(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115461068 A
(43)申请公布日 2022.12.09

(21)申请号 202180031498.6

(22)申请日 2021.02.25

(30)优先权数据
62/981,453 2020.02.25 US
63/002,249 2020.03.30 US
62/706,225 2020.08.05 US
63/091,291 2020.10.13 US

(85)PCT国际申请进入国家阶段日
2022.10.25

(86)PCT国际申请的申请数据
PCT/US2021/019739 2021.02.25

(87)PCT国际申请的公布数据
WO2021/173879 EN 2021.09.02

(71)申请人 利甘达尔股份有限公司
地址 美国加利福尼亚州

(72)发明人 安德烈·罗纳德·沃森
亚历山大·伊兹沃尔斯基

(74)专利代理机构 北京商专永信知识产权代理事务所(普通合伙) 11400
专利代理师 邬玥 方挺

(51)Int.Cl.
A61K 38/16(2006.01)
A61K 39/42(2006.01)
A61P 31/14(2006.01)
C07K 16/10(2006.01)
G01N 33/53(2006.01)

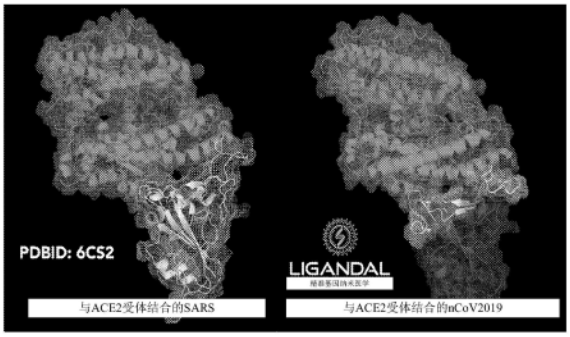
权利要求书3页 说明书57页
序列表81页 附图56页

(54)发明名称

仿生病毒肽的鉴定及其用途

(57)摘要

公开了衍生自SARS-CoV-2刺突蛋白和ACE2受体中每一个的结合界面的小肽,包含其的组合物,以及该肽和该组合物的预防和治疗用途。还公开了一种基于计算机模拟识别,设计,和修饰小肽的新型方案。



CN 115461068 A

1.一个支架,其包含来自SARS-CoV-2刺突蛋白和ACE2受体中每一个的结合界面的截短肽片段,其中,所述支架基本上保持天然SARS-CoV-2刺突蛋白或ACE2受体的结构,构象,或结合亲和力。

2.根据权利要求1所述的支架,其中,所述支架的大小为10至200个氨基酸残基,约50至约100个氨基酸残基,约55至约95个氨基酸残基,约60至约90个氨基酸残基,约65至约85个氨基酸残基,约70至约80个氨基酸残基。

3.根据权利要求1或权利要求2所述的支架,其中,所述支架的大小为小于约120个氨基酸残基,小于约110个氨基酸残基,小于约100个氨基酸残基,小于约90个氨基酸残基,小于约80个氨基酸残基,小于约70个氨基酸残基,小于约60个氨基酸残基,或小于50个氨基酸残基。

4.根据权利要求1-3中任一项所述的支架,其中,所述支架具有与SEQ ID NO:2的残基433-511的氨基酸序列或与SEQ ID NO:140的残基19-84的氨基酸序列至少40%,至少45%,至少50%,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%相同的氨基酸序列。

5.根据权利要求1-4中任一项所述的支架,其中,所述支架包含来自SARS-CoV-2刺突蛋白的结合界面的截短肽片段并保持β折叠结构,或包含来自ACE2的结合界面的截短肽片段并保持α-螺旋结构。

6.根据权利要求1-5中任一项所述的支架,其中,所述支架包含第一关键结合基序,第二关键结合基序,和所述关键结合基序之间的主链区域。

7.根据权利要求6所述的支架,其中,所述主链区域的全部或部分序列被接头替代。

8.根据权利要求7所述的支架,其中,所述接头是GS接头。

9.根据权利要求7或权利要求8所述的支架,其中,所述接头具有1至20个氨基酸残基的大小。

10.根据权利要求1-9中任一项所述的支架,其中,所述支架包含一个或多个修饰,包括插入,缺失,或取代,前提是所述一个或多个修饰不会显著降低所述支架对其结合配偶体的结合亲和力。

11.根据权利要求10所述的支架,其中,所述一个或多个修饰增加所述支架对其结合配偶体的结合亲和力。

12.根据权利要求10或权利要求11所述的支架,其中,所述支架包含一个或多个Cys取代,使得二硫键可以在所述支架中想要的位置形成。

13.根据权利要求1-12中任一项所述的支架,还包含一个或多个免疫表位。

14.根据权利要求13所述的支架,其中,所述免疫表位是T细胞表位或B细胞表位。

15.根据权利要求13或权利要求14所述的支架,其中,所述免疫表位选自SEQ ID NO:7-64和67-71。

16.根据权利要求1-15中任一项所述的支架,还包含一个或多个标签,或一个或多个可缀合结构域。

17.根据权利要求16所述的支架,其中,所述标签包括His标签和C-标签。

18.根据权利要求16所述的支架,其中,所述可缀合结构域包括马来酰亚胺-硫醇缀合。

19.根据权利要求16或权利要求18所述的支架,其中,所述支架经由所述可缀合结构域

连接至纳米颗粒,芯片,另一种底物,另一种肽,或另一种治疗剂。

20.根据权利要求1-19中任一项所述的支架,还包括在N-端的极性头部,在C-端的极性尾部,或两者。

21.根据权利要求20所述的支架,其中,所述极性头部或所述极性尾部包含聚(Arg),聚(Lys),聚(His),聚(Glu),或聚(Asp)。

22.根据权利要求20或权利要求21所述的支架,其中,所述极性头部或所述极性尾部包含2-20个带电荷的氨基酸。

23.根据权利要求1-22中任一项所述的支架,其中,所述支架是线性肽。

24.根据权利要求1-22中任一项所述的支架,其中,所述支架是头对尾环状肽。

25.一种多价支架,其包含两个或更多个权利要求1-24中任一项所述的支架。

26.一种融合蛋白,其包含一个或多个权利要求1-24中任一项所述的支架和免疫反应引发结构域。

27.根据权利要求26所述的融合蛋白,其中,所述免疫反应引发结构域是Fc结构域。

28.一种缀合物,其包含一个或多个权利要求1-24中任一项所述的支架,其缀合至另一种肽或另一种治疗剂。

29.一种组合物,其包含一个或多个权利要求1-24中任一项所述的支架,一个或多个权利要求25所述的多价支架,一种或多种权利要求26或权利要求27所述的融合蛋白,和一种或多种权利要求28所述的缀合物。

30.根据权利要求29所述的组合物,还包含一种或多种药学上可接受的载体,赋形剂,或稀释剂。

31.根据权利要求29或权利要求30所述的组合物,其中,所述组合物被配制成可注射,可吸入,口服,经鼻,外用,经皮,子宫,或直肠剂型。

32.根据权利要求29-31中任一项所述的组合物,其中,所述组合物通过肠胃外,口服,肺,颊,经鼻,经皮,直肠,或眼途径施用于受试者。

33.根据权利要求29-32中任一项所述的组合物,其中,所述组合物是疫苗组合物。

34.一种治疗或预防受试者SAR-CoV-2感染的方法,包括向受试者施用治疗有效量的一个或多个权利要求1-24中任一项所述的支架,一个或多个权利要求25所述的多价支架,一种或多种权利要求26或权利要求27所述的融合蛋白,一种或多种权利要求28所述的缀合物,或一种或多种权利要求29-33中任一项所述的组合物。

35.根据权利要求34所述的方法,其中,所述受试者是哺乳动物。

36.根据权利要求34或权利要求35所述的方法,其中,所述受试者是人。

37.一种阻断SAR-CoV-2病毒进入受试者的方法,包括向受试者施用治疗有效量的的一个或多个权利要求1-24中任一项所述的支架,一个或多个权利要求25所述的多价支架,一种或多种权利要求26或权利要求27所述的融合蛋白,一种或多种权利要求28所述的缀合物,或一种或多种权利要求29-33中任一项所述的组合物。

38.根据权利要求37所述的方法,其中,所述受试者是哺乳动物。

36.根据权利要求37或权利要求38所述的方法,其中,所述受试者是人。

37.一种靶向递送一种或多种治疗剂的方法,包括将所述一种或多种治疗剂缀合至一个或多个权利要求1-24中任一项所述的支架,并将所述缀合物递送至有需要的受试者。

38.一种获得仿拟衍生支架的天然蛋白质的结合的支架的方法,包括:

产生第一结合配偶体和第二结合配偶体的三维结合模型,

基于该结合模型确定每个结合配偶体上的结合界面,

分析该结合界面以将每个结合配偶体的结构和/或构象保持在其天然,游离,或结合状态,

基于热力学计算(ΔG)确定关键结合残基,和

确定每个结合配偶体的结合界面的氨基酸序列以获得支架。

39.根据权利要求38所述的方法,其中,所述三维结合由计算机程序产生。

40.根据权利要求39所述的方法,其中,所述计算机程序是SWISS-MODEL。

41.根据权利要求38-40中任一项所述的方法,其中,所述三维结合基于所述第一结合配偶体或所述第二结合配偶体与已知序列和/或结构的蛋白质的同源性。

42.根据权利要求38-41中任一项所述的方法,还包括设计多种构象或折叠状态的支架以匹配相应的结合配偶体。

43.根据权利要求38-42中任一项所述的方法,其中,所述第一结合配偶体和所述第二结合配偶体分别是SARS-CoV-2刺突蛋白和ACE2。

仿生病毒肽的鉴定及其用途

[0001] 相关申请交叉引用

[0002] 本申请要求于2020年2月25日提交的美国临时专利申请号62/981,453,于2020年3月30日提交的美国临时专利申请号63/002,249,于2020年8月5日提交的美国临时专利申请号62/706,225,于2020年10月13日提交的美国临时专利申请号63/091,291的优先权,其内容均通过引用整体并入本文。

[0003] 序列表

[0004] 本申请包含经由EFS-Web以ASCII格式提交并通过引用整体并入本文的序列表。ASCII副本创建于2021年2月25日,名为2021-02-25Ligandal 8009.W000序列表,大小为160KB。

背景技术

[0005] 导致COVID-19的SARS-CoV-2是一种全球性大流行病。SARS-CoV-2和其他冠状病毒,包括MERS和SARS,会导致人类严重的呼吸疾病,并且被认为是在蝙蝠和啮齿动物中传播的病毒的共同起源。一些以动物为宿主的冠状病毒已获得突变,从而将其宿主范围扩大至包括人类。截至2020年3月,SARS-CoV-2已在整个人类物种中发生变异和扩展;共鉴定出214个单倍型(即序列变异)和344个不同的株。通过突变,重组,和自然选择获得的这些变异中的大多数已在刺突(Spike)(S)蛋白中发现。这种变异可能导致更具感染性和强毒性的株。探索与病毒蛋白相关联的序列空间是一个难题,对进化生物学和疾病预测具有至关重要的意义。虽然过去的几项研究试图解决病毒进化的问题,但很少有能够获得与为SARS-CoV-2编制的数据集相似的数据集或目前可供研究人员使用的来自数据科学,数学,和生物物理学的丰富的新型分析工具集。

[0006] SARS-CoV-2感染对康复个体的长期健康后果仍有待观察,但它们包括一系列后遗症,从神经学到血液学,血管性,免疫学,炎性,肾性,呼吸性,甚至可能是自身免疫性。当考虑到SARS-CoV-1的已知神经精神影响时,这些长期影响尤其令人担忧,233名SARS幸存者中有27.1%在康复4年后表现出符合慢性疲劳综合征诊断标准的症状。此外,40.3%的人报告有慢性疲劳问题,40%的人表现出精神疾病。当前的预防方法包括,例如,mRNA疫苗方法和包含病毒样颗粒,重组刺突蛋白片段等的重组疫苗方法。这些疫苗方法通常成本高昂,开发缓慢,并且需要减毒活,重组,或基于mRNA的方法,这些方法需要进行广泛的再设计以接近新型抗原。虽然在实验室工作台规模制造mRNA的成本超过$1000/mg,但本文公开的肽方法是一种更具成本效益的替代方法,在实验室工作台规模下约为$5/mg。

[0007] 快速和全球可扩展的疫苗开发对于保护世界免受SARS-CoV-2以及未来致命疾病爆发和大流行病至关重要。因此,迫切需要更好地了解SARS-CoV-2或任何其他新病毒等中S蛋白基因组序列的潜在变异,并在普通人群中开发一种负担得起,可全球部署,室温稳定,且可重复施用,并发症风险低的疗法,。

发明内容

[0008] 在一个方面，本文公开的是包含来自SARS-CoV-2刺突(S)蛋白的结合结构域或ACE2受体的截短肽片段的支架，其中该支架基本上保持了天然蛋白的结构，构象，和/或结合亲和力。在某些实施例中，该支架具有40至200个氨基酸残基之间的大小。在某些实施例中，该支架包含来自CoV-2刺突蛋白结合界面的两个关键结合基序。在某些实施例中，该支架包含来自ACE2结合界面的两个关键结合基序。在某些实施例中，该两个关键结合基序通过接头如GS接头连接。在某些实施例中，该接头具有1至20个氨基酸残基的大小。在某些实施例中，该支架包含一个或多个修饰，包括插入，缺失，和/或取代。在某些实施例中，该支架还包含一个或多个免疫表位，一个或多个标签，一个或多个可缀合结构域，和/或极性头部或尾部。在某些实施例中，该一个或多个支架经由一个或多个接头连接以形成多价支架。在某些实施例中，一个或多个支架附接至免疫反应引发结构域，如Fc结构域(例如，人Fc结构域或人源化Fc结构域)以形成融合蛋白。在某些实施例中，一个或多个支架附接至底物，如纳米颗粒或芯片。在某些实施例中，一个或多个支架与另一种肽或治疗剂缀合。

[0009] 在另一个方面，本文公开的是包含本文公开的一个或多个支架，一个或多个缀合物，或一个或多个融合蛋白的组合物。在某些实施例中，该组合物还包含一种或多种药学上可接受的载体，赋形剂，或稀释剂。在某些实施例中，将该组合物配制成可注射，可吸入，口服，经鼻，外用，经皮，子宫，或直肠剂型。在某些实施例中，通过肠胃外，口服，肺，颊，经鼻，经皮，直肠，或眼途径向受试者施用该组合物。在某些实施例中，该组合物是疫苗组合物。

[0010] 在另一个方面，本文公开的是一种治疗或预防受试者SAR-CoV-2感染的方法，包括向该受试者施用治疗有效量的一个或多个支架，一个或多个缀合物，一个或多个融合蛋白，或包含本文公开的一个或多个支架，一个或多个缀合物，或一个或多个融合蛋白的组合物。在某些实施例中，该受试者是哺乳动物。在某些实施例中，该受试者是人。

[0011] 在另一个方面，本文公开的是一种阻断SAR-CoV-2病毒进入受试者的方法，包括向该受试者施用治疗有效量的一个或多个支架，一个或多个缀合物，一个或多个融合蛋白，或包含本文公开的一个或多个支架，一个或多个缀合物，或一个或多个融合蛋白的组合物。在某些实施例中，该受试者是哺乳动物。在某些实施例中，该受试者是人。

[0012] 在另一个方面，本文公开的是一种靶向递送一种或多种治疗剂的方法，包括将该一种或多种治疗剂缀合至本文公开的一个或多个支架，并将该缀合物递送至有需要的受试者。

[0013] 在另一个方面，本文公开的是一种获得仿拟衍生支架的天然蛋白的结合的支架的方法。该方法包括以下步骤：产生第一结合配偶体和第二结合配偶体的三维结合模型，基于该结合模型确定每个结合配偶体上的结合界面，分析该结合界面以将每个结合配偶体的结构和/或构象保持在其天然，游离，或结合状态，基于热力学计算(ΔG)确定关键结合残基，和确定每个结合配偶体的结合界面的氨基酸序列以获得支架。在某些实施例中，该三维结合由计算机程序如SWISS-MODEL产生。在某些实施例中，该三维结合基于该第一结合配偶体或该第二结合配偶体与已知序列和/或结构的蛋白的同源性。在某些实施例中，该方法还需要设计多种构象或折叠状态的支架以匹配相应的结合配偶体。

[0014] 附图简要说明

[0015] 本申请包含至少一张彩色绘制的附图。具有一张或多张彩色附图的本申请的副本

将在请求并支付必要费用后由专利商标局(the Office)提供。

[0016] 图1显示与ACE2结合的SARS-CoV-1(PDBID 6CS2)的晶体结构(左)和与ACE2结合的SARS-CoV-2的模拟结构(右)的比较。对结合(-ΔG)有积极贡献的氨基酸残基以绿色显示,具有约0ΔG的氨基酸残基以黄色显示,排斥性(repulsory)氨基酸残基(+ΔG)以粉红色(左)或橙色(右)显示。

[0017] 图2A显示两个先前公布的SARS-CoV-1免疫表位的3D结构。图2B-2D显示基于与SARS-CoV-1的同源性所推导的CoV-2免疫表位的3D结构和位置。

[0018] 图3显示免疫表位的MHC-I结合预测结果。“免疫表位1”＝SEQ ID NO:67;“免疫表位2”＝SEQ ID NO:69;KMSECVLGQSKRV＝SEQ ID NO:71;LLFNKVTLA＝SEQ ID NO:7;SFIEDLLFNKV＝SEQ ID NO:68。

[0019] 图4显示他人鉴定的CoV-2S蛋白中CoV-2S蛋白抗体表位的位置(图示为SEQ ID NO:2的残基15-1137)。野生型蛋白中的CoV-2支架加双下划线。表位以粗体显示,而具有高抗原性评分的表位以粗体和下划线显示。

[0020] 图5显示与ACE2比对的截短CoV-2S蛋白以及抗体表位(洋红色)和ACE2结合残基(绿色)的位置。

[0021] 图6描绘结合构象中三个代表性接头的三维分子建模。主链被描绘为蓝色线圈。在PyMol中对侧链原子使用命令颜色>按链>链弓和颜色>按>元素>HNOS进行颜色编码,其中H＝白色,N＝蓝色,O＝红色,S＝黄色。

[0022] 所描绘的代表性序列是

[0023] SNNLDSKVGGNYNYLYRLFDGTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP(SEQ ID NO:116);

[0024] SNNLDSKVGGNYNYLYRLFNANDKIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP(SEQ ID NO:119);

[0025] SNNLDSKVGGNYNYLYRLFPGTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP(SEQ ID NO:122)。

[0026] 图7A示出支架#15(SEQ ID NO:86)与ACE2(SEQ ID NO:140)的残基19-169的结合。B细胞表位以洋红色显示,T细胞表位以橙色显示,ACE2结合位点以绿色显示。图7B示出具有59个氨基酸(从野生型序列消除了18个氨基酸)的经修饰的CoV-2支架保留了与ACE2的结合亲和力。图7C示出具有67个氨基酸(从野生型序列消除了10个氨基酸)的经修饰的CoV-2支架保留了与ACE2的结合亲和力。B细胞抗体免疫表位区域以洋红色显示,T细胞受体结合,MHC-1和MHC-2负载区域以橙色显示,ACE2结合区域以绿色显示。

[0027] 图8A-8B显示S蛋白支架#9(SEQ ID NO:80)可以装至ACE2(图8A;所示为ACE2的残基19-107)以基于计算机建模确定其ACE2结合亲和力和K_D预测(图8B)。

[0028] 图9A显示基于CoV-1和CoV-2的同源性与ACE2(红色)结合的CoV-1(青绿色)和CoV-2(藏青色)的计算机建模。图9B显示与ACE2(红色)结合的CoV-1(青绿色)的计算机建模。图9C显示与ACE2(红色)结合的CoV-2(藏青色)的计算机建模。

[0029] 图10A和10B分别显示ΔG计算以确定CoV-2和CoV-1的关键结合残基。

[0030] 图11A-11B显示与ACE2结合的CoV-2的热力学建模,图11B为放大的结合界面。图11C显示所确定的CoV-2的两个关键结合基序:残基437-455(SEQ ID NO:65)和残基473至

507(SEQ ID NO:66)。具有负ΔG,正ΔG,和约0ΔG的氨基酸残基分别以绿色,橙色,和黄色显示。主链残基以藏青色显示。L455和P491以洋红色显示。

[0031] 图12A-12J示出具有SEQ ID NO:72的氨基酸序列的CoV-2支架#1的折叠可能性(PyMOL中显示的center0到center9构象)。

[0032] 图13A显示CoV-2支架#9(SEQ ID NO:80)的center0与ACE2的结合。图13B显示CoV-2支架#9的center0和center9与ACE2的结合。图13C-13D显示CoV-2支架#9的center0-center9与ACE2的混沌什锦(assortment),显示了海森堡不确定性原理下合理的平均折叠和所有可能折叠状态的位置。图13D显示支架#9和ACE2的放大的结合界面。

[0033] 图14A描绘与CoV-2S蛋白结合的ACE2的模拟。图14B-14D描绘ACE2支架1(SEQ ID NO:141)(紫色),经由RaptorX模拟,覆盖有野生型hACE2(红色)。ACE2在与CoV-2S蛋白界面处的关键结合残基以绿色突出显示。

[0034] 图15A显示从ACE2蛋白截短的ACE2支架1(SEQ ID NO:141)的计算机建模。图15B描绘ACE2支架1(紫色,关键结合残基以绿色显示)与CoV-2S蛋白(蓝色,抗体结合结构域以蓝绿色箭头显示)的分子建模。支架与CoV-2S蛋白结合,同时在结合时保留S蛋白的抗体结合免疫表位区域的呈现。图15C描绘与CoV-2S蛋白结合的ACE2支架1。ACE2支架1预计不会影响CoV-2S蛋白的免疫结合结构域(粉红色)。

[0035] 图16A显示他人公布的CoV-2S蛋白的Cryo-EM结构与ACE2的结合,图16B显示基于SWISS-MODEL的CoV-2S蛋白与ACE2的结合。

[0036] 图17A显示使用他人公布的结构(上)和本公开的计算机模拟结构(下)与ACE2的模拟构象。图17B显示他人公布的CoV-2的Cryo-EM结构(左)与所公开的截短和标记的SWISS-MODEL模拟结构(右)的比较。红色点虚线椭圆表示Cryo-EM结构中缺失残基的位置。紫色区域表示由他人确定的B细胞免疫表位,而橙色区域表示ACE2排斥性区域,绿色区域表示ACE2结合区域,黄色区域表示ACE2中性区域,经由PDBePISA所确定。

[0037] 图18显示定制的肽机器人在约24分钟内完成了9个氨基酸MHC-1负载表位的合成,允许在商业扩大之前进行快速原型设计。

[0038] 图19A显示使用支架#47(“Ligandal-05”,SEQ ID NO:118)作为示例通过酰胺偶联在溶液中对侧链保护的肽进行头对尾环化。图19B显示使用支架#48(“Ligandal-06”,SEQ ID NO:119)作为示例通过酰胺偶联进行树脂上的头对尾环化。图19C显示使用支架#46(“Ligandal-04”,SEQ ID NO:117)作为示例通过NCL进行纯化的线性硫酯肽的环化。

[0039] 图20A-20I是描绘与在链霉亲和素传感器末端上捕获的ACE2-生物素(2.5nm捕获)结合的支架#4(“肽1”,SEQ ID NO:75),支架#7(“肽4”,SEQ ID NO:78),支架#8(“肽5”,SEQ ID NO:79),和支架#9(“肽6”,SEQ ID NO:80)的生物层干涉测量图以确定支架对ACE2的解离常数。所有支架在10μM浓度处均表现出对RBD与ACE2结合的有效抑制。如图20A-20D所示,随着浓度的增加,观察到每个支架与ACE2的明确结合(减去空白值)。如图20E-20H所示,还观察到剂量反应曲线,据此RBD能够在不存在肽的情况下在35μM处与每个传感器强烈结合(绿色,顶曲线),并经历了肽剂量反应依赖性结合抑制(蓝色,青绿色,和红色分别代表10,3,和1μM浓度)。图20I对应于在链霉亲和素传感器末端上捕获的RBD-生物素(5nm捕获),随后与ACE2结合。

[0040] 图21A-21F是描绘与在抗人IgG(AHC)传感器末端上捕获的中和抗体(1nm捕获)结

合的支架的生物层干涉测量图，用于确定支架#4(“肽1”，SEQ ID NO:75)，支架#7(“肽4”，SEQ ID NO:78)，支架#8(“肽5”，SEQ ID NO:79)，和支架#9(“肽6”，SEQ ID NO:80)对中和抗体的解离常数(图21A-21D)。用抗RBD中和抗体测定了RBD浓度增加的解离常数(图21E)。图21F显示在引入与传感器结合的中和抗体之前，将117nM RBD与浓度增加的ACE2混合，以证明ACE2对与RBD结合的中和抗体的抑制。

[0041] 图22显示与支架#4(“肽1”，SEQ ID NO:75)，支架#7(“肽4”，SEQ ID NO:78)，支架#8(“肽5”，SEQ ID NO:79)，和支架#9(“肽6”，SEQ ID NO:80)共转染时，ACE2-HEK293细胞在SARS-CoV-2刺突感染后于感染后60小时的发光(RLU)。对照组包括未转染的ACE2-HEK293细胞(无病毒)和被SARS-CoV2刺突蛋白转染的ACE2-HEK293细胞。

[0042] 图23A-23D显示被SARS-CoV-2刺突蛋白或无病毒(对照)和支架#8(“肽5”，SEQ ID NO:79)(图23A)，可溶性ACE2(图23B)，SARS-CoV-2刺突蛋白的可溶性受体结合结构域(RBD)(图23C)，和SARS-CoV-2中和抗体(neuAb)(图23D)转染的ACE2-HEK293细胞中的发光(RLU)。

[0043] 图24A显示支架#4(“LGDL_NIH_001”，SEQ ID NO:75)，支架#7(“LGDL_NIH_004”，SEQ ID NO:78)，支架#8(“LGDL_NIH_005”，SEQ ID NO:79)，和支架#9(“LGDL_NIH_006”，SEQ ID NO:80)在微摩尔浓度下表现出对活病毒中病毒载量的抑制超过90％(EC90)。图24B显示图24A中所测试的支架在有效浓度下没有毒性。

[0044] 图25描绘支架#4(“肽1”，SEQ ID NO:75)的三维分子建模，其基于OpenMM中从类天然构象开始的180ns运行(单轨迹)。

[0045] 图26是绘制在所示时间点的支架#4(SEQ ID NO:75)的Rosetta评分(REU)图。

[0046] 图27A和27B是S蛋白上仅在融合期间暴露的表位的示意图。

[0047] 图28A和28B是将阻止该过程移动至下一步中和的结合位点的示意图。图28C显示放大的结合位点。

[0048] 图29A和29B(放大)描绘装至SARS-CoV-2刺突蛋白(绿色)的序列KMSECVLGQSKRV(SEQ ID NO:8)(以红色显示)的三维分子建模。SEQ ID NO:8对应于图27-28中鉴定的结合位点之一，该结合位点在预成束(pre-bundle)阶段期间位于七肽重复(HR)1(HR1)和HR2之间的铰链中。

[0049] 图30描绘外显肽插入位置的示意图。

[0050] 图31A-31D是在肽筛选和优化期间生成的示意图。

[0051] 图32A-32B显示了本文公开的代表性SARS-CoV-2S蛋白支架的序列比对。比对显示SEQ ID NO:2的氨基酸残基433-511。关键结合基序加下划线。取代加双下划线并以黄色突出显示。GS接头加粗体并以蓝色突出显示。B细胞和T细胞结合的表位加粗体加斜体并以绿色突出显示。EPEA C标记加斜体并以灰色突出显示。带多电荷的N-和C-端残基加波浪下划线并以粉红色突出显示。替代TCR表位以红色突出显示。

[0052] 图33A-33E示出使用IDT siRNA设计工具的siRNA设计过程，包括所选择的正义和反义链(SEQ ID NO:143-148)的位置和序列。

[0053] 图34描绘与血管紧张素转换酶2(ACE2)(PDB ID 6CS2;红色)结合的SARS-CoV-1的三维模拟模型以近似于SWISS-MODEL模拟的SARS-CoV-2的结合界面(左);所选择的MHC-I和MHC-II表位区域包含在支架#8中(粉红色)，代表S1刺突蛋白中的P807-K835和A1020-Y1047。右侧模型描绘SARS-CoV-2刺突蛋白(蓝色/多色)与ACE2(红色)模拟结合的受体结合

结构域(RBD)。模拟模型鉴定所预测的热力学有利(绿色),中性(黄色),和不利(橙色)相互作用。用于生成支架的氨基酸(V433-V511)的外边界在右侧以青绿色显示。

[0054] 图35描绘与支架#4,#7,#8,和#9比对的ACE2受体(红色)的3维模拟模型(上图,从左到右)。肽5的多种折叠状态在与ACE2的模拟结合中示出(下图)。所预测的结合残基以绿色表示(上图和下图)。

[0055] 图36显示SARS-CoV-2基因组序列(SEQ ID NO:1)。核苷酸21536-25357(加下划线)编码SEQ ID NO:2的S蛋白。核苷酸26218-26445(加双下划线)编码SEQ ID NO:3的包膜蛋白。

[0056] 图37显示SARS-CoV-2刺突(S)蛋白的氨基酸序列(SEQ ID NO:2)。

[0057] 图38显示ACE2的氨基酸序列(SEQ ID NO:140)。

[0058] 发明详述

[0059] 如本文所公开,通过结合来自数学数据科学,生物物理学,和实验生物学的方法,可以预测最有可能通过自然选择扩大宿主范围并增加SARS-CoV-2在人类群体中的稳定性的S蛋白序列。开发了一种计算管道来估计SARS-CoV-2S蛋白的突变形势。对所预测的序列进行实验设计,并使用生化试验和冷冻电子显微镜测量其与人受体ACE2的结合。

[0060] 受遗传算法结构的启发,开发了新的数学方法用于识别SARS-CoV-2的S蛋白的高度可能序列。具而言之,所公开的方法将来自图论,拓扑数据分析,和计算生物物理学的描述符并入结合了神经网络和遗传算法的一个新的机器学习框架。这种强大的跨学科方法允许使用来自SARS-CoV-2的已有数据来发现一些最有可能出现在其病毒S蛋白进化过程中的候选序列。这些结果通过从所获得的序列生成肽进行了实验验证。由此产生的管道提供了新的解决方案以更好地了解病毒蛋白的突变形势。

[0061] 如本文所公开,通过预测1)ACE2受体结合区域,2)T细胞受体MHC-I和MHC-II负载的免疫表位区域,和3)B细胞受体或抗体结合的免疫表位区域进行计算机模拟(in silico)分析以生成和筛选被设计用作SARS-CoV-2刺突(S)蛋白的竞争性抑制剂的新肽(“支架”)。如工作示例所示,使用三维建模和计算机模拟分析以检验新肽的所预测结构,评估了多个序列修饰(例如,通过检验候选肽的Rosetta能量单位(REU)评分),并通过计算机模拟所预测的结合模型。基于这些结果,本文提供的是通过获取肽序列(例如,SARS-CoV-2刺突蛋白),引入序列修饰,和使用三维建模技术以预测折叠或结合构象来生成和优化在疫苗开发中用作竞争性抑制剂的肽支架的方法。本文还提供了使用这些方法设计的优化的肽支架,包含这些肽支架的制剂,和使用这些肽支架和制剂以竞争性抑制病毒蛋白或治疗病毒感染的方法,以及使用这些肽支架和制剂作为疫苗来预防病毒感染。

[0062] 因此,本公开涉及用于快速疫苗原型设计的突破性方法。在一些方面,所公开的疫苗方法提供了用于仿拟T细胞受体和抗体结合表位的完全合成支架,其可以快速定制以适应新的病毒突变形式。另外地,该合成支架可用作仿拟病毒进入的靶向配体,以和治疗剂靶向病变细胞和组织。这些“微型病毒”支架可以在数小时内合成并迅速扩大至100kg以上的规模以满足全球需求。另外地,本文提供的支架可单独用于代替小分子以抑制结合裂隙(cleft)的相互作用。

[0063] 本文公开的支架是通过模拟SARS-CoV-2刺突蛋白受体结合基序(RBM)保守基序而生成的肽,具有作为针对病毒的预防剂,免疫刺激剂,和治疗剂的潜在用途。因此,本文还公

开了包含一个或多个支架的组合物,其可用于:1)抑制ACE2-刺突相互作用和病毒进入表达ACE2的细胞,2)促进与中和抗体的结合而不竞争性地代替中和抗体与该RBD的结合;和/或3)防止可溶性ACE2与该RBD结合。

[0064] 本公开详细描述了肽支架的模拟,设计,合成,和表征,这些肽支架被设计为阻断病毒与表达ACE2的细胞结合,同时刺激免疫反应并促进刺突蛋白暴露以被免疫系统识别。与寻求靶向病毒的中和抗体疗法和其他方法相比,开发了一种仿生病毒诱饵肽技术来竞争与细胞的结合,并暴露病毒使其与中和抗体结合。

[0065] I.计算机辅助的3D建模

[0066] A.分析结合界面

[0067] 在一个方面,本公开涉及计算机辅助的三维(3D)建模以研究蛋白质-蛋白质相互作用的方法。这些方法包括产生第一结合配偶体和第二结合配偶体的3D模型,确定每个结合配偶体的氨基酸序列,3D结构,和界面构象,截短每个结合配偶体的结合界面,同时保持每个结合配偶体的3D结构以获得代表每个结合配偶体的支架,基于每个残基的热力学能量计算确定该支架中每个氨基酸残基的结合亲和力,并确定该支架中关键结合基序的位置和序列。在某些实施例中,3D模型用SWISS-MODEL基于与该第一结合配偶体或该第二结合配偶体的蛋白质序列同源性而产生。可以对该支架进行多个修饰以维持或改善该支架的结构,构象,和结合亲和力。这些修饰包括但不限于该支架中一个或多个氨基酸残基的插入,缺失,或取代。如本公开所详述,可将多个接头,可缀合结构域,和/或免疫表位添加至该支架以获得多功能支架。在某些实施例中,可以缺失或取代对结合不是关键的一个或多个氨基酸。在某些实施例中,该结合配偶体是SARS-CoV-2S蛋白和ACE2。在某些实施例中,S蛋白具有SEQ ID NO:2所列的氨基酸序列。在其他实施例中,S蛋白是变体,包括但不限于B.1.1.7变体(SEQ ID NO:137),B.1.351变体(SEQ ID NO:138),或P.1变体(SEQ ID NO:139)。可以在nextstrain.org/ncov/global上找到其他冠状病毒变体。在某些实施例中,ACE2具有SEQ ID NO:140所列的氨基酸序列。

[0068] 如本文所用,术语“支架”是指位于结合配偶体的结合界面处并涉及与其他结合配偶体的结合的一段连续氨基酸序列。在某些实施例中,该支架的大小为小于约120个氨基酸残基,小于约110个氨基酸残基,小于约100个氨基酸残基,小于约90个氨基酸残基,小于约80个氨基酸残基,小于约70个氨基酸残基,小于约60个氨基酸残基,或小于50个氨基酸残基。在某些实施例中,该支架保持衍生其一个或多个结合序列的蛋白质的天然,游离,或结合状态的3D结构和/或构象。例如,该支架可被设计为当从野生型蛋白质序列截短时保持α-螺旋和/或β折叠结构。在某些实施例中,该支架可包含一个或多个修饰,如插入,缺失,和/或取代,只要这些修饰不会显著降低,并且在一些实施例中实际上增加该支架与其结合配偶体的结合亲和力。

[0069] 如本文所公开,将SARS-CoV-2刺突蛋白的蛋白质序列(SARS-CoV-2或CoV-2;SEQ ID NO:2)与SARS冠状病毒蛋白质序列(PDB ID 6CS2)进行比较以产生CoV-2与ACE2结合的3D模型(图1)。研究了CoV-2和ACE2中每一个的结合界面以确定参与结合的一段氨基酸残基。该段氨基酸序列可从剩余的蛋白质序列截短,并保持这段氨基酸序列的结构和/或构象,以模拟处于游离或结合状态的天然蛋白质的结构和/或构象,从而获得本公开的CoV-2支架或ACE2支架。

[0070] 因此,本文公开了一个CoV-2支架,其具有与SEQ ID NO:2的残基433-511的氨基酸序列(VIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV)至少40%,至少45%,至少50%,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%相同的氨基酸序列。

[0071] 在一些实施例中,该CoV-2支架具有与氨基酸序列:

[0072] VIAWNSNNLDSK*V*GGNYNYL**Y**R*L*FRKSNLKPFERDISTEI**Y**Q*A*GSTPCNGV**E** G*F***N**C*YF*PL**QSY***G***F****QP***T***N***G***V**G***Y***Q**P**YRVV(“支架#1;SEQ ID NO:72)至少40%,至少45%,至少50%,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%相同的氨基酸序列。

[0073] 在上述SEQ ID NO:72中,CoV-2S蛋白主链中的氨基酸残基以普通字母显示(包括433V-436W,F456-I472,和Y508-V511),在结合中为中性的具有约0ΔG的氨基酸残基加下划线(包括N437,S438,N440-K444,G447,N448,N450-L452,R454,S477-V483,G485,C488,P491,L492,F497,G504,和P507),为关键结合残基的具有负ΔG的氨基酸残基以粗体显示(包括N439,Y449,Y453,Q474,E484,N487,Q493-Y495,Q498,P499,N501,和Q506),和为排斥性残基的具有正ΔG的氨基酸残基以斜体显示(包括V445,G446,L455,Y473,A475,G476,F486,Y489,F490,G496,T500,G502,V503,和Y505)。

[0074] 基于计算机建模和热力学能量的计算,确定本公开的一个CoV-2支架包含第一关键结合基序,其包含SEQ ID NO:2的残基437至455,第二关键结合基序,其包含SEQ ID NO:2的残基473至507,和主链区域,其包含SEQ ID NO:2的残基456至472。该第一和第二关键结合基序在结合界面上直接与ACE2相互作用,而该主链区域包含不直接与ACE2相互作用的氨基酸残基。

[0075] 该CoV-2支架可进一步在N-端,C-端,或两者处包含来自CoV-2S蛋白主链的一个或多个氨基酸以达到想要的大小。在一些实施例中,该CoV-2支架包含约40至约200个氨基酸残基,约50至约100个氨基酸残基,约55至约95个氨基酸残基,约60至约90个氨基酸残基,约65至约85个氨基酸残基,约70至约80个氨基酸残基。在一些实施例中,该CoV-2支架包含约50个氨基酸残基,约55个氨基酸残基,约60个氨基酸残基,约65个氨基酸残基,约70个氨基酸残基,约75个氨基酸残基,约80个氨基酸残基,约85个氨基酸残基,约90个氨基酸残基,约95个氨基酸残基,或约100个氨基酸残基。

[0076] 尽管该CoV-2支架的大小可以变化,并且可以结合某些修饰,如插入,缺失,和/或取代,但该支架关于ACE2结合界面保持其天然状态的结构和/或构象。保留这种结构和/或构象允许该支架以与全长S蛋白相同或更高的亲和力结合ACE2,尽管其被截短。例如,β折叠结构得以保持,并且可以通过进一步的修饰而稳定。在一些实施例中,该CoV-2支架包含L455C和P491C取代,使得在位置455和位置491之间形成二硫键以稳定该β折叠结构。基于计算机建模,这两个位置在与ACE2结合的天然CoV-2S蛋白中似乎彼此靠近。在一些实施例中,该CoV-2支架包含一个或多个突变以替代一个或多个已有的Cys残基,使得仅剩下的Cys残基是在位置455和491处引入的那些,以避免形成二硫键的任何不想要的干扰。例如,Cys可以被Gly,Ser,或任何其他残基取代,只要该取代不会损害与ACE2的结合亲和力。替代Cys残

基的一些示例包括但不限于C480G和C488G。

[0077] 在某些实施例中，本文公开的该CoV-2支架可进一步包含环以使用接头如胺-羧基接头连接N-端残基和C-端残基以获得头对尾环化的支架。在某些实施例中，该支架的环化提供增加的稳定性，具有更低的自由能，增强的折叠，结合，或与底物的缀合，和/或增强的溶解度。该环不直接与该支架的结合配偶体相互作用。在某些实施例中，该环允许该支架附接至siRNA有效负载或其他底物。在某些实施例中，该环包含1-200个氨基酸残基。在某些实施例中，该环包含小于约150个氨基酸残基。取决于想要的该支架，接头，可缀合结构域，极性头部或尾部的构象等，可以相应地调整该环的大小。在一些实施例中，该环包含9-15个Arg和/或Lys残基。在一些实施例中，该环包含可缀合结构域，如马来酰亚胺或其他接头以将该支架缀合至底物或多氨基酸链。在一些实施例中，该环包含一种或多种免疫激活多氨基酸链或免疫反应性聚糖。N-端和C-端也可以通过形成二硫键，任何其他合适的接头(柔性或刚性)，点击化学，PEG，聚肌氨酸，或生物缀合来连接。因此，肽可被环化，稳定化，线性化，以其他方式点击化学或生物缀合，或被非天然氨基酸，类肽，糖肽，脂质，胆固醇部分，多糖，或任何增强折叠，结合，溶解度，或稳定性的物质取代。

[0078] 本文还公开了ACE2支架，其与SEQ ID NO:140的残基19-84的氨基酸序列至少40％，至少45％，至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％，至少99％，或100％相同。

[0079] 在一些实施例中，该ACE2支架与氨基酸序列：

[0080] **S**TIEE**Q**AKTFLDKFNHEAE**D**LFY**Q**SSLASWNYNTNITEENVQNMNNAGDK

WSAFLK**EQ**STLAQM**Y**P（SEQ ID NO:151）至少40％，至少45％，至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％，至少99％，或100％相同。

[0081] 在上述SEQ ID NO:151中，为关键结合残基的具有负ΔG的氨基酸残基以粗体显示(包括S19，Q24，D38，Q42，E75，Q76，和Y83)。

[0082] 基于计算机建模和热力学能量的计算，确定本公开的ACE2支架包含第一关键结合基序，其包含SEQ ID NO:140的氨基酸残基19至42，第二关键结合基序，其包含SEQ ID NO:140的残基64至84，和主链区域，其包含SEQ ID NO:140的残基43至63。该第一和第二关键结合基序在结合界面上直接与CoV-2S蛋白相互作用，而主链包含ACE2主链上的氨基酸残基且不直接与CoV-2S蛋白相互作用。

[0083] 在一些实施例中，该ACE2支架包含连接该两个关键结合基序的接头(以粗体显示)，参见例如，ACE2支架1(SEQ ID NO:141)：

STIEEQAKTFLDKFNHEAEDLFYQ**GSGSG**NAGDKWSAFLKEQSTLAQMYP

[0084] 在一些实施例中，该ACE2支架进一步包含EPEAC-标签(加下划线)，参见例如，ACE支架2(SEQ ID NO:142)：

STIEEQAKTFLDKFNHEAEDLFYQ**GSGSG**NAGDKWSAFLKEQSTLAQMYP

[0085]

EPEA

[0086] 该ACE2支架可进一步包含来自ACE2蛋白主链或在合适界面处重建ACE2与衍生自ACE2的N-端，C-端，或两者的刺突蛋白的结合作用的一个或多个氨基酸或单体单元，以达到

想要的大小，折叠，和亲和力。在一些实施例中，该ACE2支架包含约10至约200个氨基酸残基，约50至约100个氨基酸残基，约55至约95个氨基酸残基，约60至约90个氨基酸残基，约65至约85个氨基酸残基，约70至约80个氨基酸残基。在一些实施例中，该ACE2支架包含约50个氨基酸残基，约55个氨基酸残基，约60个氨基酸残基，约65个氨基酸残基，约70个氨基酸残基，约75个氨基酸残基，约80个氨基酸残基，约85个氨基酸残基，约90个氨基酸残基，约95个氨基酸残基，或约100个氨基酸残基。

[0087] 尽管该ACE2支架的大小可以变化，并且可以结合某些修饰，如插入，缺失，和/或取代，但该支架关于CoV-2S蛋白结合界面保持其天然状态的结构和/或构象。保留这种结构和/或构象允许该支架以与全长ACE2蛋白相同或更高的亲和力结合S蛋白，尽管其被截短。

[0088] 在某些实施例中，ACE2支架的N-端，C-端，或两个末端都用任意数量的生物缀合基序，接头，间隔子，标签(如his-标签和C-标签)等进行修饰。在某些实施例中，对于结合至CoV-2S蛋白非关键的一个或多个氨基酸被缺失或取代。

[0089] 可以设计额外的支架以仿拟ACE2与CoV-2S蛋白的结合。这些ACE2支架可以与CoV-2病毒结合以包被病毒，使得该病毒无法与ACE2结合，从而抑制病毒进入人体(或其他宿主)。此外，可以进一步修饰ACE2支架以包括例如，片段结晶(Fc)结构域或用于激活免疫反应的替代结构域。

[0090] 预计与野生型ACE2相比，包含第一关键结合基序，第二关键结合基序，和连接该关键结合基序的接头的ACE2支架1(SEQ ID NO:141)对CoV-2S蛋白具有更高的亲和力。另外地，与结合至CoV-2S蛋白并阻断CoV-2S蛋白的免疫表位区域的ACE2相比，预计ACE2支架1不会影响CoV-2S蛋白的免疫结合结构域并允许免疫系统识别CoV-2病毒。与本文提供的其他支架相似，ACE2支架1可以纳米颗粒或其他合适的底物提供，并且可起到聚集病毒的作用。例如，N-或C-末端可用任何数量的生物缀合基序，接头，间隔子等进行修饰；并且可具有多种底物，包括巴基球(例如，C60/C70富勒烯)，支链PEG，超支化树枝状大分子，单壁碳纳米管，双壁碳纳米管，KLH，OVA，和/或BSA。预计ACE2支架1与游离ACE2相比对病毒刺突蛋白的亲和力更高。

[0091] B.分析免疫表位

[0092] 免疫表位数据库(IEDB)用于在出现具有多种HLA等位基因的人群的多种T细胞受体(TCR)反应的临床数据之前预测关键表位。将这些所预测的表位与SARS-CoV-1中MHC-I和MHC-II反应的已知表位进行了比较。先前报道了具有氨基酸序列LPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYG(SEQ ID NO:135)的S5肽(SARS-CoV-1的残基788-820)和具有氨基酸序列ASANLAATKMSECVLGQSKRVDFCGKGYH(SEQ ID NO:136)的S6肽(SARS-CoV-1的残基1002-1030)表现出的免疫原性反应类似于使用SARS-CoV S蛋白的截短重组蛋白类似物进行的平行研究中发现的免疫原性反应(2)。S5肽是基于单价肽在其引发MHC-I反应和抗体反应的能力方面的已知免疫原性而定义，而许多其他肽仅在多价存在时具有免疫原性。S6肽代表来自SARS-CoV-1的已知MHC-II结构域。

[0093] 将SARS-CoV-1S蛋白的这些免疫表位与CoV-2S蛋白进行比对，以确定CoV-2S蛋白上可能的免疫原性位点。基于同源性，CoV-2S蛋白中相应的免疫表位鉴定如下，并且还被设计为在S1-S2界面的TMPRSS2裂解后与其融合前构象中的S2刺突区域重叠：

[0094] (QIL)PDPSKPSKRSFIED**LLFNKVTLA**DAGFIK（SEQ ID NO:67）（位置

804-835)，和

[0095] ASANLAAT**KMSECVLGQSKRV**DFCGKGY（SEQ ID NO:69）（位置1020-1047)

[0096] SARS-CoV-1免疫表位的3D结构如图2A所示，CoV-2免疫表位的3D结构及其在CoV-2S蛋白上的位置如图2B-2D所示。IEDB确定了SARS-CoV-2S蛋白的序列KMSECVLGQSKRV(SEQ ID NO:71)和LLFNKVTLA(SEQ ID NO:7)，分别代表HLA-A*02:01的MHC-II和MHC-I结合结构域，将分别具有0.9和1.2的百分等级的免疫原性。较低的百分等级代表更好的结合。图3显示这些免疫表位的MHC-I结合预测结果。因此，具有以下序列的免疫表位用于进一步研究并包含在一些支架中：

[0097] KMSECVLGQSKRV(SEQ ID NO:71)和LLFNKVTLA(SEQ ID NO:7)。

[0098] 可以从多个数据库识别额外的表位。例如，在TepiTool结果中，YLQPRTFLL(SEQ ID NO:9)，FIAGLIAIV(SEQ ID NO:22)，和FVFLVLLPL(SEQ ID NO:21)是评分最高的HLA—A*0201表位。这些评分最高的表位非常疏水。一些评分最高的表位，或者如果可用的话，在体内或体外表现免疫原性的表位可以包括在本文公开的支架中。

[0099] 图4显示他人识别的抗体表位如CoV-2S蛋白中的B细胞表位(12)与CoV-2S蛋白氨基酸序列的比对。

[0100] 如图5所示，具有SEO ID NO:4的氨基酸序列(下)的截短CoV-2S蛋白与ACE2比对以显示抗体表位(洋红色)和ACE2结合残基(绿色)的位置。

[0101] CPFGEVFNATRFASVY**AWNRKR**ISNCVADYSVLYNSASFSTFKCYGVSPTK LNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAW**N** **SNNLD**SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNC YFPLQSY**GFQPT**NGVG**YQPY**RVVVLSFELLHAPATVCG**PKKS**T（SEQ ID NO:4;一些免疫表位以粗体突出显示)。

[0102] 使用Bepipred工具进行序列搜索表明大多数受体结合基序(残基440-501)被预测为B细胞线性表位。

[0103] PDB也可用于识别B细胞表位。例如，PDB列出了先前通过实验探索过的八个表位。已证实SARS-CoV-2S蛋白上的两个线性表位可在COVID-19患者中引发中和抗体(12)。B细胞表位的一些示例包括:PSKPSKRSFIEDLLFNKV(S21P2)(SEQ ID NO:30)，TESNKKFLPFQQFGRDIA(S14P5)(SEQ ID NO:25)，PATVCGPKKSTNLVKNKC(SEQ ID NO:24)，GIAVEQDKNTQEVFAQVK(SEQ ID NO:26)，NTQEVFAQVKQIYKTPPI(SEQ ID NO:27)，PIKDFGGFNFSQILPDPS(SEQ ID NO:29)，PINLVRDLPQGFSALEPL(SEQ ID NO:23)，和VKQIYKTPPIKDFGGFNF(SEQ ID NO:28)。

[0104] 如下文详细公开的，本文公开的支架可以被修饰以包括一个或多个免疫表位，包括T细胞表位和/或B细胞表位。

[0105] 这些结果表明，可以用与Cryo-EM和其他高分辨率结构数据一致的方式预测结合袋(bindingpockets)。这种技术可用于快速处理任何已知或新病毒的未来突变，即使整个病毒的基因组数据显示只有80％的相似性。本文公开的技术还结合了用于绘制TCR和BCR/抗体表位的生物信息学驱动方法，允许蛋白质大小的“压缩算法”。与重组技术和其他方法

相比，本文公开的技术利用小肽，如约1200个氨基酸的刺突蛋白中小于70个氨基酸的肽，以生成用于ACE2结合和TCR/抗体识别的多功能支架。

[0106] II.支架/肽修饰

[0107] 本文公开了CoV-2支架或ACE2支架，其包含来自CoV-2S蛋白和ACE2中每一个的结合界面的一个或多个氨基酸序列片段，同时基本上保持处于其游离或结合状态的天然蛋白的结构和/或构象。本文公开的支架基本上保持或提高了与相应结合配偶体的结合亲和力。例如，本文公开的CoV-2支架基本上保持或提高了与野生型ACE2的结合亲和力；本文公开的ACE2支架基本上保持或提高了与野生型CoV-2S蛋白的结合亲和力。该CoV-2支架或该ACE2支架包含约10至约100个氨基酸残基，15至约30个氨基酸残基，约55至约95个氨基酸残基，约60至约90个氨基酸残基，约65至约85个氨基酸残基，约70至约80个氨基酸残基。在一些实施例中，该CoV-2支架或该ACE2支架包含约50个氨基酸残基，约55个氨基酸残基，约60个氨基酸残基，约65个氨基酸残基，约70个氨基酸残基，约75个氨基酸残基，约80个氨基酸残基，约85个氨基酸残基，约90个氨基酸残基，约95个氨基酸残基，或约100个氨基酸残基。在一些实施例中，该CoV-2支架或ACE2支架包含两个或多个序列，其增强一个或多个支架的结合，代替，或免疫原性。在一些实施例中，拟态病毒或拟态病原体的支架不需要与SARS-CoV-2及其与ACE2的结合有关，并且可以衍生自与其伴随的人或一个或多个宿主蛋白结合配偶体结合的任何病原体，包括真核生物和原核生物物种。

[0108] 在某些实施例中，本文公开的是CoV-2支架或ACE2支架，各自包含两个关键结合基序，其中该关键结合基序涉及与结合配偶体的直接结合。在一些实施例中，该支架进一步包含一个或多个主链区域，该主链区域包含不涉及与该结合配偶体直接结合的氨基酸残基。在一些实施例中，该主链区域位于该两个关键结合基序之间。在一些实施例中，该主链区域位于该第一关键结合基序的N-端。在一些实施例中，该主链区域位于该第二关键结合基序的C-端。

[0109] 在某些实施例中，本文公开的CoV-2支架包含第一关键结合基序，其与NSNNLDSKVGGNYNYLYRL(SEQ ID NO:65)的氨基酸序列具有至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少98％，或100％的同一性，和第二关键结合基序，其与YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP(SEQ ID NO:66)的氨基酸序列具有至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少98％，或100％的同一性。

