可用于筛选分析。例如,将包含本文所述的多个pMHC偶联多聚体的文库与T细胞样品接触,并测定一种或多种T细胞功能,包括但不限于T细胞增殖、T细胞毒性、抑制T细胞增殖,T细胞抑制和T细胞的细胞因子产生。

[0540] 在一些实施方式中,可以诱导功能特性的pMHC偶联多聚体可制成肽文库子集。例如,文库子集可包含pMHC偶联多聚体,其在与TCR结合时诱导T细胞增殖,在与TCR结合时诱导细胞毒性,在与TCR结合时抑制T细胞,在与TCR结合时T细胞引起抑制,在与TCR结合时产生细胞因子,或其任何组合。例如,可通过染料稀释试验(例如,CFSE稀释试验)或DNA复制定量(例如,BrdU掺入试验)来确定增殖。细胞毒性可通过例如基于死亡细胞释放细胞内酶(例如乳酸脱氢酶)的分析,染料排阻分析(例如碘化丙啶),或细胞裂解标志物(例如颗粒酶、CD107a)的表达,通过流式细胞术或qPCR来确定。例如,可以通过ELISA、多重免疫分析、胞内细胞因子染色、ELISPOT、Western印迹法或qPCR来确定细胞因子的产生。例如,可通过将T细胞克隆与效应细胞和靶抗原共同孵育,测定增殖、细胞毒性、细胞因子产生、活化标志物的表达等来确定T细胞抑制。

[0541] 在一些实施方式中,本文公开的组合物和方法用于鉴定与保护性免疫、非保护性免疫或自身免疫相关的抗原特异性T细胞效应子克隆。在一些实施方式中,本文公开的组合物和方法用于鉴定抗原特异性T细胞效应子克隆,其表现出无反应性、衰竭、耐受性、自身免疫性、炎症性或抗炎性(例如,Treg)。在一些实施方式中,本文所公开的组合物和方法用于鉴定表现出某些效应子或记忆特性(例如,未成熟、末端效应子、效应子记忆、中央记忆、驻留记忆、 $T_H1$ 、 $T_H2$ 、 $T_H17$ 、 $T_H9$ 、 $T_C1$ 、 $T_C2$ 、 $T_C17$ 、某些细胞因子的产生)的抗原特异性T细胞效应子克隆。

[0542] 在一些实施方式中,使用本文公开的组合物和方法鉴定的TCR用作治疗干预的一部分。例如,将TCR序列、TCR可变区序列或CDR序列转染或转导到T细胞中以产生具有相同抗原特异性的修饰的T细胞。修饰的T细胞可以扩增、极化为所需的效应子表型(例如 $T_H1$ 、 $T_C1$ 、Treg),并注入对象体内。在一些实施方式中,使用本文公开的组合物和方法鉴定的多个TCR用于寡克隆治疗。

[0543] 在一些实施方式中,使用本文公开的方法鉴定的肽、配体、激动剂、拮抗剂、抗原或表位用作治疗干预的一部分。在一些实施方式中,肽、抗原或表位用于例如使用抗原递呈细胞、人工抗原递呈细胞、固定化肽或可溶性肽离体扩增细胞群。在一些实施方式中,将扩增的细胞注入患者体内。在一些实施方式中,扩增外周血淋巴细胞。在一些实施方式中,扩增肿瘤浸润性淋巴细胞(TIL)。在一些实施方式中,扩增 $T_H1$ 细胞。在一些实施方式中,扩增细胞毒性T淋巴细胞。在一些实施方式中,扩增调节性T细胞。

[0544] 在一些实施方式中,本文公开的组合物和方法用于鉴定用于疫苗开发的MHC-结合的抗原肽,例如亚单位疫苗、引发覆盖一系列保护性抗原的疫苗,或通用疫苗。

[0545] 在一些实施方式中,本文公开的组合物和方法可用于诊断医学病况。在一些实施



方式中,本文公开的组合物和方法用于指导临床决策制定,例如,治疗选择、预后因素的鉴定、治疗反应或疾病进展的监测,或预防措施的实施。

[0546] 在一些实施方式中,本文公开的组合物和方法可用于选择和/或设计医疗条件的治疗,特别是用于选择抗原特异性T细胞(例如CD8+细胞毒性T细胞和/或CD4+辅助性T细胞)或由此衍生的TCR,用于过继转移T细胞治疗。例如,pMHC偶联多聚体可用于鉴定患者样品中与感兴趣抗原(如癌症抗原或病原体抗原)反应的T细胞,从而选择这些细胞进行体外扩增,然后重新引入患者体内。此外,从这种抗原特异性T细胞鉴定的TCR可以被测序并重组导入T细胞中,以增加表达与患者中具有治疗意义的抗原结合的TCR的细胞群。

[0547] XI. 组合物和试剂盒

[0548] 在另一方面,本公开包括用于本文所述方法的组合物和试剂盒。在一个实施方式中,本公开提供了pMHC偶联多聚体组合物。在一个实施方式中,pMHC偶联多聚体是pMHC偶联四聚体。在一个实施方式中,四聚体的多聚结构域是链霉亲和素或亲和素。在一个实施方式中,pMHC偶联四聚体包括四个MHC单体,它们在链霉亲和素或亲和素的生物素结合位点以外的位点与链霉亲和素或亲和素共价偶联。在一个实施方式中,四个MHC单体各自包含(即加载有)MHC结合肽,其中每个单体包含相同的MHC结合肽。在一个实施方式中,MHC偶联四聚体还包括结合到链霉亲和素或亲和素的生物素结合位点的生物素化寡核苷酸条码。在一个实施方式中,pMHC偶联多聚体(例如四聚体)是pMHC I类偶联多聚体(例如四聚体)。在另一个实施方式中,pMHC偶联多聚体(例如四聚体)是pMHC II类偶联多聚体(例如四聚体)。

[0549] 在一个实施方式中,本发明包括包含多种pMHC偶联多聚体组合物的试剂盒。在一个实施方式中,多个pMHC偶联多聚体中的每一个都是pMHC偶联四聚体。在一个实施方式中,每个四聚体的多聚结构域是链霉亲和素或亲和素。在一个实施方式中,每个pMHC偶联四聚体包括四个MHC单体,它们在链霉亲和素或亲和素的生物素结合位点以外的位点与链霉亲和素或亲和素共价偶联。在一个实施方式中,四个MHC单体各自包含MHC-结合肽,其中每一个四聚体内的每个MHC单体都包含(即,加载有)相同的MHC结合肽,并且其中多个MHC偶联四聚体中的每个MHC-偶联四聚体包含(即加载有)不同MHC-结合肽,从而形成MHC结合肽的文库。在一个实施方式中,文库内的每个MHC偶联四聚体还包括结合到链霉亲和素或亲和素的生物素结合位点的生物素化寡核苷酸条码。在一个实施方式中,多个pMHC偶联多聚体(例如四聚体)中的每一个是pMHC I类偶联多聚体(例如四聚体)。在另一个实施方式中,多个pMHC偶联多聚体(例如四聚体)中的每一个是pMHC II类偶联多聚体(例如四聚体)。

[0550] 实施例

[0551] 下面是用于实施本发明的具体实施方式的示例。所提供的实例仅用于说明目的,并不旨在限制本发明的范围。

[0552] 除非另有说明,本发明的实践将采用本领域技术范围内的蛋白质化学、生物化学、重组DNA技术和药理学的常规方法。这些技术在文献中已有充分描述。见例如T.E.Creighton,Proteins:Structures and Molecular Properties(W.H.Freeman and Company,1993);A.L.Lehninger,Biochemistry(Worth Publishers,Inc.,current addition);Sambrook,等人Molecular Cloning:A Laboratory Manual(2nd Edition,1989);Methods In Enzymology(S.Colowick和N.Kaplan编,Academic Press,Inc.);Remington's Pharmaceutical Sciences,18版(Easton,Pennsylvania:Mack Publishing



Company, 1990); Carey and Sundberg Advanced Organic Chemistry 第三版. (Plenum Press) A卷和B卷(1992)。

[0553] 除非另有说明,所有试剂和化学品均从商业来源获得,并在未经进一步纯化的情况下使用。

[0554] 实施例1-产生具有分选酶标签的可交换肽MHC I类多聚体

[0555] 在本实施例中,表达MHC I重链并与 $\beta$ 2-微球蛋白( $\beta$ 2m)和可交换肽复合,使得MHC重链包含C末端分选酶标签,该标签能够实现与链霉亲和素(SAv)的翻译后偶联,形成可条码化、可交换的MHC I四聚体。MHC I重链和SAv与C末端分选酶标签表达(其氨基酸序列分别显示在SEQ ID NO:1和3中)。然后使用分选酶(具有SEQ ID NO:6中所示氨基酸序列)将GGG-X点击手柄肽偶联到MHC I或GGG-Y点击手柄肽偶联到SAv,其中点击手柄肽包含点击部分,如炔烃(X)或叠氮化物(Y),反之亦然。随后通过铜辅助的炔烃-叠氮环加成或无铜的炔烃-叠氮环加成将MHC I与SAv化学偶联,从而产生加载可交换肽的MHC I四聚体。

[0556] HLA和 $\beta$ 2m表达和重折叠。生成了编码与Sorttag(分选标签)连接的HLA-A\*02:01的细菌表达质粒,本文称为HLA-A2-分选标签(包含C末端分选酶标签,6x-His标签)(其氨基酸序列如SEQ ID NO:1所示)和 $\beta$ 2m(其氨基酸序列如SEQ ID NO:2所示)。在大肠杆菌内的包涵体内表达HLA-A2-分选标签和 $\beta$ 2m。纯化包涵体并溶解在尿素缓冲液(20mM MES, pH 6.0, 8mM 尿素, 10mM EDTA)中,该缓冲液含有1mM或0.1mM DTT(对于HLA-A2-分选标签)或0.1mM DTT(对于 $\beta$ 2m)。化学合成了UV不稳定占位符肽(GILGFVFLJ(SEQ ID NO:7),其中J是3-氨基-3-(2-硝基)苯基丙酸)。根据先前描述的方案(Garboczi等人,PNAS, 89:3429-3433 1992; Rodenko等人,Nat Protoc., 1:1120-322006),用 $\beta$ 2m和占位符肽对HLA-A2进行了重折叠,并进行了微小修饰。简言之,在搅拌下将以下组分按以下顺序以所示最终浓度添加到预冷的重折叠缓冲液(100mM Tris, pH 8.0, 0.4M精氨酸-HCl, 2mM EDTA, 5mM还原的谷胱甘肽, 0.5mM氧化的谷胱甘肽, 0.2mM PMSF)中:肽(45 $\mu$ M),  $\beta$ 2m(3 $\mu$ M)和HLA-A2-分选标签(1.5 $\mu$ M)溶解的包涵体。重折叠反应在搅拌下于4 $^{\circ}$ C孵育过夜。次日,将 $\beta$ 2m和HLA-A2-分选标签溶解的包涵体分别加入重折叠反应中,最终浓度分别为6 $\mu$ M和3 $\mu$ M。第4天,通过离心,然后通过0.2 $\mu$ m滤器过滤,澄清重折叠反应中的任何沉淀。然后使用具有10kDa Minimate TFF Capsule(Pa11)的最小切向流过滤系统(Pa11)和具有10000Da分子量截留膜(Millipore)的Amicon Ultra-15离心过滤器浓缩重折叠反应。浓缩的重折叠反应在HiLoad 26/600Superdex 200制备级(GE Life Sciences)上通过尺寸排阻色谱(SEC)纯化,其在SEC缓冲液(20mM HEPES pH 7.2, 150mM NaCl)中预平衡。合并并浓缩对应于单体HLA-A2-分选标签/ $\beta$ 2m/肽复合物的纯化级分。HLA-A2、 $\beta$ 2m和NLVPMVATV(SEQ ID NO:8)肽(缩写为NLV)的重折叠和纯化遵循类似的程序。

[0557] 使用分选酶将点击手柄肽与HLA-A2-分选标签偶联。使用转肽酶分选酶用点击手柄肽对HLA-A2进行酶修饰。在大肠杆菌中表达含有5个增强突变的分选酶(Chen,PNAS 2011 108(28) 11399-11404)(其氨基酸序列如SEQ ID NO:6所示),并根据(Antos,Curr Protoc Protein Sci, 2009 doi:10.1002/0471140864.ps1503s56)进行纯化。将含有N末端三甘氨酸和随后的PEG接头(PEG<sub>4</sub>或PEG<sub>5</sub>)的点击手柄肽合成性连接到:1)丙炔基甘氨酸(称为GGG-炔烃、炔烃或alk),2) Sulfo-DBCO(称为GGG-DBCO或DBCO),或3)吡啶甲基叠氮化物(称为GGG-叠氮化物、叠氮化物或Az)。具有C末端酰胺化的GGG-PEG<sub>5</sub>-炔烃肽由GenScript

(Piscataway, NJ) 合成。通过点击化学工具 (Scottsdale, AZ) 合成了具有C末端酰胺化的 GGG-PEG<sub>4</sub>-叠氮化物肽和GGG-PEG<sub>4</sub>-DBCO肽。

[0558] 将HLA-A2/ $\beta$ 2m/肽单体 (100-150 $\mu$ M)、点击手柄肽 (GGG-炔烃、GGG-DBCO或GGG-叠氮化物, 6-10mM)、分选酶 (5-6 $\mu$ M) 和10mM CaCl<sub>2</sub>混合, 并在4 $^{\circ}$ C下孵育4小时, 以产生HLA-点击手柄融合物。如上所述通过SEC纯化反应混合物以除去残余分选酶和点击手柄肽。合并并浓缩对应于单体HLA-点击手柄/ $\beta$ 2m/肽复合物的纯化级分。

[0559] SA<sub>v</sub>的表达、纯化和使用分选酶将点击手柄肽与SA<sub>v</sub>偶联。通过标准方法在BL21 (DE3) 细胞中表达含有C末端分选酶-标签和6xHis标签 (其氨基酸序列如SEQ ID NO:3所示) 的全长SA<sub>v</sub>。如上所述, 通过固定化金属亲和层析 (IMAC) 和SEC从可溶性级分纯化SA<sub>v</sub>。SA<sub>v</sub>形成天然四聚体并在SDS-PAGE上作为稳定四聚体迁移 (Waner M. J. 等人, 2004, doi:10.1529/biophysj.104.047266)。合并并浓缩对应于四聚体SA<sub>v</sub>的纯化级分。通过在4C下混合SA<sub>v</sub> (70-150 $\mu$ M)、Click Handle肽 (GGG-DBCO或GGG Azide, 3-10mM)、Sortase (6 $\mu$ M) 和CaCl<sub>2</sub> (10mM) 长达4小时来产生SA<sub>v</sub>-点击手柄融合物。通过SEC纯化反应混合物以除去残余的分选酶和肽, 合并并浓缩与SA<sub>v</sub>-点击手柄融合物相对应的纯化级分。通过测定相对于不同量的未处理SA<sub>v</sub>样品的抗-6xHis反应条带强度的损失程度, 通过抗-His Western印迹分析评估与SA<sub>v</sub>的偶联程度 (图3A)。

[0560] 点击肽/MHC I类-SA<sub>v</sub>多聚体的生成。本文使用几种不同点击化学形式 (例如, 在以下文献所述的点击化学: Agard NJ, Prescher JA, Bertozzi CR J Am Chem Soc. 2004年11月24日; 126 (46): 15046-7; Hong, V. 等人, Angew Chem Int Ed Engl. 2009; 48 (52): 9879-9883. doi:10.1002/anie.200905087) 产生点击的HLA-链霉亲和素融合物。因为SA<sub>v</sub>形成SDS抗性四聚体, 所以可以使用SDS-PAGE来监测反应程度并确定SA<sub>v</sub>上HLA的价态 (Waner M. J. 等人, 2004, doi:10.1529/biophysj.104.047266)。

[0561] 1) 通过将HLA-A2-DBCO/NLV (150 $\mu$ M) 与SA<sub>v</sub>-Az (相对于SA单体为50 $\mu$ M) 混合, 并在冰上温育3小时, 进行无铜炔烃-叠氮环加成形成点击的多聚体。SDS-PAGE分析证实了附着有1、2、3和4个HLA分子的四聚体SA的形成 (图3B)。观察到副产物, 其归因于DBCO与 $\beta$ 2m或HLA-A2上半胱氨酸残基的不良副反应 (van Geel, R, Bioconjugate Chem. 2012, 23 (3): 392-398. doi.org/10.1021/bc200365k)。

[0562] 2) 还通过在室温下或在冰上 (未显示) 混合不同比例的HLA-A2-Az/NLV和SA-DBCO (3:1和2:1) 1.5-3.0小时来制备共价偶联的多聚体HLA。SDS-PAGE分析显示四聚体、三聚体、二聚体和单体HLA-A2-Az-SA<sub>v</sub>-DBCO物质的形成, 与HLA-A2-DBCO-SA<sub>v</sub>-Az相比, 不良副反应产物水平降低。(图3C)。

[0563] 3) 另一种产生共价连接的HLA-A2和SA<sub>v</sub>的方法是通过铜辅助的炔烃-叠氮环加成。通过在冰上混合以下反应组分生成HLA-A2-A1k-SA<sub>v</sub>-Az: HLA-A2-A1k/GILGFVFL (SEQ ID NO:7) / $\beta$ 2m (100-130 $\mu$ M)、SA<sub>v</sub>-Az (70-80 $\mu$ M, 相对于SA单体)、硫酸铜 (0.5mM)、BTAA (2.5mM) 和抗坏血酸 (5mM)。通过SDS-PAGE监测反应, 4小时后通过SEC纯化反应混合物以从纯化的HLA-A2-炔烃-SA<sub>v</sub>-Az多聚体中分离未反应的HLA、SA<sub>v</sub>和其他反应组分。通过SDS-PAGE分析SEC级分, 合并并浓缩对应于大多数四聚体/三聚体物质的部分。通过SDS-PAGE分析肽/HLA-A2-炔烃-SA<sub>v</sub>-Az/ $\beta$ 2m样品, 其显示了未煮沸/未还原样品的明显四聚体和三聚体物质以及非常少量的单体, 而煮沸和还原凝胶分析证实了HLA-A2-A1k和SA<sub>v</sub>-Az单体在约53kDa的共

价连接(图3D)。变性条件下的质谱也证实了HLA-A2和SAv之间形成叠氮-炔炔融合物(未显示)。如图3E所示,还生成了HLA-A\*01:01、HLA-A\*03:01和HLA-A\*24:02的HLA-炔炔SAv-Az形式。

[0564] 实施例2-产生具有内含蛋白标签的可交换肽MHC I类多聚体

[0565] 在本实施例中,MHCI重链与C末端N-内含蛋白标签表达,链霉亲和素(SA)与N末端C-内含蛋白标签表达,然后用内含蛋白介导的偶联产生加载可交换肽的MHC I四聚体。例如,Stevens等人,J. Am. Chem. Soc., 138, 2162-2165, 2016; Shah等人J. Am. Chem. Soc., 134, 11338-11341, 2012; 和Vila-Perello等人,J. Am. Chem. Soc., 135, 286-292, 2013中进一步描述了内含蛋白序列及其用于偶联蛋白质的用途,其全部内容通过引用并入本文。

[0566] HLA-A2(<sub>HLA-A\*02:01</sub>) 在BL21 (DE3) 中表达为与C末端的Npu N-内含蛋白片段的融合物(其氨基酸序列如SEQ ID NO:4所示)。链霉亲和素在BL21 (DE3) 中表达,N末端融合到Npu-C-内含蛋白片段和C末端Flag标签(其氨基酸序列如SEQ ID NO:5所示)。在细菌包涵体中表达HLA-A2-N-内含蛋白和C-内含蛋白-SAv。分离包涵体并溶解在尿素缓冲液(25mM MES、8mM尿素、10mM EDTA、0.1mM DTT, pH 6.0) 中。用β<sub>2m</sub>和UV不稳定占位符肽(GILGFVJL (SEQ ID NO:7), 其中J是3-氨基-3-(2-硝基) 苯基丙酸) 重折叠HLA-A2-N-内含蛋白。如实施例1所述,在搅拌下将下列组分加入预冷的重折叠缓冲液中。使用具有10000Da MWC0、Millipore Biomax超滤盘(Millipore) 和Amicon Ultra-15离心过滤单元10000MWC0 (Millipore) 的Amicon搅拌池浓缩重折叠反应。浓缩的重折叠反应在HiLoad 26/600Superdex 200制备级(GE Life Sciences) 上通过尺寸排阻色谱(SEC) 纯化,其在SEC缓冲液(20mM HEPES pH 7.2, 150mM NaCl) 中预平衡。合并并浓缩对应于单体HLA-A2-N-内含蛋白/β<sub>2m</sub>/肽复合物的纯化级分。用相同方法重折叠C-内含蛋白-SAv: 简单地说,将尿素溶解的C-内含蛋白-SAv注射到预冷的重折叠缓冲液中,并根据实施例1中描述的方案进行重折叠,如所述在具有10KMWC0膜的Amicon搅拌池中浓缩,并如上所述通过尺寸排阻色谱纯化。将SEC纯化的C-内含蛋白-SAv浓缩至100-200μM。

[0567] 通过向HLA-A2-内含蛋白和C-内含蛋白-SA组分中加入最终浓度为0.5mM的三(2-羧基乙基) 膦盐酸盐(TCEP), 进行HLA-A2-N-内含蛋白/β<sub>2m</sub>/肽复合物与C-内含蛋白-SAv之间的剪接反应。所有组分保持在冰上。为了有利于四聚体物质形成,在16小时内以5个增量添加链霉亲和素,直到达到与HLA-A2-内含蛋白的等摩尔量。在非还原/非煮沸条件下反应的SDS-PAGE分析显示了更高MW物质的形成,而煮沸/还原的样品显示了大约52kDa的物质,与HLA-A2-SAv融合物的预期大小一致(图4)。

[0568] 实施例3. 通过生物素化和与链霉亲和素的偶联产生可交换MHCI四聚体。

[0569] 带有C末端Avitag的HLA-A\*02重链在大肠杆菌包涵体中表达。Avitag的氨基酸序列如SEQ ID NO:161所示。根据文献方法(Altman&Davis, Curr Protoc Immunol. 2003; 第17章: 单元17.3; Rodenko等人, Nat Protoc. 2006; 1(3): 1120-32), 将纯化的包涵体溶解在尿素中并用β-2-微球蛋白和肽NLVPMVATV (SEQ ID NO:8) 或条件配体GILGFVJL (SEQ ID NO:7, 其中J是2-硝基苯基氨基酸残基) 重折叠。然后使用生物素连接酶将包含重链的SEC纯化的MHC单体、β-2-微球蛋白和肽生物素化, 然后再次SEC纯化。将链霉亲和素添加到10个单独等分试样中的生物素化MHC单体中, 以实现生物素位点相对于MHC单体量轻微摩尔过量。在生物素介导的链霉亲和素四聚体或生物素化的HLA单体上进行肽交换(如实施例4所述)。在

后者的情况下,单体在交换后与链霉亲和素四聚。

[0570] 实施例4.通过二肽或UV交换进行肽交换。

[0571] 通过两种方法中的任一种交换如以上实施例1中所述制备的HLA-A\*02:01-A1k-SAv-Az偶联四聚体以及如实施例3中所制备的生物素介导的HLA-A\*02四聚体。对于二肽交换,在室温下,存在或不存在10mM GM二肽的情况下,将加载占位符肽(例如GILGFVFL(SEQ ID NO:7))的5 $\mu$ M MHC四聚体与30倍过量的NLVPMVATV(SEQ ID NO:8)肽孵育3小时(Saini等人,PNAS2006;112(1):202-206)。对于UV交换,将2-10 $\mu$ M的MHC单体或0.5-2.5 $\mu$ M加载有占位符肽(GILGFVFL(SEQ ID NO:7))的MHC四聚体与30-100倍摩尔过量的NLVPMVATV(SEQ ID NO:8)(或其他肽)在冰上孵育1小时,然后将其暴露于距离样品2-5cm的灯发出的365nm UV光下30分钟。紫外线暴露后,有时在30 $^{\circ}$ C孵育30分钟以允许完全交换。通过差示扫描荧光法(DSF)、ELISA和细胞染色/流式细胞术监测肽交换效率。

[0572] 对于DSF,将0.25mg/ml HLA-A\*02四聚体与等体积的20X Sypro Orange(Invitrogen S6650)混合,并在qPCR仪器(例如,Applied Biosystems Quant Studio 3)中经受从25 $^{\circ}$ C到99 $^{\circ}$ C的0.05 $^{\circ}$ C/秒升温。熔体曲线的一阶导数中的峰值表示pMHC的T<sub>m</sub>。如图5A所示,如实施例1中制备的HLA-A\*02:01-A1k-SAv-Az偶联四聚体的T<sub>m</sub>在从占位符GILGFVFL(SEQ ID NO:7)肽紫外交换到NLVPMVATV(SEQ ID NO:8)后从40 $^{\circ}$ C转变到61 $^{\circ}$ C。UV交换后的T<sub>m</sub>与观察到的NLVPMVATV(SEQ ID NO:8)交换成生物素化单体,然后四聚(工业标准)或直接交换成生物素介导的四聚体的T<sub>m</sub>相同(图5B)。这些数据证实多聚体状态对UV交换效率没有影响,并且本发明的偶联四聚体具有与工业标准pMHC相同的稳定性。

[0573] 对于流式细胞术,将用NLVPMVATV(SEQ ID NO:8)(或其他肽)扩增的10<sup>5</sup>个供体T细胞用如上制备的pMHC四聚体染色。所有pMHC均在PBS加10%FBS中稀释,并使用抗CD8-BV785和抗-Flag-APC或抗链霉亲和素-PE(Biolegend)作为二抗进行染色。如图6A-F所示,在生物素介导的四聚体形式上进行的二肽交换或UV交换产生HLA-A\*02四聚体,其显示出与通过工业标准方法(生物素化单体的重折叠后四聚或UV交换后四聚)产生的那些具有相同水平的与扩增T细胞的结合。图7显示了与NLVPMVATV(SEQ ID NO:8)肽进行UV交换的HLA-A\*02:01-A1k-SAv-Az偶联四聚体与扩增的T细胞的高亲和力结合。

[0574] ELISA也用于监测四聚体上的交换,并且是pMHC稳定性的另一个指标。首先用抗链霉亲和素抗体涂覆平板,然后在pH 5.4的柠檬酸-磷酸盐缓冲液中捕获四聚体,然后使用HRP偶联的抗- $\beta$ 2-微球蛋白(Biolegend)读取。如图8A所示,一组NLVPMVATV(SEQ ID NO:8)突变肽可以有效地UV交换成HLA-A\*02:01-A1k-SAv-Az偶联四聚体,产生一系列ELISA信号。UV交换成生物素介导的HLA-A\*02四聚体的一小组类似肽也产生了一系列ELISA信号(图8C),其与DSF测量的T<sub>m</sub>呈正相关(图8B)。

[0575] 实施例5:用HLA-A\*01:01产生的偶联四聚体

[0576] 用肽STAPGJLEY(SEQ ID NO:16)重折叠的HLA-A\*01:01单体构建如上实施例1所述的HLA-A\*01:01-A1k-SAv-Az偶联四聚体和QC'd。如图9A和图9B所示,HLA-A\*01:01-A1k-SAv-Az偶联四聚体是高度多聚体,聚集体百分比低(3%)。在同源肽VTEHDTLLY(SEQ ID NO:10)存在下的UV处理导致DSF熔体曲线的特征性偏移,表明有效的肽交换(图9C)。交换后的HLA-A\*01:01-A1k-SAv-Az偶联四聚体与用VTEHDTLLY肽(SEQ ID NO:10)扩增的PBMC强烈结合,类似于用VTEHDTLLY肽(SEQ ID NO:10)重折叠的HLA-A\*01:01,其通过生物素与链霉亲

和素偶联(图9D)。正如预期的那样,在没有UV交换的情况下没有观察到结合。

[0577] 实施例6:用HLA-A\*24:02产生的偶联四聚体

[0578] 用肽VYGJVRACL (SEQ ID NO:11) 重折叠的HLA-A\*24:02单体构建如上实施例1所述的HLA-A\*24:02-Alk-SAv-Az偶联四聚体和QC'd。如图10A和图10B所示,HLA-A\*24:02-Alk-SAv-Az偶联四聚体是高度多聚体,聚集体百分比低(6%)。在同源肽QYDPVAALF (SEQ ID NO:12) 存在下的UV处理导致DSF熔体曲线的特征性偏移,表明有效的肽交换(图10C)。交换后的HLA-A\*24:02-Alk-SAv-Az偶联四聚体与用QYDPVAALF肽(SEQ ID NO:12) 扩增的PBMC强烈结合,类似于用QYDPVAALF肽(SEQ ID NO:12) 重折叠的HLA-A\*24:02,其通过生物素与链霉亲和素偶联(图10D)。正如预期的那样,在没有UV交换的情况下没有观察到结合。

