(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117677702 A
(43)申请公布日 2024.03.08

(21)申请号 202280048699.1

(22)申请日 2022.05.11

(30)优先权数据
63/186871 2021.05.11 US

(85)PCT国际申请进入国家阶段日
2024.01.09

(86)PCT国际申请的申请数据
PCT/US2022/028775 2022.05.11

(87)PCT国际申请的公布数据
WO2022/240991 EN 2022.11.17

(71)申请人 卡美纳生物科技有限公司
地址 英国萨弗伦沃尔登

(72)发明人 D·L·斯坦普 S·哈维 N·贝尔 J·艾伦 E·斯塔尔 N·法塔拉

(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 段菊兰 彭昶

(51)Int.Cl.
C12N 15/66(2006.01)
C12N 15/10(2006.01)

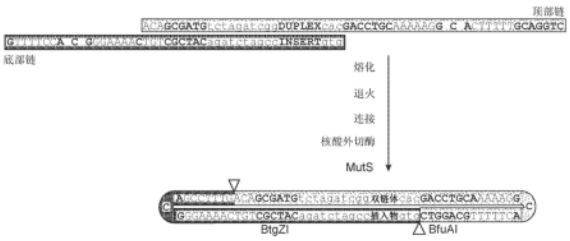
权利要求书3页 说明书29页
序列表25页 附图29页

(54)发明名称

根据寡核苷酸池的基因组装

(57)摘要

本文公开了利用包含基于阵列的寡核苷酸合成的高通量寡核苷酸合成方法来产生基于gSynth和Addamer的DNA合成方法所需的材料的组合物和方法。因此,本文公开了将gSynth和Addamer技术与高通量寡核苷酸合成组合以产生常规高保真基因(靶核酸)合成的组合物和方法。



CN 117677702 A

1.一种包括两种或更多种多个核酸分子的组合物,

其中所述多个核酸分子中的每个核酸分子包括两个或更多个种类的核酸分子,

其中不同种类的核酸分子包括不同的核酸序列,

其中在所述两种或更多种多个核酸分子中,存在至少一组对应的多个核酸分子,使得当所述至少一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的不同多个核酸分子的核酸分子杂交在一起形成至少一个种类的杂交复合物。

2.根据权利要求1所述的组合物,其中所述组合物包括约:

a)六种;

b)十种;

c)15种;

d)20种;或

e)50种

多个核酸分子。

3.根据前述权利要求中任一项所述的组合物,其中每种多个核酸分子包括至少约25个不同种类的核酸分子。

4.根据前述权利要求中任一项所述的组合物,其中每组对应的多个核酸分子包括相同数量的多个核酸分子。

5.根据前述权利要求中任一项所述的组合物,其中所述至少一个种类杂交复合物包括来自所述一组对应的多个核酸分子内的所述多个核酸分子中的每个核酸分子的一个核酸种类。

6.根据前述权利要求中任一项所述的组合物,其中所述两种或更多种多个核酸分子存在于单独的体积中。

7.根据前述权利要求中任一项所述的组合物,其中所述至少一组对应的多个核酸分子包括至少约:

a)两种;

b)三种;

c)四种;或

d)五种

多个核酸分子。

8.根据前述权利要求中任一项所述的组合物,其中对应的多个核酸分子的组数等于:

$$X!/Y!\,(X-Y)!$$

其中X等于多个核酸分子的总数,并且Y等于杂交在一起形成单一杂交复合物的核酸种类的数量。

9.根据前述权利要求中任一项所述的组合物,其中:

a)所述至少一组对应的多个核酸分子包括两种多个核酸分子,并且所述至少一种杂交复合物包括两个核酸分子;

b)所述至少一组对应的多个核酸分子包括三种多个核酸分子,并且所述至少一种杂交复合物包括三个核酸分子;或者

c)所述至少一组对应的多个核酸分子包括四种多个核酸分子,并且所述至少一种杂交复合物包括四个核酸分子。

10.根据前述权利要求中任一项所述的组合物,其中当一组对应的多个核酸分子在单个反应中组合时,形成至少约五个不同的杂交复合物种类。

11.根据前述权利要求中任一项所述的组合物,其中单一多个核酸分子中的不同种类的核酸分子彼此不互补。

12.一种产生至少一种Addamer的方法,所述方法包括:

a)提供根据前述权利要求中任一项所述的组合物;

b)将至少一组对应的多个核酸分子在单个反应体积中组合,使得形成至少一个种类的杂交复合物;

c)使所述至少一个种类的杂交复合物与至少一种连接酶接触,以形成在两个末端处被发夹封端的所述至少一种Addamer。

13.根据权利要求12所述的方法,其进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

14.根据权利要求12或权利要求13所述的方法,其进一步包括使所述至少一个种类的杂交复合物与MutS酶接触。

15.根据权利要求12至14中任一项所述的方法,其中所述Addamer包括:

a)第一II型S限制性核酸内切酶(IISRE)序列;

b)有效载荷序列;

c)至少第二IISRE序列;并且

其中所述Addamer的至少一个末端包括发夹结构。

16.根据权利要求12至15中任一项所述的方法,其中所述Addamer包括所述Addamer的两个末端处的发夹结构。

17.根据权利要求12至16中任一项所述的方法,其中所述Addamer包括:

a)第一IISRE序列;

b)第二IISRE序列;

c)有效载荷序列;以及

d)至少第三IISRE序列。

18.根据权利要求12至17中任一项所述的方法,其中所述Addamer包括:

a)第一IISRE序列;

b)第二IISRE序列;

c)有效载荷序列;

d)第三IISRE序列;以及

e)至少第四IISRE序列。

19.根据权利要求12至18中任一项所述的方法,其中所述Addamer进一步包括多克隆位点(MCS)序列,其中所述MCS序列包括一个或多个限制性核酸内切酶序列。

20.根据权利要求12至19中任一项所述的方法,其中所述IISRE序列中的至少一个序列选自MlyI序列、NgoAVII序列、SspD5I序列、AlwI序列、BccI序列、BcefI序列、PleI序列、BceAI序列、BceSIV序列、BscAI序列、BspD6I序列、FauI序列、EarI序列、BspQI序列、BfuAI序

列、PaqCI序列、Esp3I序列、BbsI序列、BbvI序列、BtgZI序列、FokI序列、BsmFI序列、BsaI序列、BcoDI序列和HgaI序列。

21.根据权利要求12至20中任一项所述的方法,其中至少一个发夹结构包括适体序列。

22.根据权利要求12至21中任一项所述的方法,其中所述适体序列选自pL1适体序列、凝血酶29聚体适体序列、S2.2适体序列、ART1172适体序列、R12.45适体序列、Rb008适体序列和38NT SELEX适体序列。

23.一种合成包括靶核酸序列的核酸分子的方法,所述方法包括:

a)提供根据前述权利要求中任一项所述的组合物,其中所述组合物包括一组对应的多个核酸分子,使得当所述一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的不同多个核酸分子的核酸分子杂交在一起形成两个或更多个种类的杂交复合物,其中所述两个或更多个种类的杂交复合物包括所述靶核酸序列的片段;

b)将所述一组对应的多个核酸分子在单个反应体积中组合,使得形成所述两个或更多个种类的杂交复合物;

c)使所述两个或更多个种类的杂交复合物与至少一种连接酶接触,以形成在两个末端处被发夹封端的两个或更多个种类的Addamer;

d)组装所述两个或更多个种类的Addamer以合成包括所述靶核酸序列的所述核酸分子。

24.根据权利要求23所述的方法,其中组装所述两个或更多个种类的Addamer包括同时或顺序地用以下各项处理所述Addamer:

a)一种或多种限制性酶;以及

b)一种或多种连接酶,

由此组装包括所述靶核酸序列的所述核酸分子。

25.根据权利要求23所述的方法,其中组装所述两个或更多个种类的Addamer包括同时或顺序地用以下各项处理所述Addmer:

a)与至少一种向导RNA结合展示切口酶活性的经修饰的Cas9;以及

b)一种或多种连接酶,

由此组装包括所述靶核酸序列的所述核酸分子。

根据寡核苷酸池的基因组装

[0001] 相关申请交叉引用

[0002] 本申请要求于2021年5月11日提交的美国临时申请第63/186,871号的优先权和权益,所述美国临时申请的内容出于所有目的通过引用整体并入本文。

[0003] 序列表

[0004] 本申请含有已经以ASCII格式经由EFS-Web递交并特此通过引用整体并入的序列表。所述ASCII副本创建于2022年5月11日,命名为“DNWR-010_001WO_SeqList.txt”,并且大小为约23,462字节。

背景技术

[0005] 本领域需要高效产生高保真度和高通量的任意双链DNA序列的基因组装。基因组装的最新进展包含被称为gSynth方法的双链DNA组装方法,其在PCT申请第PCT/US2020/051838号(公开为WO2021055962A1)中详细描述。gSynth方法产生高保真度DNA序列组装体,然而,需要增加gSynth方法的通量,因为使用高通量DNA合成技术,如基于阵列的寡核苷酸合成(参见Lipshutz,R.J.等人,高密度合成寡核苷酸阵列(High density synthetic oligonucleotide arrays.)《自然遗传学(Nature Genetics)》第21卷,第20-24页,1999)来生产gSynth方法所需的材料,需要中间扩增(Saiki R.K.等人,β-珠蛋白基因组序列的酶促扩增和限制性位点分析用于诊断镰状细胞贫血(Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.)《科学(Science)》1985年12月20日:第230卷,第4732期,第1350-1354页)和加工双链体元件。

[0006] 除了gSynth方法之外,最近开发了环境友好的(例如,“绿色”)双链DNA组装/合成方法,其使用被称为“Addamer”的双链DNA构建体作为基础构建。Addamer是双链双发夹结构,携带DNA有效载荷以及各种允许操纵序列的控制元件。Addamer元件可以包含但不限于限制性核酸内切酶(RE)的结合位点、II型S限制性核酸内切酶(IISRE)的结合位点、有效载荷DNA序列和范围从简单GNA基序(Yoshizawa S等人,GNA三核苷酸环序列产生非常稳定的DNA小发夹(GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins.)《生物化学(Biochemistry)》1997,36,16,4761-4767)到具有高亲和力配体结合的复杂三维适体的发夹结构(Ellington.A.D.和Szostak,J.W.与特定配体结合的RNA分子的体外选择(In vitro selection of RNA molecules that bind specific ligands.)《自然(Nature)》346,第818-822页,1990;Tuerk C等人,通过指数富集的配体的系统进化:RNA配体至噬菌体T4 DNA聚合酶(Systematic evolution of ligands by exponential enrichment:RNA ligands to bacteriophage T4 DNA polymerase.)《科学》1990年8月03日:第249卷,第4968期,第505-510页)。

[0007] 本文公开了利用包含基于阵列的寡核苷酸合成的高通量寡核苷酸合成方法来产生文和本文所描述的基于gSynth和Addamer的DNA合成方法所需的材料的组合物和方法。因此,本文公开了将gSynth和Addamer技术与高通量寡核苷酸合成组合以产生常规高保真基

因(靶核酸)合成的组合物和方法。

发明内容

[0008] 本公开提供了包括两种或更多种多个核酸分子的组合物,其中所述多个核酸分子中的每个核酸分子包括两个或更多个种类的核酸分子,其中不同种类的核酸分子包括不同的核酸序列,其中在所述两种或更多种多个核酸分子中,存在至少一组对应的多个核酸分子,使得当所述至少一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的不同多个核酸分子的核酸分子杂交在一起,形成至少一个种类的杂交复合物。

[0009] 在一些方面,本公开的组合物包括约:a)六种;b)十种;c)15种;d)20种;或e)50种多个核酸分子。

[0010] 在本公开的组合物的一些方面,每种多个核酸分子包括至少约25个不同种类的核酸分子。

[0011] 在本公开的组合物的一些方面,每组对应的多个核酸分子包括相同数量的多个核酸分子。

[0012] 在本公开的组合物的一些方面,所述至少一个种类杂交复合物包括来自所述一组对应的多个核酸分子内的所述多个核酸分子中的每个核酸分子的一个核酸种类。

[0013] 在本公开的组合物的一些方面,所述两种或更多种多个核酸分子存在于单独的体积中。

[0014] 在本公开的组合物的一些方面,所述至少一组对应的多个核酸分子包括至少约:a)两种;b)三种;c)四种;或d)五种多个核酸分子。

[0015] 在本公开的组合物的一些方面,对应的多个核酸分子的组数等于:X!/Y!(X-Y)!,其中X等于多个核酸分子的总数,并且Y等于杂交在一起形成单一杂交复合物的核酸种类的数量。

[0016] 在本公开的组合物的一些方面,a)所述至少一组对应的多个核酸分子包括两种多个核酸分子,并且所述至少一种杂交复合物包括两个核酸分子;b)所述至少一组对应的多个核酸分子包括三种多个核酸分子,并且所述至少一种杂交复合物包括三个核酸分子;或者c)所述至少一组对应的多个核酸分子包括四种多个核酸分子,并且所述至少一种杂交复合物包括四个核酸分子。

[0017] 在本公开的组合物的一些方面,当一组对应的多个核酸分子在单个反应中组合时,形成至少约五个不同的杂交复合物种类。

[0018] 在本公开的组合物的一些方面,单一多个核酸分子中的不同种类的核酸分子彼此不互补。

[0019] 本公开提供了产生至少一种Addamer的方法,所述方法包括:a)提供本公开的组合物;b)将至少一组对应的多个核酸分子在单个反应体积中组合,使得形成至少一个种类的杂交复合物;c)使所述至少一个种类的杂交复合物与至少一种连接酶接触,以形成在两个末端处被发夹封端的所述至少一种Addamer。

[0020] 在一些方面,本公开的方法进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

[0021] 在一些方面,本公开的方法进一步包括使所述至少一个种类的杂交复合物与MutS

酶接触。

[0022] 在一些方面,所述Addamer包括:a)第一II型S限制性核酸内切酶(IISRE)序列;b)有效载荷序列;c)至少第二IISRE序列;并且所述Addamer的至少一个末端包括发夹结构。

[0023] 在一些方面,Addamer包括所述Addamer的两个末端处的发夹结构。

[0024] 在一些方面,Addamer包括:a)第一IISRE序列;b)第二IISRE序列;c)有效载荷序列;以及d)至少第三IISRE序列。

[0025] 在一些方面,Addamer包括:a)第一IISRE序列;b)第二IISRE序列;c)有效载荷序列;d)第三IISRE序列;以及e)至少第四IISRE序列。

[0026] 在一些方面,Addamer进一步包括多克隆位点(MCS)序列,其中所述MCS序列包括一个或多个限制性核酸内切酶序列。

[0027] 在一些方面,所述IISRE序列中的至少一个序列选自MlyI序列、NgoAVII序列、SspD5I序列、AlwI序列、BccI序列、BcefI序列、PleI序列、BceAI序列、BceSIV序列、BscAI序列、BspD6I序列、FauI序列、EarI序列、BspQI序列、BfuAI序列、PaqCI序列、Esp3I序列、BbsI序列、BbvI序列、BtgZI序列、FokI序列、BsmFI序列、BsaI序列、BcoDI序列和HgaI序列。

[0028] 在一些方面,至少一个发夹结构包括适体序列。在一些方面,所述适体序列选自pL1适体序列、凝血酶29聚体适体序列、S2.2适体序列、ART1172适体序列、R12.45适体序列、Rb008适体序列和38NT SELEX适体序列。

[0029] 本公开提供了合成包括靶核酸序列的核酸分子的方法,所述方法包括:a)提供根据前述权利要求中任一项所述的组合物,其中所述组合物包括一组对应的多个核酸分子,使得当所述一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的不同多个核酸分子的核酸分子杂交在一起形成两个或更多个种类的杂交复合物,其中所述两个或更多个种类的杂交复合物包括所述靶核酸序列的片段;b)将所述一组对应的多个核酸分子在单个反应体积中组合,使得形成所述两个或更多个种类的杂交复合物;c)使所述两个或更多个种类的杂交复合物与至少一种连接酶接触,以形成在两个末端处被发夹封端的两个或更多个种类的Addamer;d)组装所述两个或更多个种类的Addamer以合成包括所述靶核酸序列的所述核酸分子。

[0030] 在本公开的方法的一些方面,组装所述两个或更多个种类的Addamer包括同时或顺序地用以下各项处理所述Addamer:a)一种或多种限制性酶;以及b)一种或多种连接酶,由此组装包括所述靶核酸序列的所述核酸分子。

[0031] 在本公开的方法的一些方面,组装所述两个或更多个种类的Addamer包括同时或顺序地用以下各项处理所述Addmer:a)与至少一种向导RNA结合展示切口酶活性的经修饰的Cas9;以及b)一种或多种连接酶,由此组装包括所述靶核酸序列的所述核酸分子。

[0032] 上文和本文所描述的任何方面和/或实施例可以与上文和本文所描述的任何其它方面和/或实施例组合。

[0033] 除非另有定义,否则本文使用的所有技术和科学术语具有与本公开所属领域的普通技术人员通常理解的含义相同的含义。在本说明书中,除非上下文另外明确规定,否则单数形式也包含复数形式;作为实例,术语“一个/一种(a/an)”和“所述”被理解为单数或复数,并且术语“或”被理解为是包含的。举例来说,“要素”意指一个或多个要素。贯穿本说明书,词语“包括(comprising)”或如“包括(comprises)”或“包括(comprising)”等变体应当

被理解为暗示包含所陈述要素、整数或步骤或要素组、整数组或步骤组,但不排除任何其它要素、整数或步骤或要素组、整数组或步骤组。约可以被理解为在所陈述值的10％、9％、8％、7％、6％、5％、4％、3％、2％、1％、0.5％、0.1％、0.05％或0.01％范围内。在整个说明书中,描述为“介于x与y之间”或“介于x至y之间”的值的范围的叙述,其中x和y是两个值,被理解为包含x和y。除非上下文另有明确说明,否则本文提供的所有数值均由术语“约”修饰。

[0034] 尽管类似或等效于本文所描述的方法和材料的方法和材料可以用于实践或测试本公开,但是下面描述了合适的方法和材料。本文提及的所有出版物、专利申请、专利和其它参考出于所有目的通过引用整体并入。不承认本文所引用的参考文献是所要求保护的发明的现有技术。在有矛盾的情况下,将以本说明书(包含定义)为准。另外,材料、方法和实例仅是说明性的并且不旨在是限制性的。根据以下详细描述和权利要求书,本公开的其它特征和优点将显而易见。

附图说明

[0035] 当结合附图时,将从以下详细描述更清楚地理解上述特征和另外的特征。

[0036] 图1示出了描绘了由化学合成的寡核苷酸对产生Addamer的示意图。图1描绘了由两个单链DNA分子构建DNA Addamer。顶部链和底部链、双链体插入物和两个切口位点如图所示。图1中呈现的十个或更多个核苷酸的核苷酸序列在SEQ ID NO:1-7中给出。

[0037] 图2描绘了一种双链系统,其具有成对池组合以产生用于基因组装的Addamer集合。左侧的图描绘了用于Addamer产生的顶部链和底部链的池位置。6个池的每个组合提供了互补的顶部链和底部链的独特组合。从6个池中有15个可能的成对组合,这对应于基因A-O。右侧的图示出了池的数量和可能的基因组装体的数量、每个池的Addamer基因组的数量与每个池的不同寡核苷酸的总数之间的关系。