[0110] 在某些实施例中，本文公开的ACE2支架包含第一关键结合基序，其与STIEEQAKTFLDKFNHEAEDLFYQ(SEQ ID NO:149)的氨基酸序列具有至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少98％，或100％的同一性，和第二关键结合基序，其与NAGDKWSAFLKEQSTLAQMYP(SEQ ID NO:150)的氨基酸序列具有至少50％，至少55％，至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少98％，或100％的同一性。

[0111] 本文公开的支架可具有不同的大小，这取决于在该两个关键结合基序之间，在该第一关键结合基序的N-端处，和/或在该第二关键结合基序的C-端处所包含的来自主链的氨基酸残基的数量。参见，例如，支架#1(SEQ ID NO:72)(上图)和支架#10(SEQ ID NO:81)(下图)，氨基酸序列比对如下。

SEQ ID NO:72 VIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV

SEQ ID NO:81 ----NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY---

[0112] 在某些实施例中,该支架中的一个或多个氨基酸残基被缺失或取代。例如,一个或多个具有正ΔG的排斥性氨基酸残基被缺失或取代,一个或多个具有约0ΔG的中性氨基酸残基被缺失或取代,和/或关键结合基序之外的一个或多个氨基酸残基被缺失或取代,例如,在主链区域。尽管并非想要的,但一个或多个具有负ΔG的关键氨基酸残基可被缺失或取代。

[0113] 在某些实施例中,通过用具有多个长度的接头如GS接头替代该关键结合基序之间的该主链区域中的氨基酸残基来修饰该支架。在一些实施例中,该支架包含具有约1至约20个氨基酸残基的接头,例如,该接头具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20个氨基酸残基。可以优化该接头的大小以实现想要的该支架的结构和/或构象。参见,例如,支架#1(SEQ ID NO:72),支架#3(SEQ ID NO:74),支架#11(SEQ ID NO:82),支架#12(SEQ ID NO:83),支架#13(SEQ ID NO:84),和支架#14(SEQ ID NO:85),按照出现的顺序从上至下,氨基酸序列比对如下。GS接头以粗体显示。

[0114]

```
VIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV
VIAWNSNNLDSKVGGNYNYLYRLGSGSG              QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV
    NSNNLDSKVGGNYNYLYRLGSGSGS             QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
    NSNNLDSKVGGNYNYLYRLGSGSG              QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
    NSNNLDSKVGGNYNYLYRLGSGS               QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
    NSNNLDSKVGGNYNYLYRLGSG                QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
```

[0115] 在某些实施例中,该支架包含一个或多个免疫表位,如一个或多个T细胞表位,一个或多个B细胞表位,或两者。该免疫表位可以包含在非界面环结构内,该非界面环结构替代了该支架的该两个关键结合基序之间的该主链区域的全部或部分序列。例如,该主链区域中的一个或多个氨基酸残基可以被一个或多个免疫表位替代。在另一个示例中,该关键结合基序中的一个或多个氨基酸残基,优选排斥性或中性氨基酸残基,可以被一个或多个免疫表位替代。根据该支架的想要的大小和结构,可以选择哪些氨基酸残基被一个或多个免疫表位替代。

[0116] 在某些实施例中,该免疫表位是9或13个氨基酸残基长,对应于MHC-I和MHC-II结合。例如,T细胞表位包括但不限于KMSECVLGQSKRV(SEQ ID NO:8),和LLFNKVTLA(SEQ ID NO:7)。该支架中也可包括其他已知的表位。例如,显性TCR表位包括KLWAQCVQL(SEQ ID NO:10)(ORF1ab,3886-3894,17.7nM,主要用于A*02),YLQPRTFLL(SEQ ID NO:9)(S,269-277,5.4nM,主要用于A*02),和LLYDANYFL(SEQ ID NO:11)(ORF3a,139-147,主要用于A*02)(3)。其他已知的TCR表位包括PRWYFYYLGTGP(SEQ ID NO:12)(核衣壳),SPRWYFYYL(SEQ ID NO:13)(核衣壳,主要用于B*07:02,A*11:01,A*03:01),WSFNPETN(SEQ ID NO:14)(膜蛋白),QPPGTGKSH(SEQ ID NO:15)(ORF1ab多蛋白),和VYTACSHAAVDALCEKA(SEQ ID NO:16)(ORF1ab多蛋白)(1,4-6)。一些表位很强但可能是HLA限制性的,如KTFPPTEPK(SEQ ID NO:17)(N蛋白;20.8nM),CTDDNALAYY(SEQ ID NO:18)(ORF1ab;5.3nM),TTDPSFLGRY(SEQ ID NO:19)(ORF1ab;7.2nM),和FTSDYYQLY(SEQ ID NO:20)(ORF3a;3.2nM)。

[0117] 在某些实施例中,B细胞表位包括但不限于FDEDDS(SEQ ID NO:63),IQKEIDRL(SEQ ID NO:62),KYFKNHTSP(SEQ ID NO:61),MAYR(SEQ ID NO:56),NVLYENQ(SEQ ID NO:57),QSKR(SEQ ID NO:58),YQPY(SEQ ID NO:45),SEFR(SEQ ID NO:36),TPGDSS(SEQ ID NO:

38),TTKR(SEQ ID NO:64),YYHKNNKSWM(SEQ ID NO:35),ASTEK(SEQ ID NO:33),AWNRKR(SEQ ID NO:41),DPSKPSKRSF(SEQ ID NO:55),DQLTPTWRVY(SEQ ID NO:50),EQDKNTQ(SEQ ID NO:54),ESNKK(SEQ ID NO:47),FPQSA(SEQ ID NO:59),GFQPT(SEQ ID NO:44),GTNTSN(SEQ ID NO:49),HVNNSY(SEQ ID NO:51),IADTTDAVRDPQT(SEQ ID NO:48),IYSKHT(SEQ ID NO:37),KYNENGT(SEQ ID NO:39),LDSKTQ(SEQ ID NO:34),LKPFERDI(SEQ ID NO:43),LTTRTQLPPAYTNS(SEQ ID NO:31),NSNNLD(SEQ ID NO:42),PKKS(SEQ ID NO:46),QTSNFRVQPT(SEQ ID NO:40),SMTKT(SEQ ID NO:53),TNGTKRFD(SEQ ID NO:32),VPAQEKNFT(SEQ ID NO:50),和YQTQTNSPRRAR(SEQ ID NO:52)。CoV-2S蛋白中B细胞表位的位置如图4所示。

[0118] 参见,例如,支架#10(SEQ ID NO:81),支架#15(SEQ ID NO:86),支架#16(SEQ ID NO:87),支架#17(SEQ ID NO:88),支架#18(SEQ ID NO:89),支架#19(SEQ ID NO:90),和支架#20(SEQ ID NO:91),按照出现的顺序从上至下,氨基酸序列比对如下。支架#1(SEQ ID NO:72),支架#3(SEQ ID NO:74),支架#11(SEQ ID NO:82),支架#12(SEQ ID NO:83),支架#13(SEQ ID NO:84),支架#14(SEQ ID NO:85),按照出现的顺序从上至下,氨基酸序列比对如下。该关键结合基序中的氨基酸残基加下划线,免疫表位以粗体显示。取决于该支架的想要的大小和/或结构,该免疫表位可以替代该关键结合基序之间的该主链区域的全部或部分序列,和/或可以替代该关键结合基序的部分序列,特别是该关键结合基序中的排斥性和/或中性氨基酸残基。

[0119]

NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEI YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRL **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRLFR **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRLFRKS **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRLFRKSN **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRLFRKSNL **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY
NSNNLDSKVGGNYNYLYRLFRKSNLK **KMSECVLGQSKRVQALLFNKVTLA**GFNCYFPLQSYGFQPTNGVGYQPY

[0120] 在某些实施例中,该支架包含一个或多个Cys取代,使得可以经由二硫键在想要的位置处形成Cys-Cys桥。例如,进行L455C和P491C取代以引入Cys-Cys桥以维持或稳定该支架的β折叠结构。在一些实施例中,其他位置处的Cys残基可以被Gly或其他残基取代以避免干扰想要的位置处的Cys-Cys桥。在其他实施例中,其他点击化学或联硒化物化学技术可用于桥接一个或多个支架的两个氨基酸或单体区域以重建想要的结构。

[0121] 在某些实施例中,该支架进一步包含头部和/或尾部,其包含附接至N-端,C-端,或两者的一个或多个带电荷的氨基酸,如聚(Arg),聚(Lys),聚(His),聚(Glu),或聚(Asp)。添加这些阳离子或阴离子序列以制备本文公开的支架的静电纳米颗粒。

[0122] 在某些实施例中,该支架包含一个或多个氨基酸取代以增加ACE2结合亲和力,抗体亲和力,或两者。例如,增加ACE2结合亲和力的取代包括但不限于:N439R,L452K,T470N,E484P,Q498Y,N501T。例如,改变抗体亲和力的取代包括但不限于:A372T,S373F,T393S,I402V,S438T,N439R,L441I,S443A,G446T,K452K,L455Y,F456L,S459G,T470N,E471V,Y473F,Q474S,S477G,E484P,F490W,Q493N,S494D,Q498Y,P499T,和N501T。增加ACE2结合亲和力同时降低或可能代替抗体结合和B细胞结合至这些序列的取代可能有助于免疫逃避和免疫逃逸。在某些实施例中,该支架包含一个或多个氨基酸取代,包括N501Y,N501T,E484K,S477N,T478K,L452R,和N439K,序列和不一定来自S蛋白的其他序列片段,与具有增强的致

病性或传播性的选择压力，或有助于抗原漂移/逃逸的活性序列相对。这些肽可以在全球传播之前被快速设计和分布，以在新株出现时覆盖特定区域，这一原理也可以应用于其他病原体(包括细菌，真菌，原生动物等)和病毒。增强ACE2结合的一些示例性致病性变体可能会也可能不会相应地增加感染性，致病性，和抗体逃逸。例如，可以维持N426-F443和Y460-Y491。

[0123] 在某些实施例中，该支架包含具有EPEA的氨基酸序列的His标签或C-标签。

[0124] 本文公开的支架可以是线性肽。替代地，本文公开的支架可以是环状肽，例如，线性肽可以经由酰胺键头对尾环化。头对尾环状支架的一些示例包括具有以下氨基酸序列的支架#43(SEQ ID NO:114)和支架#44(SEQ ID NO:115)(GS接头以粗体显示)：

[0125] SNNLDSKVGGNYNYLYRL**GSGSG**QAGSTPCNGVEGFNCYFPLQSYGFQPT NGVGYQP（SEQ ID NO:114）和

[0126] SNNLDSKVGGNYNYLYRC**GSGSG**QAGSTPGNGVEGFNGYFCLQSYGFQPT NGVGYQP（SEQ ID NO:115）。

[0127] 预计这些环状支架是非聚集和无毒的，并且具有与线性支架相当或更好的结合亲和力。在某些实施例中，可使用替代接头来进一步优化该环状支架。头对尾环状支架的一些示例具有以下氨基酸序列(接头为粗体)：

[0128] SNNLDSKVGGNYNYLYRLF**DG**TEIYQAGSTPCNGVEGFNCYFPLQSYGFQP TNGVGYQP(支架#45;SEQ ID NO:116)；和

[0129] SNNLDSKVGGNYNYLYRLF**PKP**EIYQAGSTPCNGVEGFNCYFPLQSYGFQP TNGVGYQP(支架#53;SEQ ID NO:124)。

[0130] 在某些实施例中，可以向该支架添加额外的氨基酸残基以实现想要的大小或结构。同样地，这些支架可以是线性肽或头对尾环状肽。该支架的一些示例具有以下氨基酸序列(所添加的残基以粗体显示)：

[0131] SNNLDSKVGGNYNYLYRLF**NAND**EIYQAGSTPCNGVEGFNCYFPLQSYGF QPTNGVGYQP(支架#46;SEQ ID NO:117)；

[0132] SNNLDSKVGGNYNYLYRLF**NAHDK**IYQAGSTPCNGVEGFNCYFPLQSYGF QPTNGVGYQP(支架#47;SEQ ID NO:118)；

[0133] SNNLDSKVGGNYNYLYRLF**NANDK**IYQAGSTPCNGVEGFNCYFPLQSYGF QPTNGVGYQP(支架#48;SEQ ID NO:119)；和

[0134] SNNLDSKVGGNYNYLYRLF**DAHDK**IYQAGSTPCNGVEGFNCYFPLQSYGF QPTNGVGYQP(支架#49;SEQ ID NO:120)。

[0135] 在某些实施例中，该支架被修饰以包括含有Pro残基的接头以获得更刚性的结构。这种刚性支架的一些示例具有以下氨基酸序列(含Pro的接头以粗体显示)：

[0136] SNNLDSKVGGNYNYLYRLF**PKP**EQAGSTPCNGVEGFNCYFPLQSYGFQPTN

GVGYQP (支架#50;SEQ ID NO:121)；

[0137] SNNLDSKVGGNYNYLYRLF**PG**TEIYQAGSTPCNGVEGFNCYFPLQSYGFQP TNGVGYQP (支架#51;SEQ ID NO:122)；

[0138] SNNLDSKVGGNYNYLYRLF**PA**TEIYQAGSTPCNGVEGFNCYFPLQSYGFQP TNGVGYQP (支架#52;SEQ ID NO:123)；

[0139] SNNLDSKVGGNYNYLYRLF**PGTD**IYQAGSTPCNGVEGFNCYFPLQSYGFQP TNGVGYQP (支架#54;SEQ ID NO:125);和

[0140] SNNLDSKVGGNYNYLYRLF**PAHDK**IYQAGSTPCNGVEGFNCYFPLQSYGF QPTNGVGYQP (支架#55;SEQ ID NO:126)。

[0141] 图6描绘结合构象中三个代表性接头的三维分子建模。

[0142] 在某些实施例中,该支架包含PEG链,如PEG2000(45个单位)以允许与二聚体ACE2的两个单元的结合。在一些实施例中,该PEG链的长度在30至60个单位之间,35至55个单位之间,或40至50个单位之间。在一些实施例中,该PEG链的长度为约30,约31,约32,约33,约34,约35,约36,约37,约38,约39,约40,约41,约42,约43,约44,约45,约46,约47,约48,约49,约50,约51,约52,约53,约54,约55,约56,约57,约58,约59,或约60个单位。

[0143] 在某些实施例中,该支架包含一个或多个被亲水性氨基酸或聚合序列取代的氨基酸以减少聚集。在某些实施例中,该支架包含修饰以增加亲水性氨基酸的数量。在某些实施例中,该支架被配置有面向内的疏水性氨基酸。在一些实施例中,可以添加PEG,聚(肌氨酸),或亲水性聚合物序列以增加支架溶解度。此类支架的一些示例具有以下氨基酸序列(K取代以粗体显示):

[0144] V**K**AWNSNNLDSKVGGNYNYLYRLGSGSGQAGSTPCNGVEGFNCYFPLQS YGFQPTNGVGYQPYRVV (支架#41;SEQ ID NO:112);和

[0145] VI**K**WNSNNLDSKVGGNYNYLYRLGSGSGQAGSTPCNGVEGFNCYFPLQSY GFQPTNGVGYQPYRVV (支架#42;SEQ ID NO:113)。

[0146] 可以使用合适的接头将本文公开的支架连接至一个或多个额外的支架或其他肽以生成多聚体结构。在某些实施例中,二聚体可通过用接头将两个支架或肽连接在一起而形成。在某些实施例中,三聚体可通过用两个接头将三个支架或肽连接而形成。可生成更大的多聚体结构,例如,包括四个,五个,六个,七个,或八个连接在一起的支架或肽的组装体。接头的示例包括但不限于PEG或聚(肌氨酸)。

[0147] 在某些实施例中,该支架包含在N-端的天然F残基,在C-端的残基IYQ,或两者。在一些实施例中,该支架包含在残基YQP处的头对尾环状肽的闭合。

[0148] 衍生自SARS-CoV-2S蛋白的支架的代表性示例列于下表1。关键结合基序加下划线,免疫表位加粗体和斜体,接头加粗体。这些代表性支架序列的比对在图32A-32B中列出。所获得的支架可以装至ACE2以研究结合亲和力,如图7A-7C所示。

[0149] 如图8A-8B所示,无论是否经过修饰,都可以将该CoV-2支架装至ACE2,以基于计算

机建模确定其ACE2结合亲和力和K_D预测。

[0150] 所选择的支架经受进一步的结构分析和修饰以实现更高的结合亲和力,更好的功效,和/或改进的稳定性。本文公开的支架,无论是否经过修饰,都可以通过任何已有技术获得,例如,通过肽合成或通过重组技术。

[0151] III.支架的体外试验

[0152] 如工作示例所示,进行生物层干涉测量试验以筛选在体外对ACE2具有高结合亲和力并显著抑制CoV-2感染的支架。生物层干涉测量("BLI")是一种在传感器末端表面上负载配体和/或可溶性分析物与该传感器末端表面上的该配体结合后测量入射白光的波长位移的方法。该波长位移对应于存在的分析物的量并且可用于确定多种分析物与所固定的配体周期之间的解离常数和竞争。该波长位移对应于存在的分析物的量并且可用于确定多种分析物与所固定的配体之间的解离常数和竞争。

[0153] 具体而言,该支架与ACE2和SARS-CoV-2中和抗体的相互作用通过生物层干涉测量,以及ACE2-HEK293细胞的假型慢病毒感染来表征。如工作示例所证明,支架#8(SEQ ID NO:79)的剂量低至30nM时观察到感染的统计学显著抑制,在6.66μM范围内具有95%或更大的感染抑制。

[0154] ACE2,通常被称为SARS-CoV-2的病毒进入受体,以膜结合和可溶性形式存在。虽然ACE2在以可溶性形式存在时防止体外感染,但它可能有助于病毒在体内的免疫遮蔽和免疫逃避特性,基本上屏蔽处于其开放构象的刺突蛋白使其不被适应性免疫系统识别。如工作示例所示,ACE2浓度低至4nM时观察到对SARS-CoV-2假型慢病毒感染的统计学显著抑制。然而,工作示例进一步表明,可溶性ACE2阻止了中和抗体与刺突蛋白的受体结合结构域(RBD)结合。

[0155] 在心力衰竭患者中,可溶性ACE2的血浆浓度为16.6-41.1ng/mL(第一和第四四分位数范围),对应于大约193-478pM,而一些研究报告称急性心力衰竭患者中的浓度为7.9ng/mL,健康志愿者中为4.8ng/mL,分别对应于大约92pM和大约56pM(4,6)。

[0156] 其他研究报告称患有1型糖尿病(大约27.0ng/mL)并存糖尿病肾病(大约25.6ng/mL)和/或冠心病(大约35.5ng/mL)的男性和女性患者的循环ACE2浓度高于男性对照组(大约27.0ng/mL),动脉硬度较高和微血管或大血管疾病与可溶性ACE2浓度呈正相关(14)。在这样的范围内,ACE2可能会由于在开放构象中闭塞S1刺突蛋白的受体结合结构域而增强体内感染,鉴于单个病毒刺突仅在暴露于弗林蛋白酶(在生物合成期间)和TMPRSS2(在膜结合期间)后才呈现这种"开放"构象(7;17)。另外地,心血管,糖尿病,肾脏,和血管疾病患者中较高浓度的ACE2可能进一步与SARS-CoV-2的致病性增加相关联。由于ACE2对SARS-CoV-2受体结合结构域(RBD)表现出极强的结合亲和力,这可能会干扰中和抗体与病毒的结合,因此该病毒可能会因可溶性ACE2而避免被免疫系统检测到。与支气管肺泡灌洗,纤维支气管镜刷活检,痰液,鼻拭子,咽拭子,和粪便相比,临床试样血液中SARS-CoV-2病毒滴度较低(平均降低$2^{4.6}$,而循环阈值为30,对应于<$2.6x10^4$拷贝/mL),对应于血液中大约1000个病毒拷贝/mL(20)。假设每个病毒约有100个刺突,如果所有刺突均处于开放构象,这对应于每毫升血液中大约100,000个可能的ACE2结合位点。然而,鉴于开放构象仅在TMPRSS2裂解后发生,必须假定每个刺突的起始位置是闭合的,并且在任何给定时间点,这些约100,000个位点中可能只有一小部分暴露与ACE2或中和抗体结合。因此,大约193-478pM可溶性ACE2浓度对应

于1.6x10^{14}至2.9x10^{14}分子/mL，当将大约720pM至1.2nM Kd的ACE2与开放构象中的刺突蛋白偶联时，表明SARS-CoV-2主要存在于血液中，其“开放”刺突被ACE2闭塞。预计ACE2与某些SARS-CoV-2RBD突变体结合低至110至130pM Kd，重要的是-当处于完全“开放”构象时-SARS-CoV-2刺突蛋白表现出与竞争同一结合位点的中和抗体相当的结合亲和力(25,26)。当考虑到ACE2阻碍中和抗体与该位点结合的能力，并且中和抗体是B细胞成熟的产物，这尤其令人不安，藉此B细胞必须使抗体和BCR成熟以达到与ACE2刺突结合的强度相当的个位数纳摩尔或皮摩尔的结合亲和力。

[0157] 事实上，根据本文提供的结果，SARS-CoV-2对ACE2的结合亲和力甚至与强效中和抗体相当。如本文所示，ACE2在体外严重消除了抗体与SARS-CoV-2刺突RBD的结合并作为表达ACE2的细胞中SARS-CoV-2假型慢病毒感染的有效抑制剂。总之，这些数据表明ACE2既具有针对感染的保护功能，也具有对病毒免疫识别的抑制功能，充当针对刺突蛋白的中和抗体识别的竞争性抑制剂，结合亲和力范围为约676pM至约33.97nM(1)。

[0158] 如工作示例所示，SARS-CoV-2刺突的受体结合结构域(RBD)以约3nM的亲和力与ACE2结合，而ACE2能够阻止中和抗体与RBD的结合，否则RBD将具有约6nM的结合亲和力。总之，在生理ACE2浓度下，ACE2与“开放”构象刺突蛋白的结合是抑制中和抗体形成和与刺突蛋白RBD结合的可行机制，因此病毒具有多种机制来避免被中和抗体检测到。

[0159] 最近的刺突蛋白突变体D614G相对于原始序列似乎进一步增加了表面上“开放”刺突蛋白的密度，以及一般刺突的密度，这显著地使这种突变体可能相对于含有天冬氨酸(D)的变体对中和抗体更为敏感，同时还增加了感染性(9,23)。事实上，通过SARS-CoV-2假型慢病毒感染试验，D614G变体在表达ACE2的细胞中的感染性似乎增加了>1/2log^{10}(～3x)(11)。

[0160] 随着SARS-CoV-2和COVID19继续肆虐世界，重要的是通过避免中和抗体识别或“开放”构象中刺突蛋白的识别来监测多种突变体的出现和对“免疫遮蔽”的敏感性。

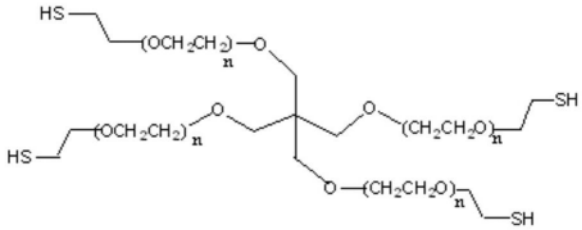
[0161] IV.支架/肽和包含其的组合物的应用

[0162] 本文还公开了包含一个或多个支架的组合物，包含一个或多个支架的缀合物，或包含一个或多个支架的融合蛋白。在一些实施例中，该组合物还包含一种或多种药学上可接受的载体，赋形剂，或稀释剂。在一些实施例中，该组合物可以配制成可注射，可吸入，口服，经鼻，外用，经皮，子宫，润滑的，油状，糖果，小熊软糖，和/或阴道和直肠剂型。在一些实施例中，该组合物通过肠胃外，口服，肺，颊，经鼻，经皮，直肠，阴道，导管，尿道，或眼途径施用于受试者。

[0163] 如本文件所公开，可以通过将包含2-150个氨基酸残基的极性头部或极性尾部添加至N-端或C-端来修饰该支架，例如，包含聚(Arg)，聚(K)，聚(His)，聚(Glu)，或聚(Asp)，以及非天然氨基酸和其他聚合形态，包括糖肽，多糖，线型和支链聚合物等。可适用于本公开的支架的非天然氨基酸和其他聚合形态的示例包括美国临时专利申请号62/889,496中描述的聚合分子，其通过引用并入本文。还可以经由接头将重组膜融合结构域添加至支架。因此，该支架可以组装成静电纳米颗粒。另外地，该支架可以固定在芯片上用于表面等离子共振(SPR)。该支架可用作多种治疗剂的靶向递送配体，如siRNA，基于CRISPR的技术，和小分子，作为合成和天然/重组衍生递送系统，基于基因和蛋白质的有效负载等的一部分。该支架可以是合成的或重组的，并且可以包括接头和对N-端或C-端的合成或重组修饰，以进一步增强膜融合或递送底物融合。任选地，靶向递送可基于纳米颗粒。本领域已知的多个标

签也可以附接至该支架，例如，His-标签和C-标签。

[0164] 在本文件中还公开了该支架可包含允许使用已有肽缀合物技术连接可缀合结构域的环。在一些实施例中，本文公开的支架可经由马来酰亚胺进行缀合，马来酰亚胺通常用于生物缀合，并且与硫醇反应，硫醇是Cys残基侧链中的反应性基团。马来酰亚胺可用于将本文公开的支架附接至任何含SH的表面，如下所示：

[0165] 

[0166] 本文公开的支架可包含一个或多个免疫表位。此外，本文公开的一个或多个支架可经由接头或其他可缀合结构域缀合在一起以获得多表位，多价支架。另外地，本文公开的支架可以附接至其他免疫反应引发结构域或片段。在一些实施例中，本文公开的一个或多个支架可以附接至Fc片段以形成融合蛋白。

[0167] 本文公开的ACE-2支架和CoV-2支架均可在组合物中用作“包被(coating)”以阻断或抑制病毒进入。在一些实施例中，该ACE-2支架可以与CoV-2病毒的RBD结合以包被病毒，从而阻断病毒进入人体。在一些实施例中，该CoV-2支架可以与ACE2结合结构域结合以包被ACE2受体，从而阻断病毒进入人体。

[0168] 本文公开的支架是大小小于100个氨基酸的小肽，例如，约70个氨基酸残基或更少，并且包含：1)T细胞受体MHC-I和MHC-II负载的免疫表位区域，2)B细胞受体或抗体结合的免疫表位区域，和3)ACE2受体结合区域。这些合成或重组支架不仅可以作为ACE2与SARS-CoV-2病毒结合的竞争性抑制剂，而且还可以被设计为触发免疫学习，并能够在多种免疫活性支架和佐剂上呈现。另外地，这些支架可以容易地与多种免疫佐剂以及用于多价展示的已知和新型底物缀合。这些支架也可用于多种感染性疾病的致病因子，包括细菌，真菌，原生动物，变形虫，寄生虫，病毒，性传播疾病等。

[0169] 另外地，所公开的技术允许靶向递送多种治疗剂，包括沉默RNA，CRISPR和其他基于基因编辑的技术，和小分子试剂至病毒感染的细胞。例如，本文公开的支架可用作治疗设计和开发中基于纳米颗粒的siRNA递送和小分子缀合物方法的配体。此外，包含免疫表位的支架还可以通过MHC-I和MHC-II负载呈现抗体和T细胞受体进行免疫表位识别的关键残基，如通过基于SARS-CoV-1与中和抗体的晶体结构数据的整个SARS-CoV-2蛋白最远侧环结构的所预测抗体结合区域所确定，除IEDB免疫表位预测方法之外。

[0170] 由于它们与针对RBD的中和抗体结合，本文公开的支架也有望增强对SARS-CoV-2的免疫反应，而不是将其钝化。相比之下，ACE2仿拟和抗体疗法等方法可能会降低中和抗体对病毒的反应，因为它们包被病毒并阻止适应性免疫系统与所结合部分的结合，这是B细胞受体(BCR)成熟为靶向其“开放”构象中刺突蛋白RBD的中和抗体所必需的刺突蛋白的同一片段。

[0171] 重要的是，本文公开的支架预计不会干扰ACE2的活性，因为其与不代谢血管紧张素II的酶的表面结合。对于疫苗设计和免疫反应促进至关重要，这些肽还被设计为具有MHC-I和MHC-II识别的模块化表位，除了肽序列内包含抗体结合表位之外，还可以根据多种

患者群体的单倍型进行定制。充满希望地是,康复的COVID-19患者对一组保守的表位形成显性CD8+T细胞反应,在6种HLA类型的24名筛查患者中,94%表现出对1个或多个显性表位的T细胞反应,53%的患者表现出对全部3个显性表位的反应(5;27)。此外,先前的研究表明,具有多种HLA基因型的患者对不同SARS-CoV-2表位形成MHC-I介导的反应,这可以通过生物信息学方法进行预测(10)。虽然对应于多种HLA基因型的MHC负载的生物信息学预测并不能预言性地显示哪些肽序列将被负载或不会被负载,但它们确实为经验验证创建了可能的状态空间的全面概述。本文公开的支架被设计为展示用于启动选择性TCR组库(repertoire)的克隆扩增的模块化基序,这可以通过对恢复的患者TCR组库进行测序,插入这些支架,并评估目标群体的HLA基因型来促进(28)。这为快速疫苗和解毒剂设计提供了一种便利的方法,将生物信息学与结构和患者衍生的组学数据相结合,以创建治疗感染因子的迭代设计方法。

[0172] 体外研究为抗体识别和有效病毒阻断提供了原理证明。支架的合成性质提供了经由点击化学将这些肽连系至多种底物的效用,其中包括但不限于C60巴克敏斯特富勒烯,单壁和多壁碳纳米管,树枝状大分子,传统疫苗底物,如KLH,OVA,和BSA等-尽管在本示例中检验了末端为炔烃的裸肽。这些肽的合成性质,计算机模拟筛选,和精确构象允许快速合成,而不受重组,减毒活,基因递送系统,病毒载体,或灭活病毒疫苗方法的传统限制。由于这些肽的点击化学性质,它们还可以通过静电序列修饰或通过点击化学或膜融合到脂质颗粒上作为药物和基因递送载体。本文提供的包含本文公开的支架的组合物和这些肽的未来排列可用于有助于针对多种感染性因子的精确治疗剂和疫苗的设计,开发,和扩大,作为更广泛的生物防御计划的一部分。肽不必是合成的,并且可任选地包含重组变体或重组蛋白与选自合成肽,聚合物,类肽,糖蛋白,多糖,脂肽,和脂糖的部分之间的融合体。

[0173] 本文公开的支架旨在克服与抗体疗法,ACE2-Fc疗法,和其他抗病毒疗法相关联的许多限制。尽管中和抗体可用作“权宜之计”疗法来预防疾病的进展,但所施用抗体的短暂性使有机体易于再次感染。此外,如本示例所示,ACE2是中和抗体与SARS-CoV-2刺突蛋白受体结合结构域结合的有效的抑制剂。仿拟ACE2并屏蔽这一关键表位的疗法可能会使抗体形成偏向脱靶(off-target)位点,这可能导致抗体依赖性增强(ADE),与疫苗相关联的增强型呼吸性疾病(VAERD),和反复病毒攻击后的一系列其他免疫问题。这些关键问题在疫苗开发中也很重要,因为在接种SARS-CoV-1的动物中有增强型呼吸性疾病的先例(29)。对于SARS-CoV-1,在感染6年后的患者中观察到外周记忆B细胞反应明显缺乏(30)。因此,任何促进特异性和中和免疫反应的方法,无论是独立的还是与另一种疫苗方法或感染相结合,都应被视为免疫抑制剂和潜在的脱靶抗体形成方法的替代方法。

[0174] 特别是,任何有可能限制内源性抗体形成的方法都应仔细重新考虑,因为病毒免疫逃避技术跨越了一系列机制,包括但不限于刺突蛋白在“开放”和“封闭”构象之间转换,重糖基化限制可及区域,以及由于被感染细胞的MHC下调和SARS-COV-2刺突蛋白的潜在MHC-II结合限制CD4+T细胞反应而导致的T细胞逃避的表现,这些都可能是导致被感染患者T细胞耗竭和无效和/或短暂的抗体和记忆B细胞反应的因素。理想的治疗策略应该增强中和抗体的形成,而不是将其钝化,同时还可以防止病毒进入细胞和复制(31,32,33,34)。事实上,重症和危重病人表现出极端的B细胞活化,可能还有抗体反应。然而,临床结果不佳,这表明免疫逃避和/或脱靶抗体形成占主导地位(35;36)。

[0175] 对多种因素单独在何种程度上促成这一现象仍然知之甚少。当然,COVID19本身就是一种多因素疾病,具有一连串的有害影响。此外,不同疾病严重程度的队列再感染的可能性仍有待充分阐明,尽管大量临床和轶事报告表明对冠状病毒的免疫力明显短暂,经常观察到α和β冠状病毒再感染易感性存在季节性变化,一些抗体反应持续时间不超过3个月(37)。

[0176] 特别是对于SARS-CoV-2,出现中度抗体反应的患者在短短50天内就会出现无法检测到的抗体(38)。另外地,一项针对149名康复者的研究报告称,33%的研究参与者在症状出现后39天未产生可检测到的中和抗体,并且该队列的大多数的中和抗体活性不高(39)。

[0177] 重要的是,由于存在抗体形成的关键表位和在这些实验中展现一个MHC-I表位和一个MHC-II表位的支架#8(SEQ ID NO:79)的性能,本文提供的结果表明本文公开的支架可用作治疗剂和疫苗。MHC-I和MHC-II结构域可被灵活地取代以匹配不同群体中的HLA类型,或在展现多个结构域的肽组中汇集。因为所公开的支架仿拟病毒,而不是与病毒结合,而且由于其能够代替ACE2以遮蔽病毒刺突蛋白,所以本文提供的组合物可被证明为被感染患者的有效免疫增强策略,具有作为预防性疫苗的额外的潜力。

[0178] 因此,本文公开的支架和包含其的组合物可用于防止病毒与ACE2结合和感染,同时还有助于降低可溶性ACE2对病毒的屏蔽。相对于本文提供的支架(约1μM K_D),抗体与病毒(本文研究的与中和抗体为约6nM K_D)的热力学有利相互作用表明支架可以解离ACE2,在感染期间促进针对病毒的抗体形成,并且优先训练免疫系统以消除病毒。

[0179] 示例1.在没有结构数据的情况下模拟和对接SARS-CoV-2刺突(S)蛋白

[0180] 此工作示例演示了在没有确定SARS-CoV-2的原子分辨率结构的晶体学或cryo-EM数据以及没有任何关于CoV-2病毒与ACE2受体的结合裂隙的数据的情况下,使用SWISS-MODEL基于SARS-CoV-2刺突蛋白序列(UniProt ID P0DTC2)及其与SARS-CoV-1(PDB ID 6CS2)的同源性构建新型病毒SARS-CoV-2的结构模拟。

[0181] 为了阐明CoV-2受体结合结构域(RBD)的结合基序,在没有结构数据的情况下,依赖于先前SARS-CoV-1与ACE2的晶体学实验的结果。在2020年2月Cryo-EM或X射线晶体学数据可用之前,SWISS-MODEL被用于生成SARS-CoV-2刺突蛋白结构(20-3)。

[0182] 使用PyMOL(PyMOL分子图形系统,版本2.3.5,薛定谔公司)将SARS-CoV-2刺突蛋白结构和与ACE2结合的SARS-CoV-1刺突蛋白(PDB ID 6CS2)比对。然后通过PDBePISA运行该结构以确定吉布斯自由能(ΔG)并预测SARS-CoV-2刺突蛋白与ACE2受体之间的氨基酸相互作用(10)。PyMOL还用于将处于其天然构象SARS-CoV-1的截短序列(位置322-515)与SARS-CoV-2S蛋白的ACE2受体(位置336-531)比对,并使用PDBePISA对SARS-CoV-2S蛋白与ACE2的模拟结合袋进行热力学ΔG计算。根据结构数据的可用性,比较并确定该方法已正确识别出与ACE2结合所必需的氨基酸片段,如示例4所详述。

[0183] 如图1所示,将SARS冠状病毒(“SARS-CoV”;“SARS-CoV-1”;或“CoV-1”)蛋白序列(PDB ID 6CS2)与SARS-CoV-2S蛋白序列(以下简称“CoV-2”;SEQ ID NO:2,由SEQ ID NO:1的核苷酸21536-25357编码)进行比较,并使用SWISS-MODEL生成同源模型,然后将其作为PDB文件导入PyMOL。

[0184] 将CoV-2的A链与SARS-CoV-1(PDB ID 6CS2)的A链在ACE2受体的结合状态下比对(参见图9A-9C)。PDB-PISA在CoV-2S蛋白与ACE2受体的结合界面上运行以确定关键结合残

基。图10A和10B分别显示CoV-2和CoV-1结合界面上每个残基的ΔG计算结果。如图11A-11C所示,与ACE2结合的CoV-2S蛋白的关键残基(负ΔG)以绿色突出显示,而具有约0ΔG的残基以黄色显示,具有正ΔG的排斥性残基以橙色显示。基于计算机模型,L455和P491(以洋红色显示)接近,因此可能被Cys残基替代以在这两个位置之间引入二硫键以稳定CoV-2结合界面中的β折叠结构。

[0185] 示例2.仿拟S蛋白受体结合基序的合成肽支架的设计和模拟

[0186] 在模拟SARS-CoV-2 S蛋白RBD的结构和结合之后,设计了包含截短受体结合基序(RBM)的肽支架。该序列被设计为重建该基序中大蛋白的结构,并进行关键修饰以有助于β折叠形成。多种基于深度学习的方法被用来模拟肽支架的结构。例如,基于SUMMIT超级计算机的建模方法可用于模拟假定支架的结合。另外地,可以将PDBePISA和Prodigy组合方法添加至超级计算机的试探法以评估结合裂隙相互作用。建模技术还可以包括CD147-SPIKE相互作用作为组件,使得超级计算机的分子动力学模拟可以在没有预偏置比对的情况下进行预测。此外,使用或不使用超级计算机的建模可以与使用RaptorX(或AlphaFold或等效物)相结合,以对随机肽序列的自由能折叠状态进行建模。这允许使用超级计算机对随机肽序列进行组合筛选,然后在目标受体上运行分子动力学模拟。这些折叠技术与同源性建模有独特的区别,其不考虑肽的自由能和具有与较大蛋白的自由能不同的较小截短蛋白片段的可能折叠状态的全域(gamut)。本文提供的方法允许从头生成肽序列,然后模拟它们的折叠和结合状态。

[0187] 为了说明,使用RaptorX模拟有或无修饰的肽支架,RaptorX是一种高效且准确的蛋白结构预测软件包,基于强大的深度学习技术(19)。给定一个序列,RaptorX用于运行同源搜索工具HHblits以查找其序列同源物并建立多序列比对(MSA),然后导出序列谱(sequence profile)和残基间协同进化信息(13)。之后,RaptorX用于将序列谱和协同进化信息馈送到一个非常深的卷积残差神经网络(约100个卷积层)以预测原子间距离(即Ca-Ca,Cb-Cb,和N-O距离)和预测下蛋白质的残基间取向分布。为了预测原子间距离分布,RaptorX将两个原子之间的欧几里得距离离散化为47个区间:0-2,2-2.4,2.4-2.8,2.8-3.2.....19.6-20,和>20A。为了预测残基间取向分布,RaptorX将先前定义的取向角度(13)离散化为10个度的分箱(bin)。最后,RaptorX从所预测分布导出距离和取向的可能性,并通过最小化可能性来构建蛋白质的3D模型。实验验证表明,这种深度学习技术能够预测比以往更多的蛋白质的正确折叠,并且优于比较建模,除非预测下的蛋白质在蛋白质数据库(PDB)中具有非常接近的同源物。

[0188] 用RaptorX分析本文公开的支架以获得它们可能的折叠状态。图12A-12J示出具有以下氨基酸序列的支架#1的折叠可能性(PyMOL中显示的center0到center9构象):

[0189] VIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV(SEQ ID NO:72)。

[0190] 使用PyMOL比对命令,将处于其10种可能折叠状态的每个支架与和ACE2对接的CoV-2RBD重叠以模仿结合。图13A-13D示出支架#9与ACE2结合的多个构象。支架#9的氨基酸序列为:

[0191] EEVIAWNSNNLDSKVGGNYNYLYRCGSGSGQAGSTPGNGVEGFNGYFCLQSYGFQPTNGVGYQPYRVVRRR(SEQ ID NO:80)。

[0192] 示例3.仿拟ACE2结合结构域的合成肽支架的设计和模拟

[0193] ACE2的结合界面以与示例2公开的类似方式进行研究。来自ACE2(SEQ ID NO:140)位置19-84的一段氨基酸序列似乎与结合至CoV-2S蛋白有关。关键结合残基包括S19,Q24,D38,Q42,E75,Q76,和Y83,在图14A-14D中以绿色显示。基于该分析,合成了具有以下氨基酸序列的ACE2支架1 STIEEQAKTFLDKFNHEAEDLFYQ*GSGSG*NAGDKWSAFLKEQSTLAQMYP (SEQ ID NO:141)(GS接头加斜体并加下划线)。ACE2支架1似乎具有两个关键结合基序:STIEEQAKTFLDKFNHEAEDLFYQ(SEQ ID NO:149)(位置19-42)和NAGDKWSAFLKEQSTLAQMYP(SEQ ID NO:150)(位置64-84)。图15A显示从ACE2蛋白截短的ACE2支架1的计算机建模,图15B-15C显示ACE2支架1与CoV-2S蛋白结合的模拟。

[0194] 示例4.Cryo-EM结构与所模拟结构的比较

[0195] 将示例1公开的计算机模拟的结合模型与他人确定和公布的CoV-2S蛋白的实际Cryo-EM溶液结构(例如,22;华盛顿大学Veesler实验室(faculty.washington.edu/dveesler/publications/)进行比较(图16A-16B)。发现结合界面在涉及与ACE2结合的一段氨基酸残基方面基本一致,但在精确的关键氨基酸方面存在一些差异。令人惊讶的是,他人公布的cryo-EM结构缺少这些通过计算机建模识别的关键结合残基(图17A-17B)。所公布的结构不包含残基444-502,因此,缺少从位置437至453和从位置473至507的关键结合基序。

[0196] 此示例表明,基于同源结合支架的蛋白质结合界面的模拟是快速设计结合支架,抑制剂,和帮助药物发现的有效手段。

[0197] 示例5.产生CoV-2支架

[0198] 本示例说明了CoV-2支架的设计和产生。

[0199] SARS-CoV-2 S蛋白的模拟及其ACE2结合区域的确定。SWISS-MODEL用于创建CoV-2病毒与SARS-CoV-1(PDB ID 6CS2)相比较的结构模拟。接下来,PyMOL用于将处于其天然构象的SARS-CoV-1的截短序列(位置322-515)与SARS-CoV-2的ACE2受体(位置336-531)进行比对。

[0200] 绘制最小界面序列。使用PDBePISA对SARS-CoV-2S蛋白与ACE2的模拟结合袋进行热力学ΔG计算以确定与ACE2结合的CoV-2支架和该支架中的关键结合残基。

[0201] 绘制免疫表位。将刺突糖蛋白的整个序列以及先前定义的一段SARS-CoV-1免疫原性位点与SARS-CoV-2上的类似位点进行了比较。IEDB用于推荐2.22抗原性评分以确定SARS-CoV-2上的同源位点是否具有免疫原性。

[0202] 模拟截短的CoV-2支架。SWISS-MODEL和额外的深度学习驱动的蛋白质模拟方法用于对新型支架进行结构模拟。对支架进行了多种修饰,如添加接头,替代SARS-CoV-2糖蛋白最近侧ACE2界面片段的非ACE2结合区域和非抗体结合区域以整合抗体结合区域。这些结构域可为可变的,并且可以并行进行,以涵盖对已知和预测的免疫原性位点的整体筛选。

[0203] 肽合成。肽支架序列由内部设计和合成或由第三方商业供应商如sb-PEPTIDE(法国)定制合成。质谱法用于确认合适的肽分子量。

[0204] 在内部制造靶向配体的情况下,方法和材料如下。使用定制的肽机器人,使用基于标准Fmoc的固相肽合成(SPPS)合成肽,演示了9个氨基酸序列的每个氨基酸偶联约需要120秒。以前,只需2小时即可合成30-50个氨基酸肽(图18)。合成可以通过任何合适的方式进

行。例如,在肽机器人的替代物中,酵母可用于合成蛋白质。肽在Rink-酰胺AM树脂上合成。用O-(1H-6-氯苯并三氮唑-1-基)-1,1,3,3-四甲基脲六氟磷酸酯(HCTU)偶联剂和N-甲基吗啡啉(NMM)在二甲基甲酰胺(DMF)中进行氨基酸偶联。用三氟乙酸(TFA),三异丙基硅烷(TIPS),和水对肽进行去保护和裂解。通过反相HPLC(RP-HPLC)纯化粗肽混合物。将纯肽级分冷冻并冻干以产生纯化的肽。

[0205] 示例6.CoV-2支架的环化

[0206] 此示例说明CoV-2支架的头对尾环化的多种策略,包括:(1)通过酰胺偶联在溶液中对侧链保护的肽进行头对尾环化(图19A),(2)通过酰胺偶联进行树脂上的头对尾环化(图19B),和(3)通过NCL进行纯化的线性硫酯肽的环化(图19C)。对于策略(1),以整体去保护肽的HPLC纯度为～30％完成合成。对于策略(2),合成完成,在去烯丙基化和整体去保护后,HPLC纯度为～25％。对于策略(3),以～20％的粗产物带有O-烯丙基保护基团完成微波合成。在树脂上去烯丙基化后,尝试用PyBOP/DIPEA环化16小时。没有观察到想要的产物质量,但硫代酯化将是下一步。

[0207] 示例7.CoV-2支架的生物层干涉测量

[0208] 生物层干涉测量(“BLI”)直接询问两种或多种分析物之间的结合。此示例演示了使用BLI对CoV-2支架进行体外分析以通过确定与二聚体ACE2结合的支架的解离常数来表征结合动力学,和支架对ACE2与受体结合结构域(“RBD”)结合的抑制作用。BLI还用于确定支架与IgG中和抗体(NAb)结合的解离常数。

[0209] Octet®RED384生物层干涉仪(颇尔富迪(Fortebio))与传感器末端一起使用,在96孔板中显示抗人IgGFc(ACH),链霉亲和素(SA),镍-次氮基三乙酸(NTA),或抗五his(HIS1K)。对于链霉亲和素末端,在用给定的所固定配体饱和后,使用1mM生物素阻断表面。在用带有His标签的ACE2和RBD的变体与带有生物素标签的ACE2和RBD的变体进行方案优化后,溶液中的支架分析物表现出与具有NTA和HIS1K末端的传感器末端表面的非特异性结合,而生物素化表面最大限度地减少了这种非特异性结合。此外,ACE2-His(义翘神州(Sino Biological))和RBD-his(义翘神州)表现出与HIS1K末端的极弱结合。因此,在SA末端使用二聚体ACE 2-生物素(UCSF)和RBD生物素(USSF),所有研究均在AHC末端使用针对SARS-CoV-2刺突糖蛋白的中和单克隆IgG抗体(CR3022,抗体在线(antibodies-online))。仍然观察到支架#8(SEQ ID NO:79)与AHC末端上的中和抗体结合的非特异性结合,这使得使用支架#8作为分析物测定K_D的工作与中和抗体配体相比更加复杂。所有储备溶液均在含有0.2％BSA和0.02％Tween20的1X PBS中制备。研究了以下配体和分析物:

[0210] 1)将二聚体ACE2-生物素固定在SA末端上(～2.5nm捕获)。

[0211] a.以1,3,和10μM浓度将支架#4(“肽1”,SEQ ID NO:75),支架#7(“肽4”,SEQ ID NO:78),支架#8(“肽5”,SEQ ID NO:79),和支架#9(“肽6”,SEQ ID NO:80)引入所固定的ACE2(图20A至20D)。

[0212] b.从肽溶液中取出传感器末端并将其引入35μM RBD-His(义翘神州)(图20E至20H)。

[0213] 2)将RBD-生物素固定在SA末端上(～5nm捕获)。

[0214] a.以1.3,3.9,11.7,35,和105μM浓度将ACE2-His(义翘神州)引入所固定的RBD(图20I)。

[0215] 3)将中和IgG抗体固定在AHC末端上(～1nm捕获)。

[0216] a.以0.37,1.1,3.33,和10μM浓度将支架#4,支架#7,支架#8,和支架#9引入所固定的ACE2(图20A至20D)。

[0217] b.以1,3,9,27,和81μM浓度将RBD-His(义翘神州)引入所固定的中和抗体(CR3022,抗体在线)。

[0218] c.117nMRBD-His(义翘神州)与ACE2-His在0(仅RBD),2.88,8.63,25.9,和77.7μM浓度下混合。接下来,引入了所固定的中和抗体(CR3022,抗体在线)。

[0219] BLI用于确定与二聚体ACE2结合的所选择的支架的解离常数,以及支架对ACE2与RBD结合的抑制作用。如图20A-20I所示,本实验中测试的CoV-2支架以浓度依赖方式阻止ACE2与S蛋白RBD的结合。所有测试的支架在10μM浓度下均表现出对RBD与ACE2结合的有效抑制。支架在1,3,和10μM浓度下与ACE2结合直至饱和(图20A-20D)。在与ACE2结合后,在没有支架的情况下测量了35μM处SARS-CoV-2RBD与ACE2的结合(图20E-20H),即使在盐水+FBS洗涤后,支架也显示出强拮抗剂的作用。有趣的是,在1μM和3μM下,ACE2与支架#8的结合增强了RBD结合,而10μM浓度则强烈地消除了结合(图20G)。所有其他肽在阻止RBD结合方面表现出剂量反应样行为,包括在1μM和3μM浓度下(图20E,20G,和20H)。为了评估竞争性不可逆拮抗作用,支架不包括在35μM RBD的最终溶液中,如图20I所示。

[0220] BLI还用于测定与IgG中和抗体结合的所选择的支架的解离常数。支架#8表现出与传感器末端的非特异性结合(图21C),阻止了针对中和抗体的K_D的测定。在所有未使用生物素阻断传感器表面的生物素化底物的研究中都观察到了支架#8的这种非特异性结合。然而,用中和抗体测定了所有其他支架的单微摩尔结合亲和力(图21A,21B,和21D)。接下来,测量了用抗RBD中和抗体增加RBD浓度的解离常数(图21E)。为了检查ACE2对中和抗体与RBD结合的抑制作用,在引入所固定的中和抗体之前,将117nM RBD与浓度增加的ACE2混合(图21F)。当RBD浓度为117nM时,抑制RBD与中和抗体之间相互作用的ACE2的半数最大抑制浓度(IC50)的插值为ACE2约30-35nM(图21F)。这些数据表明,ACE2比中和抗体更有效结合RBD,而且即使在SARS-CoV-2刺突RBD的分数摩尔浓度下,可溶性ACE2也可以充当有效的“斗篷”对抗中和抗体识别。本实验中测试的所有支架均具有免疫原性。

[0221] 使用上图所示的BLI数据,确定了1)支架#4,支架#7,支架#8,和支架#9与ACE2结合,2)ACE2和中和抗体与RBD的结合,3)支架与中和抗体的结合,和4)在ACE2浓度增加的情况下RBD与中和抗体的结合的解离常数(K_D)和RMax值(稳态结合分析)。K_D和RMax值列于下表2。

[0222] 表2

[0223]

结合配偶体	RMax	K_D
支架#4与ACE2的结合	0.12111318±0.0239083	1.80E-06±1.1E-06M
支架#7与ACE2的结合	0.22716674±0.0339753	5.20E-06±1.7E-06M
支架#8与ACE2的结合	0.57363623±0.1333544	4.00E-06±2.2E-06M
支架#9与ACE2的结合	0.13006174±0.032393	2.40E-06±1.7E-06M
支架#4与NAb的结合	0.71962807±0.0759471	4.30E-06±1.0E-06M
支架#7与NAb的结合	0.22716674±0.0339753	5.20E-06±1.7E-06M
支架#8与NAb的结合	∞	N/A

支架#9与NAb的结合	1.18656192±0.0552815	3.30E-06±3.9E-07M
ACE2与RBD的结合	0.57847430±0.0155693	2.30E-09±3.0E-10M
NAb与RBD的结合	4.40220793±0.159029	8.60E-09±1.1E-09M

[0224] 表2中提供的数据表明支架#4,7,8,和9与ACE2的解离常数分别为1.8±1.1μM(支架#4),5.2±1.7μM(支架#7),2.4±1.7μM(支架#8),和2.4±1.7μM(支架#9),与中和抗体的解离常数分别为4.3±1.0μM(支架#4),5.2±1.7μM(支架#7),未知(支架#8),和3.3±1.19μM(支架#9)。由于与传感器末端的非特异性相互作用引起的技术错误,支架#8与中和抗体的结合未被测定。ACE2与RBD的解离常数为2.3±0.3nM,而中和抗体与RBD的解离常数为8.6±1.1nM。这些数据表明,支架#4对中和抗体和ACE2都表现出最强的亲和力。

[0225] 示例8.SARS-CoV-2刺突蛋白假型慢病毒感染ACE2-HEK293细胞

[0226] 用显示SARS-CoV-2刺突糖蛋白(BPS生物科学)的假型慢病毒转导ACE2-HEK293细胞(BPS生物科学),并在感染后60小时评估了荧光素酶活性和台盼蓝毒性。抗SARS-CoV-2刺突糖蛋白(CR3022,抗体在线),ACE2(义翘神州),刺突糖蛋白(义翘神州)的受体结合结构域(RBD)的中和单克隆IgG抗体和本公开的所选择的支架被用作感染抑制剂。经由生物发光定量感染,并利用Synergy™ H1 BioTek分光光度计经由台盼蓝吸光度试验表征毒性。