[0579] 实施例7:用HLA-B\*07:02产生的偶联四聚体

[0580] 用肽AARGJTLAM (SEQ ID NO:14) 重折叠的HLA-B\*07:02单体构建如上实施例1所述的HLA-B\*07:02-Alk-SAv-Az偶联四聚体和QC'd。如图11A和图11B所示,HLA-B\*07:02-Alk-SAv-Az偶联四聚体是多聚体,没有可检测的聚集体。在同源肽RPHENGFTVL (SEQ ID NO:13) 存在下进行UV处理后,交换的HLA-B\*07:02-Alk-SAv-Az偶联四聚体与用RPHENGFTVL肽(SEQ ID NO:13) 扩增的PBMC强烈结合,类似于用RPHENGFTVL肽(SEQ ID NO:13) 重折叠的HLA-B\*07:02,该肽通过生物素与链霉亲和素偶联(图11C)。正如预期的那样,在没有UV交换的情况下没有观察到结合。

[0581] 实施例8:UV交换的四聚体的条码化和合并

[0582] 由于链霉亲和素上的生物素结合位点是空的,交换的HLA-A\*02:01-Alk-SAv-Az偶联四聚体很容易用识别寡核苷酸标签(条码)标记。以2:1寡聚物:四聚体摩尔比加入5'生物素化寡核苷酸,并在4°C下孵育30分钟,然后在4°C用生物素以400:1的生物素:四聚体摩尔比淬灭30分钟。通过在4-12%的双-tris凝胶上电泳,然后用硝酸纤维素印迹并用抗-Flag抗体(Invitrogen#MA1-91878-D800)染色,确认条码化。如图12所示,相对于四聚体起始物质的凝胶位移表明用寡核苷酸条码正确标记。

[0583] 实施例9:用合并的条码化UV交换四聚体单细胞测序

[0584] 用对于192种不同的NLVPMVATV (SEQ ID NO:8) 的APL变体UV交换的单个HLA-A\*02:01-Alk-SAv-Az偶联四聚体样品分别与寡核苷酸标记物偶联,合并,在NLVPMVAT (SEQ ID NO:8) 扩增的T细胞上染色,并进行单细胞测序。分析结果显示在图13的热图中,表明APL变体子集的克隆型特异性结合。

[0585] 实施例10:用于条码化UV交换四聚体池高通量生产的多孔水凝胶制备

[0586] 水凝胶珠是通过在不同相对浓度下混合丙烯酸胺单体单元和双丙烯酸胺交联剂单元以及丙烯酸化寡核苷酸引物的混合物,使用微流体Drop-maker封装在液滴中,并孵育混合物直到交联完成而制备的。在该实施例中,预交联水性混合物包括10%TEBST (Tris-EDTA缓冲盐水加吐温-20) 中的0.75%双丙烯酸胺、3%丙烯酸胺、25 $\mu$ M 5'-丙烯酸化正向引物、0.5%过硫酸铵。混合并搅拌水性混合物的所有试剂。向混合物中添加1.5%的TEMED和1%的008-FluoroSurfactant,封装在液滴中,在室温下孵育1小时,然后转移到60°C的烘箱中过夜孵育,从而形成水凝胶。水凝胶珠用20%的1H,1H,2H,2H-全氟-1-辛醇(PFO)洗涤一次,然后用TEBST洗涤3次,然后用低TE (1mM Tris-Cl pH 7.5,0.1mM EDTA) 洗涤3次。水凝胶珠4°C储存在TEBST中直至使用。

[0587] 实施例11:单模板PCR以产生编码肽的扩增子

[0588] 编码SUMO结构域-肽融合物的线性DNA模板在单模板条件下以液滴形式PCR扩增到水凝胶珠上,每滴最多得到一个DNA模板。将实施例10中制备的1.4mL水凝胶珠与如下的PCR组分在2mL反应体积中混合:400 $\mu$ L Q5反应缓冲液(新英格兰生物实验室(New England Biolabs))、40 $\mu$ L 10mM dNTP、40 $\mu$ L 1 $\mu$ M正向引物、40 $\mu$ L 25 $\mu$ M5'-生物素化反向引物、40 $\mu$ L 0.1pg/ $\mu$ l线性DNA模板(或模板混合物)、8 $\mu$ L 20%IGEPAL和20 $\mu$ L Q5 DNA聚合酶(新英格兰生物实验室(New England Biolabs))。以液滴形式包封混合物,并进行35轮PCR。通过添加等体积的100%全氟辛醇(PFO)使液滴裂解后,用10体积的低TE洗涤水凝胶五次。用XbaI(在扩增子内切割)在37 $^{\circ}$ C下消化水凝胶珠的等分(10 $\mu$ L每等份)1小时,并与PCR上清液一起在1.2%琼脂糖凝胶上跑胶,定量扩增子的产量和质量(图14)。通过用链霉亲和素-PE标记水凝胶证明了单模板条件有效,其中仅23%的液滴扩增的水凝胶被染色,相对于100%批量扩增的水凝胶(图15)。

[0589] 实施例12:将可条码化交换就绪的偶联四聚体加载到水凝胶上

[0590] 将PCR扩增的水凝胶与50-500nM HLA-A\*02:01-Alk-SAv-Az偶联四聚体以体积比1:1混合,所述四聚体加载有UV不稳定肽(例如GILGFVJL(SEQ ID NO:7),隔绝环境光,并在冰上孵育2小时。通过用抗-Flag-APC或抗- $\beta$ 2M-Alexa488洗涤和染色来确认HLA-A\*02:01-Alk-SAv-Az偶联四聚体的加载,如图16A所示。如图16B所示,通过用Benzonase酶或SmaI(其在扩增子内切割)释放,然后用抗-链霉亲和素捕获和抗-Flag-HRP或抗 $\beta$ 2M-HRP检测的ELISA来定量加载的四聚体的量。

[0591] 实施例13:肽的液滴体外转录/翻译(IVTT)和UV交换到加载四聚体中

[0592] 120 $\mu$ L水凝胶珠与240 $\mu$ L IVTT主混合液共同包封在液滴中,包括120 $\mu$ L PURExpress溶液A(新英格兰生物实验室)、90 $\mu$ L PURExpress溶液B(NEB)、6 $\mu$ L RNA酶OUT(Invitrogen)、以及12 $\mu$ L Ulp1蛋白酶(Invitrogen)液滴在30 $^{\circ}$ C下孵育4小时,不振摇,然后通过30分钟暴露于365nm UV光进行UV交换,所述UV光来自离样品2-5cm的灯。紫外线暴露后,在30 $^{\circ}$ C孵育30分钟以允许完全交换。将D-生物素添加到IVTT反应中,使其最终浓度达到500 $\mu$ M,然后破碎液滴,最后添加等体积的100%PFO。用10体积的PBS加2%的BSA洗涤水凝胶珠五次。可以从PCR扩增子产生足够的肽以产生功能性交换四聚体。如图17A和17B所示。

[0593] 实施例14:单链多聚肽MHC的释放与分析

[0594] 通过用SmaI消化从洗涤的水凝胶中释放UV交换的pMHC,SmaI在肽编码区上游的扩增子内切割,释放具有自识别寡核苷酸标签(条码)释放的四聚体,如图16B所示。释放的pMHC如图16B所示通过ELISA定量,并如图18所示在抗原特异性CD8+T细胞上染色。在图19中示意性地总结了液滴生产的整个过程。

[0595] 实施例15:偶联肽/MHC II类-SAv多聚体的产生

[0596] 使用分选酶将点击手柄肽与MHC II-分选标签结合。MHC II $\alpha$ 链和 $\beta$ 链的序列重组表达如下: $\alpha$ -链胞外结构域序列与C末端分选酶标签一起表达,该标签能够实现与链霉亲和素(SAv)的翻译后偶联,形成可条码化可交换MHC II多聚体。 $\alpha$ 链还包含用于诊断目的的Myc标签。具有分选标签和Myc标签的 $\alpha$ 链胞外结构域的氨基酸序列如SEQ ID NO:191所示。 $\beta$ 链用N末端低亲和力占位符肽(CLIP肽,其序列如SEQ ID NO:189所示),随后是柔性接头、 $\beta$ 链胞外结构域和组氨酸纯化标签重组表达。具有占位符肽、柔性接头和His标签的 $\beta$ 链胞外结

构域的氨基酸序列如SEQ ID NO:192所示。柔性接头包含一个切割位点,该位点允许通过特定蛋白酶破坏肽和β链之间的连接,从而实现随后的肽交换。其中加载有共价占位符肽的MHCII分子在本文中称为p\*MHCII。

[0597] p\*MHCII $\alpha$ -和β-链在CHO细胞中共表达,并作为稳定的异二聚体分泌到表达介质中。CHO表达后,通过固定化金属离子亲和层析和尺寸排阻层析(SEC)纯化p\*MHCII。然后使用分选酶将GGG-X肽偶联到p\*MHCII $\alpha$ 链(图20,步骤1),其中X可以是叠氮化物、炔烃或任何可点击的化学部分。为了执行化学偶联反应,p\*MHCII(30-50 $\mu$ M),点击手柄肽(GGG-炔烃、GGG-DBCO或GGG-叠氮,6-10mM)、分选酶(5-6 $\mu$ M)和10mM CaCl<sub>2</sub>混合并在4°C下温育2小时,以生成p\*MHCI I-点击手柄融合物。通过SEC纯化反应混合物以除去残留的分选酶和点击手柄肽。合并并浓缩对应于p\*MHCII-点击手柄融合物的纯化级分。点击手柄的添加导致偶联蛋白大小改变,证实成功的分选酶介导的连接(图21A)。

[0598] 偶联p\*MHCII-SAv多聚体的产生。使用分选酶的点击手柄p\*MHCII到SAv的表达、纯化和偶联如图20步骤2所示,并且基本上按照实施例1中对MHCI多聚体的描述进行。铜辅助的炔烃-叠氮环加成用于生成共价连接的p\*MHCII和SAv(图20,步骤3)。p\*MHCII-Alk-SAv-Az通过在冰上混合以下反应组分而产生:MHC II-Alk(50 $\mu$ M)、SAv-Az(相对于SA单体为25 $\mu$ M),硫酸铜(0.5mM),BTAA(2.5mM)和抗坏血酸(5mM)。通过SDS-PAGE监测反应(图21B),4小时后通过SEC纯化反应混合物以从纯化的p\*MHCII-Alk-SAv-Az多聚体中分离未反应的HLA、SAv和其他反应组分(图21C)。SAv和β链分别含有FLAG和His标签,能够区分对应于多聚体种类的级分(图21D和21E)。多聚体级分显示出明显的四聚体和三聚体物质。更重要的是,在SDS-PAGE和western印迹分析下,在取自多聚体级分的煮沸样品中未观察到游离SAv物质(图21D)。这表明优势物质是四聚体,其中每个SAv亚基与p\*MHCII亚基共价连接。

[0599] 实施例16:pMHC II多聚体可交换并结合同源表位特异性TCR

[0600] 接头消化和肽交换

[0601] p\*MHCII-Alk-SAv-Az多聚体(下文中为p\*MHCII-SAv)在4°C、1mM CaCl<sub>2</sub>存在下以5:1(w/w)的比例由因子Xa(NEB)过夜消化(图20,步骤4)。然后,根据厂商建议(Sigma-Aldrich),通过添加1,5-丹磺酰-Glu-Gly-Arg氯甲基酮抑制剂使蛋白酶不可逆失活。消化的样品比未消化样品迁移更快,这表明在SDS-PAGE变性条件下去除了新切割的肽(图22A)。

[0602] 为了测试切割的p\*MHCII-SAv(下文为-p↓MHCII-SAv)是否结合交换的肽,进行ELISA结合测定。在该测定中,加载来自甲型流感病毒的生物素化肽表位(血凝素,HA,其氨基酸序列如SEQ ID NO:193所示),同时在弱酸性pH条件下去除切割的占位符肽(图20,步骤5)。然后通过监测链霉亲和素-HRP与新交换的生物素化肽的结合来确定交换水平。在交换反应之前,链霉亲和素分子上的游离生物素结合位点被过量游离生物素封闭。该步骤确保了任何检测到的生物素化肽只能与肽结合口袋结合。交换缓冲液组合物如下:100mM柠檬酸钠,pH 5.5,50mM氯化钠,1%辛基葡糖苷(v/v),1x SIGMAFAST蛋白酶抑制剂混合物(Sigma-Aldrich)和0.1mM DTT。在96孔板中制备150 $\mu$ l肽交换反应物,其中每个孔由以下组成:1x交换缓冲液,30nM p↓MHCII-SAv和5倍系列稀释的HA-生物素化肽、HA-非生物素化肽或缓冲液。6nM p↓MHCII单体与5倍系列稀释的HA-生物素化肽一起孵育,作为阳性对照。加入1:15(v/v)的1M Tris-HCl(pH10)中和酸性pH,在37°C孵育过夜后停止交换反应。使用96通道台式移液器,将来自每个孔的100 $\mu$ l转移到ELISA板上,该ELISA板预涂有(100ng/孔)L243构象



敏感抗体 (Abcam), 洗涤 ( $3 \times \text{PBS-T}$ ) 并用补充有 2% (v/v) BSA 的 PBS-T 封闭。在室温下孵育 1 小时后, 洗涤平板 ( $3 \times \text{PBS-T}$ ), 在黑暗中与 SA-HRP 孵育 30 分钟, 再次洗涤 ( $3 \times \text{PBS-T}$ ) 并使用 HRP 底物和终止溶液显影。对于两种单体 p $\downarrow$ MHCII 和 p $\downarrow$ MHCII-SAv, 观察到肽浓度与 SA-HRP 结合水平之间成正比 (图 21B)。这表明两种物质都用占位符肽交换了生物素化-HA 肽。用非生物素化肽或缓冲液孵育未产生可检测信号, 表明生物素化表位结合是特异性的。与单体 p $\downarrow$ MHCII 相比, p $\downarrow$ MHCII-SAv 的曲线向右偏移, 在较高的肽浓度下未达到饱和。多聚体的大小至少大 4 倍, 这可能会阻断与捕获抗体和/或 SA-HRP 读取探针的结合。

[0603] 交换的 p $\downarrow$ MHCII-SAv 与可溶性 TCR 结合

[0604] F11, 一种 HA-肽表位特异性可溶性 TCR, 与 FC 结构域融合并如 Wagner 等人 J Biol Chem., 294:5790-5804, 2019 所述生产 (图 20, 步骤 6)。简而言之, 将编码 F11 胞外  $\alpha$  链和  $\beta$  链的 DNA 克隆到 pDT5 质粒, 小鼠 IgGk 链前导序列下游。人 TCR 恒定结构域包含额外的链间二硫键。C- $\alpha$  结构域之后是人 IgG1 的上较链序列 (VEPKSC; SEQ ID NO:270)、核心和下较链, 然后是 Fc 结构域。天然 IgG1 轻链半胱氨酸插入 C- $\beta$  的 C 末端, 以与上较链半胱氨酸配对, 进一步稳定 TCR 异二聚化。其他修改包括去除 N-连接的糖基化位点。通过瞬时转染在 Expi-CHO 细胞中表达编码  $\alpha$ -Fc 和  $\beta$  结构域的质粒, 并通过蛋白 A 亲和层析从澄清上清液中纯化产物。

[0605] 交换反应如上实施例 1 所述进行, 但有两个不同: 使用单根试管代替 96 孔板, 蛋白质浓度变化。1.75  $\mu\text{M}$  p $\downarrow$ MHCII-SAv 与 100  $\mu\text{M}$  HA 肽在交换缓冲液存在下孵育。停止反应并保持在冰上后, 使用 Octet RED96 仪器 (ForteBio) 在 30°C 的 BLI 缓冲液 (PBS+0.02% 吐温 20、0.1% BSA、0.05% 叠氮化钠) 中进行生物层干涉法 (BLI) 测定。将 F11 TCR 加载到抗-hIgG Fc 捕获生物传感器 (Molecular Devices) 上至 0.6nm 加载信号。用 BLI 缓冲液洗涤后, 将生物传感器转移到含有 14nM 交换的 p $\downarrow$ MHCII-SAv、125nM 非交换 p\*MHCII-SAv 或 BLI 缓冲液的孔中, 以测量结合动力学 (图 22C)。为了测量解离动力学, 将生物传感器转移回不含多聚体的 BLI 缓冲液中。观察到对于 HA-交换的 p $\downarrow$ MHCII-SAv, BLI 响应信号有显著增加, 表明与 F11 TCR 有强结合 (图 22C)。相反, 非交换的 p\*MHCII-SAv 显示出非常少的结合, 表明 F11-TCR 和 HA 展示多聚体之间的相互作用是特异性的。当生物传感器浸入 BLI 缓冲液中时, 未观察到结合。HA 交换的 p $\downarrow$ MHCII-SAv 表现出与 F11-TCR 非常轻微的解离。这一结果表明 TCR-MHC II 结合紧密, 这是高亲合力多聚体相互作用的特征。

[0606] 交换的 pp $\downarrow$ MHCII-SAv 文库与抗原特异性 CD4<sup>+</sup>T 细胞的结合

[0607] 用寡核苷酸标记单独的 DRB1\*01:01-SAv 偶联四聚体样品并合并, 该样品与流感血凝素 (HA) 肽 (SEQ ID NO:281) 和 9 种其他对照肽进行了 UV 交换。随后, 使用该合并物对 HA 扩增的 CD4<sup>+</sup>T 细胞进行染色, 这些细胞在掺入对照表位 ELAIGIGILTV (SEQ ID NO:282)-扩增的细胞 (已用 HLA-A\*02:01 四聚体合并物染色) 后进行分选并进行单细胞测序。分析结果显示在图 23 中的热图中, 表明加载 HA 的 DRB1\*01:01 四聚体的克隆型特异性结合。

[0608] 通过引用纳入

[0609] 本申请中引用的每篇专利、出版物和非专利文献通过引用全文纳入本文, 就如同各自单独通过引用纳入。

[0610] 序列表总结



[0611]

SEQ ID NO:	说明
1	MSGSHSMRYFFTSVSRPGRGEPRIA VGYVDDTQFVRFSDAASQRMEPRAPWIEQE GPEYWDGETRKVK AHSQTHRVDLGTLRGYYNQSEAGSHTVQRMYGCDVGS DWRF LRGYHQYAYDGKD YIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLEGT CVEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALS FYPAEITLTWQ RDGEDQTQDTEL VETRPAGDGT FQKWA AVVVP SGQE QRYTCHVQHEGLPKPLTLR WEPSSGSGSAGGSGSGGGSLPETGGHHHHHH (HLA-A2- 分选标签和 His6 标签)
2	MIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFS KDWSFYLLYYTEFTPEKDEYACRVNHVTL SQPKIVKWRDRM (β-2 微球蛋白, 具有额外的 N 末端甲硫氨酸)
3	MDYKDDDDK GSSGDP SKDSKAQV SAAEAGITGTWYNQLG STFIVTAGADGALTGT YESAVGNAESRYVLTGRYDSAPATDGS GTALGWTVAWKNNYRNAHSATTWSGQY VGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAASIDA AKKAGVNNG NPLDAVQQLPETGGHHHHHH (SAv -分选标签和 His6 标签)
4	MGSMSMRYFFTSVSRPGRGEPRIA VGYVDDTQFVRFSDAASQRMEPRAPWIEQE GPEYWDGETRKVK AHSQTHRVDLGTLRGYYNQSEAGSHTVQRMYGCDVGS DWRF LRGYHQYAYDGKD YIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLEGT CVEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALS FYPAEITLTWQ RDGEDQTQDTEL VETRPAGDGT FQKWA AVVVP SGQE QRYTCHVQHEGLPKPLTLR WEPSSGSGSAGGSGFESGPGA EYCLSYETEILTVEYGLLP IGTKIVEKRIE CTVYSVDNNG NIYTQPVAQWHD RGEQEVFEYCLE DGLIRATKDHKFMTVDGQMLPIDEIFERELDL MRVDNLPN (HLA-A2-N-内含蛋白)
5	MIKIATRKYLGKQNVYDIGVERDHN FALKNGFIASNCFNVDDPSKDSKAQV SAAEA GITGTWYNQLG STFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGS GT ALGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLV GHDTFTKVKPSAASIDA AKKAGVNNGNPLDAVQQGSTGDYKDDDDK (C-内含蛋白-Sav 和 Flag Tag)
6	MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPA TREQLNRGV SFAEENESLDDQNIS IAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVEVLDEQ K GKDKQLTLITCDDYNEETGVWETRKIFVATEVKLEHHHHHH (具有 His6 标签的分选酶)
7	GILGFV FJL (A02:01 占位符肽)
8	NLVPMVATV (A02:01 结合肽)
9	NLVPMVGTV (A02:01 结合肽)
10	VTEHDTLLY (A01:01 结合肽)
11	VYGJVRACL (A24:02 占位符肽)
12	QYDPVAALF (A24:02 结合肽)
13	RIPHERNGFTVL (B7:02 结合肽)
14	AARGJTLAM

[0612]

	(B7:02 占位符肽)
15	KILGFVFJV (A2:01 占位符肽)
16	STAPGJLEY (A1:01 占位符肽)
17	RIYRJGATR (A3:01 占位符肽)
18	RVFAJSFIK (A11:01 占位符肽)
19	KPIVVVLJGY (B35:01 占位符肽)
20	FVYGJSKTSL (C3:04 占位符肽)
21	FLRGRAJGL (B8:01 占位符肽)
22	VRIJHLYIL (C7:02 占位符肽)
23	QYDJAVYKL (C4:01 占位符肽)
24	ILGPJGSVY (B15:01 占位符肽)
25	TEADVQJWL (B40:01 占位符肽)
26	ISARGQJLF (B58:01 占位符肽)
27	KAAJDLSHFL (C8:01 占位符肽)
28	MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F F T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q K M E P R A P W I E Q E G P E Y W D Q E T R N M K A H S Q T D R A N L G T L R G Y Y N Q S E D G S H T I Q I M Y G C D V G P D G R F L R G Y R Q D A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T K R K W E A V H A A E Q R R V Y L E G R C V D G L R R Y L E N G K E T L Q R T D P P K T H M T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E L S S Q T I P V G H I A G L V L L G A V I T G A V V A A V M W R R K S S D R K G G S Y T Q A A S S D S A Q G S D V S L T A C K V (HLA-A*01:01 全长)
29	MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F F T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E G P E Y W D Q E T R N V K A Q S Q T D R V D L G T L R G Y Y N Q S E A G S H T I Q I M Y G C D V G S D G R F L R G Y R Q D A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T K R K W E A A H E A E Q L R A Y L D G T C V E W L R R Y L E N G K E T L Q R T D P P K T H M T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E L S S Q T I P V G H I A G L V L L G A V I T G A V V A A V M W R R K S S D R K G G S Y T Q A A S S D S A Q G S D V S L T A C K V (HLA-A*03:01 全长)
30	MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E G P E Y W D Q E T R N V K A Q S Q T D R V D L G T L R G Y Y N Q S E D G S H T I Q I M Y G C D V G P D G R F L R G Y R Q D A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T K R K W E A A H A A E Q Q R A Y L E G R C V E W L R R Y L E N G K E T L Q R T D P P K T H M T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E L S S Q T I P V G H I A G L V L L G A V I T G A V V A A V M W R R K S S D R K G G S Y T Q A A S S D S A Q G S D V S L T A C K V (HLA-A*11:01 全长)
31	MAVMAPRTL V L L L S G A L A L T Q T W A G S H S M R Y F S T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E G P E Y W D E E T G K V K A H S Q T D R E N L R I A L R Y Y N Q S

[0613]

	EAGSHTLQMMFGCDVGS DGRFLRGYHQYAYDGKDYIALKEDLRSWTAADMAAQI TKRKWEAAHVAEQQRAYLEGTCVDGLRRYLENGKETLQRTDPPKTHMTHHPISDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQKWA AVVVPSG EEQRYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIIAGLVLLGAVITGAVVAAVMWR RNSSDRKGGSYSQAASSDSAQGS DVSLTACKV (HLA-A*24:02 全长)
32	MLVMAPRTVLLLLSAALAL TETWAGSHSMRYFYTSVSRPGRGEPFISVGYVDDTQ FVRFSDAASPREEPRAPWIEQEGPEYWRNTQIYKAQAQTDRESLRNLRGYYNQS EAGSHTLQSMY GCDVGP DGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQIT QRKWEAAREAEQRRAYLEGE CVEWLRRYLENGKDKLERADPPKTHVTHHPISDHE ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSGE EQRYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCR RKSSGGKGGSYSQAACSDSAQGS DVSLTA (HLA-B*07:02 全长)
33	MRVMAPRTLILLLSGALAL TETWAGSHSMRYFSTSVSWPGRGEPFIAVGYVDDTQ FVRFSDAASPRGEPREP WVEQEGPEYWDRETQKYKRQAQADRVNLRKLRGYYNQ SEDSHTLQRMFGCDLGP DGRLLRGYNQFAYDGKDYIALNEDLRSWTAADTAAQIT QRKWEAAREAEQRRAYLEGT CVEWLRRYLENGKETLQRAEHPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQWDGEDQTQDTEL VETRPAGDGT FQKWA AVVVPSGE EQRYTCHVQHEGLPEPLTLRWKPSQSTIPIVGIVAGLAVLAVLAVLGAMVAVVMC RRKSSGGKGGSCSQAASSNSAQGS DESLIACKA (HLA-C*04:01 全长)
34	MRVMAPRALLLLLSGGLAL TETWACSHSMRYFDTA VSRPGRGEPFISVGYVDDTQ FVRFSDAASPRGEPREP WVEQEGPEYWDRETQKYKRQAQADRVSLRNLRGYYNQ SEDSHTLQRMSGCDLGP DGRLLRGYDQSA YDGKDYIALNEDLRSWTAADTAAQIT QRKLEAARAAEQRLRAYLEGT CVEWLRRYLENGKETLQRAEPPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQKWA AVVVPSGQ EQRYTCHMQHEGLQEPLT LSWEPSSQSTIPIMGIVAGLAVLVVLA VLGAVVTAMMC RRKSSGGKGGSCSQAACSN SAQGS DESLITCKA (HLA-C*07:02 全长)
35	MLVMAPRTVLLLLSAALAL TETWAGSHSMRYFDTAMSRPGRGEPFISVGYVDDTQ FVRFSDAASPREEP RAPWIEQEGPEYWRNTQIFKTNTQTDRESLRNLRGYYNQSE AGSHTLQSMY GCDVGP DGRLLRGHNQYAYDGKDYIALNEDLRSWTAADTAAQITQ RKWEAARVAEQDRAYLEGT CVEWLRRYLENGKDTLERADPPKTHVTHHPISDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCR KSSGGKGGSYSQAACSDSAQGS DVSLTA (HLA-B*08:01 全长)
36	MRVTAPRTVLLLLWGAVAL TETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWRNTQIFKTNTQTYRESLRNLRGYYNQS EAGSHIIQRMYGCDLGP DGRLLRGHDQSA YDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQRLRAYLEGL CVEWLRRYLENGKETLQRADPPKTHVTHHPVSDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSQSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRK SSGGKGGSYSQAASSDSAQGS DVSLTA (HLA-B*35:01 全长)
37	MRVTAPRTVLLLLWGAVAL TETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRMAPRAPWIEQEGPEYWDGETRNMKASAQTYRENLR LIALRYYN QSEAGSHIIQVMY GCDVGP DGRLLRGHDQSA YDGKDYIALNEDLSSWTAADTAAQI TQRKWEAARVAEQRLRAYLEGL CVEWLRRYLENGKETLQRADPPKTHVTHHPISDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSG EEQRYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMC RRKSSGGKGGSYSQAACSDSAQGS DVSLTA (HLA-B*57:01 全长)
38	MRVTAPRTVLLLLWGAVAL TETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT

[0614]