[0038] 图3描绘了由多个寡核苷酸阵列池产生Addamer:用于Addamer产生的双链和三链设计。对于300个碱基的链,重叠互补区是长的,对于双链系统是200个碱基对,并且对于三链系统,在左侧链与中间链之间是200个碱基对,并且对于中间链和右侧链是100个碱基对。对于双链系统,成功杂交产生两个切口,并且三链系统产生三个切口。这些切口被T4 DNA连接酶快速高效地分解。

[0039] 图4描绘了三链系统,其具有池的三向组合以产生用于基因组装的Addamer集合。左侧的图描绘了用于Addamer产生的左侧链、中间链和右侧链的池位置。6个池的每个组合提供了互补的左侧链、中间链和右侧链的独特组合。从6个池中有20个可能的成对组合,这对应于基因A-T。右侧的图示出了池的数量和可能的基因组装体的数量、每个池的Addamer基因组的数量与每个池的不同寡核苷酸的总数之间的关系。

[0040] 图5描绘了三链系统乘以六个池阵列布局。图5在左侧示出了超过六个寡核苷酸阵列池的三链系统的组合图,并且在右侧示出了示例阵列布局。三个池的每个组合产生Addamer的单个基因组。在此,基因S(浅紫色)由每个池1、池4和池6上的五个不同寡核苷酸(A、B、C、D和E)编码的五个Addamer构建。

[0041] 图6描绘了示出了使用池化方法来组合Addamer以产生高保真基因序列的示意图。内部IISRE位点用黑色斑点表示。gSynth位点由编号方块表示。

[0042] 图7是Addamer的示例性示意图,其是在两个末端处包括发夹的双链核酸分子。

[0043] 图8示出了本公开的各种Addamer设计的示例性示意图。示出了Addamer设计的若干实例,每种Addamer类型都具有带有侧接IISRE位点的有效载荷、连接到固体载体的方式和促进核酸外切酶抗性的双发夹。键中示出了一些可能的IISRE和RE位点以及凝血酶适体发夹。

[0044] 图9是本公开的包括使用本公开的Addamer的核酸合成方法的示例性示意图。最初,使用DNA连接将第1个和第2个Addamer连接到带有连接柱的MCS,所述连接柱已经装载到固体载体上。供体构建体和受体构建体在单独体积中产生。用不同的IISRE处理供体构建体和受体构建体以产生可连接的末端。在此图中,受体是通过用R1 IISRE消化产生的,释放端和酶通过漂洗被丢弃。通过用紫色L2酶消化产生供体构建体。将供体构建体溶液(携带L2酶)转移到受体孔中,并且使用T4 DNA连接酶进行连接,所述连接酶对于2、3或4碱基的粘性末端连接具有>80%的高效率。将孔用核酸外切酶处理并冲洗。然后所得的连接的Addamer构建体为后续延伸循环做好准备。

[0045] 图10A、图10B、图10C、图10D、图10E、图10F、图10G和图10H是本公开的核酸合成方法的示例性示意图,所述核酸合成方法包括使用本公开的Addamer来合成27个核苷酸长的靶核酸分子。图10A中呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:8中提出的核苷酸序列。图10E中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:9-10中提出的那些核苷酸序列。图10G中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:11-22中提出的那些核苷酸序列。图10H中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:23-33中提出的那些核苷酸序列。

[0046] 图11A-11F是本公开的双链几何合成(gSynth)的示意图。

[0047] 图11A是使用本公开的双链gSynth方法合成的序列。粗体和下划线的序列的部分对应于已被选择的4聚体突出端,因此定义了将用于合成整个序列的片段。图11A中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:34。

[0048] 图11B示出了图11A所示的序列的单独双链核酸片段,其将用于本公开的双链gSynth方法中以构建图11A所示的序列。基于图11A中选择的位点选择这些片段。图11B中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:35-62。

[0049] 图11C是二叉树的示意图,其示出了组装图11B中的片段以产生图11A所示的序列的顺序。

[0050] 图11D是用于合成图11A所示的序列的双链gSynth方法中第一轮连接的示意图。在第一轮连接中,片段1和2、片段3和4、片段5和6、片段7和8、片段9和10、片段11和12以及片段13和14通过它们互补的5'突出端杂交,并且然后连接在一起以产生片段1+2、片段3+4、片段5+6、片段7+8、片段9+10、片段11+12和片段13+14。图11D中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:35-62。

[0051] 图11E是用于合成图11A所示的序列的双链gSynth方法中第二轮连接的示意图。在第二轮连接中,片段1+2和3+4、片段5+6和7+8以及片段11+12和13+14通过它们互补的5'突出端杂交,并且然后连接在一起以产生片段1+2+3+4、片段5+6+7+8和片段11+12+13+14。图11E中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:63-76。

[0052] 图11F是用于合成图11A所示的序列的双链gSynth方法中第三轮连接的示意图。在第三轮连接中,片段1+2+3+4和5+6+7+8以及片段9+10和11+12+13+14通过它们互补的5'突

出端杂交,并且然后连接在一起以产生片段1+2+3+4+5+6+7+8和片段9+10+11+12+13+14。图11F中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:77-84。

[0053] 图11G是用于合成图11A所示的序列的双链gSynth方法中第四轮连接和最后一轮连接的示意图。在第四轮连接中,片段1+2+3+4+5+6+7+8和9+10+11+12+13+14通过它们互补的5'突出端杂交并连接在一起,由此产生图1A所示的序列。图11G中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:85-88。

[0054] 图12示出了对要使用本公开的方法合成的靶核酸序列的示例性处理和分析。431bp的全长序列被分成五个不同大小的片段(F1-F5)。这些片段是用于产生最终序列的Addamer的有效载荷。使用计算机程序可靠地预测在连接反应中相容的间隔良好的4个碱基的突出端位点,以及其它特征,如最佳GC含量。对整个序列进行了分析,以确定是否存在IISRE位点(BsmFI、FokI、BtgZI、SfaNI)。出于组装目的,靶序列中存在的位点不包括某些IISRE。

[0055] 图13示出了对应于图12中鉴定的片段的Addamer。Addamer的有效载荷是预测的片段(蓝色箭头表示5'->3'朝向),其侧接有IISRE位点(例如,内部片段的BsaI)。4个碱基的突出端在浅橙色框(上链)和浅绿色框(下链)中示出。对于末端片段F1和F5,外部IISRE位点是不同的(BbsI),以允许序列克隆后的后续组装。片段F1和F5还具有用于扩增的正向引物位点和反向引物位点(延伸的T7和延伸的T3)以及用于克隆的RE位点(XbaI和EcoRI)。图11G中所呈现的十个或更多个核苷酸的核苷酸序列对应于SEQ ID NO:89-108。

[0056] 图14示出了通过测序对使用图13呈现的Addamer组装的靶核酸序列的验证。挑选克隆并使用毛细管测序对质粒进行测序。反向序列和正向序列完全对准,从而验证了来自寡核苷酸池的基于Addamer的组装的准确性。

[0057] 图15示出了使用本公开的方法组装的另外的靶核酸序列的组装程序的设计。

具体实施方式

[0058] 本公开涉及利用汇集的核酸分子产生包含但不限于Addamer的双链或部分双链核酸分子的组合物和方法,用于双链DNA组装/合成方法,包含但不限于基于gSynth的方法和基于Addamer的方法。也就是说,本文所描述的组合物和方法通过利用汇集的核酸分子的非常高的寡核苷酸产生可能性,改进了现有的基于gSynth和基于Addamer的DNA组装/合成方法,而不需要双链体的介入扩增,后者需要大量纯化的核苷酸和易于出错的DNA聚合酶。在一些方面,这些汇集的核酸分子可以使用基于阵列的寡核苷酸合成方法来产生。

[0059] 一方面,可以产生许多寡核苷酸池,然后组合以允许从在这些组合池中产生的Addamer独特地构建单独基因。例如,使用gSynth算法,任何基因都可以被解构为相对少量的片段,其中每个连接处都有独特的高保真悬垂序列,允许直接的粘性末端连接产生最终产物,类似于Golden Gate组装方法(Pryor J.M.等人,使用数据优化的组装设计实现空前复杂的一锅Golden Gate组装体(Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design.)《公共科学图书馆综合(PLoS ONE)》15(9):e0238592,2020)。值得注意的是,Golden Gate组装最适合克隆序列,具有均匀消化,选择高保真重叠。如果双链体序列是500个碱基对(假设两个300个碱基的寡聚物结合产生100个碱基的Addamer作为控制元件,剩下500个碱基或250个碱基对的

双链体序列),那么五个连续序列的连接将产生一个2.5kb基因。在一些方面,本公开的方法的使用者可以广泛地改变任何单独基因片段的长度,以最佳地适应组装、IISRE位点可用性和二级结构参数。

[0060] 在一些方面,Addamer(其结构在本文中进一步详细的描述)可以通过一对互补DNA链的独特组合来产生,所述链彼此杂交,从而形成具有一对未解析切口的双链双发夹结构。这些切口可以通过应用DNA连接酶,如T4DNA连接酶来修复。一旦切口被解析,可能的错配可以通过His标记的MutS蛋白结合亲和试剂来去除(Wang J等人,用组氨酸标记的嗜热栖热菌MutS直接找出基因组DNA中的细微突变(Directly fishing out subtle mutations in genomic DNA with histidine-tagged Thermus thermophilus MutS.)第547卷,第1-2期,2004年3月22日,第41-47页)。在一些方面,通过应用T7核酸外切酶去除所有非Addamer DNA。

[0061] DNA Addamer可以由两个单链DNA分子构建。双链体插入物可以很大,大约250个碱基对(这允许从估计的300个碱基的单链寡核苷酸大小中得到100个碱基的控制序列)。由于总序列量可能很高,就选择性和双链体回收率两者而言,杂交的保真度也将很高。退火可以产生具有稳定发夹的几乎完全双链序列。这两个“切口”可以通过用T4 DNA连接酶进行的短时间处理而高效且快速地修复。限制性连接可以与高保真杂交组合,这意味着唯一的T7核酸外切酶抗性物质可以是期望的Addamer产物。另外,可以用His标记的Taq MutS蛋白进行处理,以结合可能的错配碱基对,随后与Ni-NTA(镍次氮基乙酸)琼脂糖亲和树脂接触,以除去含有MutS结合错配的DNA,留下基本上富集的纯的期望产物(参见图1)。

[0062] 也就是说,在一些方面,本公开的Addamer可以通过杂交第一单链核酸分子和第二单链核酸分子来产生,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第二区互补的第一区和自互补的第二区,并且所述第二单链核酸分子包括自互补的第一区和与所述第一单链核酸分子上的第一区互补的第二区,如图3的上图所示。第一单链核酸分子和第二单链核酸分子杂交在一起产生杂交复合物。然后,杂交复合物可以任选地与MutS酶接触,所述酶与错配的碱基结合并且将DNA暴露于核酸外切酶消化。然后,杂交复合物可以与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。在与连接酶接触后,产物可以与T7核酸外切酶接触以纯化和富集适当形成的Addamer。此方法在本文中被称为“双链Addamer组装方法”。

[0063] 因此,本公开提供了产生本文所描述的Addamer的方法,所述方法包括:a)提供第一单链核酸分子和第二单链核酸分子,其中所述第一单链核酸分子和所述第二单链核酸分子的序列包括要产生的Addamer的部分,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第二区互补的第一区和自互补的第二区,并且所述第二单链核酸分子包括自互补的第一区和与所述第一单链核酸分子上的第一区互补的第二区;b)将所述第一单链核酸分子和所述第二单链核酸分子杂交;以及c)使部分双链核酸分子与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。

[0064] 在一些方面,前述方法可以进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

[0065] 在一些方面,前述方法可以进一步包括在步骤(b)之后和步骤(c)之前,使所述部分双链核酸分子与MutS酶接触。

[0066] 在一些方面,Addamer可以通过将第一单链核酸分子、第二单链核酸分子和第三单链核酸分子杂交来产生,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第一区互补的第一区和自互补的第二区,所述第二单链核酸分子包括与所述第一单链核酸分子的所述第一区互补的第一区和与所述第三单链核酸分子的第一区互补的第二区,并且所述第三单链核酸分子包括与所述第二单链核酸分子的所述第二区互补的第一区和自互补的第二区,如图3的中间图所示。在图3中,所述第一单链核酸分子被称为左链,所述第二单链核酸分子被称为中间链,并且所述第三单链核酸分子被称为右链。所述第一单链核酸分子、所述第二单链核酸分子和所述第三单链核酸分子杂交在一起以产生杂交复合物。然后,杂交复合物可以任选地与MutS酶接触,所述酶与错配的碱基结合并且将DNA暴露于核酸外切酶消化。然后,杂交复合物可以与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。在与连接酶接触后,产物可以与T7核酸外切酶接触以纯化和富集适当形成的Addamer。在三链系统中,可以调整中间链相对于左链和右链的位置,以平衡链间的杂交程度。然而,在形成发夹的自杂交的量与形成完整Addamer结构的交叉杂交的量之间存在平衡。此方法在本文中被称为“三链Addamer组装方法”。

[0067] 因此,本公开提供了产生本文所描述的Addamer的方法,所述方法包括:a)提供第一单链核酸分子、第二单链核酸分子和第三单链核酸分子,其中所述第一单链核酸分子、所述第二单链核酸分子和所述第三单链核酸分子的序列包括要产生的Addamer的部分,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第一区互补的第一区和自互补的第二区,所述第二单链核酸分子包括与所述第一单链核酸分子的所述第一区互补的第一区和与所述第三单链核酸分子的第一区互补的第二区,并且所述第三单链核酸分子包括与所述第二单链核酸分子的所述第二区互补的第一区和自互补的第二区;b)将所述第一单链核酸分子、所述第二单链核酸分子和所述第三单链核酸分子杂交;以及c)使部分双链核酸分子与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。

[0068] 在一些方面,前述方法可以进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

[0069] 在一些方面,前述方法可以进一步包括在步骤(b)之后和步骤(c)之前,使所述部分双链核酸分子与MutS酶接触。

[0070] 在一些方面,Addamer可以通过将第一单链核酸分子、第二单链核酸分子、第三单链核酸分子和第四单链核酸分子杂交来产生,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第一区互补的第一区和自互补的第二区,所述第二单链核酸分子包括与所述第一单链核酸分子的所述第一区互补的第一区和与所述第三单链核酸分子的第一区互补的第二区,所述第三单链核酸分子包括与所述第二单链核酸分子的所述第二区互补的第一区和与所述第四单链核酸分子上的第一区互补的第二区,并且所述第四单链核酸分子包括与所述第三单链核酸分子的所述第二区互补的第一区和自互补的第二区域,如图3的下图所示。所述第一单链核酸分子、所述第二单链核酸分子、所述第三单链核酸分子和所述第四单链核酸分子杂交在一起以产生杂交复合物。然后,杂交复合物可以任选地与MutS酶接触,所述酶与错配的碱基结合并且将DNA暴露于核酸外切酶消化。然后,杂交复合物可以与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。在与连接酶接触后,产物可以与T7核酸外切酶接触以纯化和富集适当形成的Addamer。在三链系统中,可以调整

中间链相对于左链和右链的位置,以平衡链间的杂交程度。然而,在形成发夹的自杂交的量与形成完整Addamer结构的交叉杂交的量之间存在平衡。此方法在本文中被称为“四链Addamer组装方法”。

[0071] 因此,本公开提供了产生本文所描述的Addamer的方法,所述方法包括:a)提供第一单链核酸分子、第二单链核酸分子、第三单链核酸分子和第四单链核酸分子,其中所述第一单链核酸分子、所述第二单链核酸分子、所述第三单链核酸分子和所述第四单链核酸分子的序列包括要产生的Addamer的部分,其中所述第一单链核酸分子包括与所述第二单链核酸分子上的第一区互补的第一区和自互补的第二区,所述第二单链核酸分子包括与所述第一单链核酸分子的所述第一区互补的第一区和与所述第三单链核酸分子的第一区互补的第二区,所述第三单链核酸分子包括与所述第二单链核酸分子的所述第二区互补的第一区和与所述第四单链核酸分子上的第一区互补的第二区,并且所述第四单链核酸分子包括与所述第三单链核酸分子的所述第二区互补的第一区和自互补的第二区;b)将所述第一单链核酸分子、所述第二单链核酸分子和所述第三单链核酸分子杂交;以及c)使部分双链核酸分子与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。

[0072] 在一些方面,前述方法可以进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

[0073] 在一些方面,前述方法可以进一步包括在步骤(b)之后和步骤(c)之前,使所述部分双链核酸分子与MutS酶接触。

[0074] 在双链Addamer组装方法、三链Addamer组装方法和/或四链Addamer组装方法的上下文中,用于每种方法的单独单链核酸分子可以在单独的多个核酸分子(本文中也被称为“池”)中单独提供,其中所述单独的多个核酸分子是使用如基于阵列的寡核苷酸合成等方法产生的。如技术人员所理解的,基于阵列的寡核苷酸合成方法包含但不限于电化学方法、基于光的化学方法、喷墨印刷方法或其任何组合。

[0075] 也就是说,基因片段组可以分布在若干寡核苷酸池中,使得独特的池对可以组合产生表示要合成的最终基因(靶核酸)构建体的每个片段的Addamer。作为实例,对于六个寡核苷酸阵列池,其中每个基因由五个片段构建,总共有15个可能的基因构建物,并且每个池含有五个不同基因的顶部或底部五个片段。因此,各自具有25个寡核苷酸的6个池,其中寡核苷酸为约300个碱基长,将足以构建15个2.5kb长的基因。

[0076] 通过成对组合,很明显当前阵列池的容量未得到充分利用。即使有50个库,其中可以组装1225个基因,每个池中不同寡核苷酸的总数也仅有245个(参加图2)。虽然,随着每个池中不同寡核苷酸数量的减少,每种特定寡核苷酸的量增加,但有效组装给定基因所需的每种寡核苷酸的总量是有限的。通过增加用于产生Addamer的链的数量,每个池的基因和独特寡核苷酸的总数显著增加,特别是对于高数量的池。

[0077] 另一方面,可以采用三链Addamer组装方法(参见图3),而不是两链。在三链系统的情况下,池的可能组合的数量显著增加。对于双链设计和三链设计来说,形成特定的Addamer,而不是随机组合,具有很强的选择优势。首先,使用来自阵列池的长寡聚物,例如每条链300个碱基,重叠互补区非常长,这确保了在正常条件下的稳健杂交。其次,因为所述过程包含T7核酸外切酶处理,以去除非Addamer DNA,切口的连接成为选择性事件。因此,如果杂交不正确,Addamer将不会连接在一起形成双发夹结构,并且将被T7核酸外切酶降解。

[0078] 在另一个优选实施例中,可以使用四链系统。每个基因具有四条链和四个池,池数量为10和以上时,基因数量增加。例如,在15个库的情况下,总共可以产生1365个基因(参见表1)。池的复杂性的增加提供了寡核苷酸阵列的更高效利用。

[0079] 表1

[0080]