[0227] 如图22所示,通过感染后60小时的荧光素酶活性评估,支架#4和支架#7在浓度低于20μM的情况下没有阻断ACE2-HEK293细胞的SARS-CoV-2刺突蛋白假型病毒感染。然而,支架#8和支架#9在6.66μM时均能阻止病毒感染,而支架#8在纳摩尔范围内显著表现出这种阻断作用(80nM和30nM,$p<0.05$,仅与病毒进行t检验比较)。(*,$p<0.05$;***,$p<0.001$;未配对学生t检验,技术一式三份)。

[0228] 图23A-23D显示可溶性RBD和可溶性ACE2在0.33uM时几乎完全抑制SARS-CoV-2刺突蛋白假型病毒感染,而SARS-CoV-2中和抗体在浓度低至6nM处对感染的抑制程度相似。有趣的是,12nMRBD增强了感染。(*,pv0.05;***,$p<0.001$;未配对学生t检验,技术一式三份)。

[0229] 重要的是,除了20μM剂量的支架#8导致细胞死亡并导致溶液中支架的可见聚集和166nM中和抗体增强细胞存活外,添加不同浓度的支架,可溶性ACE2,可溶性RBD,或SARS-CoV-2中和抗体并未导致病毒存在的情况下细胞存活率发生统计学显著变化,约为50%。

[0230] 因此,本文公开的新型合成肽支架已被证明可阻断病毒与细胞的结合,同时也显示抗体和T细胞受体形成的表位。该实验表明,所测试的支架有效地阻断了>95%的假型慢病毒感染,显示SARS-CoV-2刺突蛋白感染表达ACE2的细胞在EC95剂量下没有毒性,并且所测试的支架即使在极高的RBD浓度(35μM)下也阻止了SARS-CoV-2受体结合结构域(RBD)与ACE2的结合。所测试的支架在亚微摩尔范围内表现出IC50,在30nM时具有统计学显著的病毒抑制。

[0231] 示例9.CoV-2支架在活病毒中的作用

[0232] 在CaCo2细胞的活病毒中测试了支架#4,支架#7,支架#8,和支架#9的抑制作用,然后进行毒性试验。关于所测试的支架对SARS-CoV-2的抗病毒活性见表3,所示为通过对Vero 76细胞的终点稀释而确定的50%细胞培养感染剂量(CCID50),和通过对Caco-2细胞的中性红染料摄取而确定的所测试的支架的百分比毒性。

[0233] 图24A显示所测试的支架在微摩尔浓度下对活病毒中的病毒载量(EC90)表现出超过90%的抑制。图24B显示SARS-CoV-2活病毒对病毒载量的抑制高达2.5log而无毒性。

[0234] 示例10.支架的分子动力学和建模

[0235] 如图25所示,分子动力学建模用于对具有氨基酸序列

[0236] VIAWNSNNLDSKVGGNYNYLYRCFRKSNLKPFERDISTEIYQAGSTPGNGVEGFNGYFCLQSYGFQPTNGVGYQPYRVV(SEQ ID NO:75)的支架#4的折叠进行建模。研究了评分最高的结构在溶液中是否自身稳定,以及其灵活性如何。肽有时可以很快重新折叠;如果起点不在局部最小值附近,则从建模中可以明显看出。肽是否在中短时间范围内稳定是比这是否是整体最小值更容易的问题。

[0237] 定性地,大环(RKSNLKPFERDISTE;SEQ ID NO:128)向上和向下折叠直至其接触中间的β折叠,<1ns;TNG结合环向下折叠并向中间折叠,约20ns;然后结构在模拟的其余部分基本稳定,但仍然非常灵活。结合环非常灵活-它们不断展开和重新折叠,动作为5-6Armsd;并且结构的中间相当刚性,<1Armsd。折叠期间的Rosetta评分如图26所示。这些数据表明,相对于理想结构,损失了约20个单位。Rosetta具有不完全的物理能量函数,该函数针对具有稳定折叠的有序蛋白质进行优化,但不能完美地模拟溶剂效应,从而驱动环折叠。可能没有特别好的特异性结合相互作用,并且分析无序区域存在困难。在运行接近尾声时仔细退火结构可能会找到更好的Rosetta评分。

[0238] 可以在复制物中通过上面公开的步骤运行所有支架结构。可以进行改进以有效地对更长的时间范围进行采样。例如,与普通MD相比,AMD可以提供100x的有效加速,因此甚至可以分析从头开始的折叠或追踪结合路径。

[0239] 在设计肽时,刚性可能是最重要的考虑因素。通过H键或通过共价键(例如,钉合肽)的交联链可以增加准备结合的构象中肽的有效浓度,并降低由于弯曲而导致肽未结合的可能性。可能存在使生物设计更加灵活的强大选择压力,尤其是在表面暴露区域。在后续的设计中考虑了灵活性,例如,通过添加多个脯氨酸,或者确定如何将两个β折叠位制成一个更大的β折叠。一些示例性肽包括支架#4,5,6,7,8,和9(分别为SEQ ID NO:75-80)用于进一步的结构分析和修饰。

[0240] 最初在没有受体结合结构域(RBD)的情况下测试了支架的序列或部分序列,以确定其是否产生预期的结构。可以使用序列

[0241] CKMSECVLGQ SKRVQALLFNKVTLAGFNGYFC(SEQ ID NO:129)进行初始测试,该序列是仅来自支架#8的环,在N-端和C-端具有半胱氨酸残基以确保闭合;KMSECVLGQSKRV**DFC**(SEQ ID NO:70)的部分序列也成环,其通过在C-端添加三个氨基酸残基(加粗体和下划线)而略大于免疫表位。

[0242] 如图27所示,对仅在融合期间暴露的S蛋白的独特表位进行了检验。还检验了将阻止该过程移动至下一步中和的结合位点,并且无限期地暴露一些隐藏的表位(图28)。研究了来自蛋白质数据库(PDB,SARS-CoV-2刺突蛋白的不同构象形状,www.rcsb.org/structure/6XRA)的序列6XRA,它是S蛋白在融合期间的成束构型。KMSECVLGQSKRV(SEQ ID NO:71)的序列与图29A描绘的蛋白质结构相匹配,并在图29B中放大显示。确定这正是图27-28中确定的其中一个结合位点的位置,该位点在预成束阶段位于HR1和HR2之间的铰链中,即图29B中放大的结合位点。因此,预测支架#8在纳摩尔浓度下阻止了与假型病毒的融合/感染,因为其中一些在该位点结合,使用完全独立于ACE2的机制。这一假设得到了支架#8与ACE2的结合并不比其他肽好多少这一认定的支持。但在极低(nM)浓度下的效果明显不同,

这表明了第二种作用机制。第二种作用机制仅对实际的刺突蛋白和实际的病毒起作用。因此,它可能与刺突蛋白结合。这个结合袋被束中的其他两条链包围。任何设法进入该结合袋的肽都可能具有超紧密的结合,可能在个位数纳摩尔或更低的浓度下,它也会完全破坏融合。需要额外的分析来确定刺突蛋白从其原始折叠形式变为结合形式所经历的一系列重排。可能有多种途径,只有其中一些途径可能会在结合期间暂时打开此位点。此外,还有许多其他结合位点,其中6XRA结构的小片段可能与整个结构竞争。每10-20个氨基酸线性片段的计算机模拟或体外筛选可能会发现更多位点。

[0243] 虽然支架#8本身不太可能是最佳的,因为RBD位似乎在这里除了提供体积或空间位阻之外没有做任何事情,这可能会使结合不那么紧密,尽管也会更多地破坏发夹结构,但它可以作为概念证明。最初测试了具有二硫键的KMSECVLGQSKRVDFC(SEQ ID NO:70)的序列并随后进行优化。

[0244] 也利用了如前所述的自催化的遗传编码的环状肽(16)。选择的外显肽插入图30标识的区域,位置1有Cys或Ser残基,这是内含肽剪接所必需的。序列的示例如下:

[0245] HHHHHHGENLYFKLQAMGMIKIATRKYLGKQNVYDIGVERYHNFALKN GFIASN**CAAAAA**CLSYDTEILTVEYGILPIGKIVEKRIECTVYSVDNNGNI YTQPVAQWHDRGEQEVFEYCLEDGCLIRATKDHKFMTVDGQMMPIDEI FERELDLMRVDNLPNGTAANDENYALA（SEQ ID NO:152）。

[0246] 粗体序列是生成的环状肽,其余的自行拼接-6个氨基酸或更长的环,第一个氨基酸可以是Cys或Ser。内含肽-外显肽融合可用作将肽或重组/合成序列与自催化和自拼接序列融合以在两个肽序列之间产生融合的机制。

[0247] 示例11.支架序列的进一步优化

[0248] 基于最高评分的一致性确定了用于设计支架的其他序列,评分是稳定性和结合亲和力的组合,重点强调亲和力。

[0249] 对于充当具有非常高亲和力的“超级结合剂”的肽而言,需要有一个较长的环,该环在ACE2侧突出并与它有更多的接触。目标是在不损害结合亲和力的情况下提高支架本身的稳定性。优选地,通过将相同的结合残基轻推到更好的位置来提高结合亲和力。

[0250] 图31A-31D示出如何筛选和优化肽序列。步骤1:RLxxxxxQA中长度为5的环,约60k次尝试。F在第一位置最普遍。图31A。步骤2:修复第一个F-此时-YQA似乎比-QA更靠近下一个残基,因此,构建了长度为5的环RLFxxxxxYQA,约15k次尝试。非常好,这再现了天然的-**TEI**YQA位;来自该运行的单个最佳序列是RLF**DG**TEIYQA。图31B。步骤3:第二个残基的几何形状与脯氨酸不太相容,但环构建算法在插入它时遇到了麻烦;尝试将脯氨酸固定在那个位置。使用RLFPxxxxYQA构建长度为4的环,约22k次尝试。包括RLF**PG**TEIYQA的环序列得分等于RLF**DG**TEIYQA。包括RLF**PG**T**D**IYQA的环序列也很好。图31C。步骤4:使用RLFxxxxxIYQA构建一个稍长的环,5个残基,约41k次尝试。这次运行没有那么决定性。评分函数在每个位置都有利于D或E。这可能是因为当环面向溶剂时,它们可以与自己的主链形成氢键。此外,它们似乎不一定与不相邻的残基相互作用,但它们仍可能稳定环。这

批次中最好的候选者是RLFNANDKIYQA或RLFNANDEIYQA。图31D。

[0251] 示例12.支架在siRNA递送中的用途

[0252] 使用IDT的沉默RNA设计工具为SARS-CoV-2的包膜蛋白设计了siRNA。包膜蛋白由SEQ ID NO:1的nts 26,191-26,288编码。

[0253] 使用以下序列:13.4正义(SEQ ID NO:143)和13.4反义(SEQ ID NO:144)

[0254] (对应于SEQ ID NO:1的nts 26,200-26,224);13.10正义(SEQ ID NO:145)

[0255] 和13.10反义(SEQ ID NO:146)(对应于SEQ ID NO:1的nts26,235-26,259);

[0256] 和13.5正义(SEQ ID NO:147)和13.5反义(SEQ ID NO:148)(对应于SEQ ID NO:1的nts 26,207-26,231)。图33A-33E示出使用IDT siRNA设计工具进行设计的过程,包括所选择的正义链和反义链的位置和序列。

[0257] 根据先前开发的方法将CoV-2支架与siRNA混合,无论是否经过修饰,或有或没有一个或多个免疫表位,以创建具有a)免疫启动活性和疫苗行为,和b)病毒复制的沉默RNA行为的基因载体。参见美国临时专利申请号62/889,496。这种方法还可用于基因编辑,RNA编辑,和其他基于蛋白质的Cas工具,以治疗多种病毒。

[0258] 示例13.CoV-2支架与ACE2结合的计算机模拟

[0259] 此示例演示支架#4,支架#7,支架#8,和支架#9与ACE2结合的模拟。

[0260] PyMOL中使用了“比对”命令,SARS-CoV-1结合至ACE2(PDB ID 6CS2)以近似于SWISS-MODEL模拟的SARS-CoV-2的结合界面(左);所选择的MHC-I和MHC-II表位区域包含在支架#4中为粉红色,代表S1刺突蛋白中的P807-K835和A1020-Y1047,并通过IEDB免疫表位分析进一步细化。图34。接下来,将图34右侧所示的SARS-CoV-2S1刺突蛋白与ACE2(红色)结合的受体结合结构域(RBD,蓝色和多色)从较大的结构截短。所得的RBD结构通过PDBePISA运行以确定相互作用的残基。在右侧的模型中(图34),绿色残基表示ACE2和S1刺突蛋白RBD之间预测的热力学有利相互作用,而黄色表示预测的热力学中性,橙色表示预测的热力学不利相互作用。青绿色残基表示用于生成SARS-BLOCK™肽的氨基酸(V433-V511)的外边界。虽然预测的结合残基没有与随后经过经验验证的序列完全重叠,但模拟基序中反映的一段氨基酸准确地反映了结合行为,其中所公开的PDBePISA模拟表明N439,Y449,Y453,Q474,G485,N487,Y495,Q498,P499,和Q506是关键的ACE2界面残基。其他诱变研究已确定G446,Y449,Y453,L455,F456,Y473,A475,G476,E484,F486,N487,Y489,F490,Q493,G496,Q498,T500,N501,G502,和Y505对于S425-Y508段内的结合至关重要。(40)。因此,本文提供的残基预测可以被评估为精确并且准确到实际结合行为的几个氨基酸内-并且代表了当使用足够长的氨基酸序列时预测没有结构的结合蛋白段的快速且计算极简的方法。

[0261] 使用PyMOL中的“比对”命令将经由RaptorX模拟的支架与ACE2受体(红色,PDBePISA预测的结合界面为绿色)比对。参见图35,从左至右(上图)显示的是支架#4,支架#7,支架#8,和支架#9。支架#4和支架#7包括具有两个交联基序取代的野生型序列。支架#8包括MHC-I和MHC-II表位,支架#9在其非ACE2界面环区域之一中包括GSGSG接头(白色)。考虑到每个肽产生的所有可能的折叠状态(下图支架#8所示),这些简单的比对命令可以考虑每个肽的多种潜在构象,并且可以作为未来研究的基础,探索更先进的分子动力学方法以松弛和模拟结合界面处的分子内相互作用。许多可能折叠状态的叠加表示可能状态的电子分布云,可以模拟它们的最小界面自由能,与建模缺乏已有结构的从头肽或蛋白质-蛋白质界

面的结合袋相比,所需的计算资源要少得多。

[0262] 从前述可以理解,本发明的具体实施例已在本文出于说明的目的进行描述,但在不偏离本发明的范围的情况下可以进行多种修改。因此,本发明除所附权利要求外不受限制。

[0263] 参考文献

[0264] 1.Chi et al.Humanized Single Domain Antibodies Neutralize SARS-CoV-2by Targeting Spike Receptor Binding Domain.Nat Commun 11(1):4528(2020)

[0265] 2.Choy et al.Synthetic peptide studies on the Severe Acute Respiratory Syndrome (SARS)coronavirus spike glycoprotein:Perspective for SARS vaccine development.Clinical Chemistry50(6):1036-1042(2004)

[0266] 3.Bertoni et al.Modeling protein quaternary structure of homo-and hetero-oligomers beyond binary interactions by homology.Scientific Reports 7:10480(2017)

[0267] 4.Epelman et al.Detection of soluble angiotensin-converting enzyme 2 in heart failure:insights into the endogenous counter-regulatory pathway of the renin-angiotensin-aldosterone system.J Am Coll Cardiol 52(9):750-754(2008)

[0268] 5.Ferretti et al.COVID-19 patients form memory CD8+T cells that recognize a small set of shared immunodominant epitopes in SARS-CoV-2.Immunity 53(5):P1095-1107(2020)

[0269] 6.Hisatake et al.Serum angiotensin-converting enzyme 2 concentration and angiotensin-(1-7)concentration in patients with acute heart failure patients requiring emergency hospitalization.Heart and Vessels 32(3):303-308(2017)

[0270] 7.Hoffmann et al.Amultibasic cleavage site in the spike protein of SARS-CoV-2is essential for infection of human lung cells.Mol Cell 78(4):779-784(2020)

[0271] 8.Krissinel E and Henrick K.Interference of macromolecular assemblies from crystalline state.J Mol Biol 372(3):774-797(2007)

[0272] 9.Mansbach et al.The SARS-CoV-2Spike Variant D614G Favors an Open Conformational State.bioRxiv(preprint)2020.07.26.219741(2020)

[0273] 10.Nguyen et al.Human leukocyte antigen susceptibility map for Severe Acute Respiratory Syndrome coronavirus 2.J Virol94(13):e00510-20(2020)

[0274] 11.Ogawa et al.The D614G muiation inthe SARS-Co V2 Spike pro tein increases infectivity in an ACE2 receptor dependent manner.bioRxiv(preprint)2020.07.21.214932(2020)

[0275] 12.Poh et al.Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralizing antibodies in COVID-19 patients.Nature Commun 11:2806(2020)

[0276] 13.Remmert et al.HHblits:lightning-fast iterative protein sequence searcinng by HMM-HMM alignment.Nature Methods 9:173-175(2012)

[0277] 14.Soro-Paavonen et al.Circulating ACE2 activity is increased in patients with type 1diabetes and vascular complications.J HYPertens 30(2):375-383(2012)

[0278] 15.Studer et al.QMEANDisCo-distance constraints applied on model quality estimation.Bioinformatics 36(6):1765-1771(2020)

[0279] 16.Townend&Tavassoli.Traceless production of cyclic peptide libraries in E.coli.ACS Chemical Biology 11:1624-1630(2016)

[0280] 17.Walls et al.Structure,function,and antigenicity of the SARS-CoV-2 spike glycoprotein.Cell 181(2):281-292(2020)

[0281] 18.Wang et al.Identification of an HLA-A*0201-restricted CD8+T-cellepitope SSp-1of SARS-CoV spike protein.Blood 104(1):200-206(2004)

[0282] 19.Wang et al.Accurate de novo prediction of pro tein contact map by ultra-deeplearning model.PLoS Computational Biology13(1):e1005324(2017)

[0283] 20.Wang et al.Detection of SARS-CoV-2 in different types of clinical specimens.JAMA 323(18):1843-1844(2020)

[0284] 21.Waterhouse et al.SWISS-MODEL:homology modelling of protein structures and complexes.Nucl Acids Res46:W296-W303(2018)

[0285] 22.Wrapp et al.Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.Science 367:1260-1263(2020)

[0286] 23.Zhang et al.The D614G mutation inthe SARS-CoV-2spike protein reduces S1 shedding and increases infectivity.bioRxiv(preprint)2020.06.12.148726(2020)

[0287] 24.Zheng&Song.Novel antibody epitopes dominate the antigenicity of spikeglycoprotein in SARS-CoV-2 compared to SARS-CoV.Cell Mol Immunol 17:538-538(2020)

[0288] 25.Ou et al.Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein(2020)

[0289] 26.Nami et al.The effect ofACE2 inhibitor MLN-4760on the interaction of SARS-CoV-2spike protein with humanACE2:a molecular dynamics study(2020)

[0290] 27.Chour et al.Shared antigen-specific CD8+Tcell Responses against the SARS-CoV-2 spike protein in HLA-A*02:01 COVID-19participants.medRxiv preprint 2020.05.04.20085779(2020)

[0291] 28.Gutierrez et al.Decipheringthe TCR repertoire to solve the COVID-19mystery.Trends in Pharmacological Science 41(8)(2020)

[0292] 29.Tseng et al.Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus.PLoS ONE 7(4):

e35421(2012)

[0293] 30.Tang et al.Lack of peripheral memory B cell responses in recovered patients with severe acute respiratorysy ndrome:a six-year follow-up study.The Journal of Immunology 186(12):7264-7268(2011)

[0294] 31.Zhang et al.The ORF8 Protein of SARS-CoV-2Mediates Immune Evaston through Potently Downregulating MHC-I.bioRxiv preprint 2020.05.24.111823 (2020)

[0295] 32.Diao et al.Reduction and functional exhaustion of Tcells in patients withcoronavirus disease 2019(COVID-19).Frontiers in Immunology 11 (827)(2020)

[0296] 33.Zheng et al.Elevated exhaustion levels and reduced functional diversity of Tcells in peripheral blood may predict severe progression in COVID-19patients.Cellular&molecular immunology 17(5):541-543(2020)

[0297] 34.Kumar et al.An in-silico based clinical insight on the effect of noticeable CD4 conserved residues of SARS-CoV-2on the CD4-MHC-II interactions.bioRxiv preprint 2020.06.19.161802(2020)

[0298] 35.Woodruff et al.Critically ill SARS-CoV-2 patients display lupus-itke hallmarks of extrafollicular B cell activation.medRxiv preprint 2020.04.29.20083717(2020)

[0299] 36.Woodruff et al.Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19.Nature Immunol 21:1506-1516 (2020)

[0300] 37.Kellam&Barclay.The dynamics of humoral immune responses following SARS-CoV-2 infection and the potential for reinfection.Journal of General Virology 101:791-797(2020)

[0301] 38.Seow et al.Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection.medRxiv preprint 2020.07.09.20148429(2020)

[0302] 39.Robbiani et al.Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals.Nature 584(7821):437-442(2020)

[0303] 40.Yi et al.Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies.Cellular&Molecular Immunology 17:621-630(2020)

[0304]

表 1：代表性 SARS-CoV-2 支架

衍生自 SARS-CoV-2 S 蛋白的支架的代表性示例列于下表 1。关键结合基序加下划线，免疫表位加粗体和斜体，接头加粗体，取代加双下划线，带多电荷的 N-和 C-端氨基酸残基加波浪下划线，EPEA C-端标签加斜体。

名称	序列	SEQ ID NO	修饰
支架#1	VIAW***NSNNLD***SKVGGNYNYLYRLFRKS N***LKPFERDI***STEIYQAGSTPCNGVEGFNC YFPLQSY***GFQPT***NGVG***YQPY***RVV	72	对应于野生型 SARS-CoV-2 S 蛋白的残基 433–511。
支架#2	VIAW***NSNNLD***SKVGGNYNYLYRL***KMSE CVLGQSKRV***QA***LLFNKVTLA***GFNCYFPL QSY***GFQPT***NGVG***YQPY***RVV	73	对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代。
支架#3	VIAW***NSNNLD***SKVGGNYNYLYRL***GSGS*** ***GQAGSTPCNGVEGFNCYFPLQSY GFQPT*** NGVG***YQPY***RVV	74	对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。 与支架#41 相同但不存在 I434K 取代。

[0305]

[0306]

			与支架#42 相同但不存在 A435K 取代。
支架#4	VIAW***NSNNLD***SKVGGNYNYLYRCFRKS ***N*LKPFERDI***STEIYQAGSTPGNGVEGFN GYFCLQSY***GFQPT***NGVG***YQPY***RVV	75	对应于 CoV-2 S 蛋白的残基 433–511，但具有 L455C 和 P491C 取代以引入二硫键，和 C480G 和 C488G 取代。 与支架#7 相同，但不存在添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。
支架#5	VIAW***NSNNLD***SKVGGNYNYLYRC***KMSE*** ***CVLGQSKRV***QA***LLFNKVTLA***GFNGYFCL QSY***GFQPT***NGVG***YQPY***RVV	76	对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，L455C 和 P491C 取代以引入二硫键，和 C488G 取代。
支架#6	VIAW***NSNNLD***SKVGGNYNYLYRC***GSGS*** ***GQAGSTPGNGVEGFNGYFCLQSY GFQP TNGVG YQPY***RVV	77	对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS

[0307]

			接头替代，排斥性残基 Y473 缺失，L455C 和 P491C 取代以引入二硫键，和 C480G 和 C488G 取代。 与支架#9 相同，但不存在添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。
支架#7	EEVIAW***NSNNLD***SKVGGNYNYLYRCFR KSN***LKPFERDI***STEIYQAGSTPGNGVEGF NGYFCLQSY***GFQPT***NGVG***YQPY***RVVRR R	78	对应于 CoV-2 S 蛋白的残基 433–511，但具有 L455C 和 P491C 取代以引入二硫键，C480G 和 C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。 与支架#4 相同，但添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。

[0308]

支架#8	EEVIAW***NSNNLD***SKVGGNYNYLYRC***KM*** ***SECVLGQSKRVQALLFNKVTLA***GFNGYF CLQSY***GFQPT***NGVG***YQPY***RVVRRR	79 对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，L455C 和 P491C 取代以引入二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。 与支架#40 相同，但不存在第一插入的免疫表位中的 C 至 S 取代。
支架#9	EEVIAW***NSNNLD***SKVGGNYNYLYRC**GS** **GSGQ**AGSTPGNGVEGFNGYFCLQSY***GF*** ***QPT***NGVG***YQPY***RVVRRR	80 对应于 CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS 接头替代，排斥性残基 Y473 缺失，L455C 和 P491C 取代以引入二硫键，C480G 和 C488G 取代，和添加至 N-端的两个带多电

[0309]

			荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。 与支架#6 相同，但添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。
支架#10	***NSNNLD***SKVGGNYNYLYRLFRKSN***LKP******FERDI***STEIYQAGSTPCNGVEGFNCYFPLQSY***GFQPT***NGVG***YQPY***	81	对应于野生型 SARS-CoV-2 S 蛋白的残基 437–508。
支架#11	***NSNNLD***SKVGGNYNYLYRL**GSGSGS**QAGSTPCNGVEGFNCYFPLQSY***GFQPT***NGVG***YQPY***	82	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。
支架#12	***NSNNLD***SKVGGNYNYLYRL**GSGSG**QAGSTPCNGVEGFNCYFPLQSY***GFQPT***NGVG***YQPY***	83	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。

[0310]

支架#13	84	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。
支架#14	85	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。
支架#15	86	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代。 与支架#2 相同，但不存在 N-端 VIAW 和 C-端 RVV。
支架#16	87	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的大部分主

[0311]

			链区域和第二关键结合基序的部分序列被两个免疫表位替代。
支架#17	***NSNNLD***SKVGGNYNYLYRLFRKS***KMSE*** ***CVLGQSKRV***QA***LLFNKVTLA***GFNCYFPL QSY***GFQPT***NGVG***YQPY***	88	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的大部分主链区域和第二关键结合基序的部分序列被两个免疫表位替代。
支架#18	***NSNNLD***SKVGGNYNYLYRLFRKSN***KMS*** ***ECVLGQSKRV***QA***LLFNKVTLA***GFNCYFP LQSY***GFQPT***NGVG***YQPY***	89	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的大部分主链区域和第二关键结合基序的部分序列被两个免疫表位替代。
支架#19	***NSNNLD***SKVGGNYNYLYRLFRKSNL***KMS*** ***ECVLGQSKRV***QA***LLFNKVTLA***GFNCYFPL QSY***GFQPT***NGVG***YQPY***	90	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的大部分主链区域和第二关键结合基序的部分序列被两个免疫表位替代。

[0312]

支架#20	91	对应于 SARS-CoV-2 S 蛋白的残基 437–508，但两个关键结合基序之间的大部分主链区域和第二关键结合基序的部分序列被两个免疫表位替代。
支架#21	92	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但具有 N439R，L452K，T470N，E484P，和 N501T 取代以增加对 ACE2 和抗体的亲和力。
支架#22	93	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，L455C 和 P491C 取代以引入二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。

[0313]

			与支架#8 相同，但添加第二插入的免疫表位和第二关键结合基序之间的 QA。 与支架#24 相同，但不存在 EPEA C-端标签。
支架#23	EEVIAW***NSNNLD***SKVGGNYNYLYRCFR KSN***LKPFERDI***STEIYQAGSTPCNGVEGF NGYFCLQSY***GFQPT***NGVG***YQPY***RVVRR R	94	对应于 CoV-2 S 蛋白的残基 433–511，但具有 L455C 和 P491C 取代以引入二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。 与支架#7 相同，但不存在 C480G 取代。
支架#24	EEVIAW***NSNNLD***SKVGGNYNYLYRC***KM*** ***SECVLGQSKRV***QA***LLFNKVTLA***QAGFN GYFCLQSY***GFQPT***NGVG***YQPY***RVVRRR *EPEA*	95	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，L455C 和 P491C 取代以引入

[0314]

			二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。 与支架#22 相同，但添加 EPEA C-端标签。
支架#25	EEVIAW***NSNNLD***SKVGGNYNYLYRC***KM******SESVLGQSKRV***QA***LLFNKVTLA***QAGFNG YFCLQSY***GFQPT***NGVG***YQPY***RVVRRR***E*** *PEA*	96	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，第一插入的免疫表位中的 C 至 S 取代，L455C 和 P491C 取代以引入二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。
支架#26	EEVIAW***NSNNLD***SKVGGNYNYLYRL***KM*** ***SECVLGQSKRV***QA***LLFNKVTLA***QAGFN	97	游离自折叠肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域

[0315]

	GYFCLQSY***GFQPT***NGVG***YQPY***RVVRRR *EPEA*		和第二关键结合基序的部分序列被两个免疫表位替代，P491C 取代，C488G 取代，添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基，和 EPEA C-标签。
支架#27	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRLF**GSGSGSGSGSGSGS**YQAGS TPCNGVEGFNSYFPLQSY***GFQPT***NGVG***Y*** ***QPY***RVVRVRFRVRVR*EPEA*	98	游离自折叠肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被 GS 接头替代，C488S 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#28 相同，但不存在 L455C 和 P491C 取代。
支架#28	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRCF**GSGSGSGSGSGSGS**YQAGS	99	游离的二硫键肽。

[0316]

	TPCNGVEGFNSYFCLQSY***GFQPT***NGVG***Y******QPY***RVVVRVRFRVRVR***EPEA***		对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被 GS 接头替代，L455C 和 P491C 取代以引入二硫键，C488S 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#27 相同，但添加 L455C 和 P491C 取代。 与支架#30 相同，但用 GS 接头代替三个插入的 TCR 表位。
支架#29	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRLFKLWAQCVQLYLQPRTFLLLLY DANYFLYQAGSTPCNGVEGFNSYFPLQS ***YGFQPT***NGVG***YQPY***RVVVRVRFRVRVR***E*** ***PEA***	100	游离自折叠肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被三个 TCR 表位（KLWAQCVQL，LLYDANYFL，和 YLQPRTFLL）替代，

[0317]

支架#30	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRCFKLWAQCVQLYLQPRTFLLLLY DANYFLYQAGSTPCNGVEGFNSYFCLQS Y***GFQPT***NGVG***YQPY***RVVRVRFRVRVR*E* *PEA*	101	游离的二硫键肽。 对应于 SARS-CoV-2 S 蛋白的残基 433-511，但两个关键结合基序之间的大部分主链区域被三个 TCR 表位（KLWAQCVQL，LLYDANYFL，和 YLQPRTFLL）替代，L455C 和 P491C 取代以引入二硫键，C488S 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#28 相同，但用三个 TCR 表位代替插入的 GS 接头。
			C488S 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#30 相同，但不存在 L455C 和 P491C 替代。

[0318]

			与支架#29 相同，但添加 L455C 和 P491C 取代。
支架#31	CEVEVEFEVEVIAW***NSRNLD***SKVGGNY NYKYRLFKLWAQCVQLYLQPRTFLLLL YDANYFLYQAGSTPCNGVEGFNSYFPLQ SY***GFQPT***TGVG***YQPY***RRR***EPEA***	102	可缀合肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被三个 TCR 表位（KLWAQCVQL，LLYDANYFL，和 YLQPRTFLL）替代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#32 相同，但用三个 TCR 表位代替插入的 GS 接头。
支架#32	CEVEVEFEVEVIAW***NSRNLD***SKVGGNY NYKYRLF**GSGSGSGSGSGSGSGSYQAG**	103	可缀合肽。

[0319]

	STPCNGVEGFNSYFPLQSY***GFQPT***TGVG ***YQPY***RRR*EPEA*		对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被 GS 接头替代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#31 相同，但用 GS 接头代替三个插入的 TCR 表位。
支架#33	CEVEVEFEVEVIAW***NSRNLD***SKVGGNY NYKYRLFKLWAQCVQLYLQPRTFLLLL YDANYFLNEIYQAGSTPCNGVEGFNSYF PLQSY***GFQPT***TGVG***YQPY***RRR*EPEA*	104	可缀合肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被三个 TCR 表位（KLWAQCVQL，LLYDANYFL，和 YLQPRTFLL）和 N 残基替代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。

[0320]

			与支架#34 相同，但用三个 TCR 表位代替插入的 GS 接头。
支架#34	CEVEVEFEVEVIAW***NSRNLD***SKVGGNY NYKYRLF**GSGSGSGSGSGSGSGS**NEIYQ AGSTPCNGVEGFNSYFPLQSY***GFQPT***TG VG***YQPY***RRR*EPEA*	105	可缀合肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被 GS 接头和 N 残基替代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。 与支架#33 相同，但用 GS 接头代替三个插入的 TCR 表位。
支架#35	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRLF***KMSESVLGQSKRV***QA***LLFNKVT L***AQYQAGSTPCNGVEGFNSYFPLQSY***GF QPT***NGVG***YQPY***RVVRVRFRVRVR*EPEA*	106	游离自折叠肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被两个免疫表位替代，第一插入的

[0321]

			免疫表位中的C至S取代，C488S取代，添加至N-和C-末端的填充残基，和EPEA C-标签。 与支架#36相同，但不存在L455C和P491C取代。
		107	游离的二硫键肽。 对应于SARS-CoV-2 S蛋白的残基433–511，但两个关键结合基序之间的大部分主链区域被两个免疫表位替代，第一插入的免疫表位中的C至S取代，L455C和P491C取代以引入二硫键，C488S取代，添加至N-和C-末端的填充残基，和EPEA C-标签。 与支架#35相同，但添加L455C和P491C取代。
	支架#36	EVEVEFEVEVIAW***NSNNLD***SKVGGNYN YLYRCF***KMSESVLGQSKRV***QA***LLFNKVT*** ***LA***QYQAGSTPCNGVEGFNSYFCLQSY***GF*** ***QPT***NGVG***YQPY***RVVRVRFRVRVR***EPEA***	

[0322]

支架#37	108	可缀合肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被两个免疫表位替代，第一插入的免疫表位中的 C 至 S 取代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。
支架#38	109	可缀合肽。 对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域被两个免疫表位替代，第一插入的免疫表位中的 C 至 S 取代，N439R，L452K，C488S，和 N501T 取代，添加至 N-和 C-末端的填充残基，和 EPEA C-标签。

[0323]

支架#39	110	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS 接头替代，排斥性残基 Y473 缺失，L455C 和 P491C 取代以引入二硫键，C488G 取代，添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基，和 EPEA C-标签。
支架#40	111	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域和第二关键结合基序的部分序列被两个免疫表位替代，第一插入的免疫表位中的 C 至 S 取代，L455C 和 P491C 取代以引入二硫键，C488G 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。

[0324]

			与支架#8 相同，但包含第一插入的免疫表位中的 C 至 S 取代。
支架#41	VKAW***NSNNLD***SKVGGNYNYLYRL**GSG** **SG**QAGSTPCNGVEGFNCYFPLQSY***GFQP*** ***T***NGVG***YQPY***RVV	112	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS 接头替代，排斥性残基 Y473 缺失，和 I434K。 与支架#3 相同但有 I434K 取代。
支架#42	VIKW***NSNNLD***SKVGGNYNYLYRL**GSGS** **G**QAGSTPCNGVEGFNCYFPLQSY***GFQPT*** NGVG***YQPY***RVV	113	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的主链区域被 GS 接头替代，排斥性残基 Y473 缺失，和 A435K。 与支架#3 相同但有 A435K 取代。

[0325]

支架#43	114	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的主链区域被 GS 接头替代和排斥性残基 Y473 缺失。 支架#3，#12，#41，和#42 的截短版本。
支架#44	115	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的主链区域被 GS 接头替代，排斥性残基 Y473 缺失，L455C 和 P491C 取代以引入二硫键，和 C480G 和 C488G 取代。
支架#45	116	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 DGT 替代。

[0326]

支架#46	***SNNLD***SKVGGNYNYLYRLFNANDEIYQ AGSTPCNGVEGFNCYFPLQSY***GFQPT***NG VG***YQP***	117	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 NAND 替代。
支架#47	***SNNLD***SKVGGNYNYLYRLFNAHDKIYQ AGSTPCNGVEGFNCYFPLQSY***GFQPT***NG VG***YQP***	118	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 NAHDK 替代。
支架#48	***SNNLD***SKVGGNYNYLYRLFNANDKIYQA GSTPCNGVEGFNCYFPLQSY***GFQPT***NGVG ***YQP***	119	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 NANDK 替代。
支架#49	***SNNLD***SKVGGNYNYLYRLFDAHDKIYQ AGSTPCNGVEGFNCYFPLQSY***GFQPT***NG VG***YQP***	120	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 DAHDK 替代。

[0327]

支架#50	***SNNLD***SKVGGNYNYLYRLFPKPEQAGS TPCNGVEGFNCYFPLQSY***GFQPT***NGVG***Y*** ***QP***	121	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PKPE 替代和排斥性 Y473 缺失。
支架#51	***SNNLD***SKVGGNYNYLYRLFPGTEIYQAG STPCNGVEGFNCYFPLQSY***GFQPT***NGVG ***YQP***	122	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PGT 替代。
支架#52	***SNNLD***SKVGGNYNYLYRLFPATEIYQAG STPCNGVEGFNCYFPLQSY***GFQPT***NGVG ***YQP***	123	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PAT 替代。
支架#53	***SNNLD***SKVGGNYNYLYRLFPKPEIYQAGS TPCNGVEGFNCYFPLQSY***GFQPT***NGVG***Y*** ***QP***	124	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PKP 替代。

[0328]

支架#54	***SNNLD***SKVGGNYNYLYRLFPGTDIYQA GSTPCNGVEGFNCYFPLQSY***GFQPT***NGV ***GYQP***	125	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PGTD 替代。
支架#55	***SNNLD***SKVGGNYNYLYRLFPAHDKIYQ AGSTPCNGVEGFNCYFPLQSY***GFQPT***NG VG***YQP***	126	对应于 SARS-CoV-2 S 蛋白的残基 438–507，但两个关键结合基序之间的大部分主链区域被 PAHDK 替代。
支架#56	EEVIAW***NSNNLD***SKVGGNYNYLYRLFX XXXXXXXXXXXXXXXXXXXXXXXX XXXXAGFNGYFSLQSY***GFQPT***NGVG***YQ*** ***PY***RVVRRR	127	对应于 SARS-CoV-2 S 蛋白的残基 433–511，但两个关键结合基序之间的大部分主链区域和第二关键结合基序的部分序列被 5–30 个 AA 替代，C488G 和 P491S 取代，和添加至 N-端的两个带多电荷的氨基酸残基和添加至 C-端的三个带多电荷的氨基酸残基。

[0329]

表 3：多个 CoV-2 支架的毒性和抗病毒活性

表 3. 配体化合物对 SARS-CoV-2 的毒性和抗病毒活性

百分比毒性					病毒滴度 - CCID50/mL (Log10)					百分比毒性	病毒滴度 - CCID50/mL (Log10)
浓度 (μg/ml)	支架#4 (SEQ ID NO:75)	支架#7 (SEQ ID NO:78)	支架#8 (SEQ ID NO:79)	支架#9 (SEQ ID NO:80)	支架#4 (SEQ ID NO:75)	支架#7 (SEQ ID NO:78)	支架#8 (SEQ ID NO:79)	支架#9 (SEQ ID NO:80)	浓度 (μg/ml)	M128533	M128533
20	6.3%	21.0%	0.0%	0.0%	4.3	3.0	4.3	2.5	**100**	16.1%	<1.7
6.3	17.7%	11.9%	0.0%	0.0%	4.0	3.7	4.7	4.3	**32**	0.0%	<1.7
2	13.8%	10.8%	0.0%	0.0%	4.7	5.3	5.3	4.7	**10**	0.0%	3.0
0.63	1.9%	0.0%	10.7%	0.8%	4.7	5.3	5.3	4.7	**3.2**	0.0%	4.3
0.2	15.1%	22.2%	6.9%	9.2%	4.7	5.0	5.0	5.0	**1**	8.4%	4.7
0.063	4.9%	13.4%	6.5%	2.8%	5.0	4.7	4.7	5.3	**0.32**	10.7%	5.0
0.02	11.3%	9.8%	13.2%	18.2%	4.5	5.0	5.5	5.3	**0.1**	9.5%	5.0
0.0063	0.0%	7.8%	12.3%	0.0%	4.7	4.7	5.0	5.0	**0.032**	10.0%	4.7
				病毒对照	**4.5**	**4.7**	**4.5**	**5.0**			

通过对 Caco-2 细胞的中性红染料摄取而确定的百分比毒性

通过对 Vero 76 细胞的终点稀释而确定的 50%细胞培养感染剂量（CCID50）

序列表

<110> 利甘达尔股份有限公司

<120> 仿生病毒肽的鉴定及其用途

<130> 134554.8009.WO00

<150> US 62/981,453

<151> 2020-02-25

<150> US 62/002,249

<151> 2020-03-30

<150> US 62/706,225

<151> 2020-08-05

<150> US 62/091,291

<151> 2020-10-13

<160> 152

<170> PatentIn version 3.5

<210> 1

<211> 29848

<212> DNA

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 1

aaaggtttat accttcccag gtaacaaacc aaccaacttt cgatctcttg tagatctgtt 60
ctctaaacga actttaaaat ctgtgtggct gtcactcggc tgcatgctta gtgcactcac 120
gcagtataat taataactaa ttactgtcgt tgacaggaca cgagtaactc gtctatcttc 180
tgcaggctgc ttacggtttc gtccgtgttg cagccgatca tcagcacatc taggtttcgt 240
ccgggtgtga ccgaaaggta agatggagag ccttgtccct ggtttcaacg agaaaacaca 300
cgtccaactc agtttgcctg ttttacaggt tcgcgacgtg ctcgtacgtg gcttttcaga 360
ggcacgtcaa catcttaaag atggcacttg tggcttagta gaagttgaaa aaggcgtttt 420
gcctcaactt gaacagccct atgtgttcat caaacgttcg gatgctcgaa ctgcacctca 480
tggtcatgtt atggttgagc tggtagcaga actcgaaggc attcagtacg gtcgtagtgg 540
tgagacactt ggtgtccttg tccctcatgt gggcgaaata ccagtggctt accgcaaggt 600
tcttcttcgt aagaacggta ataaaggagc tggtggccat agttacggcg ccgatctaaa 660
gtcatttgac ttaggcgacg agcttggcac tgatccttat gaagattttc aagaaaactg 720
gaacactaaa catagcagtg gtgttacccg tgaactcatg cgtgagctta acggaggggc 780
atacactcgc tatgtcgata acaacttctg tggccctgat ggctaccctc ttgagtgcat 840
taaagacctt ctagcacgtg ctggtaaagc ttcatgcact ttgtccgaac aactggactt 900
tattgacact aagaggggtg tatactgctg ccgtgaacat gagcatgaaa ttgcttggta 960
cacggaacgt tctgaaaaga gctatgaatt gcagacacct tttgaaatta aattggcaaa 1020
gaaatttgac accttcaatg gggaatgtcc aaattttgta tttcccttaa attccataat 1080
caagactatt caaccaaggg ttgaaaagaa aaagcttgat ggctttatgg gtagaattcg 1140

atctgtctat ccagttgcgt caccaaatga atgcaaccaa atgtgccttt caactctcat 1200
gaagtgtgat cattgtggtg aaacttcatg gcagacgggc gattttgtta aagccacttg 1260
cgaattttgt ggcactgaga atttgactaa agaaggtgcc actacttgtg gttacttacc 1320
ccaaaatgct gttgttaaaa tttattgtcc agcatgtcac aattcagaag taggacctga 1380
gcatagtctt gccgaatacc ataatgaatc tggcttgaaa accattcttc gtaagggtgg 1440
tcgcactatt gcctttggag gctgtgtgtt ctcttatgtt ggttgccata acaagtgtgc 1500
ctattgggtt ccacgtgcta gcgctaacat aggttgtaac catacaggtg ttgttggaga 1560
aggttccgaa ggtcttaatg acaaccttct tgaaatactc caaaaagaga aagtcaacat 1620
caatattgtt ggtgacttta aacttaatga agagatcgcc attattttgg catcttttc 1680
tgcttccaca agtgcttttg tggaaactgt gaaaggtttg gattataaag cattcaaaca 1740
aattgttgaa tcctgtggta attttaaagt tacaaaagga aaagctaaaa aaggtgcctg 1800
gaatattggt gaacagaaat caatactgag tcctctttat gcatttgcat cagaggctgc 1860
tcgtgttgta cgatcaattt tctctcgcac tcttgaaact gctcaaaatt ctgtgcgtgt 1920
tttacagaag gccgctataa caatactaga tggaatttca cagtattcac tgagactcat 1980
tgatgctatg atgttcacat ctgatttggc tactaacaat ctagttgtaa tggcctacat 2040
tacaggtggt gttgttcagt tgacttcgca gtggctaact aacatctttg gcactgttta 2100
tgaaaaactc aaacccgtcc ttgattggct tgaagagaag tttaaggaag gtgtagagtt 2160
tcttagagac ggttgggaaa ttgttaaatt tatctcaacc tgtgcttgtg aaattgtcgg 2220
tggacaaatt gtcacctgtg caaaggaaat taaggagagt gttcagacat tctttaagct 2280
tgtaaataaa tttttggctt tgtgtgctga ctctatcatt attggtggag ctaaacttaa 2340
agccttgaat ttaggtgaaa catttgtcac gcactcaaag ggattgtaca gaaagtgtgt 2400
taaatccaga gaagaaactg gcctactcat gcctctaaaa gccccaaaag aaattatctt 2460
cttagaggga gaaacacttc ccacagaagt gttaacagag gaagttgtct tgaaaactgg 2520
tgatttacaa ccattagaac aacctactag tgaagctgtt gaagctccat tggttggtac 2580
accagtttgt attaacgggc ttatgttgct cgaaatcaaa gacacagaaa agtactgtgc 2640
ccttgcacct aatatgatgg taacaaacaa taccttcaca ctcaaaggcg gtgcaccaac 2700
aaaggttact tttggtgatg acactgtgat agaagtgcaa ggttacaaga gtgtgaatat 2760
cacttttgaa cttgatgaaa ggattgataa agtacttaat gagaagtgct ctgcctatac 2820
agttgaactc ggtacagaag taaatgagtt cgcctgtgtt gtggcagatg ctgtcataaa 2880
aactttgcaa ccagtatctg aattacttac accactgggc attgatttag atgagtggag 2940
tatggctaca tactacttat ttgatgagtc tggtgagttt aaattggctt cacatatgta 3000
ttgttctttc taccctccag atgaggatga agaagaaggt gattgtgaag aagaagagtt 3060
tgagccatca actcaatatg agtatggtac tgaagatgat taccaaggta aacctttgga 3120
atttggtgcc acttctgctg ctcttcaacc tgaagaagag caagaagaag attggttaga 3180
tgatgatagt caacaaactg ttggtcaaca agacggcagt gaggacaatc agacaactac 3240
tattcaaaca attgttgagg ttcaacctca attagagatg gaacttacac cagttgttca 3300
gactattgaa gtgaatagtt ttagtggtta tttaaaactt actgacaatg tatacattaa 3360
aaatgcagac attgtggaag aagctaaaaa ggtaaaacca acagtggttg ttaatgcagc 3420
caatgtttac cttaaacatg gaggaggtgt tgcaggagcc ttaaataagg ctactaacaa 3480

tgccatgcaa gttgaatctg atgattacat agctactaat ggaccactta aagtgggtgg 3540
tagttgtgtt ttaagcggac acaatcttgc taaacactgt cttcatgttg tcggcccaaa 3600
tgttaacaaa ggtgaagaca ttcaacttct taagagtgct tatgaaaatt ttaatcagca 3660
cgaagttcta cttgcaccat tattatcagc tggtattttt ggtgctgacc ctatacattc 3720
tttaagagtt tgtgtagata ctgttcgcac aaatgtctac ttagctgtct ttgataaaaa 3780
tctctatgac aaacttgttt caagcttttt ggaaatgaag agtgaaaagc aagttgaaca 3840
aaagatcgct gagattccta aagaggaagt taagccattt ataactgaaa gtaaaccttc 3900
agttgaacag agaaaacaag atgataagaa aatcaaagct tgtgttgaag aagttacaac 3960
aactctggaa gaaactaagt tcctcacaga aaacttgtta ctttatattg acattaatgg 4020
caatcttcat ccagattctg ccactcttgt tagtgacatt gacatcactt tcttaaagaa 4080
agatgctcca tatatagtgg gtgatgttgt tcaagagggt gttttaactg ctgtggttat 4140
acctactaaa aaggctggtg gcactactga aatgctagcg aaagctttga gaaaagtgcc 4200
aacagacaat tatataacca cttacccggg tcagggttta aatggttaca ctgtagagga 4260
ggcaaagaca gtgcttaaaa agtgtaaaag tgccttttac attctaccat ctattatctc 4320
taatgagaag caagaaattc ttggaactgt ttcttggaat ttgcgagaaa tgcttgcaca 4380
tgcagaagaa acacgcaaat taatgcctgt ctgtgtggaa actaaagcca tagtttcaac 4440
tatacagcgt aaatataagg gtattaaaat acaagagggt gtggttgatt atggtgctag 4500
attttacttt tacaccagta aaacaactgt agcgtcactt atcaacacac ttaacgatct 4560
aaatgaaact cttgttacaa tgccacttgg ctatgtaaca catggcttaa atttggaaga 4620
agctgctcgg tatatgagat ctctcaaagt gccagctaca gtttctgttt cttcacctga 4680
tgctgttaca gcgtataatg gttatcttac ttcttcttct aaaacacctg aagaacattt 4740
tattgaaacc atctcacttg ctggttccta taaagattgg tcctattctg gacaatctac 4800
acaactaggt atagaatttc ttaagagagg tgataaaagt gtatattaca ctagtaatcc 4860
taccacattc cacctagatg gtgaagttat cacctttgac aatcttaaga cacttctttc 4920
tttgagagaa gtgaggacta ttaaggtgtt tacaacagta gacaacatta acctccacac 4980
gcaagttgtg gacatgtcaa tgacatatgg acaacagttt ggtccaactt atttggatgg 5040
agctgatgtt actaaaataa aacctcataa ttcacatgaa ggtaaaacat tttatgtttt 5100
acctaatgat gacactctac gtgttgaggc ttttgagtac taccacacaa ctgatcctag 5160
ttttctgggt aggtacatgt cagcattaaa tcacactaaa aagtggaaat acccacaagt 5220
taatggttta acttctatta aatgggcaga taacaactgt tatcttgcca ctgcattgtt 5280
aacactccaa caaatagagt tgaagtttaa tccacctgct ctacaagatg cttattacag 5340
agcaagggct ggtgaagctg ctaacttttg tgcacttatc ttagcctact gtaataagac 5400
agtaggtgag ttaggtgatg ttagagaaac aatgagttac ttgtttcaac atgccaattt 5460
agattcttgc aaaagagtct tgaacgtggt gtgtaaaact tgtggacaac agcagacaac 5520
ccttaagggt gtagaagctg ttatgtacat gggcacactt tcttatgaac aatttaagaa 5580
aggtgttcag ataccttgta cgtgtggtaa acaagctaca aaatatctag tacaacagga 5640
gtcacctttt gttatgatgt cagcaccacc tgctcagtat gaacttaagc atggtacatt 5700
tacttgtgct agtgagtaca ctggtaatta ccagtgtggt cactataaac atataacttc 5760
taaagaaact ttgtattgca tagacggtgc tttacttaca aagtcctcag aatacaaagg 5820