	<p>QFVRFSDAASPRMAPRAPWIEQEGPEYWDGETRNMKASAQTYRENLRALRYYN                  QSEAGSHIIQVMYGCDVGPDRLLRGNQYAYDGKDIALNEDLSSWTAADTAAQI                  TQRKWEAARVAEQLRAYLEGLCVWLRRYLENGKETLQRADPPKTHVTHHPISDH                  EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWAAVVVPSG                  EEQRYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAVMC                  RRKSSGGKGGSYSQAACSDSAQGSVDVSLTA                  (HLA-B*57:03 全长)</p>
39	<p>MVDGTLLLLLSEALALTQTWAGSHSLKYFHTSVSRPGRGEPFISVGYVDDTQFVRF                  DNDAASPRMVPAPWMEQEGSEYWDRETRSDRTAQIFRVNLRTRLRGYYNQSEAG                  SHTLQWMHGCELGPDGRFLRGYEQFAYDGKDYLTNEDLRSWTAADTAAQISEQK                  SNDASEAEHQRAYLEDTCVEWLHKYLEKGGKETLLHLEPPKTHVTHHPISDHEATLR                  CWALGFYPAEITLTWQQDGEGHTQDTEL VETRPAGDGTTFQKWAAVVVPSGEEQRY                  TCHVQHEGLPEPVTLRWKPASQPTIPIVGIIAGLVLLGSSVVSAGVVAAVIWRKKSSGG                  KGGSYSKAEWSDSAQGSSESHL                  (HLA-E 全长)</p>
40	<p>MRVMAPRTLILLLSGALALTETWACSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQ                  FVRFSDAASPRGEPAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQ                  SEAGSHTLQWMYGCDLGPDRLLRGYDQSAAYDGKDYIALNEDLRSWTAADTAAQI                  TQRKWEAARAAEQRAYLEGTCVEWLRRYLENGKETLQRAEHPKTHVTHHLVSD                  HEATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTTFQKWAAVVVPS                  GEEQRYTCHVQHEGLPEPLTLRWEPSQPTIPIVGIVAGLAVLAVLAVLGAVVAVVM                  CRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA                  (HLA-C*16:01 全长)</p>
41	<p>MRVMAPRTLILLLSGALALTETWACSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQ                  FVQFSDAASPRGEPAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQ                  SEAGSHTLQRMYGCDLGPDRLLRGYNQFAYDGKDYIALNEDLRSWTAADKAAQI                  TQRKWEAAREAEQRAYLEGTCVEWLRRYLENGKKTQRAEHPKTHVTHHPVSDH                  EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTTFQKWAAVVVPSG                  EEQRYTCHVQHEGLPEPLTLRWGPSSQPTIPIVGIVAGLAVLAVLAVLGAVMAVVM                  CRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA                  (HLA-C*08:02 全长)</p>
42	<p>MRVMAPRALLLLSGGLALTETWACSHSMRYFDTA VSRPGRGEPFISVGYVDDTQ                  FVRFSDAASPRGEPAPWVEQEGPEYWDRETQNYKRQAQADRSLRNLRGYYNQ                  SEDGSHTLQRMYGCDLGPDRLLRGYDQSAAYDGKDYIALNEDLRSWTAADTAAQI                  TQRKLEAARAAEQRAYLEGTCVEWLRRYLENGKETLQRAEPPKTHVTHHPISDHE                  ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTTFQKWAAVVVPSGQ                  EQRYTCHMQHEGLQEPLTSLWEPSSQPTIPIVGIVAGLAVLVVLA VLGAVVTAMMC                  RRKSSGGKGGSCSQAACNSAQGSDES LITCKA                  (HLA-C*07:01 全长)</p>
43	<p>MRVMAPRTLILLLSGALALTETWACSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQ                  FVQFSDAASPRGEPAPWVEQEGPEYWDRETQKYKRQAQTDRVNLRLKRGYYNQ                  SEAGSHTLQRMYGCDLGPDRLLRGYNQFAYDGKDYIALNEDLRSWTAADKAAQI                  TQRKWEAAREAEQRAYLEGTCVEWLRRYLENGKKTQRAEHPKTHVTHHPVSDH                  EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTTFQKWAAVVVPSG                  EEQRYTCHVQHEGLPEPLTLRWGPSSQPTIPIVGIVAGLAVLAVLAVLGAVMAVVM                  CRRKSSGGKGGSCSQAASSNSAQGSDES LITCKA                  (HLA-C*05:01 全长)</p>
44	<p>MRVTAPRTLILLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFITVGYVDDT                  LFVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRENLRALRYYNQS                  EAGSHIIQRMYGCDVGPDRLLRGYDQDAYDGKDYIALNEDLSSWTAADTAAQITQ                  RKWEAARVAEQDRAYLEGLCVESLRYLENGKETLQRADPPKTHVTHHPISDHEVT                  LRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWAAVVVPSGEEQ                  RYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAVMCRRK                  SSGGKGGSYSQAACSDSAQGSVDVSLTA                  (HLA-B*44:02 全长)</p>

[0615]

45	<p>MAVMAPRTLLLLLLGALALTQTWAGSHSMRYFTTSVSRPGRGEPFIAVGYVDDTQ                  FVRFSDAASQRMEPRAPWIEQEGPEYWDLQTRNVKAQSQTDRANLGLTRGYYNQ                  SEAGSHTIQMMYGCDVGS DGRFLRGYRQDAYDGKDIALNEDLRSWTAADMAAQI                  TQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHAVSD                  HEATLRCWALSFPYAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQK WASVVVPSG                  QEQR YTCHVQHEGLPKPLTLR WEPSSQPTIPIVGIIAGLVLF GAVFAGAVVA AVRWR                  RKSSDRKGGSYSQAASSDSAQGS DMSLTACKV                  (HLA-A*29:02 全长)</p>
46	<p>MRVTAPRTLLLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFITVGYVDDT                  LFVRFSDATS PRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRENLR TALRYYNQS                  EAGSHIIQRMYGCDVGP DGRLLRGYDQDAYDGKDIALNEDLSSWTAADTAAQITQ                  RKWEAARVAEQLRAYLEGLCVESLRRYLENGKETLQRADPPKTHVTHHPISDHEVT                  LRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQK WAAVVVPSGEEQ                  RYTCHVQHEGLPKPLTLR WEPSSQSTVPIVGIVAGLAVLAVVVIGAVVA AVMCRRK                  SSGGKGGSYSQAACSDSAQGS DVSLTA                  (HLA-B*44:03 全长)</p>
47	<p>MRVMAPRTLILLLSGALALTETWAGSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQ                  FVRFSDAASPRGEP RAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQ                  SEAGSHIIQRMYGCDVGP DGRLLRGYDQYAYDGKDIALNEDLRSWTAADTAAQIT                  QRKWEAAREAEQLRAYLEGLCVESLRRYLKNGKETLQRAEHPKTHVTHHPVSDHE                  ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQK WAAVVVPSGE                  EQRYTCHVQHEGLPEPLTLR WEPSSQPTIPIVGIVAGLAVLAVLAVLGAVVAVVMCR                  RKSSGGKGGSCSQAASSNSAQGS DESLIACKA                  (HLA-C*03:04 全长)</p>
48	<p>MRVTAPRTVLLLLSAALALTETWAGSHSMRYFHTAMSRPGRGEPFITVGYVDDTL                  FVRFSDATS PRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNLRGYYNQSE                  AGSHTLQRMYGCDVGP DGRLLRGHNQYAYDGKDIALNEDLRSWTAADTAAQISQ                  RKLEAARVAEQLRAYLEGECVEWLRRYLENGKDKLERADPPKTHVTHHPISDHEAT                  LRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQK WAAVVVPSGEEQ                  RYTCHVQHEGLPKPLTLR WEPSSQSTVPIVGIVAGLAVLAVVVIGAVVA AVMCRRK                  SSGGKGGSYSQAACSDSAQGS DVSLTA                  (HLA-B*40:01 全长)</p>
49	<p>MRVMAPRTLILLLSGALALTETWACSHSMRYFDTAVSRPGRGEPFISVGYVDDTQF                  VRFSDAASPRGEP RAPWVEQEGPEYWDRETQKYKRQAQADR VNLRKLRGYYNQ                  SEDGSHTLQWMYGCDLGP DGRLLRGYDQSAYDGKDIALNEDLRSWTAADTAAQI                  TQRKWEAAREAEQWRAYLEGTCVEWLRRYLENGKETLQRAEHPKTHVTHHPVSD                  HEATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQK WAAVVVPS                  GEEQRYTCHVQHEGLPEPLTLR WEPSSQPTIPIVGIVAGLAVLAVLAVLGAVMAVV                  MCRRKSSGGKGGSCSQAASSNSAQGS DESLIACKA                  (HLA-C*06:02 全长)</p>
50	<p>MRVTAPRTVLLLLSGALALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQ                  FVRFSDAASPRMAPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNLRGYYNQSE                  AGSHTLQRMYGCDVGP DGRLLRGHDQSAYDGKDIALNEDLSSWTAADTAAQITQ                  RKWEAAREAEQWRAYLEGLCVESLRRYLENGKETLQRADPPKTHVTHHPISDHEA                  TLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQK WAAVVVPSGEE                  QRYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRK                  SSGGKGGSYSQAASSDSAQGS DVSLTA                  (HLA-B*15:01 全长)</p>
51	<p>MRVMAPRTLILLLSGALALTETWAGSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQ                  FVRFSDAASPRGEP RAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQ                  SEARSHIIQRMYGCDVGP DGRLLRGYDQYAYDGKDIALNEDLRSWTAADTAAQIT                  QRKWEAAREAEQLRAYLEGLCVESLRRYLKNGKETLQRAEHPKTHVTHHPVSDHE                  ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGT FQK WAAVVVPSGE                  EQRYTCHVQHEGLPEPLTLR WEPSSQPTIPIVGIVAGLAVLAVLAVLGAVVAVVMCR                  RKSSGGKGGSCSQAASSNSAQGS DESLIACKA</p>

[0616]

	(HLA-C*03:03 全长)
52	MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F S T S V S R P G S G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E R P E Y W D Q E T R N V K A Q S Q T D R V D L G T L R G Y Y N Q S E A G S H T I Q I M Y G C D V G S D G R F L R G Y E Q H A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T Q R K W E A A R W A E Q L R A Y L E G T C V E W L R R Y L E N G K E T L Q R T D P P K T H M T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E L S S Q T P I V G I I A G L V L L G A V I T G A V V A A V M W R R K S S D R K G G S Y T Q A A S S D S A Q G S D V S L T A C K V (HLA-A*30:01 全长)
53	M R V T A P R T L L L L L W G A V A L T E T W A G S H S M R Y F Y T A M S R P G R G E P R F I T V G Y V D D T Q F V R F D S D A T S P R M A P R A P W I E Q E G P E Y W D R E T Q I S K T N T Q T Y R E N L R T A L R Y Y N Q S E A G S H T W Q T M Y G C D L G P D G R L L R G H N Q L A Y D G K D Y I A L N E D L S S W T A A D T A A Q I T Q L K W E A A R V A E Q L R A Y L E G E C V E W L R R Y L E N G K E T L Q R A D P P K T H V T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D R T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E P S S Q S T V P I V G I V A G L A V L A V V V I G A V V A A V M C R R K S S G G K G G S Y S Q A A C S D S A Q G S D V S L T A (HLA-B*13:02 全长)
54	M R V M A P R T L I L L L S G A L A L T E T W A C S H S M R Y F Y T A V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S P R G E P R A P W V E Q E G P E Y W D R E T Q K Y K R Q A Q A D R V S L R N L R G Y Y N Q S E A G S H T L Q W M Y G C D L G P D G R L L R G Y D Q S A Y D G K D Y I A L N E D L R S W T A A D T A A Q I T Q R K W E A A R E A E Q W R A Y L E G T C V E W L R R Y L E N G K E T L Q R A E H P K T H V T H H P V S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P E P L T L R W E P S S Q T P I V G I V A G L A V L A V L A V L G A V M A V V M C R R K S S G G K G G S C S Q A A S S N S A Q G S D E S L I A C K A (HLA-C*12:03 全长)
55	M A V M A P R T L V L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E G P E Y W D R N T R N V K A H S Q T D R A N L G T L R G Y Y N Q S E D G S H T I Q R M Y G C D V G P D G R F L R G Y Q Q D A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T Q R K W E T A H E A E Q W R A Y L E G R C V E W L R R Y L E N G K E T L Q R T D A P K T H M T H H A V S D H E A T L R C W A L S F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A S V V V P S G Q E Q R Y T C H V Q H E G L P K P L T L R W E P S S Q T P I V G I I A G L V L F G A V I A G A V V A A V M W R R K S S D R K G G S Y S Q A A S S D S A Q G S D M S L T A C K V (HLA-A*26:01 全长)
56	M L V M A P R T V L L L L S A A L A L T E T W A G S H S M R Y F Y T S V S R P G R G E P R F I S V G Y V D D T Q F V R F D S D A A S P R E E P R A P W I E Q E G P E Y W D R N T Q I C K T N T Q T Y R E N L R I A L R Y Y N Q S E A G S H T L Q R M Y G C D V G P D G R L L R G H N Q F A Y D G K D Y I A L N E D L S S W T A A D T A A Q I T Q R K W E A A R V A E Q L R T Y L E G T C V E W L R R Y L E N G K E T L Q R A D P P K T H V T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D R T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E P S S Q S T V P I V G I V A G L A V L A V V V I G A V V A A V M C R R K S S G G K G G S Y S Q A A S S D S A Q G S D V S L T A (HLA-B*38:01 全长)
57	M L V M A P R T V L L L L S A A L A L T E T W A G S H S M R Y F Y T A V S R P G R G E P R F I S V G Y V D D T Q F V R F D S D A A S P R E E P R A P W I E Q E G P E Y W D R N T Q I C K T N T Q T D R E S L R N L R G Y Y N Q S E A G S H T L Q W M Y G C D V G P D G R L L R G Y N Q F A Y D G K D Y I A L N E D L S S W T A A D T A A Q I T Q R K W E A A R E A E Q L R A Y L E G T C V E W L R R H L E N G K E T L Q R A D P P K T H V T H H P I S D H E A T L R C W A L G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D R T F Q K W A A V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W E P S S Q S T V P I V G I V A G L A V L A V V V I G A V V A A V M C R R K S S G G K G G S Y S Q A A S S D S A Q G S D V S L T A (HLA-B*14:02 全长)
58	M A V M A P R T L L L L L L G A L A L T Q T W A G S H S M R Y F T T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S Q R M E P R A P W I E Q E G P E Y W D R N T R N V K A H S Q I D R V D L G T L R G Y Y N Q S E A G S H T I Q M M Y G C D V G S D G R F L R G Y Q Q D A Y D G K D Y I A L N E D L R S W T A A D M A A Q I T Q R K W E A A R V A E Q L R A Y L E G T C V E W L R R H L E N G K E T L Q R T D P P R T H M T H H A V S D H E A T L R C W A L S F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A S V V V P S G Q E Q R Y T C H V Q H E G L P K P L T L R W E P S S Q T P I V G I I A G L V L F G A V F A G A V V A A V R W R



[0617]

	RKSSDRKGGSYSQAASSDSAQGSMSLTACKV (HLA-A*33:01 全长)
59	MAVMAPRTLVLGALALQTWAGSHSMRYFSTSVSRPGRGEPFIAVGYVDDTQ FVRFSDAASQRMEPRAPWIEQEGPEYWDEETGKVKAHSQTDRENLRALRYYNQS EAGSHTLQMMFGCDVGS DGRFLRGYHQAAYDGKDIALKEDLRSWTAADMAAQI TQRKWEAARVAEQLRAYLEGTCVDGLRRYLENGKETLQRTDPPKTHMTHHPISDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVVPSG EEQRYTCHVQHEGLPKPLTLR WEPSSQPTVHIVGIIAGLVLLGAVITGAVVA VMWR RNSSDRKGGSYSQAASSDSAQGSMSLTACKV (HLA-A*23:01 全长)
60	MAVMAPRTLVLGALALQTWAGSHSMRYFYTSVSRPGRGEPFIAVGYVDDT QFVRFSDAASQRMEPRAPWIEQEGPEYWDRNTRNVKAHSQTDRESLRALRYYNQ SEDSHTIQRMYGCDVGP DGRFLRGYQDAYDGKDIALNEDLRSWTAADMAAQI TQRKWETAHEAEQWRAYLEGRCVWLRRYLENGKETLQRTDAPKTHMTHHAVSD HEATLRCWALSFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVVPSG QEQRYTCHVQHEGLPKPLTLR WEPSSQPTIPIVGIAGLVLF GAVIAGAVVA VMWR RKSSDRKGGSYSQAASSDSAQGSMSLTACKV (HLA-A*25:01 全长)
61	MRVTAPRTL LLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEPFISVGYVDDGTQ FVRFSDAASPRTEPRAPWIEQEGPEYWDRNTQISK TNTQTYRESLRNLRGYYNQSE AGSHTLQRMYGCDVGP DGRLLRGHDQSA YDGKDIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRAYLEGTCVEWLRRLHLENGKETLQRADPPKTHVTHHPISDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIAGLAVLAVVVIGAVVATVMCRRK SSGKGGSYSQAASSDSAQGSMSLTA (HLA-B*18:01 全长)
62	MRVTAPRTL LLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEPFISVGYVDDGTQ FVRFSDAASPRTEPRAPWIEQEGPEYWDRETQISK TNTQTYREDLRTLRYYNQSE AGSHTIQRMSGCDVGP DGRLLRGYNQFAYDGKDIALNEDLSSWTAADTAAQITQR KWEAARVAEQDRAYLEGTCVEWLRRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSGEEQ RYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIAGLAVLAVVVIGAVVATVMCRRKS SGGKGGSYSQAASSDSAQGSMSLTA (HLA-B*37:01 full length)
63	MRVTAPRTVL LLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWDRNTQIFK TNTQTYRENLRALRYYNQS EAGSHTWQTMYGCDVGP DGRLLRGHNQYAYDGKDIALNEDLSSWTAADTAAQI TQRKWEAAREAEQLRAYLEGTCVEWLRRLHLENGKETLQRADPPKTHVTHHPVSDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVVPSG EEQRYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIAGLAVLAVVVIGAVVATVMCR RKSSGKGGSYSQAASSDSAQGSMSLTA (HLA-B*51:01 全长)
64	MRVMAPRTL LLLLSGALAL TETWACSHSMRYFSTSVSRPGRGEPFIAVGYVDDTQF VRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQS EAGSHTLQWMFGCDL GPDGRLLRGYDQSA YDGKDIALNEDLRSWTAADTAAQIT QRKWEAAREAEQRRAYLEGTCVEWLRRYLENGKETLQRAEHPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQWDGEDQTQDTEL VETRPAGDGTQKWA AVVVPSGE EQRYTCHVQHEGLPEPLTLR WEPSSQPTIPIVGIAGLAVLAVLAVLAVLAVVAVVMCR RKSSGKGGSCSQAASSNSAQGSDES LIACKA (HLA-C*14:02 全长)
65	MRVMAPRTL LLLLSGALAL TETWACSHSMRYFYTAVSRPGRGEPHFI VGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQNYKRQAQTDRVNLRLRGYYNQ SEAGSHIIQRMYGCDL GPDGRLLRGHDQLAYDGKDIALNEDLRSWTAADTAAQIT QRKWEAAREAEQLRAYLEGTCVEWLRRYLENGKETLQRAEHPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVVPSGE

[0618]

	EQRYTCHVQHEGLPEPLTLRWEPSSQPTIPIVGIVAGLAVLAVLAVLGAVMAVVMC RRKSSGGKGGSCSQAASSNSAQGSDESLIACKA (HLA-C*15:02 全长)
66	MRVMAPRTLTLGALALTETWACSHSMRYFYTAVSRPSRGEPHFIAVGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQTDRVNLKLRGYYNQ SEAGSHTLQRMYGCDLGPDRLLRGYDQSAYDGKDYIALNEDLRSWTAADTAAQI TQRKWEAAREAEQWRAYLEGECEVWLRRYLENGKETLQRAEHPKTHVTHHPVSD HEATLRCWALGFYPTTEITLWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVPS GEEQRYTCHVQHEGLPEPLTLRWEPSSQPTIPIVGIVAGLAVLAVLAVLGAVVAVVM CRRKSSGGKGGSCSQAASSNSAQGSDESLIACKA (HLA-C*02:02 全长)
67	MRVTAPRTLTLGALALTETWAGSHSMRYFHTSVSRPGRGEPFITVGYVDDTL FVRFSDAASPREPRAPWIEQEGPEYWDRETQICKAKAQTDRDLRLLRYYNQSE AGSHTLQNMYGCDVGPDRLLRGYHQDAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRAYLEGECEVWLRRYLENGKETLQRADPPKTHVTHHPISDHEA TLRCWALGFYPAEITLWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVVIGAVVAAVMCR KSSGGKGGSYSQAACSDSAQGSVSLTA (HLA-B*27:05 全长)
68	MAVMAPRTLTLGALALTQTWAGSHSMRYFHTSVSRPGRGEPFIAVGYVDDTQ FVRFSDAASQRMPEPRAPWIEQERPEYWDQETRNKKAHSQIDRVLDLGRGYYNQ EAGSHTIQMMYGCDVGS DGRFLRGYQDAYDGKDYIALNEDLRSWTAADMAAQI TQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDPPKTHMTHHAVSD HEATLRCWALSFPYPAEITLWQRDGEDQTQDTEL VETRPAGDGTQKWA SVVPSG QEQRYTCHVQHEGLPKPLTLRWEPSSQPTIPIVGIAGLVLFGAVFAGAVVAAVRWR RKSSDRKGGSYSQAASSDSAQGS DMSLTACKV (HLA-A*31:01 全长)
69	MAVMAPRTLTLGALALTQTWAGSHSMRYFSTSVSRPGSGEPFIAVGYVDDTQ FVRFSDAASQRMPEPRAPWIEQERPEYWDQETRNKKAHSQTDREN LGTLRGYYNQ SEAGSHTIQIMYGCDVGS DGRFLRGYEQHAYDGKDYIALNEDLRSWTAADMAAQIT QRKWEAARRAEQLRAYLEGTCVEWLRRYLENGKETLQRTDPPKTHMTHHPISDHE ATLRCWALGFYPAEITLWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVPSGE EQRYTCHVQHEGLPKPLTLRWELSSQPTIPIVGIAGLVLLGAVITGAVVAAVMWR KSSDRKGGSYTQAASSDSAQGS DVS LTACKV (HLA-A*30:02 全长)
70	MLVMAPRTVLLLLSAALALTETWAGSHSMRYFYTSVSRPGRGEPFISVGYVDDTQ FVRFSDAASPREPRAPWIEQEGPEYWD RNTQIYKAQAQTDRSLRNLRGYYNQS EAGSHTLQSMYGCDVGPDRLLRGHNQYAYDGKDYIALNEDLRSWTAADTAAQIT QRKWEAARVAEQDRAYLEGTCVEWLRRYLENGKDTLERADPPKTHVTHHPISDHE ATLRCWALGFYPAEITLWQRDGEDQTQDTEL VETRPAGDRTFQKWA AVVPSGE EQRYTCHVQHEGLPKPLTLRWEPSSQSTVPIVGIVAGLAVLAVVIGAVVAAVMCR RKSSGGKGGSYSQAACSDSAQGS DVS LTAA (HLA-B*42:01 全长)
71	MRVMAPQALLLTLGALAL IETWAGSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQADR VNLKLRGYYN QSEAGSHTIQRMYGCDLGPDRLLRGYNQFAYDGKDYIALNEDLRSWTAADTAAQI SQRKLEAAREAEQLRAYLEGECEVWLRGYLENGKETLQRAERP KTHVTHHPVSDH EATLRCWALGFYPAEITLWQRDGEDQTQDTEL VETRPAGDGTQKWA AVVPSG QEQRYTCHVQHEGLQEPCTLRWKPSQPTIPNLGIVSGPAVLAVLAVLAVLAVLGA VVA AVIHRKSSGGKGGSCSQAASSNSAQGSDES LIACKA (HLA-C*17:01 全长)
72	MRVTAPRTVLLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWD RNTQIFKTNTQTYRESLRNLRGYYNQS EAGSHIQRMYGCDLGPDRFLRGHNQYAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRAYLEGLCEVWLRRYLENGKETLQRADPPKTHVTHHPVSDHEA



[0619]

	TLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSQSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRK SSGGKGGSYSQAASSDSAQGSVSLTA (HLA-B*35:02 全长)
73	MLVMAPRTVLLLLSAALALTETWAGSHSMRYFYTSVSRPGRGEPFISVGYVDDTQ FVRFSDAASPREEPRAPWIEQEGPEYWDRNTQICKTNTQTDRESLRNLRGYYNQSE AGSHTWQTMYGCDVGPDRLLRGNQFAYDGKDIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRQTYLEGTCVEWLRRLRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVPSGEEQ RYTCHVQHEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRRK SSGGKGGSYSQAASSDSAQGSVSLTA (HLA-B*39:06 全长)
74	MRVMAPRTLILLLSGALALTETWAGSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQ SEAGSHILQRMYGCDVGPDRLLRQYDQSAYDGKDIALNEDLRSWTAADTAAQIT QRKWEAAREAEQLRAYLEGLCVEWLRRLRYLKNKGTQLQRAEHPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQWDGEDQTQDTELVETRPAGDGTQFQKWAAVVPSGE EQRYTCHVQHEGLPEPLTLRWEPSQSTIPIVGIVAGLAVLAVLAVLGAVVAVVMCR RKSSGGKGGSCSAASSNSAQGSDESLIACKA (HLA-C*03:02 全长)
75	MRVTAPRTVLLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWDGETRNMKASAQTYRENLRALRYYNQ SEAGSHIIQRMYGCDLGPDRLLRQHDQSAYDGKDIALNEDLSSWTAADTAAQIT QRKWEAARVAEQLRAYLEGLCVEWLRRLRYLENGKETLQRADPPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWAAVVPSGE EQRYTCHVQHEGLPKPLTLRWEPSQSTIPIVGIVAGLAVLAVVVIGAVVATVMCRR KSSGGKGGSYSQAASSDSAQGSVSLTA (HLA-B*58:01 全长)
76	MAVMAPRTLILLLLGALALTQWAGSHSMRYFSTVSRPGRGEPFIAVGYVDDTQ FVRFSDAASQRMPEPRAPWIEQEGPEYWDRNTRNVKAHSQIDRVLDLGTLRGYYNQ SEAGSHTIQMMYGCDVGSDFRFLRGYQQDAYDGKDIALNEDLRSWTAADMAAQI TQRKWEAARVAEQLRAYLEGLCVEWLRRLRYLENGKETLQRTDPPKTHMTHHAVSD HEATLRCWALSFPYPAEITLTWQRDGEDQTQDTELVETRPAGDGTQFQKWASVVVPSG QEQRYTCHVQHEGLPKPLTLRWEPSQSTIPIVGIIAGLVLFAGVAVVAVRWR RKSSDRKGGSYSQAASSDSAQGSMDSLTACKV (HLA-A*33:03 全长)
77	MAVMAPRTLVLILLSGALALTQWAGSHSMRYFYTSMSRPGRGEPFIAVGYVDDT QFVRFSDAASQRMPEPRAPWIEQEGPEYWDRNTRNVKAQSQTDRVDLGTLRGYYN QSEAGSHTIQRMYGCDVGPDRFLRGYHQYAYDGKDIALKEDLRSWTAADMAA QTTKHKWEAAHVAEQWRAYLEGLCVEWLRRLRYLENGKETLQRTDAPKTHMTHHA VSDHEATLRCWALSFPYPAEITLTWQRDGEDQTQDTELVETRPAGDGTQFQKWAVV VPSGQEQRYTCHVQHEGLPKPLTLRWEPSQSTIPIVGIIAGLVLFAGAVITGAVVAV MWRRKSSDRKGGSYSQAASSDSAQGSVSLTACKV (HLA-A*68:02 全长)
78	MRVMAPRTLILLLSGALALTETWACSHSMKYFFTSVSRPGRGEPFISVGYVDDTQF VRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQTDRVSLRNLRGYYNQS EAGSHTLQWMCGLDGPDRLLRQYDQYAYDGKDIALNEDLRSWTAADTAAQIT QRKWEAAREAEQRRAYLEGLCVEWLRRLRYLENGKETLQRAEHPKTHVTHHPVSDHE ATLRCWALGFYPAEITLTWQWDGEDQTQDTELVETRPAGDGTQFQKWAAVMVPSGE EQRYTCHVQHEGLPEPLTLRWEPSQSTIPIVGIVAGLAVLAVLAVLGAVVAVVMCR RKSSGGKGGSCSAASSNSAQGSDESLIACKA (HLA-C*01:02 全长)
79	MRVMAPRALLLLSGGLALTETWACSHSMRYFDTA VSRPGRGEPFISVGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQADRSLRNLRGYYNQ SEDSHTFQRMYGCDLGPDRLLRQYDQFAYDGKDIALNEDLRSWTAADTAAQI

[0620]