池	基因	基因集合/池	寡聚物/池
6	15	10	50
10	210	84	420
15	1365	364	1820
20	4845	969	4845
50	230300	18424	92120

[0081] 一些商业寡核苷酸阵列池可以用任意数量的不同寡核苷酸产生。在此情况下,对于固定的阵列表面,不同寡核苷酸的数量越少,产生的每个寡核苷酸序列的总质量越大。考虑到具有六个池的三链系统,每个池将具有50种不同的寡核苷酸种类,因此在组合的三个池中,将有150种不同的寡核苷酸要区分(参见图5)。组合的池可以用于产生任意数量的Addamer,但为了避免突出端的不适当连接,最好将组装体中的Addamer数量保持在最小值。

[0082] 因此,本公开提供了一种包括两种或更多种多个核酸分子的组合物,其中所述多个核酸分子中的每个核酸分子包括两个或更多个种类的核酸分子,其中不同种类的核酸分子包括不同的核酸序列,其中在所述两种或更多种多个核酸分子中,存在至少一组对应的多个核酸分子,使得当所述一组对应的多个核酸分子在单个反应体积中组合时,来自所述组中的多个核酸分子中至少一个核酸分子的至少一个种类的核酸分子与来自所述组中的其它多个核酸分子中至少一个核酸分子的至少一个种类的核酸分子杂交,以形成至少一个种类的杂交复合物。

[0083] 因此,本公开提供了一种包括两种或更多种多个核酸分子的组合物,其中所述多个核酸分子中的每个核酸分子包括两个或更多个种类的核酸分子,其中不同种类的核酸分子包括不同的核酸序列,其中在所述两种或更多种多个核酸分子中,存在至少一组对应的多个核酸分子,使得当所述一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的不同多个核酸分子的核酸分子杂交在一起,形成至少一个种类的杂交复合物。

[0084] 因此,本公开提供了一种包括两种或更多种多个核酸分子的组合物,其中所述多个核酸分子中的每个核酸分子包括两个或更多个种类的核酸分子,其中不同种类的核酸分子包括不同的核酸序列,其中在所述两种或更多种多个核酸分子中,存在至少一组对应的多个核酸分子,使得当所述一组对应的多个核酸分子在单个反应体积中组合时,来自所述组内的每种不同多个核酸分子的核酸分子杂交在一起,形成至少一个种类的杂交复合物。

[0085] 在前述组合物的一些方面,杂交复合物可以是图3所示的任何杂交复合物。

[0086] 在前述组合物的一些方面,在每种多个核酸分子中,单一种类的核酸分子存在于多个核酸分子中(即,在多个核酸分子中存在所述种类的核酸分子的超过一个拷贝)。

[0087] 在一些方面,前述组合物可以包括至少约一种、或至少约两种、或至少约三种、或至少约四种、或至少约五种、或至少约六种、或至少约七种、或至少约八种、或至少约九种、或至少约10种、或至少约11种、或至少约12种、或至少约13种、或至少约14种、或至少约15种、或至少约16种、或至少约17种、或至少约18种、或至少约19种、或至少约20种、或至少约

25种、或至少约30种、或至少约35种、或至少约40种、或至少约45种、或至少约50种、或至少约55种、或至少约60种、或至少约65种、或至少约70种、或至少约75种、或至少约80种、或至少约85种、或至少约90种、或至少约95种、或至少约100种、或至少约150种、或至少约200种、或至少约250种、或至少约300种、或至少约350种、或至少约400种、或至少约450种、或至少约500种、或至少约550种、或至少约600种、或至少约650种、或至少约700种、或至少约750种、或至少约800种、或至少约850种、或至少约900种、或至少约950种、或至少约1000种、或至少约1500种、或至少约2000种、或至少约2500种、或至少约3000种、或至少约3500种、或至少约4000种、或至少约4500种、或至少约5000种、或至少约5500种、或至少约6000种、或至少约6500种、或至少约7000种、或至少约7500种、或至少8000种、或至少约8500种、或至少约9000种、或至少约9500种或至少约10000种多个核酸分子。

[0088] 在一些方面,前述组合物可以包括可以包括约一种、或约两种、或约三种、或约四种、或约五种、或约六种、或约七种、或约八种、或约九种、或约10种、或约11种、或约12种、或约13种、或约14种、或约15种、或约16种、或约17种、或约18种、或约19种、或约20种、或约25种、或约30种、或约35种、或约40种、或约45种、或约50种、或约55种、或约60种、或约65种、或约70种、或约75种、或约80种、或约85种、或约90种、或约95种、或约100种、或约150种、或约200种、或约250种、或约300种、或约350种、或约400种、或约450种、或约500种、或约550种、或约600种、或约650种、或约700种、或约750种、或约800种、或约850种、或约900种、或约950种、或约1000种、或约1500种、或约2000种、或约2500种、或约3000种、或约3500种、或约4000种、或约4500种、或约5000种、或约5500种、或约6000种、或约6500种、或约7000种、或约7500种、或8000种、或约8500种、或约9000种、或约9500种或约10000种多个核酸分子。

[0089] 在前述组合物的一些方面,每种多个核酸分子可以包括至少约一个、或至少约两个、或至少约三个、或至少约四个、或至少约五个、或至少约六个、或至少约七个、或至少约八个、或至少约九个、或至少约10个、或至少约11个、或至少约12个、或至少约13个、或至少约14个、或至少约15个、或至少约16个、或至少约17个、或至少约18个、或至少约19个、或至少约20个、或至少约25个、或至少约30个、或至少约35个、或至少约40个、或至少约45个、或至少约50个、或至少约55个、或至少约60个、或至少约65个、或至少约70个、或至少约75个、或至少约80个、或至少约85个、或至少约90个、或至少约95个、或至少约100个、或至少约150个、或至少约200个、或至少约250个、或至少约300个、或至少约350个、或至少约400个、或至少约450个、或至少约500个、或至少约550个、或至少约600个、或至少约650个、或至少约700个、或至少约750个、或至少约800个、或至少约850个、或至少约900个、或至少约950个、或至少约1000个、或至少约1500个、或至少约2000个、或至少约2500个、或至少约3000个、或至少约3500个、或至少约4000个、或至少约4500个、或至少约5000个、或至少约5500个、或至少约6000个、或至少约6500个、或至少约7000个、或至少约7500个、或至少8000个、或至少约8500个、或至少约9000个、或至少约9500个、或至少约10000个、或至少约20000个、或至少约30000个、或至少约40000个、或至少约50000个、或至少约60000个、或至少约70000个、或至少约80000个、或至少约90000个、或至少约100000个不同种类的核酸分子。

[0090] 在前述组合物的一些方面,每种多个核酸分子可以包括约一个、或约两个、或约三个、或约四个、或约五个、或约六个、或约七个、或约八个、或约九个、或约10个、或约11个、或约12个、或约13个、或约14个、或约15个、或约16个、或约17个、或约18个、或约19个、或约20

个、或约25个、或约30个、或约35个、或约40个、或约45个、或约50个、或约55个、或约60个、或约65个、或约70个、或约75个、或约80个、或约85个、或约90个、或约95个、或约100个、或约150个、或约200个、或约250个、或约300个、或约350个、或约400个、或约450个、或约500个、或约550个、或约600个、或约650个、或约700个、或约750个、或约800个、或约850个、或约900个、或约950个、或约1000个、或约1500个、或约2000个、或约2500个、或约3000个、或约3500个、或约4000个、或约4500个、或约5000个、或约5500个、或约6000个、或约6500个、或约7000个、或约7500个、或8000个、或约8500个、或约9000个、或约9500个、或约10000个、或约20000个、或约30000个、或约40000个、或约50000个、或约60000个、或约70000个、或约80000个、或约90000个或约100000个不同种类的核酸分子。

[0091] 在前述组合物的一些方面,单一多个核酸分子中的不同种类的核酸分子彼此不互补。

[0092] 在前述组合物的一些方面,每组对应的多个核酸分子包括相同数量的多个核酸分子。在一些方面,所述至少一个种类的杂交复合物包括来自所述一组对应的多个核酸分子内的所述多个核酸分子中的每个核酸分子的一个核酸种类。

[0093] 在前述组合物的一些方面,所述两种或更多种多个核酸分子存在于单独的体积中(即,它们在物理上彼此分开,如在不同的容器中或在阵列的物理上不同的部分上)。

[0094] 在前述组合物的一些方面,所述一组对应的多个核酸分子可以包括至少约两种多个核酸分子、至少约三种多个核酸分子、至少约四种多个核酸分子、至少约五种多个核酸分子、至少约六种多个核酸分子、至少约七种多个核酸分子、至少约八种多个核酸分子、至少约九种多个核酸分子或至少约十种多个核酸分子。

[0095] 在前述组合物的一些方面,所述一组对应的多个核酸分子可以包括约两种多个核酸分子、约三种多个核酸分子、约四种多个核酸分子、约五种多个核酸分子、约六种多个核酸分子、七种多个核酸分子、约八种多个核酸分子、约九种多个核酸分子或约十种多个核酸分子。

[0096] 在前述组合物的一些方面,所述一组对应的多个核酸分子可以包括两种多个核酸分子、三种多个核酸分子、四种多个核酸分子、五种多个核酸分子、六种多个核酸分子、七种多个核酸分子、八种多个核酸分子、九种多个核酸分子或十种多个核酸分子。

[0097] 在前述组合物的一些方面,对应的多个核酸分子的组数可以等于:

[0098] $X!/Y!(X-Y)!$

[0099] 其中X等于多个核酸分子的总数,并且Y等于杂交在一起形成单一杂交复合物的核酸种类的数量。例如,包括10种多个核酸的组合物,其中每种杂交复合物包含三个种类的核酸分子,将具有120组对应的多个核酸分子。

[0100] 在非限制性实例中,如果一组对应的多个核酸分子包括两种多个核酸分子,那么杂交复合物将包括两个核酸分子,一个核酸分子来自两种多个核酸分子中的每一种。在非限制性实例中,如果一组对应的多个核酸分子包括三种多个核酸分子,那么杂交复合物将包括三个核酸分子,一个核酸分子来自三种多个核酸分子中的每一种。在非限制性实例中,如果一组对应的多个核酸分子包括四种多个核酸分子,那么杂交复合物将包括四个核酸分子,一个核酸分子来自四种多个核酸分子中的每一种。

[0101] 在前述组合物的一些方面,当一组对应的多个核酸分子在单个反应中组合时,可

以形成至少约一个、或至少约两个、或至少约三个、或至少约四个、或至少约五个、或至少约六个、或至少约七个、或至少约八个、或至少约九个、或至少约10个、或至少约11个、或至少约12个、或至少约13个、或至少约14个、或至少约15个、或至少约16个、或至少约17个、或至少约18个、或至少约19个、或至少约20个、或至少约25个、或至少约30个、或至少约35个、或至少约40个、或至少约45个、或至少约50个、或至少约55个、或至少约60个、或至少约65个、或至少约70个、或至少约75个、或至少约80个、或至少约85个、或至少约90个、或至少约95个或至少约100个不同的杂交复合物种类。

[0102] 在前述组合物的一些方面,当一组对应的多个核酸分子在单个反应中组合时,可以形成约一个、或约两个、或约三个、或约四个、或约五个、或约六个、或约七个、或约八个、或约九个、或约10个、或约11个、或约12个、或约13个、或约14个、或约15个、或约16个、或约17个、或约18个、或约19个、或约20个、或约25个、或约30个、或约35个、或约40个、或约45个、或约50个、或约55个、或约60个、或约65个、或约70个、或约75个、或约80个、或约85个、或约90个、或约95个或约100个不同的杂交复合物种类。

[0103] 在前述组合物的一些方面,所形成的至少一个种类的杂交复合物对应于本公开的Addamer,使得如果所述至少一个种类的杂交复合物与合适的连接酶和任选地MutS酶接触,则本公开的Addamer将由杂交复合物形成。

[0104] 因此,本公开提供了一种产生本公开的至少一种Addamer的方法,所述方法包括:a)提供本公开的组合物;b)将至少一组对应的多个核酸分子在单个反应体积中组合,使得形成至少一个种类的杂交复合物;c)使所述至少一个种类的杂交复合物与连接酶接触,以形成在两个末端处被发夹封端的双链Addamer结构。在一些方面,所述至少一个种类的杂交复合物包括两个单链核酸分子,如本文所描述的双链Addamer组装方法中提出的。在一些方面,所述至少一个种类的杂交复合物包括三个单链核酸分子,如本文所描述的三链Addamer组装方法中提出的。在一些方面,所述至少一个种类的杂交复合物包括四个单链核酸分子,如本文所描述的四链Addamer组装方法中提出的。

[0105] 在一些方面,前述方法可以进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的Addamer。

[0106] 在一些方面,前述方法可以进一步包括在步骤(b)之后和步骤(c)之前,使所述部分双链核酸分子与MutS酶接触。

[0107] 另外,本公开提供了一种产生本公开的至少一个双链片段的方法,所述方法包括:a)提供本公开的组合物;b)将至少一组对应的多个核酸分子在单个反应体积中组合,使得形成至少一个种类的杂交复合物,其中所述至少一个种类的杂交复合物包括所述至少一个双链片段。在一些方面,如本文所描述,所述至少一个双链片段是用于gSynth合成方法的片段。

[0108] 本公开的基于Addamer的方法和组合物

[0109] Addamer

[0110] 本公开提供了一种包括至少一种Addamer的组合物。如本文所使用的,术语Addamer用于描述在两个末端处包括发夹结构的双链核酸分子。在一些方面,其中Addamer被固定到固体表面,Addamer可以包括位于分子的末端处的不与固体表面连接的单个发夹。固定到固体表面的一个Addamer和两个Addamer的示例性示意图在图7中示出。Addamer可以

包括本文所描述的一个或多个特征。

[0111] 在一些方面,Addamer可以包括DNA、基本上由其组成或由其组成。

[0112] 在一些方面,Addamer可以包括一个或多个多克隆位点(MCS)序列。在一些方面,MCS序列可以包括一个或多个限制性核酸内切酶(RE)序列,其可以用对应的限制性核酸内切酶切割以产生3'突出端、5'突出端或平端。

[0113] 正如技术人员所理解的,"平端"用于描述DNA片段的末端,其中没有不成对的核苷酸。

[0114] 如熟练技术人员所理解的,术语5'突出端用于指部分双链核酸分子的位于一条链的5'末端处的单链部分。

[0115] 如熟练技术人员所理解的,术语3'突出端用于指部分双链核酸分子的位于一条链的3'末端处的单链部分。

[0116] 在一些方面,Addamer可以包括至少一个可以用对应的II型S限制性核酸内切酶(下文被称为"IISRE")切割的偏移切割II型S限制性核酸内切酶(IISRE)序列(下文被称为"IISRE序列")。在一些方面,Addamer可以包括至少一个IISRE序列。在一些方面,Addamer可以包括至少三个IISRE序列。在一些方面,Addamer可以包括至少四个IISRE序列。

[0117] 在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生"平端"。

[0118] 在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为1个核苷酸的5'突出端。在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为2个核苷酸的5'突出端。在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为3个核苷酸的5'突出端。在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为4个核苷酸的5'突出端。在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为5个核苷酸的5'突出端。在一些方面,IISRE序列是这样的序列,使得用对应的IISRE切割导致产生长度为约1个核苷酸至约5个核苷酸的5'突出端。

[0119] IISRE序列的非限制性实例,连同它们对应的IISRE和对通过切割IISRE序列和对应的IISRE产生的突出端/平端的描述在表2中示出。因此,Addamer可以包括表2中列出的一个或多个IISRE序列。

[0120] 表2:示例性IISRE序列和对应IISRE

[0121]

IISRE 名称	IISRE 序列	产生的突出端/平端
MlyI	GAGTC (5/5)	平端
NgoAVII	GCCGC (7/7)	平端
SspD5I	GGTGA (8/8)	平端
AlwI	GGATC (4/5)	1 nt 5'突出端
BccI	CCATC (4/5)	1 nt 5'突出端
BcefI	ACGGC (12/13)	1 nt 5'突出端
PleI	GAGTC (4/5)	1 nt 5'突出端
BceAI	ACGGC (12/14)	2 nt 5'突出端
BceSIV	GCTGC (9/11)	2 nt 5'突出端
BscAI	GCATC (4/6)	2 nt 5'突出端
BspD6I	GAGTC (4/6)	2 nt 5'突出端
FauI	CCCGC (4/6)	2 nt 5'突出端
EarI	CTCTTC (1/4)	3 nt 5'突出端
BspQI	GCTCTTC (1/4)	3 nt 5'突出端
BfuAI	ACCTGC (4/8)	4 nt 5'突出端
PaqCI	CACCTGC (4/8)	4 nt 5'突出端
Esp3I	CGTCTC (1/5)	4 nt 5'突出端
BbsI	GAAGAC (2/6)	4 nt 5'突出端
BbvI	GCAGC (8/12)	4 nt 5'突出端
BtgZI	GCGATG (10/14)	4 nt 5'突出端
FokI	GGATG (9/13)	4 nt 5'突出端
BsmFI	GGGAC (10/14)	4 nt 5'突出端
BsaI	GGTCTC (1/5)	4 nt 5'突出端
BcoDI	GTCTC (1/5)	4 nt 5'突出端

[0122]

HgaI	GACGC (5/10)	5 nt 5'突出端

[0123] 在一些方面,位于Addamer的末端处的发夹结构或发夹(可互换使用)可以包括至少约1个、或至少约两个、或至少约三个、或至少约四个、或至少约五个、或至少约六个、或至少约七个、或至少约八个、或至少约九个、或至少约十个、或至少约11个、或至少约12个、或至少约13个、或至少约14个、或至少约15个、或至少约16个、或至少约17个、或至少约18个、或至少约19个、或至少约20个、或至少约21个、或至少约22个、或至少约23个、或至少约24

个、或至少约25个、或至少约26个、或至少约27个、或至少约28个、或至少约29个、或至少约30个、或至少约31个、或至少约32个、或至少约33个、或至少约34个、或至少约35个、或至少约36个、或至少约37个、或至少约38个、或至少约39个、或至少约40个、或至少约41个、或至少约42个、或至少约43个、或至少约44个、或至少约45个、或至少约46个、或至少约47个、或至少约48个、或至少约49个或至少约50个核苷酸。

[0124] 如上文所描述,Addamer在任一个末端处被发夹结构封端。发夹结构有若干作用。首先,发夹为Addamer提供了防止核酸外切酶消化的保护。这允许在本公开的方法中从给定的反应中清除未反应的中间体,这提供了产生期间的Addamer和Addamer延伸之后的产物两者的纯度。其次,发夹结构提供了用于将Addamer与固体载体连接的手段。这些连接直接通过适体的结合,所述适体与特定的固体载体结合的配体结合;在用常规RE如BamHI消化之后通过MCS连接;通过λ噬菌体cos位点连接;或者通过与单链固体载体结合的锚定NA杂交而产生。最后,本文所描述的发夹允许Addamer与固体载体(例如,珠粒)连接,而不需要非天然修饰,如生物素。因此,本公开的Addamer可以使用完全天然的方式合成,不需要小规模和/或大规模的亚磷酰胺合成。因此,本公开的Addamer和方法可以允许核酸分子的更快和更便宜的合成,并且产生更少的毒性废物。