tcctattacg gatgttttct acaaagaaaa cagttacaca acaaccataa aaccagttac 5880
ttataaattg gatggtgttg tttgtacaga aattgaccct aagttggaca attattataa 5940
gaaagacaat tcttatttca cagagcaacc aattgatctt gtaccaaacc aaccatatcc 6000
aaacgcaagc ttcgataatt ttaagtttgt atgtgataat atcaaatttg ctgatgattt 6060
aaaccagtta actggttata agaaacctgc ttcaagagag cttaaagtta catttttccc 6120
tgacttaaat ggtgatgtgg tggctattga ttataaacac tacacaccct cttttaagaa 6180
aggagctaaa ttgttacata aacctattgt ttggcatgtt aacaatgcaa ctaataaagc 6240
cacgtataaa ccaaatacct ggtgtatacg ttgtctttgg agcacaaaac cagttgaaac 6300
atcaaattcg tttgatgtac tgaagtcaga ggacgcgcag ggaatggata atcttgcctg 6360
cgaagatcta aaaccagtct ctgaagaagt agtggaaaat cctaccatac agaaagacgt 6420
tcttgagtgt aatgtgaaaa ctaccgaagt tgtaggagac attatactta aaccagcaaa 6480
taatagttta aaaattacag aagaggttgg ccacacagat ctaatggctg cttatgtaga 6540
caattctagt cttactatta agaaacctaa tgaattatct agagtattag gtttgaaaac 6600
ccttgctact catggtttag ctgctgttaa tagtgtccct tgggatacta tagctaatta 6660
tgctaagcct tttcttaaca aagttgttag tacaactact aacatagtta cacggtgttt 6720
aaaccgtgtt tgtactaatt atatgcctta tttctttact ttattgctac aattgtgtac 6780
ttttactaga agtacaaatt ctagaattaa agcatctatg ccgactacta tagcaaagaa 6840
tactgttaag agtgtcggta aattttgtct agaggcttca tttaattatt tgaagtcacc 6900
taatttttct aaactgataa atattataat ttggttttta ctattaagtg tttgcctagg 6960
ttctttaatc tactcaaccg ctgctttagg tgttttaatg tctaatttag gcatgccttc 7020
ttactgtact ggttacagag aaggctattt gaactctact aatgtcacta ttgcaaccta 7080
ctgtactggt tctatacctt gtagtgtttg tcttagtggt ttagattctt tagacaccta 7140
tccttcttta gaaactatac aaattaccat ttcatctttt aaatgggatt taactgcttt 7200
tggcttagtt gcagagtggt ttttggcata tattcttttc actaggtttt tctatgtact 7260
tggattggct gcaatcatgc aattgttttt cagctatttt gcagtacatt ttattagtaa 7320
ttcttggctt atgtggttaa taattaatct tgtacaaatg gccccgattt cagctatggt 7380
tagaatgtac atcttctttg catcatttta ttatgtatgg aaaagttatg tgcatgttgt 7440
agacggttgt aattcatcaa cttgtatgat gtgttacaaa cgtaatagag caacaagagt 7500
cgaatgtaca actattgtta atggtgttag aaggtccttt tatgtctatg ctaatggagg 7560
taaaggcttt tgcaaactac acaattggaa ttgtgttaat tgtgatacat tctgtgctgg 7620
tagtacattt attagtgatg aagttgcgag agacttgtca ctacagttta aaagaccaat 7680
aaatcctact gaccagtctt cttacatcgt tgatagtgtt acagtgaaga atggttccat 7740
ccatctttac tttgataaag ctggtcaaaa gacttatgaa agacattctc tctctcattt 7800
tgttaactta gacaacctga gagctaataa cactaaaggt tcattgccta ttaatgttat 7860
agtttttgat ggtaaatcaa aatgtgaaga atcatctgca aaatcagcgt ctgtttacta 7920
cagtcagctt atgtgtcaac ctatactgtt actagatcag gcattagtgt ctgatgttgg 7980
tgatagtgcg gaagttgcag ttaaaatgtt tgatgcttac gttaatacgt tttcatcaac 8040
ttttaacgta ccaatggaaa aactcaaaac actagttgca actgcagaag ctgaacttgc 8100
aaagaatgtg tccttagaca atgtcttatc tacttttatt tcagcagctc ggcaagggtt 8160

tgttgattca gatgtagaaa ctaaagatgt tgttgaatgt cttaaattgt cacatcaatc 8220
tgacatagaa gttactggcg atagttgtaa taactatatg ctcacctata acaaagttga 8280
aaacatgaca ccccgtgacc ttggtgcttg tattgactgt agtgcgcgtc atattaatgc 8340
gcaggtagca aaaagtcaca acattgcttt gatatggaac gttaaagatt tcatgtcatt 8400
gtctgaacaa ctacgaaaac aaatacgtag tgctgctaaa aagaataact tacctttaa 8460
gttgacatgt gcaactacta gacaagttgt taatgttgta acaacaaaga tagcacttaa 8520
gggtggtaaa attgttaata attggttgaa gcagttaatt aaagttacac ttgtgttcct 8580
ttttgttgct gctattttct atttaataac acctgttcat gtcatgtcta aacatactga 8640
cttttcaagt gaaatcatag gatacaaggc tattgatggt ggtgtcactc gtgacatagc 8700
atctacagat acttgttttg ctaacaaaca tgctgatttt gacacatggt ttagccagcg 8760
tggtggtagt tatactaatg acaaagcttg cccattgatt gctgcagtca taacaagaga 8820
agtgggtttt gtcgtgcctg gtttgcctgg cacgatatta cgcacaacta atggtgactt 8880
tttgcatttc ttacctagag tttttagtgc agttggtaac atctgttaca caccatcaaa 8940
acttatagag tacactgact ttgcaacatc agcttgtgtt ttggctgctg aatgtacaat 9000
ttttaaagat gcttctggta agccagtacc atattgttat gataccaatg tactagaagg 9060
ttctgttgct tatgaaagtt tacgccctga cacacgttat gtgctcatgg atggctctat 9120
tattcaattt cctaacacct accttgaagg ttctgttaga gtggtaacaa cttttgattc 9180
tgagtactgt aggcacggca cttgtgaaag atcagaagct ggtgtttgtg tatctactag 9240
tggtagatgg gtacttaaca atgattatta cagatcttta ccaggagttt tctgtggtgt 9300
agatgctgta aatttactta ctaatatgtt tacaccacta attcaaccta ttggtgcttt 9360
ggacatatca gcatctatag tagctggtgg tattgtagct atcgtagtaa catgccttgc 9420
ctactatttt atgaggttta gaagagcttt tggtgaatac agtcatgtag ttgcctttaa 9480
tactttacta ttccttatgt cattcactgt actctgttta acaccagttt actcattctt 9540
acctggtgtt tattctgtta tttacttgta cttgacattt tatcttacta atgatgtttc 9600
ttttttagca catattcagt ggatggttat gttcacacct ttagtacctt tctggataac 9660
aattgcttat atcatttgta tttccacaaa gcatttctat tggttcttta gtaattacct 9720
aaagagacgt gtagtcttta atggtgtttc ctttagtact tttgaagaag ctgcgctgtg 9780
caccttttg ttaaataaag aaatgtatct aaagttgcgt agtgatgtgc tattacctct 9840
tacgcaatat aatagatact tagctcttta taataagtac aagtatttta gtggagcaat 9900
ggatacaact agctacagag aagctgcttg ttgtcatctc gcaaaggctc tcaatgactt 9960
cagtaactca ggttctgatg ttctttacca accaccacaa acctctatca cctcagctgt 10020
tttgcagagt ggttttagaa aaatggcatt cccatctggt aaagttgagg gttgtatggt 10080
acaagtaact tgtggtacaa ctacacttaa cggtctttgg cttgatgacg tagtttactg 10140
tccaagacat gtgatctgca cctctgaaga catgcttaac cctaattatg aagatttact 10200
cattcgtaag tctaatcata atttcttggt acaggctggt aatgttcaac tcagggttat 10260
tggacattct atgcaaaatt gtgtacttaa gcttaaggtt gatacagcca atcctaagac 10320
acctaagtat aagtttgttc gcattcaacc aggacagact ttttcagtgt tagcttgtta 10380
caatggttca ccatctggtg tttaccaatg tgctatgagg cccaatttca ctattaaggg 10440
ttcattcctt aatggttcat gtggtagtgt tggttttaac atagattatg actgtgtctc 10500

ttttgttac atgcaccata tggaattacc aactggagtt catgctggca cagacttaga 10560
aggtaacttt tatggacctt ttgttgacag gcaaacagca caagcagctg gtacggacac 10620
aactattaca gttaatgttt tagcttggtt gtacgctgct gttataaatg gagacaggtg 10680
gtttctcaat cgatttacca caactcttaa tgactttaac cttgtggcta tgaagtacaa 10740
ttatgaacct ctaacacaag accatgttga catactagga cctctttctg ctcaaactgg 10800
aattgccgtt ttagatatgt gtgcttcatt aaaagaatta ctgcaaaatg gtatgaatgg 10860
acgtaccata ttgggtagtg ctttattaga agatgaattt acacctttg atgttgttag 10920
acaatgctca ggtgttactt tccaaagtgc agtgaaaaga acaatcaagg gtacacacca 10980
ctggttgtta ctcacaattt tgacttcact tttagtttta gtccagagta ctcaatggtc 11040
tttgttcttt tttttgtatg aaaatgcctt tttacctttt gctatgggta ttattgctat 11100
gtctgctttt gcaatgatgt ttgtcaaaca taagcatgca tttctctgtt tgtttttgtt 11160
accttctctt gccactgtag cttattttaa tatggtctat atgcctgcta gttgggtgat 11220
gcgtattatg acatggttgg atatggttga tactagtttg tctggtttta agctaaaaga 11280
ctgtgttatg tatgcatcag ctgtagtgtt actaatcctt atgacagcaa gaactgtgta 11340
tgatgatggt gctaggagag tgtggacact tatgaatgtc ttgacactcg tttataaagt 11400
ttattatggt aatgctttag atcaagccat ttccatgtgg gctcttataa tctctgttac 11460
ttctaactac tcaggtgtag ttacaactgt catgtttttg gccagaggta ttgtttttat 11520
gtgtgttgag tattgcccta ttttcttcat aactggtaat acacttcagt gtataatgct 11580
agtttattgt ttcttaggct atttttgtac ttgttacttt ggcctctttt gtttactcaa 11640
ccgctacttt agactgactc ttggtgttta tgattactta gtttctacac aggagtttag 11700
atatatgaat tcacagggac tactcccacc caagaatagc atagatgcct tcaaactcaa 11760
cattaaattg ttgggtgttg gtggcaaacc ttgtatcaaa gtagccactg tacagtctaa 11820
aatgtcagat gtaaagtgca catcagtagt cttactctca gttttgcaac aactcagagt 11880
agaatcatca tctaaattgt gggctcaatg tgtccagtta cacaatgaca ttctcttagc 11940
taaagatact actgaagcct ttgaaaaaat ggtttcacta ctttctgttt tgctttccat 12000
gcagggtgct gtagacataa acaagctttg tgaagaaatg ctggacaaca gggcaacctt 12060
acaagctata gcctcagagt ttagttccct tccatcatat gcagcttttg ctactgctca 12120
agaagcttat gagcaggctg ttgctaatgg tgattctgaa gttgttctta aaaagttgaa 12180
gaagtctttg aatgtggcta aatctgaatt tgaccgtgat gcagccatgc aacgtaagtt 12240
ggaaaagatg gctgatcaag ctatgaccca aatgtataaa caggctagat ctgaggacaa 12300
gagggcaaaa gttactagtg ctatgcagac aatgcttttc actatgctta gaaagttgga 12360
taatgatgca ctcaacaaca ttatcaacaa tgcaagagat ggttgtgttc ccttgaacat 12420
aatacctctt acaacagcag ccaaactaat ggttgtcata ccagactata acacatataa 12480
aaatacgtgt gatggtacaa catttactta tgcatcagca ttgtgggaaa tccaacaggt 12540
tgtagatgca gatagtaaaa ttgttcaact tagtgaaatt agtatggaca attcacctaa 12600
tttagcatgg cctcttattg taacagcttt aagggccaat tctgctgtca aattacagaa 12660
taatgagctt agtcctgttg cactacgaca gatgtcttgt gctgccggta ctacacaaac 12720
tgcttgcact gatgacaatg cgttagctta ctacaacaca acaaagggag gtaggtttgt 12780
acttgcactg ttatccgatt tacaggattt gaaatgggct agattcccta agagtgatgg 12840

aactggtact atctatacag aactggaacc accttgtagg tttgttacag acacacctaa 12900
aggtcctaaa gtgaagtatt tatactttat taaaggatta aacaacctaa atagaggtat 12960
ggtacttggt agtttagctg ccacagtacg tctacaagct ggtaatgcaa cagaagtgcc 13020
tgccaattca actgtattat ctttctgtgc ttttgctgta gatgctgcta aagcttacaa 13080
agattatcta gctagtgggg gacaaccaat cactaattgt gttaagatgt tgtgtacaca 13140
cactggtact ggtcaggcaa taacagttac accggaagcc aatatggatc aagaatcctt 13200
tggtggtgca tcgtgttgtc tgtactgccg ttgccacata gatcatccaa atcctaaagg 13260
attttgtgac ttaaaaggta agtatgtaca aatacctaca acttgtgcta atgaccctgt 13320
gggttttaca cttaaaaaca cagtctgtac cgtctgcggt atgtggaaag gttatggctg 13380
tagttgtgat caactccgcg aacccatgct tcagtcagct gatgcacaat cgtttttaaa 13440
cgggtttgcg gtgtaagtgc agcccgtctt acaccgtgcg gcacaggcac tagtactgat 13500
gtcgtataca gggcttttga catctacaat gataaagtag ctggttttgc taaattccta 13560
aaaactaatt gttgtcgctt ccaagaaaag gacgaagatg acaatttaat tgattcttac 13620
tttgtagtta agagacacac tttctctaac taccaacatg aagaaacaat ttataattta 13680
cttaaggatt gtccagctgt tgctaaacat gacttcttta agtttagaat agacggtgac 13740
atggtaccac atatatcacg tcaacgtctt actaaataca caatggcaga cctcgtctat 13800
gctttaaggc attttgatga aggtaattgt gacacattaa aagaaatact tgtcacatac 13860
aattgttgtg atgatgatta tttcaataaa aaggactggt atgattttgt agaaaaccca 13920
gatatattac gcgtatacgc caacttaggt gaacgtgtac gccaagcttt gttaaaaaca 13980
gtacaattct gtgatgccat gcgaaatgct ggtattgttg gtgtactgac attagataat 14040
caagatctca atggtaactg gtatgatttc ggtgatttca tacaaaccac gccaggtagt 14100
ggagttcctg ttgtagattc ttattattca ttgttaatgc ctatattaac cttgaccagg 14160
gctttaactg cagagtcaca tgttgacact gacttaacaa agccttacat taagtgggat 14220
ttgttaaaat atgacttcac ggaagagagg ttaaaactct ttgaccgtta ttttaaatat 14280
tgggatcaga cataccaccc aaattgtgtt aactgtttgg atgacagatg cattctgcat 14340
tgtgcaaact ttaatgtttt attctctaca gtgttcccac ctacaagttt tggaccacta 14400
gtgagaaaaa tatttgttga tggtgttcca tttgtagttt caactggata ccacttcaga 14460
gagctaggtg ttgtacataa tcaggatgta aacttacata gctctagact tagttttaag 14520
gaattacttg tgtatgctgc tgaccctgct atgcacgctg cttctggtaa tctattacta 14580
gataaacgca ctacgtgctt ttcagtagct gcacttacta acaatgttgc ttttcaaact 14640
gtcaaacccg gtaattttaa caaagacttc tatgactttg ctgtgtctaa gggtttcttt 14700
aaggaaggaa gttctgttga attaaaacac ttcttctttg ctcaggatgg taatgctgct 14760
atcagcgatt atgactacta tcgttataat ctaccaacaa tgtgtgatat cagacaacta 14820
ctatttgtag ttgaagttgt tgataagtac tttgattgtt acgatggtgg ctgtattaat 14880
gctaaccaag tcatcgtcaa caacctagac aaatcagctg gttttccatt taataaatgg 14940
ggtaaggcta gactttatta tgattcaatg agttatgagg atcaagatgc acttttcgca 15000
tatacaaaac gtaatgtcat ccctactata actcaaatga atcttaagta tgccattagt 15060
gcaaagaata gagctcgcac cgtagctggt gtctctatct gtagtactat gaccaataga 15120
cagtttcatc aaaaattatt gaaatcaata gccgccacta gaggagctac tgtagtaatt 15180

ggaacaagca aattctatgg tggttggcac aacatgttaa aaactgttta tagtgatgta 15240
gaaaaccctc accttatggg ttgggattat cctaaatgtg atagagccat gcctaacatg 15300
cttagaatta tggcctcact tgttcttgct cgcaaacata caacgtgttg tagcttgtca 15360
caccgtttct atagattagc taatgagtgt gctcaagtat tgagtgaaat ggtcatgtgt 15420
ggcggttcac tatatgttaa accaggtgga acctcatcag gagatgccac aactgcttat 15480
gctaatagtg tttttaacat ttgtcaagct gtcacggcca atgttaatgc acttttatct 15540
actgatggta acaaaattgc cgataagtat gtccgcaatt tacaacacag actttatgag 15600
tgtctctata gaaatagaga tgttgacaca gactttgtga atgagtttta cgcatatttg 15660
cgtaaacatt tctcaatgat gatactctct gacgatgctg ttgtgtgttt caatagcact 15720
tatgcatctc aaggtctagt ggctagcata aagaacttta agtcagttct ttattatcaa 15780
aacaatgttt ttatgtctga agcaaaatgt tggactgaga ctgaccttac taaaggacct 15840
catgaatttt gctctcaaca tacaatgcta gttaaacagg gtgatgatta tgtgtacctt 15900
ccttacccag atccatcaag aatcctaggg gccggctgtt ttgtagatga tatcgtaaaa 15960
acagatggta cacttatgat tgaacggttc gtgtctttag ctatagatgc ttacccactt 16020
actaaacatc ctaatcagga gtatgctgat gtctttcatt tgtacttaca atacataaga 16080
aagctacatg atgagttaac aggacacatg ttagacatgt attctgttat gcttactaat 16140
gataacactt caaggtattg ggaacctgag ttttatgagg ctatgtacac accgcataca 16200
gtcttacagg ctgttggggc ttgtgttctt tgcaattcac agacttcatt aagatgtggt 16260
gcttgcatac gtagaccatt cttatgttgt aaatgctgtt acgaccatgt catatcaaca 16320
tcacataaat tagtcttgtc tgttaatccg tatgtttgca atgctccagg ttgtgatgtc 16380
acagatgtga ctcaacttta cttaggaggt atgagctatt attgtaaatc acataaacca 16440
cccattagtt ttccattgtg tgctaatgga caagtttttg gtttatataa aaatacatgt 16500
gttggtagcg ataatgttac tgactttaat gcaattgcaa catgtgactg gacaaatgct 16560
ggtgattaca ttttagctaa cacctgtact gaaagactca agctttttgc agcagaaacg 16620
ctcaaagcta ctgaggagac atttaaactg tcttatggta ttgctactgt acgtgaagtg 16680
ctgtctgaca gagaattaca tctttcatgg gaagttggta aacctagacc accacttaac 16740
cgaaattatg tctttactgg ttatcgtgta actaaaaaca gtaaagtaca aataggagag 16800
tacacctttg aaaaaggtga ctatggtgat gctgttgttt accgaggtac aacaacttac 16860
aaattaaatg ttggtgatta ttttgtgctg acatcacata cagtaatgcc attaagtgca 16920
cctacactag tgccacaaga gcactatgtt agaattactg gcttataccc aacactcaat 16980
atctcagatg agttttctag caatgttgca aattatcaaa aggttggtat gcaaaagtat 17040
tctacactcc agggaccacc tggtactggt aagagtcatt ttgctattgg cctagctctc 17100
tactaccctt ctgctcgcat agtgtataca gcttgctctc atgccgctgt tgatgcacta 17160
tgtgagaagg cattaaaata tttgcctata gataaatgta gtagaattat acctgcacgt 17220
gctcgtgtag agtgttttga taaattcaaa gtgaattcaa cattagaaca gtatgtcttt 17280
tgtactgtaa atgcattgcc tgagacgaca gcagatatag ttgtctttga tgaaatttca 17340
atggccacaa attatgattt gagtgttgtc aatgccagat tacgtgctaa gcactatgtg 17400
tacattggcg accctgctca attacctgca ccacgcacat tgctaactaa gggcacacta 17460
gaaccagaat atttcaattc agtgtgtaga cttatgaaaa ctataggtcc agacatgttc 17520

ctcggaactt gtcggcgttg tcctgctgaa attgttgaca ctgtgagtgc tttggtttat 17580
gataataagc ttaaagcaca taaagacaaa tcagctcaat gctttaaaat gttttataag 17640
ggtgttatca cgcatgatgt ttcatctgca attaacaggc cacaaatagg cgtggtaaga 17700
gaattcctta cacgtaaccc tgcttggaga aaagctgtct ttatttcacc ttataattca 17760
cagaatgctg tagcctcaaa gattttggga ctaccaactc aaactgttga ttcatcacag 17820
ggctcagaat atgactatgt catattcact caaaccactg aaacagctca ctcttgtaat 17880
gtaaacagat ttaatgttgc tattaccaga gcaaaagtag gcatactttg cataatgtct 17940
gatagagacc tttatgacaa gttgcaattt acaagtcttg aaattccacg taggaatgtg 18000
gcaactttac aagctgaaaa tgtaacagga ctctttaaag attgtagtaa ggtaatcact 18060
gggttacatc ctacacaggc acctacacac ctcagtgttg acactaaatt caaaactgaa 18120
ggtttatgtg ttgacatacc tggcatacct aaggacatga cctatagaag actcatctct 18180
atgatgggtt ttaaaatgaa ttatcaagtt aatggttacc ctaacatgtt tatcacccgc 18240
gaagaagcta taagacatgt acgtgcatgg attggcttcg atgtcgaggg gtgtcatgct 18300
actagagaag ctgttggtac caatttacct ttacagctag gtttttctac aggtgttaac 18360
ctagttgctg tacctacagg ttatgttgat acacctaata atacagattt ttccagagtt 18420
agtgctaaac caccgcctgg agatcaattt aaacacctca taccacttat gtacaaagga 18480
cttctttgga atgtagtgcg tataaagatt gtacaaatgt taagtgacac acttaaaaat 18540
ctctctgaca gagtcgtatt tgtcttatgg gcacatggct ttgagttgac atctatgaag 18600
tattttgtga aaataggacc tgagcgcacc tgttgtctat gtgatagacg tgccacatgc 18660
ttttccactg cttcagacac ttatgcctgt tggcatcatt ctattggatt tgattacgtc 18720
tataatccgt ttatgattga tgttcaacaa tggggtttta caggtaacct acaaagcaac 18780
catgatctgt attgtcaagt ccatggtaat gcacatgtag ctagttgtga tgcaatcatg 18840
actaggtgtc tagctgtcca cgagtgcttt gttaagcgtg ttgactggac tattgaatat 18900
cctataattg gtgatgaact gaagattaat gcggcttgta gaaaggttca acacatggtt 18960
gttaaagctg cattattagc agacaaattc ccagttcttc acgacattgg taaccctaaa 19020
gctattaagt gtgtacctca agctgatgta gaatggaagt tctatgatgc acagccttgt 19080
agtgacaaag cttataaaat agaagaatta ttctattctt atgccacaca ttctgacaaa 19140
ttcacagatg gtgtatgcct attttggaat tgcaatgtcg atagatatcc tgctaattcc 19200
attgtttgta gatttgacac tagagtgcta tctaacctta acttgcctgg ttgtgatggt 19260
ggcagtttgt atgtaaataa acatgcattc cacacaccag cttttgataa aagtgctttt 19320
gttaatttaa aacaattacc atttttctat tactctgaca gtccatgtga gtctcatgga 19380
aaacaagtag tgtcagatat agattatgta ccactaaagt ctgctacgtg tataacacgt 19440
tgcaatttag gtggtgctgt ctgtagacat catgctaatg agtacagatt gtatctcgat 19500
gcttataaca tgatgatctc agctggcttt agcttgtggg tttacaaaca atttgatact 19560
tataacctct ggaacacttt tacaagactt cagagtttag aaaatgtggc ttttaatgtt 19620
gtaaataagg gacactttga tggacaacag ggtgaagtac cagtttctat cattaataac 19680
actgtttaca caaaagttga tggtgttgat gtagaattgt ttgaaaataa aacaacatta 19740
cctgttaatg tagcatttga gctttgggct aagcgcaaca ttaaaccagt accagaggtg 19800
aaaatactca ataatttggg tgtggacatt gctgctaata ctgtgatctg ggactacaaa 19860

agagatgctc cagcacatat atctactatt ggtgtttgtt ctatgactga catagccaag 19920
aaaccaactg aaacgatttg tgcaccactc actgtctttt ttgatggtag agttgatggt 19980
caagtagact tatttagaaa tgcccgtaat ggtgttctta ttacagaagg tagtgttaaa 20040
ggtttacaac catctgtagg tcccaaacaa gctagtctta atggagtcac attaattgga 20100
gaagccgtaa aaacacagtt caattattat aagaaagttg atggtgttgt ccaacaatta 20160
cctgaaactt actttactca gagtagaaat ttacaagaat ttaaacccag gagtcaaatg 20220
gaaattgatt tcttagaatt agctatggat gaattcattg aacggtataa attagaaggc 20280
tatgccttcg aacatatcgt ttatggagat tttagtcata gtcagttagg tggtttacat 20340
ctactgattg gactagctaa acgttttaag gaatcacctt ttgaattaga agattttatt 20400
cctatggaca gtacagttaa aaactatttc ataacagatg cgcaaacagg ttcatctaag 20460
tgtgtgtgtt ctgttattga tttattactt gatgattttg ttgaaataat aaaatcccaa 20520
gatttatctg tagtttctaa ggttgtcaaa gtgactattg actatacaga aatttcattt 20580
atgctttggt gtaaagatgg ccatgtagaa acattttacc caaaattaca atctagtcaa 20640
gcgtggcaac cgggtgttgc tatgcctaat ctttacaaaa tgcaaagaat gctattagaa 20700
aagtgtgacc ttcaaaatta tggtgatagt gcaacattac ctaaaggcat aatgatgaat 20760
gtcgcaaaat atactcaact gtgtcaatat ttaaacacat taacattagc tgtaccctat 20820
aatatgagag ttatacattt tggtgctggt tctgataaag gagttgcacc aggtacagct 20880
gttttaagac agtggttgcc tacgggtacg ctgcttgtcg attcagatct taatgacttt 20940
gtctctgatg cagattcaac tttgattggt gattgtgcaa ctgtacatac agctaataaa 21000
tgggatctca ttattagtga tatgtacgac cctaagacta aaaatgttac aaaagaaaat 21060
gactctaaag agggtttttt cacttacatt tgtgggttta tacaacaaaa gctagctctt 21120
ggaggttccg tggctataaa gataacagaa cattcttgga atgctgatct ttataagctc 21180
atgggacact tcgcatggtg gacagccttt gttactaatg tgaatgcgtc atcatctgaa 21240
gcatttttaa ttggatgtaa ttatcttggc aaaccacgcg aacaaataga tggttatgtc 21300
atgcatgcaa attacatatt ttggaggaat acaaatccaa ttcagttgtc ttcctattct 21360
ttatttgaca tgagtaaatt tccccttaaa ttaaggggta ctgctgttat gtctttaaaa 21420
gaaggtcaaa tcaatgatat gattttatct cttcttagta aaggtagact tataattaga 21480
gaaaacaaca gagttgttat ttctagtgat gttcttgtta acaactaaac gaacaatgtt 21540
tgtttttctt gttttattgc cactagtctc tagtcagtgt gttaatctta caaccagaac 21600
tcaattaccc cctgcataca ctaattcttt cacacgtggt gtttattacc ctgacaaagt 21660
tttcagatcc tcagttttac attcaactca ggacttgttc ttacctttct tttccaatgt 21720
tacttggttc catgctatac atgtctctgg gaccaatggt actaagaggt ttgataaccc 21780
tgtcctacca tttaatgatg gtgtttattt tgcttccact gagaagtcta acataataag 21840
aggctggatt tttggtacta ctttagattc gaagacccag tccctactta ttgttaataa 21900
cgctactaat gttgttatta aagtctgtga atttcaattt tgtaatgatc catttttggg 21960
tgtttattac cacaaaaaca acaaaagttg gatggaaagt gagttcagag tttattctag 22020
tgcgaataat tgcacttttg aatatgtctc tcagcctttt cttatggacc ttgaaggaaa 22080
acagggtaat ttcaaaaatc ttagggaatt tgtgtttaag aatattgatg gttattttaa 22140
aatatattct aagcacacgc ctattaattt agtgcgtgat ctccctcagg gtttttcggc 22200

tttagaacca ttggtagatt tgccaatagg tattaacatc actaggtttc aaactttact 22260
tgctttacat agaagttatt tgactcctgg tgattcttct tcaggttgga cagctggtgc 22320
tgcagcttat tatgtgggtt atcttcaacc taggactttt ctattaaaat ataatgaaaa 22380
tggaaccatt acagatgctg tagactgtgc acttgaccct ctctcagaaa caaagtgtac 22440
gttgaaatcc ttcactgtag aaaaaggaat ctatcaaact tctaacttta gagtccaacc 22500
aacagaatct attgttagat ttcctaatat tacaaacttg tgccctttg gtgaagtttt 22560
taacgccacc agatttgcat ctgtttatgc ttggaacagg aagagaatca gcaactgtgt 22620
tgctgattat tctgtcctat ataattccgc atcattttcc acttttaagt gttatggagt 22680
gtctcctact aaattaaatg atctctgctt tactaatgtc tatgcagatt catttgtaat 22740
tagaggtgat gaagtcagac aaatcgctcc agggcaaact ggaaagattg ctgattataa 22800
ttataaatta ccagatgatt ttacaggctg cgttatagct tggaattcta acaatcttga 22860
ttctaaggtt ggtggtaatt ataattacct gtatagattg tttaggaagt ctaatctcaa 22920
acctttgag agagatattt caactgaaat ctatcaggcc ggtagcacac cttgtaatgg 22980
tgttgaaggt tttaattgtt actttccttt acaatcatat ggtttccaac ccactaatgg 23040
tgttggttac caaccataca gagtagtagt actttctttt gaacttctac atgcaccagc 23100
aactgtttgt ggacctaaaa agtctactaa tttggttaaa aacaaatgtg tcaatttcaa 23160
cttcaatggt ttaacaggca caggtgttct tactgagtct aacaaaaagt ttctgccttt 23220
ccaacaattt ggcagagaca ttgctgacac tactgatgct gtccgtgatc cacagacact 23280
tgagattctt gacattacac catgttcttt tggtggtgtc agtgttataa caccaggaac 23340
aaatacttct aaccaggttg ctgttcttta tcaggatgtt aactgcacag aagtccctgt 23400
tgctattcat gcagatcaac ttactcctac ttggcgtgtt tattctacag gttctaatgt 23460
ttttcaaaca cgtgcaggct gtttaatagg ggctgaacat gtcaacaact catatgagtg 23520
tgacataccc attggtgcag gtatatgcgc tagttatcag actcagacta attctcctcg 23580
gcgggcacgt agtgtagcta gtcaatccat cattgcctac actatgtcac ttggtgcaga 23640
aaattcagtt gcttactcta ataactctat tgccataccc acaaatttta ctattagtgt 23700
taccacagaa attctaccag tgtctatgac caagacatca gtagattgta caatgtacat 23760
ttgtggtgat tcaactgaat gcagcaatct tttgttgcaa tatggcagtt tttgtacaca 23820
attaaaccgt gctttaactg gaatagctgt tgaacaagac aaaaacaccc aagaagtttt 23880
tgcacaagtc aaacaaattt acaaaacacc accaattaaa gattttggtg gttttaattt 23940
ttcacaaata ttaccagatc catcaaaacc aagcaagagg tcatttattg aagatctact 24000
tttcaacaaa gtgacacttg cagatgctgg cttcatcaaa caatatggtg attgccttgg 24060
tgatattgct gctagagacc tcatttgtgc acaaaagttt aacggcctta ctgttttgcc 24120
acctttgctc acagatgaaa tgattgctca atacacttct gcactgttag cgggtacaat 24180
cacttctggt tggacctttg gtgcaggtgc tgcattacaa ataccatttg ctatgcaaat 24240
ggcttatagg tttaatggta ttggagttac acagaatgtt ctctatgaga accaaaaatt 24300
gattgccaac caatttaata gtgctattgg caaaattcaa gactcacttt cttccacagc 24360
aagtgcactt ggaaaacttc aagatgtggt caaccaaaat gcacaagctt taaacacgct 24420
tgttaaacaa cttagctcca attttggtgc aatttcaagt gttttaaatg atatcctttc 24480
acgtcttgac aaagttgagg ctgaagtgca aattgatagg ttgatcacag gcagacttca 24540

aagtttgcag acatatgtga ctcaacaatt aattagagct gcagaaatca gagcttctgc 24600
taatcttgct gctactaaaa tgtcagagtg tgtacttgga caatcaaaaa gagttgattt 24660
ttgtggaaag ggctatcatc ttatgtcctt ccctcagtca gcacctcatg gtgtagtctt 24720
cttgcatgtg acttatgtcc ctgcacaaga aaagaacttc acaactgctc ctgccatttg 24780
tcatgatgga aaagcacact ttcctcgtga aggtgtcttt gtttcaaatg gcacacactg 24840
gtttgtaaca caaaggaatt tttatgaacc acaaatcatt actacagaca acacatttgt 24900
gtctggtaac tgtgatgttg taataggaat tgtcaacaac acagtttatg atcctttgca 24960
acctgaatta gactcattca aggaggagtt agataaatat tttaagaatc atacatcacc 25020
agatgttgat ttaggtgaca tctctggcat taatgcttca gttgtaaaca ttcaaaaaga 25080
aattgaccgc ctcaatgagg ttgccaagaa tttaaatgaa tctctcatcg atctccaaga 25140
acttggaaag tatgagcagt atataaaatg gccatggtac atttggctag gttttatagc 25200
tggcttgatt gccatagtaa tggtgacaat tatgctttgc tgtatgacca gttgctgtag 25260
ttgtctcaag ggctgttgtt cttgtggatc ctgctgcaaa tttgatgaag acgactctga 25320
gccagtgctc aaaggagtca aattacatta cacataaacg aacttatgga tttgtttatg 25380
agaatcttca caattggaac tgtaactttg aagcaaggtg aaatcaagga tgctactcct 25440
tcagattttg ttcgcgctac tgcaacgata ccgatacaag cctcactccc tttcggatgg 25500
cttattgttg gcgttgcact tcttgctgtt tttcagagcg cttccaaaat cataaccctc 25560
aaaaagagat ggcaactagc actctccaag ggtgttcact ttgtttgcaa cttgctgttg 25620
ttgtttgtaa cagtttactc acacctttg ctcgttgctg ctggccttga agccccttt 25680
ctctatcttt atgctttagt ctacttcttg cagagtataa actttgtaag aataataatg 25740
aggctttggc tttgctggaa atgccgttcc aaaaacccat tactttatga tgccaactat 25800
tttctttgct ggcatactaa ttgttacgac tattgtatac cttacaatag tgtaacttct 25860
tcaattgtca ttacttcagg tgatggcaca acaagtccta tttctgaaca tgactaccag 25920
attggtggtt atactgaaaa atgggaatct ggagtaaaag actgtgttgt attacacagt 25980
tacttcactt cagactatta ccagctgtac tcaactcaat tgagtacaga cactggtgtt 26040
gaacatgtta ccttcttcat ctacaataaa attgttgatg agcctgaaga acatgtccaa 26100
attcacacaa tcgacggttc atccggagtt gttaatccag taatggaacc aatttatgat 26160
gaaccgacga cgactactag cgtgcctttg taagcacaag ctgatgagta cgaacttatg 26220
tactcattcg tttcggaaga gacaggtacg ttaatagtta atagcgtact tctttttctt 26280
gctttcgtgg tattcttgct agttacacta gccatcctta ctgcgcttcg attgtgtgcg 26340
tactgctgca atattgttaa cgtgagtctt gtaaaacctt ctttttacgt ttactctcgt 26400
gttaaaaatc tgaattcttc tagagttcct gatcttctgg tctaaacgaa ctaaatatta 26460
tattagtttt tctgtttgga actttaattt tagccatggc agattccaac ggtactatta 26520
ccgttgaaga gcttaaaaag ctccttgaac aatggaacct agtaataggt ttcctattcc 26580
ttacatggat ttgtcttcta caatttgcct atgccaacag gaataggttt ttgtatataa 26640
ttaagttaat tttcctctgg ctgttatggc cagtaacttt agcttgtttt gtgcttgctg 26700
ctgtttacag aataaattgg atcaccggtg gaattgctat cgcaatggct tgtcttgtag 26760
gcttgatgtg gctcagctac ttcattgctt ctttcagact gtttgcgcgt acgcgttcca 26820
tgtggtcatt caatccagaa actaacattc ttctcaacgt gccactccat ggcactattc 26880

tgaccagacc gcttctagaa agtgaactcg taatcggagc tgtgatcctt cgtggacatc 26940
ttcgtattgc tggacaccat ctaggacgct gtgacatcaa ggacctgcct aaagaaatca 27000
ctgttgctac atcacgaacg ctttcttatt acaaattggg agcttcgcag cgtgtagcag 27060
gtgactcagg ttttgctgca tacagtcgct acaggattgg caactataaa ttaaacacag 27120
accattccag tagcagtgac aatattgctt tgcttgtaca gtaagtgaca acagatgttt 27180
catctcgttg actttcaggt tactatagca gagatattac taattattat gaggactttt 27240
aaagtttcca tttggaatct tgattacatc ataaacctca taattaaaaa tttatctaag 27300
tcactaactg agaataaata ttctcaatta gatgaagagc aaccaatgga gattgattaa 27360
acgaacatga aaattattct tttcttggca ctgataacac tcgctacttg tgagctttat 27420
cactaccaag agtgtgttag aggtacaaca gtacttttaa aagaaccttg ctcttctgga 27480
acatacgagg gcaattcacc atttcatcct ctagctgata acaaatttgc actgacttgc 27540
tttagcactc aatttgcttt tgcttgtcct gacggcgtaa aacacgtcta tcagttacgt 27600
gccagatcag tttcacctaa actgttcatc agacaagagg aagttcaaga actttactct 27660
ccaatttttc ttattgttgc ggcaatagtg tttataacac tttgcttcac actcaaaaga 27720
aagacagaat gattgaactt tcattaattg acttctattt gtgctttta gcctttctgc 27780
tattccttgt tttaattatg cttattatct tttggttctc acttgaactg caagatcata 27840
atgaaacttg tcacgcctaa acgaacatga aatttcttgt tttcttagga atcatcacaa 27900
ctgtagctgc atttcaccaa gaatgtagtt tacagtcatg tactcaacat caaccatatg 27960
tagttgatga cccgtgtcct attcacttct attctaaatg gtatattaga gtaggagcta 28020
gaaaatcagc acctttaatt gaattgtgcg tggatgaggc tggttctaaa tcacccattc 28080
agtacatcga tatcggtaat tatacagttt cctgtttacc ttttacaatt aattgccagg 28140
aacctaaatt gggtagtctt gtagtgcgtt gttcgttcta tgaagacttt ttagagtatc 28200
atgacgttcg tgttgtttta gatttcatct aaacgaacaa actaaaatgt ctgataatgg 28260
accccaaaat cagcgaaatg caccccgcat tacgtttggt ggaccctcag attcaactgg 28320
cagtaaccag aatggagaac gcagtggggc gcgatcaaaa caacgtcggc cccaaggttt 28380
acccaataat actgcgtctt ggttcaccgc tctcactcaa catggcaagg aagaccttaa 28440
attccctcga ggacaaggcg ttccaattaa caccaatagc agtccagatg accaaattgg 28500
ctactaccga agagctacca gacgaattcg tggtggtgac ggtaaaatga aagatctcag 28560
tccaagatgg tatttctact acctaggaac tgggccagaa gctggacttc cctatggtgc 28620
taacaaagac ggcatcatat gggttgcaac tgagggagcc ttgaatacac caaaagatca 28680
cattggcacc cgcaatcctg ctaacaatgc tgcaatcgtg ctacaacttc ctcaaggaac 28740
aacattgcca aaaggcttct acgcagaagg gagcagaggc ggcagtcaag cctcttctcg 28800
ttcctcatca cgtagtcgca acagttcaag aaattcaact ccaggcagca gtaggggaac 28860
ttctcctgct agaatggctg gcaatggcgg tgatgctgct cttgctttgc tgctgcttga 28920
cagattgaac cagcttgaga gcaaaatgtc tggtaaaggc caacaacaac aaggccaaac 28980
tgtcactaag aaatctgctg ctgaggcttc taagaagcct cggcaaaaac gtactgccac 29040
taaagcatac aatgtaacac aagctttcgg cagacgtggt ccagaacaaa cccaaggaaa 29100
ttttggggac caggaactaa tcagacaagg aactgattac aaacattggc cgcaaattgc 29160
acaatttgcc cccagcgctt cagcgttctt cggaatgtcg cgcattggca tggaagtcac 29220

accttcggga acgtggttga cctacacagg tgccatcaaa ttggatgaca aagatccaaa 29280
tttcaaagat caagtcattt tgctgaataa gcatattgac gcatacaaaa cattcccacc 29340
aacagagcct aaaaaggaca aaaagaagaa ggctgatgaa actcaagcct taccgcagag 29400
acagaagaaa cagcaaactg tgactcttct tcctgctgca gatttggatg atttctccaa 29460
acaattgcaa caatccatga gcagtgctga ctcaactcag gcctaaactc atgcagacca 29520
cacaaggcag atgggctata taaacgtttt cgcttttccg tttacgatat atagtctact 29580
cttgtgcaga atgaattctc gtaactacat agcacaagta gatgtagtta actttaatct 29640
cacatagcaa tctttaatca gtgtgtaaca ttagggagga cttgaaagag ccaccacatt 29700
ttcaccgagg ccacgcggag tacgatcgag tgtacagtga acaatgctag ggagagctgc 29760
ctatatggaa gagccctaat gtgtaaaatt aattttagta gtgctatccc catgtgattt 29820
taatagcttc ttaggagaat gacaaaaa 29848

<210> 2
<211> 1273
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<220>
<221> MISC_FEATURE
<222> (1)..(12)
<223> 信号肽
<220>
<221> MISC_FEATURE
<222> (13)..(685)
<223> S1区域
<220>
<221> MISC_FEATURE
<222> (13)..(1213)
<223> 胞外结构域
<220>
<221> MISC_FEATURE
<222> (686)..(1273)
<223> S2区域
<220>
<221> MISC_FEATURE
<222> (1214)..(1234)
<223> 细胞质结构域
<300>
<308> https://www.ncbi.nlm.nih.gov/protein/NP_001358344.1
<309> 2020-07-18
<313> (1)..(1273)

<300>
<308> YP_009724390.1
<309> 2020-07-18
<313> (1)..(1273)
<400> 2

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15
Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80
Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95
Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110
Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125
Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140
Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160
Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175
Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190
Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205
Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255
Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285
Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300
Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335
Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
370 375 380
Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385 390 395 400
Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe

580 585 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val

595 600 605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile

610 615 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser

625 630 635 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val

645 650 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala

675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser

690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile

705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu

740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr

755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln

770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe

785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser

805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly

820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp

835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly

865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile

885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940
Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
1010 1015 1020
Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
1025 1030 1035
Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
1040 1045 1050
Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
1055 1060 1065
Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080
Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095
Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110
Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125
Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140
Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155
His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170
Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185
Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu

1190 1195 1200
Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215
Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230
Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
1250 1255 1260
Val Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 3
<211> 75
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<300>
<308> QPK75943.1
<309> 2020-12-03
<313> (1)..(75)
<400> 3
Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu Ile Val Asn Ser
1 5 10 15
Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala
20 25 30
Ile Leu Thr Ala Leu Arg Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn
35 40 45
Val Ser Leu Val Lys Pro Ser Phe Tyr Val Tyr Ser Arg Val Lys Asn
50 55 60
Leu Asn Ser Ser Arg Val Pro Asp Leu Leu Val
65 70 75

<210> 4
<211> 196
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 4
Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr
1 5 10 15
Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val
20 25 30

Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser
35 40 45
Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser
50 55 60
Phe Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr
65 70 75 80
Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly
85 90 95
Cys Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly
100 105 110
Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro
115 120 125
Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro
130 135 140
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
145 150 155 160
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
165 170 175
Val Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro
180 185 190
Lys Lys Ser Thr
195

<210> 5
<211> 202
<212> PRT
<213> 人工序列(Artificial Sequence)
<220>
<223> 带有N-端C6的截短刺突(S)蛋白
<400> 5
Cys Cys Cys Cys Cys Cys Cys Pro Phe Gly Glu Val Phe Asn Ala Thr
1 5 10 15
Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg Lys Arg Ile Ser Asn Cys
20 25 30
Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Ala Ser Phe Ser Thr Phe
35 40 45
Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu Asn Asp Leu Cys Phe Thr
50 55 60
Asn Val Tyr Ala Asp Ser Phe Val Ile Arg Gly Asp Glu Val Arg Gln
65 70 75 80

Ile Ala Pro Gly Gln Thr Gly Lys Ile Ala Asp Tyr Asn Tyr Lys Leu
85 90 95
Pro Asp Asp Phe Thr Gly Cys Val Ile Ala Trp Asn Ser Asn Asn Leu
100 105 110
Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg
115 120 125
Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr
130 135 140
Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr
145 150 155 160
Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr
165 170 175
Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu His Ala Pro
180 185 190
Ala Thr Val Cys Gly Pro Lys Lys Ser Thr
195 200

<210> 6
<211> 201
<212> PRT
<213> 人工序列
<220>
<223> 带有C-端H6的截短刺突(S)蛋白
<400> 6

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

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
130 135 140
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
145 150 155 160
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
165 170 175
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
180 185 190
Lys Ser Thr His His His His His His
195 200

<210> 7
<211> 9
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 7
Leu Leu Phe Asn Lys Val Thr Leu Ala
1 5

<210> 8
<211> 13
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 8
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1 5 10

<210> 9
<211> 9
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 9
Tyr Leu Gln Pro Arg Thr Phe Leu Leu
1 5

<210> 10
<211> 9
<212> PRT
<213> 人工序列
<220>
<223> T细胞表位
<400> 10
Lys Leu Trp Ala Gln Cys Val Gln Leu

1 5

<210> 11

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> T细胞表位

<400> 11

Leu Leu Tyr Asp Ala Asn Tyr Phe Leu

1 5

<210> 12

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> T细胞表位

<400> 12

Pro Arg Trp Tyr Phe Tyr Tyr Leu Gly Thr Gly Pro

1 5 10

<210> 13

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> T细胞表位

<400> 13

Ser Pro Arg Trp Tyr Phe Tyr Tyr Leu

1 5

<210> 14

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> T细胞表位

<400> 14

Trp Ser Phe Asn Pro Glu Thr Asn

1 5

<210> 15

<211> 9

<212> PRT
<213> 人工序列
<220>
<223> T细胞表位
<400> 15
Gln Pro Pro Gly Thr Gly Lys Ser His
1 5
<210> 16
<211> 17
<212> PRT
<213> 人工序列
<220>
<223> T 细胞表位
<400> 16
Val Tyr Thr Ala Cys Ser His Ala Ala Val Asp Ala Leu Cys Glu Lys
1 5 10 15
Ala
<210> 17
<211> 9
<212> PRT
<213> 人工序列
<220>
<223> T细胞表位
<400> 17
Lys Thr Phe Pro Pro Thr Glu Pro Lys
1 5
<210> 18
<211> 10
<212> PRT
<213> 人工序列
<220>
<223> T细胞表位
<400> 18
Cys Thr Asp Asp Asn Ala Leu Ala Tyr Tyr
1 5 10
<210> 19
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> T细胞表位
<400> 19
Thr Thr Asp Pro Ser Phe Leu Gly Arg Tyr
1 5 10
<210> 20
<211> 9
<212> PRT
<213> 人工序列
<220>
<223> T细胞表位
<400> 20
Phe Thr Ser Asp Tyr Tyr Gln Leu Tyr
1 5
<210> 21
<211> 9
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 21
Phe Val Phe Leu Val Leu Leu Pro Leu
1 5
<210> 22
<211> 9
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 22
Phe Ile Ala Gly Leu Ile Ala Ile Val
1 5
<210> 23
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 23
Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
1 5 10 15
Pro Leu
<210> 24
<211> 18
<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 24
Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr Asn Leu Val Lys Asn
1 5 10 15
Lys Cys
<210> 25
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 25
Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln Gln Phe Gly Arg Asp
1 5 10 15
Ile Ala
<210> 26
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 26
Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
1 5 10 15
Val Lys
<210> 27
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 27
Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln Ile Tyr Lys Thr Pro
1 5 10 15
Pro Ile
<210> 28
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 28
Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
1 5 10 15
Asn Phe
<210> 29
<211> 18

<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 29
Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser Gln Ile Leu Pro Asp
1 5 10 15
Pro Ser
<210> 30
<211> 18
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 30
Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu Asp Leu Leu Phe Asn
1 5 10 15
Lys Val
<210> 31
<211> 14
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 31
Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser
1 5 10
<210> 32
<211> 8
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 32
Thr Asn Gly Thr Lys Arg Phe Asp
1 5
<210> 33
<211> 5
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 33
Ala Ser Thr Glu Lys
1 5
<210> 34
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 34
Leu Asp Ser Lys Thr Gln
1 5
<210> 35
<211> 10
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 35
Tyr Tyr His Lys Asn Asn Lys Ser Trp Met
1 5 10
<210> 36
<211> 4
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 36
Ser Glu Phe Arg
1
<210> 37
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 37
Ile Tyr Ser Lys His Thr
1 5
<210> 38
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 38
Thr Pro Gly Asp Ser Ser
1 5
<210> 39
<211> 7
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 39
Lys Tyr Asn Glu Asn Gly Thr
1 5
<210> 40

<210> 10
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 40
Gln Thr Ser Asn Phe Arg Val Gln Pro Thr
1 5 10
<210> 41
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 41
Ala Trp Asn Arg Lys Arg
1 5
<210> 42
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 42
Asn Ser Asn Asn Leu Asp
1 5
<210> 43
<211> 8
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 43
Leu Lys Pro Phe Glu Arg Asp Ile
1 5
<210> 44
<211> 5
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 44
Gly Phe Gln Pro Thr
1 5
<210> 45
<211> 4
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 45

Tyr Gln Pro Tyr
1
<210> 46
<211> 4
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 46
Pro Lys Lys Ser
1
<210> 47
<211> 5
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 47
Glu Ser Asn Lys Lys
1 5
<210> 48
<211> 13
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 48
Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro Gln Thr
1 5 10
<210> 49
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 49
Gly Thr Asn Thr Ser Asn
1 5
<210> 50
<211> 10
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 50
Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr
1 5 10
<210> 51
<211> 6

<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 51
His Val Asn Asn Ser Tyr
1 5
<210> 52
<211> 12
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 52
Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg
1 5 10
<210> 53
<211> 5
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 53
Ser Met Thr Lys Thr
1 5
<210> 54
<211> 7
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 54
Glu Gln Asp Lys Asn Thr Gln
1 5
<210> 55
<211> 10
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 55
Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe
1 5 10
<210> 56
<211> 4
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 56
Met Ala Tyr Arg

1

<210> 57

<211> 7

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 57

Asn Val Leu Tyr Glu Asn Gln

1 5

<210> 58

<211> 4

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 58

Gln Ser Lys Arg

1

<210> 59

<211> 5

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 59

Phe Pro Gln Ser Ala

1 5

<210> 60

<211> 9

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 60

Val Pro Ala Gln Glu Lys Asn Phe Thr

1 5

<210> 61

<211> 9

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 61

Lys Tyr Phe Lys Asn His Thr Ser Pro

1 5

<210> 62

<211> 8

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 62
Ile Gln Lys Glu Ile Asp Arg Leu
1 5
<210> 63
<211> 6
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 63
Phe Asp Glu Asp Asp Ser
1 5
<210> 64
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> B细胞表位
<400> 64
Thr Thr Lys Arg
1
<210> 65
<211> 19
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 65
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu
<210> 66
<211> 35
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 66
Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys
1 5 10 15
Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly
20 25 30
Tyr Gln Pro
35

<210> 67
<211> 32
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 67
Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu
1 5 10 15
Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys
20 25 30
<210> 68
<211> 11
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 68
Ser Phe Ile Glu Asp Leu Leu Phe Asn Lys Val
1 5 10
<210> 69
<211> 28
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 69
Ala Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly
1 5 10 15
Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr
20 25
<210> 70
<211> 16
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 70
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe Cys
1 5 10 15
<210> 71
<211> 13
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 71
Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1 5 10

<210> 72
<211> 79
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 72

```
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1               5                  10                  15
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
            20                  25                  30
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
        35                  40                  45
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
    50                  55                  60
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
65                  70                  75
```

<210> 73
<211> 74
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 73

```
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1               5                  10                  15
Tyr Asn Tyr Leu Tyr Arg Leu Lys Met Ser Glu Cys Val Leu Gly Gln
            20                  25                  30
Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gly
        35                  40                  45
Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn
    50                  55                  60
Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
65                  70
```

<210> 74
<211> 66
<212> PRT
<213> 人工序列
<220>

<223> S蛋白支架序列
<400> 74
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Leu Gly Ser Gly Ser Gly Gln Ala Gly Ser
20 25 30
Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln
35 40 45
Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg
50 55 60
Val Val
65
<210> 75
<211> 79
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 75
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Cys Phe Arg Lys Ser Asn Leu Lys Pro Phe
20 25 30
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Gly
35 40 45
Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly
50 55 60
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
65 70 75
<210> 76
<211> 74
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 76
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Cys Val Leu Gly Gln

20 25 30
Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gly
35 40 45
Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn
50 55 60
Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
65 70
<210> 77
<211> 66
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 77
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Cys Gly Ser Gly Ser Gly Gln Ala Gly Ser
20 25 30
Thr Pro Gly Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys Leu Gln
35 40 45
Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg
50 55 60
Val Val
65
<210> 78
<211> 84
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 78
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Phe Arg Lys Ser Asn Leu Lys
20 25 30
Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr
35 40 45
Pro Gly Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser
50 55 60

Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val
65 70 75 80
Val Arg Arg Arg
<210> 79
<211> 79
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 79
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Cys Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe Gln Pro
50 55 60
Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg Arg
65 70 75
<210> 80
<211> 71
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 80
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Gly Ser Gly Ser Gly Gln Ala
20 25 30
Gly Ser Thr Pro Gly Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys
35 40 45
Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60
Tyr Arg Val Val Arg Arg Arg
65 70
<210> 81
<211> 72

<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 81

Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile
20 25 30
Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu
35 40 45
Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr
50 55 60
Asn Gly Val Gly Tyr Gln Pro Tyr
65 70

<210> 82
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 82

Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Gly Ser Gly Ser Gly Ser Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
35 40 45
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr
50 55 60

<210> 83
<211> 59
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 83

Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15

Tyr Arg Leu Gly Ser Gly Ser Gly Gln Ala Gly Ser Thr Pro Cys Asn
20 25 30
Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe
35 40 45
Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr
50 55

<210> 84
<211> 58
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 84
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Gly Ser Gly Ser Gln Ala Gly Ser Thr Pro Cys Asn Gly
20 25 30
Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln
35 40 45
Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr
50 55

<210> 85
<211> 57
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 85
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Gly Ser Gly Gln Ala Gly Ser Thr Pro Cys Asn Gly Val
20 25 30
Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro
35 40 45
Thr Asn Gly Val Gly Tyr Gln Pro Tyr
50 55