	TQRKLEAARAAEQDRAYLEGTCVEWLRRYLENGKKTQLQRAEPPKTHVTHHPLSDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDGTQKWA AVVVPSG QEQRYSCHMQHEGLQEPLTSLWEPSSQPTIPIMGIVAGLAVLVVLA VLGA VVTAMM CRRKSSGGKGGSCSQAACSNSA QGSDESLITCKA (HLA-C*07:04 全长)
80	MAVMAPRTLVL LLSGALALTQTWAGSHSMRYFYTSVSRPGRGEPFIAVGYVDDT QFVRFSDAASQRMEPRAPWIEQEGPEYWDNRNTRNVKAQSQTDRVDLGT LRGYYN QSEAGSHTIQMMYGCDVGS DGRFLRGYRQDAYDGKDYIALKEDLRSWTAADMAA QTTKHKWEAAHVAEQWRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHA VSDHEATLRCWALS FYP AEITLTWQRDGEDQTQDTELVETRPAGDGTQKWA AVV VPSGQEQRYSCHVQHEGLPKPLTLR WEPSSQPTIPIVGHAGLVLF GAVITGAVVA AV MWRKSSDRKGGSYSQAASSDSA QGSDVSLTACKV (HLA-A*68:01 全长)
81	MAVMAPRTLL LLLL GALALTQTWAGSHSMRYFFTSVSRPGRGEPFIAVGYVDDTQ FVRFSDAASQRMEPRAPWIEQEGPEYWDQETRNVKAHSQTDRESLR IALRYYNQS EAGSHTIQMMYGCDV GPDGRLLRGYQQDAYDGKDYIALNEDLRSWTAADMAAQI TQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHAVSD HEATLRCWALS FYP AEITLTWQRDGEDQTQDTELVETRPAGDGTQKWA SVVPSG QEQRYSCHVQHEGLPKPLTLR WEPSSQPTIPIVGHAGLVLF GAVITGAVVA AV RKS SDRKGGSYSQAASSDSA QGSDMSLTACKV (HLA-A*32:01 全长)
82	MRVTAPRTVL LLLS AALALTETWAGSHSMRYFHTAMSRPGRGEPFITVGYVDDTL FVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRENLR IALRYYNQSE AGSHTWQRMYGCDL GPDGRLLRGYNQLAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAAREAEQLRAYLEGLCVEWLRRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWA AVVVPSGEEQ RYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIVAGLAVLAVV VIGAVVATVMCRRKS SGGKGGSYSQAASSDSA QGSDVSLTA (HLA-B*49:01 全长)
83	MRVTAPRTVL LLLW GAVALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWDNRNTQIFKTNTQTYRENLR IALRYYNQS EAGSHIIQRMYGCDL GPDGRLLRGHDQSAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRAYLEGLCVEWLRRYLENGKETLQRADPPKTHVTHHPVSDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIVAGLAVLAVV VIGAVVATVMCRRK SSGGKGGSYSQAASSDSA QGSDVSLTA (HLA-B*53:01 全长)
84	MRVTAPRTVL LLLS AALALTETWAGSHSMRYFHTAMSRPGRGEPFITVGYVDDTL FVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNL RGYYNQSE AGSHTWQRMYGCDL GPDGRLLRGYNQLAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAAREAEQLRAYLEGLCVEWLRRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRTFQKWA AVVVPSGEEQ RYTCHVQHEGLPKPLTLR WEPSSQSTIPIVGIVAGLAVLAVV VIGAVVATVMCRRKS SGGKGGSYSQAASSDSA QGSDVSLTA (HLA-B*50:01 全长)
85	MAVMAPRTLVL LLSGALALTQTWAGSHSMRYFYTSVSRPGRGEPFIAVGYVDDT QFVRFSDAASRRMEPRAPWIEQEGPEYWDGETRNVKAHSQTHRVDLGT LRGYYN QSEAGSHTLQRMYGCDVGS DWRFLRGYHQA YDGKDYIALKEDLRSWTAADMAA QTTKHKWEAAHVAEQWRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHA VSDHEATLRCWALS FYP AEITLTWQRDGEDQTQDTELVETRPAGDGTQKWA AVV VPSGQEQRYSCHVQHEGLPKPLTLR WEPSSQPTIPIVGHAGLVLF GAVITGAVVA AV MWRKSSDRKGGSYSQAASSDSA QGSDVSLTACKV (HLA-A*02:05 全长)
86	MRVTAPRTLL LLLW GALALTETWAGSHSMRYFYTAMSRPGRGEPFIAVGYVDDT QFVRFSDAASPREEPRAPWIEQEGPEYWDNRNTQIYKAQAQT DRESLRNL RGYYNQ

[0621]

	SEAGSHTWQTMYGCDLGPDRLLRGHNQLAYDGKDYIALNEDLSSWTAADTAAQI TQRKWEAAREAEQLRAYLEGTCVEWLRRYLENGKETLQRADPPKTHVTHHPISDHE ATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSGE EQRYTCHVQHEGLPKPLTLRWEPSQSSTIPIVGIVAGLAVLAVVVIGAVVATVMCRR KSSGGKGGSYSQAASSDSAQGS DVS LTA (HLA-B*55:01 全长)
87	MRVTAPRTVLLLLSAALALTETWAGSHSMRYFHTAMSRPGRGEPFRFITVGYVDDTL FVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNLRGYYNQSE AGSHTWQRMYGCDLGPDRLLRGYNQLAYDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQDRAYLEGLCVESLRRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSGEEQ RYTCHVQHEGLPKPLTLRWEPSQSSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRKS SGGKGGSYSQAASSDSAQGS DVS LTA (HLA-B*45:01 全长)
88	MRVTAPRTVLLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFRFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWDRETQISK TNTQTYRENLR LIALRYYNQS EAGSHTWQTMYGCDVGPDRLLRGHNQYAYDGKDYIALNEDLSSWTAADTAAQI TQRKWEAAREAEQLRAYLEGLCVESLRRHLENGKETLQRADPPKTHVTHHPVSDH EATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSG EEQRYTCHVQHEGLPKPLTLRWEPSQSSTIPIVGIVAGLAVLAVVVIGAVVATVMCR RKSSGGKGGSYSQAASSDSAQGS DVS LTA (HLA-B*52:01 全长)
89	MRVMAPRTLILLLSGALALTETWACSHSMRYFYTAVSRPGRGEPFRFIAVGYVDDTQ FVRFSDAASPRGEPRAPWVEQEGPEYWDRETQKYKRQAQADR VSLRNLRGYYNQ SEAGSHTLQRMYGCDLGPDRLLRGYDQSA YDGKDYIALNEDLRSWTAADTAAQI TQRKWEAAREAEQWRAYLEGTCVEWLRRYLENGKETLQRAEHPKTHVTHHPVSD HEATLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDGT FQKWA AVVVPS GEEQRYTCHVQHEGLPEPLTLRWEPSQSSTIPIVGIVAGLAVLAVLAVLGAVMAVV MCRRKSSGGKGGSCSQAASSNSA QGSDES LIACKA (HLA-C*12:02 全长)
90	MRVTAPRTVLLLLWGAVALTETWAGSHSMRYFYTAMSRPGRGEPFRFIAVGYVDDT QFVRFSDAASPRTEPRAPWIEQEGPEYWD RNTQIFK TNTQTYRESLRNLRGYYNQS EAGSHIQRMYGCDLGPDRLLRGHDQFA YDGKDYIALNEDLSSWTAADTAAQITQ RKWEAARVAEQLRAYLEGLCVESLRRYLENGKETLQRADPPKTHVTHHPVSDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSQSSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRK SSGGKGGSYSQAASSDSAQGS DVS LTA (HLA-B*35:03 全长)
91	MRVTAPRTLLLLLWGAVALTETWAGSHSMRYFHTSVSRPGRGEPFRFITVGYVDDTL FVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNLRGYYNQSE AGSHTLQSMYGCDVGPDRLLRGHNQYAYDGKDYIALNEDLRSWTAADTAAQITQ RKWEAARVAEQLRAYLEGECEVWLRRYLENGKETLQRADPPKTHVTHHPISDHEA TLRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSGEE QRYTCHVQHEGLPKPLTLRWEPSQSSTVPIVGIVAGLAVLAVVVIGAVVA VMCR KSSGGKGGSYSQAACSDSAQGS DVS LTA (HLA-B*40:02 全长)
92	MRVTAPRTVLLLLSGALALTETWAGSHSMRYFYTAMSRPGRGEPFRFISVGYVDDTQ FVRFSDAASPREEPRAPWIEQEGPEYWDRETQISK TNTQTYRESLRNLRGYYNQSE AGSHTLQRMYGCDVGPDRLLRGHDQSA YDGKDYIALNEDLSSWTAADTAAQITQ RKWEAAREAEQLRAYLEGLCVESLRRYLENGKETLQRADPPKTHVTHHPISDHEAT LRCWALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDRFTFQKWA AVVVPSGEEQ RYTCHVQHEGLPKPLTLRWEPSQSSTIPIVGIVAGLAVLAVVVIGAVVATVMCRRKS SGGKGGSYSQAASSDSAQGS DVS LTA (HLA-B*15:03 全长)
93	MAVMAPRTLILLLSGALALTQTRAGSHSMRYFFTSVSRPGRGEPFRFIAVGYVDDTQ

[0622]

	FVRFSDAASQRMEPRAPWIEQEGPEYWDQETRNVKAHSQTDRVDLGLTRGYYNQ SEAGSHTIQMMYGCDVGPDRLLRQYQDAYDYGKDYIALNEDLRSWTAADMAAQ ITQRKWEAARVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHAVSD HEATLRCWALSFPYAEITLTWQRDGEDQTQDTELVEVTRPAGDGTQFQK WASVVVPSG QEQRYSYCHVQHEGLPKPLTLRWEPSQPTIPIVGIAGLVLFAMFAGAVVAVRWR RKSSDRKGGSSQAASSDSAQGSMSLTACKV (HLA-A*74:01 全长)
94	GSHSMRYFFTSVSRPGRGEPFRFIAVGIVDDTQFVRFSDAASQKMEPRAPWIEQEGP EYWDQETRNVKAHSQTDRANLGLTRGYYNQSEAGSHTIQIMYGCDVGPDRFLRG YRQDAYDYGKDYIALNEDLRSWTAADMAAQITKRKWEAVHAAEQRRVYLEGRCDV GLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSS (HLA-A*01:01 可溶性)
95	GSHSMRYFFTSVSRPGRGEPFRFIAVGIVDDTQFVRFSDAASQRMEPRAPWIEQEGP EYWDQETRNVKAQSQTDRVDLGLTRGYYNQSEAGSHTIQIMYGCDVGSDFRFLRG YRQDAYDYGKDYIALNEDLRSWTAADMAAQITKRKWEAAHEAEQLRAYLDGTCVE WLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSS (HLA-A*03:01 可溶性)
96	GSHSMRYFYTSVSRPGRGEPFRFIAVGIVDDTQFVRFSDAASQRMEPRAPWIEQEGP EYWDQETRNVKAQSQTDRVDLGLTRGYYNQSEAGSHTIQIMYGCDVGPDRFLRG YRQDAYDYGKDYIALNEDLRSWTAADMAAQITKRKWEAAHAAEQRAYLEGRCVE WLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSS (HLA-A*11:01 可溶性)
97	GSHSMRYFSTSVSRPGRGEPFRFIAVGIVDDTQFVRFSDAASQRMEPRAPWIEQEGP EYWDEETGKVKKAHSQTDRENLRALRYYNQSEAGSHTLQMMFGCDVGSDFRFLRG YHQYAYDYGKDYIALKEDLRSWTAADMAAQITKRKWEAAHVAEQRAYLEGTCVD GLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-A*24:02 可溶性)
98	GSHSMRYFYTSVSRPGRGEPFRFISVGIVDDTQFVRFSDAASPREPRAPWIEQEGPE YWDRNTQIYKAQAQTDRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDRLLRGH DQYAYDYGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGECEVW LRRYLENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVEVTRPAGDRTFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*07:02 可溶性)
99	GSHSMRYFSTSVSWPGRGEPFRFIAVGIVDDTQFVRFSDAASPRGEPREPWVEQEGP EYWDRETQKYKRQAQADRVLNRKLRGYYNQSEAGSHTLQRMFGCDLGPDRLLR GYNQFAYDYGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGTCVE WLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWKPS S (HLA-C*04:01 可溶性)
100	CSHSMRYFDTA VSRPGRGEPFRFISVGIVDDTQFVRFSDAASPRGEPREPWVEQEGP EYWDRETQKYKRQAQADRVSRLNLRGYYNQSEAGSHTLQRMSGCDLGPDRLLR GYDQSAYDYGKDYIALNEDLRSWTAADTAAQITQRKLEAARAAEQRAYLEGTCVE WLRRYLENGKETLQRAEPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVEVTRPAGDGTQFQK WAAVVVPSGEEQRYTCHMQHEGLQEPLTWSWEP S (HLA-C*07:02 可溶性)
101	GSHSMRYFDTAMSRPGRGEPFRFISVGIVDDTQFVRFSDAASPREPRAPWIEQEGP EYWDRNTQIFKTNTQTDRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDRLLRG HNQYAYDYGKDYIALNEDLRSWTAADTAAQITQRKWEAARVAEQDRAYLEGTCVE WLRRYLENGKDTLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGE

[0623]

	DQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*08:01 可溶性)
102	GSHSMRYFYTAMSRPGRGEPRIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDRNTQIFKTNQTYRESLRNLRGYYNQSEAGSHIIQRMYGCDLGPDRLLR DQSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQLRAYLEGLCWE RRLRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGEDQ TQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*35:01 可溶性)
103	GSHSMRYFYTAMSRPGRGEPRIAVGYVDDTQFVRFSDAASPRMAPRAPWIEQEG PEYWDGETRNMKASAQTYRENLRALRYYNQSEAGSHIIQVMYGCDVGPDRLLR GHDQSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQLRAYLEGLCWE WLRRLRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*57:01 可溶性)
104	GSHSMRYFYTAMSRPGRGEPRIAVGYVDDTQFVRFSDAASPRMAPRAPWIEQEG PEYWDGETRNMKASAQTYRENLRALRYYNQSEAGSHIIQVMYGCDVGPDRLLR GHNQYAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQLRAYLEGLCWE WLRRLRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*57:03 可溶性)
105	GSHSLKYFHTSVSRPGRGEPRIAVGYVDDTQFVRFSDAASPRMVPRAPWMEQEG SEYWDRETRSARDTAQIFRVNLRRLRGYYNQSEAGSHTLQWMHGCELGPDRFLR GYEQFAYDGKDYIALNEDLRSWTAADTAAQISEQKSNDAEAEHQRAYLEDTCVE WLHKYLEKGGKETLLHLEPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQQDGE GHTQDTELVETRPAGDGTQFQKWAAVVVPSGEEQRYTCHVQHEGLPEPVTLRWKPA S (HLA-E 可溶性)
106	CSHSMRYFYTAVSRPGRGEPRIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEG PEYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHTLQWMYGCDLGPDRLLR RGYDQSAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAARAAEQRAYLEGTC VEWLRRLRYLENGKETLQRAEHPKTHVTHHLVSDHEATLRCWALGFYPAEITLTWQR DGEDQTQDTELVETRPAGDGTQFQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRW EPSS (HLA-C*16:01 可溶性)
107	CSHSMRYFYTAVSRPGRGEPRIAVGYVDDTQFVQFSDAASPRGEPRAPWVEQEG PEYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHTLQRMYGCDLGPDRLLR GYNQFAYDGKDYIALNEDLRSWTAADKAAQITQRKWEAAREAEQRAYLEGTCVE WLRRLRYLENGKKTQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRD EDQTQDTELVETRPAGDGTQFQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWG PS (HLA-C*08:02 可溶性)
108	CSHSMRYFDTAVSRPGRGEPRIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEGP EYWDRETQNYKRQAQADRVSRLRNLRGYYNQSEAGSHTLQRMYGCDLGPDRLLR GYDQSAYDGKDYIALNEDLRSWTAADTAAQITQRKLEAARAAEQRAYLEGTCVE WLRRLRYLENGKETLQRAEPPKTHVTHHPLSDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDGTQFQKWAAVVVPSGEEQRYTCHMQHEGLQEPLTLRWEPS S (HLA-C*07:01 可溶性)
109	CSHSMRYFYTAVSRPGRGEPRIAVGYVDDTQFVQFSDAASPRGEPRAPWVEQEG PEYWDRETQKYKRQAQTDRVNLRLRGYYNQSEAGSHTLQRMYGCDLGPDRLLR RGYNQFAYDGKDYIALNEDLRSWTAADKAAQITQRKWEAAREAEQRAYLEGTCV EWLRRLRYLENGKKTQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRD GEDQTQDTELVETRPAGDGTQFQKWAAVVVPSGEEQRYTCHVQHEGLPEPLTLRW G PSS (HLA-C*05:01 可溶性)



[0624]

110	GSHSMRYFYTAMSRPGRGEPFRFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRENLR TALRYYNQSEAGSHIIQRMYGCDVGPDRLLRGY DQDAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQDRAYLEGLCVESL RRYLENGKETLQRADPPKTHVTHHPISDHEVTLRCWALGFYPAEITLTWQRDGEDQ TQDTEL VETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*44:02 可溶性)
111	GSHSMRYFTTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRM EPRAPWIEQEGP EYWDLQTRNVKAQSQTDRANLGLTRGYYNQSEAGSHTIQMMYGCDVGS DGRFLR GYRQDAYDGKDYIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCV EWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALS FYPAEITLTWQRD GEDQTQDTEL VETRPAGDGTQK WASVVVPSGQEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-A*29:02 可溶性)
112	GSHSMRYFYTAMSRPGRGEPFRFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRENLR TALRYYNQSEAGSHIIQRMYGCDVGPDRLLRGY DQDAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLCVESL RRYLENGKETLQRADPPKTHVTHHPISDHEVTLRCWALGFYPAEITLTWQRDGEDQ TQDTEL VETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*44:03 可溶性)
113	GSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQFVRFSDAASPRGEP RAPWVEQEG PEYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHIIQRMYGCDVGPDRLLR GYDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVE WLRRYLKNGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTEL VETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*03:04 可溶性)
114	GSHSMRYFHTAMSRPGRGEPFRFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRESLRNLRGYYNQSEAGSHTLQRMYGCDVGPDRLLRG HNQYAYDGKDYIALNEDLRSWTAADTAAQISQRKLEAARVAEQLRAYLEGE CV LRRYLENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTEL VETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*40:01 可溶性)
115	CSHSMRYFDTAVSRPGRGEPFISVGYVDDTQFVRFSDAASPRGEP RAPWVEQEGP EYWDRETQKYKRQAQADRVNLRKLRGYYNQSEDSHTLQWMYGCDLGPDRLLR GYDQSAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQWRAYLEGTCV EWLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRD GEDQTQDTEL VETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWE PSS (HLA-C*06:02 可溶性)
116	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRMAPRAPWIEQEG PEYWDRETQISK TNTQTYRESLRNLRGYYNQSEAGSHTLQRMYGCDVGPDRLLRG HDQSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAAREAEQWRAYLEGLC VE WLRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTEL VETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*15:01 可溶性)
117	GSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQFVRFSDAASPRGEP RAPWVEQEG PEYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEARSHIIQRMYGCDVGPDRLLR GYDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGLC VE WLRRYLKNGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTEL VETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*03:03 可溶性)
118	GSHSMRYFSTSVSRPGSGEPFIAVGYVDDTQFVRFSDAASQRM EPRAPWIEQERP EYWDQETRNKVAQSQTDRVDLGLTRGYYNQSEAGSHTIQIMYGCDVGS DGRFLRG YEQHAYDGKDYIALNEDLRSWTAADMAAQITQRKWEAARWAEQLRAYLEGTCVE

[0625]

	WLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSS (HLA-A*30:01 可溶性)
119	GSHSMRYFYTAMSRPGRGEPFITVGYVDDTQFVRFDS DATSPRMAPRAPWIEQEGP EYWDRETQISK TNTQTYRENLR TALRYYNQSEAGSHTWQTMYGCDLGPDRLLRG HNQLAYDGKDIALNEDLSSWTAADTAAQITQLKWEAARVAEQLRAYLEGECEVW LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSS (HLA-B*13:02 可溶性)
120	C SHSMRYFYTAVSRPGRGEPFIAVGYVDDTQFVRFDS DAASPRGEPRAPWVEQEG PEYWDRETQYK RQAQADRVSLRNLRGYYNQSEAGSHTLQWMYGCDLGPDRLL RGYDQSA YDGKDIALNEDLRSWTAADTAAQITQRKWEAAREAEQWRAYLEGTC VEWLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQR DGEDQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRW EPSS (HLA-C*12:03 可溶性)
121	GSHSMRYFYTSVSRPGRGEPFIAVGYVDDTQFVRFDS DAASQRMEPRAPWIEQEGP EYWDRNTRNVKAHSQTDRANLGTLRGYYNQSEDSHTIQRMYGCDVGPDRFLRG YQQDAYDGKDIALNEDLRSWTAADMAAQITQRKWETAHEAEQWRAYLEGRCVE WLRRYLENGKETLQRTDAPKTHMTHH AVSDHEATLRCWALSFP AEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA SVVVPSGQE QRYTCHVQHEGLPKPLTLRWEP S (HLA-A*26:01 可溶性)
122	GSHSMRYFYTSVSRPGRGEPFISVGYVDDTQFVRFDS DAASPREEPRAPWIEQEGPE YWDRNTQICK TNTQTYRENLR IALRYYNQSEAGSHTLQRMYGCDVGPDRLLRGH NQFAYDGKDIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRTYLEGTCVEWL RRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQ TQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSS (HLA-B*38:01 可溶性)
123	GSHSMRYFYTAVSRPGRGEPFISVGYVDDTQFVRFDS DAASPREEPRAPWIEQEGP EYWDRNTQICK TNTQTDRESLRNLRGYYNQSEAGSHTLQWMYGCDVGPDRLLRG YNQFAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGTCVEW LRRHLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSS (HLA-B*14:02 可溶性)
124	GSHSMRYFTTSVSRPGRGEPFIAVGYVDDTQFVRFDS DAASQRMEPRAPWIEQEGP EYWDRNTRNVKAHSQIDRV DLGTLRGYYNQSEAGSHTIQMMYGCDVGS DGRFLRG YQQDAYDGKDIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCVE WLRRHLENGKETLQRTDPPRTHMTHH AVSDHEATLRCWALSFP AEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA SVVVPSGQE QRYTCHVQHEGLPKPLTLRWEP S (HLA-A*33:01 可溶性)
125	GSHSMRYFSTSVSRPGRGEPFIAVGYVDDTQFVRFDS DAASQRMEPRAPWIEQEGP EYWDEETGKVKKAHSQTDRENLR IALRYYNQSEAGSHTLQMMFGCDVGS DGRFLRG YHQYAYDGKDIALKEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCVD GLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPSS (HLA-A*23:01 可溶性)
126	GSHSMRYFYTSVSRPGRGEPFIAVGYVDDTQFVRFDS DAASQRMEPRAPWIEQEGP EYWDRNTRNVKAHSQTDRESLR IALRYYNQSEDSHTIQRMYGCDVGPDRFLRG YQQDAYDGKDIALNEDLRSWTAADMAAQITQRKWETAHEAEQWRAYLEGRCVE WLRRYLENGKETLQRTDAPKTHMTHH AVSDHEATLRCWALSFP AEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA SVVVPSGQE QRYTCHVQHEGLPKPLTLRWEP S (HLA-A*25:01 可溶性)

[0626]

127	GSHSMRYFHTSVSRPGRGEPFISVGYVDGTQFVRFSDAASPRTEPRAPWIEQEGPE YWDRNTQISKNTQTYRESLRNLRGYYNQSEAGSHTLQRMYGCDVGPDRLLRGH DQSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQLRAYLEGTCVEWL RRHLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQ TQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*18:01 可溶性)
128	GSHSMRYFHTSVSRPGRGEPFISVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGPE YWDRETQISKNTQTYREDLRTLRLRYYNQSEAGSHTIQRMSGCDVGPDRLLRGIN QFAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQDRAYLEGTCVEWLR RYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGEDQT QDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*37:01 可溶性)
129	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDRNTQIFKNTQTYRENLRALRYYNQSEAGSHTWQTMYGCDVGPDRLLRG HNQYAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVE WLRRHLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDG EDQTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS S (HLA-B*51:01 可溶性)
130	CSHSMRYFSTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEGP EYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHTLQWMFGCDLGPDRLLR GYDQSA YDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGTCVE WLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*14:02 可溶性)
131	CSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEG PEYWDRETQNYKRQAQTDRVNLRLKRGYYNQSEAGSHIIQRMYGCDLGPDRLLR GHDQLAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGTCVE WLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*15:02 可溶性)
132	CSHSMRYFYTAVSRPSRGEPHFIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEGP EYWDRETQKYKRQAQTDRVNLRLKRGYYNQSEAGSHTLQRMYGCDLGPDRLLR GYDQSA YDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQWRAYLEGECV EWLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPTTEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*02:02 可溶性)
133	GSHSMRYFHTSVSRPGRGEPFITVGYVDDTLFVRFSDAASPREPRAPWIEQEGPE YWDRETQICKAKAQTDRDLRTLRLRYYNQSEAGSHTLQNMYGCDVGPDRLLRGIN HQDAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQLRAYLEGECVEW LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*27:05 可溶性)
134	GSHSMRYFTTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMPEPRAPWIEQERP EYWDQETRNKVAHSQIDRVLDGLTRGYYNQSEAGSHTIQMMYGCDVGS DGRFLRG YQQDAYDGKDYIALNEDLRSWTAADMAAQITQRKWEARVAEQLRAYLEGTCVE WLRRYLENGKETLQRTDPPKTHMTHHAVSDHEATLRCWALSFPYPAEITLTWQRDG EDQTQDTELVETRPAGDGTQKWA SVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS S (HLA-A*31:01 可溶性)
135	GSHSMRYFSTSVSRPGSGEPFIAVGYVDDTQFVRFSDAASQRMPEPRAPWIEQERP EYWDQETRNKVAHSQTDRENGLTRGYYNQSEAGSHTIQIMYGCDVGS DGRFLRG



[0627]

	YEQHAYDGKDYIALNEDLRSWTAADMAAQITQRKWEARRAEQLRAYLEGTCVE WLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVE TRPAGDGT FQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWELSS (HLA-A*30:02 可溶性)
136	GSHSMRYFYTSVSRPGRGEPFISVGYVDDTQFVRFSDAASPREEPRAPWIEQEGPE YWRNTQIYKAQAQTDRESLRNLRGYYNQSEAGSHTLQSMYGC DVGPDGRLLRGH NQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEARVAEQDRAYLEGTCVEW LRRYLENGKDTLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVE TRPAGDRTFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*42:01 可溶性)
137	GSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEG PEYWDRETQKYKRQAQADRVNLRKLRGYYNQSEAGSHTIQRMYGCDLGPDRLLR GYNQFAYDGKDYIALNEDLRSWTAADTAAQISQRKLEAAREAEQLRAYLEGECEVE WLRGYLENGKETLQRAERP KTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGE EDQTQDTELVE TRPAGDGT FQK WAAVVVPSGEEQRYTCHVQHEGLQEPCTLRWKP SS (HLA-C*17:01 可溶性)
138	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWRNTQIFKTNTQTYRESLRNLRGYYNQSEAGSHIIQRMYGCDLGPDRFLRGH NQYAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQDRAYLEGLCVEW LRRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVE TRPAGDRTFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*35:02 可溶性)
139	GSHSMRYFYTSVSRPGRGEPFISVGYVDDTQFVRFSDAASPREEPRAPWIEQEGPE YWRNTQICKTNTQTDRESLRNLRGYYNQSEAGSHTWQTM YGCDVGPDRLLRG HNQFAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQDRAYLEGLCVEW LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVE TRPAGDRTFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*39:06 可溶性)
140	GSHSMRYFYTAVSRPGRGEPHFIAVGYVDDTQFVRFSDAASPRGEPRAPWVEQEG PEYWDRETQKYKRQAQTD RVS LRNLRGYYNQSEAGSHILQRMYGCDVGPDRLLR GYDQSA YDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVE WLRRYLKNGETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTELVE TRPAGDGT FQK WAAVVVPSGEEQRYTCHVQHEGLPEPLTLRWEPS S (HLA-C*03:02 可溶性)
141	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDGETRNMKASAQTYRENLR IALRYYNQSEAGSHIIQRMYGCDLGPDRLLRG HDQSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEARVAEQDRAYLEGLCVEW LRRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVE TRPAGDRTFQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*58:01 可溶性)
142	GSHSMRYFTTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMEPRAPWIEQEGP EYWRNTRNVKAHSQIDRV DLGTLRGYYNQSEAGSHTIQMMYGCDVGS DGRFLRG YQQDAYDGKDYIALNEDLRSWTAADMAAQITQRKWEARVAEQDRAYLEGTCVE WLRRYLENGKETLQRTDPPKTHMTHHAVSDHEATLRCWALSFP AEITLTWQRDGE EDQTQDTELVE TRPAGDGT FQK WASVVVPSGEEQRYTCHVQHEGLPKPLTLRWEPS S (HLA-A*33:03 可溶性)
143	GSHSMRYFYTSMSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMEPRAPWIEQEG PEYWRNTRNVKAQSQTDRVDLGT LRGYYNQSEAGSHTIQRMYGCDVGPDRFLR GYHQYAYDGKDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQWRAYLEGTC VEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALSFP AEITLTWQR DGEDQTQDTELVE TRPAGDGT FQK WAAVVVPSGEEQRYTCHVQHEGLPKPLTLRW EPSS