[0125] 在一些方面,位于Addamer的末端的发夹可以包括允许Addamer的亲和纯化和/或Addamer与固体载体(例如,珠粒)的连接的结构序列。

[0126] 在一些方面,位于Addamer的末端的发夹可以包括允许受控自动切割的酶序列(例如,DNA酶序列)。

[0127] 在一些方面,位于Addamer的末端的发夹可以包括一个或多个限制性酶位点。不希望受理论束缚,发夹中的一个或多个限制性酶位点可以用对应的限制性酶切割,以产生至少一个单链突出端,其随后可以用于将切割的Addamer与包括与至少一个单链突出端互补的核酸的固体载体(例如,珠粒)杂交和/或连接。

[0128] 在一些方面,位于Addamer的末端的发夹可以包括适体序列。不希望受理论束缚,适体序列可用于亲和纯化和/或与固体载体(例如,珠粒)连接。适体序列的非限制性实例在表3中示出。

[0129] 表3:示例性适体

[0130]

适体名称	配体	序列
pL1 中	抗 *Pv*LDH	TCGATTGGATTGTGCCGGAAGTGCTGGCTCGA（SEQ ID NO: 109）
凝血酶 29 聚体	抗凝血酶	AGTCCGTGGTAGGGCAGGTTGGGGTGACT（SEQ ID NO: 110）
S2.2	抗 Muc1	CAGTTGATCCTTTGGATACCCTG（SEQ ID NO: 111）
ART1172	抗 VWF	GGCGTGCAGTGCCTTCGGCCGTGCGGTGCCTCCGTCACGCCT（SEQ ID NO: 112）
R12.45	抗阿特拉津	ACCGTCTGAGCGATTCGTACTTTATTCGGGAGGTATCAGCGGG（SEQ ID NO: 113）
Rb008	抗 ATX	CCTGGACGGAACCAGAATACTTTTGGTCTCCAGG（SEQ ID NO: 114）
38NT SELEX	抗 HIV_RT	AAATACCCCCCCTTCGGTGCAAAGCACCGAAGGGGGGGTATTT（SEQ ID NO: 115）

[0131] 在一些方面,Addamer可以包括λ噬菌体cos位点。

[0132] 在一些方面,Addamer可以包括“N聚体序列”,所述序列包括将使用本文所描述方法的之一合成的核酸的片段。术语“N聚体序列”、“有效载荷”、“有效载荷序列”、“N聚体有效载荷”和“双链体插入物”在本文中可互换使用。

[0133] 在一些方面,N聚体序列的长度可以为约3个核苷酸。在一些方面,N聚体序列的长度为3个核苷酸。长度为3个核苷酸的N聚体序列在本文中被称为3聚体。

[0134] 在一些方面,N聚体序列的长度可以为约4个核苷酸。在一些方面,N聚体序列的长度为4个核苷酸。长度为4个核苷酸的N聚体序列在本文中被称为4聚体。

[0135] 在一些方面,N聚体序列的长度可以为约5个核苷酸。在一些方面,N聚体序列的长度为5个核苷酸。长度为5个核苷酸的N聚体序列在本文中被称为5聚体。

[0136] 在一些方面,N聚体序列的长度可以为约6个核苷酸。在一些方面,N聚体序列的长度为6个核苷酸。长度为6个核苷酸的N聚体序列在本文中被称为6聚体。

[0137] 在一些方面,N聚体序列的长度可以为任何数量的核苷酸。在一些方面,N聚体序列的长度可以为约至少25个核苷酸、或至少约50个核苷酸、或至少约75个核苷酸、或至少约100个核苷酸、或至少约125个核苷酸、或至少约150个核苷酸、或至少约175个核苷酸、或至少约200个核苷酸、或至少约225个核苷酸、或至少约250个核苷酸、或至少约275个核苷酸或至少约300个核苷酸。

[0138] 在一些方面,N聚体序列的长度可以为任何数量的核苷酸。在一些方面,N聚体序列的长度可以为约25个核苷酸、或约50个核苷酸、或约75个核苷酸、或约100个核苷酸、或约125个核苷酸、或约150个核苷酸、或约175个核苷酸、或约200个核苷酸、或约225个核苷酸、或约250个核苷酸、或约275个核苷酸或约300个核苷酸。

[0139] 在一些方面,Addamer可以包括MCS序列、第一IISRE序列、N聚体序列和至少第二IISRE序列。在一些方面,Addamer可以包括MCS序列,随后是第一IISRE序列、随后是N聚体序列,随后是至少第二IISRE序列。前述Addamer的示例性示意图在图8中显示为Addamer设计#1-#4。在图8所示的Addamer设计#1-#3的非限制性实例中,第一IISRE序列是当被切割时产生4个核苷酸长的5’突出端的IISRE序列,N聚体序列是3聚体序列,并且至少第二IISRE序列

是当被切割时产生平端的IISRE序列。在图8所示的Addamer设计#4的非限制性实例中,第一IISRE序列是当被切割时产生平端的IISRE序列,N聚体序列是3聚体序列,并且至少第二IISRE序列是当被切割时产生3个核苷酸长的5'突出端的IISRE序列。

[0140] 在一些方面,Addamer可以包括MCS序列、第一IISRE序列、第二IISRE序列、N聚体序列、第三IISRE序列和至少第四IISRE序列。在一些方面,可以包括MCS序列,随后是第一IISRE序列,随后是第二IISRE序列,随后是N聚体序列,随后是第三IISRE序列,随后是至少第四IISRE序列。前述Addamer的示例性示意图在图8中显示为Addamer设计#5。在图8所示的Addamer设计#5的非限制性实例中,第一IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,第二IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,N聚体序列是3聚体序列,第三IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,并且至少第四IISRE序列是当被切割时产生平端的IISRE序列。

[0141] 在一些方面,Addamer可以包括第一MCS序列、第一IISRE序列、N聚体序列、至少第二IISRE序列和至少第二MCS序列。在一些方面,Addamer可以包括第一MCS序列,随后是第一IISRE序列,随后是N聚体序列,随后是至少第二IISRE序列,随后是至少第二MCS序列。前述Addamer的示例性示意图在图8中显示为Addamer设计#6。在图8所示的Addamer设计#6的非限制性实例中,第一IISRE序列是当被切割时产生平端的IISRE序列,N聚体序列是3聚体序列,并且至少第二IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列。

[0142] 在一些方面,Addamer可以包括第一MCS序列、第一IISRE序列、第二IISRE序列、N聚体序列、第三IISRE序列、至少第四IISRE序列和至少第二MCS序列。在一些方面,Addamer可以包括第一MCS序列,随后是第一IISRE序列,随后是第二IISRE序列,随后是N聚体序列,随后是第三IISRE序列,随后是至少第四IISRE序列,随后是至少第二MCS序列。前述Addamer的示例性示意图在图8中显示为Addamer设计#7。在图8所示的Addamer设计#7的非限制性实例中,第一IISRE序列是当被切割时产生平端的IISRE序列,第二IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,N聚体序列是3聚体序列,第三IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,并且所述至少第四IISRE序列是当被切割时产生平端的IISRE序列。

[0143] 在一些方面,Addamer可以包括发夹,所述发夹包括适体序列、第一IISRE序列、N聚体序列、至少第二IISRE序列和MCS序列。在一些方面,Addamer可以包括发夹,所述发夹包括适体序列,随后是第一IISRE序列,随后是N聚体序列,随后是至少第二IISRE序列,随后是MCS序列。前述Addamer的示例性示意图在图8中显示为Addamer设计#7。在图8所示的Addamer设计#7的非限制性实例中,适体序列是凝血酶适体序列,第一IISRE序列是当被切割时产生4个核苷酸长的5'突出端的IISRE序列,N聚体序列是3聚体序列,并且至少第二IISRE序列是当被切割时产生平端的IISRE。

[0144] 当Addamer中包含两个彼此相邻的IISRE序列时,这些IISRE序列可以被称为"巢套式IISRE序列"或"巢套式IISRE位点"。在一些方面,巢套式IISRE序列可以包括两个彼此直接相邻的IISRE序列。在一些方面,巢套式IISRE序列可以包括两个彼此相邻但相隔约1个至约10个核苷酸的IISRE序列。

[0145] 不希望受理论束缚,巢套式IISRE位点是可能的,因为一些位点已经切割了与它们的识别位点充分间隔的位点,以适合第一位点与有效载荷之间的其它IISRE位点。因此,在

一些方面,Addamer可以在有效载荷(N聚体序列)的任一侧上包括独特的平端切割器位点和4碱基突出端位点。不希望受理论束缚,这显著减少了进行常规核酸生成所需的Addamer试剂的数量。

[0146] 不希望受理论束缚,在Addamer中包含巢套式IISRE序列提供了在本公开的方法中在相同位置切割两个不同位点的几种选择。选择在相同的位置切割两个不同的位点可以减少文库(参见下文)中一般核酸合成所需的不同Addamer的数量。

[0147] 在一些方面,Addamer可以包括本领域已知的任何促进克隆的元件,包含但不限于扩增引物的同源序列。不希望受理论束缚,在Addamer中包含扩增引物的同源序列可以允许恢复用于克隆繁殖的特定Addamer设计。

[0148] 在一些方面,Addamer可以包括本领域已知的任何元件,以促进在质粒或噬菌体中通过发酵大规模生产Addamer。此类元件包含但不限于对应于DNA酶瘢痕的序列和/或在使用某些DNA酶切除之后促进Addamer平滑折叠的序列(参见例如Praetorius等人,《自然》,2017,552,84-87,其通过引用整体并入本文)。

[0149] 使用本公开的Addamer的核酸合成方法

[0150] 本文所描述的Addamer可以用于本文所描述的方法中,以合成包括任何靶核酸序列的核酸分子。靶核酸序列在本文中也被称为“靶核酸”或“基因”。

[0151] 在一些方面,靶核酸序列的长度可以为至少约100个、或至少约200个、或至少约300个、或至少约500个、或至少约600个、或至少约700个、或至少约800个、或至少约900个、或至少约1000个、或至少约1500个、或至少约2000个、或至少约2500个、或至少约3000个、或至少约3500个、或至少约4000个、或至少约4500个或至少约5000个核苷酸。在一些方面,靶双链核酸可以包括至少一个均聚物序列。

[0152] 在一些方面,靶核酸序列可以包括至少一个均聚物序列。如本文所使用的,术语均聚物序列用于指任何类型的重复核酸序列,包含但不限于单核苷酸的重复序列或小基序的重复序列。在一些方面,均聚物序列的长度可以为至少约10个核苷酸、或至少约20个核苷酸、或至少约30个核苷酸、或至少约40个核苷酸、或至少约50个核苷酸、或至少约60个核苷酸、或至少约70个核苷酸、或至少约80个核苷酸、或至少约90个核苷酸或至少约100个核苷酸。

[0153] 在一些方面,靶核酸序列的GC含量可以为至少约10%、或至少约20%、或至少约50%或至少约。

[0154] 作为本公开的合成方法的一部分,一种或多种Addamer可以固定到固体载体。固体载体可以是本领域已知的任何固体载体,包含但不限于至少一种珠粒。在一些方面,所述至少一种珠粒可以包括聚丙烯酰胺、聚苯乙烯、琼脂糖或其任何组合。在一些方面,所述至少一种珠粒可以是磁性的。在一些方面,固体载体包括孔或室。在一些方面,固体载体可以包括多个孔或室。在一些方面,多个孔包括多孔板。在一些方面,固体载体可以包括玻璃。在一些方面,固体载体可以包括载玻片。在一些方面,固体载体可以包括石英。在一些方面,固体载体可以包括石英载片。在一些方面,固体载体可以包括聚苯乙烯。在一些方面,固体载体可以包括聚苯乙烯载片。在一些方面,固体载体可以包括涂层,其中涂层防止不需要的蛋白质、不需要的核酸或其它不需要的生物分子的非特异性结合。在一些方面,涂层可以包括聚乙二醇(PEG)。在一些方面,涂层可以包括三乙二醇(TEG)。

[0155] 在一些方面,其中Addamer包括包含适体序列的发夹,Addamer可以通过与适体序列结合而固定到固体载体。也就是说,固体载体可以包括至少一个与Addamer上的适体序列结合的部分。因此,在其中Addamer包括包含表3中列出的适体序列之一的发夹的非限制性实例中,固体载体可以包括表3中列出的对应配体。

[0156] 在其中Addamer包括MCS序列的一些方面,Addamer可以通过一种方法固定到固体载体,所述方法包括:a)使Addamer与至少一种对应的限制性核酸内切酶接触以切割MCS序列,由此产生5'突出端或3'突出端;以及b)将5'突出端或3'突出端与固体载体上的互补单链核酸分子杂交,由此将Addamer固定到固体载体。前述方法可以进一步包括使与固体载体上的互补单链核酸分子杂交的Addamer与连接酶接触,由此将Addamer和固体载体上的互补单链核酸分子连接。

[0157] 在其中Addamer包括MCS序列的一些方面,Addamer可以通过一种方法固定到固体载体,所述方法包括:a)使Addamer与至少一种对应的限制性核酸内切酶接触以切割MCS序列,由此产生平端;以及b)将Addamer的平端与位于固体载体上的核酸分子连接,由此将Addamer固定到固体载体。

[0158] 在一些方面,已经与固体载体连接的Addamer在本文中可以被称为"连接柱"。

[0159] 本公开的基于Addamer的核酸组装/合成方法的示意图在图9中示出。

[0160] 在所述方法的第一步骤中,提供固定到固体载体(在图9中表示为珠粒或表面)上的Addamer。此Addamer在本文中被称为连接柱,并且使用上文所描述的任何方法在一个末端处连接到固体载体,并且在另一个末端处用发夹封端。连接柱也包括MCS序列。在所述方法的下一步骤中,使连接柱与切割MCS序列的限制性核酸内切酶接触,由此产生3'突出端、5'突出端或平端。在所述方法的下一步骤中,将包括MCS序列、第一IISRE序列(在图9中表示为"L1")、第一N聚体序列(在图9中被称为"有效载荷#1")和第二IISRE序列(在图9中表示为"R1")的第一Addamer与切割MCS序列的限制性核酸内切酶接触,由此产生3'突出端、5'突出端或平端。

[0161] 在所述方法的下一步骤中,通过接触切割的连接柱、切割的Addamer和连接酶,将切割的第一Addamer与切割的连接柱连接,由此产生第一连接产物,所述第一连接产物被固定到固体载体并且包括MCS序列、第一IISRE序列、有效载荷#1序列和第二IISRE序列(参见图9的左侧)。然后用核酸外切酶处理第一连接产物,以去除任何未连接的连接柱和/或第一Addamer。

[0162] 上文所描述的步骤然后用固定到固体载体上的另一种Addamer和包括MCS序列、第三IISRE序列(在图9中表示为"R2")、第二N聚体序列(在图9中称为"有效载荷#2")和第四IISRE序列(在图9中表示为"L2")的第二Addamer重复,以产生固定到固体载体的并且包括MCS序列、第三IISRE序列、有效载荷#2序列和第四IISRE序列(参见图9的右侧)的第二连接产物。

[0163] 在所述方法的下一步骤中,第1连接产物与切割第二IISRE序列(R1)的IISRE(在图9中表示为"R1酶")接触,由此产生3'突出端、5'突出端或平端,从而产生:a)第1切割产物,其固定到固体载体并且包括MCS序列、第一IISRE序列(R1)、有效载荷#1序列和3'突出端、5'突出端或平端;以及b)包括第二IISRE序列(R1)的第2切割产物。然后通过洗涤丢弃第2切割产物。

[0164] 在所述方法的下一步骤中,第2连接产物与切割第四IISRE序列(L2)的IISRE(在图9中表示为“L2酶”)接触,由此产生3’突出端、5’突出端或平端,从而产生:a)第3切割产物,其释放到溶液中并且包括一个末端处的发夹、第3IISRE序列(R2)、有效载荷#2序列和3’突出端、5’突出端或平端;以及b)固定到固体载体并且包括MCS序列和第四IISRE序列(L2)的第4切割产物。

[0165] 在下一步骤中,通过使第1切割产物、第3切割产物和连接酶接触,将第1切割产物和第3切割产物连接在一起(例如,将包括第3切割产物的溶液转移到包括固定到固体表面的第1切割产物的溶液中,并且向溶液中添加连接酶),由此产生第3连接产物,其固定到固体表面并且包括MCS序列、第一IISRE序列(L1)、有效载荷#1序列、有效载荷#2序列和第三IISRE序列(R2)。然后用核酸外切酶处理该连接反应,以去除任何未连接的第1切割产物和/或第3切割产物。

[0166] 可以重复上文所描述的步骤,直到合成了靶核酸序列。

[0167] 示例性的27个核苷酸长的靶核酸序列的合成的示意性概述在图10A-10H中示出。要合成的序列在图10A的顶部处示出。所述序列被细分成十一个6聚体片段,所述片段通过3个核苷酸或4个核苷酸重叠,所述核苷酸将被掺入到将被连接在一起以合成靶核酸序列的Addamer中。图10B示出了示例性靶核酸序列的组装树,所述组装树绘制了包括6聚体片段的Addamer被连接以高效地合成靶核酸序列的顺序。尽管存在若干种不同的方法来遍历奇数与偶数突出端的组装和放置,组装顺序是由IISRE酶位点与要产生的序列的相容性决定的。在图10B中,带编号的6聚体(1)-(11)对应于图10C-10H中的带编号的6聚体。树的每个节点处的数字对应于组装的每个步骤处的有效载荷长度。‘4’和‘3’指示所使用的突出端的长度。所得的有效载荷序列的长度为,长度=a+b-n,其中‘a’和‘b’是输入有效载荷的长度,并且‘n’是突出端的长度。

[0168] 合成靶核酸序列的第一步骤示出在图9C中,所述图描绘了装载包括GAC的3聚体序列的Addamer和包括ATC的3聚体序列的Addamer,以形成包括GACATG 6聚体的Addamer,其为图10B中的6聚体#1。为了产生GACATG六聚体,将包括MCS序列的第一连接柱以及包括MCS序列、第一IISRE序列(在图10C中表示为“L1”)、包含序列GAC的3聚体序列和第二IISRE序列(在图10C中表示为“R1”)的第一Addamer与一种或多种限制性核酸内切酶接触以切割MCS序列,由此产生互补突出端。然后将这些互补突出端杂交,并且通过使杂交复合物与连接酶接触来将Addamer和连接柱连接在一起以产生固定到固体表面并且包括MCS序列、第一IISRE序列(L1)、3聚体序列GAC和第二IISRE序列(R1)的连接产物#1。相同的过程可以用包括MCS序列的第二连接柱以及包括MCS序列、第三IISRE序列(在图10C中表示为“L2”)、包含序列ATG的3聚体测序和第四IISRE序列(在图10C中表示为“R2”)的第二Addamer重复,以产生固定到固体表面并且包括MCS序列、第三IISRE序列(L2)、3聚体序列ATG和第四IISRE序列(R2)的连接产物#2。接下来,连接产物#1与切割第二IISRE序列(R1)的IISRE(在图10C中表示为“R1酶”)接触,由此产生固定到固体表面并且包括MCS序列、第一IISRE序列(L1)和3聚体序列GAC以及平端的切割产物#1。类似地,连接产物#2与切割第三IISRE序列(L2)的IISRE(在图10C中表示为“L2酶”)接触,由此产生释放到溶液中并且包括平端、3聚体序列ATG和第四IISRE序列(R2)的切割产物#2。然后使用连接酶将切割产物#1和切割产物#2连接在一起,以产生固定到固体载体并且包括MCS序列、第一IISRE序列(L1)、6聚体序列GACATG和第四