<210> 86
<211> 67
<212> PRT

<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 86
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
20 25 30
Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gly Phe Asn Cys Tyr
35 40 45
Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr
50 55 60
Gln Pro Tyr
65
<210> 87
<211> 69
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 87
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Phe Arg Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
20 25 30
Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gly Phe Asn
35 40 45
Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val
50 55 60
Gly Tyr Gln Pro Tyr
65
<210> 88
<211> 71
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 88
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu

1 5 10 15
Tyr Arg Leu Phe Arg Lys Ser Lys Met Ser Glu Cys Val Leu Gly Gln
20 25 30
Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gly
35 40 45
Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn
50 55 60
Gly Val Gly Tyr Gln Pro Tyr
65 70

<210> 89
<211> 72
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 89

Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Phe Arg Lys Ser Asn Lys Met Ser Glu Cys Val Leu Gly
20 25 30
Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu Ala
35 40 45
Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr
50 55 60
Asn Gly Val Gly Tyr Gln Pro Tyr
65 70

<210> 90
<211> 73
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 90

Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Met Ser Glu Cys Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45

Ala Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro
50 55 60
Thr Asn Gly Val Gly Tyr Gln Pro Tyr
65 70
<210> 91
<211> 74
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 91
Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu
1 5 10 15
Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Lys Met Ser Glu Cys Val
20 25 30
Leu Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr
35 40 45
Leu Ala Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln
50 55 60
Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr
65 70
<210> 92
<211> 79
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 92
Val Ile Ala Trp Asn Ser Arg Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Lys Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
20 25 30
Glu Arg Asp Ile Ser Asn Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
35 40 45
Asn Gly Val Pro Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
50 55 60
Phe Gln Pro Thr Thr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val
65 70 75
<210> 93

<211> 81
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 93

Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Cys Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gln Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe
50 55 60
Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg
65 70 75 80
Arg

<210> 94
<211> 84
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 94

Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Phe Arg Lys Ser Asn Leu Lys
20 25 30
Pro Phe Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr
35 40 45
Pro Cys Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser
50 55 60
Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val
65 70 75 80
Val Arg Arg Arg

<210> 95
<211> 85
<212> PRT
<213> 人工序列

<220>
<223> S蛋白支架序列
<400> 95
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Cys Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gln Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe
50 55 60
Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg
65 70 75 80
Arg Glu Pro Glu Ala
85
<210> 96
<211> 85
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 96
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Ser Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gln Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe
50 55 60
Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg
65 70 75 80
Arg Glu Pro Glu Ala
85
<210> 97
<211> 85
<212> PRT
<213> 人工序列
<220>

<223> S蛋白支架序列
<400> 97
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Lys Met Ser Glu Cys Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gln Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe
50 55 60
Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg
65 70 75 80
Arg Glu Pro Glu Ala
85
<210> 98
<211> 101
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 98
Glu Val Glu Val Glu Phe Glu Val Glu Val Ile Ala Trp Asn Ser Asn
1 5 10 15
Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu
20 25 30
Phe Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser Gly
35 40 45
Ser Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn
50 55 60
Ser Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val
65 70 75 80
Gly Tyr Gln Pro Tyr Arg Val Val Arg Val Arg Phe Arg Val Arg Val
85 90 95
Arg Glu Pro Glu Ala
100
<210> 99
<211> 101
<212> PRT
<213> 人工序列

<220>
<223> S蛋白支架序列
<400> 99

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

<210> 100
<211> 112
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 100

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

<210> 101

<211> 112
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 101

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

<210> 102
<211> 104
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 102

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

Tyr Arg Arg Arg Glu Pro Glu Ala
100

<210> 103
<211> 93
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 103

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

<210> 104
<211> 107
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 104

Cys Glu Val Glu Val Glu Phe Glu Val Glu Val Ile Ala Trp Asn Ser
1 5 10 15
Arg Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Lys Tyr Arg
20 25 30
Leu Phe Lys Leu Trp Ala Gln Cys Val Gln Leu Tyr Leu Gln Pro Arg
35 40 45
Thr Phe Leu Leu Leu Leu Tyr Asp Ala Asn Tyr Phe Leu Asn Glu Ile
50 55 60
Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Ser
65 70 75 80
Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Thr Gly Val Gly

85 90 95

Tyr Gln Pro Tyr Arg Arg Arg Glu Pro Glu Ala
100 105

<210> 105
<211> 96
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 105

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

<210> 106
<211> 110
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 106

Glu Val Glu Val Glu Phe Glu Val Glu Val Ile Ala Trp Asn Ser Asn
1 5 10 15
Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu
20 25 30
Phe Lys Met Ser Glu Ser Val Leu Gly Gln Ser Lys Arg Val Gln Ala
35 40 45
Leu Leu Phe Asn Lys Val Thr Leu Ala Gln Tyr Gln Ala Gly Ser Thr
50 55 60
Pro Cys Asn Gly Val Glu Gly Phe Asn Ser Tyr Phe Pro Leu Gln Ser
65 70 75 80

Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val
85 90 95
Val Arg Val Arg Phe Arg Val Arg Val Arg Glu Pro Glu Ala
100 105 110

<210> 107
<211> 110
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 107

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

<210> 108
<211> 102
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 108

Cys Glu Val Glu Val Glu Phe Glu Val Glu Val Ile Ala Trp Asn Ser
1 5 10 15
Arg Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Lys Tyr Arg
20 25 30
Leu Phe Lys Met Ser Glu Ser Val Leu Gly Gln Ser Lys Arg Val Gln
35 40 45
Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gln Tyr Gln Ala Gly Ser

50 55 60

Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Ser Tyr Phe Pro Leu Gln
65 70 75 80

Ser Tyr Gly Phe Gln Pro Thr Thr Gly Val Gly Tyr Gln Pro Tyr Arg
85 90 95

Arg Arg Glu Pro Glu Ala
100

<210> 109
<211> 105
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 109

Cys Glu Val Glu Val Glu Phe Glu Val Glu Val Ile Ala Trp Asn Ser
1 5 10 15

Arg Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Lys Tyr Arg
20 25 30

Leu Phe Lys Met Ser Glu Ser Val Leu Gly Gln Ser Lys Arg Val Gln
35 40 45

Ala Leu Leu Phe Asn Lys Val Thr Leu Ala Gln Asn Glu Ile Tyr Gln
50 55 60

Ala Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Ser Tyr Phe
65 70 75 80

Pro Leu Gln Ser Tyr Gly Phe Gln Pro Thr Thr Gly Val Gly Tyr Gln
85 90 95

Pro Tyr Arg Arg Arg Glu Pro Glu Ala
100 105

<210> 110
<211> 75
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 110

Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15

Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Gly Ser Gly Ser Gly Gln Ala
20 25 30

Gly Ser Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Gly Tyr Phe Cys
35 40 45
Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60
Tyr Arg Val Val Arg Arg Arg Glu Pro Glu Ala
65 70 75
<210> 111
<211> 79
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 111
Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Cys Lys Met Ser Glu Ser Val Leu
20 25 30
Gly Gln Ser Lys Arg Val Gln Ala Leu Leu Phe Asn Lys Val Thr Leu
35 40 45
Ala Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe Gln Pro
50 55 60
Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Arg Arg Arg
65 70 75
<210> 112
<211> 66
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 112
Val Lys Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Leu Gly Ser Gly Ser Gly Gln Ala Gly Ser
20 25 30
Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln
35 40 45
Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg
50 55 60
Val Val

65

<210> 113

<211> 66

<212> PRT

<213> 人工序列

<220>

<223> S蛋白支架序列

<400> 113

Val Ile Lys Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
1 5 10 15
Tyr Asn Tyr Leu Tyr Arg Leu Gly Ser Gly Ser Gly Gln Ala Gly Ser
20 25 30
Thr Pro Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln
35 40 45
Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg
50 55 60
Val Val
65

<210> 114

<211> 57

<212> PRT

<213> 人工序列

<220>

<223> S蛋白支架序列

<400> 114

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Gly Ser Gly Ser Gly Gln Ala Gly Ser Thr Pro Cys Asn Gly
20 25 30
Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln
35 40 45
Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 115

<211> 57

<212> PRT

<213> 人工序列

<220>

<223> S蛋白支架序列

<400> 115

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Cys Gly Ser Gly Ser Gly Gln Ala Gly Ser Thr Pro Gly Asn Gly
20 25 30
Val Glu Gly Phe Asn Gly Tyr Phe Cys Leu Gln Ser Tyr Gly Phe Gln
35 40 45
Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 116
<211> 59
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 116

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Asp Gly Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
35 40 45
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 117
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 117

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Asn Ala Asn Asp Glu Ile Tyr Gln Ala Gly Ser Thr Pro
20 25 30
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
35 40 45
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60

<210> 118
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 118

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Asn Ala His Asp Lys Ile Tyr Gln Ala Gly Ser Thr Pro
20 25 30
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
35 40 45
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60

<210> 119
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 119

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Asn Ala Asn Asp Lys Ile Tyr Gln Ala Gly Ser Thr Pro
20 25 30
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
35 40 45
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60

<210> 120
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 120

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15

Arg Leu Phe Asp Ala His Asp Lys Ile Tyr Gln Ala Gly Ser Thr Pro
20 25 30
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
35 40 45
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60

<210> 121
<211> 57
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 121

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Lys Pro Glu Gln Ala Gly Ser Thr Pro Cys Asn Gly
20 25 30
Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln
35 40 45
Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 122
<211> 59
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 122

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Gly Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
35 40 45
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 123
<211> 59
<212> PRT

<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 123
Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Ala Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
35 40 45
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55
<210> 124
<211> 59
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 124
Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Lys Pro Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
35 40 45
Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55
<210> 125
<211> 59
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 125
Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Gly Thr Asp Ile Tyr Gln Ala Gly Ser Thr Pro Cys
20 25 30
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly

35 40 45

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55

<210> 126
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<400> 126

Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr Leu Tyr
1 5 10 15
Arg Leu Phe Pro Ala His Asp Lys Ile Tyr Gln Ala Gly Ser Thr Pro
20 25 30
Cys Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr
35 40 45
Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
50 55 60

<210> 127
<211> 87
<212> PRT
<213> 人工序列
<220>
<223> S蛋白支架序列
<220>
<221> MISC_FEATURE
<222> (27)..(31)
<223> 每个残基是任何氨基酸
<220>
<221> MISC_FEATURE
<222> (32)..(56)
<223> 每个残基是任何氨基酸或不存在
<400> 127

Glu Glu Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly
1 5 10 15
Gly Asn Tyr Asn Tyr Leu Tyr Arg Leu Phe Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ala Gly Phe Asn Gly Tyr Phe Ser
50 55 60
Leu Gln Ser Tyr Gly Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro
65 70 75 80
Tyr Arg Val Val Arg Arg Arg
85

<210> 128
<211> 15
<212> PRT
<213> 人工序列
<220>
<223> 从S蛋白支架#1,4,7,10,和23的环
<400> 128
Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp Ile Ser Thr Glu
1 5 10 15

<210> 129
<211> 32
<212> PRT
<213> 人工序列
<220>
<223> 从S蛋白支架#5,8的环
<400> 129
Cys Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Gln Ala
1 5 10 15
Leu Leu Phe Asn Lys Val Thr Leu Ala Gly Phe Asn Gly Tyr Phe Cys
20 25 30

<210> 130
<211> 32
<212> PRT
<213> 人工序列
<220>
<223> 从S蛋白支架#40的环
<400> 130
Cys Lys Met Ser Glu Ser Val Leu Gly Gln Ser Lys Arg Val Gln Ala
1 5 10 15
Leu Leu Phe Asn Lys Val Thr Leu Ala Gly Phe Asn Gly Tyr Phe Cys
20 25 30

<210> 131
<211> 22

<212> PRT
<213> 人工序列
<220>
<223> 肽
<400> 131
Thr Phe Gly Asn Pro Val Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala
1 5 10 15
Ala Thr Glu Lys Ser Asn
20
<210> 132
<211> 23
<212> PRT
<213> 人工序列
<220>
<223> 肽
<400> 132
Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro Ala Gln Asp Ile
1 5 10 15
Trp Gly Thr Ser Ala Ala Ala
20
<210> 133
<211> 21
<212> PRT
<213> 人工序列
<220>
<223> 肽
<400> 133
Ile Ser Pro Cys Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr
1 5 10 15
Asn Ala Ser Ser Glu
20
<210> 134
<211> 21
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 134
Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp
1 5 10 15
Lys Tyr Phe Lys Asn

20

<210> 135
<211> 33
<212> PRT
<213> 人工序列
<220>
<223> 肽
<400> 135

Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile Glu Asp Leu
1 5 10 15
Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met Lys Gln Tyr
20 25 30
Gly

<210> 136
<211> 29
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 136

Ala Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly
1 5 10 15
Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His
20 25

<210> 137
<211> 1270
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 137

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

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr
100 105 110
Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile Lys Val
115 120 125
Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr His Lys
130 135 140
Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala
145 150 155 160
Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu
165 170 175
Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys
180 185 190
Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Asn
195 200 205
Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu Pro Leu Val
210 215 220
Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala
225 230 235 240
Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr
245 250 255
Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro Arg Thr Phe
260 265 270
Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala Val Asp Cys
275 280 285
Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr
290 295 300
Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr
305 310 315 320
Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe Gly
325 330 335
Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg
340 345 350
Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser
355 360 365
Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu
370 375 380
Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg
385 390 395 400
Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Lys Ile Ala

405 410 415
Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala
420 425 430
Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr
435 440 445
Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp
450 455 460
Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val
465 470 475 480
Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro
485 490 495
Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe
500 505 510
Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr
515 520 525
Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr
530 535 540
Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln
545 550 555 560
Gln Phe Gly Arg Asp Ile Asp Asp Thr Thr Asp Ala Val Arg Asp Pro
565 570 575
Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly Val
580 585 590
Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val Ala Val Leu
595 600 605
Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp
610 615 620
Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe
625 630 635 640
Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser
645 650 655
Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln
660 665 670
Thr Gln Thr Asn Ser His Arg Arg Ala Arg Ser Val Ala Ser Gln Ser
675 680 685
Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr
690 695 700
Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile Ser Val Thr
705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr
725 730 735
Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln
740 745 750
Tyr Gly Ser Phe Cys Ile Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala
755 760 765
Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln
770 775 780
Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser
785 790 795 800
Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu
805 810 815
Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys
820 825 830
Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys
835 840 845
Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
850 855 860
Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr
865 870 875 880
Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
885 890 895
Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
900 905 910
Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
915 920 925
Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys
930 935 940
Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945 950 955 960
Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
965 970 975
Ile Leu Ala Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp Arg
980 985 990
Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln
995 1000 1005
Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala
1010 1015 1020
Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp

1025 1030 1035
Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1040 1045 1050
Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln
1055 1060 1065
Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys
1070 1075 1080
Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His
1085 1090 1095
Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr
1100 1105 1110
Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly
1115 1120 1125
Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp
1130 1135 1140
Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser
1145 1150 1155
Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val
1160 1165 1170
Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys
1175 1180 1185
Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
1190 1195 1200
Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile
1205 1210 1215
Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys
1220 1225 1230
Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly
1235 1240 1245
Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys
1250 1255 1260
Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 138
<211> 1273
<212> PRT
<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)
<400> 138
Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15
Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30
Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45
His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80
Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95
Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110
Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125
Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140
Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160
Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175
Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190
Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205
Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220
Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255
Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270
Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285
Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300
Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335
Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
370 375 380
Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385 390 395 400
Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Asn Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540
Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605
Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620
His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser

625 630 635 640
Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
645 650 655
Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
660 665 670
Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685
Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Val Glu Asn Ser
690 695 700
Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720
Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750
Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765
Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780
Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800
Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815
Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830
Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845
Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895
Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910
Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925
Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960
Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975
Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
980 985 990
Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005
Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
1010 1015 1020
Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
1025 1030 1035
Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
1040 1045 1050
Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
1055 1060 1065
Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080
Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095
Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110
Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125
Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140
Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155
His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170
Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185
Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200
Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215
Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230
Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys

1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro

1250 1255 1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr

1265 1270

<210> 139

<211> 1273

<212> PRT

<213> 严重急性呼吸综合征冠状病毒2(SARS-CoV-2)

<400> 139

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val

1 5 10 15

Asn Phe Thr Asn Arg Thr Gln Leu Pro Ser Ala Tyr Thr Asn Ser Phe

20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu

35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp

65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu

85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser

100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile

115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Tyr Pro Phe Leu Gly Val Tyr

130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr

145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu

165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Ser Glu Phe

180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu

210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr

225 230 235 240
Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255
Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270
Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
275 280 285
Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
290 295 300
Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305 310 315 320
Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
325 330 335
Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350
Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365
Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
370 375 380
Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385 390 395 400
Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415
Thr Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430
Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445
Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480
Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495
Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510
Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560
Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575
Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
595 600 605
Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
610 615 620
His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625 630 635 640
Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu Tyr Val
645 650 655
Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
660 665 670
Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685
Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700
Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720
Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735
Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750
Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765
Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780
Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800
Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815
Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830
Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845
Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu

850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
930 935 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945 950 955 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
965 970 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
980 985 990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
1010 1015 1020

Leu Ala Ala Ile Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
1025 1030 1035

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
1040 1045 1050

Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
1055 1060 1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170
Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185
Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200
Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215
Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230
Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245
Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
1250 1255 1260
Val Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 140
<211> 805
<212> PRT
<213> 智人(Homo sapiens)
<300>
<308> NP_001358344.1
<309> 2021-02-20
<313> (1)..(805)
<400> 140

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

Ser Lys Arg Leu Asn Thr Ile Leu Asn Thr Met Ser Thr Ile Tyr Ser
115 120 125
Thr Gly Lys Val Cys Asn Pro Asp Asn Pro Gln Glu Cys Leu Leu Leu
130 135 140
Glu Pro Gly Leu Asn Glu Ile Met Ala Asn Ser Leu Asp Tyr Asn Glu
145 150 155 160
Arg Leu Trp Ala Trp Glu Ser Trp Arg Ser Glu Val Gly Lys Gln Leu
165 170 175
Arg Pro Leu Tyr Glu Glu Tyr Val Val Leu Lys Asn Glu Met Ala Arg
180 185 190
Ala Asn His Tyr Glu Asp Tyr Gly Asp Tyr Trp Arg Gly Asp Tyr Glu
195 200 205
Val Asn Gly Val Asp Gly Tyr Asp Tyr Ser Arg Gly Gln Leu Ile Glu
210 215 220
Asp Val Glu His Thr Phe Glu Glu Ile Lys Pro Leu Tyr Glu His Leu
225 230 235 240
His Ala Tyr Val Arg Ala Lys Leu Met Asn Ala Tyr Pro Ser Tyr Ile
245 250 255
Ser Pro Ile Gly Cys Leu Pro Ala His Leu Leu Gly Asp Met Trp Gly
260 265 270
Arg Phe Trp Thr Asn Leu Tyr Ser Leu Thr Val Pro Phe Gly Gln Lys
275 280 285
Pro Asn Ile Asp Val Thr Asp Ala Met Val Asp Gln Ala Trp Asp Ala
290 295 300
Gln Arg Ile Phe Lys Glu Ala Glu Lys Phe Phe Val Ser Val Gly Leu
305 310 315 320
Pro Asn Met Thr Gln Gly Phe Trp Glu Asn Ser Met Leu Thr Asp Pro
325 330 335
Gly Asn Val Gln Lys Ala Val Cys His Pro Thr Ala Trp Asp Leu Gly
340 345 350
Lys Gly Asp Phe Arg Ile Leu Met Cys Thr Lys Val Thr Met Asp Asp
355 360 365
Phe Leu Thr Ala His His Glu Met Gly His Ile Gln Tyr Asp Met Ala
370 375 380
Tyr Ala Ala Gln Pro Phe Leu Leu Arg Asn Gly Ala Asn Glu Gly Phe
385 390 395 400
His Glu Ala Val Gly Glu Ile Met Ser Leu Ser Ala Ala Thr Pro Lys
405 410 415
His Leu Lys Ser Ile Gly Leu Leu Ser Pro Asp Phe Gln Glu Asp Asn

420 425 430

Glu Thr Glu Ile Asn Phe Leu Leu Lys Gln Ala Leu Thr Ile Val Gly

435 440 445

Thr Leu Pro Phe Thr Tyr Met Leu Glu Lys Trp Arg Trp Met Val Phe

450 455 460

Lys Gly Glu Ile Pro Lys Asp Gln Trp Met Lys Lys Trp Trp Glu Met

465 470 475 480

Lys Arg Glu Ile Val Gly Val Val Glu Pro Val Pro His Asp Glu Thr

485 490 495

Tyr Cys Asp Pro Ala Ser Leu Phe His Val Ser Asn Asp Tyr Ser Phe

500 505 510

Ile Arg Tyr Tyr Thr Arg Thr Leu Tyr Gln Phe Gln Phe Gln Glu Ala

515 520 525

Leu Cys Gln Ala Ala Lys His Glu Gly Pro Leu His Lys Cys Asp Ile

530 535 540

Ser Asn Ser Thr Glu Ala Gly Gln Lys Leu Phe Asn Met Leu Arg Leu

545 550 555 560

Gly Lys Ser Glu Pro Trp Thr Leu Ala Leu Glu Asn Val Val Gly Ala

565 570 575

Lys Asn Met Asn Val Arg Pro Leu Leu Asn Tyr Phe Glu Pro Leu Phe

580 585 590

Thr Trp Leu Lys Asp Gln Asn Lys Asn Ser Phe Val Gly Trp Ser Thr

595 600 605

Asp Trp Ser Pro Tyr Ala Asp Gln Ser Ile Lys Val Arg Ile Ser Leu

610 615 620

Lys Ser Ala Leu Gly Asp Lys Ala Tyr Glu Trp Asn Asp Asn Glu Met

625 630 635 640

Tyr Leu Phe Arg Ser Ser Val Ala Tyr Ala Met Arg Gln Tyr Phe Leu

645 650 655

Lys Val Lys Asn Gln Met Ile Leu Phe Gly Glu Glu Asp Val Arg Val

660 665 670

Ala Asn Leu Lys Pro Arg Ile Ser Phe Asn Phe Phe Val Thr Ala Pro

675 680 685

Lys Asn Val Ser Asp Ile Ile Pro Arg Thr Glu Val Glu Lys Ala Ile

690 695 700

Arg Met Ser Arg Ser Arg Ile Asn Asp Ala Phe Arg Leu Asn Asp Asn

705 710 715 720

Ser Leu Glu Phe Leu Gly Ile Gln Pro Thr Leu Gly Pro Pro Asn Gln

725 730 735

Pro Pro Val Ser Ile Trp Leu Ile Val Phe Gly Val Val Met Gly Val
740 745 750
Ile Val Val Gly Ile Val Ile Leu Ile Phe Thr Gly Ile Arg Asp Arg
755 760 765
Lys Lys Lys Asn Lys Ala Arg Ser Gly Glu Asn Pro Tyr Ala Ser Ile
770 775 780
Asp Ile Ser Lys Gly Glu Asn Asn Pro Gly Phe Gln Asn Thr Asp Asp
785 790 795 800
Val Gln Thr Ser Phe
805

<210> 141
<211> 50
<212> PRT
<213> 人工序列
<220>
<223> ACE2支架序列
<400> 141
Ser Thr Ile Glu Glu Gln Ala Lys Thr Phe Leu Asp Lys Phe Asn His
1 5 10 15
Glu Ala Glu Asp Leu Phe Tyr Gln Gly Ser Gly Ser Gly Asn Ala Gly
20 25 30
Asp Lys Trp Ser Ala Phe Leu Lys Glu Gln Ser Thr Leu Ala Gln Met
35 40 45
Tyr Pro
50

<210> 142
<211> 54
<212> PRT
<213> 人工序列
<220>
<223> ACE2支架序列
<400> 142
Ser Thr Ile Glu Glu Gln Ala Lys Thr Phe Leu Asp Lys Phe Asn His
1 5 10 15
Glu Ala Glu Asp Leu Phe Tyr Gln Gly Ser Gly Ser Gly Asn Ala Gly
20 25 30
Asp Lys Trp Ser Ala Phe Leu Lys Glu Gln Ser Thr Leu Ala Gln Met
35 40 45
Tyr Pro Glu Pro Glu Ala

50

<210> 143
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的正义链
<400> 143
gcugaugagu acgaacuuau guact 25
<210> 144
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的反义链
<400> 144
aguacauaag uucguacuca ucagcuu 27
<210> 145
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的正义链
<400> 145
ggaagagaca gguacguuaa uagtt 25
<210> 146
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的反义链
<400> 146
aacuauuaac guaccugucu cuuccga 27
<210> 147
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的正义链

<400> 147
caagcugaug aguacgaacu uaugt 25
<210> 148
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 用于siRNA设计的反义链
<400> 148
acauaaguuc guacucauca gcuugug 27
<210> 149
<211> 24
<212> PRT
<213> 智人
<400> 149
Ser Thr Ile Glu Glu Gln Ala Lys Thr Phe Leu Asp Lys Phe Asn His
1 5 10 15
Glu Ala Glu Asp Leu Phe Tyr Gln
20
<210> 150
<211> 21
<212> PRT
<213> 智人
<400> 150
Asn Ala Gly Asp Lys Trp Ser Ala Phe Leu Lys Glu Gln Ser Thr Leu
1 5 10 15
Ala Gln Met Tyr Pro
20
<210> 151
<211> 66
<212> PRT
<213> 人工序列
<220>
<223> ACE2支架序列
<400> 151
Ser Thr Ile Glu Glu Gln Ala Lys Thr Phe Leu Asp Lys Phe Asn His
1 5 10 15
Glu Ala Glu Asp Leu Phe Tyr Gln Ser Ser Leu Ala Ser Trp Asn Tyr
20 25 30

Asn Thr Asn Ile Thr Glu Glu Asn Val Gln Asn Met Asn Asn Ala Gly
35 40 45
Asp Lys Trp Ser Ala Phe Leu Lys Glu Gln Ser Thr Leu Ala Gln Met
50 55 60
Tyr Pro
65

<210> 152
<211> 174
<212> PRT
<213> 人工序列
<220>
<223> 环状肽
<400> 152

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

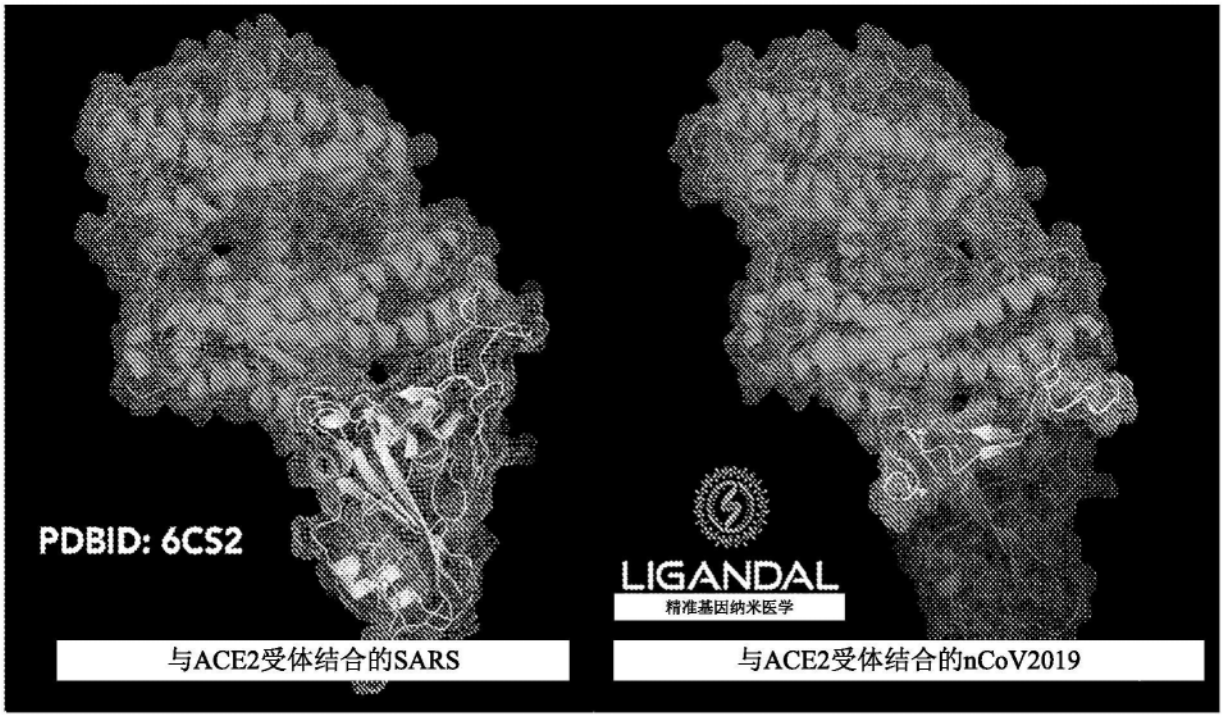

PDBID: 6CS2
LIGANDAL
精准基因纳米医学
与ACE2受体结合的SARS
与ACE2受体结合的nCoV2019

图1

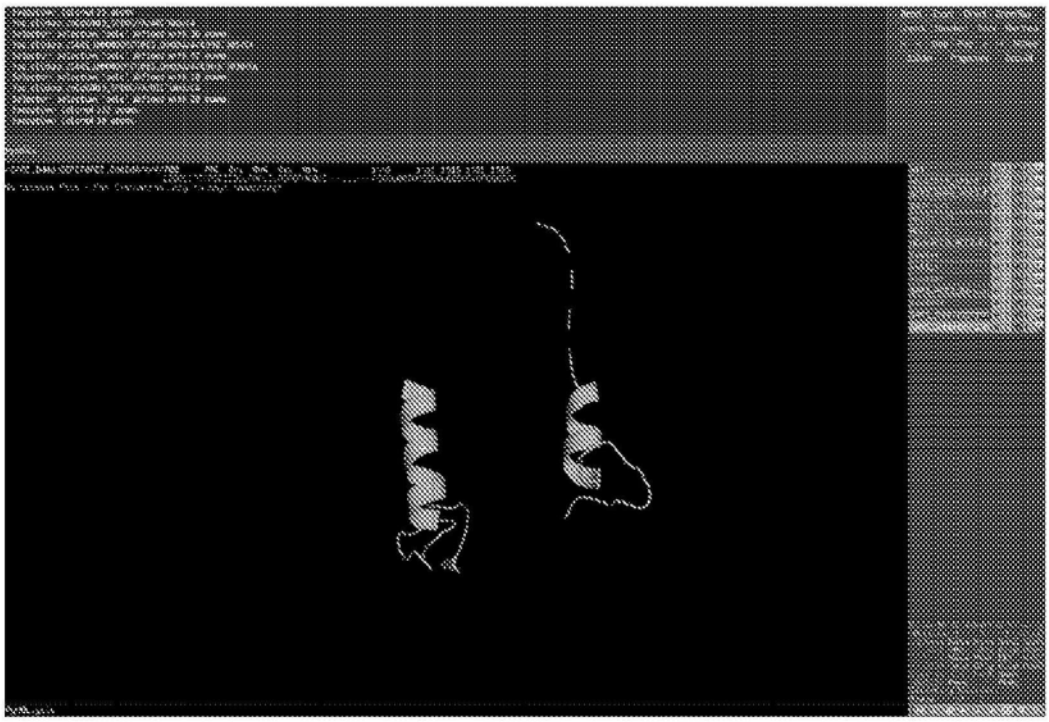


图2A

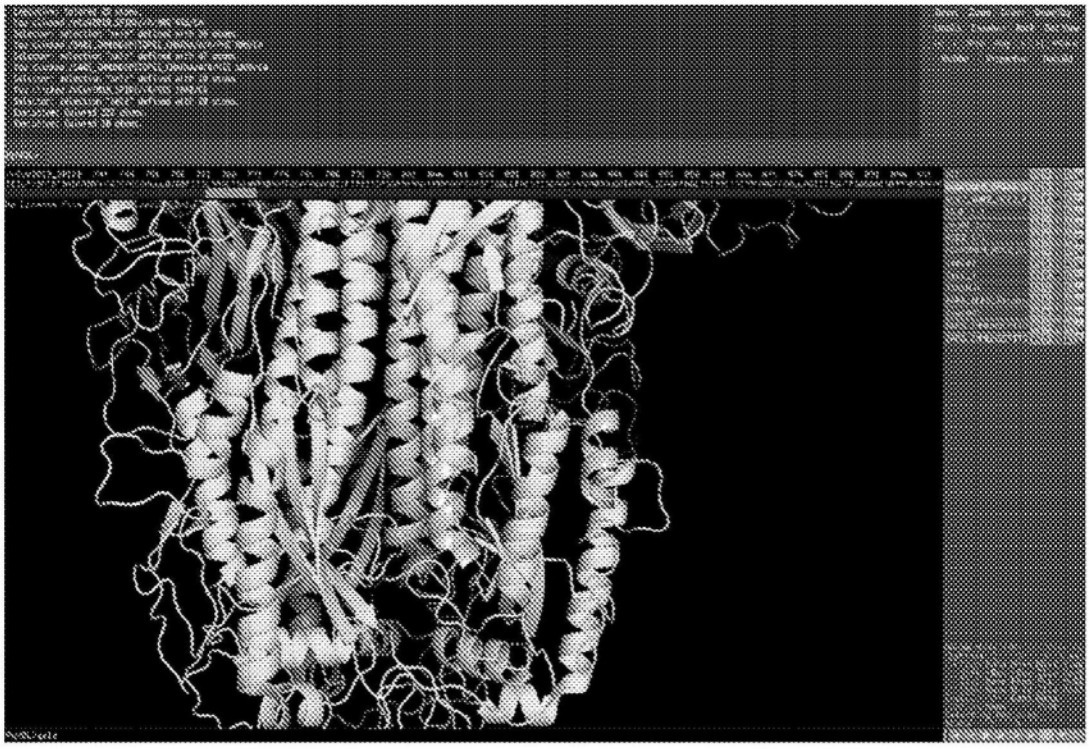


图2B

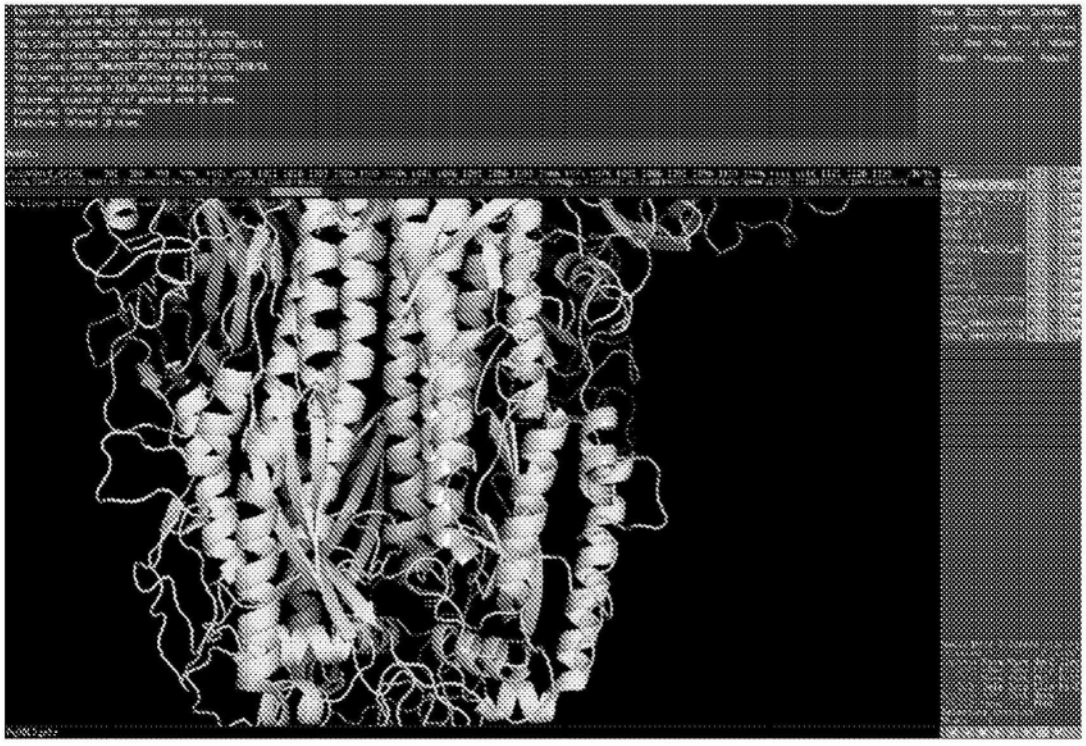


图2C

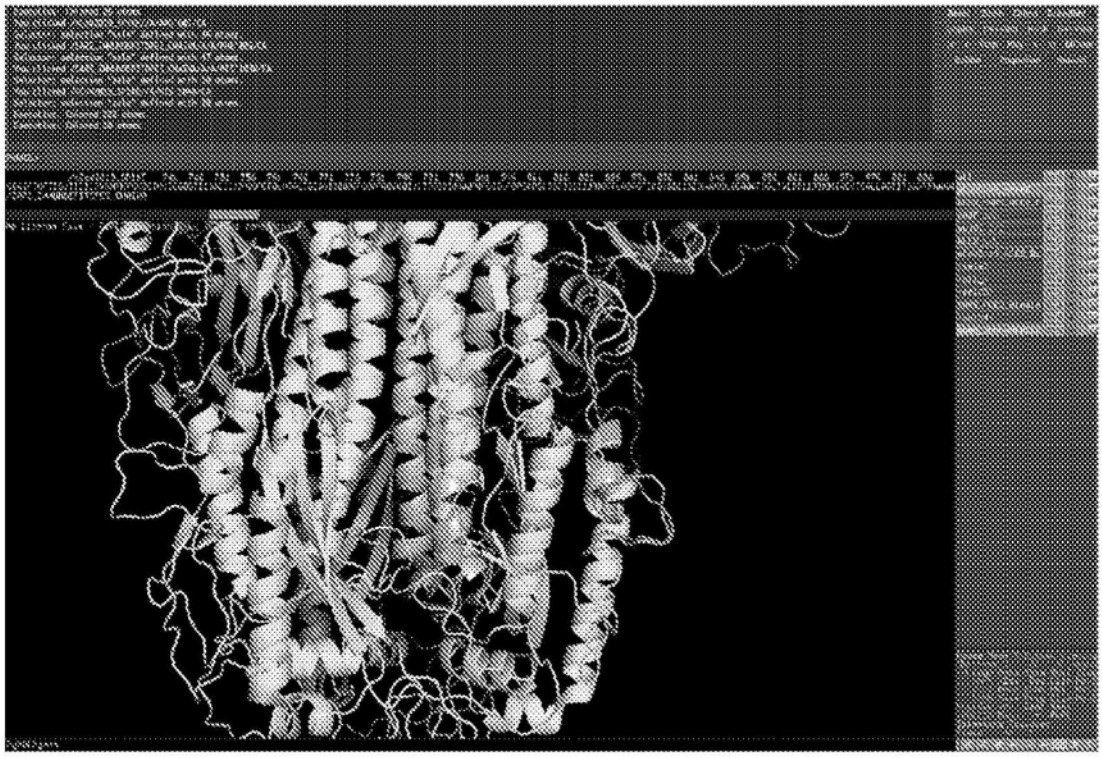


图2D

MHC-I结合预测结果

输入序列

	名称	序列
1	免疫表位1	QILPDPSKPSKRSFIEDLLFNKVTLADAGFIK
2	免疫表位2	ASANLAATKMSECVLGQSKRVDFCGKGY

预测方法：IEDB推荐2.22 | 低百分等级=好的结合

下载结果

引用

检查以扩大结果

等位基因		起始	结束	长度	肽	使用的方法	百分等级
HLA-A*02:01	2	9	21	13	KMSECVLGQSKRV	ann	0.9
HLA-A*02:01	1	18	26	9	LLFNKVTLA	共有序列 (ann/comblib_sidney2008/smm)	1.2
HLA-A*02:01	1	13	23	11	SFIEDLLFNKV	共有序列 (ann/smm)	1.4

图3

CVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSG
TNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE
FQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKN
LREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLT
PGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTV
EKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSV
LYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLP
DDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVE
GFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF
NFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVN
NSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTN
FTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKN
TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQY
GDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIP
FAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQN
AQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA
EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEK
NFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGI
VNNTV (SEQ ID NO:2,残基15-1137)

图4

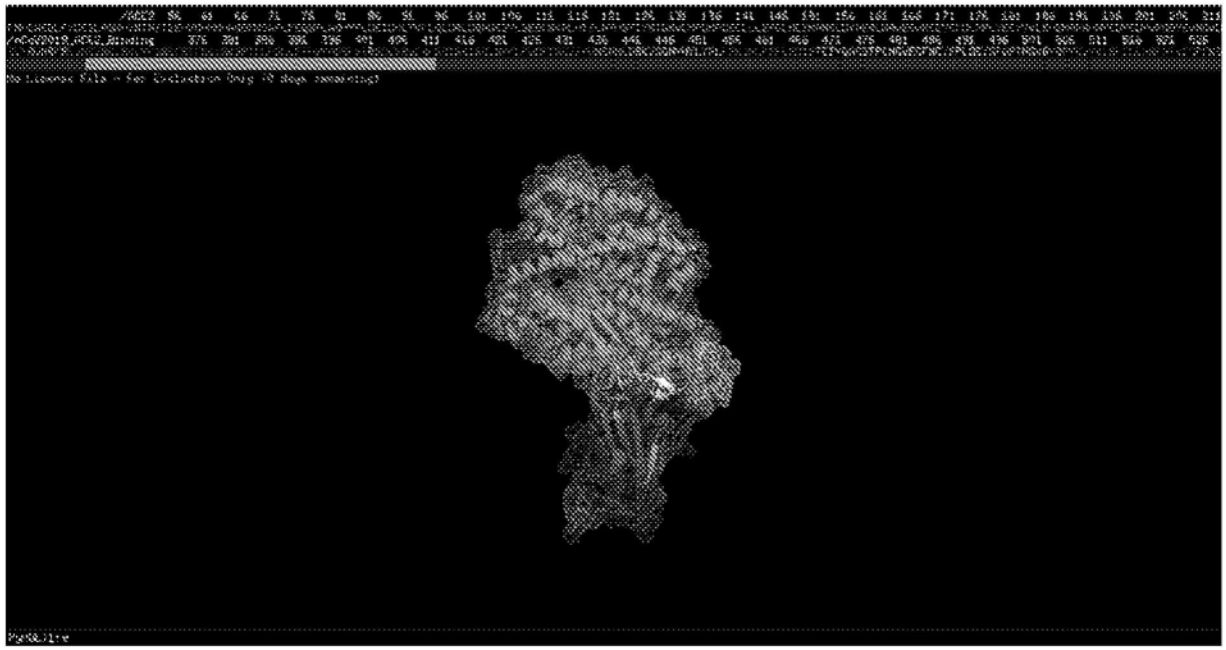


图5

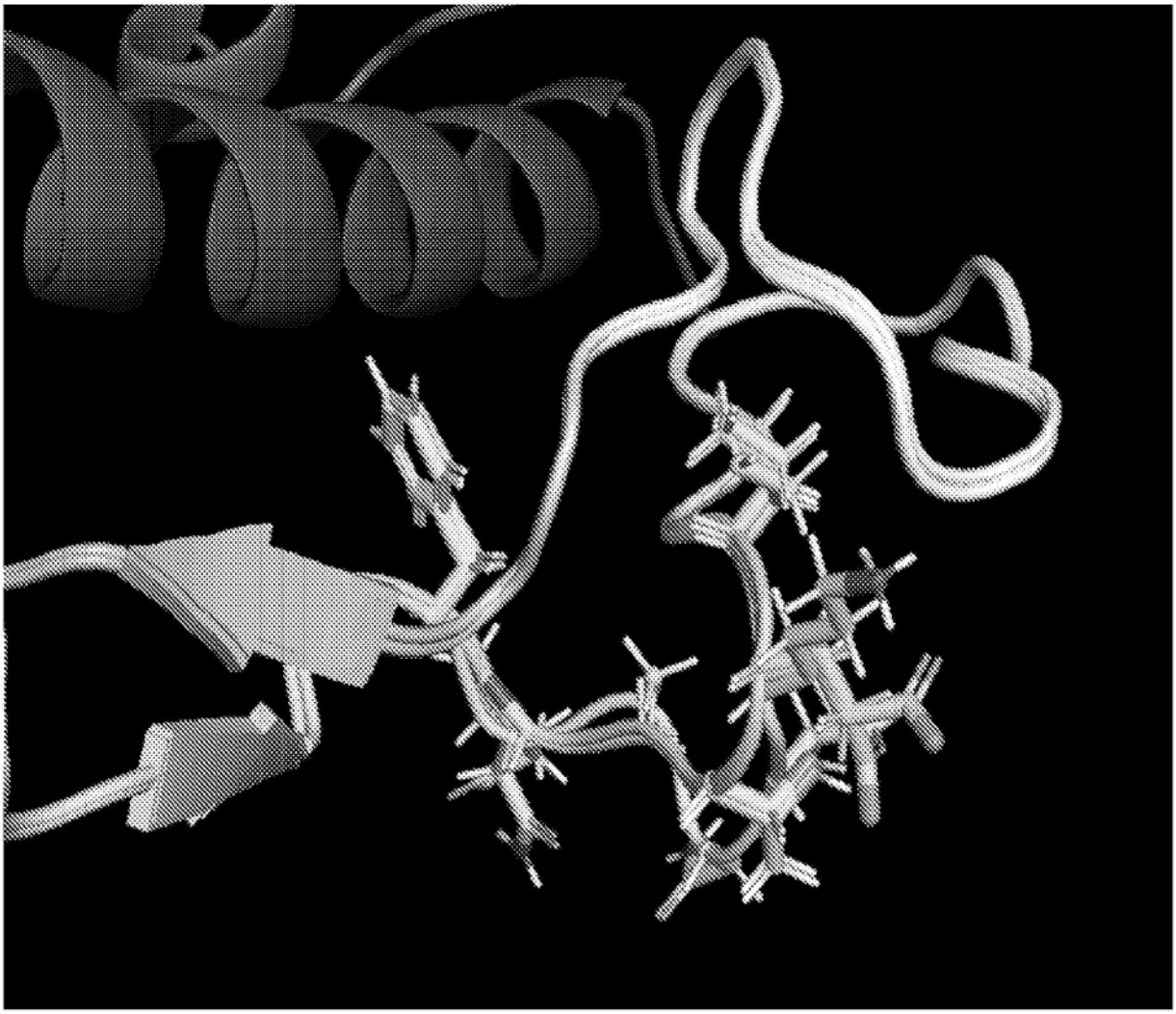


图6

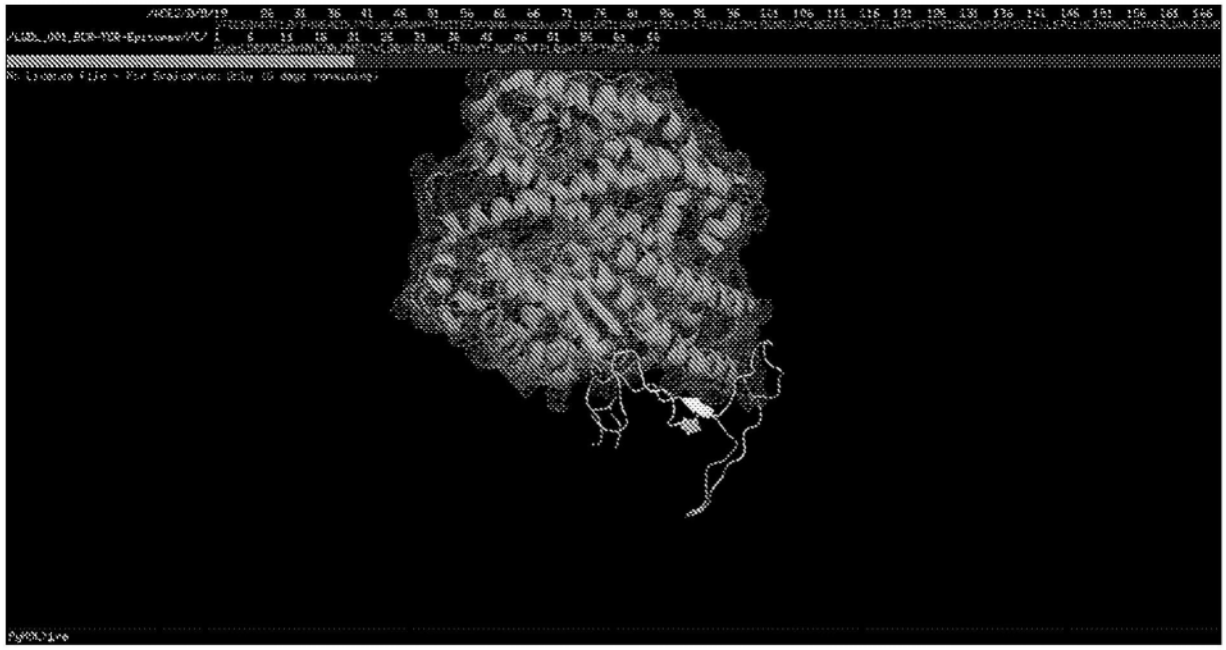


图7A

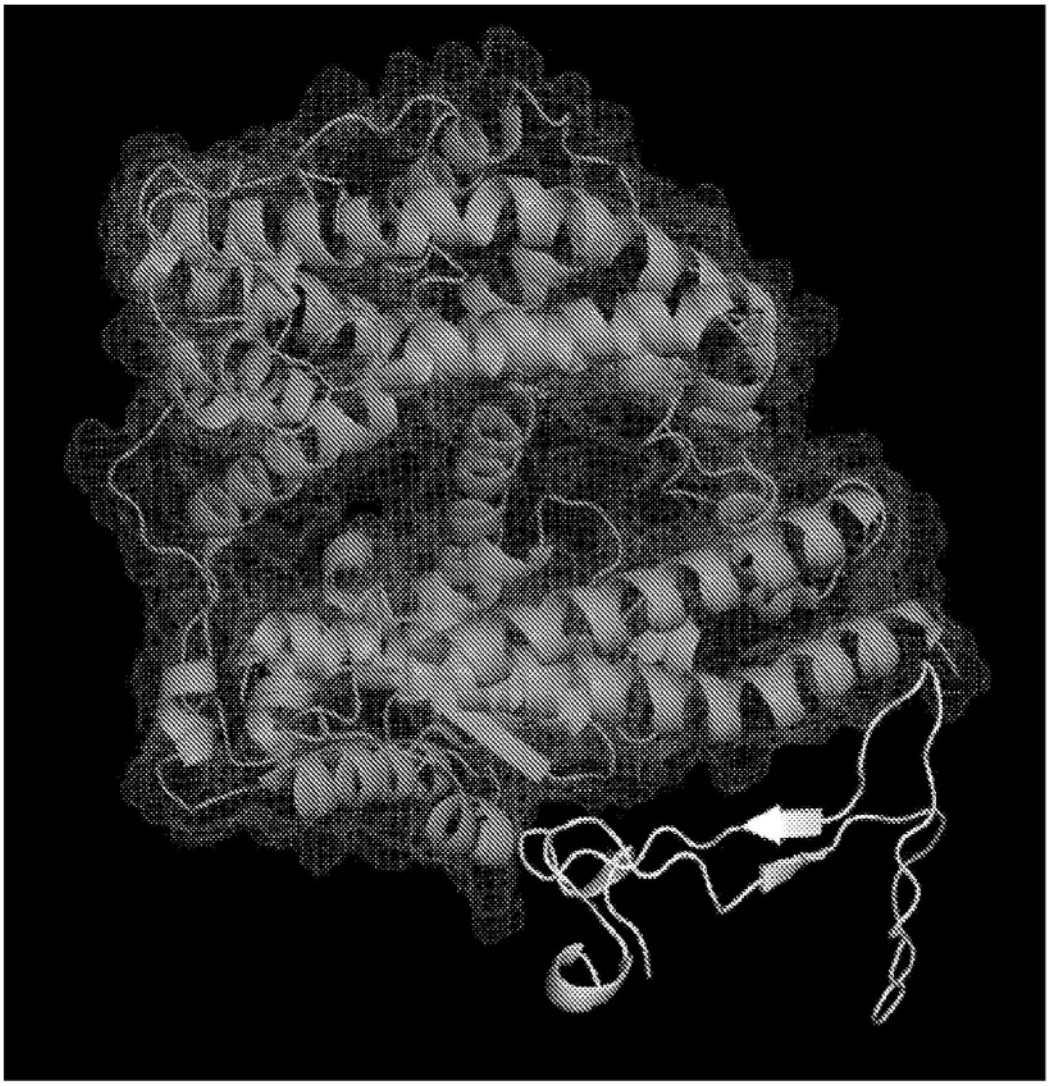


图7B

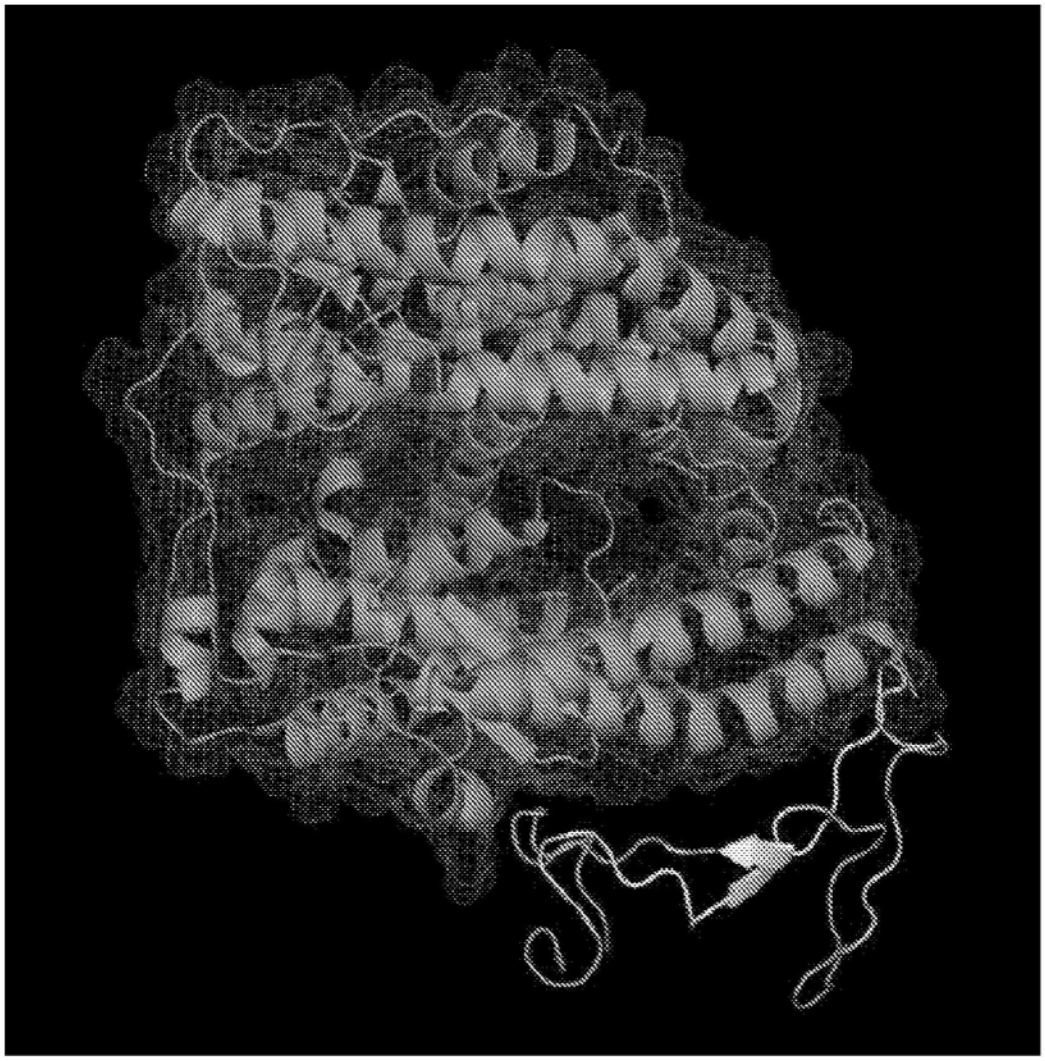
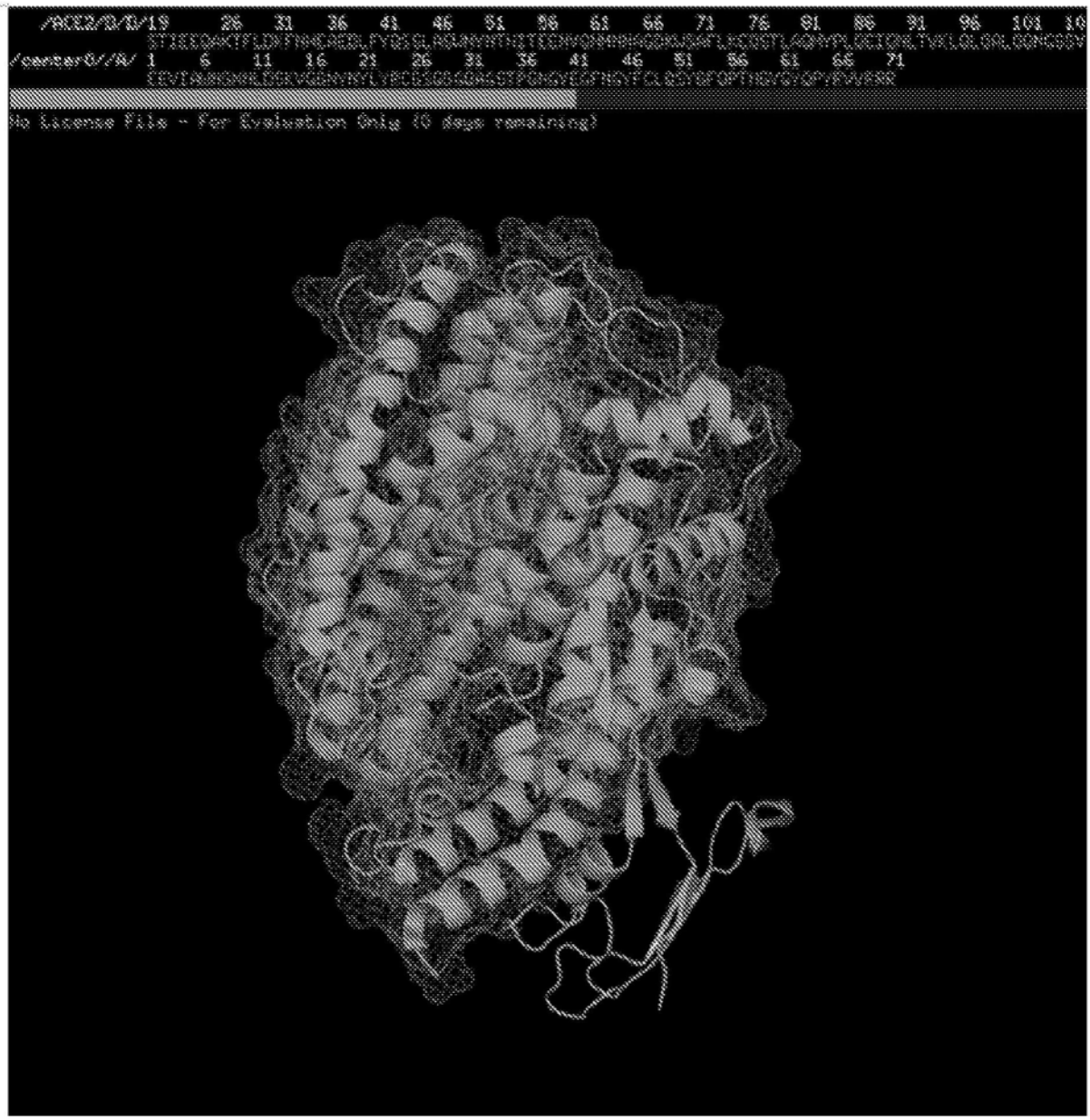