[0628]

	(HLA-A*68:02 可溶性)
144	CSHSMKYFFTSVSRPGRGEPFISVGYVDDTQFVRFSDAASPRGEPRAPWVEQEGP EYWDRETQKYKRQAQTDRVSLRNLRGYYNQSEAGSHTLQWMCGLDGPDRLLR GYDQYAYDGKDIALNEDLRSWTAADTAAQITQRKWEAAREAEQRRAYLEGTCVE WLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQWDG EDQTQDTELVETRPAGDGTQKWAAMVPSGEEQRYTCHVQHEGLPEPLTLRWEP SS (HLA-C*01:02 可溶性)
145	CSHSMRYFDTA VSRPGRGEPFISVGYVDDTQFVRFSDAASPRGEPRAPWVEQEGP EYWDRETQKYKRQAQADR VSLRNLRGYYNQSEAGSHTFQRMYGCDLGPDRLLR GYDQFAYDGKDIALNEDLRSWTAADTAAQITQRKLEAARAAEQDRAYLEGTCVE WLRRYLENGKKTQRAEPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDGTQKWAAMVPSGEEQRYTCHMQHEGLQEPLTLRWEPS S (HLA-C*07:04 可溶性)
146	GSHSMRYFYTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMPEPRAPWIEQEGP EYWDRNTRNVKAQSQTDRVLDGTLRGYYNQSEAGSHTIQMMYGCDVGS DGRFLR GYRQDAYDGKDIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQWRAYLEGTC VEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALSFPYAEITLTWQR DGEDQTQDTELVETRPAGDGTQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRW EPSS (HLA-A*68:01 可溶性)
147	GSHSMRYFFTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMPEPRAPWIEQEGP EYWDQETRNVKAHSQTDRESLRIALRYYNQSEAGSHTIQMMYGCDVGPDRLLRG YQDDAYDGKDIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCVE WLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALSFPYAEITLTWQRDGE EDQTQDTELVETRPAGDGTQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRWEPS S (HLA-A*32:01 可溶性)
148	GSHSMRYFHTAMSRPGRGEPFITVGYVDDTLFVRFSDATS PRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRENLRALRYYNQSEAGSHTWQRMYGCDLGPDRLLRG YNQLAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLC VEWL LRRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*49:01 可溶性)
149	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDRNTQIFK TNTQTYRENLRALRYYNQSEAGSHIIQRMYGCDLGPDRLLR GH DQSAYDGKDIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLC VEWL RRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGEDQ TQDTELVETRPAGDRTFQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*53:01 可溶性)
150	GSHSMRYFHTAMSRPGRGEPFITVGYVDDTLFVRFSDATS PRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRESLRLRGYYNQSEAGSHTWQRMYGCDLGPDRLLRG YNQLAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLC VEWL LRRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRWEPS (HLA-B*50:01 可溶性)
151	GSHSMRYFYTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASRRMEPRAPWIEQEGP EYWDGETRKVKAHSQTHRVDLGTLRGYYNQSEAGSHTLQRMYGCDVGS DWRFLR GYHQYAYDGKDIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQWRAYLEGTC VEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALSFPYAEITLTWQR DGEDQTQDTELVETRPAGDGTQKWAAMVPSGEEQRYTCHVQHEGLPKPLTLRW EPSS (HLA-A*02:05 可溶性)
152	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPREPRAPWIEQEGP

[0629]

	EYWDRNTQIYKAQAQTDRESLRNLRGYYNQSEAGSHTWQTMYGCDLGPDRLLR GHNQLAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGTCVE WLRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGE DQTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*55:01 可溶性)
153	GSHSMRYFHTAMSRPGRGEPFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGP EYWDRETQISK TNTQTYRESLRNLRGYYNQSEAGSHTWQRMYGCDLGPDRLLRG YNQLAYDGKDIALNEDLSSWTAADTAAQITQRKWEAARVAEQDRAYLEGLCVES LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*45:01 可溶性)
154	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDRETQISK TNTQTYRENLRALRYYNQSEAGSHTWQTMYGCDVGPDRLLRG HNQYAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVE WLRRLHLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGE DQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*52:01 可溶性)
155	CSHSMRYFYTAVSRPGRGEPFIAVGYVDDTQFVRFSDAASPRGEPAPWVEQEG PEYWDRETQKYKRQAQADRVSRLNLRGYYNQSEAGSHTLQRMYGCDLGPDRLL RGYDQSA YDGKDIALNEDLRSWTAADTAAQITQRKWEAAREAEQWRA YLEGTC VEWLRRYLENGKETLQRAEHPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQR DGEDQTQDTELVETRPAGDGTQFQKWA AVVVPSGEEQRYTCHVQHEGLPEPLTLR EPSS (HLA-C*12:02 可溶性)
156	GSHSMRYFYTAMSRPGRGEPFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGP EYWDRNTQIFK TNTQTYRESLRNLRGYYNQSEAGSHIQRMYGCDLGPDRLLRGH DQFAYDGKDIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLCVEWL RRYLENGKETLQRADPPKTHVTHHPVSDHEATLRCWALGFYPAEITLTWQRDGEDQ TQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*35:03 可溶性)
157	GSHSMRYFHTSVSRPGRGEPFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGP YWDRETQISK TNTQTYRESLRNLRGYYNQSEAGSHTLQSMYGCDVGPDRLLRGH NQYAYDGKDIALNEDLRSWTAADTAAQITQRKWEAARVAEQLRAYLEGECEVW LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*40:02 可溶性)
158	GSHSMRYFYTAMSRPGRGEPFISVGYVDDTQFVRFSDAASPREPRAPWIEQEGP EYWDRETQISK TNTQTYRESLRNLRGYYNQSEAGSHTLQRMYGCDVGPDRLLRG HDQSAYDGKDIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVEW LRRYLENGKETLQRADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDGED QTQDTELVETRPAGDRTFQKWA AVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-B*15:03 可溶性)
159	GSHSMRYFFTSVSRPGRGEPFIAVGYVDDTQFVRFSDAASQRMPEPRAPWIEQEGP EYWDQETRNVKAHSQTDRLVGLTRGYYNQSEAGSHTIQMMYGCDVGPDRLLR GYQQDAYDGKDIALNEDLRSWTAADMAAQITQRKWEAARVAEQLRAYLEGTCV EWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCWALSFPYPAEITLTWQRD GEDQTQDTELVETRPAGDGTQFQKWASVVVPSGEEQRYTCHVQHEGLPKPLTLRWE PSS (HLA-A*74:01 可溶性)
160	MSRSVALAVLALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNLCYVSGFHPSDIEVD LL KNGERIEKVEHSDLSFSKDW SFYLLYYTEFTPTEKDEYACRVNHVTL SQPKIVKWDR DM (全长人β-2-微球蛋白)

[0630]

161	GSGSAGGGLNDIFEAQKIEWHEGSTGHHHHHHHDYKDDDDK (具有 his6 标签和 Flag 标签的 Avitag 序列)
162	GSGSAGGSGSGGGSLPETGGHHHHHH (具有 His6 标签的分选标签序列)
163	LPXTG, wherein X is any amino acid (分选标签基序)
164	IPKTG (分选标签基序)
165	MPXTG, 其中 X 是任意氨基酸 (分选标签基序)
166	LAETG (分选标签基序)
167	LPXAG, 其中 X 是任意氨基酸 (分选标签基序)
168	LPESG (分选标签基序)
169	LPELG (分选标签基序)
170	LPEVG (分选标签基序)
171	XPKTG, 其中 X 是任意氨基酸 (分选标签基序)
172	APKTG (分选标签基序)
173	DPKTG (分选标签基序)
174	SPKTG (分选标签基序)
175	LPEXG, 其中 X 是任意氨基酸 (分选标签基序)
176	LPEAG (分选标签基序)
177	LPECG (分选标签基序)
178	LPEXG, ), 其中 X = A, C 或 S (分选标签基序)
179	WTWTW (ligation control motif)
180	GSGSAGGSFESGPGAIEYCFNVDDPSKDSKAQVSAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAV GNAESRYVLTGRYDSAPATDGSALTGWTVAWKNNYRHAHSATTWSGQYVGGAEARINTQWLLTSG TTEANAWKSTLVGHDTFTKVKPSAASIDAACKAGVNNGNPLDAVQQGSTGDYKDDDDK (N-内含蛋白序列, 具有 Flag 标签)
181	(GGGGS) <sub>n</sub> , 其中 n=1-6 (接头)
182	SSSSGSSSSGSAA (接头)
183	GGGGG (接头)
184	S(GGGGS) <sub>n</sub> , 其中 n=1-10 (接头)
185	(GGSG) <sub>n</sub> , 其中 n=1-5

[0631]

	(接头)
186	GSAT (接头)
187	(GGSGGS) <sub>n</sub> , 其中 n=1-5 (接头)
188	DDDDK (肠激酶, 切割位点)
189	KPVSKMRMATPLLMQA (CLIP 肽)
190	QIYKANSKFIGITEL (TT p2 肽)
191	IKEEHVIIQAEFYLNPDQSGEFMFDFDGDGEIFHVDMAKKETVWRLEEFGRFASFEAQ GALANIAVDKANLEIMTKRSNYTPITNVPPEVTVLTNSPVELREPNVLICFIDKFTPPV VNVTWLRNGKPVTTGVSETVFLPREDHLFRKFHYLPFLPSTEDVYDCRVEHWGLDE PLLKHWEFDAPSPLPETTEGSEQKLISEEDLPETGG (HLA-DRA*01:01 Myc 标签和分选标签)
192	KPVSKMRMATPLLMQAGGGGSIEGRGSGGGSGDTRPRFLWQLKFECHFFNGTERVR LLERCIYNQEESVRFSDVGEYRAVTELGRPDAEYWNSQKDLLEQRRAAVDITYCRH NYGVGESFTVQRRVEPKVTVYPSKTQPLQHNNLLVCSVSGFYPGSIEVRWFRNGQE EKAGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSE SAQSKLGGLNDFEAQKIEWHEHHHHHH (HLA-DRB1*01:01, 具有 N 末端 CLIP 肽、可消化接头和 C 末端 AviTag 和 His6 标签)
193	生物素-PKYVKQNTLKLAT (来自甲型流感病毒的 HA 肽)
194	MAISGVPVLGFFIIAVLMSAQESWAIKEEHVIIQAEFYLNPDQSGEFMFDFDGDGEIFHV DMAKKETVWRLEEFGRFASFEAQGALANIAVDKANLEIMTKRSNYTPITNVPPEV VLTNSPVELREPNVLICFIDKFTPPVNVTWLRNGKPVTTGVSETVFLPREDHLFRKF HYLPFLPSTEDVYDCRVEHWGLDEPLLKHWEFDAPSPLPETTENVVCALGLTVGLV GIIIGTIFIIGVVRKSNAEERRGPL (HLA-DRA*01:01 全长)
195	MVCLKLPGGSCMTALTVTLMVLSSPLALAGDTRPRFLWQLKFECHFFNGTERVRLL ERCIYNQEESVRFSDVGEYRAVTELGRPDAEYWNSQKDLLEQRRAAVDITYCRHN YGVGESFTVQRRVEPKVTVYPSKTQPLQHNNLLVCSVSGFYPGSIEVRWFRNGQEE KAGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*01:01 全长)
196	MVCLKLPGGSCMTALTVTLMVLSSPLALAGDTRPRFLWQLKFECHFFNGTERVRLL ERCIYNQEESVRFSDVGEYRAVTELGRPDAEYWNSQKDLLEQRRAAVDITYCRHN YGAVESFTVQRRVEPKVTVYPSKTQPLQHNNLLVCSVSGFYPGSIEVRWFRNGQEE KAGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*01:02 全长)
197	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRYL DRYFHNQEENVRFSDVGEFRAVTELGRPDAEYWNSQKDLLEQKRGRVDNYCRHN YGVVESFTVQRRVHPKVTYPSKTQPLQHNNLLVCSVSGFYPGSIEVRWFRNGQEE KTGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*03:01 全长)
198	MVCLKFPGGSCMAALTVTLMVLSSPLALAGDTRPRFLEQVKHECHFFNGTERVRFL DRYFYHQEEYVRFSDVGEYRAVTELGRPDAEYWNSQKDLLEQKRAAVDITYCRH NYGVGESFTVQRRVPEVTVYPAKTQPLQHNNLLVCSVNGFYPGSIEVRWFRNGQE EKTGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSLTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*04:01 全长)

[0632]

199	MVCLKFPGGSCMAALTVTLMVLSSPLALAGDTRPRFLEQVKHECHFFNGTERVRFL DRYFYHQEEYVRFSDVGEYRAVTELGRPDAEYWNSQKDLLEQRRAAVDTYCRHN YGVVESFTVQRRVYPEVTVYPAKTQPLQHHNLLVCSVNGFYPGSIEVRWFRNGQEE KTGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSLTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*04:04 全长)
200	MVCLKLPGGSCMAALTVTLMVLSSPLALAGDTQPRFLWQGYKCHFFNGTERVQF LERLFYNQEEFVRFSDVGEYRAVTELGRPVAESWNSQKDILEDRRGQVDTVCRHN YGVGESFTVQRRVHPEVTVYPAKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEE KAGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVMSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*07:01 全长)
201	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTGECYFFNGTERVRFL DRYFYNQEEYVRFSDVGEYRAVTELGRPSAEYWNSQKDFLEDRRALVDTYCRHN YGVGESFTVQRRVHPKVTYVPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEE KTGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWSARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*08:01 全长)
202	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEEVKFECHFFNGTERVRLLE RRVHNQEEYARYDSDVGEYRAVTELGRPDAEYWNSQKDLLERRRAAVDTYCRHN YGVGESFTVQRRVQPKVTYVPSKTQPLQHHNLLVCSVNGFYPGSIEVRWFRNGQEE KTGVVSTGLIQNGDWFQTLVMLETVPQSGEVYTCQVEHPSVMSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLPPTGFLS (HLA-DRB1*10:01 全长)
203	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRFLD RYFYNQEEYVRFSDVGEFRAVTELGRPDEEYWNSQKDFLEDRAAVDTYCRHNY GVGESFTVQRRVHPKVTYVPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEK TGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESA QSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*11:01 全长)
204	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRFLD RYFYNQEEYVRFSDVGEFRAVTELGRPDEEYWNSQKDFLEDRAAVDTYCRHNY GVVESFTVQRRVHPKVTYVPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEK TGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESA QSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*11:04 全长)
205	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRFLD RYFHNQEEENVRFSDVGEFRAVTELGRPDAEYWNSQKDILEDERRAAVDTYCRHNY GVVESFTVQRRVHPKVTYVPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEK TGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESA QSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*13:01 全长)
206	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRFLD RYFHNQEEENVRFSDVGEFRAVTELGRPDAEYWNSQKDILEDERRAAVDTYCRHNY GVGESFTVQRRVHPKVTYVPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEK TGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESA QSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*13:02 全长)
207	MVCLRLPGGSCMAVLTVTLMVLSSPLALAGDTRPRFLEYSTSECHFFNGTERVRFLD RYFHNQEEFVRFSDVGEYRAVTELGRPAAEHWNSQKDLLERRRAEVDTYCRHNY GVVESFTVQRRVHPKVTYVPSKTQPLQHYNLLVCSVSGFYPGSIEVRWFRNGQEEK TGVVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESA QSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPRGFLS (HLA-DRB1*14:01 全长)
208	MVCLKLPGGSCMTALTVTLMVLSSPLALSGLDTRPRFLWQPKRECHFFNGTERVRFL



[0633]

	DRYFYNQEESVRFDSVDVGEFRAVTELGRPDAEYWNSQKDILEQARAADVITYCRHN YGVVESFTVQRRVQPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFLNGQEE KAGMVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*15:01 全长)
209	MVCLKLPGGSCMTALTVTLMVLSSPLALSQDTRPRFLWQPKRECHFFNGTERVRFL DRHFYNQEESVRFDSVDVGEFRAVTELGRPDAEYWNSQKDILEQARAADVITYCRHN YGVVESFTVQRRVQPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFLNGQEE KAGMVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSES AQSKMLSGVGGFVLGLLFLGAGLFIYFRNQKGHSGLQPTGFLS (HLA-DRB1*15:03 全长)
210	MILNKALLL GALALTTVMSPCGGEDIVADHVASCNVLYQFYGPSGQYTHEFDGDE EFYVDLERKETAWRWPEFSKFGGFDPPQALRNMAVAKHNLNIMIKRYNSTAATNE VPEVTVFSKSPVTLGQPNTLCLVDNIFPPVNVITWLSNGQSVTEGVSETSFLSKSDHS FFKISYLTFLPSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVG LVGIVVGTVFIIQGLRSVGASRHQGPL (HLA-DQA1*01:01 全长)
211	MSWKKSLRIPGDLRVATVTLMLAILSSSLAEGRDSPEDFVYQFKGLCYFTNGTERVR GVTRHIYNREEYVRFDSVDVGYRAVTPQGRPVAEYWNSQKEVLEGARASVDRVCR HNYEVAYRGILQRRVEPTVTISPSRTEALNHHNLLICSVDYFYSQIKVRWFRNDQEE TAGVVSTPLIRNGDWFQILVMLEMTPQRGDVYTCHEHPSLQSPITVEWRAQSESA QSKMLSGVGGFVLGLIFLGLIIRQRSRKGLLH (DQB1*05:01 全长)
212	MILNKALLL GALALTTVMSPCGGEDIVADHVASCNVLYQFYGPSGQYTHEFDGDE QFYVDLERKETAWRWPEFSKFGGFDPPQALRNMAVAKHNLNIMIKRYNSTAATNE VPEVTVFSKSPVTLGQPNTLCLVDNIFPPVNVITWLSNGQSVTEGVSETSFLSKSDHS FFKISYLTFLPSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVG LMGIVVGTVFIIQGLRSVGASRHQGPL (HLA-DQA1*01:02 全长)
213	MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMICYFTNGTERV RLVTRYIYNREEYARFDSVDVGYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVC RHNYEVAFRGILQRRVEPTVTISPSRTEALNHHNLLVCSVTDYFYPGQIKVRWFRNDQ EETAGVVSTPLIRNGDWFQILVMLEMTPQRGDVYTCHEHPSLQSPITVEWRAQSE SAQSKMLSGVGGFVLGLIFLGLIIRQRSQKGLLH (HLA-DQB1*06:02 全长)
214	MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYSHEFDGD EEFYVDLERKETVWQLPLFRFRFPQFALTNIAVLKHNLNIVIKRSNSTAATNEVP EVTVFSKSPVTLGQPNTLCLVDNIFPPVNVITWLSNGHSVTEGVSETSFLSKSDHSFF KISYLTFLPSADEIYDCKVEHWGLDEPLLKHWEPEIPTPMSELTETVVCALGLSVGLV GIVVGTVLIIRGLRSVGASRHQGPL (HLA-DQA1*03:01 全长)
215	MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMICYFTNGTER VRLVTRYIYNREEYARFDSVDVGYRAVTPGPPAAEYWNSQKEVLERTRAELDTVC RHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDYFPAQIKVRWFRNDQ EETTGVVSTPLIRNGDWFQILVMLEMTPQRGDVYTCHEHPSLQNPPIIVEWRAQSE SAQSKMLSGIGGFVLGLIFLGLIHHRSQKGLLH (HLA-DQB1*03:02 全长)
216	MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGD EQFYVDLGRKETVWCLPVLQRFRFPQFALTNIAVLKHNLSLIKRSNSTAATNEVP EVTVFSKSPVTLGQPNILICLVDNIFPPVNVITWLSNGHSVTEGVSETSFLSKSDHSFF KISYLTLLPSAEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGL VGIVVGTVFIIRGLRSVGASRHQGPL (HLA-DQA1*05:01 全长)
217	MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKGMICYFTNGTER VRLVRSIYNREEIVRFDSVDVGEFRAVTLGLPAAEYWNSQKDILERKRAAVDRVCR

[0634]

	HNYQLELRRTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPAQIKVRWFRNDQE ETAGVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQSPITVEWRAQSES AQSKMLSGIGGFVLGLIFLGLGLIIHRSQKGLLH (HLA-DQB1*02:01 全长)
218	MSWKKALRIPGGLRAATVTLMLAMLSTPVAEGRDSPEDFVYQFKAMCYFTNGTER VRYVTRYIYNREEYARFSDVGYRAVTPGPPDAEYWNSQKEVLERTRAEALDTC RHNYQLELRRTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPAQIKVRWFRNDQ EETGTVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQNPITVEWRAQSE SAQSKMLSGIGGFVLGLIFLGLGLIIHRSQKGLLH (HLA-DQB1*03:01 全长)
219	MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTER VRLVTRYIYNREEYARFSDVGYRAVTPGPPDAEYWNSQKEVLERTRAEALDTC RHNYQLELRRTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPAQIKVRWFRNDQ EETGTVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQNPPIVEWRAQSE SAQSKMLSGIGGFVLGLIFLGLGLIIHRSQKGLLH (HLA-DQB1*03:03 全长)
220	MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTER VRGVTRYIYNREEYARFSDVGYRAVTPGRLDAEYWNSQKDILEEDRASVDTC RHNYQLELRRTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPAQIKVRWFRNDQ EETGTVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQNPPIVEWRAQSE SAQSKMLSGIGGFVLGLIFLGLGLIIHRSQKGLLH (HLA-DQB1*04:02 全长)
221	MSWKKSLRIPGDLRVATVTLMLAILSSSLAEGRDSPEDFVYQFKGLCYFTNGTERVR GVTRHIYNREEYVRFSDVGYRAVTPQGRPDAEYWNSQKEVLEGARASVDRVCR HNYEVAYRGILQRRVEPTVTISPSRTEALNHHNLLICSVTDVFYPSQIKVRWFRNDQEE TAGVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQSPITVEWRAQSESA QSKMLSGVGGFVLGLIFLGLGLIIRQRSRKGPPAGLLH (HLA-DQB1*05:03 全长)
222	MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMCYFTNGTER VRLVTRHIYNREEYARFSDVGYRAVTPQGRPDAEYWNSQKEVLEGTRAEALDTC CRHNYEVAFRGILQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPGQIKVRWFRND QEETAGVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQSPITVEWRAQSE ESAQSKMLSGVGGFVLGLIFLGLGLIIRQRSQKGLLH (HLA-DQB1*06:03 全长)
223	MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMCYFTNGTER VRLVTRHIYNREEYARFSDVGYRAVTPQGRPVAEYWNSQKEVLERTRAEALDTC CRHNYEVGYRGILQRRVEPTVTISPSRTEALNHHNLLVCSVTDVFYPGQIKVQWFRND QEETAGVVSTPLIRNGDWFQILVMLEMPQRGDVYTCHVEHPSLQSPITVEWRAQSE ESAQSKMLSGVGGFVLGLIFLGLGLIIRQRSQKGLLH (HLA-DQB1*06:04 全长)
224	IKEEHVIIQAEFYLNPDQSGEFMFDGDEIFHVDMAKKTVWRLEEFGRFASFEAQ GALANIAVDKANLEIMTKRSNYTPITNVPPEVTVLTNSPVELREPNVLCIFDKFTPPV VNVTWLRNGKPVTTGVSETVFLPREDHLFRKFHYLPFLPSTEDVYDCRVEHWGLDE PLLKHWEFDAPSPLPETTE (HLA-DRA*01:01 可溶性)
225	GDTRPRFLWQLKFECHFFNGTERVRLLERCIYNQEESVRFSDVGEYRAVTELGRPD AEYWNSQKDLLEQRRAAVDTYCRHNYGVGESFTVQRRVEPKVTVYPSKTQPLQHH NLLVCSVSGFYPGSIEVRWFRNGQEEKAGVVSTGLIQNGDWFQTLVMLETVPRSG EYVTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*01:01 可溶性)
226	GDTRPRFLWQLKFECHFFNGTERVRLLERCIYNQEESVRFSDVGEYRAVTELGRPD AEYWNSQKDLLEQRRAAVDTYCRHNYGAVESFTVQRRVEPKVTVYPSKTQPLQHH NLLVCSVSGFYPGSIEVRWFRNGQEEKAGVVSTGLIQNGDWFQTLVMLETVPRSG EYVTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*01:02 可溶性)



[0635]

227	GDTRPRFLEYSTSECHFFNGTERVRYLDRYFHNQEENVRFDSDVGEFRAVTELGRPDAEYWNSQKDLLEQKRGRVDNYCRHNYGVVESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*03:01 可溶性)
228	GDTRPRFLEQVKHECHFFNGTERVRFDRYFYHQEYVRFDSVGEYRAVTELGRPDAEYWNSQKDLLEQKRAAVDTYCRHNYGVGESFTVQRRVYPEVTVYPAKTQPLQHHNLLVCSVNGFYPGSIEVRWFRNGQEEKTG VVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSLTSPLTVEWRARSESAQSK (HLA-DRB1*04:01 可溶性)
229	GDTRPRFLEQVKHECHFFNGTERVRFDRYFYHQEYVRFDSVGEYRAVTELGRPDAEYWNSQKDLLEQRRAAVDTYCRHNYGVVESFTVQRRVYPEVTVYPAKTQPLQHHNLLVCSVNGFYPGSIEVRWFRNGQEEKTG VVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSLTSPLTVEWRARSESAQSK (HLA-DRB1*04:04 可溶性)
230	GDTQPRFLWQGYKCHFFNGTERVQFLERLFYNQEEFVRFDSVGEYRAVTELGRPVAESWNSQKDILEDRRGQVDTYCRHNYGVGESFTVQRRVHPEVTVYPAKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKAGVVSTGLIQNGDWFQTLVMLETVPRSGEVYTCQVEHPSVMSPLTVEWRARSESAQSK (HLA-DRB1*07:01 可溶性)
231	GDTRPRFLEYSTGECYFFNGTERVRFDRYFYHQEYVRFDSVGEYRAVTELGRPSAEYWNSQKDFLEDRRALVDTYCRHNYGVGESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWSARSESAQSK (HLA-DRB1*08:01 可溶性)
232	GDTRPRFLEEVKFECHFFNGTERVRLLEVVHNQEEYARYDSDVGEYRAVTELGRPDAEYWNSQKDLLERRRAAVDTYCRHNYGVGESFTVQRRVQPKVTVYPSKTQPLQHHNLLVCSVNGFYPGSIEVRWFRNGQEEKTG VVSTGLIQNGDWFQTLVMLETVPQS GEVYTCQVEHPSVMSPLTVEWRARSESAQSK (HLA-DRB1*10:01 可溶性)
233	GDTRPRFLEYSTSECHFFNGTERVRFDRYFYHQEYVRFDSVGEFRAVTELGRPDEEYWNSQKDFLEDRRAAVDTYCRHNYGVGESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*11:01 可溶性)
234	GDTRPRFLEYSTSECHFFNGTERVRFDRYFYHQEYVRFDSVGEFRAVTELGRPDEEYWNSQKDFLEDRRAAVDTYCRHNYGVVESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*11:04 可溶性)
235	GDTRPRFLEYSTSECHFFNGTERVRFDRYFHNQEENVRFDSDVGEFRAVTELGRPDAEYWNSQKDILEDERRAAVDTYCRHNYGVVESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*13:01 可溶性)
236	GDTRPRFLEYSTSECHFFNGTERVRFDRYFHNQEENVRFDSDVGEFRAVTELGRPDAEYWNSQKDILEDERRAAVDTYCRHNYGVGESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*13:02 可溶性)
237	GDTRPRFLEYSTSECHFFNGTERVRFDRYFHNQEEFVRFDSVGEYRAVTELGRPAAEHWNSQKDLLERRRAEVDYCRHNYGVVESFTVQRRVHPKVTVYPSKTQPLQHHNLLVCSVSGFYPGSIEVRWFRNGQEEKTG VVSTGLIHNGDWFQTLVMLETVPRSGEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*14:01 可溶性)

[0636]