IISRE序列(R2)的连接产物#3。这些产物可以任选地用核酸外切酶处理,以去除任何未连接的切割产物#1和/或切割产物#2。此段中描述的步骤可以用包括不同3聚体序列的另外Addamer重复,以产生包括图10B中所示的6聚体序列#2-#11的Addamer。

[0169] 所述方法在图10D中继续,所述图示出了包括6聚体序列#1和6聚体序列#2的Addamer的连接(参见图10B)。Addamer#1固定到固体表面并且包括MCS序列、来自图10C的第一IISRE位点(L1)、6聚体序列GACATG(来自图10B的6聚体序列#1)和来自图10C的第四IISRE序列(R2)。Addamer#2被固定到固体表面并且包括MCS序列、第五IISRE位点(在图10D中表示为"L3")、6聚体序列ATGAGG(来自图10B的6聚体序列#2)和第六IISRE位点(在图10D中表示为"R3")。Addamer#1与切割第四IISRE位点(R2)的IISRE接触,以在N聚体序列中产生单链突出端,由此产生固定到固体表面并且包括MCS序列、第一IISRE序列(L1)和具有单链突出端的N聚体序列的切割产物#3。Addamer#2与切割第五IISRE位点(L3)的IISRE接触,以在N聚体序列中产生单链突出端,由此产生释放到溶液中并且包括具有单链突出端的N聚体序列和第六IISRE序列(R3)的切割产物#4。然后使用连接酶将切割产物#3和切割产物#4连接在一起,以产生固定到固体表面并且包括MCS序列、第一IISRE序列(L1)、N聚体序列GACATGAGG(要合成的靶核酸序列中的前九个核苷酸)和第六IISRE序列(R3)的连接产物#4。连接产物#4可以任选地用核酸外切酶处理,以去除任何未连接的切割产物#1和/或切割产物#2。

[0170] 所述方法在图10E中继续,其中连接产物#4和包括6聚体序列#3的Addamer(参见图10B)用对应的IISRE处理以产生经切割产物,然后将所述产物连接在一起,以产生Addamer,所述Addamer固定到固体表面并且包括N聚体序列GACATGAGGGT(SEQ ID NO:116),即要合成的靶核酸序列中的前11个核苷酸。

[0171] 根据图10B所示的组装图,在图10F-10H中重复顺序的IISRE消化和连接,直到合成了包括对应于27个核苷酸长的靶核酸序列的N聚体序列的Addamer。在最后一个步骤中,通过用切割侧接有27个核苷酸长的靶核酸序列的IISRE序列的IISRE处理Addamer,可以从最终合成的Addamer中切除27个核苷酸长的靶核酸序列。

[0172] 前述方法可以描述如下:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)重复步骤(a)-(f)直到包括靶核酸序列的核酸分子被合成。

[0173] 前述方法可以描述如下:一种方法包括:方法包括:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后

是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)重复步骤(c)-(f)直到包括靶核酸序列的核酸分子被合成。

[0174] 前述方法可以描述如下:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)用一种或多种另外的Addamer重复步骤(a)-(f)直到包括靶核酸序列的核酸分子被合成。

[0175] 前述方法可以描述如下:一种方法包括:方法包括:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)用一种或多种另外的addamer重复步骤(c)-(f)直到包括靶核酸序列的核酸分子被合成。

[0176] 前述方法可以描述如下:a)提供固定到固体载体的本公开的第一Addamer,其中所

述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)使用步骤(f)的产物和一种或多种另外的Addamer重复步骤(a)-(f)直到包括靶核酸序列的核酸分子被合成。

[0177] 前述方法可以描述如下:一种方法包括:方法包括:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)使用步骤(f)的产物和一种或多种另外的Addamer重复步骤(c)-(f)直到包括靶核酸序列的核酸分子被合成。

[0178] 前述方法可以描述如下:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)使用步骤(f)

的产物和/或一种或多种另外的Addamer重复步骤(a)-(f)的任何组合直到包括靶核酸序列的核酸分子被合成。

[0179] 前述方法可以描述如下:一种方法包括:方法包括:a)提供固定到固体载体的本公开的第一Addamer,其中所述第一Addamer包括第一IISRE序列,随后是第一N聚体序列,随后是第二IISRE序列,随后是发夹结构;b)提供固定到固体载体的本公开的第二Addamer,其中所述第二Addamer包括第三IISRE序列,随后是第二N聚体序列,随后是第四IISRE序列,随后是发夹结构;c)使所述第一Addamer与切割位于所述第一Addamer中的第二IISRE序列的IISRE接触,由此产生第一切割产物,所述第一切割产物被固定到所述固体载体并且包括第一IISRE序列、第一N聚体序列以及3'突出端、5'突出端和平端中的至少一个;d)使所述第二Addamer与切割位于所述第二Addamer中的所述第三IISRE序列的IISRE接触,由此产生第二切割产物,所述第二切割产物被释放到溶液中并且包括所述第二N聚体序列、所述第四IISRE序列以及3'突出端、5'突出端和平端中的至少一个,并且其中所述第二切割产物在一个末端处被发夹结构封端;e)使用连接酶连接第一切割产物和第二切割产物以产生第一连接产物;f)用核酸外切酶处理步骤(e)的产物,由此去除未连接的第一切割产物和/或第二切割产物;以及g)使用步骤(f)的产物和/或一种或多种另外的Addamer重复步骤(c)-(f)的任何组合直到包括靶核酸序列的核酸分子被合成。

[0180] 另一种基于Addamer的基因合成方法,被称为"汇集合成方法"在图6中示意性地示出。如图6所示,在汇集合成方法中,在共同体积中提供了包括多个不同Addamer种类的多种Addamer。每个Addamer种类包括发夹,随后是第一IISRE序列,随后是有效载荷序列,随后是第二IISRE序列,随后是发夹。Addamer种类中的两种包括"末端IISRE序列",所述序列在方法结束时被切割以释放完全组装/合成的靶核酸序列。这些Addamer种类的示意图在图6中示出。IISRE序列显示为虚线框,其中末端IISRE序列被特别标记。有效载荷序列被表示为"A"、"B"、"C"、"D"和"E"。也就是说,在图6所示的非限制性实例中,出于组装过程的目的,靶基因已经被分成五个片段。在共同体积中提供多个Addamer后,Addamer与一个或多个切割除末端IISRE序列之外的每个IISRE序列的IISRE接触,导致产生单链突出端位点,所述单链突出端位点在图6中由标有"1"、"2"、"3"、"4"、"5"和"6"的条纹框表示。在下一步骤中,互补的单链突出端位点杂交在一起,并且所得的杂交复合物与连接酶接触以将消化的Addamer连接在一起。如图6所示,此连接步骤可以在存在IISRE的情况下进行。连接后,可以用T7核酸外切酶处理连接产物以去除任何未连接的Addamer。连接步骤和T7核酸外切酶消化步骤的结果显示在图6的底部——也就是说,靶核酸序列("A-B-C-D-E")被完全组装,并且在任一侧上侧接有发夹。然后靶核酸序列可以被此产物用一个或多个切割末端IISRE序列的IISRE切除。

[0181] 在汇集合成方法中,选择的IISRE和IISRE序列被设计成产生一组互补的单链突出端,其导致在单链突出端序列杂交后靶核酸序列的组装。在汇集合成方法的一些方面,用于组装靶核酸的多个Addamer中可用的IISRE序列的数量是两个,其中第一IISRE序列用于实现初始组装,并且第二IISRE序列用于释放组装的基因,用于后续在质粒或噬菌体中的克隆。设计中的另外的考虑因素是组装体的总长度以及每个片段的长度。需要考虑位点附近的序列,因为可能存在干扰连接的二级结构如G-四链体形成的情况。用于组装基因的片段数量也是重要的考虑因素。例如,400bp的基因组装将仅需要两个200bp的Addamer有效载

荷。同样,1kb的基因将至少需要五个200bp的Addamer有效载荷。

[0182] 因此,本公开提供了一种合成包括靶核酸序列的核酸分子的方法,其中所述靶核酸序列已经被细分为两个或更多个序列片段,所述方法包括:a)提供多个Addamer,其中所述多个Addamer包括多个不同的Addamer种类,其中每个Addamer种类包括:发夹结构,随后是第一IISRE序列,随后是有效载荷序列,随后是第二IISRE序列,随后是发夹结构,其中Addamer种类的有效载荷序列对应于所述靶核酸序列的序列片段之一,其中所述多个Addamer包括每个序列片段的至少一个Addamer;b)使多个Addamer与至少一个切割每个Addamer中的至少一个IIRSRE序列的IISRE接触,由此产生至少一个单链突出端;c)通过使切割的Addamer与至少一种连接酶接触来将切割的Addamer连接在一起,由此合成包括靶核酸序列的核酸分子。在一些方面,前述方法可以进一步包括MutS酶处理步骤。在一些方面,前述方法可以进一步包括用核酸外切酶处理步骤(c)的产物,由此纯化适当连接的产物。

[0183] 用于合成特定靶核酸的Addamer设计需要分析要组装/合成的序列,以确保不存在位点IISRE,并且其次选择用于连接单独基因片段的最适当的连接位点。连接位点来源于相容性组的集合,所述组是在开发gSynth组装工艺的过程中发现的。相容性组可以少至五个4碱基5'突出端位点(参见PCT申请第PCT/US2020/051838号,公开为WO2021055962A1)和多至35个4碱基5'突出端位点(Potapov V等人,T4 DNA连接酶对四个碱基的突出端连接保真度的综合分析以及其在DNA组装中的应用(Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly.)《ACS合成生物学(ACS Synth.Biol.)》2018,7,11,2665-2674和Pryor J.M.等人,使用数据优化的组装设计来实现空前复杂的一锅Golden Gate组装(Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design.)《公共科学图书馆综合》15(9):e0238592,2020),其已经用于Golden Gate组装。

[0184] 在本公开的方法的一些方面,连接酶可以是人DNA连接酶III(hLig3)。如技术人员将理解的,hLig3表现出高平端连接效率(>60%)。在本公开的方法的一些方面,连接酶可以是T4 DNA连接酶。如技术人员将理解的,T4DNA连接酶对包括2、3或4个核苷酸长的3'或5'突出端的核酸片段显示出高连接效率(>80%)。连接酶可以是本领域已知的任何连接酶。

[0185] 在本公开的方法和的一些方面,核酸外切酶可以是T7核酸外切酶。

[0186] 已经证明经修饰的Cas9酶可以用于产生切口而不是完全双链断裂(Mali P等人,用于靶特异性筛选的CAS9转录激活因子和用于协作基因组工程的成对切口酶,《自然生物技术(Nature Biotechnology)》第31卷,第833-838页,2013以及Ann Ran,F.等人,通过RNA引导的CRISPR Cas9进行双切口以增强基因组编辑特异性(Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.)《细胞(Cell)》,第154卷,第6期,第1380-1389页,2013年9月12日)。此类Cas9切口酶活性可以通过单一向导RNA(sgRNA)引导至靶DNA中的特定位点。另外,这种切口酶活性可以以相反的方式引导至不同的正和负DNA链,导致切口在消化之后暴露突出的单链DNA。对此类相对切口位点的研究表明,10-15个碱基的突出端产生最好的连接结果,有趣的是,10个碱基对的间距是双螺旋一圈的距离(Wang.R.Y.等人,基于切口酶系统的DNA片段组装(DNA Fragments Assembly Based on Nicking Enzyme System.)《公共科学图书馆综合》,2013年3月,第8卷,第3期,e57943)。因此,在基于Addamer的合成方法的一些方面,与靶向Addamer内的有效载荷的末

端的sgRNA组合的突变体Cas9可以用于产生突出端,所述突出端可以被连接以组装所关注基因,而不是IIRE序列和IIRE。

[0187] 在本公开的方法的一些方面,合成的靶核酸序列的纯度为至少约80%、或至少约85%、或至少约90%、或至少约95%或至少约99%。

[0188] 在一些方面,合成的靶核酸序列的纯度是指作为单次连接反应或多轮连接反应的一部分形成的与正确的/期望的连接产物相对应的总连接产物的百分比。不希望受理论束缚,包括核酸分子产物的连接的本公开的方法可以产生多种连接产物,所述多种连接产物中的一些连接产物对应于正确的/期望的连接产物,并且所述多种连接产物中的一些连接产物是不期望的(副反应、不正确的连接等)。连接产物或正在合成的靶分子的纯度可以以百分比的形式表示,所述百分比对应于所形成的与正确/期望的连接产物相对应的总连接产物的百分比。

[0189] 本公开的gSynth方法和组合物

[0190] 本文公开的汇集的寡核苷酸合成方法和组合物可以与gSynth方法(其公开为WO2021055962A1的PCT申请第PCT/US2020/051838号中详细描述)组合使用。gSynth方法也被称为“双链几何合成(gSynth)”和与其相关的组合物,用于合成长的任意双链核酸序列。在双链gSynth组装反应中,通过计算将靶序列(即,要合成的序列)分成一组相邻的双链核酸片段,然后在系统组装方法中,在一次一对的连接反应中将这些相邻的双链核酸片段连接在一起。这些片段具有3’和/或5’突出的单链N聚体位点,具有三个特性:1)N聚体位点在连接反应中不是自杂交或自反应的。2)片段的一个末端处的N聚体位点不与另一个末端处的N聚体位点交叉杂交或交叉反应。最后,3)相邻片段对的每个片段上有一个N聚体位点,它将在连接反应中与相邻片段杂交和连接,从而产生新的更长的双链片段。公开为WO2021055962A1的PCT申请第PCT/US2020/051838号提供了优选的N聚体位点,其促进更高效和准确的连接反应,由此使得本公开的双链gSynth方法可用于合成前所未有长度的核酸序列,这是使用现有的核酸组装和合成技术无法实现的。本公开的双链片段可以使用本文所描述的方法产生。

[0191] 图11A-11G展示了双链gSynth组装反应的非限制性实例。图11A示出了要使用本公开的双链gSynth方法合成的靶序列(命名为“5050Seq03”)。粗体和下划线的序列的部分对应于已被选择的4聚体突出端,因此定义了将用于合成整个序列的片段。图11B示出了图11A所示的序列的单独双链核酸片段,其将用于本公开的双链gSynth方法中以构建5050Seq03。图11D是用于合成图11A所示的序列的双链gSynth方法中第一轮连接的示意图。在第一轮连接中,片段1和2、片段3和4、片段5和6、片段7和8、片段9和10、片段11和12以及片段13和14通过它们互补的5’突出端杂交,并且然后连接在一起以产生片段1+2、片段3+4、片段5+6、片段7+8、片段9+10、片段11+12和片段13+14。图1E是用于合成图11A所示的序列的双链gSynth方法中第二轮连接的示意图。在第二轮连接中,片段1+2和3+4、片段5+6和7+8以及片段11+12和13+14通过它们互补的5’突出端杂交,并且然后连接在一起以产生片段1+2+3+4、片段5+6+7+8和片段11+12+13+14。图1F是用于合成图11A所示的序列的双链gSynth方法中第三轮连接的示意图。在第三轮连接中,片段1+2+3+4和5+6+7+8以及片段9+10和11+12+13+14通过它们互补的5’突出端杂交,并且然后连接在一起以产生片段1+2+3+4+5+6+7+8和片段9+10+11+12+13+14。图1G是用于合成图11A所示的序列的双链gSynth方法中第四轮和最后一轮连接

的示意图。在第四轮连接中,片段1+2+3+4+5+6+7+8和9+10+11+12+13+14通过它们互补的5'突出端杂交并连接在一起,由此产生图11A所示的序列。

[0192] 因此,本文公开的汇集的寡核苷酸合成方法和组合物可以用于产生双链片段,所述双链片段用于上文所描述的gSynth方法和公开为WO2021055962A1的PCT申请第PCT/US2020/051838号中。

[0193] 另外的示例性实施例

[0194] 1.一种双链Addamer,其中所述Addamer包括:

[0195] a)允许产生能够与互补突出端连接的突出端的第一序列;

[0196] b)有效载荷序列;

[0197] c)允许产生能够与互补突出端连接的突出端的第二序列;以及

[0198] d)Addamer,其中所述Addamer的至少一个末端包括发夹结构。

[0199] 2.根据权利要求1所述的Addamer,其中所述Addamer包括所述Addamer的两个末端处的发夹结构。

[0200] 3.根据权利要求1所述的Addamer,其中允许产生能够与互补突出端连接的突出端的所述第一序列和所述第二序列是限制性核酸内切酶切口酶识别位点。

[0201] 4.根据权利要求1所述的Addamer,其中允许产生能够与互补突出端连接的突出端的所述第一序列和所述第二序列是在存在sgRNA的情况下支持突变体Cas9的切口活性的位点。

[0202] 5.根据权利要求1所述的Addamer,其中产生的突出端为0个、1个、2个、3个、4个或5个碱基。

[0203] 6.根据权利要求1所述的Addamer,其中产生的突出端为10个至15个碱基。

[0204] 7.根据权利要求1所述的Addamer,其中产生的突出端对应于特定相容性组的4碱基突出端。

[0205] 8.一种寡核苷酸池,其中每个寡核苷酸是Addamer的链,但不与所述池中的任何其它寡核苷酸直接互补。

[0206] 9.两种、若干种或许多种寡核苷酸池,其中每个寡核苷酸是Addamer的链,其中任何两个、三个、四个或五个池的混合物独特地导致包括基因或基因的片段的Addamer或一组Addamer的形成。

[0207] 10.一种用于产生Addamer的方法,其中在有利于杂交的条件下,两个或更多个寡核苷酸池的混合物导致Addamer结构的形成,所述方法包括以下步骤:

[0208] a)杂交

[0209] b)连接以解析固有的切口

[0210] c)用核酸外切酶处理以去除非Addamer DNA。

[0211] 11.根据权利要求10所述的方法,其中所述混合物用His标记的Taq MutS和Ni-NTA琼脂糖处理,以去除含有错配碱基对的Addamer。

[0212] 12.一种基因组装的方法,其中在共同体积中具有内部和外部消化位点的一组Addamer:

[0213] a)用核酸内切酶或一组酶或与短核酸组合的酶处理以去除每个Addamer的内部位点的一个或两个发夹

[0214] b)在适当的缓冲液中用连接酶处理

[0215] c)用核酸外切酶处理以去除未连接的材料。

[0216] 实例

[0217] 实例1——使用汇集的寡核苷酸产生addamer和长序列组装体

[0218] 在此实例中,寡核苷酸池(在本文中也被称为多个)用于产生一组Addamer,并且这些Addamer后续可以被组装成全长靶核酸序列。

[0219] 在此实例中,使用了图论方法将431bp的序列程序化地分成五个片段(F1-F5),优化各片段间大致相等的GC含量,并且选择大小为约100bp的片段(参见图12)。预测的片段具有相容的单链突出端(CATC、AACG、TTGA、CAGA),其中估计的连接保真度为100%(参见图13)。