图7C



蛋白质-蛋白质复合物	ΔG (kcal mol^{-1})	K_d (M) at 37.0 ℃
ACE2与SEQ12_center0	-16.1	4.7E-12
ACE2与COVID刺突（Cryo-EM）		~15E-9

图8A


返回PRODIGY主页
您如何评价您使用门户网站的体验?
结合亲和力和K_D预测
结合亲和力(ΔG)和解离常数（K_d）预测值为:
蛋白质-蛋白质复合物
ΔG (kcal mol^{-1})
K_d (M) at 37.0 °C
ACE2与SEQ12_center0
-15.1
4.7E-12
预测细节
每个属性的相互作用接触数（ICs）:
ICs 带电荷-带电荷: 1
ICs 带电荷-极性: 24
ICs 带电荷-无极性: 45
ICs 极性-极性: 5
ICs 极性-无极性: 31
ICs 无极性-无极性: 28
每个属性的非相互作用表面（NIS）:
NIS 带电荷: 24.84%
NIS 无极性: 38.00%
界面处的ICs表:
界面处的残基-残基对列表（.txt）
下载输出

图8B

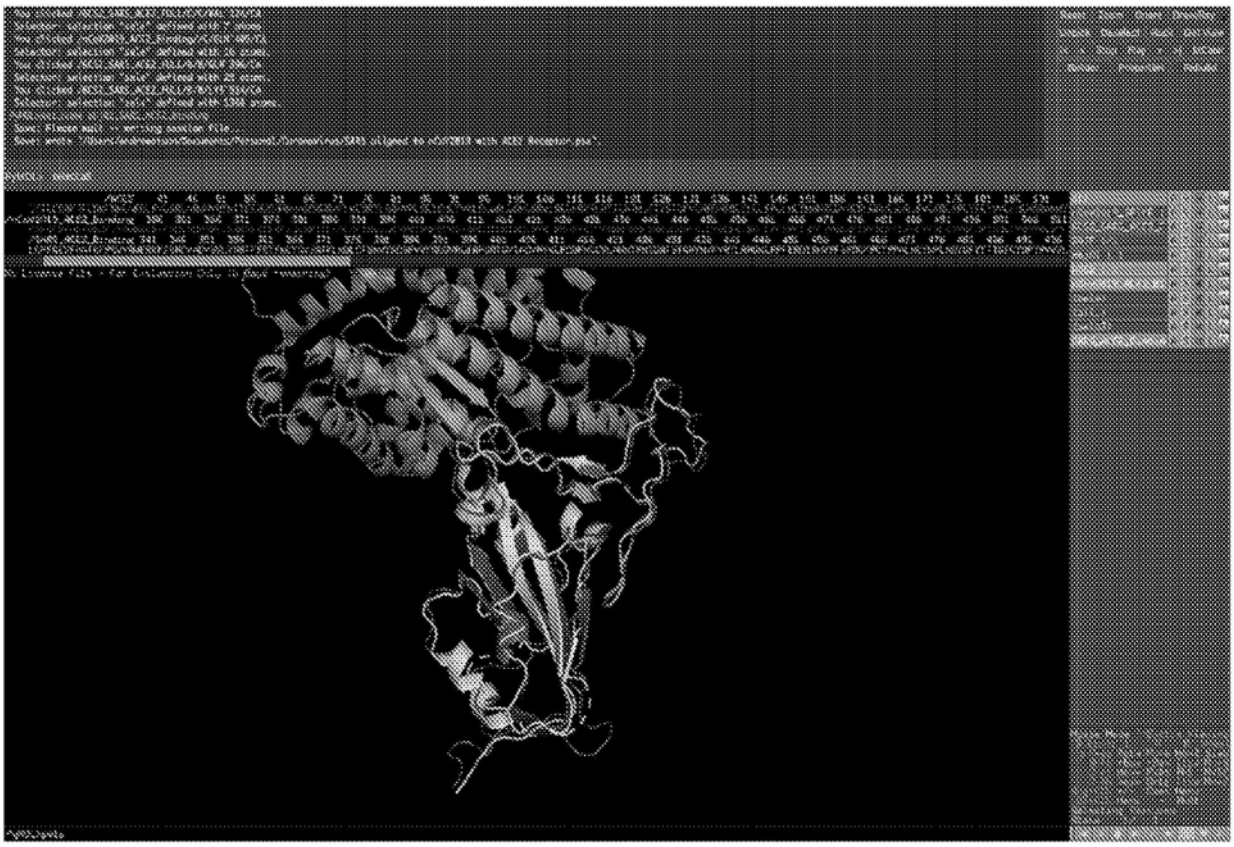


图9A

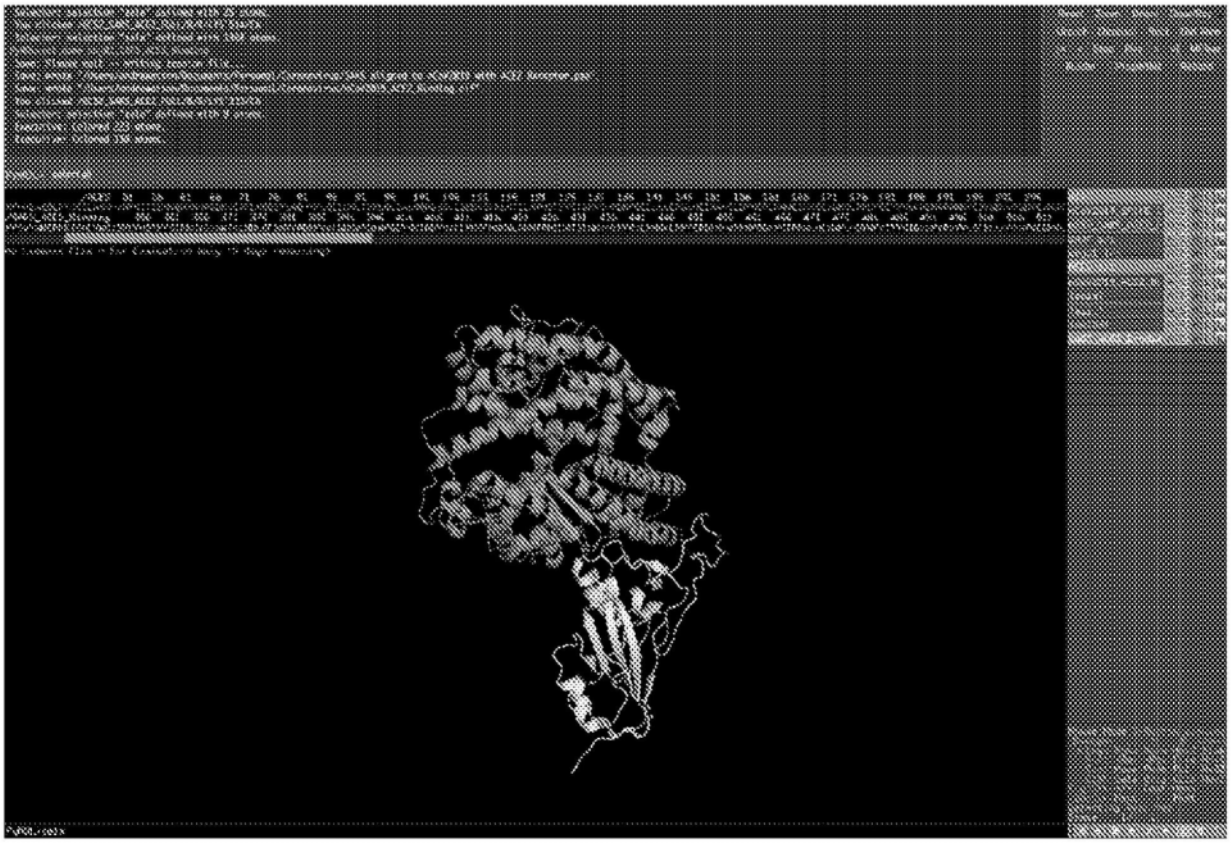


图9B

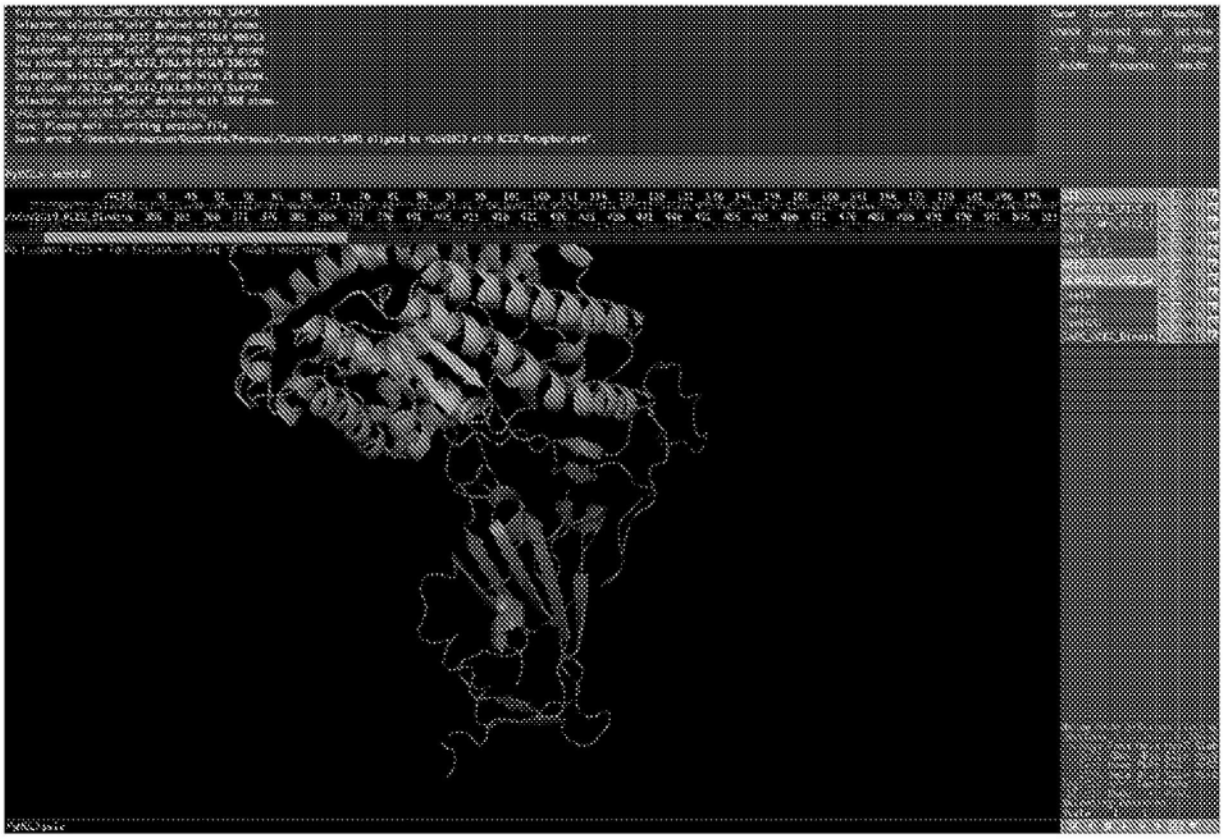


图9C

103	C:SER 438	0.27	0.00		0.00
104	C:ASN 439	51.82	4.52		-0.05
105	C:ASN 440	120.39	0.00		0.00
106	C:LEU 441	116.15	0.00		0.00
107	C:ASP 442	4.47	0.00		0.00
108	C:SER 443	7.57	0.00		0.00
109	C:LYS 444	134.20	0.00		0.00
110	C:VAL 445	104.95	12.88		0.07
111	C:GLY 446	74.09	2.01		0.03
112	C:GLY 447	18.65	0.00		0.00
113	C:ASN 448	25.97	0.00		0.00
114	C:TYR 449	144.06	50.09		-0.10
115	C:ASN 450	121.67	0.00		0.00
116	C:TYR 451	7.04	0.00		0.00
117	C:LEU 452	52.59	0.00		0.00
118	C:TYR 453	24.51	21.72		-0.22
119	C:ARG 454	6.56	0.00		0.00
120	C:LEU 455	56.10	29.48		0.47
121	C:PHE 456	62.19	46.43		0.74

138	C:TYR 473	41.14	3.71		0.06
139	C:GLN 474	63.86	26.79		-0.31
140	C:ALA 475	81.87	31.56		0.33
141	C:GLY 476	59.93	12.84		0.20
142	C:SER 477	71.37	0.00		0.00
143	C:THR 478	89.05	0.00		0.00
144	C:PRO 479	111.89	0.00		0.00
145	C:CYS 480	27.81	0.00		0.00
146	C:ASN 481	116.94	0.00		0.00
147	C:GLY 482	61.22	0.00		0.00
148	C:VAL 483	117.18	0.00		0.00
149	C:GLU 484	154.65	8.34		-0.10
150	C:GLY 485	26.83	0.00		0.00
151	C:PHE 486	197.05	88.29		1.41
152	C:ASN 487	110.05	94.19		-0.66
153	C:CYS 488	20.79	0.00		0.00
154	C:TYR 489	124.82	111.79		0.80
155	C:PHE 490	124.78	8.42		0.06
156	C:PRO 491	3.85	0.00		0.00
157	C:LEU 492	42.63	0.00		0.00
158	C:GLN 493	84.98	69.90		-0.57
159	C:SER 494	26.29	4.65		-0.05
160	C:TYR 495	16.26	0.74		-0.01
161	C:GLY 496	36.64	36.19		0.15
162	C:PHE 497	4.14	0.00		0.00
163	C:GLN 498	69.21	59.55		-0.64
164	C:PRO 499	59.66	19.01		-0.22
165	C:THR 500	107.29	86.86		0.39
166	C:ASN 501	91.21	90.92		-0.58
167	C:GLY 502	48.57	39.96		0.36
168	C:VAL 503	112.45	36.14		0.57
169	C:GLY 504	25.00	0.00		0.00
170	C:TYR 505	128.65	71.01		0.36
171	C:GLN 506	24.48	19.48		-0.22

图10A

437	B:PHE 460		30.59	7.35	0.12
438	B:SER 461		10.03	0.00	0.00
439	B:PRO 462		76.94	63.14	0.57
440	B:ASP 463		80.73	38.19	0.38
441	B:GLY 464		68.21	0.00	0.00
442	B:LYS 465		131.47	0.00	0.00
443	B:PRO 466		137.10	0.00	0.00
444	B:CYS 467		19.47	0.00	0.00
445	B:THR 468		107.80	0.00	0.00
446	B:PRO 469		37.31	0.00	0.00
447	B:PRO 470		139.42	13.43	0.00
448	B:ALA 471		58.97	5.31	-0.01
449	B:LEU 472		164.11	62.91	1.01
450	B:ASN 473		76.00	60.35	-0.65
451	B:CYS 474		14.37	0.37	-0.00
452	B:TYR 475		111.73	106.54	0.88
453	B:TRP 476		131.58	8.60	0.05
454	B:PRO 477		1.34	0.00	0.00
455	B:LEU 478		21.83	0.00	0.00
456	B:ASN 479		56.74	47.94	-0.44
457	B:ASP 480		64.39	0.25	-0.00
458	B:TYR 481		28.45	1.84	-0.02
459	B:GLY 482		27.92	24.56	-0.00
460	B:PHE 483		5.53	0.00	0.00
461	B:TYR 484		75.42	62.15	0.64
462	B:THR 485		59.52	6.13	-0.07
463	B:THR 486		110.16	84.38	0.28
464	B:THR 487		39.80	39.22	0.45
465	B:GLY 488		38.02	34.00	0.33
466	B:ILE 489		138.74	55.15	0.86
467	B:GLY 490		14.06	0.00	0.00
468	B:TYR 491		107.76	60.64	0.34
469	B:GLN 492		6.70	5.68	-0.06

图10B

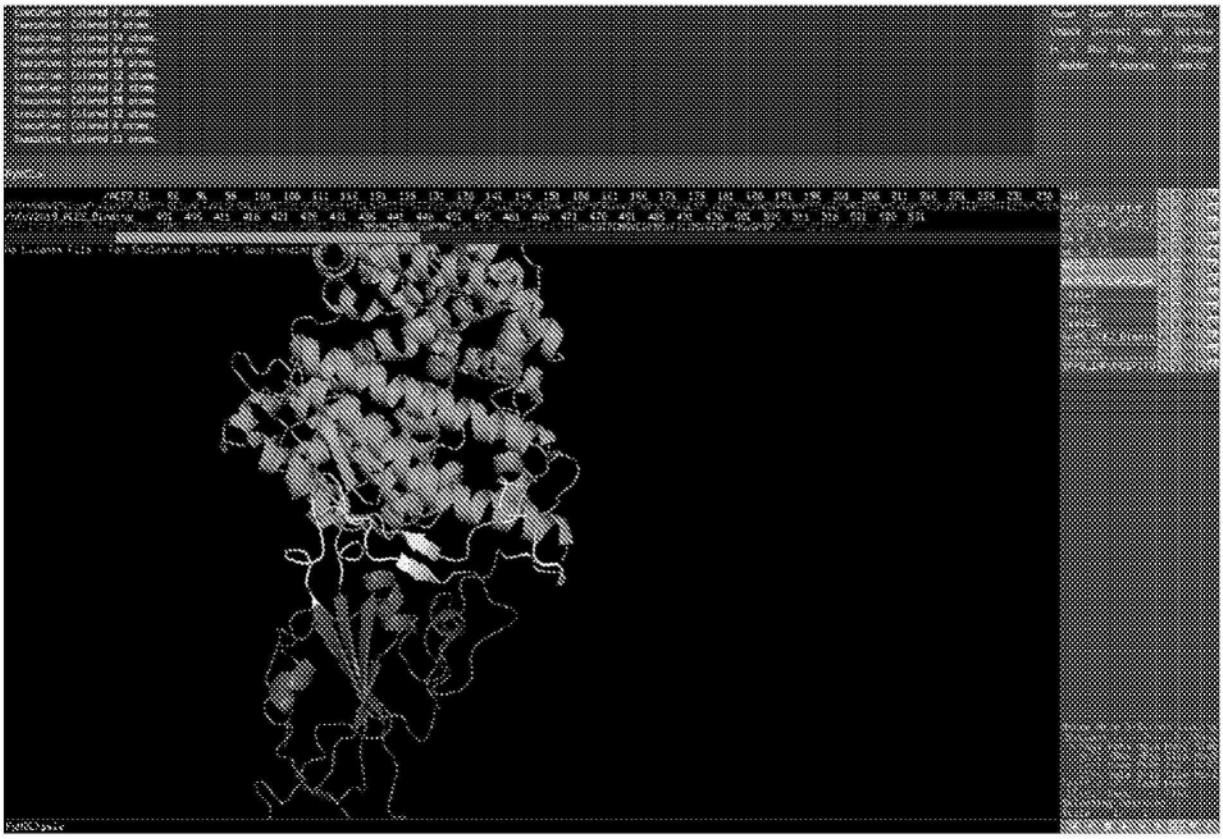


图11A

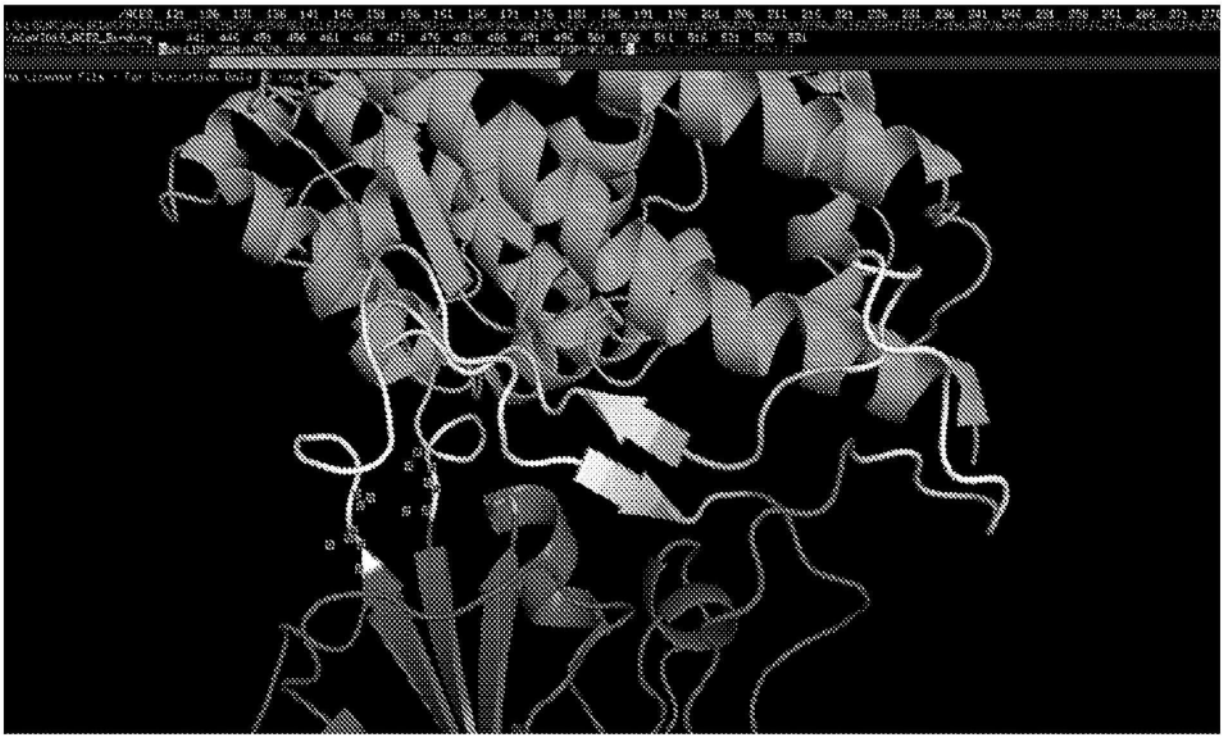


图11B

CoV-2的关键结合残基：



NSNNLDSKVGGNYNYLYRL (437-455)

YQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP (473-507)

图11C

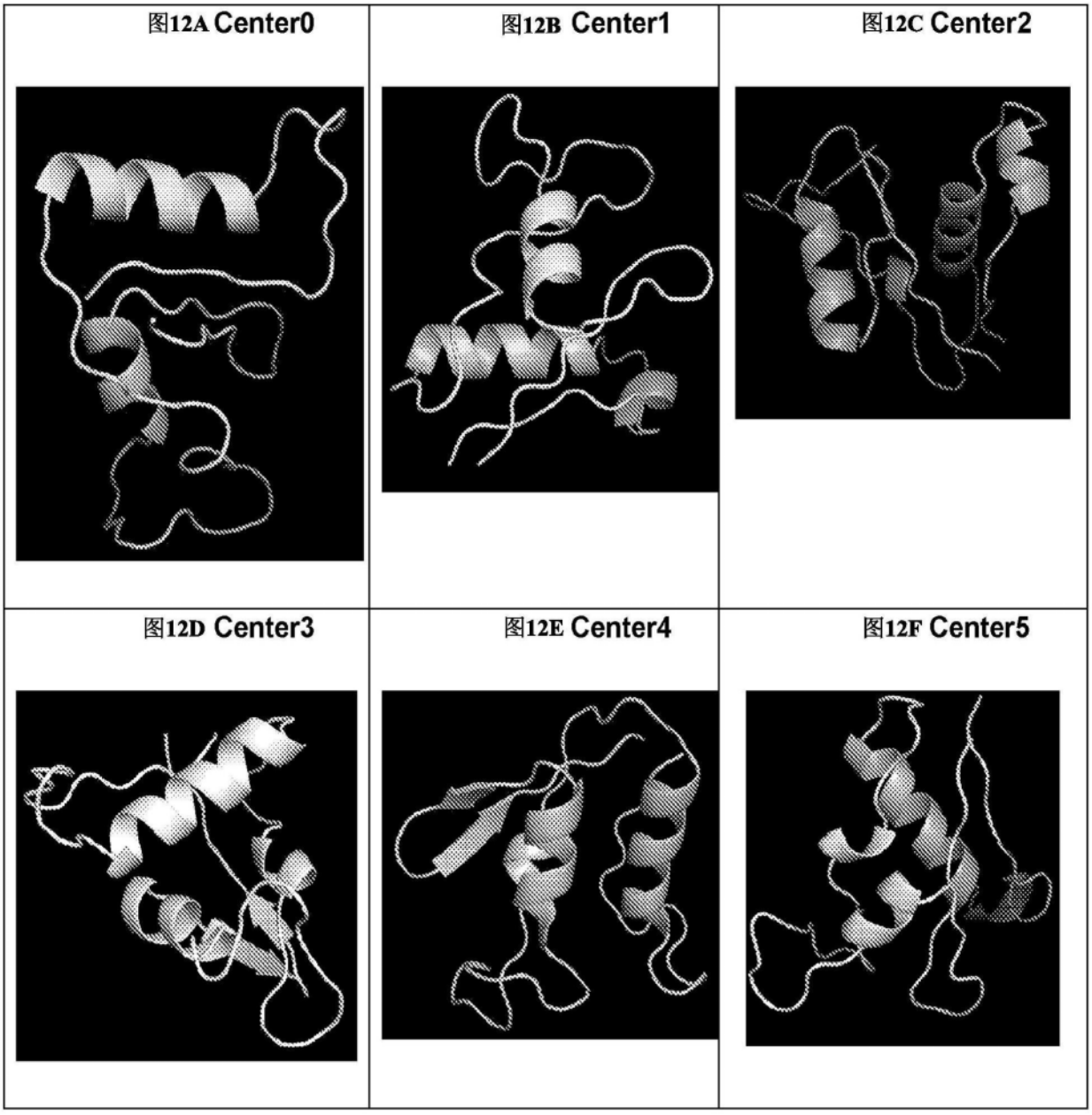


图12G Center6

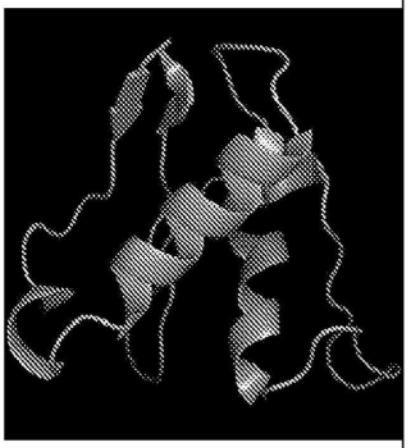


图12H Center7

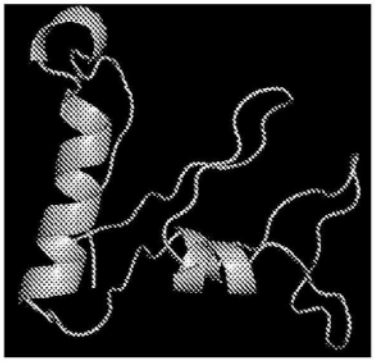


图12I Center8

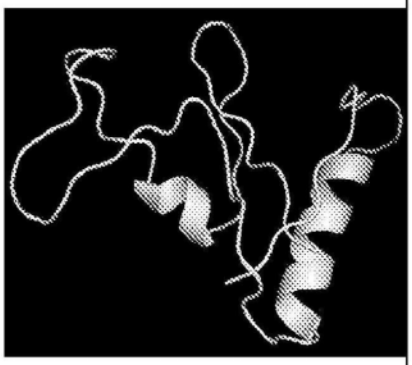


图12J Center9

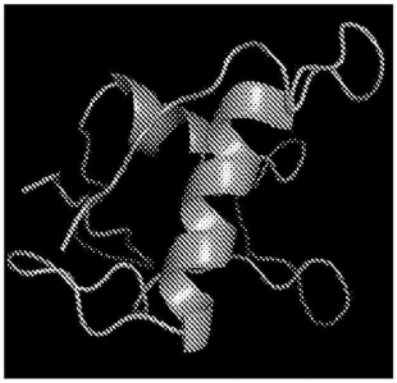


图13A

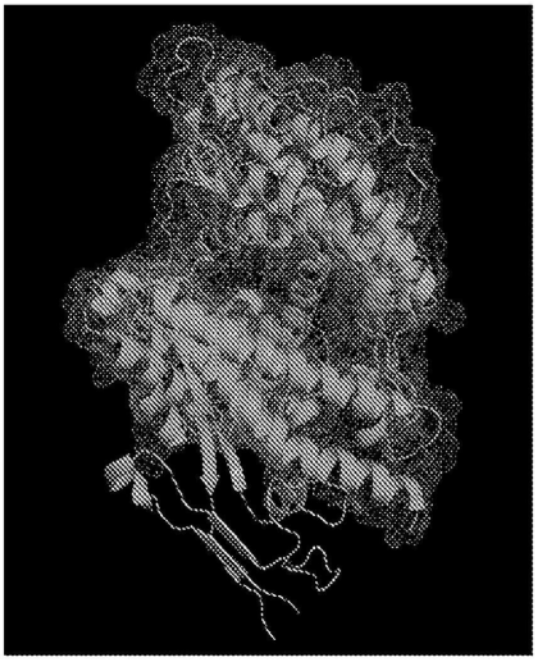


图13B



图13C

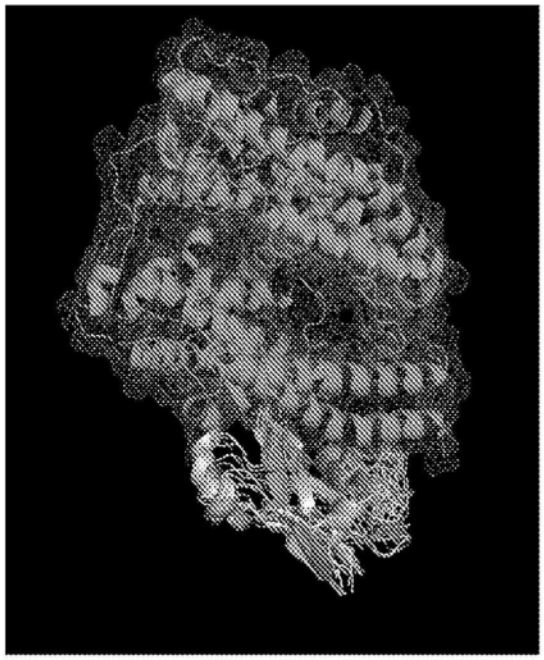
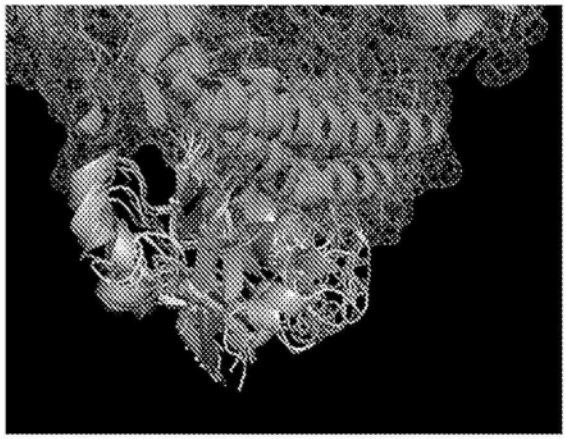