238	GDTRPRFLWQPKRECHFFNGTERVRFDRYFYNQEESVRFSDVGEFRAVTELGRP DAEYWNSQKDILEQARAAVDTYCRHNYGVVESFTVQRRVQPKVTVYPSKTQPLQH HNLLVCSVSGFYPGSIEVRWFLNGQEEKAGMVSTGLIQNGDWFQTLVMLETVPRS GEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*15:01 可溶性)
239	GDTRPRFLWQPKRECHFFNGTERVRFDRHFYFNQEESVRFSDVGEFRAVTELGRP DAEYWNSQKDILEQARAAVDTYCRHNYGVVESFTVQRRVQPKVTVYPSKTQPLQH HNLLVCSVSGFYPGSIEVRWFLNGQEEKAGMVSTGLIQNGDWFQTLVMLETVPRS GEVYTCQVEHPSVTSPLTVEWRARSESAQSK (HLA-DRB1*15:03 可溶性)
240	EDIVADHVASCVNLYQFYGPSGQYTHEFDGDEEFYVDLERKETAWRWPEFSKFGG FDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVD NIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFLPSADEIYDCKVEHWG LDQPLLKHWEPEIPAPMSELTET (HLA-DQA1*01:01 可溶性)
241	GRDSPEDFVYQFKGLCYFTNGTERVRGVTRHIYNREEYVRFSDVGVYRAVTPQGR PVAEYWNSQKEVLEGARASVDRVCRHNYEVAYRGILQRRVEPTVTISPSRTEALNH HNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLIRNGDWFQILVMLEMPQRG DVYTCHVEHPSLQSPITVEWRAQSESAQSK (HLA-DQB1*05:01 可溶性)
242	EDIVADHVASCVNLYQFYGPSGQYTHEFDGDEQFYVDLERKETAWRWPEFSKFG GFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVD DNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFLPSADEIYDCKVEHW GLDQPLLKHWEPEIPAPMSELTET (HLA-DQA1*01:02 可溶性)
243	GRDSPEDFVYQFKGMCYFTNGTERVRLVTRYIYNREEYARFSDVGVYRAVTPQGR PDAEYWNSQKEVLEGRTRAEALDTCRHNIEVAFRGILQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLIRNGDWFQILVMLEMPQRG DVYTCHVEHPSLQSPITVEWRAQSESAQSK (HLA-DQB1*06:02 可溶性)
244	EDIVADHVASYGVNLYQSYGPSGQYSHEFDGDEEFYVDLERKETVWQLPLFRRFR FDPQFALTNI AVLKHNLNIVIKRSNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNI FPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFLPSADEIYDCKVEHWGL DEPLLKHWEPEIPTPMSELTET (HLA-DQA1*03:01 可溶性)
245	GRDSPEDFVYQFKGMCYFTNGTERVRLVTRYIYNREEYARFSDVGVYRAVTPQGR PAAEYWNSQKEVLERTRAEALDTCRHNIEVAFRGILQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPAQIKVRWFRNDQEETAGVVSTPLIRNGDWFQILVMLEMPQRG DVYTCHVEHPSLQNPITVEWRAQSESAQSK (HLA-DQB1*03:02 可溶性)
246	EDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYVDLGRKETVWCLPVLQRFRF DPQFALTNI AVLKHNLNLSLIKRSNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNIF PPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLLPSAEESYDCKVEHWGLD KPLLKHWEPEIPAPMSELTET (HLA-DQA1*05:01 可溶性)
247	GRDSPEDFVYQFKGMCYFTNGTERVRLVRSIYNREEIVRFSDVGEFRAVTLGLP AAEYWNSQKDILERKRAAVDRVCRHNYQLELRITTLQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPAQIKVRWFRNDQEETAGVVSTPLIRNGDWFQILVMLEMPQRG DVYTCHVEHPSLQNPITVEWRAQSESAQSK (HLA-DQB1*02:01 可溶性)
248	GRDSPEDFVYQFKAMCYFTNGTERVRYVTRYIYNREEYARFSDVGVYRAVTPQGR PDAEYWNSQKEVLERTRAEALDTCRHNIEVAFRGILQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPAQIKVRWFRNDQEETAGVVSTPLIRNGDWFQILVMLEMPQHG DVYTCHVEHPSLQNPITVEWRAQSESAQSK (HLA-DQB1*03:01 可溶性)

[0637]

249	GRDSPEDFVYQFKGMCYFTNGTERVRLVTRYIYNREEYARFSDSDVGVYRAVTPPLGP PDAEYWNSQKEVLERTRAELDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLIRNGDWTQILVMLEMPQRG DVYTCHVEHPSLQNPPIVEWRAQSESAQSK (HLA-DQB1*03:03 可溶性)
250	GRDSPEDFVYQFKGMCYFTNGTERVRGVTRYIYNREEYARFSDSDVGVYRAVTPPLGR LDAEYWNSQKDILEEDRASVDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLIRNGDWTQILVMLEMPQRG DVYTCHVEHPSLQNPPIVEWRAQSESAQSK (HLA-DQB1*04:02 可溶性)
251	GRDSPEDFVYQFKGLCYFTNGTERVRGVTRHIYNREEYVRFSDSDVGVYRAVTPQGR PDAEYWNSQKEVLEGARASVDRVCRHNYEVA YRGILQRRVEPTVTISPSRTEALNH HNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLIRNGDWTQILVMLEMPQRG DVYTCHVEHPSLQSPITVEWRAQSESAQSK (HLA-DQB1*05:03 可溶性)
252	GRDSPEDFVYQFKGMCYFTNGTERVRLVTRHIYNREEYARFSDSDVGVYRAVTPQGR PDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFRGILQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLIRNGDWTQILVMLEMPQRG DVYTCHVEHPSLQSPITVEWRAQSESAQSK (HLA-DQB1*06:03 可溶性)
253	GRDSPEDFVYQFKGMCYFTNGTERVRLVTRHIYNREEYARFSDSDVGVYRAVTPQGR PVAEYWNSQKEVLERTRAELDTVCRHNYE VGYRGILQRRVEPTVTISPSRTEALNHH NLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLIRNGDWTQILVMLEMPQRG DVYTCHVEHPSLQSPITVEWRAQSESAQSK (HLA-DQB1*06:04 可溶性)
254	GLNDIFEAQKIEWHEGSGEQKLISEEDLHHHHHHH (avitag-Myc-His (生物素介导))
255	GLNDIFEAQKIEWHEGSGEQKLISEEDL (avitag-Myc (生物素介导))
256	GSGSAGSGSGGGSLPETGGHHHHHHH (分选标签-His (点击偶联))
257	GSGSAGSGSGGGSLPETGG (分选标签 (点击偶联))
258	GKPIPPLLGLDST (V5)
259	LTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAA (Fos)
260	RIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNH (Jun)
261	TTAPSAQLEKELQALQKENAQLEWELQALEKELAQ (酸性亮氨酸拉链)
262	TTAPSAQLKKKLQALKKKNAQLKWKLQALKKKLAQ (碱性亮氨酸拉链)
263	EPKSADKTHTCPPCAPEAAGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEV KFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA LGAPIEKTKAKGQPREPQVYTLPPCRDELTKNQVSLWCLVKGFYPSDIAVEWESN GQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCFSVMHEALHNHYTQKS LSLSPGK (杵 (杵白))
264	EPKSADKTHTCPPCAPEAAGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEV KFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKA LGAPIEKTKAKGQPREPQVCTLPPSRDELTKNQVSLSCAVKGFYPSDIAVEWESNG QPENNYKTPPVLDSDGSFFLVSKLTVDKSRWQQGNVFCFSVMHEALHNHYTQKSL

[0638]

	SLSPGK (白 (杵白))
265	RGVPHIVMVDAYKRYK (spyttag)
266	VTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEEDGRELAGATMELRDSSGKTISTWI SDGHVKDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVDGEATEGDAHT (spycatcher)
267	DYKDDDDK (flag)
268	WSHPQFEK (strep-tag)
269	EDQVDPRLIDGK (蛋白 C 标签)
270	VEPKSC (人 IgG1 的上铰链序列)
271	SVRDJLARI (A02:03 占位符肽)
272	LTAJFLIFL (A02:06 占位符肽)
273	LLDSDJERL (A02:07 占位符肽)
274	KMDIJVPLL (A02:11 占位符肽)
275	FYVJGAANR (A33:03 占位符肽)
276	ILGPPGJVY (B15:02 占位符肽)
277	EEFGAAJSF (B44:05 占位符肽)
278	KMKEIAJAY (B46:01 占位符肽)
279	KPWDJIPMV (B55:02 占位符肽)
280	ATPLLMQALPMGA (CLIP 肽)
281	PKYVKQNTLKLAT (流感血凝素表位)
282	ELAGIGILTV (对照表位)

示范性点击化学手柄和反应。

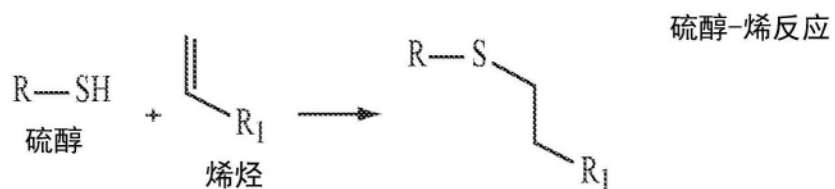
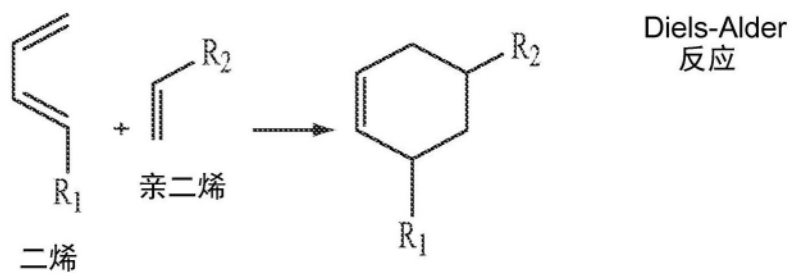
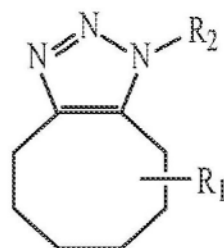
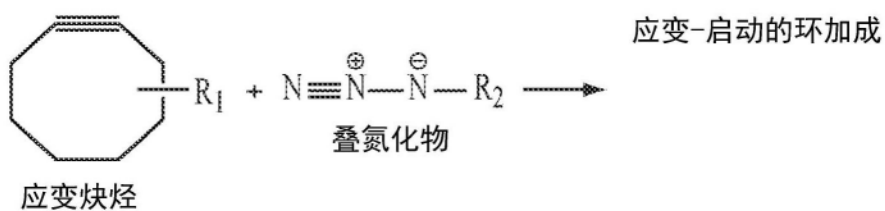
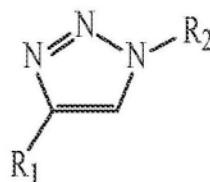
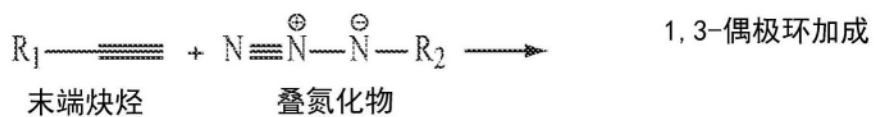


图1

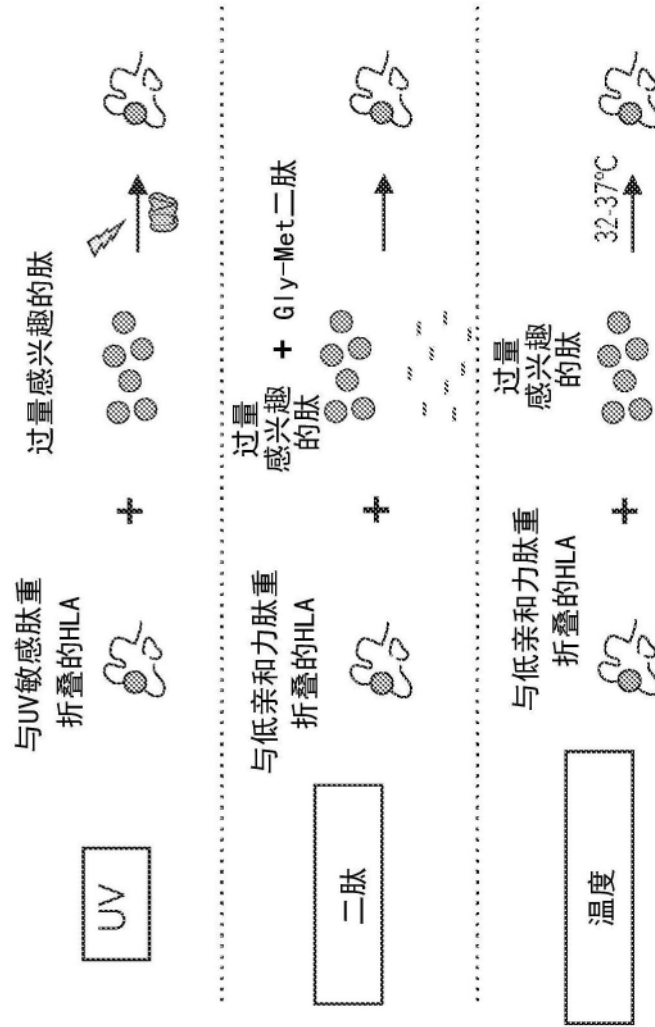


图2

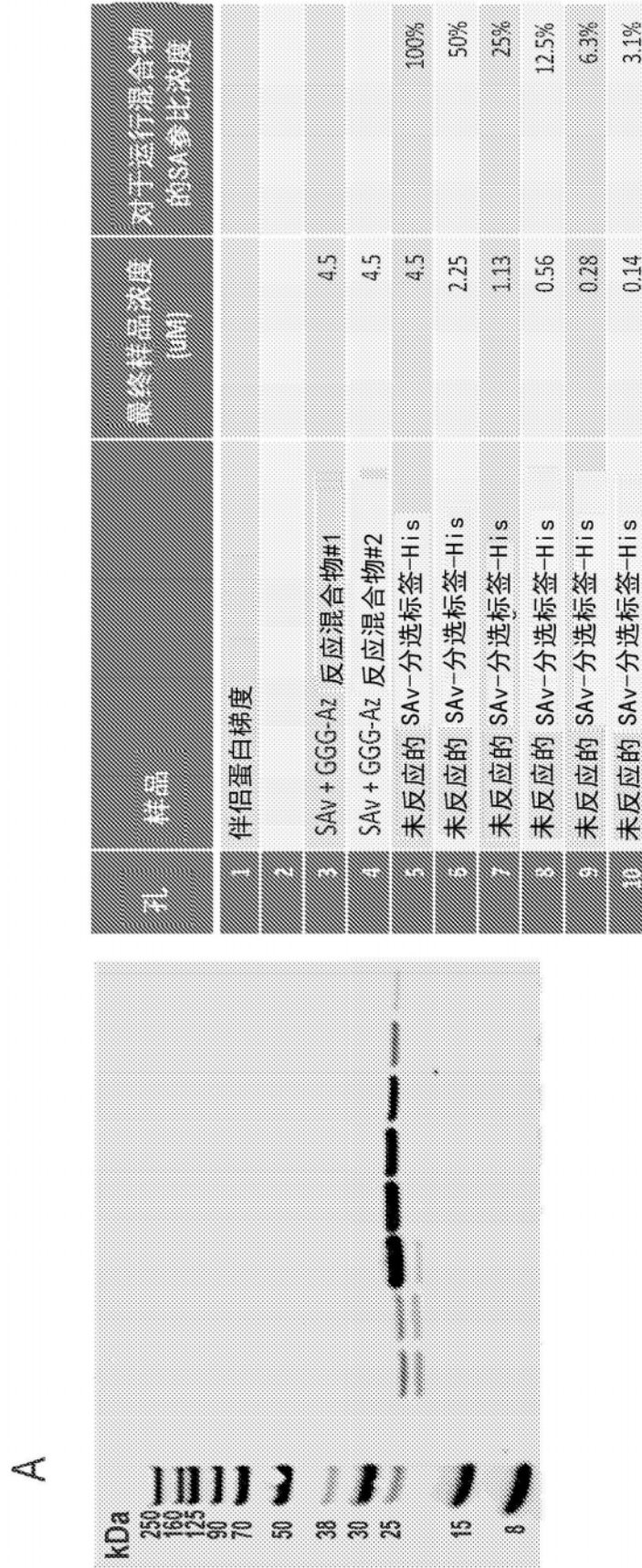


图3

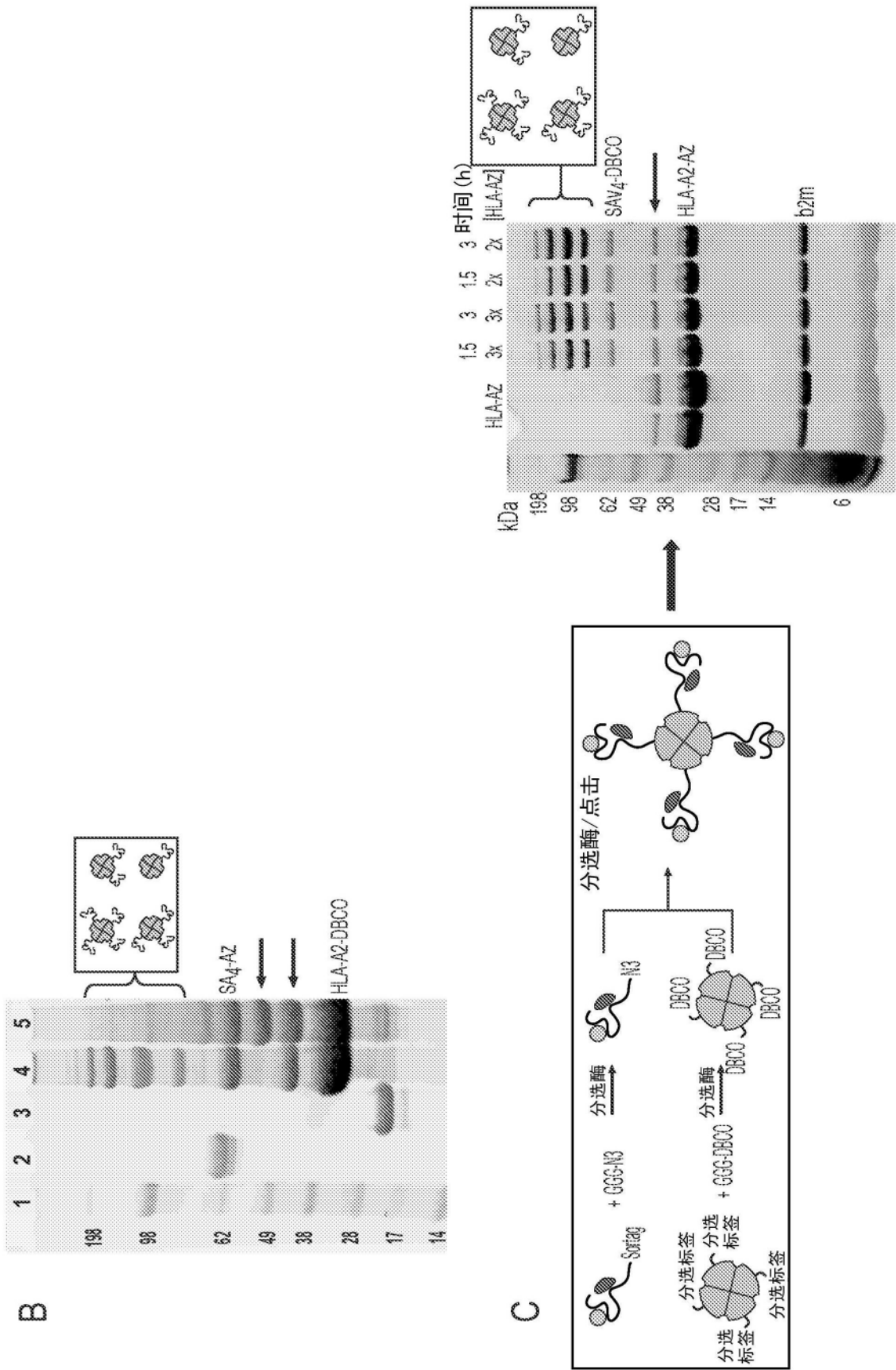


图3(续)



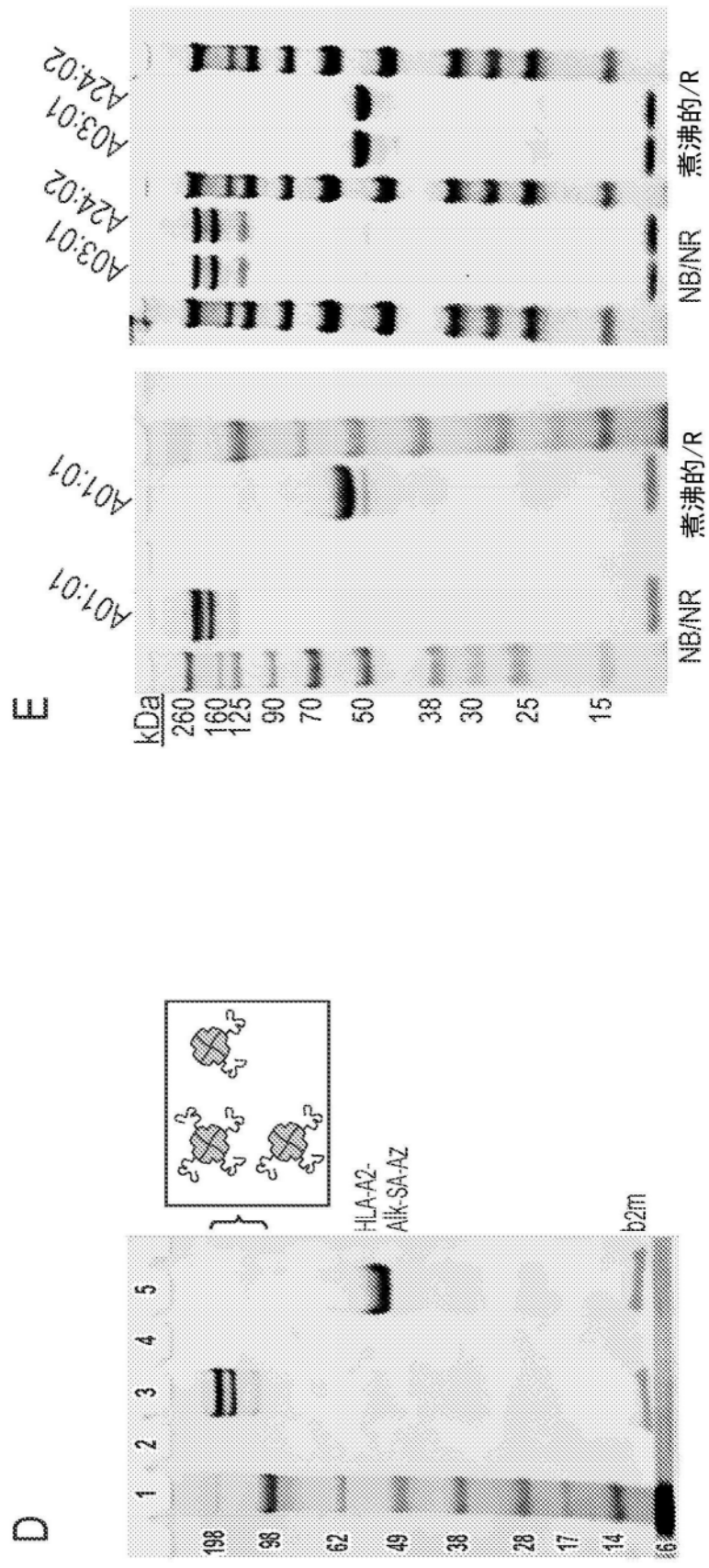


图3(续)