[0220] 每个预测的片段序列用于创建在有效载荷的任一个末端处包含II型S限制性核酸内切酶位点(IISRE)的Addamer。末端片段F1和F5还含有用于扩增的PCR引物位点和用于后续克隆步骤的限制性核酸内切酶(RE)位点(参见图13)。

[0221] 如本文所描述,每个Addamer由一个(多个)核酸分子池提供的三个寡核苷酸序列形成。然后使用图6中描述的汇集合成方法组装Addamer,并且在上文中进一步详述。如图6所示,用IISRE解封端每个Addamer,留下相容的4核苷酸单链突出端(在此施例中,IISRE是BsaI)。F1开头和F5末端的帽保持完整,使得片段将组装成单个大的Addamer,即组装的产物。这个较大的Addamer序列被设计成在随后的步骤中通过用不同的IISRE(在此实例中为BbsI)处理留下的突出端组装成更长的序列。扩增较大的Addamer产物,用REs XbaI和EcoRI切割,并且连接到克隆载体中。对质粒克隆进行毛细管测序。最后,通过比对毛细管测序的结果,证实了组装的基因具有期望的靶序列(参见图14)。

[0222] 上文所描述的方案也用于产生另外的靶序列,以证明所述方法的普遍适用性。图15示出了这些另外靶序列的每个组装程序的设计,包含靶序列被细分成的片段数量(“片段数量”栏)以及使用了多少单独核酸分子来产生对应于每个片段的Addmer(“每片段的寡聚物”栏)。这些实验证明,多达六个片段可以用Addamer组合产生更大的序列。另外,用于产生组分Addamer的寡核苷酸序列的数量可以多达四个,并且可以同时用于具有多达22个不同寡核苷酸序列的混合池中,如序列F的情况(参见图15)。

[0223] 这些结果证明,本文所描述的组合物和方法可以用于使用基于Addamer的组装方法来组装和合成靶核酸序列,其中单独的Addamer是使用利用汇集的寡核苷酸合成产生的核酸分子池(多个)来产生的。

<110> 卡美纳生物科学有限公司(Camena Bioscience Limited)
<120> 根据寡核苷酸池的基于Addamer的基因组装
<130> DNWR-010/001WO 327431-2036
<150> US 63/186,871
<151> 2021-05-11
<160> 116
<170> PatentIn第3.5版
<210> 1
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 1
acagcgatgt ctagatcgg 19
<210> 2
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 2
cacgacctgc aaaaaggcac tttttgcagg tc 32
<210> 3
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 3
gttttccacg ggaaaactgt cgctacagat ctagcc 36
<210> 4
<211> 28
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 4

cagcctttga cagcgatgtc tagatcgg 28

<210> 5

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 5

cacgacctgc aaaaaggc 18

<210> 6

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 6

gggaaaactg tcgctacaga tctagcc 27

<210> 7

<211> 17

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 7

gtgctggacg tttttca 17

<210> 8

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 8

gacatgaggg tgcccgctca gctcctg 27

<210> 9

<211> 11

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 9
gacatgaggg t 11
<210> 10
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 10
ctgtactccc a 11
<210> 11
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 11
gggtgcccgc tc 12
<210> 12
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 12
cccacgggcg ag 12
<210> 13
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 13
gctcagctcc tg 12
<210> 14
<211> 12
<212> DNA
<213> 人工序列
<220>

<223> 合成序列
<400> 14
cgagtcgagg ac 12
<210> 15
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 15
gggtgcccgc tc 12
<210> 16
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 16
cccacgggcg ag 12
<210> 17
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 17
cccacgggcg ag 12
<210> 18
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 18
gctcagctcc tg 12
<210> 19
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 合成序列
<400> 19
cgagtcgagg ac 12
<210> 20
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 20
gctcagctcc tg 12
<210> 21
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 21
gggtgcccgc tcagctcctg 20
<210> 22
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 22
cccacgggcg agtcgaggac 20
<210> 23
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 23
gacatgaggg t 11
<210> 24
<211> 11
<212> DNA

<213> 人工序列
<220>
<223> 合成序列
<400> 24
ctgtactccc a 11
<210> 25
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 25
ctgtactccc a 11
<210> 26
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 26
gggtgcccgc tcagctcctg 20
<210> 27
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 27
cccacgggcg agtcgaggac 20
<210> 28
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 28
gggtgcccgc tcagctcctg 20
<210> 29
<211> 16

<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 29
cgggcgagtc gaggac 16
<210> 30
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 30
gacatgaggg tgcccgctca gctcctg 27
<210> 31
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 31
ctgtactccc acgggcgagt cgaggac 27
<210> 32
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 32
gacatgaggg tgcccgctca gct 23
<210> 33
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 33
actcccacgg gcgagtcgag gac 23
<210> 34

<211> 300
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 34
cgacgtgtct atcgaatacg ttgtctgcca acatctgtcg caaccaatgt aggccggcaa 60
tgttaagctg tcgacgatgc aagtcaataa gcgaacgagt gctccgtctc tctcaacctc 120
atagtgacta ggaacgcgta atgcatgaac aatgccgtca accgtcaggt gattaggcgt 180
gtcgattctc cgtccttgta caagacgtgg tagaaccggc aatgttcgac tctgtcgaag 240
aacgctgtca gagtcaatca acatctgctc gtgtagttct gtgcaggtat gacggattcc 300
<210> 35
<211> 10
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 35
gtgtctatcg 10
<210> 36
<211> 10
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 36
gctgcacaga 10
<210> 37
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 37
aatacgttgt ctgccaacat ctgt 24
<210> 38
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 合成序列
<400> 38
tagcttatgc aacagacggt tgta 24
<210> 39
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 39
cgcaaccaat gtaggccggc 20
<210> 40
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 40
gacagcgttg gttacatccg 20
<210> 41
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 41
aatgttaagc tgtcgacgat gcaagtc 27
<210> 42
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 42
gccgttacaa ttcgacagct gctacgt 27
<210> 43
<211> 22
<212> DNA

<213> 人工序列
<220>
<223> 合成序列
<400> 43
aataagcgaa cgagtgctcc gt 22
<210> 44
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 44
tcagttattc gcttgctcac ga 22
<210> 45
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 45
ctctctcaac ctcatagtga ctaggaa 27
<210> 46
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 46
ggcagagaga gttggagtat cactgat 27
<210> 47
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 47
cgcgtaatgc atgaacaatg c 21
<210> 48
<211> 21

<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 48
ccttgcgcat tacgtacttg t 21
<210> 49
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 49
cgtcaaccgt caggtgatta ggc 23
<210> 50
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 50
tacggcagtt ggcagtccac taa 23
<210> 51
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 51
gtgtcgattc tccgtcctt 19
<210> 52
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 52
tccgcacagc taagaggca 19
<210> 53

<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 53
gtacaagacg tggtagaacc ggca 24
<210> 54
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 54
ggaacatgtt ctgcaccatc ttgg 24
<210> 55
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 55
atgttcgact ctgtcgaaga ac 22
<210> 56
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 56
ccgttacaag ctgagacagc tt 22
<210> 57
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 57
gctgtcagag tcaatcaac 19

<210> 58
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 58
cttgcgacag tctcagtta 19
<210> 59
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 59
atctgctcgt gtagttctgt g 21
<210> 60
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 60
gttgtagacg agcacatcaa g 21
<210> 61
<211> 17
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 61
caggtatgac ggattcc 17
<210> 62
<211> 17
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 62

acacgtccat actgcct 17

<210> 63

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 63

gtgtctatcg aatacgttgt ctgccaacat ctgt 34

<210> 64

<211> 34

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 64

gctgcacaga tagcttatgc aacagacggt tgta 34

<210> 65

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 65

cgcaaccaat gtaggccggc aatgttaagc tgtcgacgat gcaagtc 47

<210> 66

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 66

gacagcgttg gttacatccg gccgttacaa ttcgacagct gctacgt 47

<210> 67

<211> 49

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 67
aataagcgaa cgagtgctcc gtctctctca acctcatagt gactaggaa 49
<210> 68
<211> 49
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 68
tcagttattc gcttgctcac gaggcagaga gagttggagt atcactgat 49
<210> 69
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 69
cgcgtaatgc atgaacaatg ccgtcaaccg tcaggtgatt aggc 44
<210> 70
<211> 44
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 70
ccttgcgcat tacgtacttg ttacggcagt tggcagtcca ctaa 44
<210> 71
<211> 43
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 71
gtgtcgattc tccgtccttg tacaagacgt ggtagaaccg gca 43
<210> 72
<211> 43
<212> DNA
<213> 人工序列
<220>

<223> 合成序列
<400> 72
tccgcacagc taagaggcag gaacatgttc tgcaccatct tgg 43
<210> 73
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 73
atgttcgact ctgtcgaaga acgctgtcag agtcaatcaa c 41
<210> 74
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 74
ccgttacaag ctgagacagc ttcttgcgac agtctcagtt a 41
<210> 75
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 75
atctgctcgt gtagttctgt gcaggtatga cggattcc 38
<210> 76
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 76
gttgtagacg agcacatcaa gacacgtcca tactgcct 38
<210> 77
<211> 81
<212> DNA
<213> 人工序列

<220>
<223> 合成序列
<400> 77
gtgtctatcg aatacgttgt ctgccaacat ctgtcgcaac caatgtaggc cggcaatgtt 60
aagctgtcga cgatgcaagt c 81
<210> 78
<211> 81
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 78
gctgcacaga tagcttatgc aacagacggt tgtagacagc gttggttaca tccggccgtt 60
acaattcgac agctgctacg t 81
<210> 79
<211> 93
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 79
aataagcgaa cgagtgctcc gtctctctca acctcatagt gactaggaac gcgtaatgca 60
tgaacaatgc cgtcaaccgt caggtgatta ggc 93
<210> 80
<211> 93
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 80
tcagttattc gcttgctcac gaggcagaga gagttggagt atcactgatc cttgcgcatt 60
acgtacttgt tacggcagtt ggcagtccac taa 93
<210> 81
<211> 43
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 81

gtgtcgattc tccgtccttg tacaagacgt ggtagaaccg gca 43

<210> 82

<211> 43

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 82

tccgcacagc taagaggcag gaacatgttc tgcaccatct tgg 43

<210> 83

<211> 79

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 83

atgttcgact ctgtcgaaga acgctgtcag agtcaatcaa catctgctcg tgtagttctg 60

tgcaggtatg acggattcc 79

<210> 84

<211> 79

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 84

ccgttacaag ctgagacagc ttcttgcgac agtctcagtt agttgtagac gagcacatca 60

agacacgtcc atactgcct 79

<210> 85

<211> 81

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 85

gtgtctatcg aatacgttgt ctgccaacat ctgtcgcaac caatgtaggc cggcaatgtt 60

aagctgtcga cgatgcaagt c 81

<210> 86

<211> 85

<212> DNA

<213> 人工序列
<220>
<223> 合成序列
<400> 86
gctgcacaga tagcttatgc aacagacggt tgtagacagc gttggttaca tccggccgtt 60
acaattcgac agctgctacg ttcag 85
<210> 87
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 87
gtgtcgattc tccgtccttg tacaagacgt ggtagaaccg gcaatgttcg actctgtcg 59
<210> 88
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 88
tccgcacagc taagaggcag gaacatgttc tgcaccatct tggccgttac aagctgagac 60
agc 63
<210> 89
<211> 43
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 89
cagcagaaga gcaatacgac tcactatagt ctagagaaga cgc 43
<210> 90
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 90
gcgtcttctc gttatgctga gtgatatcag atctcttctg cggttg 46

<210> 91
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 91
catccgagac cgcgc 15
<210> 92
<211> 10
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 92
gctctggcga 10
<210> 93
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 93
cagcggtctc t 11
<210> 94
<211> 14
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 94
gcgccagaga gtag 14
<210> 95
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 95

aacgggagac cgcgc 15

<210> 96

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 96

cctctggcga 10

<210> 97

<211> 11

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 97

cagcggtctc t 11

<210> 98

<211> 14

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 98

gcgccagaga ttgc 14

<210> 99

<211> 15

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 99

ttgaggagac cgcgc 15

<210> 100

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 合成序列

<400> 100
cctctggcga 10
<210> 101
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 101
cagcggtctc g 11
<210> 102
<211> 14
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 102
gcgccagagc aact 14
<210> 103
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 103
cagatgagac cgcgc 15
<210> 104
<211> 10
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 104
actctggcga 10
<210> 105
<211> 11
<212> DNA
<213> 人工序列
<220>

<223> 合成序列
<400> 105
cagcggtctc a 11
<210> 106
<211> 14
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 106
gcgccagagt gtct 14
<210> 107
<211> 47
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 107
tgttcagtct tcgaattcct ttagtgaggg ttaatgctct tcagcgc 47
<210> 108
<211> 42
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 108
gtcagaagct taaggaaatc actcccaatt acgagaagtc ga 42
<210> 109
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 109
tcgattggat tgtgccggaa gtgctggctc ga 32
<210> 110
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 合成序列
<400> 110
agtccgtggt agggcaggtt ggggtgact 29
<210> 111
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 111
cagttgatcc tttggatacc ctg 23
<210> 112
<211> 42
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 112
ggcgtgcagt gccttcggcc gtgcggtgcc tccgtcacgc ct 42
<210> 113
<211> 43
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 113
accgtctgag cgattcgtac tttattcggg aggtatcagc ggg 43
<210> 114
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 114
cctggacgga accagaatac ttttggtctc cagg 34
<210> 115
<211> 43
<212> DNA

<213> 人工序列
<220>
<223> 合成序列
<400> 115
aaataccccc ccttcggtgc aaagcaccga agggggggta ttt 43
<210> 116
<211> 11
<212> DNA
<213> 人工序列
<220>
<223> 合成序列
<400> 116
gacatgaggg t 11

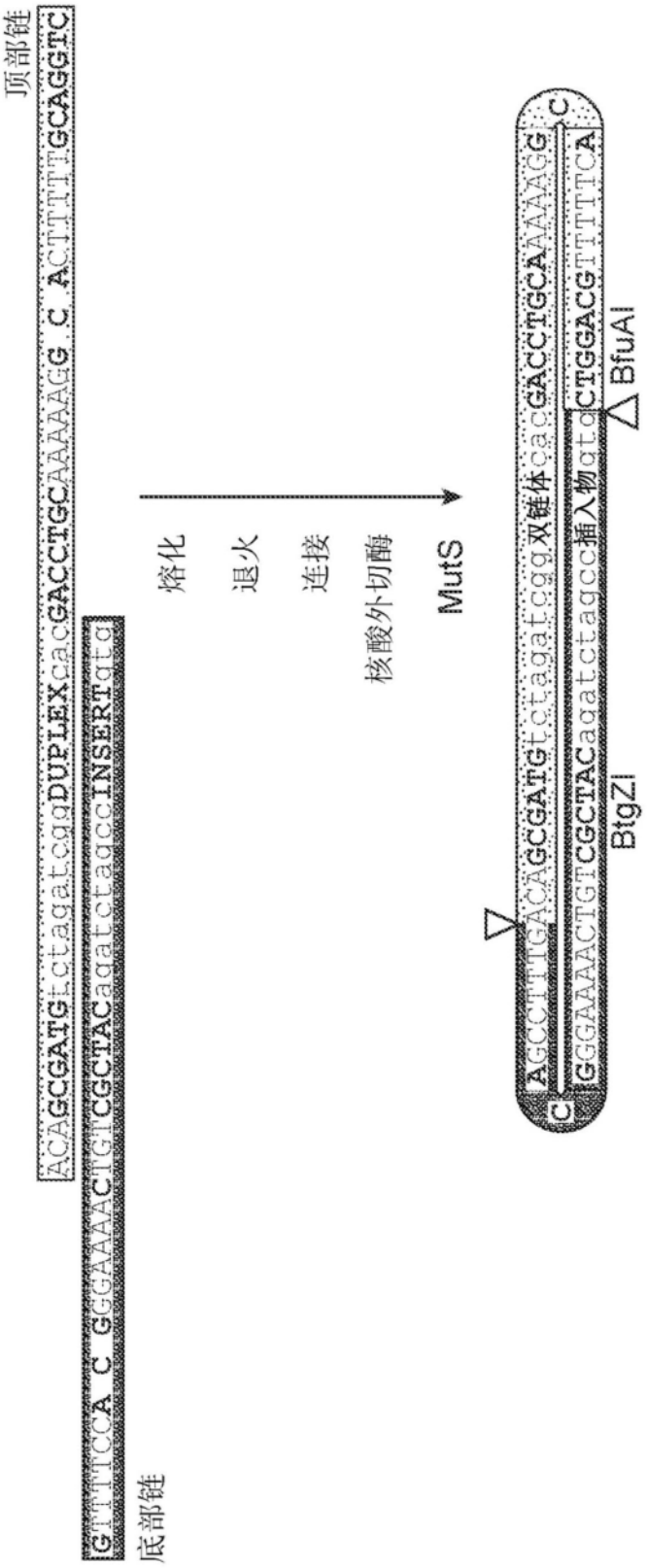

顶部链
ACAGCGATGtctagatcggDUPLEXcacGACCTGCAAAAAGG C ACTTTTGCAGGTC
IGTTTTCCA C GGGAAAACTGTCGCTACagatctagccINSERTgtg
底部链
熔化
退火
连接
核酸外切酶
MutS
AGCCTTTGACAGCGATGtctagatcgg双链体caccGACCTGCAAAAAGG
C
GGGAAAACTGTCGCTACagatctagcc插入物gtg CTGGACGTTTTTCA
C
BtgZI
BfuAI