图13D



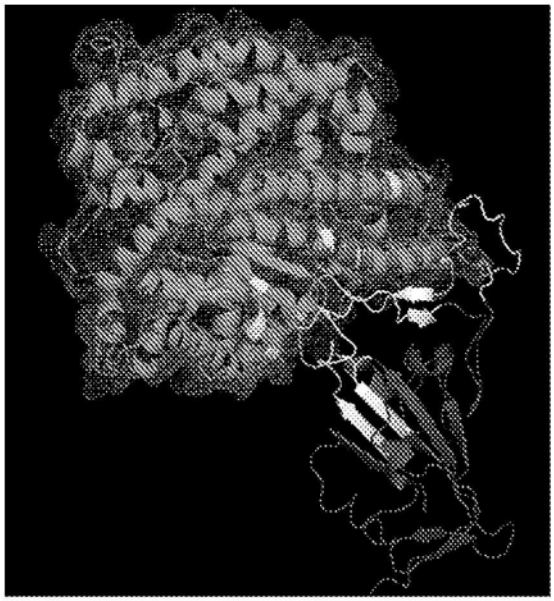


图14A

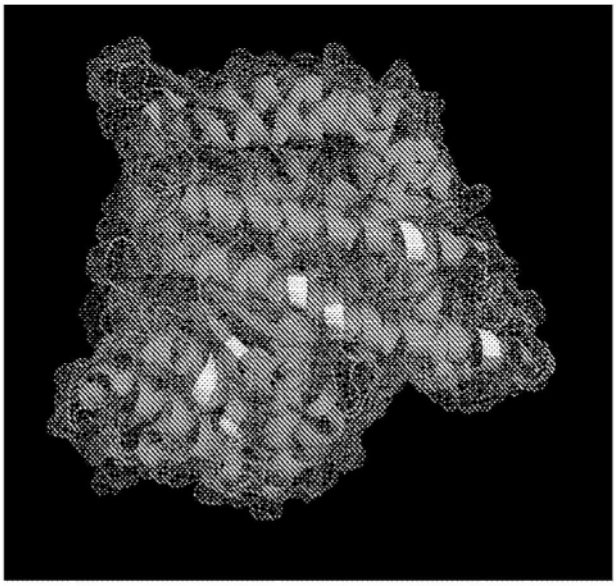


图14B

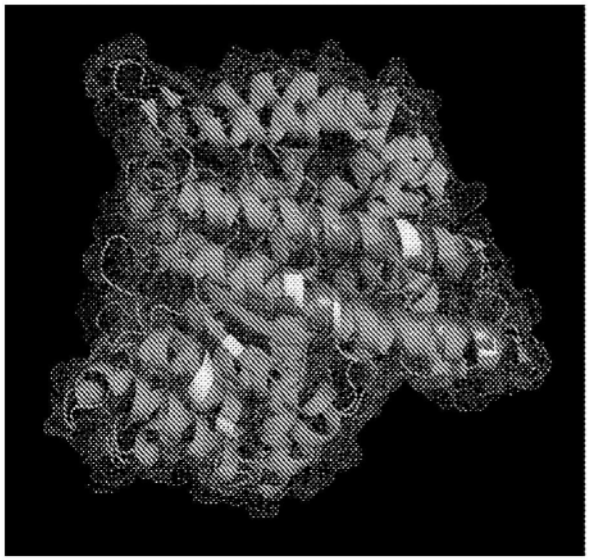


图14C

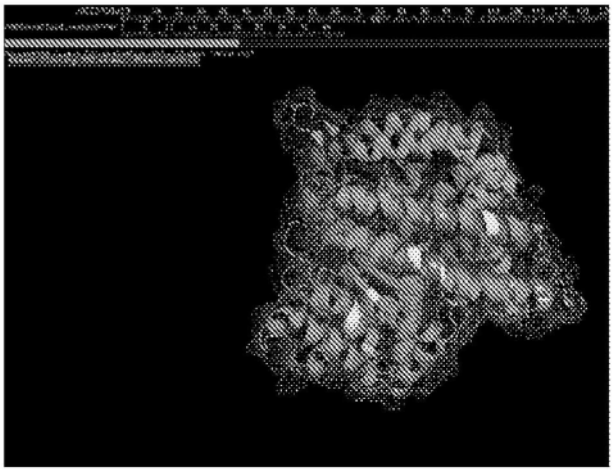


图14D

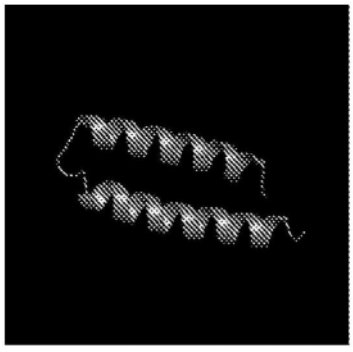


图15A



图15B

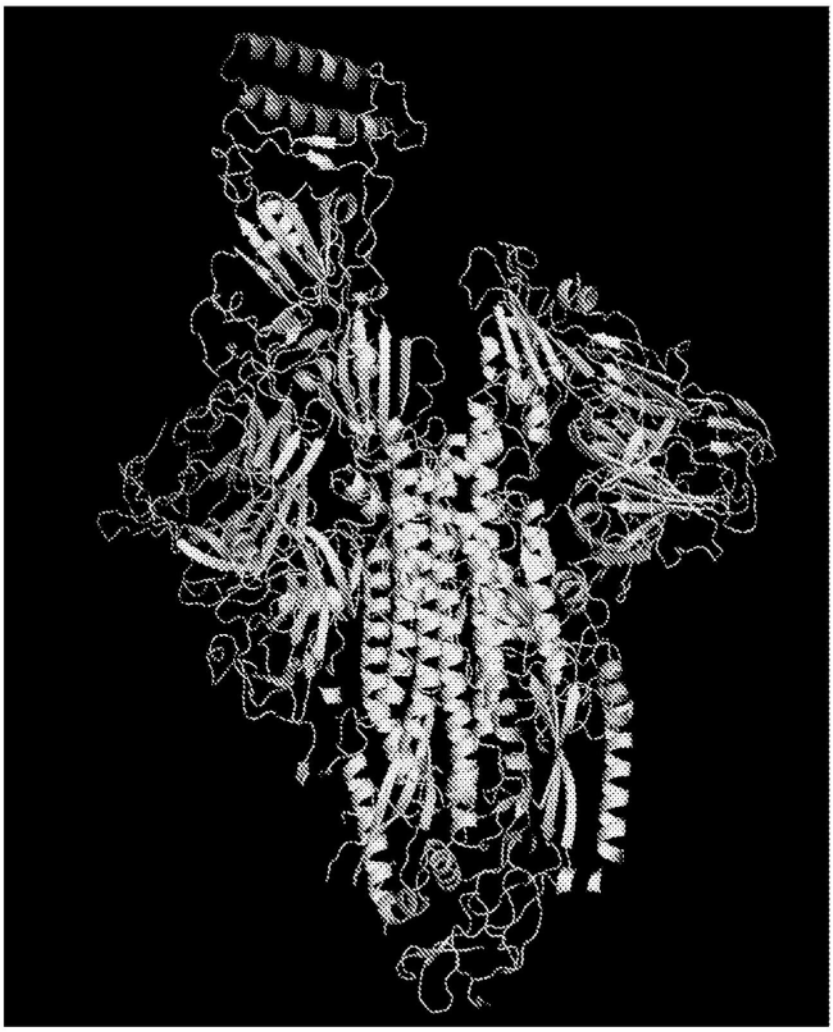


图15C

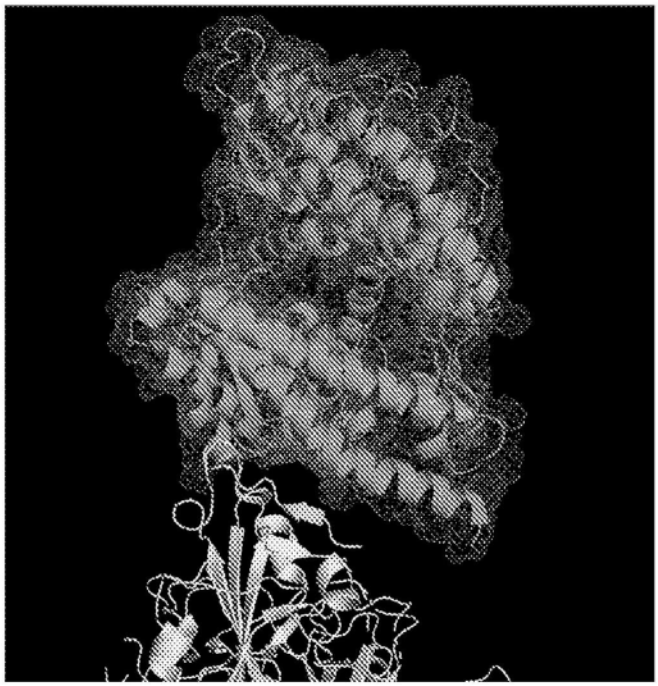


图16A

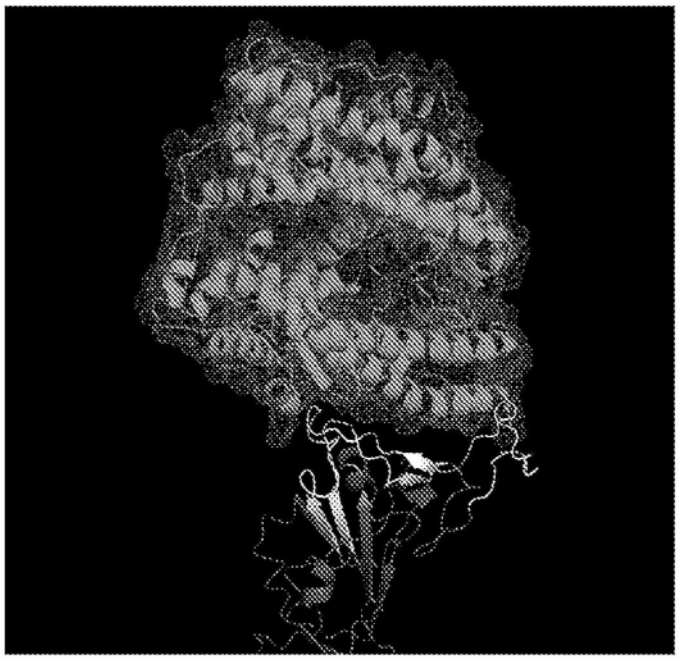


图16B

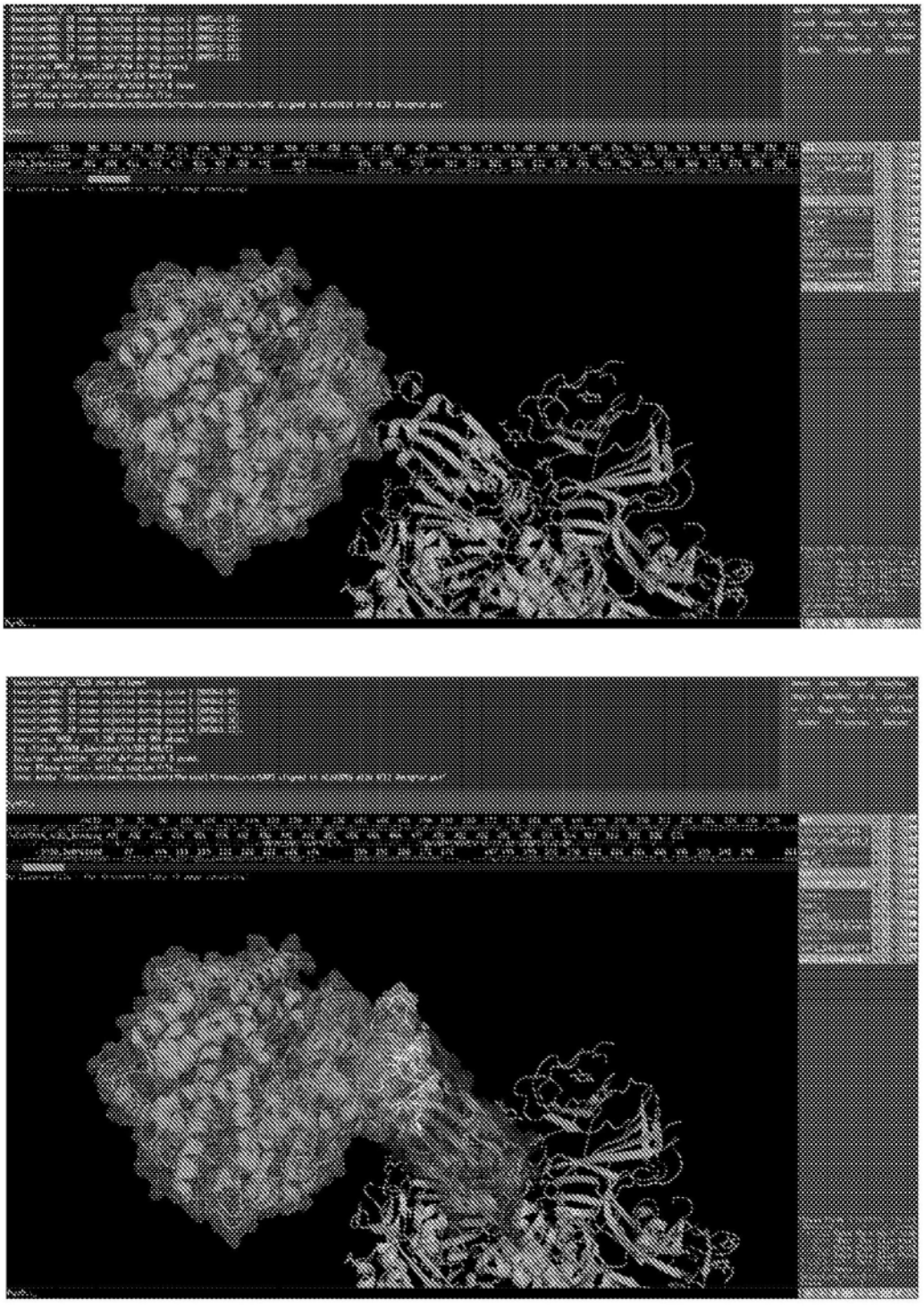


图17A

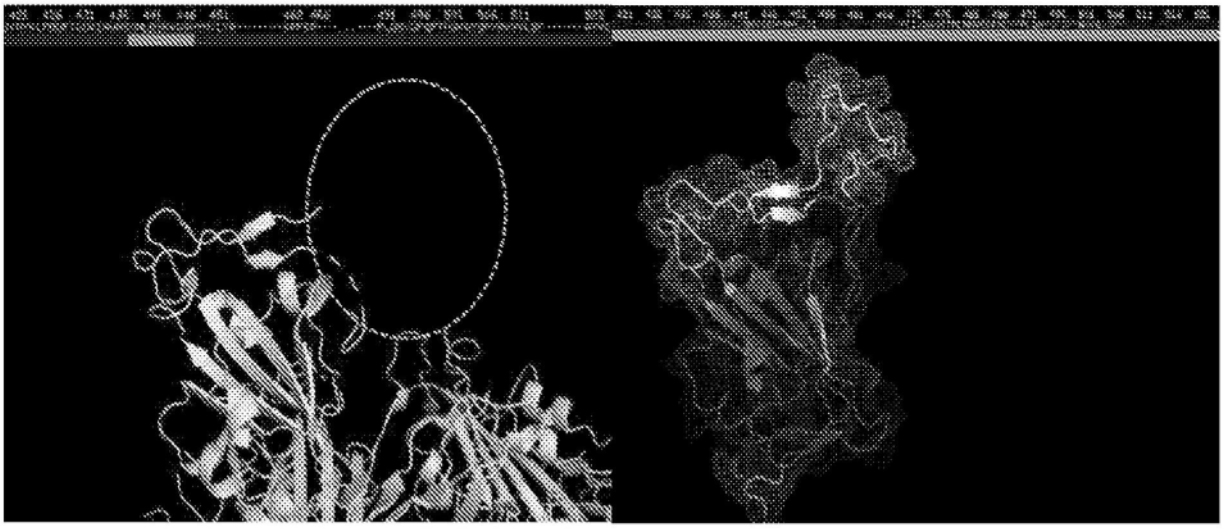


图17B

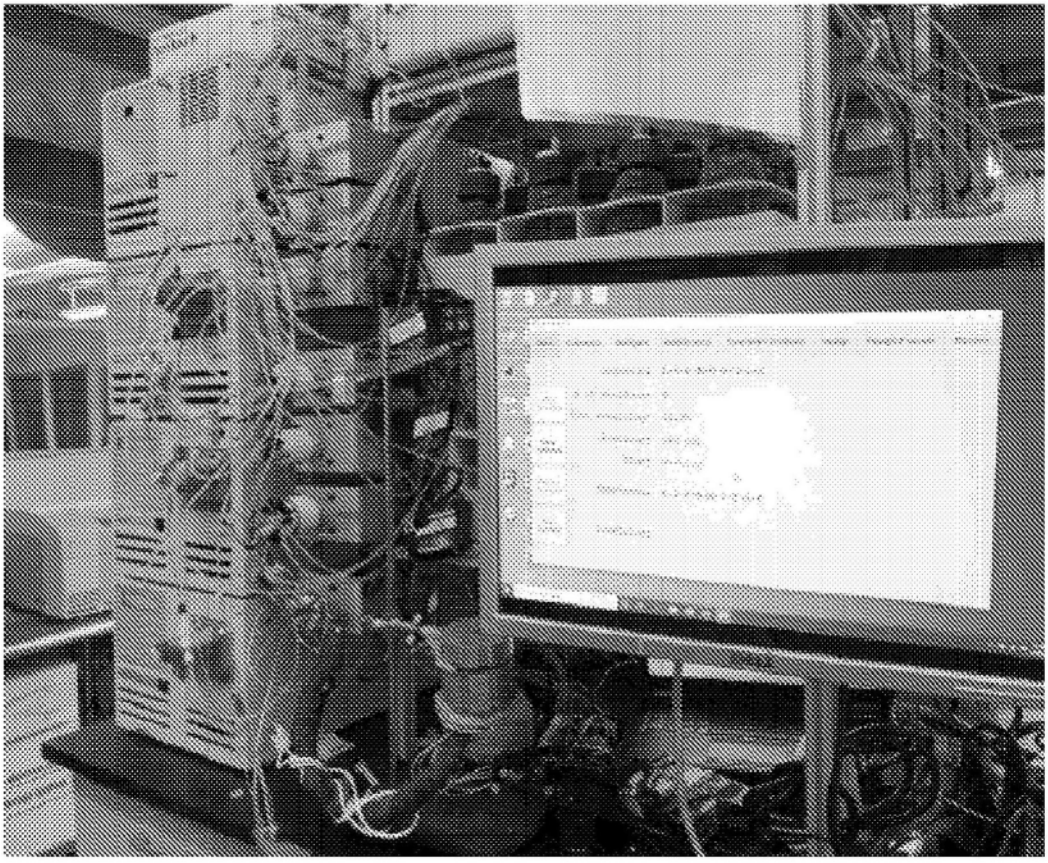


图18

Ligandal-05作为示例

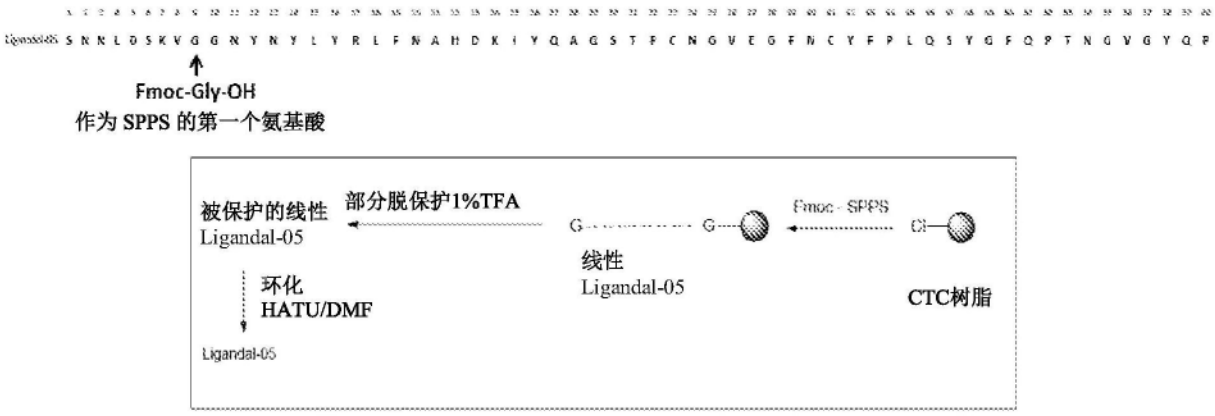


图19A

Ligandal-06作为示例

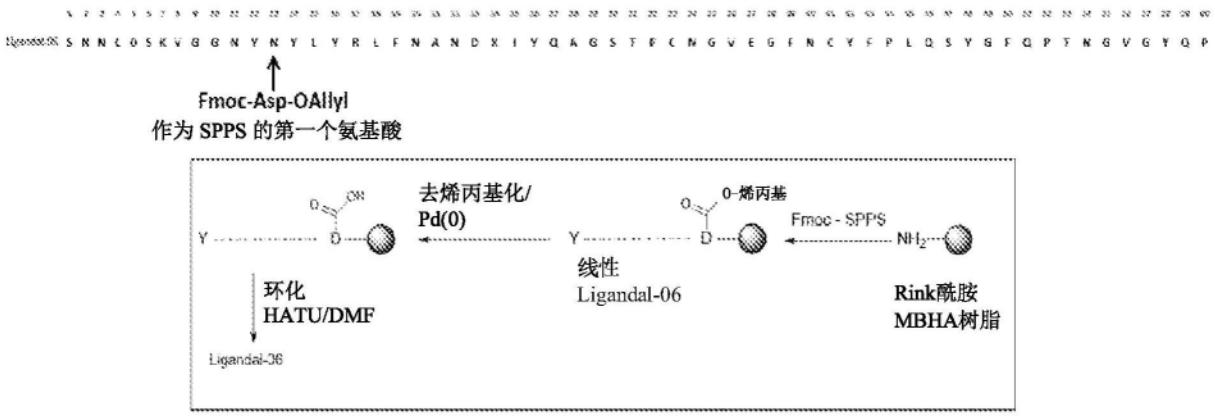


图19B

Ligandal-04作为示例

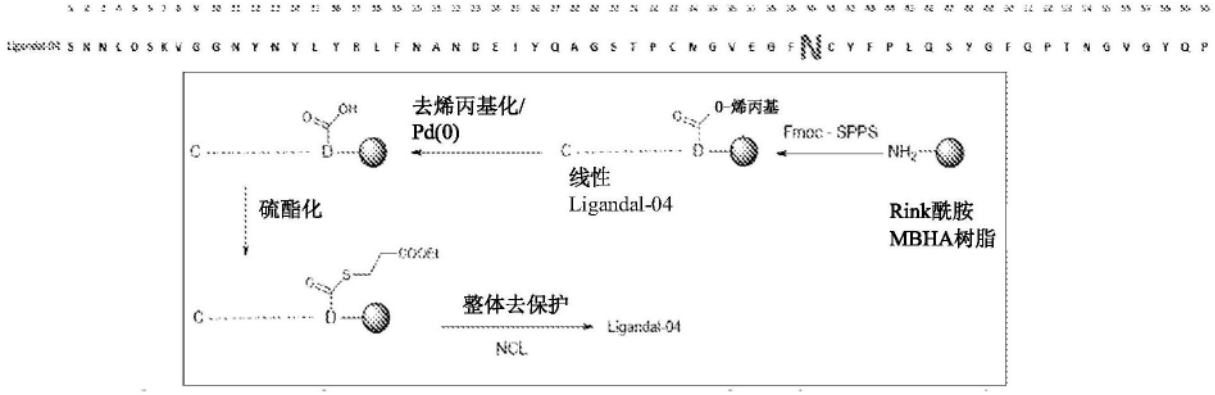


图19C

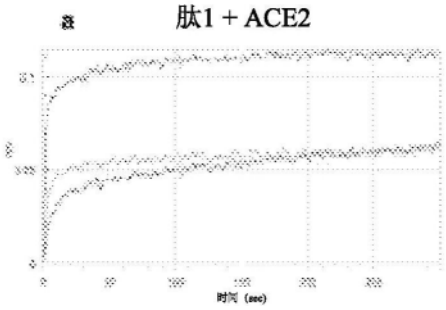


图20A

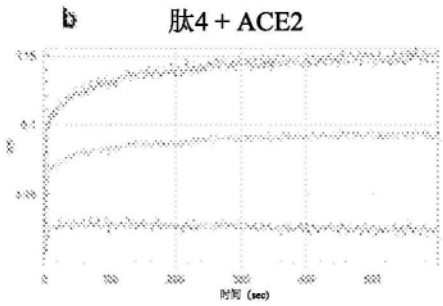


图20B

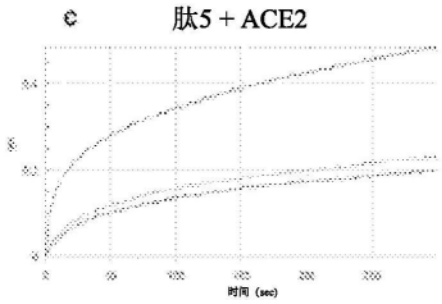


图20C

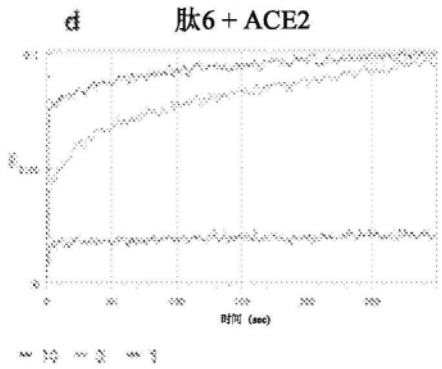


图20D

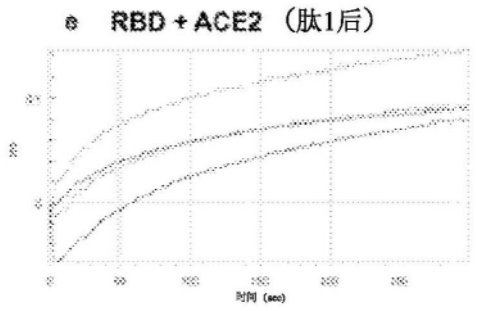


图20E

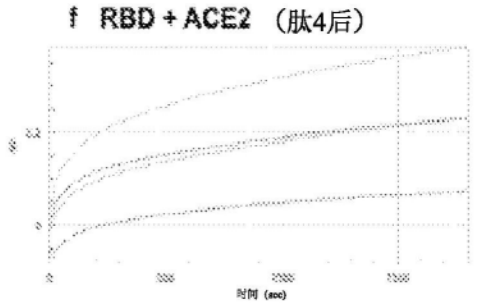


图20F

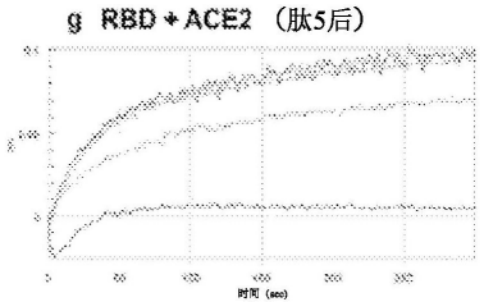


图20G

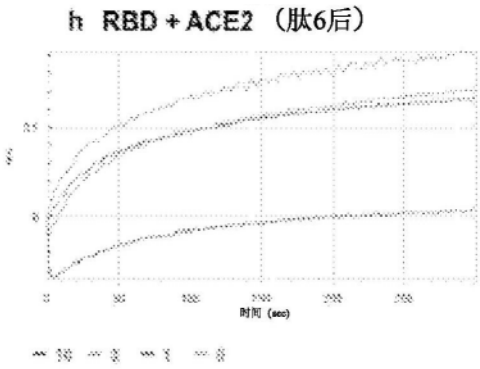


图20H

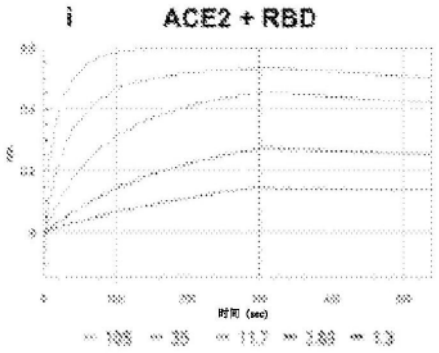


图20I

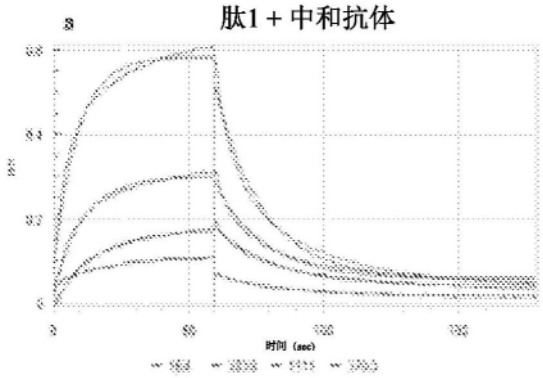


图21A

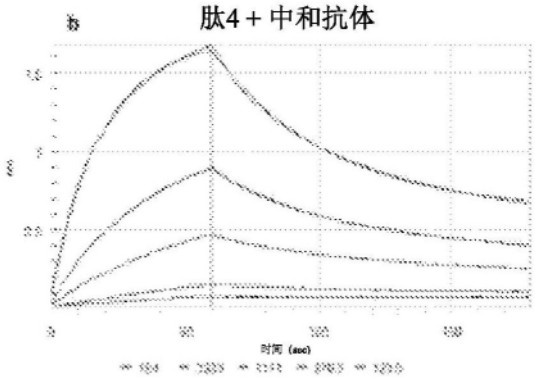


图21B

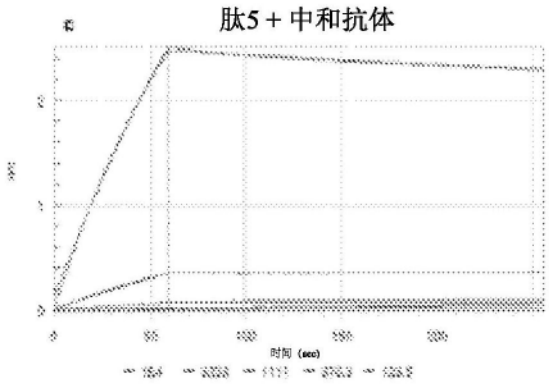


图21C

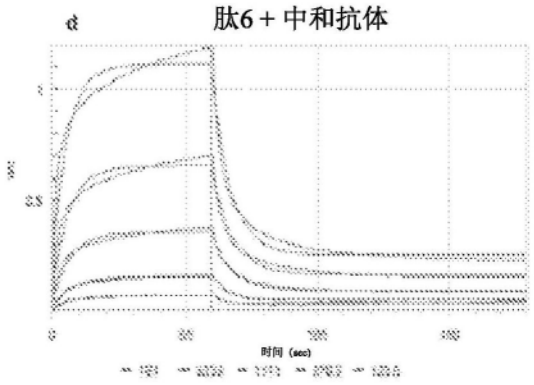


图21D

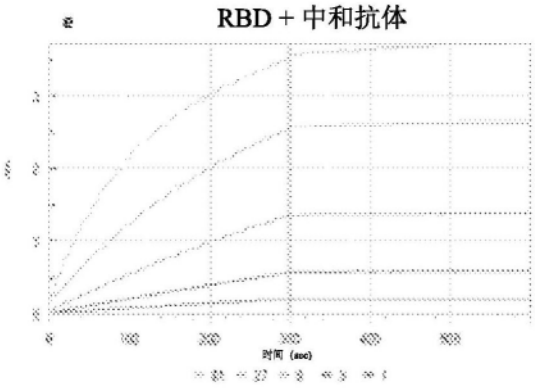


图21E

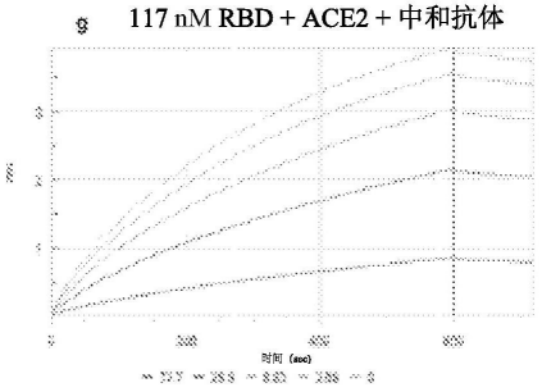

117 nM RBD + ACE2 + 中和抗体

图21F

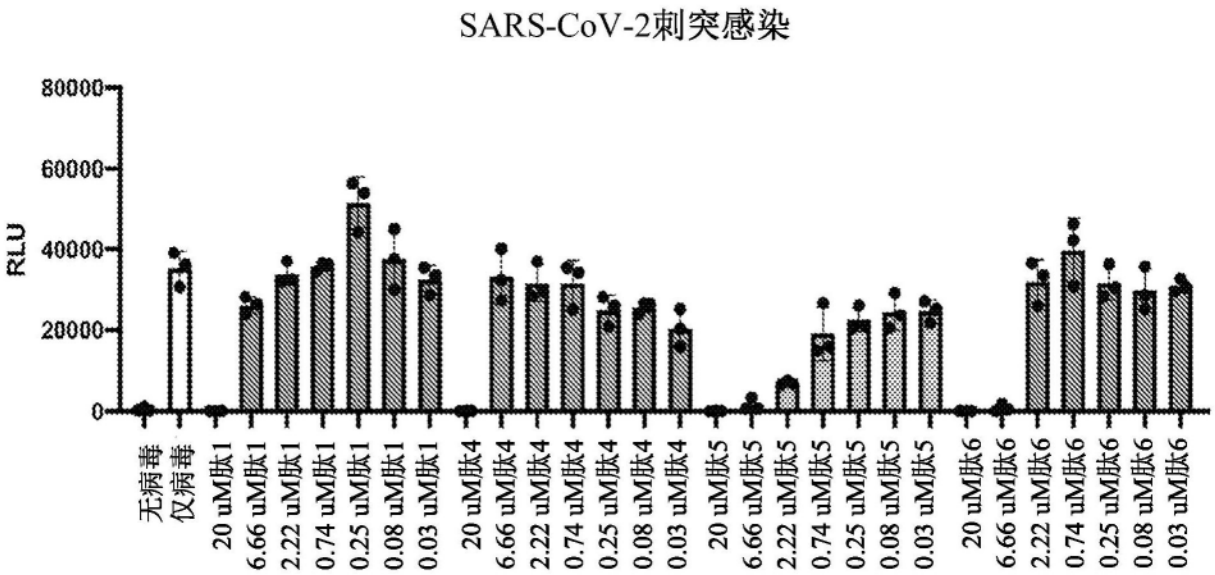

SARS-CoV-2刺突感染
RLU
80000
60000
40000
20000
0
无病毒
仅病毒
20 uM肽1
6.66 uM肽1
2.22 uM肽1
0.74 uM肽1
0.25 uM肽1
0.08 uM肽1
0.03 uM肽1
20 uM肽4
6.66 uM肽4
2.22 uM肽4
0.74 uM肽4
0.25 uM肽4
0.08 uM肽4
0.03 uM肽4
20 uM肽5
6.66 uM肽5
2.22 uM肽5
0.74 uM肽5
0.25 uM肽5
0.08 uM肽5
0.03 uM肽5
20 uM肽6
6.66 uM肽6
2.22 uM肽6
0.74 uM肽6
0.25 uM肽6
0.08 uM肽6
0.03 uM肽6

图22

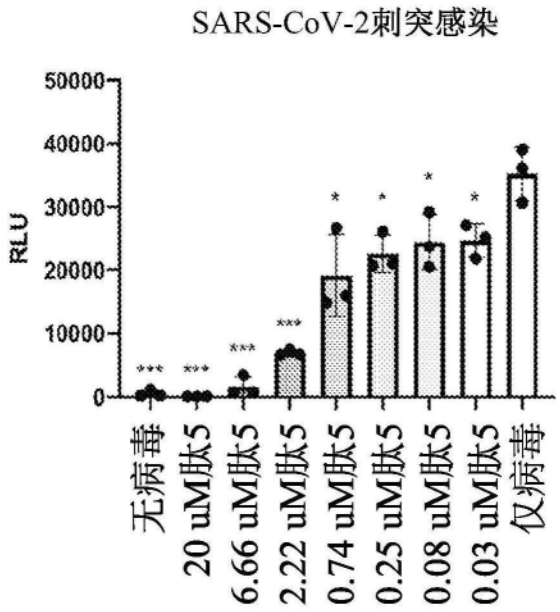

SARS-CoV-2刺突感染
RLU
50000
40000
30000
20000
10000
0
无病毒
20 uM肽5
6.66 uM肽5
2.22 uM肽5
0.74 uM肽5
0.25 uM肽5
0.08 uM肽5
0.03 uM肽5
仅病毒

图23A

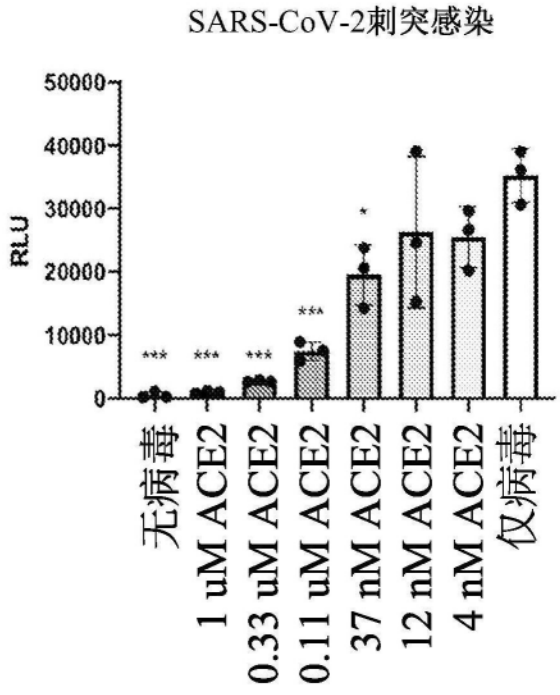

SARS-CoV-2刺突感染
RLU
50000
40000
30000
20000
10000
0
无病毒
1 uM ACE2
0.33 uM ACE2
0.11 uM ACE2
37 nM ACE2
12 nM ACE2
4 nM ACE2
仅病毒

图23B

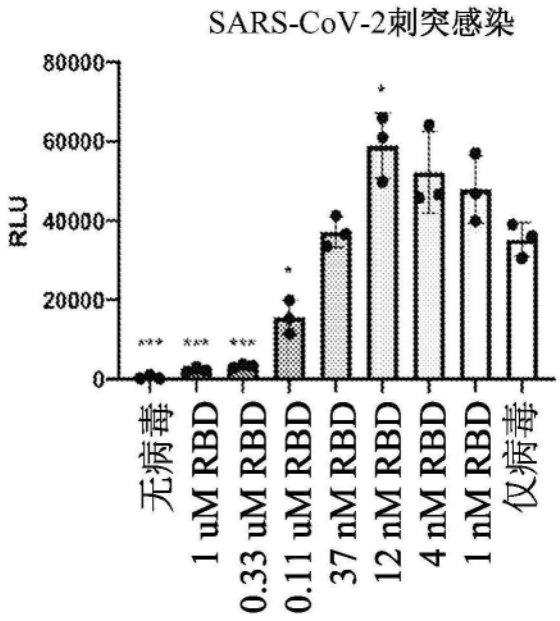

SARS-CoV-2刺突感染
RLU
80000
60000
40000
20000
0
无病毒
1 uM RBD
0.33 uM RBD
0.11 uM RBD
37 nM RBD
12 nM RBD
4 nM RBD
1 nM RBD
仅病毒

图23C

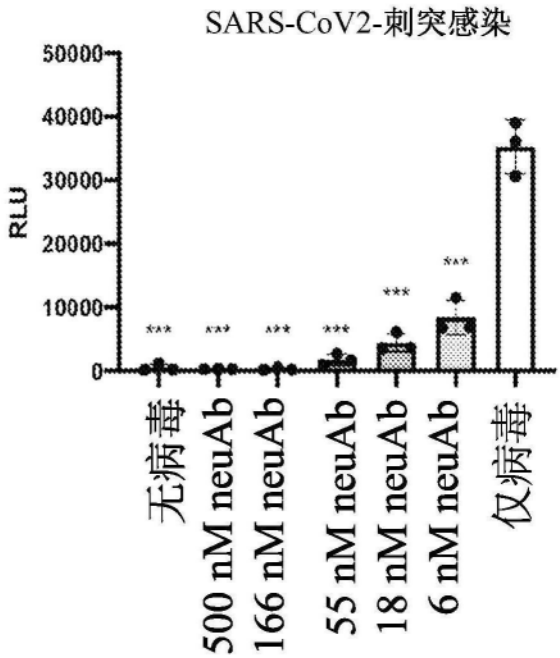

SARS-CoV2-刺突感染
RLU
50000
40000
30000
20000
10000
0
无病毒
500 nM neuAb
166 nM neuAb
55 nM neuAb
18 nM neuAb
6 nM neuAb
仅病毒

图23D

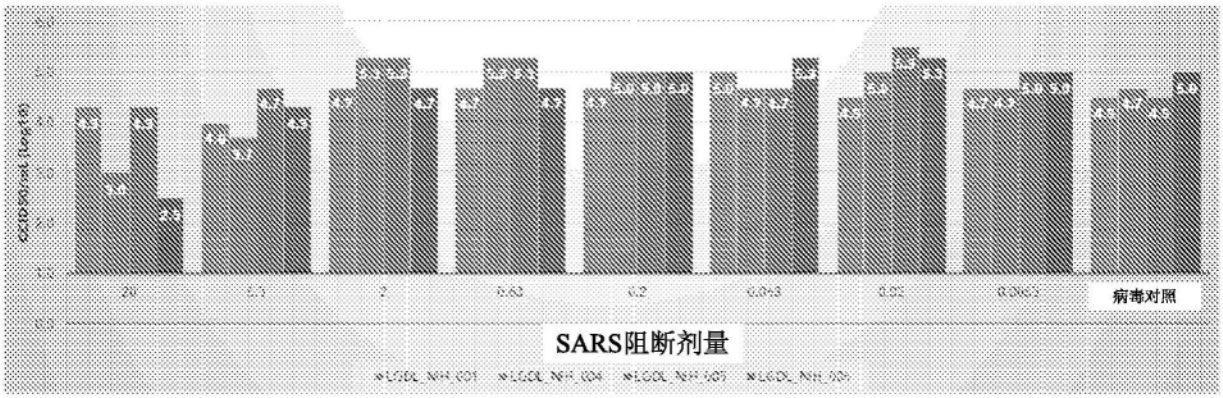


图24A

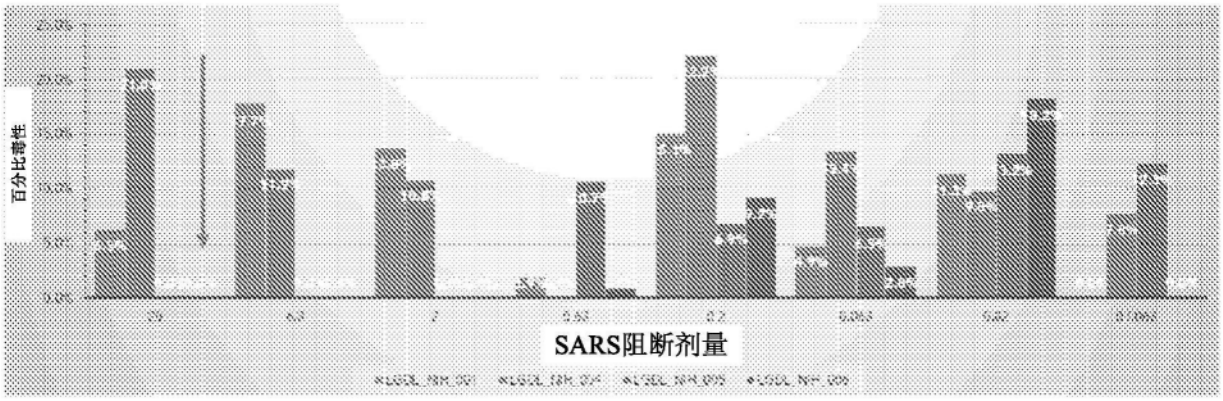


图24B

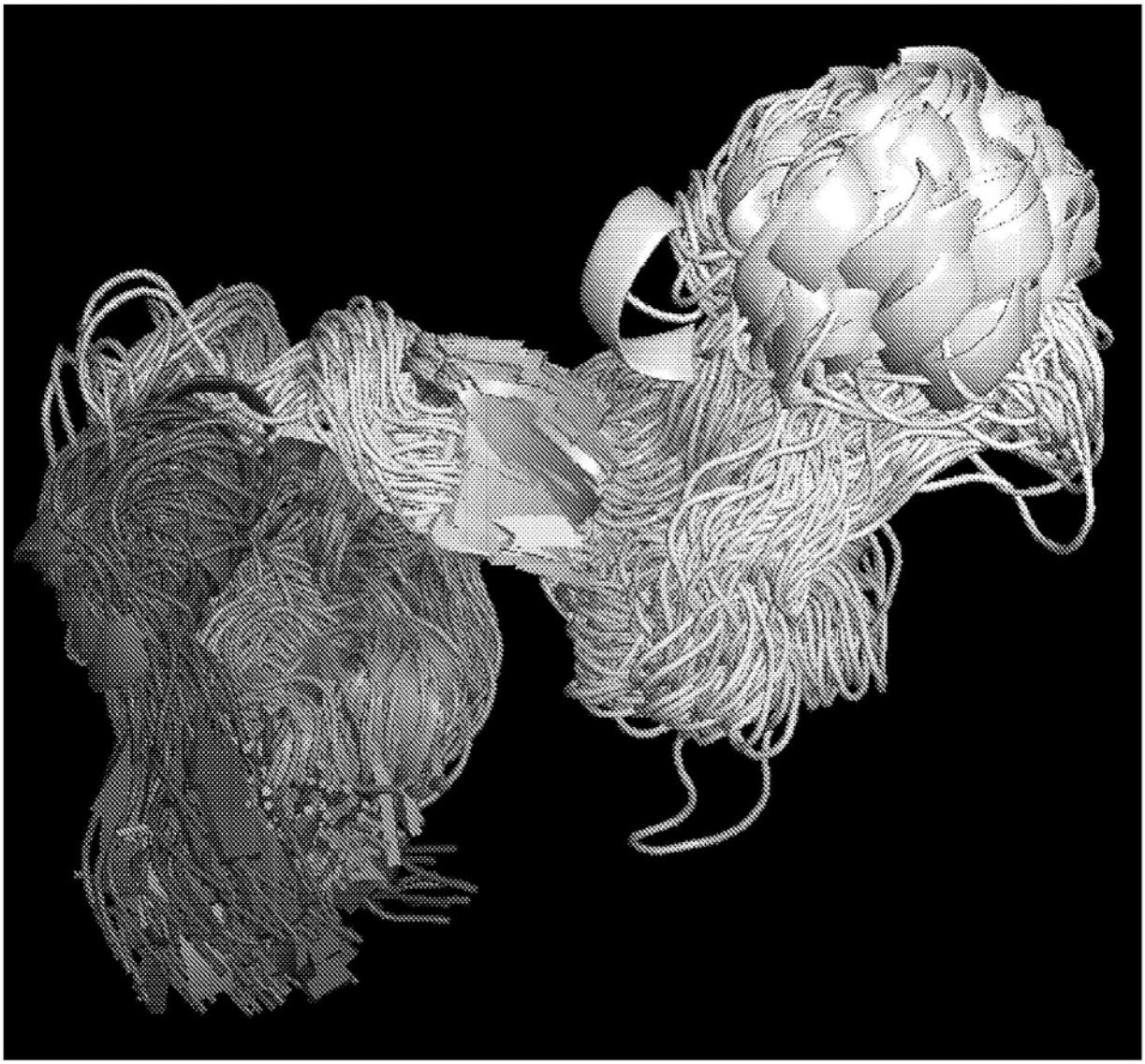


图25

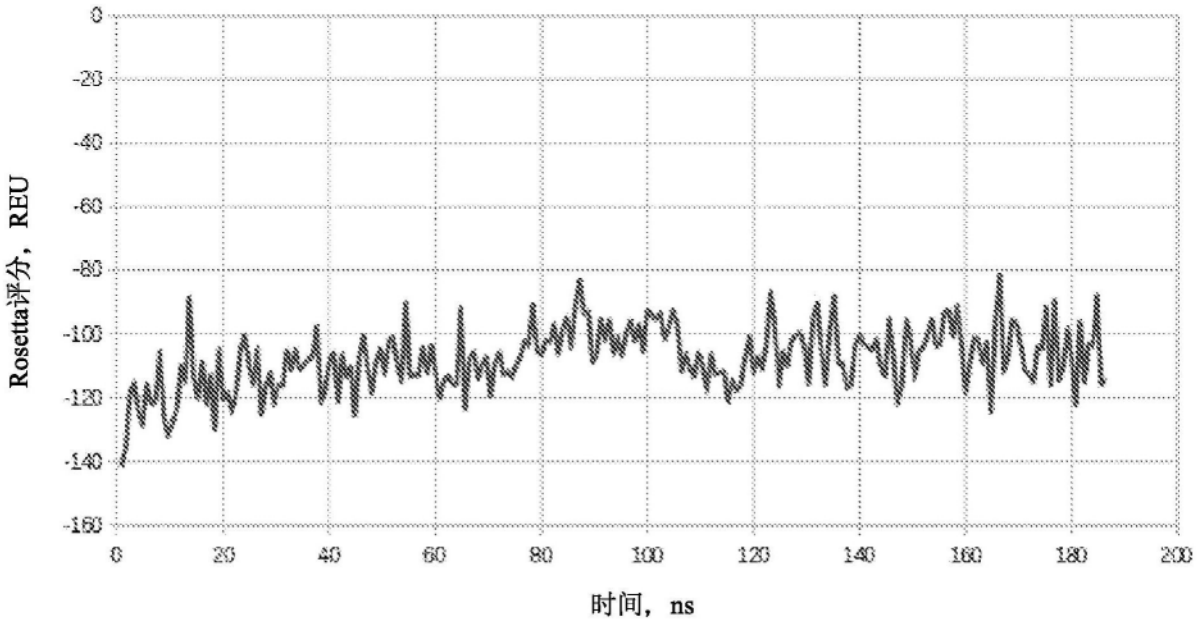


图26

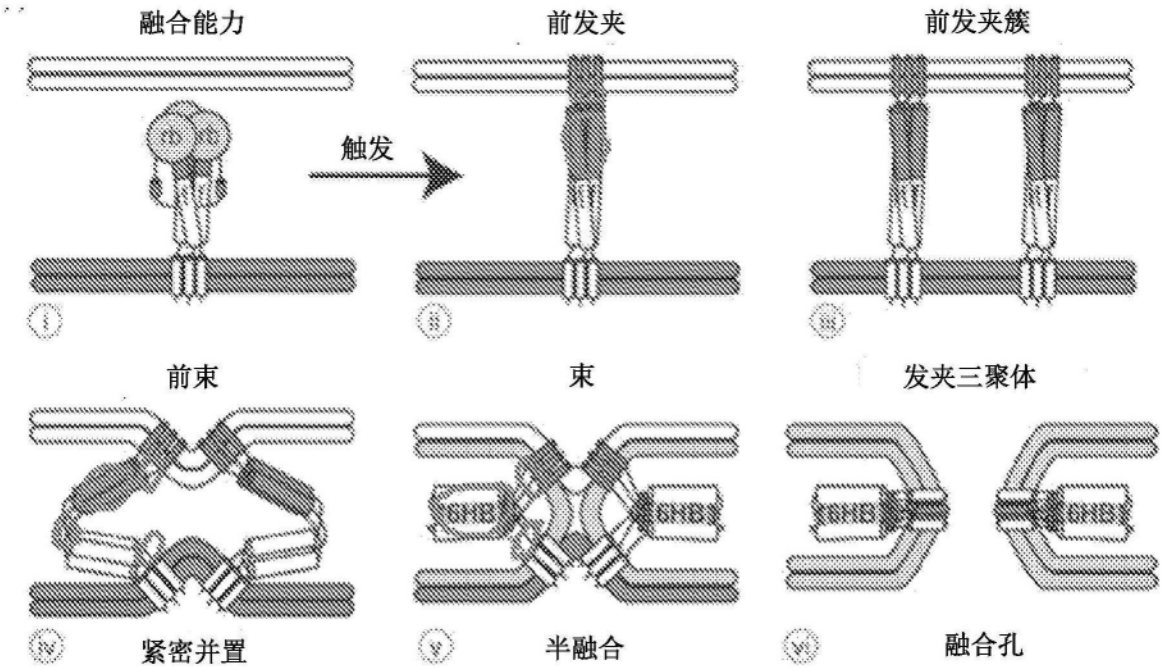


图27A



图27B

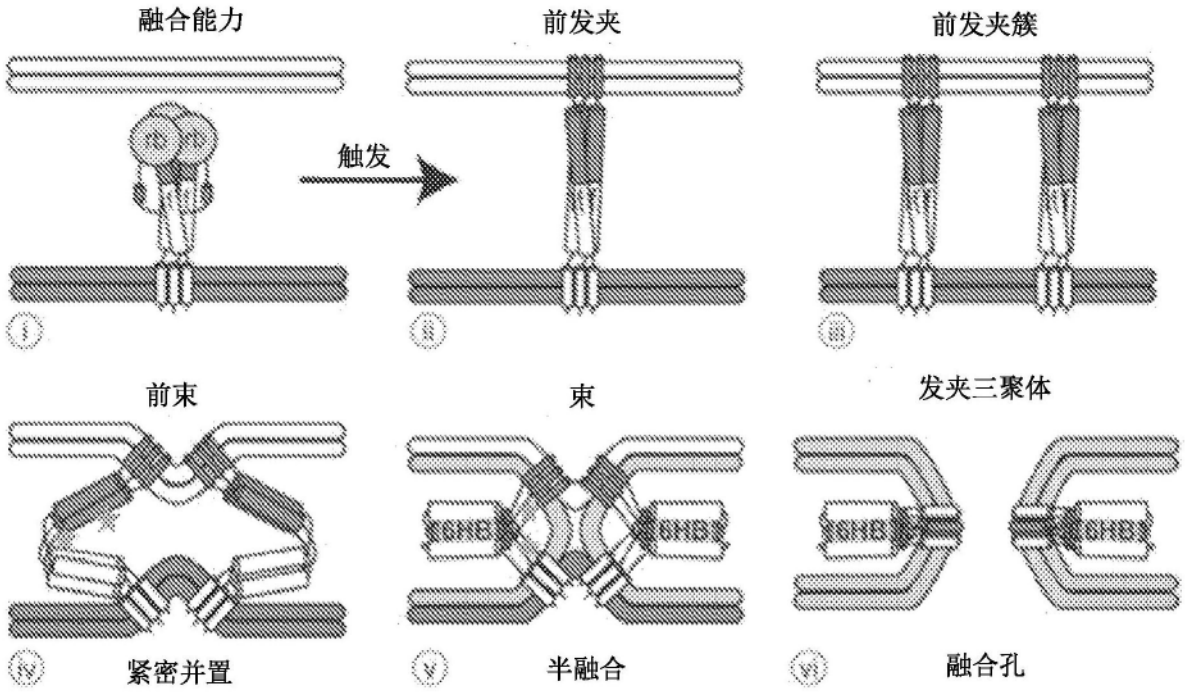


图28A

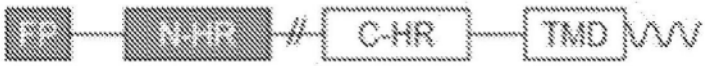


图28B



图28C

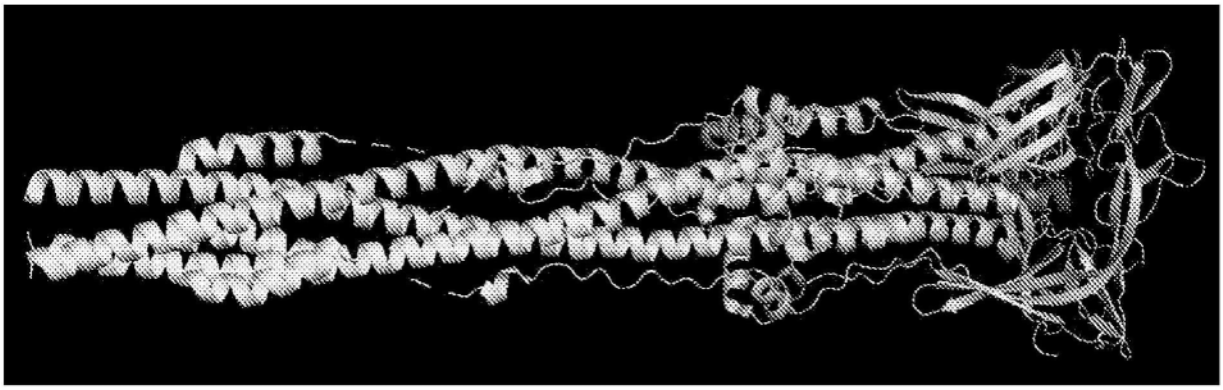


图29A

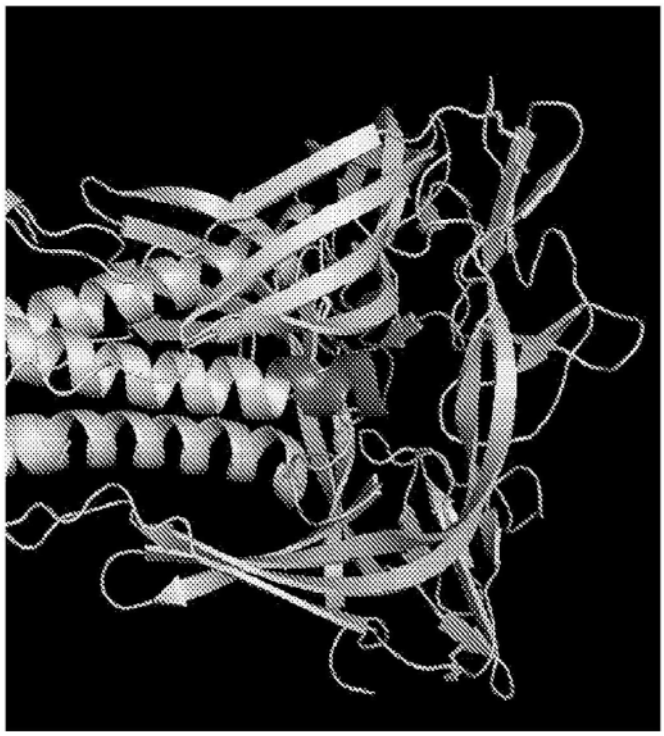


图29B



图30

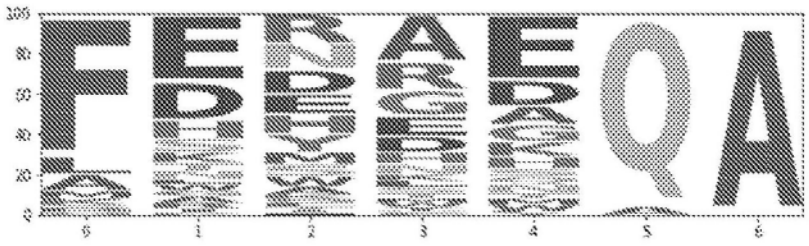


图31A

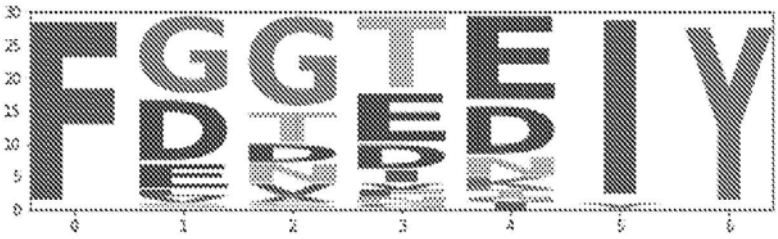


图31B

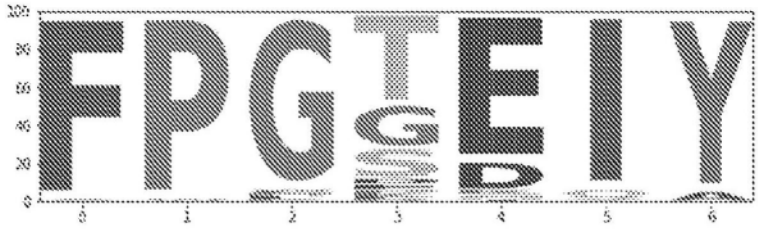


图31C

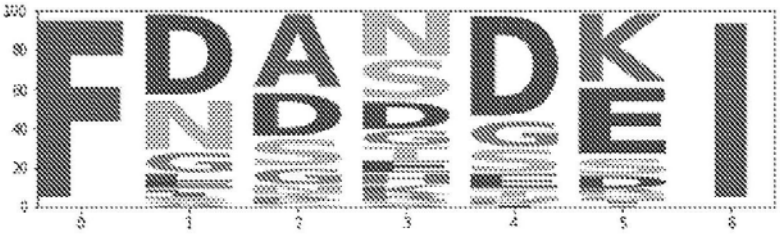


图31D

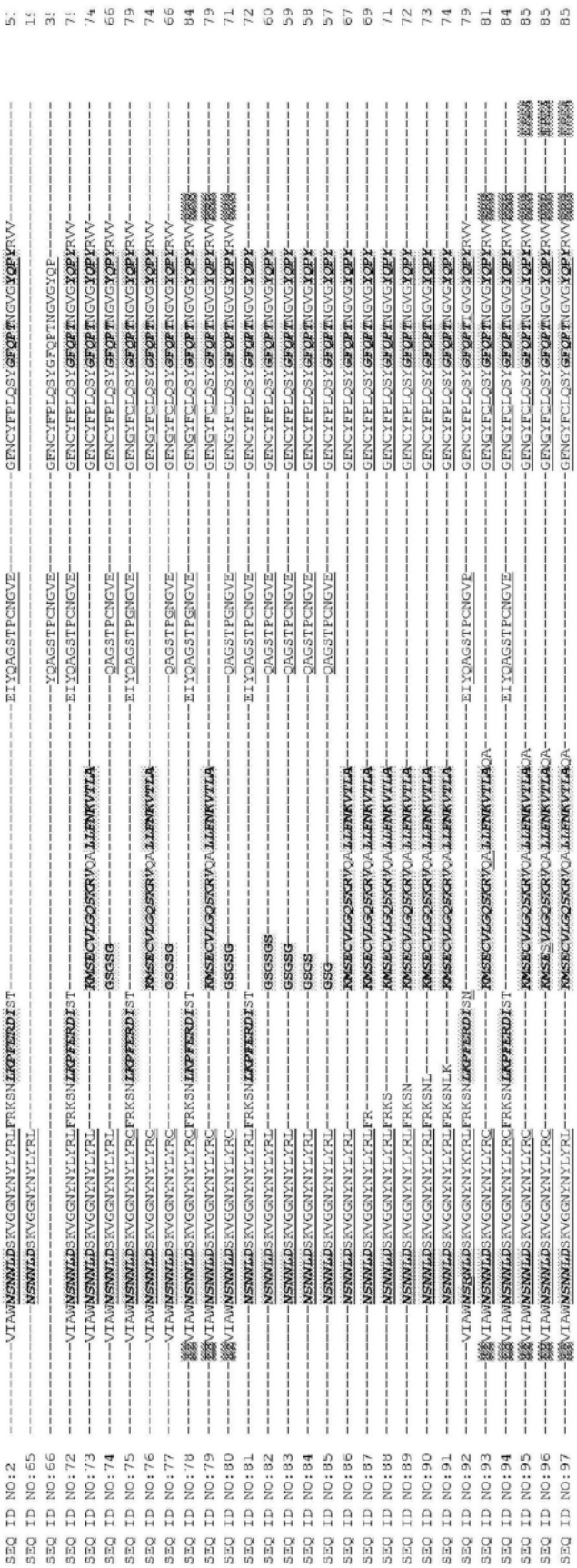


图32A

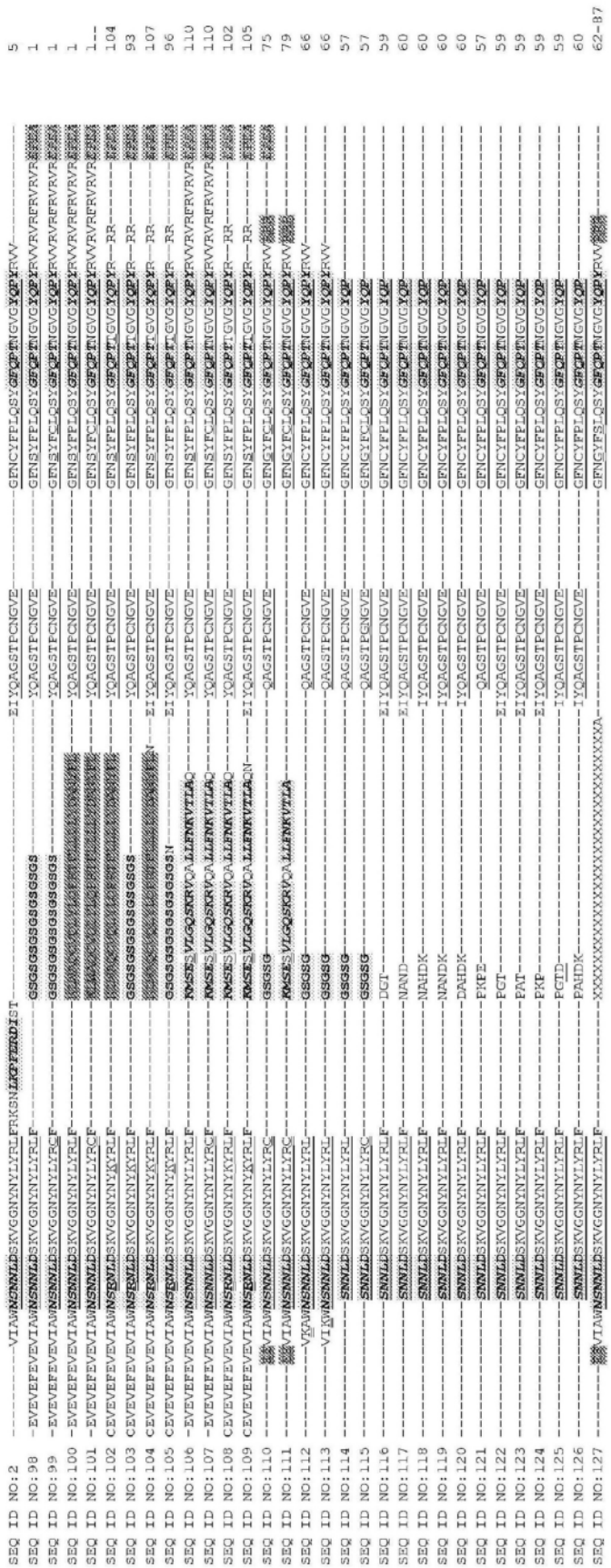


图32B


从结果中排除转录物
搜索
选择全部
无
选择全部结果
对您感兴趣的物种进行BLAST搜索：从下拉菜单中选择BLAST
定制
搜索输入
定制输入
交叉反应物种
人
设计ID
添加到购物车
显示产品细节+
更多

图33A


搜索输入
交叉反应物种
设计ID
定制输入
人
$95.00 USD
添加到购物车
隐藏产品细节-
1
114
228
交叉反应转录物 无
序列位置 10-33
序列细节 正义和反义序列
序列
链

图33B


DsiRNA 定制
搜索输入
交叉反应物种
设计ID
定制输入
人
CD.RI.218115.13.10
$95.00 USD
添加到购物车
隐藏产品细节-
1
114
228
交叉反应转录物 无
序列位置 45-70
序列细节 正义和反义序列
序列
链

图33C


DsiRNA 定制
搜索输入
交叉反应物种
设计ID
定制输入
人
CD.RI.218115.13.5
$95.00 USD
添加到购物车
隐藏产品细节-
1
114
228
交叉反应转录物 无
序列位置 7-32
序列细节 正义和反义序列
序列
链

图33D


IDT
INTEGRATED DNA TECHNOLOGIES
客户服务： 800-328-2661 or 319-626-8400
客户服务： custcare@idtdna.com
技术支持： 800-328-2661 or 319-626-8400
订单信息
网络订单 #19244165
感谢您的订单。您订购的物品如下所示。请打印或保存副本以供联系客户服务 1-800-328-2661。
订购更多
返回主页
登出
打印
1 : CD.RL.218115.13.4
Qty: 1
$95.00
产品 DsiRNA, 2 nmol
期望交货日期 2/17/2020
纯化 标准脱盐
保证产量 2 nmol
序列 rGrCrU rGrArU rGrArG rUrArC rGrArA rCrUrU rArUrG rUrAC T
序列2 rArGrU rArCrA rUrArA rGrUrU rCrGrU rArCrU rCrArU rCrArG rCrUrU
2 : CD.RL.218115.13.5
Qty: 1
$95.00
产品 DsiRNA, 2 nmol
期望交货日期 2/17/2020
纯化 标准脱盐
保证产量 2 nmol
序列 rCrArA rGrCrU rGrArU rGrArG rUrArC rGrArA rCrUrU rArUrG T
序列2 rArCrA rUrArA rGrUrU rCrGrU rArCrU rCrArU rCrArG rCrUrU rGrUrG
3 CD.RL.218115.13.10
Qty: 1
$95.00
产品 DsiRNA, 2 nmol
期望交货日期 2/17/2020
纯化 标准脱盐
保证产量 2 nmol
序列 rGrCrA rArGrA rGrArC rArGrG rUrArC rGrUrU rArArU rArGT T
序列2 rArArC rUrArU rUrArA rCrGrU rArCrC rUrGrU rCrUrC rUrUrC rCrGrA
我们做的怎么样?进行 1 分钟的调查
订单摘要
小计 $285.00 USD
S & H $16.00 USD
税 $25.59 USD
总计 $301.00 USD
支付信息
支付类型 PO
订购单号码 Ligandal_nCoV_001
订单选择
交货完毕 Yes
交货单据 No