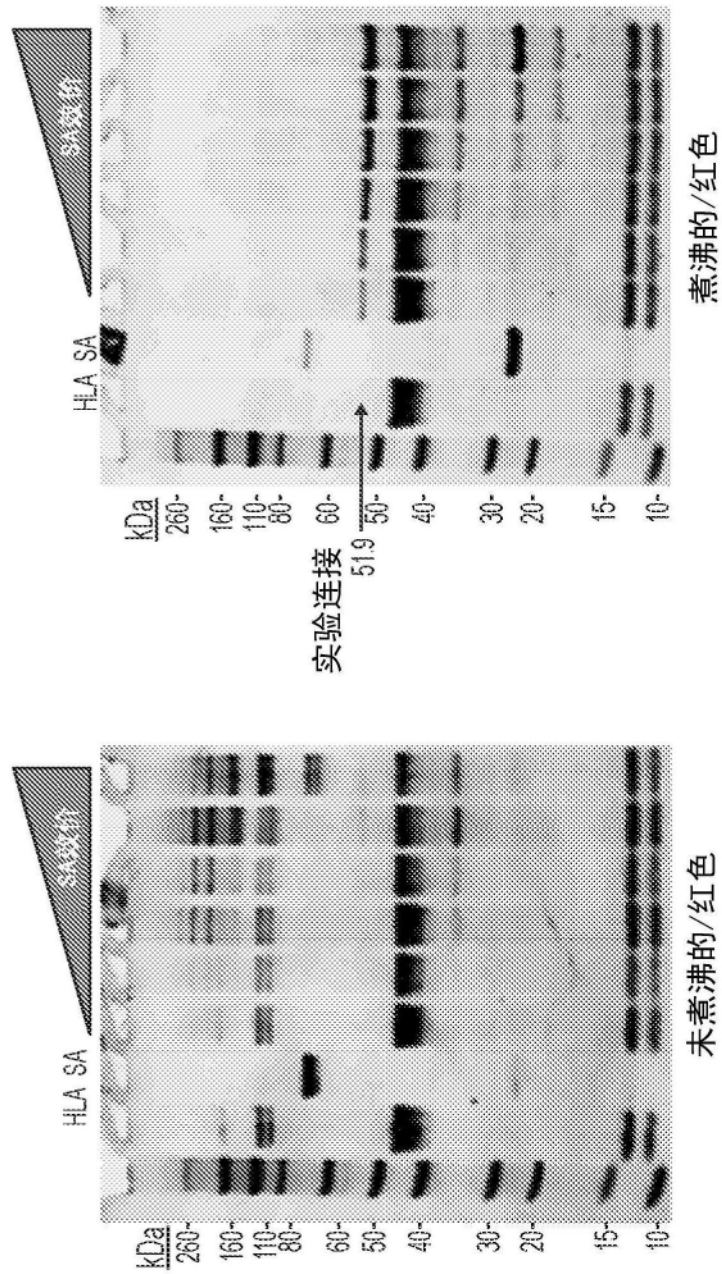


图4

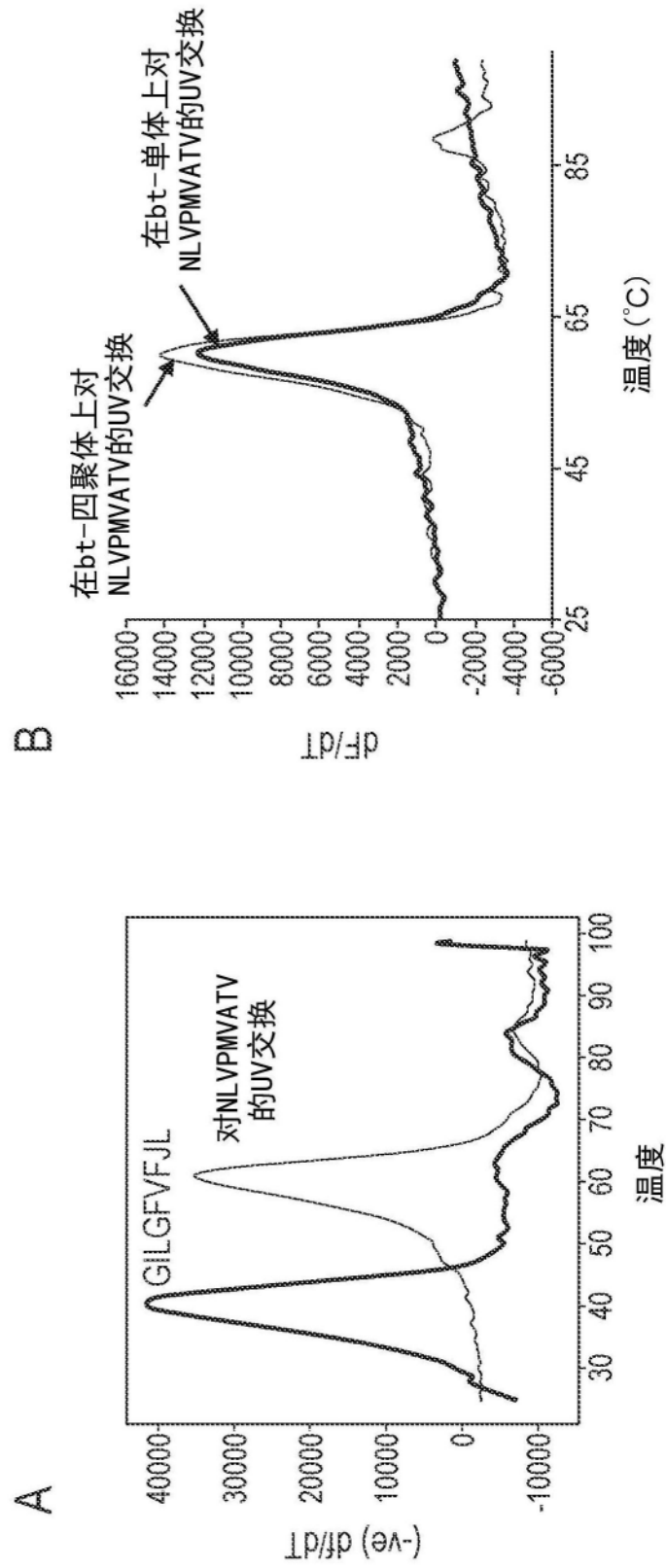


图5

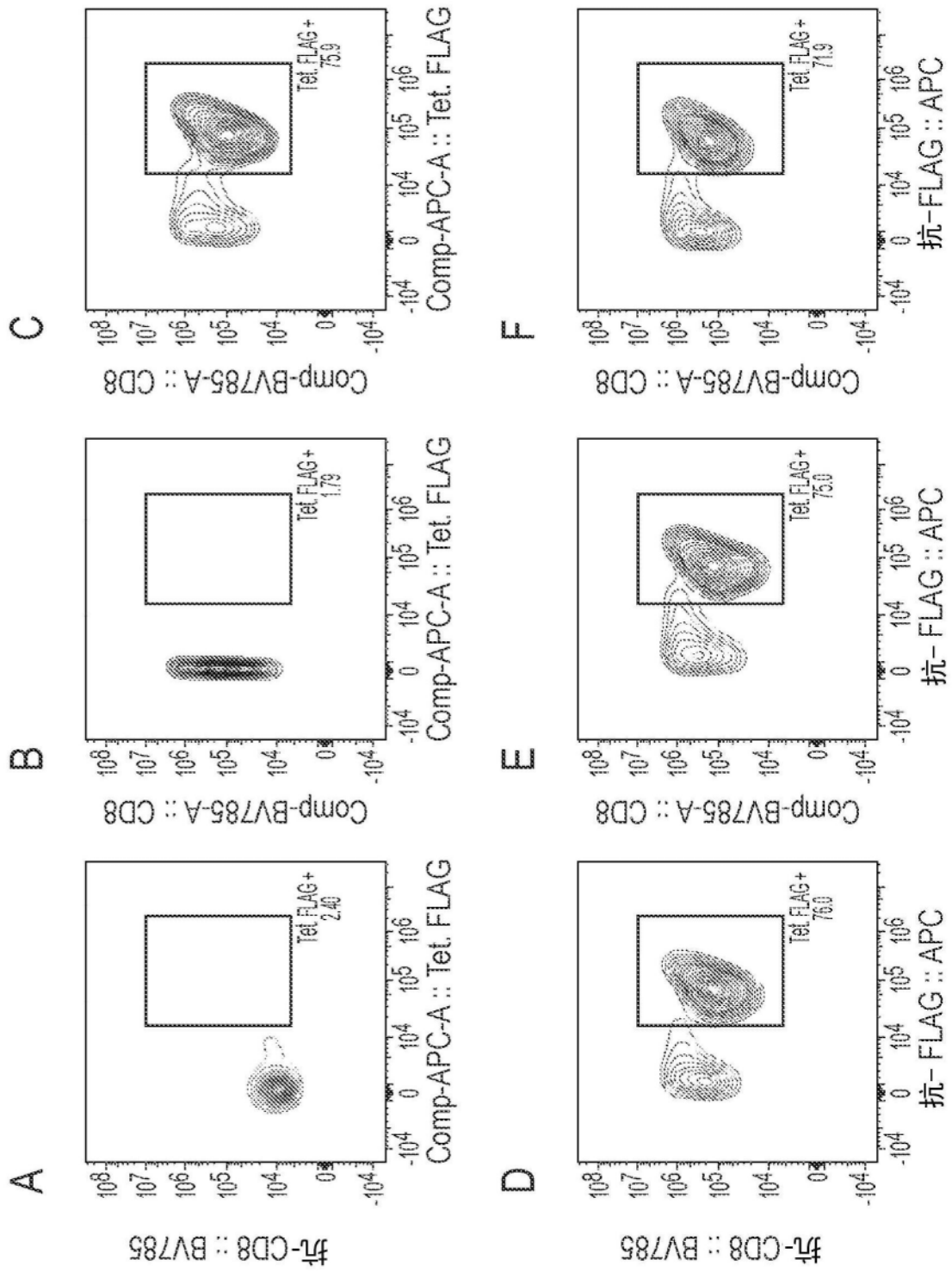


图6

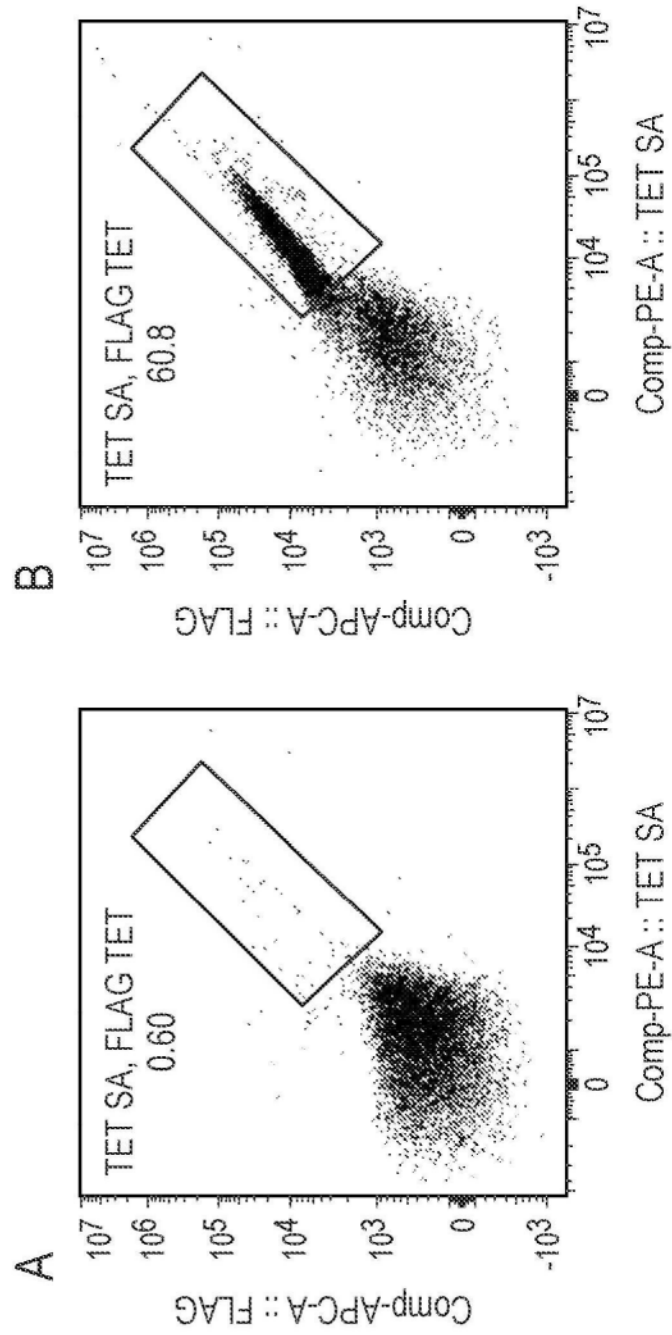


图7

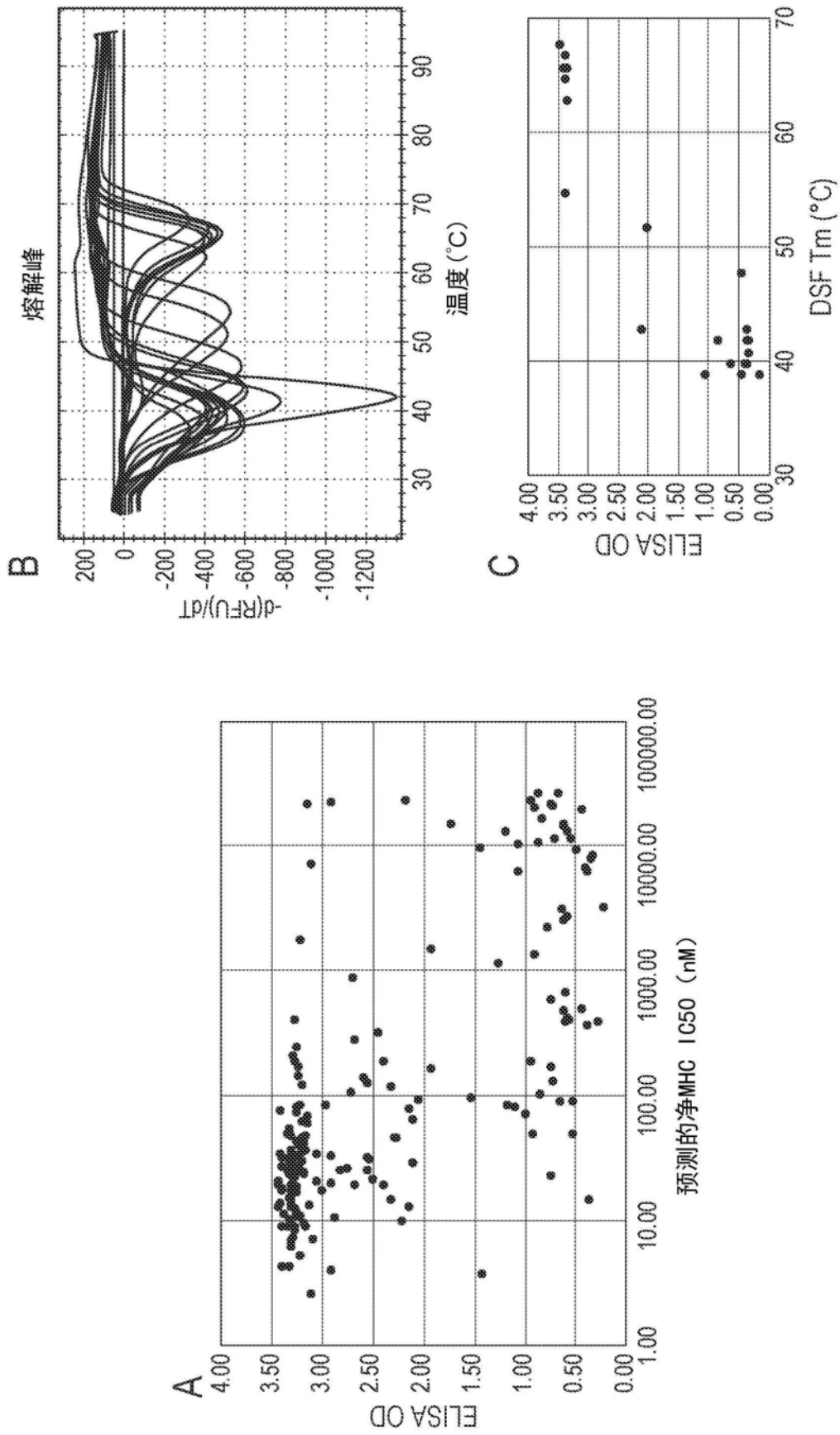


图8

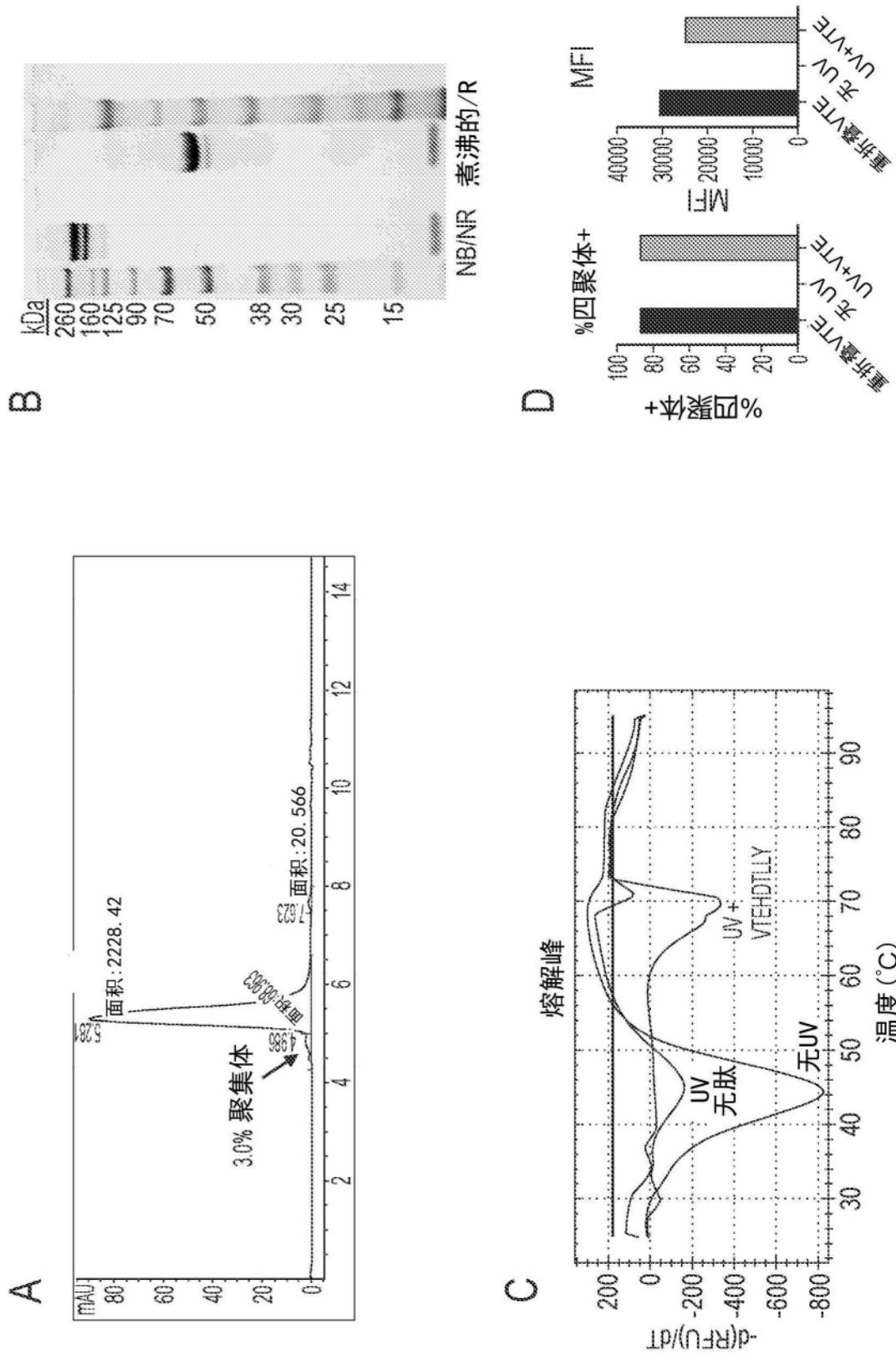


图9



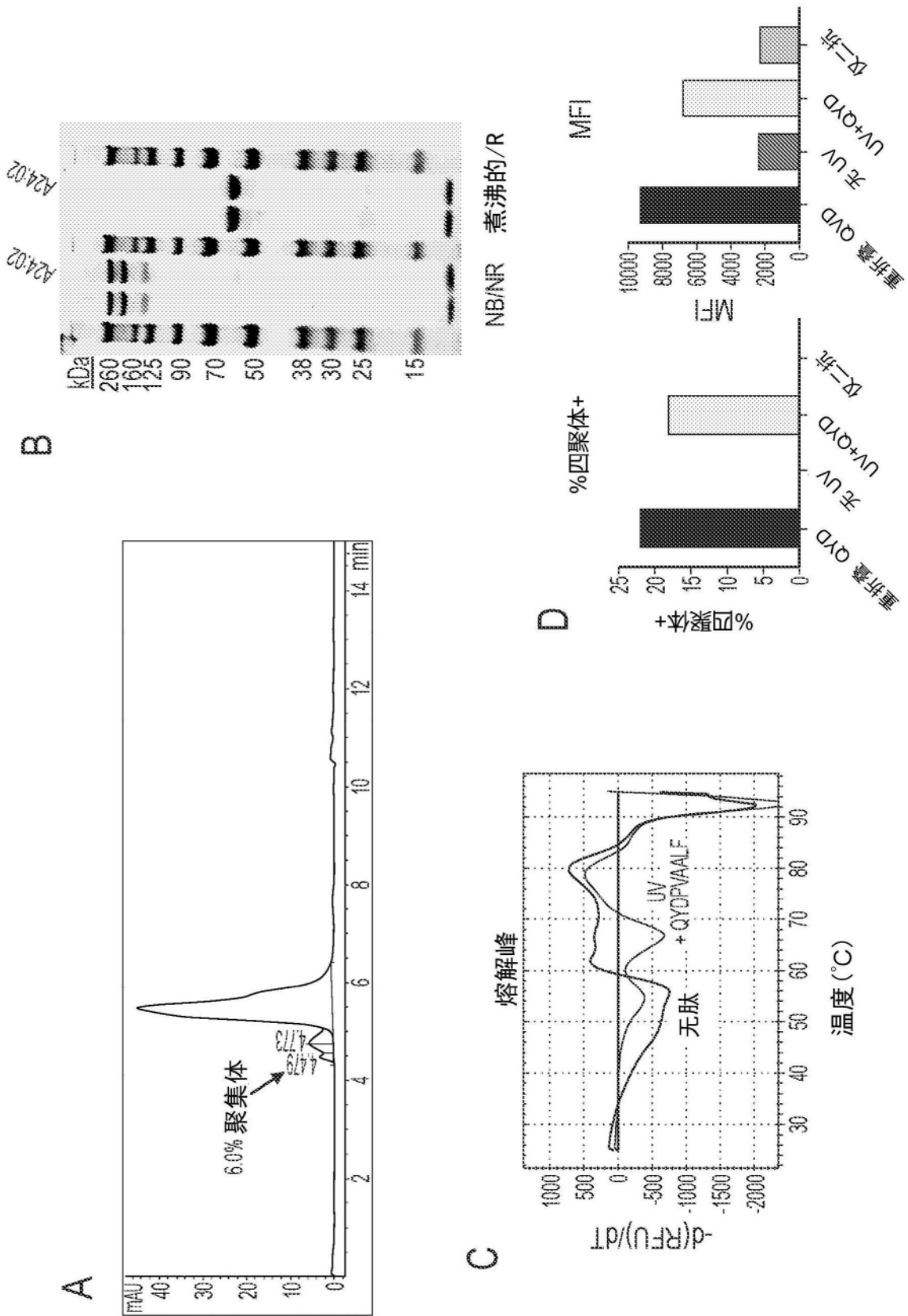


图10

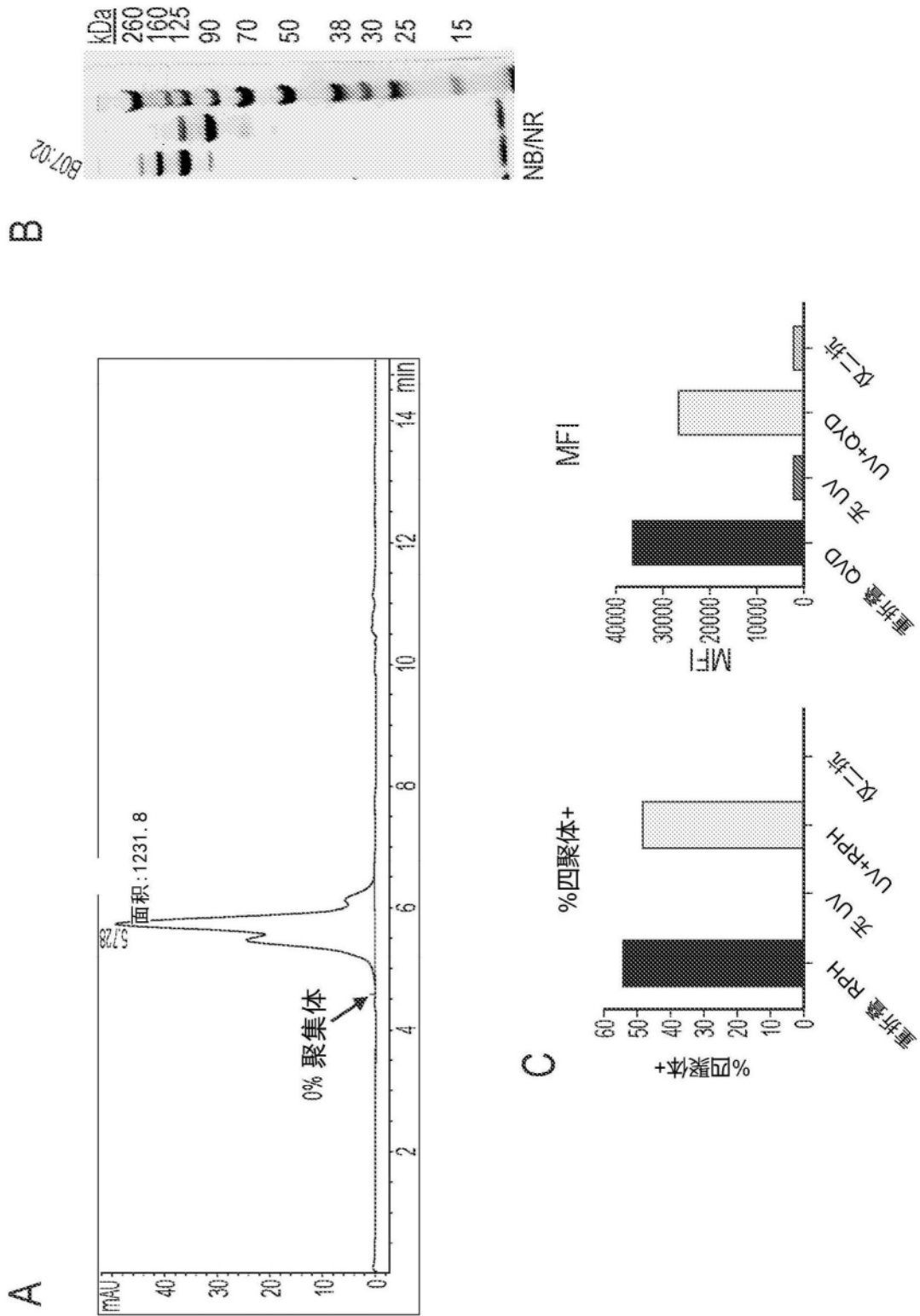


图11

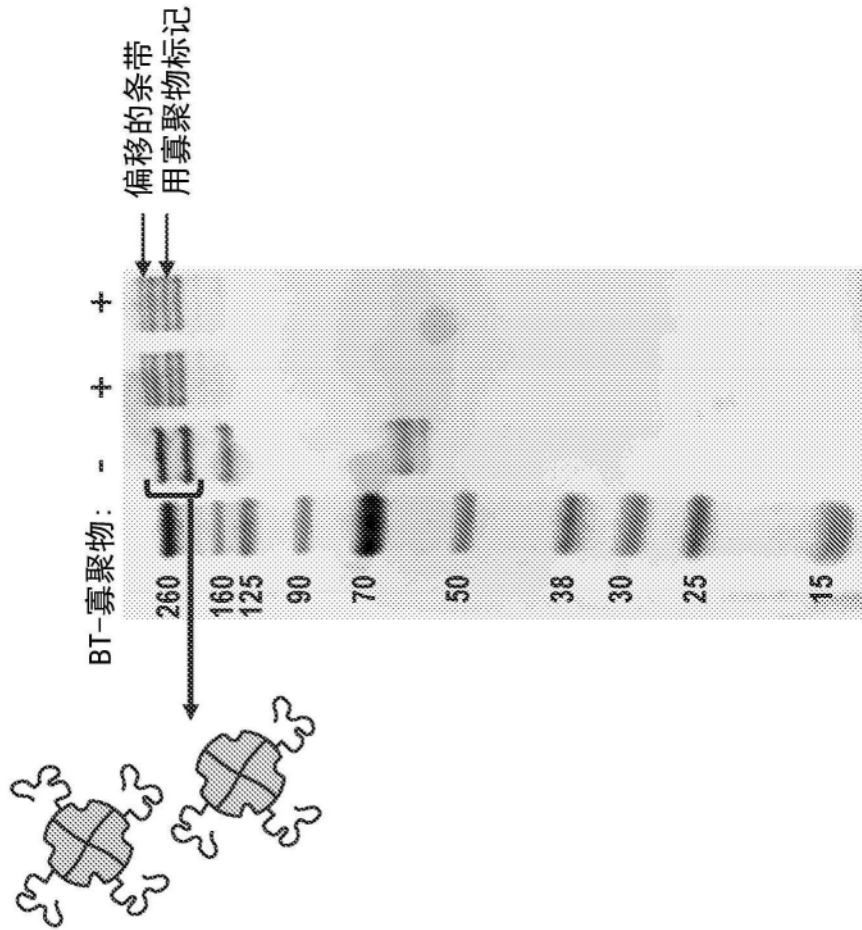


图12

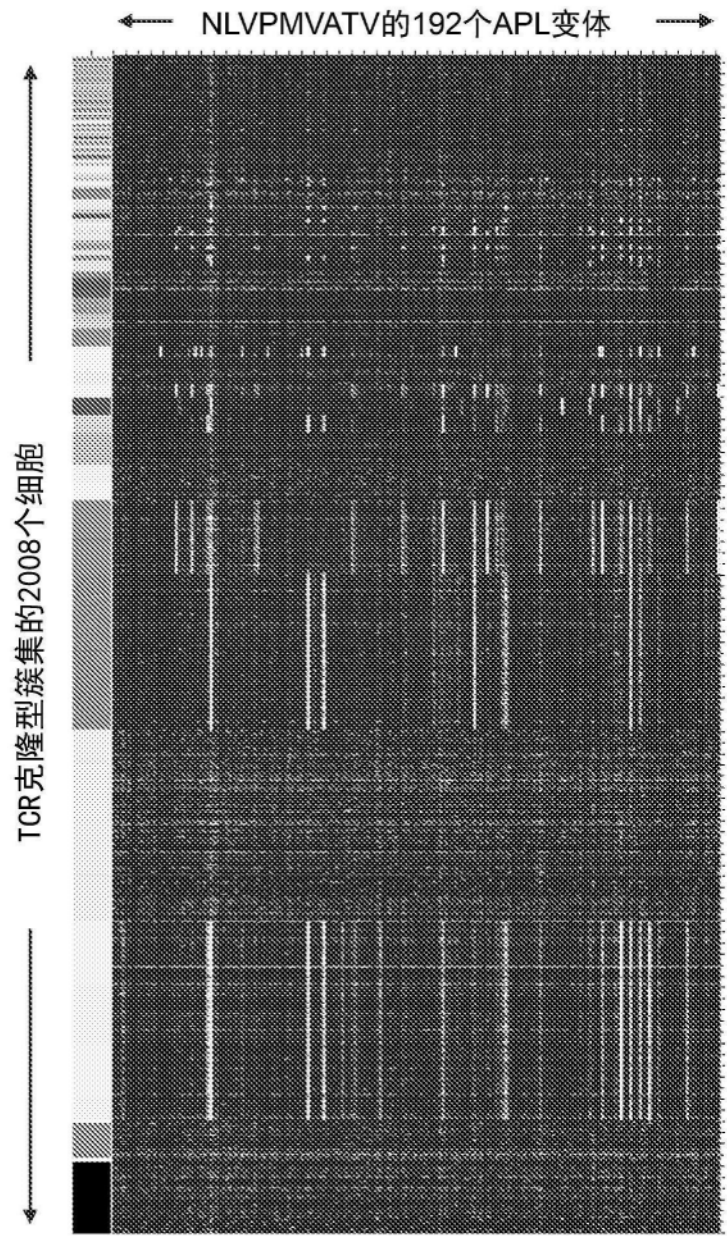


图13

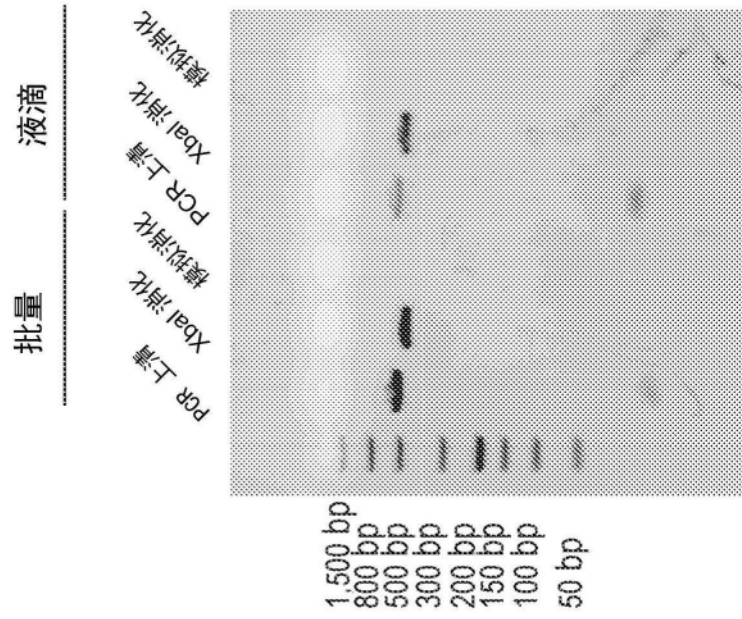