图1

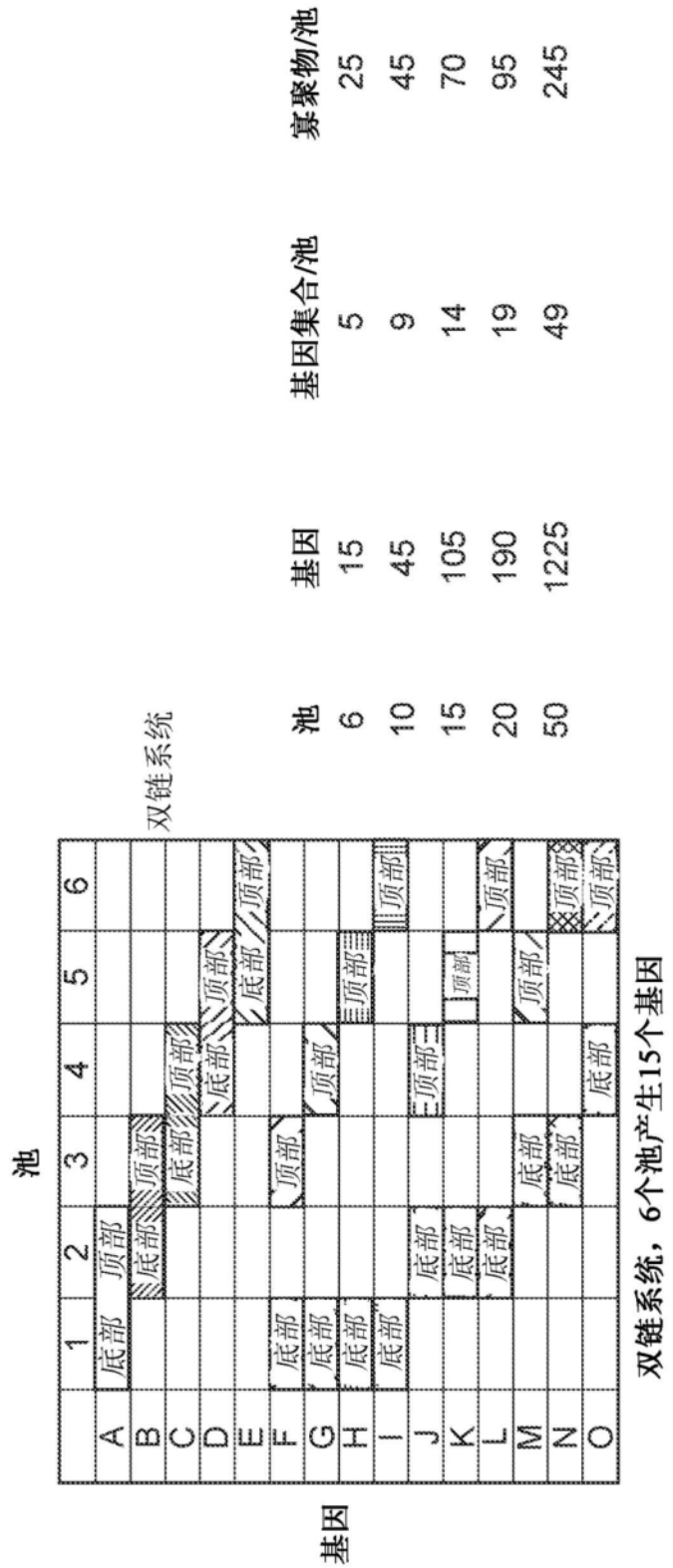

池
基因
双链系统
池 基因 基因集合/池 寡聚物/池
6 15 5 25
10 45 9 45
15 105 14 70
20 190 19 95
50 1225 49 245
双链系统，6个池产生15个基因

图2

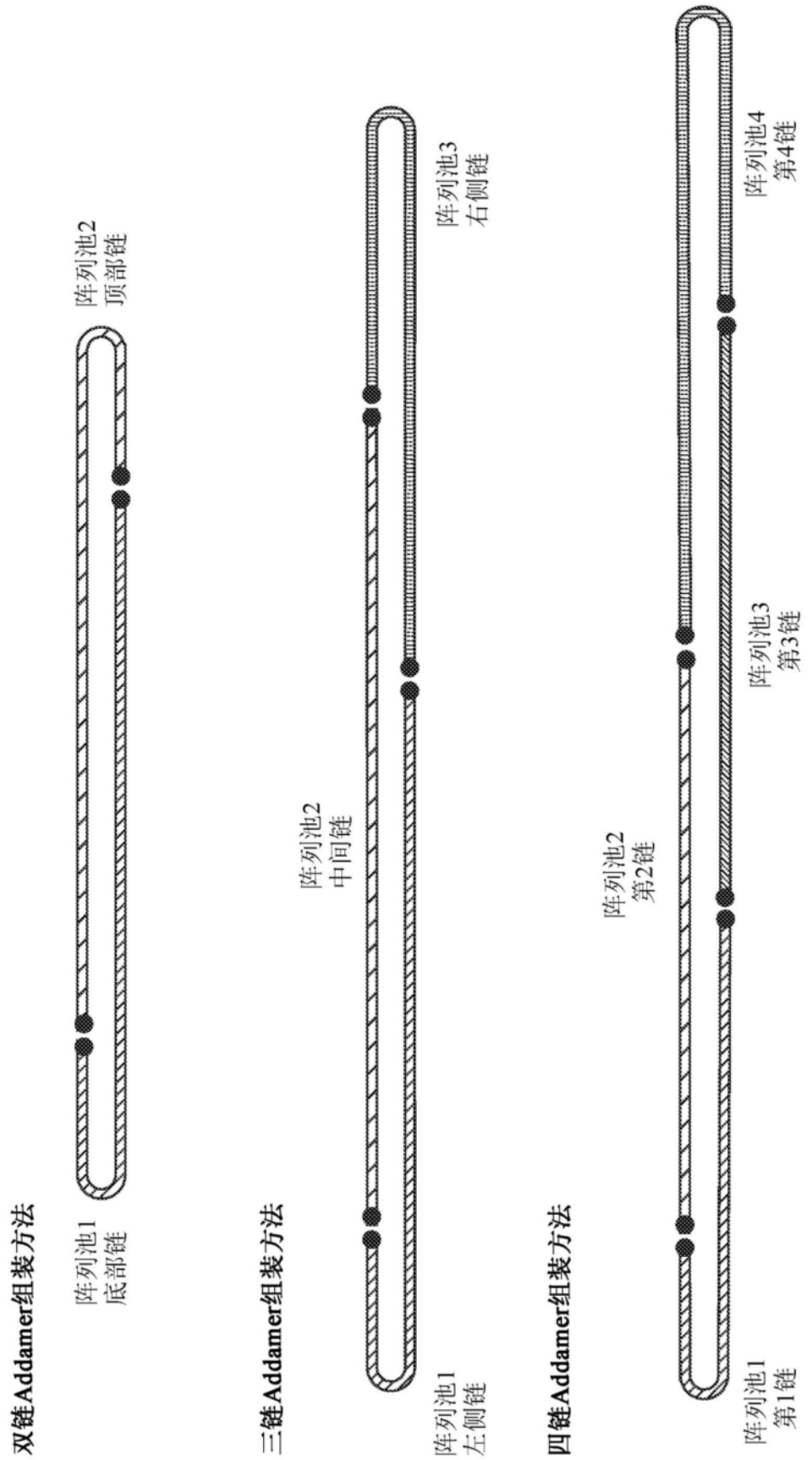

双链Addamer组装方法
阵列池1
底部链
阵列池2
顶部链
三链Addamer组装方法
阵列池1
左侧链
阵列池2
中间链
阵列池3
右侧链
四链Addamer组装方法
阵列池1
第1链
阵列池2
第2链
阵列池3
第3链
阵列池4
第4链

图3

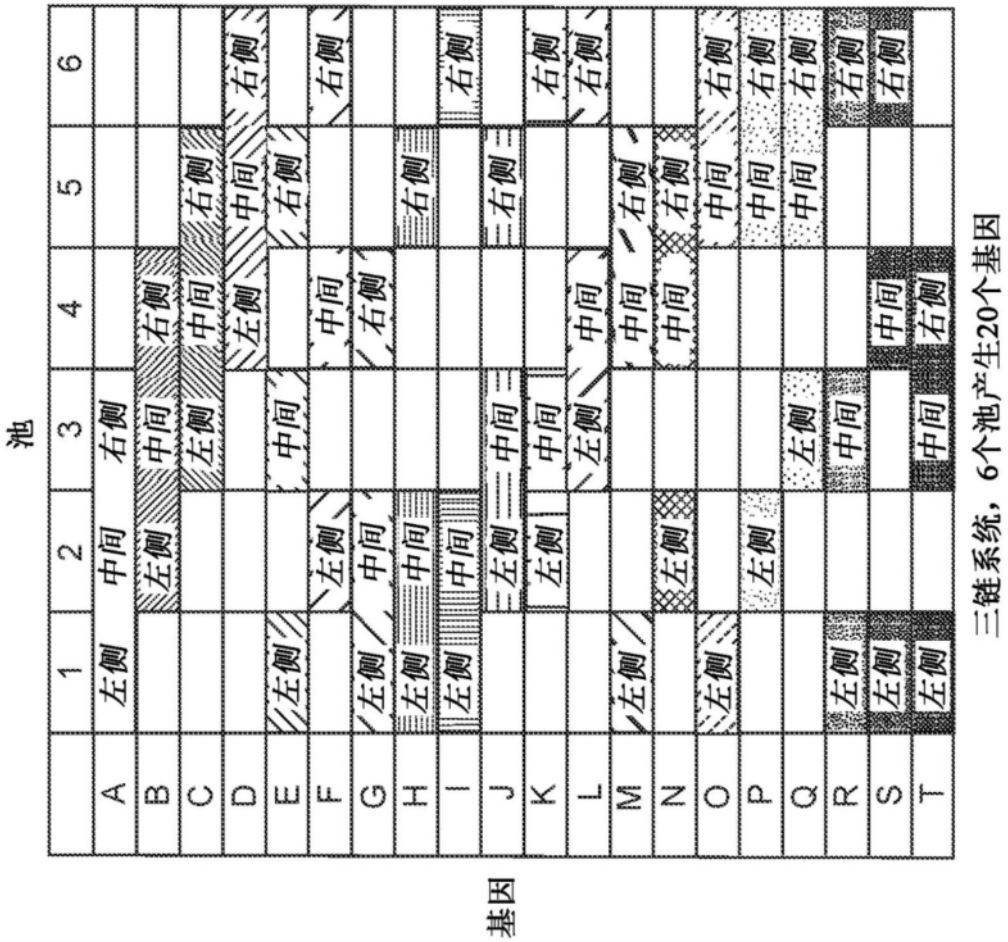
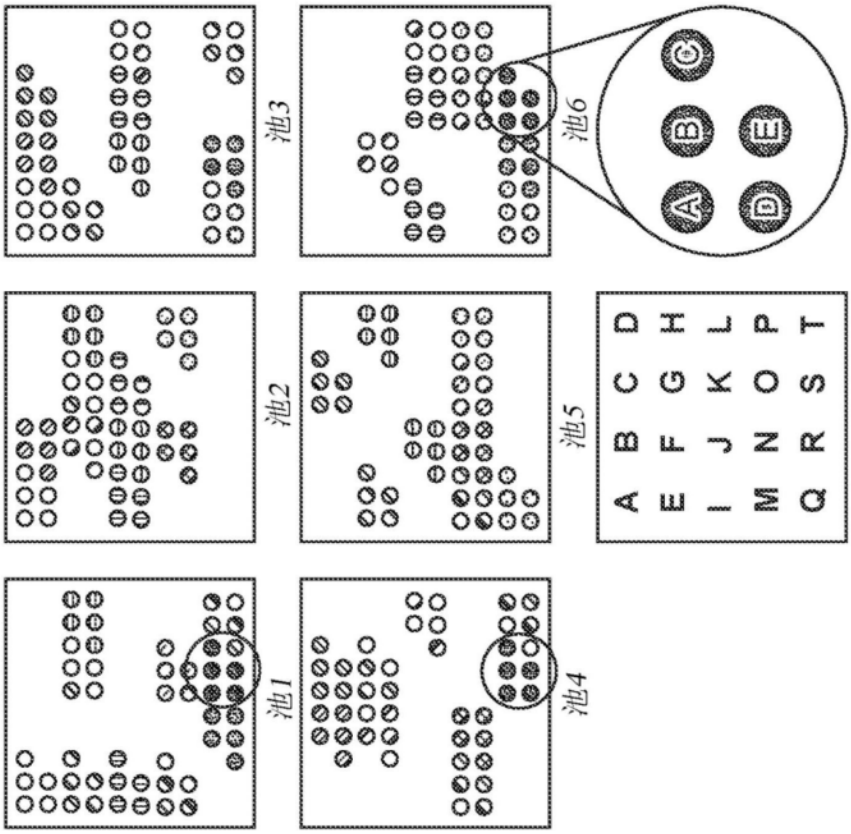

池
1 2 3 4 5 6
A 左侧 中间 右侧
B 左侧 中间 右侧
C 左侧 中间 右侧
D 左侧 中间 右侧
E 左侧 中间 右侧
F 左侧 中间 右侧
G 左侧 中间 右侧
H 左侧 中间 右侧
I 左侧 中间 右侧
J 左侧 中间 右侧
K 左侧 中间 右侧
L 左侧 中间 右侧
M 左侧 中间 右侧
N 左侧 中间 右侧
O 左侧 中间 右侧
P 左侧 中间 右侧
Q 左侧 中间 右侧
R 左侧 中间 右侧
S 左侧 中间 右侧
T 左侧 中间 右侧
基因
池 基因 基因集合/池 寡聚物/池
6 20 10 50
10 120 36 180
15 455 91 455
20 1140 171 855
50 19600 1176 5880
三链系统，6个池产生20个基因

图4


池1
池2
池3
池4
池5
池6
A B C D
E F G H
I J K L
M N O P
Q R S T

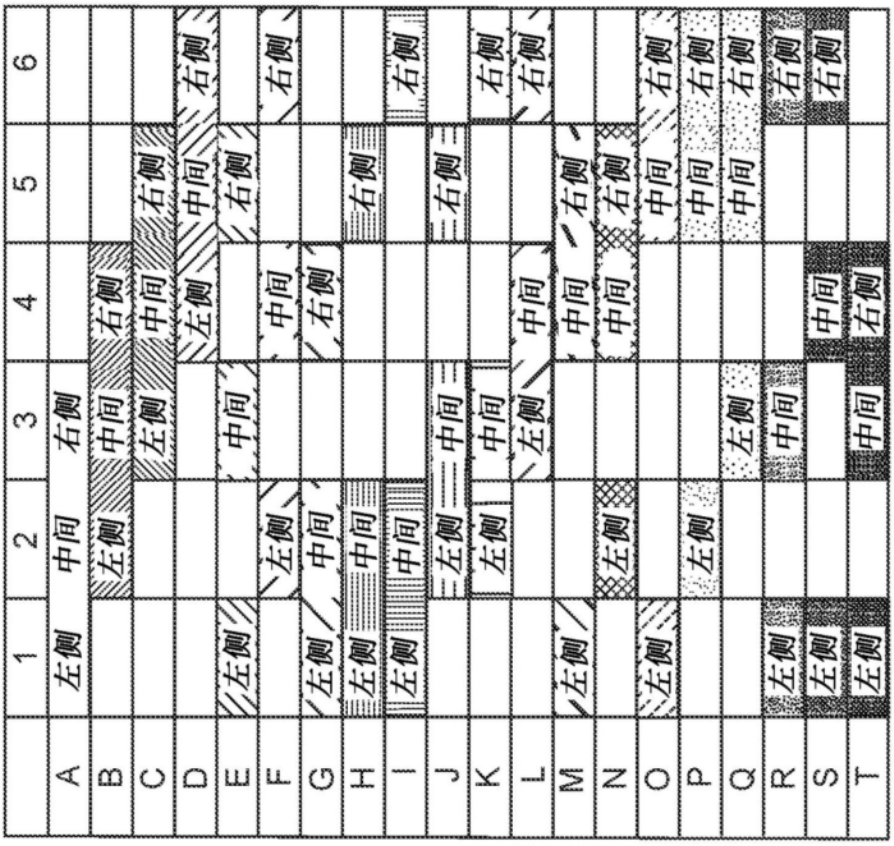

1 2 3 4 5 6
A 左侧 中间 右侧
B 左侧 中间 右侧
C 左侧 中间 右侧
D 左侧 中间 右侧
E 左侧 中间 右侧
F 左侧 中间 右侧
G 左侧 中间 右侧
H 左侧 中间 右侧
I 左侧 中间 右侧
J 左侧 中间 右侧
K 左侧 中间 右侧
L 左侧 中间 右侧
M 左侧 中间 右侧
N 左侧 中间 右侧
O 左侧 中间 右侧
P 左侧 中间 右侧
Q 左侧 中间 右侧
R 左侧 中间 右侧
S 左侧 中间 右侧
T 左侧 中间 右侧

图5

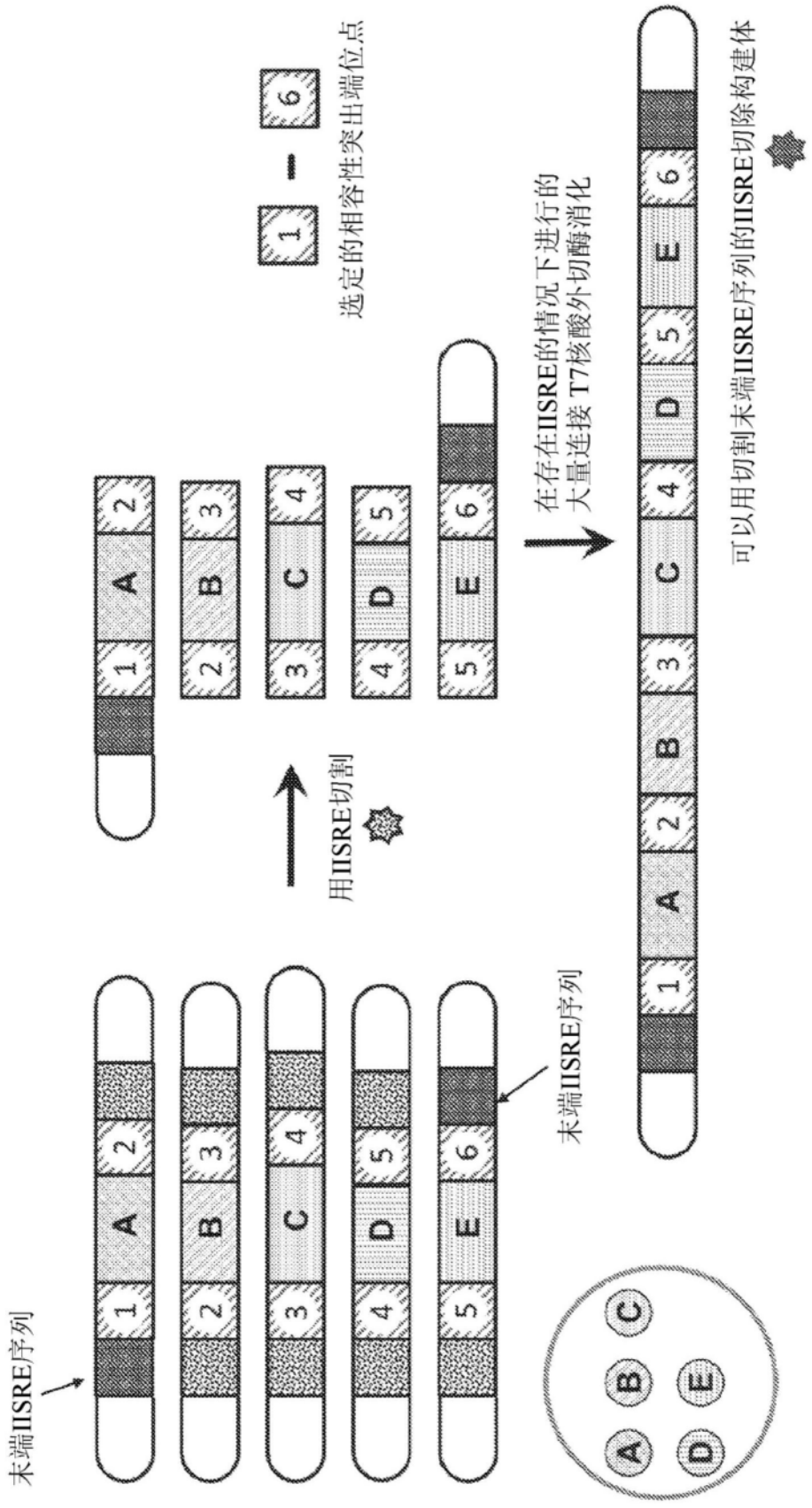

末端IISRE序列
用IISRE切割
选定的相容性突出端位点
在存在IISRE的情况下进行的
大量连接 T7核酸外切酶消化
末端IISRE序列
可以用切割末端IISRE序列的IISRE切除构建体