图33E

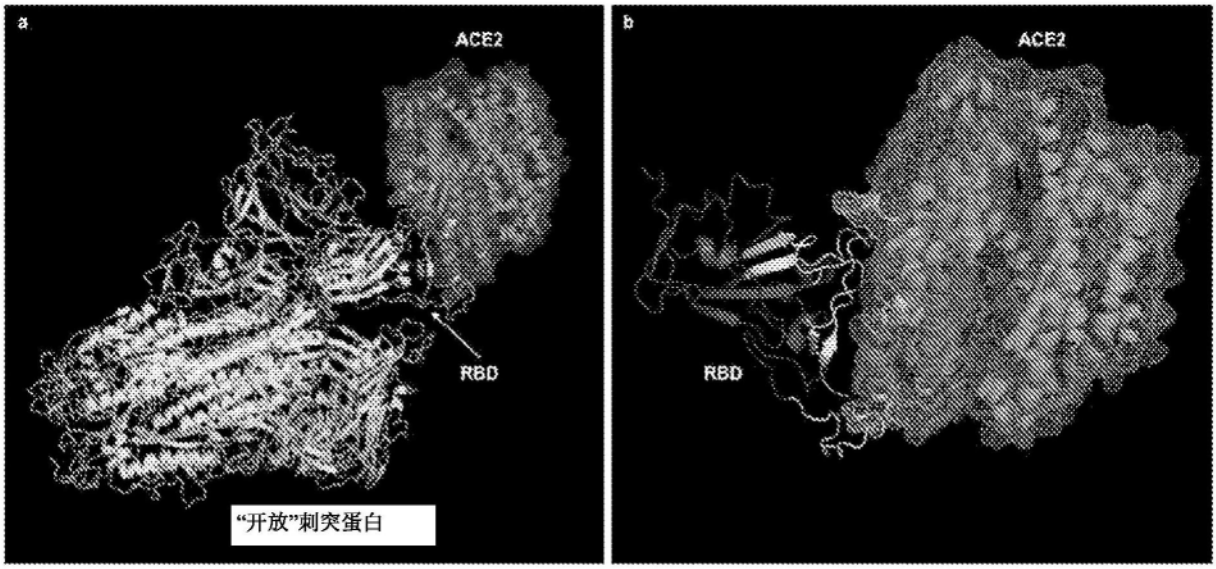


图34

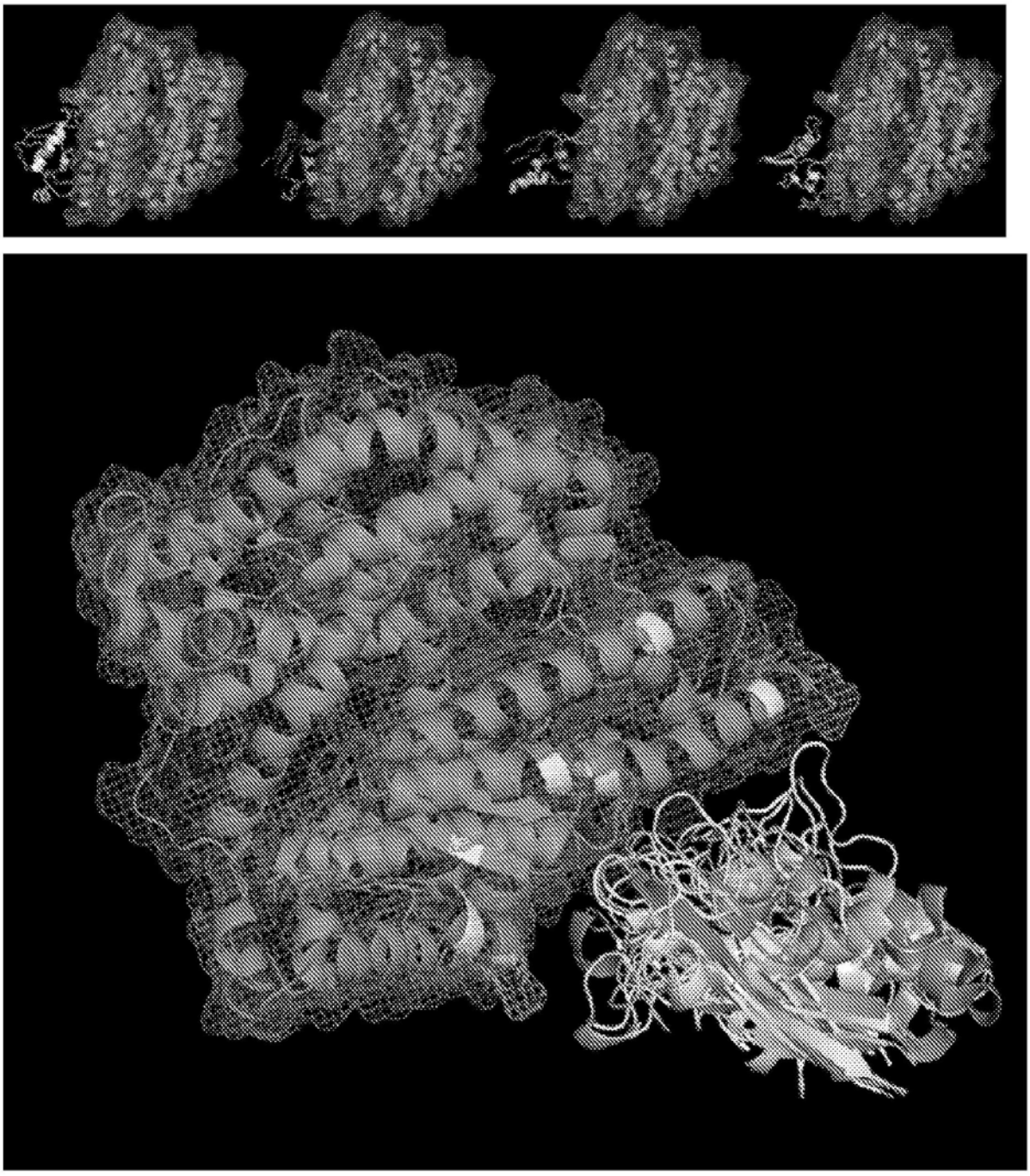


图35

SEQ ID NO: 1

aaaggtttataccttcccaggtaacaaaccaaccaactttcgatctcttgtagatctgttctctaaacgaactttaa
aatctgtgtggctgtcactcggctgcatgcttagtgcactcacgcagtataattaataactaattactgtcgttgac
aggacacgagtaactcgtctatcttctgcaggctgcttacggtttcgtccgtgttgcagccgatcatcagcacatct
aggtttcgtccgggtgtgaccgaaaggtaagatggagagccttgtccctggtttcaacgagaaaacacacgtccaac
tcagtttgcctgttttacaggttcgcgacgtgctcgtacgtggcttttcagaggcacgtcaacatcttaaagatggc
acttgtggcttagtagaagttgaaaaaggcgttttgcctcaacttgaacagccctatgtgttcatcaaacgttcgga
tgctcgaactgcacctcatggtcatgttatggttgagctggtagcagaactcgaaggcattcagtacggtcgtagtg
gtgagacacttggtgtccttgtccctcatgtgggcgaaataccagtggcttaccgcaaggttcttcttcgtaagaac
ggtaataaaggagctggtggccatagttacggcgccgatctaaagtcatttgacttaggcgacgagcttggcactga
tccttatgaagattttcaagaaaactggaacactaaacatagcagtggtgttacccgtgaactcatgcgtgagctta
acggaggggcatacactcgctatgtcgataacaacttctgtggccctgatggctaccctcttgagtgcattaaagac
cttctagcacgtgctggtaaagcttcatgcactttgtccgaacaactggactttattgacactaagaggggtgtata
ctgctgccgtgaacatgagcatgaaattgcttggtacacggaacgttctgaaaagagctatgaattgcagacacctt
ttgaaattaaattggcaaagaaatttgacaccttcaatggggaatgtccaaattttgtatttcccttaaattccata
atcaagactattcaaccaagggttgaaaagaaaaagcttgatggctttatgggtagaattcgatctgtctatccagt
tgcgtcaccaaatgaatgcaaccaaatgtgcctttcaactctcatgaagtgtgatcattgtggtgaaacttcatggc
agacgggcgattttgttaaagccacttgcgaattttgtggcactgagaatttgactaaagaaggtgccactacttgt
ggttacttaccccaaaatgctgttgttaaaatttattgtccagcatgtcacaattcagaagtaggacctgagcatag
tcttgccgaataccataatgaatctggcttgaaaaccattcttcgtaagggtggtcgcactattgcctttggaggct
gtgtgttctcttatgttggttgccataacaagtgtgcctattgggttccacgtgctagcgctaacataggttgtaac
catacaggtgttgttggagaaggttccgaaggtcttaatgacaaccttcttgaaatactccaaaaagagaaagtcaa
catcaatattgttggtgactttaaacttaatgaagagatcgccattattttggcatctttttctgcttccacaagtg
cttttgtggaaactgtgaaaggtttggattataaagcattcaaacaaattgttgaatcctgtggtaattttaaagtt
acaaaaggaaaagctaaaaaaggtgcctggaatattggtgaacagaaatcaatactgagtcctctttatgcatttgc
atcagaggctgctcgtgttgtacgatcaattttctctcgcactcttgaaactgctcaaaattctgtgcgtgttttac
agaaggccgctataacaatactagatggaatttcacagtattcactgagactcattgatgctatgatgttcacatct
gatttggctactaacaatctagttgtaatggcctacattacaggtggtgttgttcagttgacttcgcagtggctaac
taacatctttggcactgtttatgaaaaactcaaacccgtccttgattggcttgaagagaagtttaaggaaggtgtag
agtttcttagagacggttgggaaattgttaaatttatctcaacctgtgcttgtgaaattgtcggtggacaaattgtc
acctgtgcaaaggaaattaaggagagtgttcagacattctttaagcttgtaaataaatttttggctttgtgtgctga
ctctatcattattggtggagctaaacttaaagccttgaatttaggtgaaacatttgtcacgcactcaaagggattgt
acagaaagtgtgttaaatccagagaagaaactggcctactcatgcctctaaaagccccaaaagaaattatcttctta
gagggagaaacacttcccacagaagtgttaacagaggaagttgtcttgaaaactggtgatttacaaccattagaaca
acctactagtgaagctgttgaagctccattggttggtacaccagtttgtattaacgggcttatgttgctcgaaatca
aagacacagaaaagtactgtgcccttgcacctaatatgatggtaacaaacaataccttcacactcaaaggcggtgca
ccaacaaaggttacttttggtgatgacactgtgatagaagtgcaaggttacaagagtgtgaatatcacttttgaact
tgatgaaaggattgataaagtacttaatgagaagtgctctgcctatacagttgaactcggtacagaagtaaatgagt
tcgcctgtgttgtggcagatgctgtcataaaaactttgcaaccagtatctgaattacttacaccactgggcattgat
ttagatgagtggagtatggctacatactacttatttgatgagtctggtgagtttaaattggcttcacatatgtattg
ttctttctaccctccagatgaggatgaagaagaaggtgattgtgaagaagaagagtttgagccatcaactcaatatg
agtatggtactgaagatgattaccaaggtaaacctttggaatttggtgccacttctgctgctcttcaacctgaagaa
gagcaagaagaagattggttagatgatgatagtcaacaaactgttggtcaacaagacggcagtgaggacaatcagac
aactactattcaaacaattgttgaggttcaacctcaattagagatggaacttacaccagttgttcagactattgaag
tgaatagttttagtggttatttaaaacttactgacaatgtatacattaaaaatgcagacattgtggaagaagctaaa
aaggtaaaaccaacagtggttgttaatgcagccaatgtttaccttaaacatggaggaggtgttgcaggagccttaaa
taaggctactaacaatgccatgcaagttgaatctgatgattacatagctactaatggaccacttaaagtgggtggta
gttgtgttttaagcggacacaatcttgctaaacactgtcttcatgttgtcggcccaaatgttaacaaaggtgaagac
attcaacttcttaagagtgcttatgaaaattttaatcagcacgaagttctacttgcaccattattatcagctggtat
ttttggtgctgaccctatacattctttaagagtttgtgtagatactgttcgcacaaatgtctacttagctgtctttg
ataaaaatctctatgacaaacttgtttcaagctttttggaaatgaagagtgaaaagcaagttgaacaaaagatcgct
gagattcctaaagaggaagttaagccatttataactgaaagtaaaccttcagttgaacagagaaaacaagatgataa

图36

gaaaatcaaagcttgtgttgaagaagttacaacaactctggaagaaactaagttcctcacagaaaacttgttacttt
atattgacattaatggcaatcttcatccagattctgccactcttgttagtgacattgacatcactttcttaaagaaa
gatgctccatatatagtgggtgatgttgttcaagagggtgttttaactgctgtggttatacctactaaaaaggctgg
tggcactactgaaatgctagcgaaagctttgagaaaagtgccaacagacaattatataaccacttacccgggtcagg
gtttaaatggttacactgtagaggaggcaaagacagtgcttaaaaagtgtaaaagtgccttttacattctaccatct
attatctctaatgagaagcaagaaattcttggaactgtttcttggaatttgcgagaaatgcttgcacatgcagaaga
aacacgcaaattaatgcctgtctgtgtggaaactaaagccatagtttcaactatacagcgtaaatataagggtatta
aaatacaagagggtgtggttgattatggtgctagattttacttttacaccagtaaaacaactgtagcgtcacttatc
aacacacttaacgatctaaatgaaactcttgttacaatgccacttggctatgtaacacatggcttaaatttggaaga
agctgctcggtatatgagatctctcaaagtgccagctacagtttctgtttcttcacctgatgctgttacagcgtata
atggttatcttacttcttcttctaaaacacctgaagaacattttattgaaaccatctcacttgctggttcctataaa
gattggtcctattctggacaatctacacaactaggtatagaatttcttaagagaggtgataaaagtgtatattacac
tagtaatcctaccacattccacctagatggtgaagttatcacctttgacaatcttaagacacttctttctttgagag
aagtgaggactattaaggtgtttacaacagtagacaacattaacctccacacgcaagttgtggacatgtcaatgaca
tatggacaacagtttggtccaacttatttggatggagctgatgttactaaaataaaacctcataattcacatgaagg
taaaacattttatgttttacctaatgatgacactctacgtgttgaggcttttgagtactaccacacaactgatccta
gttttctgggtaggtacatgtcagcattaaatcacactaaaaagtggaaatacccacaagttaatggtttaacttct
attaaatgggcagataacaactgttatcttgccactgcattgttaacactccaacaaatagagttgaagtttaatcc
acctgctctacaagatgcttattacagagcaagggctggtgaagctgctaacttttgtgcacttatcttagcctact
gtaataagacagtaggtgagttaggtgatgttagagaaacaatgagttacttgtttcaacatgccaatttagattct
tgcaaaagagtcttgaacgtggtgtgtaaaacttgtggacaacagcagacaacccttaagggtgtagaagctgttat
gtacatgggcacactttcttatgaacaatttaagaaaggtgttcagataccttgtacgtgtggtaaacaagctacaa
aatatctagtacaacaggagtcacctttttgttatgatgtcagcaccacctgctcagtatgaacttaagcatggtaca
tttacttgtgctagtgagtacactggtaattaccagtgtggtcactataaacatataacttctaaagaaactttgta
ttgcatagacggtgctttacttacaaagtcctcagaatacaaaggtcctattacggatgttttctacaaagaaaaca
gttacacaacaaccataaaaccagttacttataaattggatggtgttgtttgtacagaaattgaccctaagttggac
aattattataagaaagacaattcttatttcacagagcaaccaattgatcttgtaccaaaccaaccatatccaaacgc
aagcttcgataatttaagtttgtatgtgataatatcaaatttgctgatgatttaaaccagttaactggttataaga
aacctgcttcaagagagcttaaagttacatttttccctgacttaaatggtgatgtggtggctattgattataaacac
tacacaccctcttttaagaaaggagctaaattgttacataaacctattgtttggcatgttaacaatgcaactaataa
agccacgtataaaccaaatacctggtgtatacgttgtctttggagcacaaaaccagttgaaacatcaaattcgtttg
atgtactgaagtcagaggacgcgcagggaatggataatcttgcctgcgaagatctaaaaccagtctctgaagaagta
gtggaaaatcctaccatacagaaagacgttcttgagtgtaatgtgaaaactaccgaagttgtaggagacattatact
taaaccagcaaataatagtttaaaaattacagaagaggttggccacacagatctaatggctgcttatgtagacaatt
ctagtcttactattaagaaacctaatgaattatctagagtattaggtttgaaaaccttgctactcatggtttagct
gctgttaatagtgtcccttgggatactatagctaattatgctaagccttttcttaacaaagttgttagtacaactac
taacatagttacacggtgtttaaaccgtgtttgtactaattatatgccttatttctttacttttattgctacaattgt
gtacttttactagaagtacaaattctagaattaaagcatctatgccgactactatagcaaagaatactgttaagagt
gtcggtaaattttgtctagaggcttcatttaattatttgaagtcacctaatttttctaaactgataaatattataat
ttggtttttactattaagtgtttgcctaggttctttaatctactcaaccgctgctttaggtgttttaatgtctaatt
taggcatgccttcttactgtactggttacagagaaggctatttgaactctactaatgtcactattgcaacctactgt
actggttctataccttgtagtgtttgtcttagtggtttagattctttagacacctatccttctttagaaactataca
aattaccatttcatcttttaaatgggatttaactgcttttggcttagttgcagagtggtttttggcatatattcttt
tcactaggtttttctatgtacttggattggctgcaatcatgcaattgtttttcagctattttgcagtacatttttatt
agtaattcttggcttatgtggttaataattaatcttgtacaaatggccccgatttcagctatggttagaatgtacat
cttctttgcatcattttattatgtatggaaaagttatgtgcatgttgtagacggttgtaattcatcaacttgtatga
tgtgttacaaacgtaatagagcaacaagagtcgaatgtacaactattgttaatggtgttagaaggtccttttatgtc
tatgctaatggaggtaaaggcttttgcaaactacacaattggaattgtgttaattgtgatacattctgtgctggtag
tacatttattagtgatgaagttgcgagagacttgtcactacagtttaaaagaccaataaatcctactgaccagtctt
cttacatcgttgatagtgttacagtgaagaatggttccatccatctttactttgataaagctggtcaaaagacttat
gaaagacattctctctctcattttgttaacttagacaacctgagagctaataacactaaaggttcattgcctattaa
tgttatagtttttgatggtaaatcaaaatgtgaagaatcatctgcaaaatcagcgtctgtttactacagtcagctta
tgtgtcaacctatactgttactagatcaggcattagtgtctgatgttggtgatagtgcggaagttgcagttaaaatg
tttgatgcttacgttaatacgttttcatcaacttttaacgtaccaatggaaaaactcaaaacactagttgcaactgc

图36续

agaagctgaacttgcaaagaatgtgtccttagacaatgtcttatctacttttatttcagcagctcggcaagggtttg
ttgattcagatgtagaaactaaagatgttgttgaatgtcttaaattgtcacatcaatctgacatagaagttactggc
gatagttgtaataactatatgctcacctataacaaagttgaaaacatgacaccccgtgaccttggtgcttgtattga
ctgtagtgcgcgtcatattaatgcgcaggtagcaaaaagtcacaacattgctttgatatggaacgttaaagatttca
tgtcattgtctgaacaactacgaaaacaaatacgtagtgctgctaaaaagaataacttacctttttaagttgacatgt
gcaactactagacaagttgttaatgttgtaacaacaaagatagcacttaagggtggtaaaattgttaataattggtt
gaagcagttaattaaagttacacttgtgttccttttgttgctgctattttctatttaataacacctgttcatgtca
tgtctaaacatactgacttttcaagtgaaatcataggatacaaggctattgatggtggtgtcactcgtgacatagca
tctacagatacttgttttgctaacaaacatgctgattttgacacatggtttagccagcgtggtggtagttatactaa
tgacaaagcttgcccattgattgctgcagtcataacaagagaagtgggttttgtcgtgcctggtttgcctggcacga
tattacgcacaactaatggtgacttttgcatttcttacctagagtttttagtgcagttggtaacatctgttacaca
ccatcaaaacttatagagtacactgactttgcaacatcagcttgtgttttggctgctgaatgtacaatttttaaaga
tgcttctggtaagccagtaccatattgttatgataccaatgtactagaaggttctgttgcttatgaaagtttacgcc
ctgacacacgttatgtgctcatggatggctctattattcaatttcctaacacctaccttgaaggttctgttagagtg
gtaacaacttttgattctgagtactgtaggcacggcacttgtgaaagatcagaagctggtgtttgtgtatctactag
tggtagatgggtacttaacaatgattattacagatctttaccaggagttttctgtggtgtagatgctgtaaatttac
ttactaatatgtttacaccactaattcaacctattggtgctttggacatatcagcatctatagtagctggtggtatt
gtagctatcgtagtaacatgccttgcctactattttatgaggtttagaagagcttttggtgaatacagtcatgtagt
tgcctttaatactttactattccttatgtcattcactgtactctgtttaacaccagtttactcattcttacctggtg
tttattctgttatttacttgtacttgacattttatcttactaatgatgtttctttttagcacatattcagtggatg
gttatgttcacacctttagtacctttctggataacaattgcttatatcatttgtatttccacaaagcatttctattg
gttctttagtaattacctaaagagacgtgtagtctttaatggtgtttccttttagtacttttgaagaagctgcgctgt
gcacctttttgttaaataaagaaatgtatctaaagttgcgtagtgatgtgctattacctcttacgcaatataataga
tacttagctctttataataagtacaagtattttagtggagcaatggatacaactagctacagagaagctgcttgttg
tcatctcgcaaaggctctcaatgacttcagtaactcaggttctgatgttctttaccaaccaccacaaacctctatca
cctcagctgttttgcagagtggttttagaaaaatggcattcccatctggtaaagttgagggttgtatggtacaagta
acttgtggtacaactacacttaacggtctttggcttgatgacgtagtttactgtccaagacatgtgatctgcacctc
tgaagacatgcttaaccctaattatgaagatttactcattcgtaagtctaatcataatttcttggtacaggctggta
atgttcaactcagggttattggacattctatgcaaaattgtgtacttaagcttaaggttgatacagccaatcctaag
acacctaagtataagtttgttcgcattcaaccaggacagacttttcagtgttagcttgttacaatggttcaccatc
tggtgtttaccaatgtgctatgaggcccaatttcactattaagggttcattccttaatggttcatgtggtagtgttg
gttttaacatagattatgactgtgtctcttttttgttacatgcaccatatggaattaccaactggagttcatgctggc
acagacttagaaggtaacttttatggacctttttgttgacaggcaaacagcacaagcagctggtacggacacaactat
tacagttaatgttttagcttggttgtacgctgctgttataaatggagacaggtggtttctcaatcgatttaccacaa
ctcttaatgactttaaccttgtggctatgaagtacaattatgaacctctaacacaagaccatgttgacatactagga
cctctttctgctcaaactggaattgccgttttagatatgtgtgcttcattaaaagaattactgcaaaatggtatgaa
tggacgtaccatattgggtagtgctttattagaagatgaatttacacctttgatgttgttagacaatgctcaggtg
ttactttccaaagtgcagtgaaaagaacaatcaagggtacacaccactggttgttactcacaattttgacttcactt
ttagttttagtccagagtactcaatggtctttgttctttttttgtatgaaaatgccttttaccttttgctatggg
tattattgctatgtctgcttttgcaatgatgtttgtcaaacataagcatgcatttctctgtttgtttttgttacctt
ctcttgccactgtagcttattttaatatggtctatatgcctgctagttgggtgatgcgtattatgacatggttggat
atggttgatactagtttgtctggttttaagctaaaagactgtgttatgtatgcatcagctgtagtgttactaatcct
tatgacagcaagaactgtgtatgatgatggtgctaggagagtgtggacacttatgaatgtcttgacactcgtttata
aagtttattatggtaatgctttagatcaagccatttccatgtgggctcttataatctctgttacttctaactactca
ggtgtagttacaactgtcatgtttttggccagaggtattgtttttatgtgtgttgagtattgccctattttcttcat
aactggtaatacacttcagtgtataatgctagtttattgtttcttaggctatttttgtacttgttactttggcctct
tttgtttactcaaccgctactttagactgactcttggtgtttatgattacttagtttctacacaggagtttagatat
atgaattcacagggactactcccacccaagaatagcatagatgccttcaaactcaacattaaattgttgggtgttgg
tggcaaaccttgtatcaaagtagccactgtacagtctaaaatgtcagatgtaaagtgcacatcagtagtcttactct
cagttttgcaacaactcagagtagaatcatcatctaaattgtgggctcaatgtgtccagttacacaatgacattctc
ttagctaaagatactactgaagcctttgaaaaaatggtttcactactttctgttttgctttccatgcagggtgctgt
agacataaacaagctttgtgaagaaatgctggacaacagggcaaccttacaagctatagcctcagagtttagttccc
ttccatcatatgcagcttttgctactgctcaagaagcttatgagcaggctgttgctaatggtgattctgaagttgtt
cttaaaaagttgaagaagtctttgaatgtggctaaatctgaatttgaccgtgatgcagccatgcaacgtaagttgga

图36续

aaagatggctgatcaagctatgacccaaatgtataaacaggctagatctgaggacaagagggcaaaagttactagtg
ctatgcagacaatgcttttcactatgcttagaaagttggataatgatgcactcaacaacattatcaacaatgcaaga
gatggttgtgttcccttgaacataatacctcttacaacagcagccaaactaatggttgtcataccagactataacac
atataaaaatacgtgtgatggtacaacatttacttatgcatcagcattgtgggaaatccaacaggttgtagatgcag
atagtaaaattgttcaacttagtgaaattagtatggacaattcacctaatttagcatggcctcttattgtaacagct
ttaagggccaattctgctgtcaaattacagaataatgagcttagtcctgttgcactacgacagatgtcttgtgctgc
cggtactacacaaactgcttgcactgatgacaatgcgttagcttactacaacacaacaaagggaggtaggtttgtac
ttgcactgttatccgatttacaggatttgaaatgggctagattccctaagagtgatggaactggtactatctataca
gaactggaaccaccttgtaggtttgttacagacacacctaaaggtcctaaagtgaagtatttatactttattaaagg
attaaacaacctaaatagaggtatggtacttggtagtttagctgccacagtacgtctacaagctggtaatgcaacag
aagtgcctgccaattcaactgtattatctttctgtgcttttgctgtagatgctgctaaagcttacaaagattatcta
gctagtgggggacaaccaatcactaattgtgttaagatgttgtgtacacacactggtactggtcaggcaataacagt
tacaccggaagccaatatggatcaagaatcctttggtggtgcatcgtgttgtctgtactgccgttgccacatagatc
atccaaatcctaaaggattttgtgacttaaaaggtaagtatgtacaaatacctacaacttgtgctaatgaccctgtg
ggttttacacttaaaaacacagtctgtaccgtctgcggtatgtggaaaggttatggctgtagttgtgatcaactccg
cgaacccatgcttcagtcagctgatgcacaatcgtttttaaacgggtttgcggtgtaagtgcagcccgtcttacacc
gtgcggcacaggcactagtactgatgtcgtatacagggcttttgacatctacaatgataaagtagctggttttgcta
aattcctaaaaactaattgttgtcgcttccaagaaaaggacgaagatgacaatttaattgattcttactttgtagtt
aagagacacactttctctaactaccaacatgaagaaacaatttataatttacttaaggattgtccagctgttgctaa
acatgacttctttaagtttagaatagacggtgacatggtaccacatatatcacgtcaacgtcttactaaatacacaa
tggcagacctcgtctatgctttaaggcattttgatgaaggtaattgtgacacattaaaagaaatacttgtcacatac
aattgttgtgatgatgattatttcaataaaaaggactggtatgattttgtagaaaacccagatatattacgcgtata
cgccaacttaggtgaacgtgtacgccaagctttgttaaaaacagtacaattctgtgatgccatgcgaaatgctggta
ttgttggtgtactgacattagataatcaagatctcaatggtaactggtatgatttcggtgatttcatacaaaccacg
ccaggtagtggagttcctgttgtagattcttattattcattgttaatgcctatattaaccttgaccagggctttaac
tgcagagtcacatgttgacactgacttaacaaagccttacattaagtgggatttgttaaaatatgacttcacggaag
agaggttaaaactctttgaccgttattttaaatattgggatcagacataccacccaaattgtgttaactgtttggat
gacagatgcattctgcattgtgcaaactttaatgttttattctctacagtgttcccacctacaagttttggaccact
agtgagaaaaatatttgttgatggtgttccatttgtagtttcaactggataccacttcagagagctaggtgttgtac
ataatcaggatgtaaacttacatagctctagacttagttttaaggaattacttgtgtatgctgctgaccctgctatg
cacgctgcttctggtaatctattactagataaacgcactacgtgcttttcagtagctgcacttactaacaatgttgc
ttttcaaactgtcaaacccggtaattttaacaaagacttctatgactttgctgtgtctaagggtttctttaaggaag
gaagttctgttgaattaaaacacttcttctttgctcaggatggtaatgctgctatcagcgattatgactactatcgt
tataatctaccaacaatgtgtgatatcagacaactactatttgtagttgaagttgttgataagtactttgattgtta
cgatggtggctgtattaatgctaaccaagtcatcgtcaacaacctagacaaatcagctggttttccatttaataaat
ggggtaaggctagactttattatgattcaatgagttatgaggatcaagatgcacttttcgcatatacaaaacgtaat
gtcatccctactataactcaaatgaatcttaagtatgccattagtgcaaagaatagagctcgcaccgtagctggtgt
ctctatctgtagtactatgaccaatagacagtttcatcaaaaattattgaaatcaatagccgccactagaggagcta
ctgtagtaattggaacaagcaaattctatggtggttggcacaacatgttaaaaactgtttatagtgatgtagaaaac
cctcaccttatgggttgggattatcctaaatgtgatagagccatgcctaacatgcttagaattatggcctcacttgt
tcttgctcgcaaacatacaacgtgttgtagcttgtcacaccgtttctatagattagctaatgagtgtgctcaagtat
tgagtgaaatggtcatgtgtggcggttcactatatgttaaaccaggtggaacctcatcaggagatgccacaactgct
tatgctaatagtgtttttaacatttgtcaagctgtcacggccaatgttaatgcacttttatctactgatggtaacaa
aattgccgataagtatgtccgcaatttacaacacagactttatgagtgtctctatagaaatagagatgttgacacag
actttgtgaatgagttttacgcatatttgcgtaaacatttctcaatgatgatactctctgacgatgctgttgtgtgt
ttcaatagcacttatgcatctcaaggtctagtggctagcataaagaactttaagtcagttctttattatcaaaacaa
tgtttttatgtctgaagcaaaatgttggactgagactgaccttactaaaggacctcatgaattttgctctcaacata
caatgctagttaaacagggtgatgattatgtgtaccttccttacccagatccatcaagaatcctaggggccggctgt
tttgtagatgatatcgtaaaaacagatggtacacttatgattgaacggttcgtgtctttagctatagatgcttaccc
acttactaaacatcctaatcaggagtatgctgatgtctttcatttgtacttacaatacataagaaagctacatgatg
agttaacaggacacatgttagacatgtattctgttatgcttactaatgataacacttcaaggtattgggaacctgag
ttttatgaggctatgtacacaccgcatacagtcttacaggctgttggggcttgtgttctttgcaattcacagacttc
attaagatgtggtgcttgcatacgtagaccattcttatgttgtaaatgctgttacgaccatgtcatatcaacatcac

图36续

ataaattagtcttgtctgttaatccgtatgtttgcaatgctccaggttgtgatgtcacagatgtgactcaactttac
ttaggaggtatgagctattattgtaaatcacataaaccacccattagttttccattgtgtgctaatggacaagtttt
tggtttatataaaaatacatgtgttggtagcgataatgttactgactttaatgcaattgcaacatgtgactggacaa
atgctggtgattacattttagctaacacctgtactgaaagactcaagctttttgcagcagaaacgctcaaagctact
gaggagacatttaaactgtcttatggtattgctactgtacgtgaagtgctgtctgacagagaattacatctttcatg
ggaagttggtaaacctagaccaccacttaaccgaaattatgtctttactggttatcgtgtaactaaaaacagtaaag
tacaaataggagagtacacctttgaaaaaggtgactatggtgatgctgttgtttaccgaggtacaacaacttacaaa
ttaaatgttggtgattattttgtgctgacatcacatacagtaatgccattaagtgcacctacactagtgccacaaga
gcactatgttagaattactggcttatacccaacactcaatatctcagatgagttttctagcaatgttgcaaattatc
aaaaggttggtatgcaaaagtattctacactccagggaccacctggtactggtaagagtcattttgctattggccta
gctctctactacccttctgctcgcatagtgtatacagcttgctctcatgccgctgttgatgcactatgtgagaaggc
attaaaatatttgcctatagataaatgtagtagaattatacctgcacgtgctcgtgtagagtgttttgataaattca
aagtgaattcaacattagaacagtatgtcttttgtactgtaaatgcattgcctgagacgacagcagatatagttgtc
tttgatgaaatttcaatggccacaaattatgatttgagtgttgtcaatgccagattacgtgctaagcactatgtgta
cattggcgaccctgctcaattacctgcaccacgcacattgctaactaagggcacactagaaccagaatatttcaatt
cagtgtgtagacttatgaaaactataggtccagacatgttcctcggaacttgtcggcgttgtcctgctgaaattgtt
gacactgtgagtgctttggtttatgataataagcttaaagcacataaagacaaatcagctcaatgctttaaaatgtt
ttataagggtgttatcacgcatgatgtttcatctgcaattaacaggccacaaataggcgtggtaagagaattcctta
cacgtaaccctgcttggagaaaagctgtctttatttcaccttataattcacagaatgctgtagcctcaaagattttg
ggactaccaactcaaactgttgattcatcacagggctcagaatatgactatgtcatattcactcaaaccactgaaac
agctcactcttgtaatgtaaacagatttaatgttgctattaccagagcaaaagtaggcatactttgcataatgtctg
atagagacctttatgacaagttgcaatttacaagtcttgaaattccacgtaggaatgtggcaactttacaagctgaa
aatgtaacaggactctttaaagattgtagtaaggtaatcactgggttacatcctacacaggcacctacacacctcag
tgttgacactaaattcaaaactgaaggtttatgtgttgacatacctggcatacctaaggacatgacctatagaagac
tcatctctatgatgggttttaaaatgaattatcaagttaatggttaccctaacatgtttatcacccgcgaagaagct
ataagacatgtacgtgcatggattggcttcgatgtcgaggggtgtcatgctactagagaagctgttggtaccaattt
acctttacagctaggtttttctacaggtgttaacctagttgctgtacctacaggttatgttgatacacctaataata
cagatttttccagagttagtgctaaaccaccgcctggagatcaatttaaacacctcataccacttatgtacaaagga
cttctttggaatgtagtgcgtataaagattgtacaaatgttaagtgacacacttaaaaatctctctgacagagtcgt
atttgtcttatgggcacatggctttgagttgacatctatgaagtattttgtgaaaataggacctgagcgcacctgtt
gtctatgtgatagacgtgccacatgcttttccactgcttcagacacttatgcctgttggcatcattctattggattt
gattacgtctataatccgtttatgattgatgttcaacaatggggttttacaggtaacctacaaagcaaccatgatct
gtattgtcaagtccatggtaatgcacatgtagctagttgtgatgcaatcatgactaggtgtctagctgtccacgagt
gctttgttaagcgtgttgactggactattgaatatcctataattggtgatgaactgaagattaatgcggcttgtaga
aaggttcaacacatggttgttaaagctgcattattagcagacaaattcccagttcttcacgacattggtaaccctaa
agctattaagtgtgtacctcaagctgatgtagaatggaagttctatgatgcacagccttgtagtgacaaagcttata
aaatagaagaattattctattcttatgccacacattctgacaaattcacagatggtgtatgcctattttggaattgc
aatgtcgatagatatcctgctaattccattgtttgtagatttgacactagagtgctatctaaccttaacttgcctgg
ttgtgatggtggcagtttgtatgtaaataaacatgcattccacacaccagcttttgataaaagtgcttttgttaatt
taaaacaattaccatttttctattactctgacagtccatgtgagtctcatggaaaacaagtagtgtcagatatagat
tatgtaccactaaagtctgctacgtgtataacacgttgcaatttaggtggtgctgtctgtagacatcatgctaatga
gtacagattgtatctcgatgcttataacatgatgatctcagctggctttagcttgtgggtttacaaacaatttgata
cttataacctctggaacacttttacaagacttcagagtttagaaaatgtggcttttaatgttgtaaataagggacac
tttgatggacaacagggtgaagtaccagtttctatcattaataacactgtttacacaaaagttgatggtgttgatgt
agaattgtttgaaaataaaacaacattacctgttaatgtagcatttgagctttgggctaagcgcaacattaaaccag
taccagaggtgaaaatactcaataatttgggtgtggacattgctgctaatactgtgatctgggactacaaaagagat
gctccagcacatatatctactattggtgtttgttctatgactgacatagccaagaaaccaactgaaacgatttgtgc
accactcactgtctttttgatggtagagttgatggtcaagtagacttatttagaaatgcccgtaatggtgttctta
ttacagaaggtagtgttaaaggtttacaaccatctgtaggtcccaaacaagctagtcttaatggagtcacattaatt
ggagaagccgtaaaaacacagttcaattattataagaaagttgatggtgttgtccaacaattacctgaaacttactt
tactcagagtagaaatttacaagaatttaaacccaggagtcaaatggaaattgatttcttagaattagctatggatg
aattcattgaacggtataaattagaaggctatgccttcgaacatatcgtttatggagattttagtcatagtcagtta
ggtggtttacatctactgattggactagctaaacgttttaaggaatcacctttgaattagaagattttattcctat
ggacagtacagttaaaaactatttcataacagatgcgcaaacaggttcatctaagtgtgtgtgttctgttattgatt

图36续

tattacttgatgattttgttgaaataataaaatcccaagatttatctgtagtttctaaggttgtcaaagtgactatt
gactatacagaaatttcatttatgctttggtgtaaagatggccatgtagaaacattttacccaaaattacaatctag
tcaagcgtggcaaccgggtgttgctatgcctaatctttacaaaatgcaaagaatgctattagaaaagtgtgaccttc
aaaattatggtgatagtgcaacattacctaaaggcataatgatgaatgtcgcaaaatatactcaactgtgtcaatat
ttaaacacattaacattagctgtaccctataatatgagagttatacattttggtgctggttctgataaaggagttgc
accaggtacagctgttttaagacagtggttgcctacgggtacgctgcttgtcgattcagatcttaatgactttgtct
ctgatgcagattcaactttgattggtgattgtgcaactgtacatacagctaataaatgggatctcattattagtgat
atgtacgaccctaagactaaaaatgttacaaaagaaaatgactctaaagagggtttttcacttacatttgtgggtt
tatacaacaaaagctagctcttggaggttccgtggctataaagataacagaacattcttggaatgctgatctttata
agctcatgggacacttcgcatggtggacagcctttgttactaatgtgaatgcgtcatcatctgaagcatttttaatt
ggatgtaattatcttggcaaaccacgcgaacaaatagatggttatgtcatgcatgcaaattacatattttggaggaa
tacaaatccaattcagttgtcttcctattctttatttgacatgagtaaatttccccttaaattaaggggtactgctg
ttatgtctttaaaagaaggtcaaatcaatgatatgattttatctcttcttagtaaaggtagacttataattagagaa
aacaacagagttgttatttctagtgatgttcttgttaacaactaaacgaacaatgtttgtttttcttgttttattgc
cactagtctctagtcagtgtgttaatcttacaaccagaactcaattacccccctgcatacactaattctttcacacgt
ggtgtttattaccctgacaaagttttcagatcctcagttttacattcaactcaggacttgttcttacctttcttttc
caatgttacttggttccatgctatacatgtctctgggaccaatggtactaagaggtttgataaccctgtcctaccat
ttaatgatggtgtttattttgcttccactgagaagtctaacataataagaggctggatttttggtactactttagat
tcgaagacccagtccctacttattgttaataacgctactaatgttgttattaaagtctgtgaatttcaattttgtaa
tgatccatttttgggtgtttattaccacaaaaacaacaaaagttggatggaaagtgagttcagagtttattctagtg
cgaataattgcacttttgaatatgtctctcagccttttcttatggaccttgaaggaaaacagggtaatttcaaaaat
cttagggaatttgtgtttaagaatattgatggttatttaaaatatattctaagcacacgcctattaatttagtgcg
tgatctccctcagggtttttcggctttagaaccattggtagatttgccaataggtattaacatcactaggtttcaaa
ctttacttgctttacatagaagttatttgactcctggtgattcttcttcaggttggacagctggtgctgcagcttat
tatgtgggttatcttcaacctaggacttttctattaaaatataatgaaaatggaaccattacagatgctgtagactg
tgcacttgaccctctctcagaaacaaagtgtacgttgaaatccttcactgtagaaaaaggaatctatcaaacttcta
actttagagtccaaccaacagaatctattgttagatttcctaatattacaaacttgtgccctttggtgaagttttt
aacgccaccagatttgcatctgtttatgcttggaacaggaagagaatcagcaactgtgttgctgattattctgtcct
atataattccgcatcattttccacttttaagtgttatggagtgtctcctactaaattaaatgatctctgctttacta
atgtctatgcagattcatttgtaattagaggtgatgaagtcagacaaatcgctccagggcaaactggaaagattgct
gattataattataaattaccagatgattttacaggctgcgttatagcttggaattctaacaatcttgattctaaggt
tggtggtaattataattacctgtatagattgtttaggaagtctaatctcaaacctttgagagagatatttcaactg
aaatctatcaggccggtagcacaccttgtaatggtgttgaaggttttaattgttactttcctttacaatcatatggt
ttccaacccactaatggtgttggttaccaaccatacagagtagtagtactttcttttgaacttctacatgcaccagc
aactgtttgtggacctaaaaagtctactaatttggttaaaaacaaatgtgtcaatttcaacttcaatggtttaacag
gcacaggtgttcttactgagtctaacaaaaagtttctgcctttccaacaatttggcagagacattgctgacactact
gatgctgtccgtgatccacagacacttgagattcttgacattacaccatgttcttttggtggtgtcagtgttataac
accaggaacaaatacttctaaccaggttgctgttcttttatcaggatgttaactgcacagaagtccctgttgctattc
atgcagatcaacttactcctacttggcgtgtttattctacaggttctaatgtttttcaaacacgtgcaggctgttta
ataggggctgaacatgtcaacaactcatatgagtgtgacatacccattggtgcaggtatatgcgctagttatcagac
tcagactaattctcctcggcgggcacgtagtgtagctagtcaatccatcattgcctacactatgtcacttggtgcag
aaaattcagttgcttactctaataactctattgccatacccacaaattttactattagtgttaccacagaaattcta
ccagtgtctatgaccaagacatcagtagattgtacaatgtacatttgtggtgattcaactgaatgcagcaatctttt
gttgcaatatggcagtttttgtacacaattaaaccgtgctttaactggaatagctgttgaacaagacaaaaacaccc
aagaagtttttgcacaagtcaaacaaatttacaaaacaccaccaattaaagattttggtggttttaatttttcacaa
atattaccagatccatcaaaaccaagcaagaggtcatttattgaagatctacttttcaacaaagtgacacttgcaga
tgctggcttcatcaaacaatatggtgattgccttggtgatattgctgctagagacctcatttgtgcacaaaagttta
acggccttactgttttgccacctttgctcacagatgaaatgattgctcaatacacttctgcactgttagcgggtaca
atcacttctggttggacctttggtgcaggtgctgcattacaaataccatttgctatgcaaatggcttataggtttaa
tggtattggagttacacagaatgttctctatgagaaccaaaaattgattgccaaccaatttaatagtgctattggca
aaattcaagactcactttcttccacagcaagtgcacttggaaaacttcaagatgtggtcaaccaaaatgcacaagct
ttaaacacgcttgttaaacaacttagctccaattttggtgcaatttcaagtgttttaaatgatatcctttcacgtct
tgacaaagttgaggctgaagtgcaaattgataggttgatcacaggcagacttcaaagtttgcagacatatgtgactc
aacaattaattagagctgcagaaatcagagcttctgctaatcttgctgctactaaaatgtcagagtgtgtacttgga

图36续

caatcaaaaagagttgatttttgtggaaagggctatcatcttatgtccttccctcagtcagcacctcatggtgtagt
cttcttgcatgtgacttatgtccctgcacaagaaaagaacttcacaactgctcctgccatttgtcatgatggaaaag
cacactttcctcgtgaaggtgtctttgtttcaaatggcacacactggtttgtaacacaaaggaatttttatgaacca
caaatcattactacagacaacacatttgtgtctggtaactgtgatgttgtaataggaattgtcaacaacacagttta
tgatcctttgcaacctgaattagactcattcaaggaggagttagataaatattttaagaatcatacatcaccagatg
ttgatttaggtgacatctctggcattaatgcttcagttgtaaacattcaaaaagaaattgaccgcctcaatgaggtt
gccaagaatttaaatgaatctctcatcgatctccaagaacttggaaagtatgagcagtatataaaatggccatggta
catttggctaggttttatagctggcttgattgccatagtaatggtgacaattatgctttgctgtatgaccagttgct
gtagttgtctcaagggctgttgttcttgtggatcctgctgcaaatttgatgaagacgactctgagccagtgctcaaa
ggagtcaaattacattacacataaacgaacttatggatttgtttatgagaatcttcacaattggaactgtaactttg
aagcaaggtgaaatcaaggatgctactccttcagattttgttcgcgctactgcaacgataccgatacaagcctcact
ccctttcggatggcttattgttggcgttgcacttcttgctgtttttcagagcgcttccaaaatcataaccctcaaaa
agagatggcaactagcactctccaagggtgttcactttgtttgcaacttgctgttgttgtttgtaacagtttactca
caccttttgctcgttgctgctggccttgaagccccttttctctatctttatgctttagtctacttcttgcagagtat
aaactttgtaagaataataatgaggctttggctttgctggaaatgccgttccaaaaacccattactttatgatgcca
actattttctttgctggcatactaattgttacgactattgtataccttacaatagtgtaacttcttcaattgtcatt
acttcaggtgatggcacaacaagtcctatttctgaacatgactaccagattggtggttatactgaaaaatgggaatc
tggagtaaaagactgtgttgtattacacagttacttcacttcagactattaccagctgtactcaactcaattgagta
cagacactggtgttgaacatgttaccttcttcatctacaataaaattgttgatgagcctgaagaacatgtccaaatt
cacacaatcgacggttcatccggagttgttaatccagtaatggaaccaatttatgatgaaccgacgacgactactag
cgtgcctttgtaagcacaagctgatgagtacgaacttatgtactcattcgtttcggaagagacaggtacgttaatag
ttaatagcgtacttctttttcttgctttcgtggtattcttgctagttacactagccatccttactgcgcttcgattg
tgtgcgtactgctgcaatattgttaacgtgagtcttgtaaaaccttctttttacgtttactctcgtgttaaaaatct
gaattcttctagagttcctgatcttctggtctaaacgaactaaatattatattagtttttctgtttggaactttaat
tttagccatggcagattccaacggtactattaccgttgaagagcttaaaaagctccttgaacaatggaacctagtaa
taggtttcctattccttacatggatttgtcttctacaatttgcctatgccaacaggaataggtttttgtatataatt
aagttaattttcctctggctgttatggccagtaactttagcttgttttgtgcttgctgctgtttacagaataaattg
gatcaccggtggaattgctatcgcaatggcttgtcttgtaggcttgatgtggctcagctacttcattgcttctttca
gactgtttgcgcgtacgcgttccatgtggtcattcaatccagaaactaacattcttctcaacgtgccactccatggc
actattctgaccagaccgcttctagaaagtgaactcgtaatcggagctgtgatccttcgtggacatcttcgtattgc
tggacaccatctaggacgctgtgacatcaaggacctgcctaaagaaatcactgttgctacatcacgaacgctttctt
attacaaattgggagcttcgcagcgtgtagcaggtgactcaggttttgctgcatacagtcgctacaggattggcaac
tataaattaaacacagaccattccagtagcagtgacaatattgctttgcttgtacagtaagtgacaacagatgtttc
atctcgttgactttcaggttactatagcagagatattactaattattatgaggacttttaaagtttccatttggaat
cttgattacatcataaacctcataattaaaaatttatctaagtcactaactgagaataaatattctcaattagatga
agagcaaccaatggagattgattaaacgaacatgaaaattattcttttcttggcactgataacactcgctacttgtg
agctttatcactaccaagagtgtgttagaggtacaacagtacttttaaaagaaccttgctcttctggaacatacgag
ggcaattcaccatttcatcctctagctgataacaaatttgcactgacttgctttagcactcaatttgcttttgcttg
tcctgacggcgtaaaacacgtctatcagttacgtgccagatcagtttcacctaaactgttcatcagacaagaggaag
ttcaagaactttactctccaatttttcttattgttgcggcaatagtgtttataacactttgcttcacactcaaaaga
aagacagaatgattgaactttcattaattgacttctatttgtgctttttagcctttctgctattccttgttttaatt
atgcttattatcttttggttctcacttgaactgcaagatcataatgaaacttgtcacgcctaaacgaacatgaaatt
tcttgttttcttaggaatcatcacaactgtagctgcatttcaccaagaatgtagtttacagtcatgtactcaacatc
aaccatatgtagttgatgacccgtgtcctattcacttctattctaaatggtatattagagtaggagctagaaaatca
gcacctttaattgaattgtgcgtggatgaggctggttctaaatcacccattcagtacatcgatatcggtaattatac
agtttcctgtttacctttacaattaattgccaggaacctaaattgggtagtcttgtagtgcgttgttcgttctatg
aagacttttagagtatcatgacgttcgtgttgttttagatttcatctaaacgaacaaactaaaatgtctgataatg
gaccccaaaatcagcgaaatgcaccccgcattacgtttggtggaccctcagattcaactggcagtaaccagaatgga
gaacgcagtggggcgcgatcaaaacaacgtcggccccaaggtttacccaataatactgcgtcttggttcaccgctct
cactcaacatggcaaggaagaccttaaattccctcgaggacaaggcgttccaattaacaccaatagcagtccagatg
accaaattggctactaccgaagagctaccagacgaattcgtggtggtgacggtaaaatgaaagatctcagtccaaga
tggtatttctactacctaggaactgggccagaagctggacttccctatggtgctaacaaagacggcatcatatgggt
tgcaactgagggagccttgaatacaccaaaagatcacattggcacccgcaatcctgctaacaatgctgcaatcgtgc
tacaacttcctcaaggaacaacattgccaaaaggcttctacgcagaagggagcagaggcggcagtcaagcctcttct

图36续

cgttcctcatcacgtagtcgcaacagttcaagaaattcaactccaggcagcagtaggggaacttctcctgctagaat
ggctggcaatggcggtgatgctgctcttgctttgctgctgcttgacagattgaaccagcttgagagcaaaatgtctg
gtaaaggccaacaacaacaaggccaaactgtcactaagaaatctgctgctgaggcttctaagaagcctcggcaaaaa
cgtactgccactaaagcatacaatgtaacacaagctttcggcagacgtggtccagaacaaacccaaggaaattttgg
ggaccaggaactaatcagacaaggaactgattacaaacattggccgcaaattgcacaatttgcccccagcgcttcag
cgttcttcggaatgtcgcgcattggcatggaagtcacaccttcgggaacgtggttgacctacacaggtgccatcaaa
ttggatgacaaagatccaaatttcaaagatcaagtcattttgctgaataagcatattgacgcatacaaaacattccc
accaacagagcctaaaaaggacaaaaagaagaaggctgatgaaactcaagccttaccgcagagacagaagaaacagc
aaactgtgactcttcttcctgctgcagatttggatgatttctccaaacaattgcaacaatccatgagcagtgctgac
tcaactcaggcctaaactcatgcagaccacacaaggcagatgggctatataaacgttttcgcttttccgtttacgat
atatagtctactcttgtgcagaatgaattctcgtaactacatagcacaagtagatgtagttaactttaatctcacat
agcaatctttaatcagtgtgtaacattagggaggacttgaaagagccaccacattttcaccgaggccacgcggagta
cgatcgagtgtacagtgaacaatgctagggagagctgcctatatggaagagccctaatgtgtaaaattaattttagt
agtgctatccccatgtgattttaatagcttcttaggagaatgacaaaaa

图36续

SEQ ID NO: 2

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSG
TNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVY
YHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRD
LPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITD
AVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTG
CVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNG
VGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTT
DAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQT
RAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFT
ISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKT
PPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLL
TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQD
SLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYV
TQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA
PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKE
ELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA
GLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

图37

SEQ ID NO: 140

MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWS
AFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTILNTMSTIYSTGKVCNPDNPQEC
LLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEV
NGVDGYDYSRGQLIEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSYISPIGCLPAHLLGDMWGRFWTNLY
SLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVSVGLPNMTQGFWENSMLTDPGNVQKAVCHPTA
WDLGKGDFRILMCTKVTMDDFLTAHHEMGHIQYDMAYAAQPFLLRNGANEGFHEAVGEIMSLSAATPKH
LKSIGLLSPDFQEDNETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKKWWEMKREIVGV
VEPVPHDETYCDPASLFHVSNDYSFIRYYTRTLYQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFNML
RLGKSEPWTLALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWSTDWSPYADQSIKVRISLKSAL
GDKAYEWNDNEMYLFRSSVAYAMRQYFLKVKNQMILFGEEDVRVANLKPRISFNFFVTAPKNVSDIIPRTE
VEKAIRMSRSRINDAFRLNDNSLEFLGIQPTLGPPNQPPVSIWLIVFGVVMGVIVVGIVILIFTGIRDRKKKNK
ARSGENPYASIDISKGENNPGFQNTDDVQTSF

图38