图14

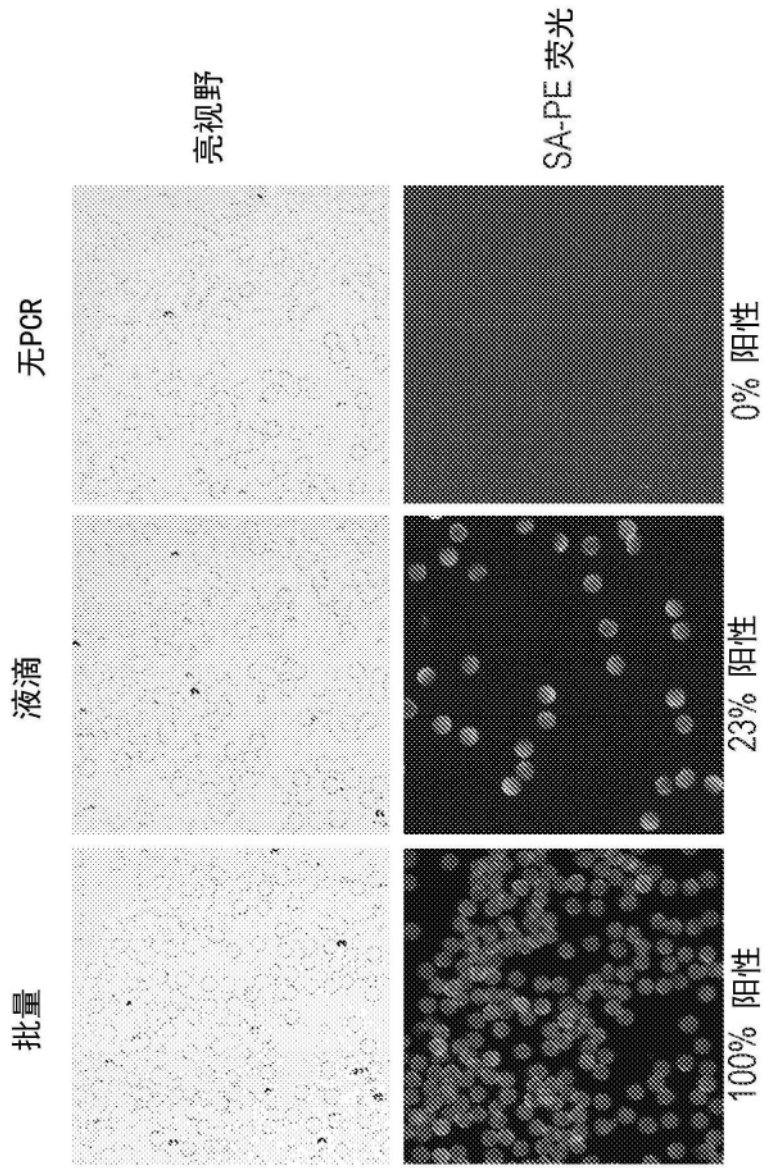


图15

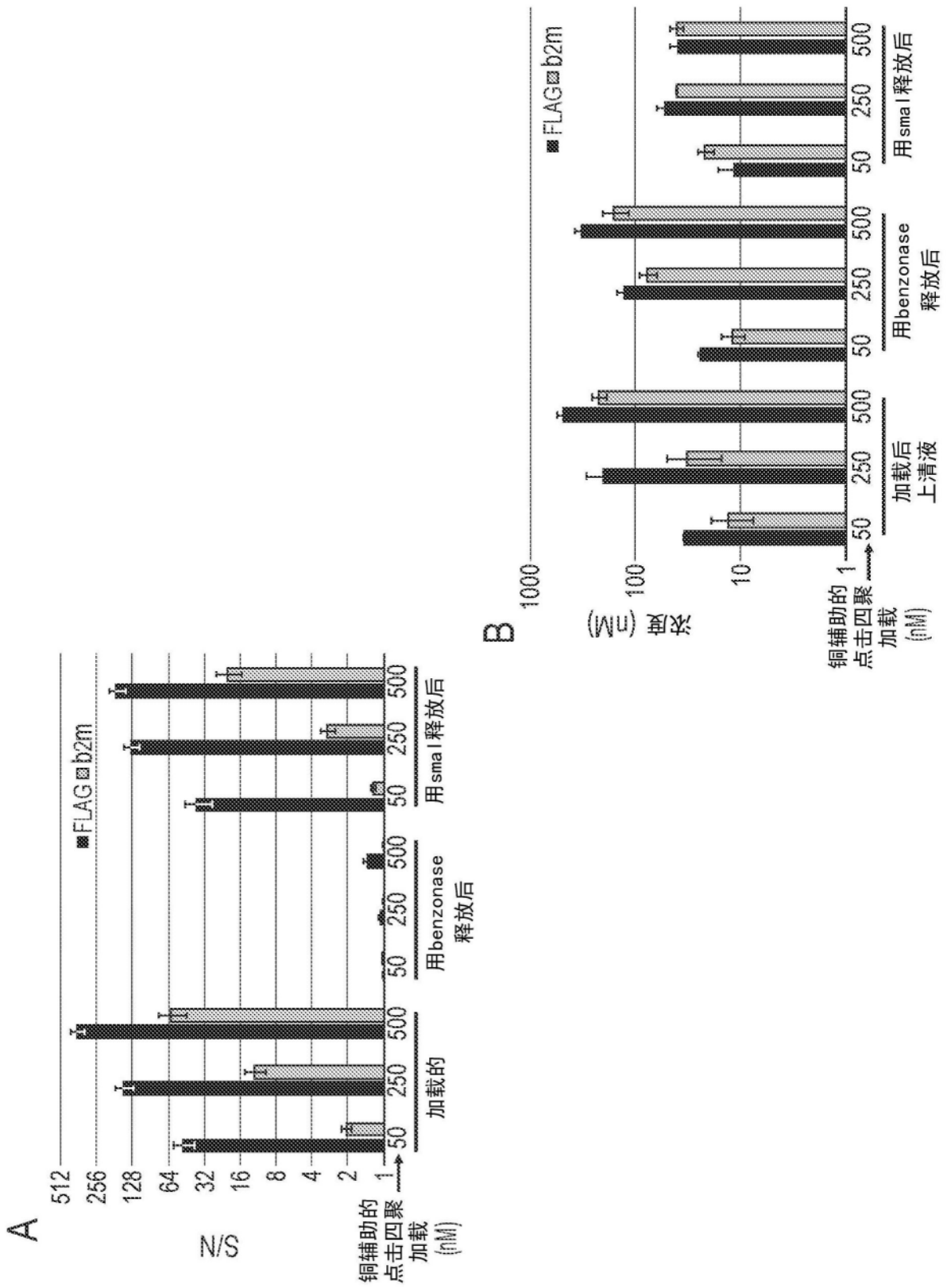


图16



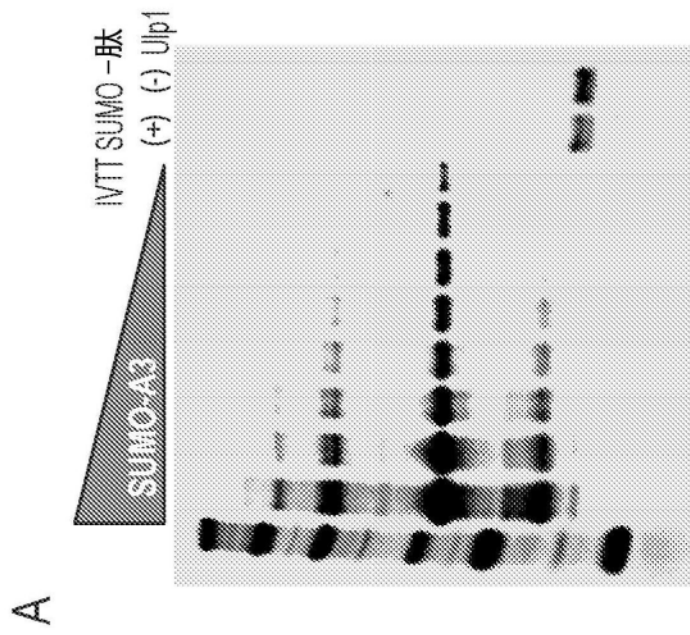
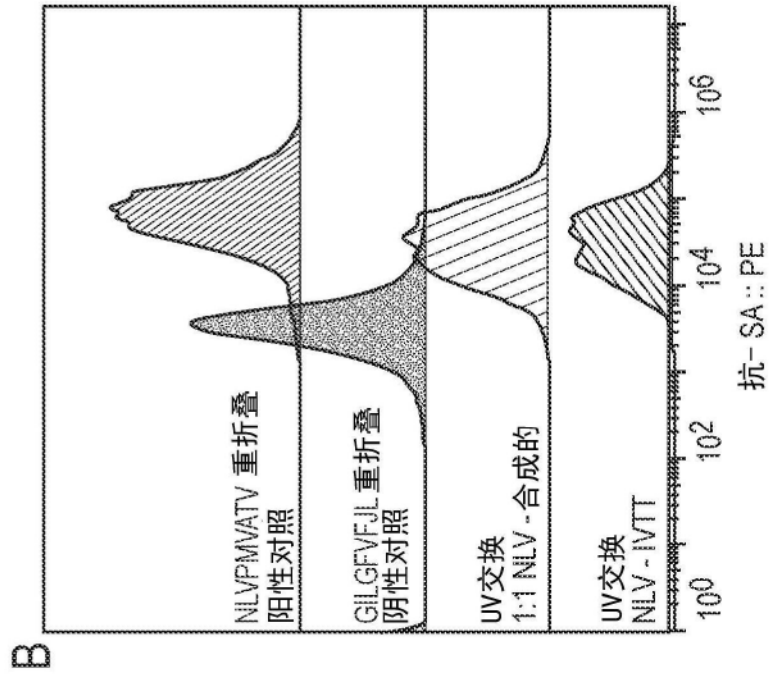


图17

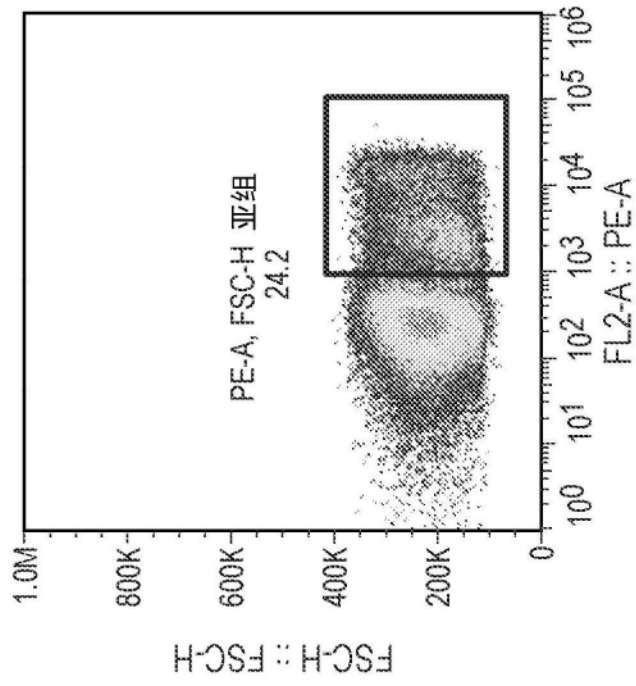


图18

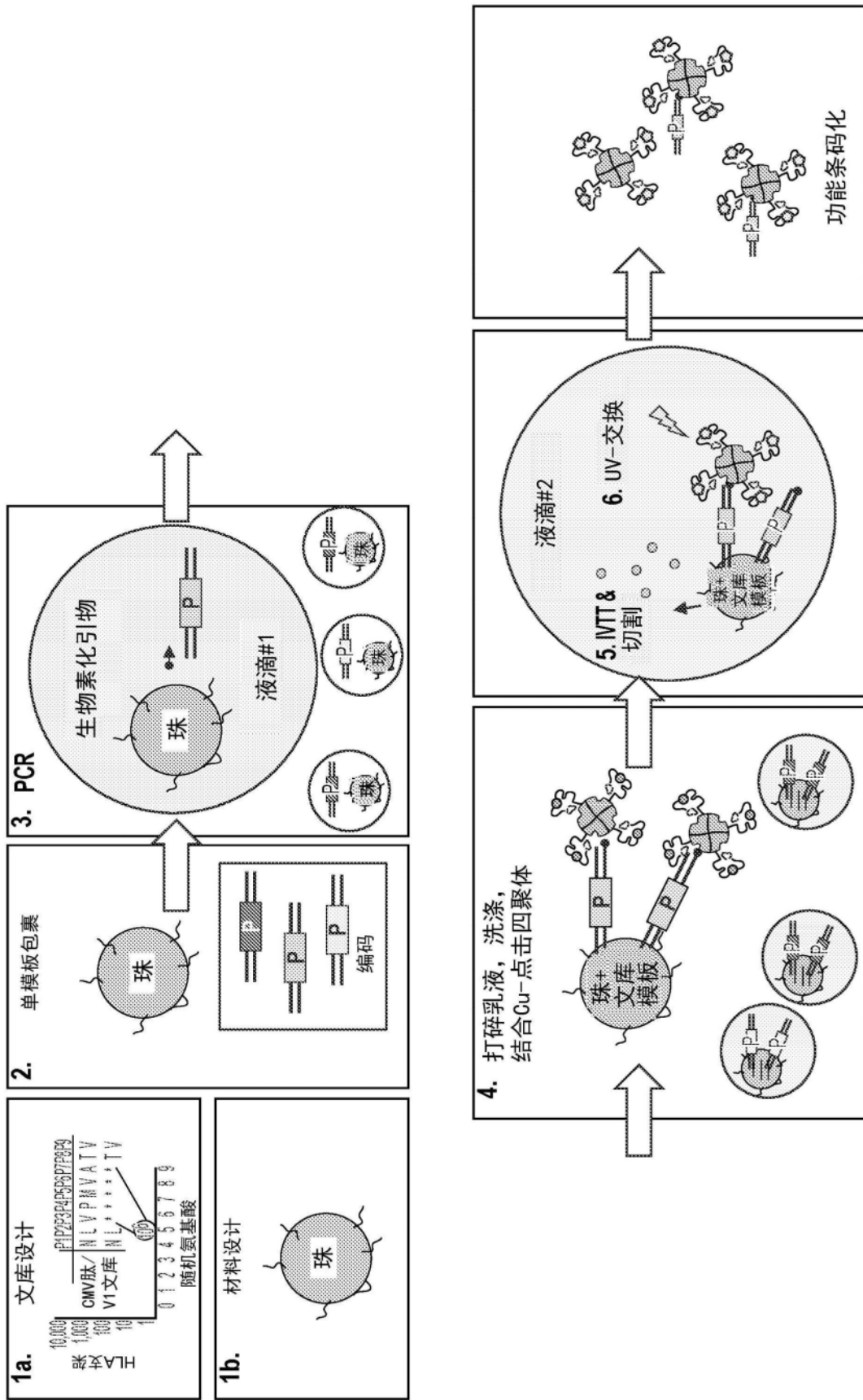


图19

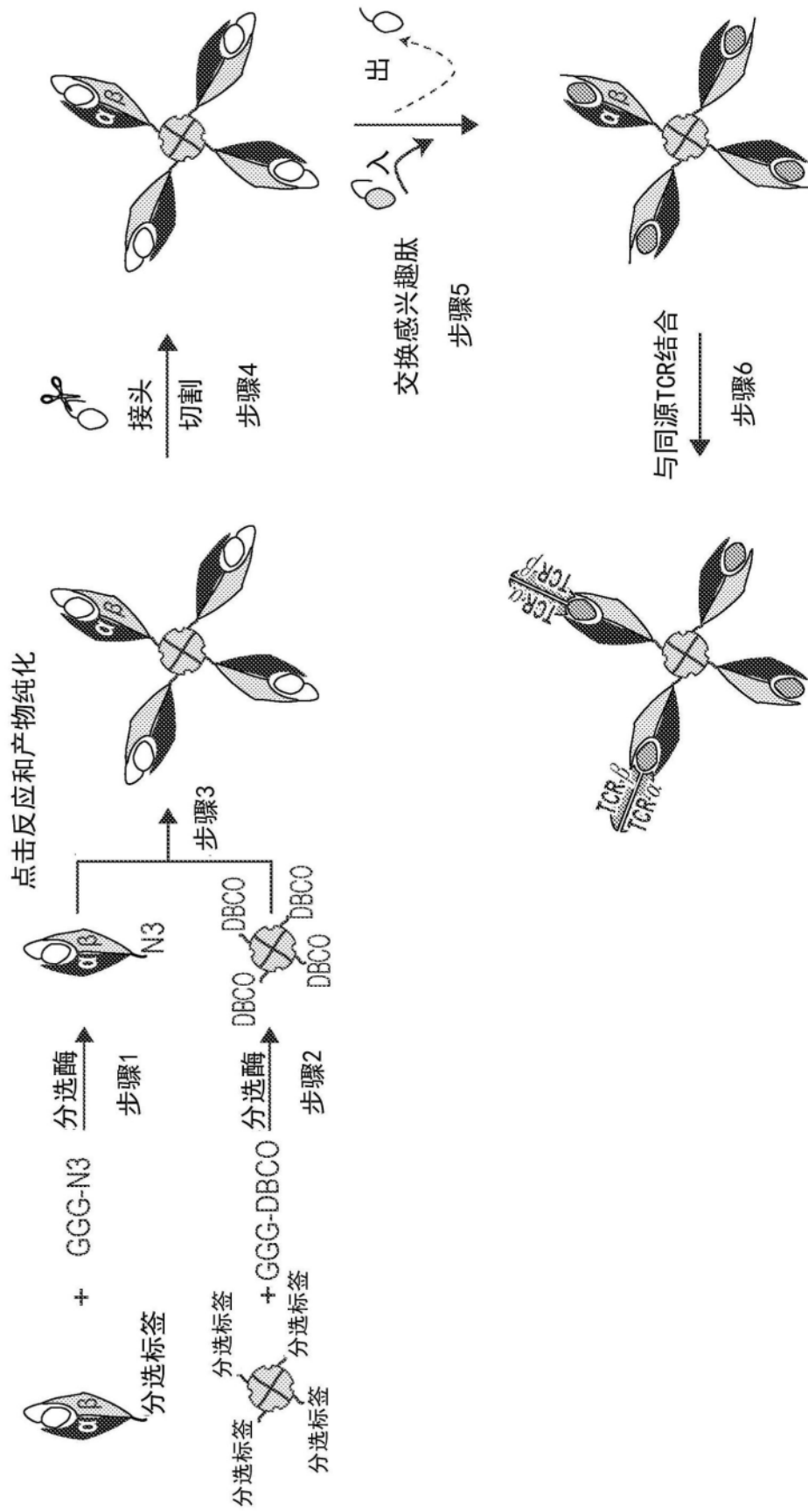


图20

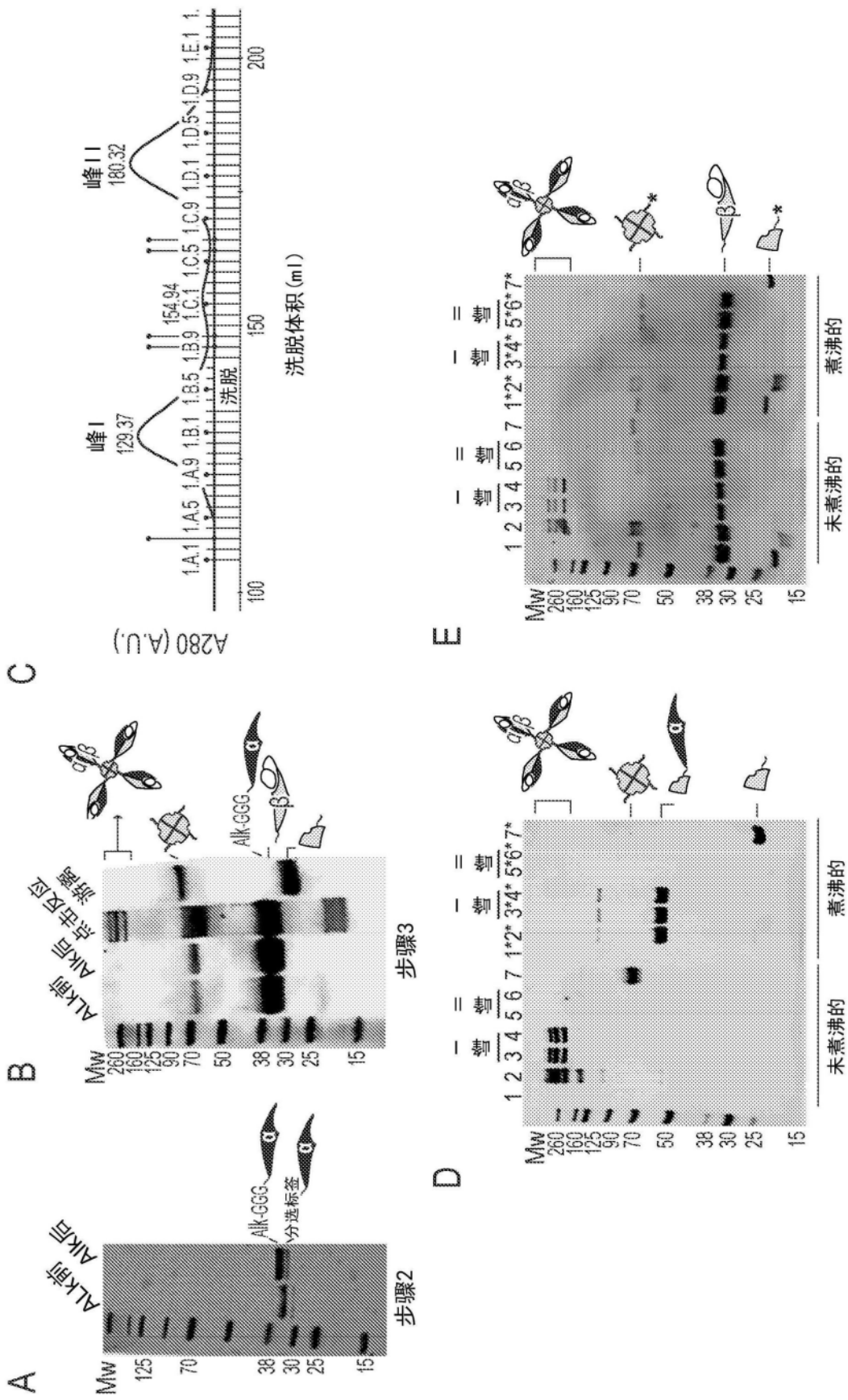


图21

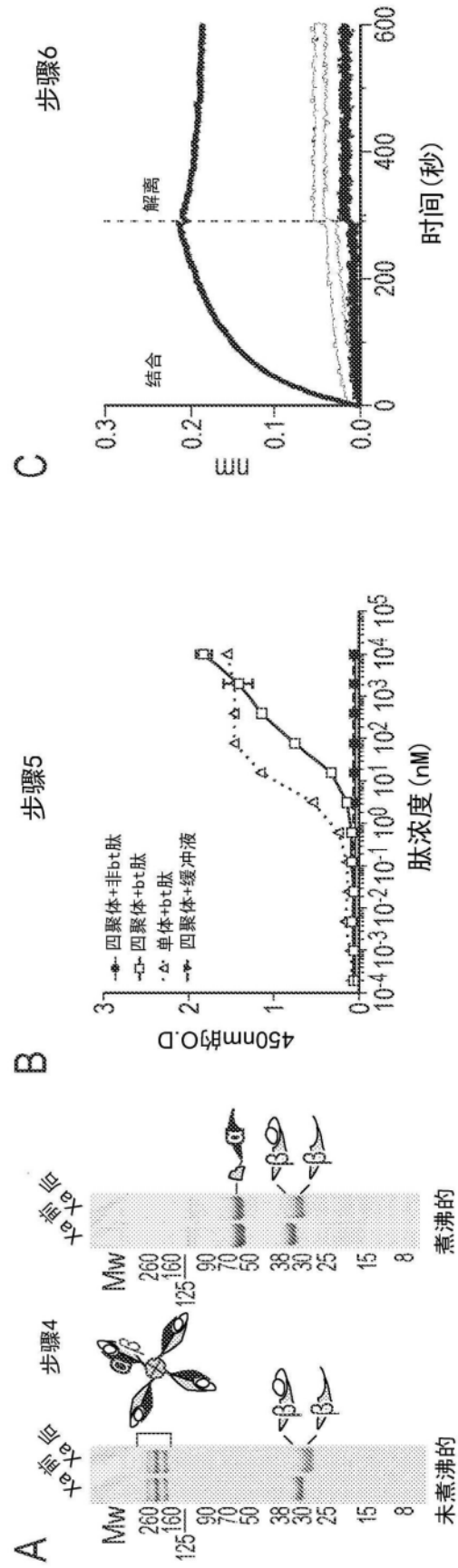


图22

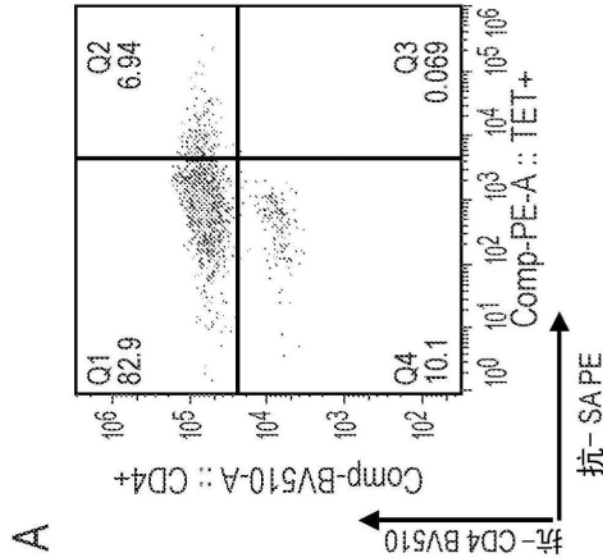
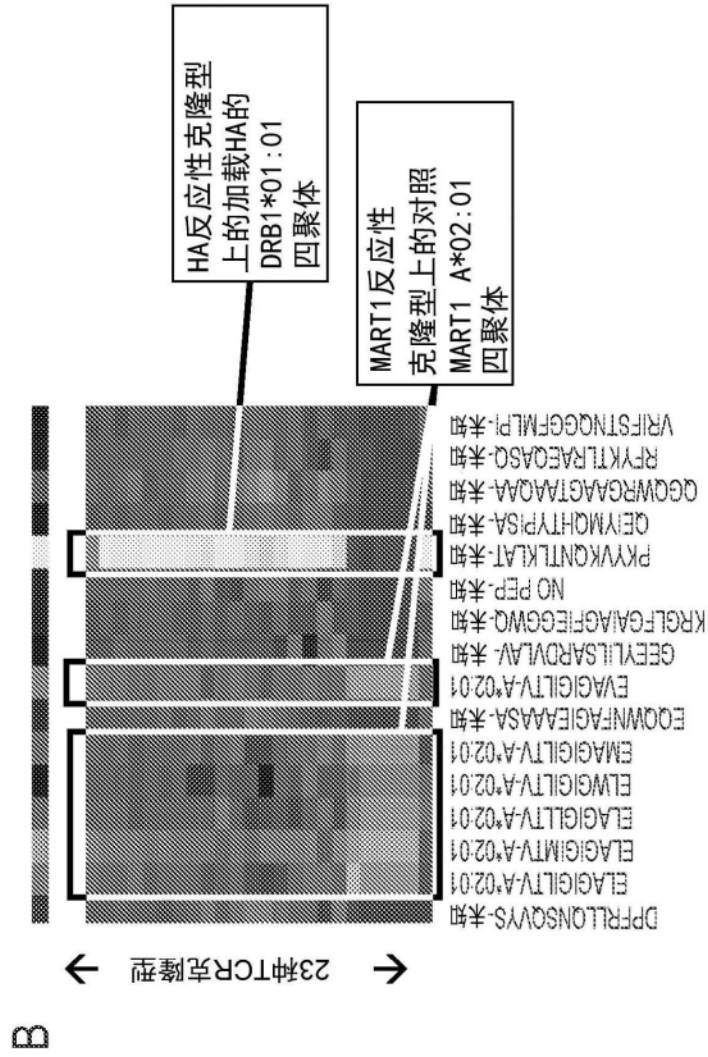


图23