图6

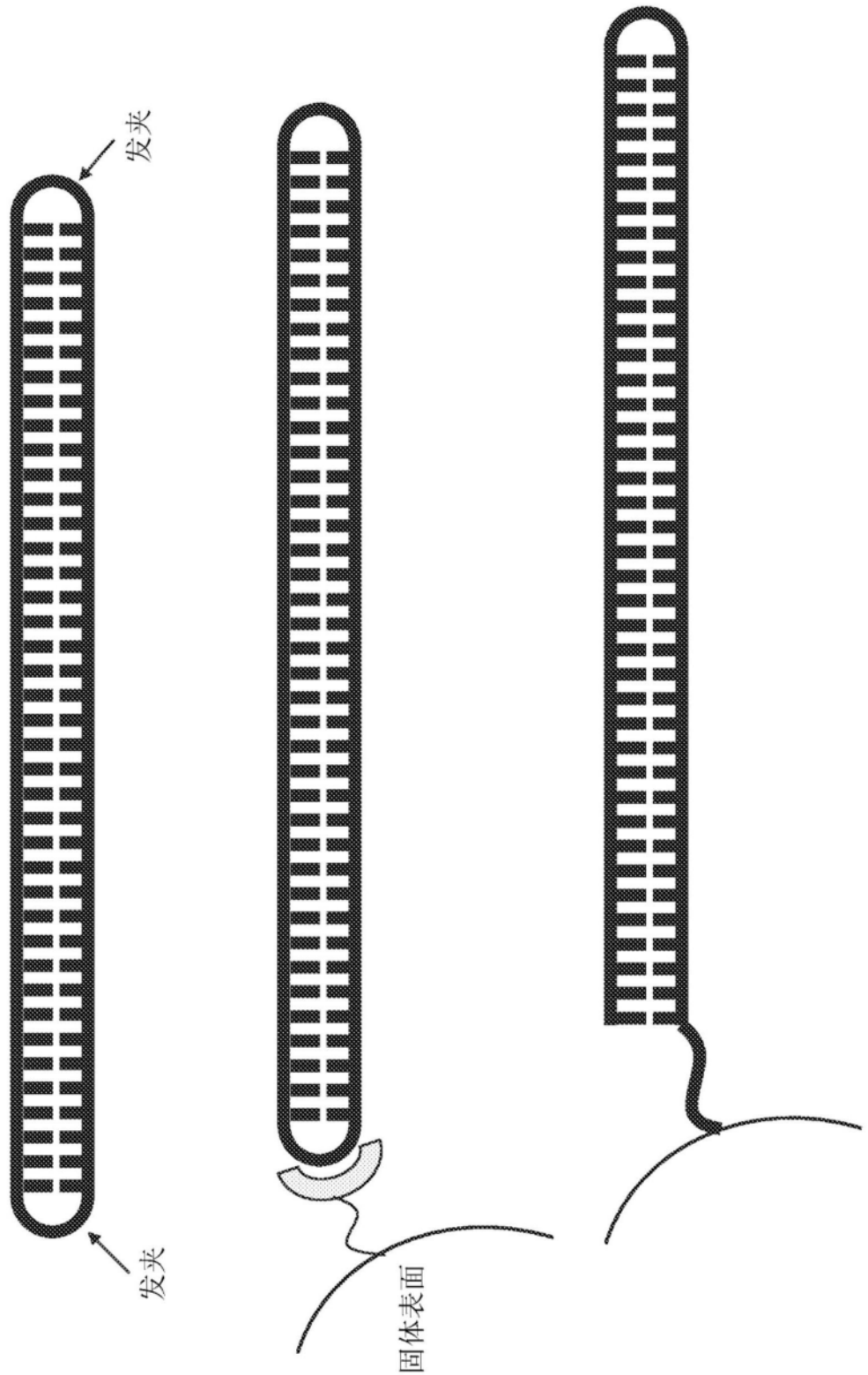


图7

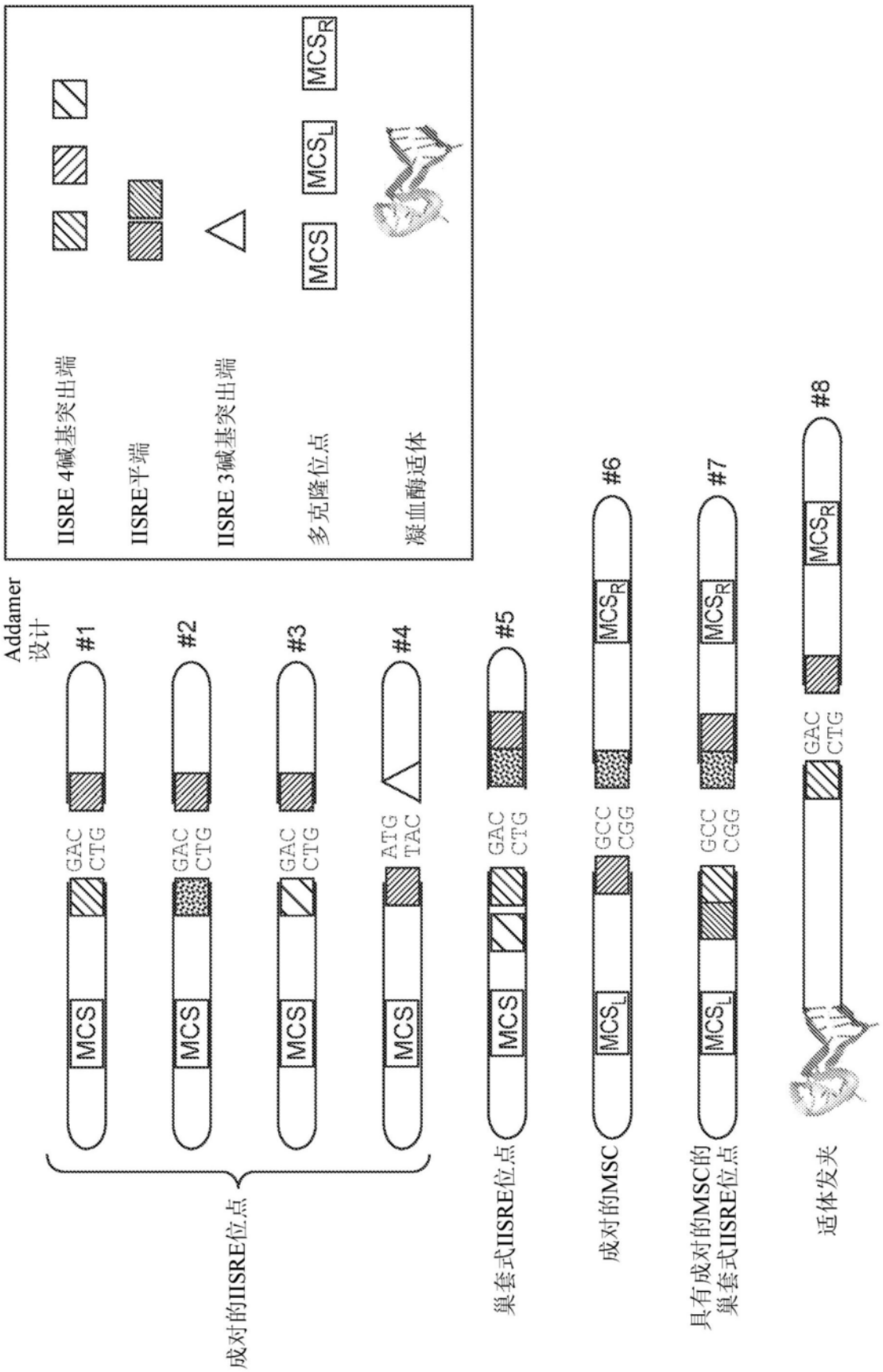

IISRE 4碱基突出端
IISRE平端
IISRE 3碱基突出端
多克隆位点
凝血酶适体
MCS
MCSL
MCSR
Addamer
设计
#1
#2
#3
#4
#5
#6
#7
#8
GAC
CTG
ATG
TAC
GCC
CGG
成对的IISRE位点
巢套式IISRE位点
成对的MSC
具有成对的MSC的
巢套式IISRE位点
适体发夹

图8

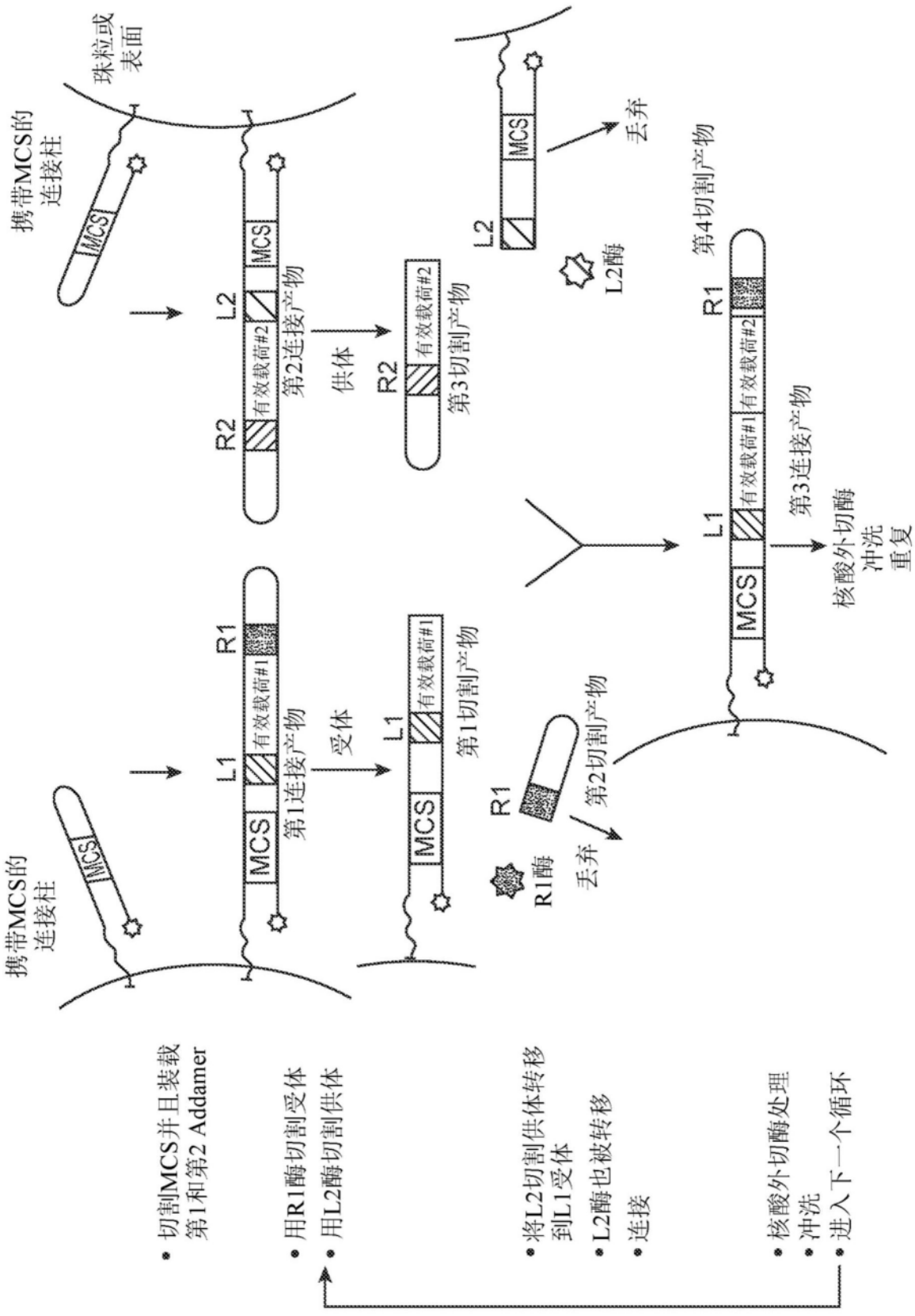

携带MCS的连接柱
MCS
珠粒或表面
L1
R1
有效载荷#1
第1连接产物
受体
第1切割产物
R1酶
丢弃
第2切割产物
L2
R2
有效载荷#2
第2连接产物
供体
第3切割产物
L2酶
第4切割产物
第3连接产物
核酸外切酶
冲洗
重复
• 切割MCS并且装载第1和第2 Addamer
• 用R1酶切割受体
• 用L2酶切割供体
• 将L2切割供体转移到L1受体
• L2酶也被转移
• 连接
• 核酸外切酶处理
• 冲洗
• 进入下一个循环

图9

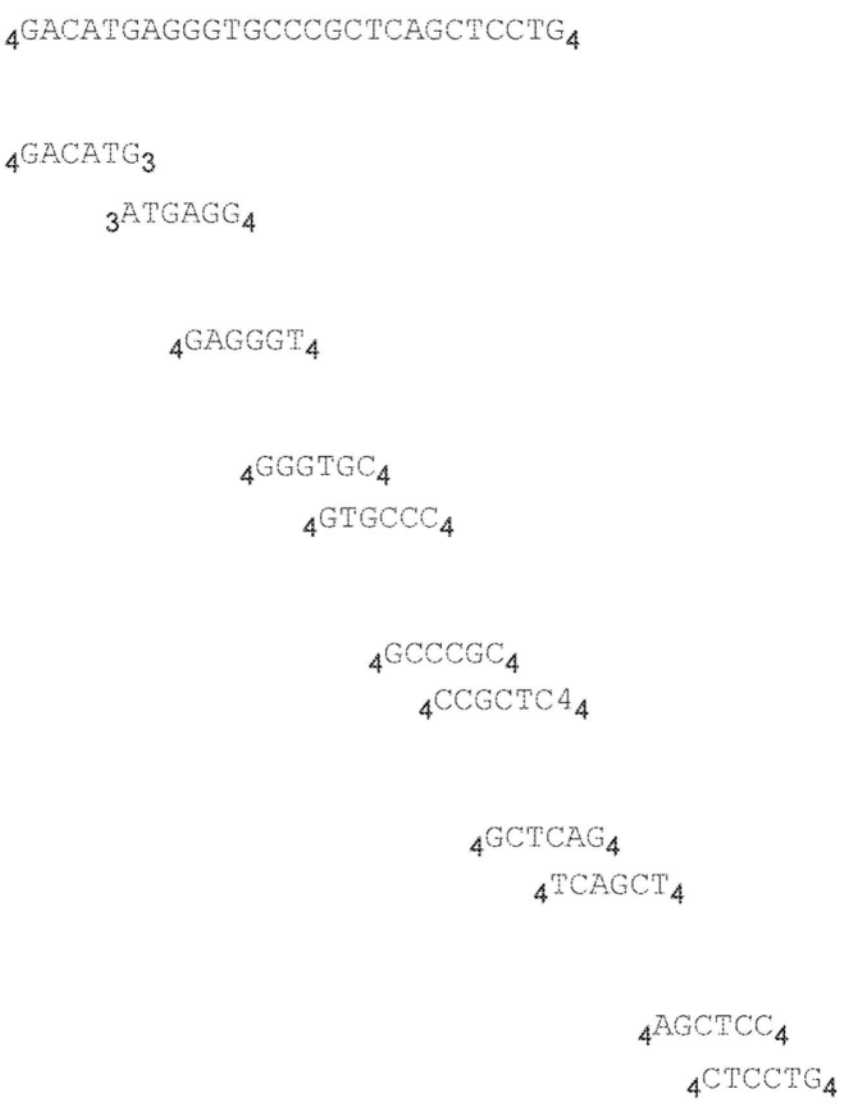

4GACATGAGGGTGCCCGCTCAGCTCCTG4
4GACATG3
3ATGAGG4
4GAGGGT4
4GGGTGC4
4GTGCCC4
4GCCCGC4
4CCGCTC44
4GCTCAG4
4TCAGCT4
4AGCTCC4
4CTCCTG4

图10A

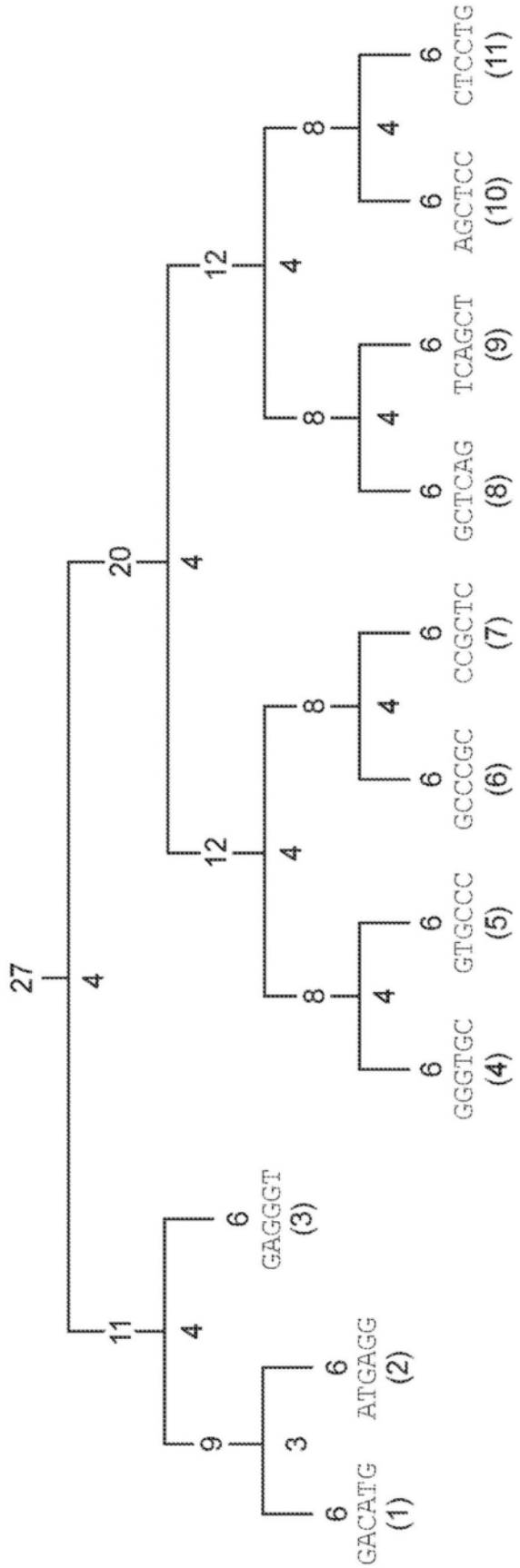

27
4
20
4
11
4
9
3
6
GACATG
(1)
6
ATGAGG
(2)
6
GAGGGT
(3)
12
4
12
4
8
4
6
GGGTGC
(4)
6
GTGCCC
(5)
8
4
6
GCCCGC
(6)
6
CCGCTC
(7)
8
4
6
GCTCAG
(8)
12
4
8
4
6
TCAGCT
(9)
6
AGCTCC
(10)
8
4
6
CTCCTG
(11)

图10B

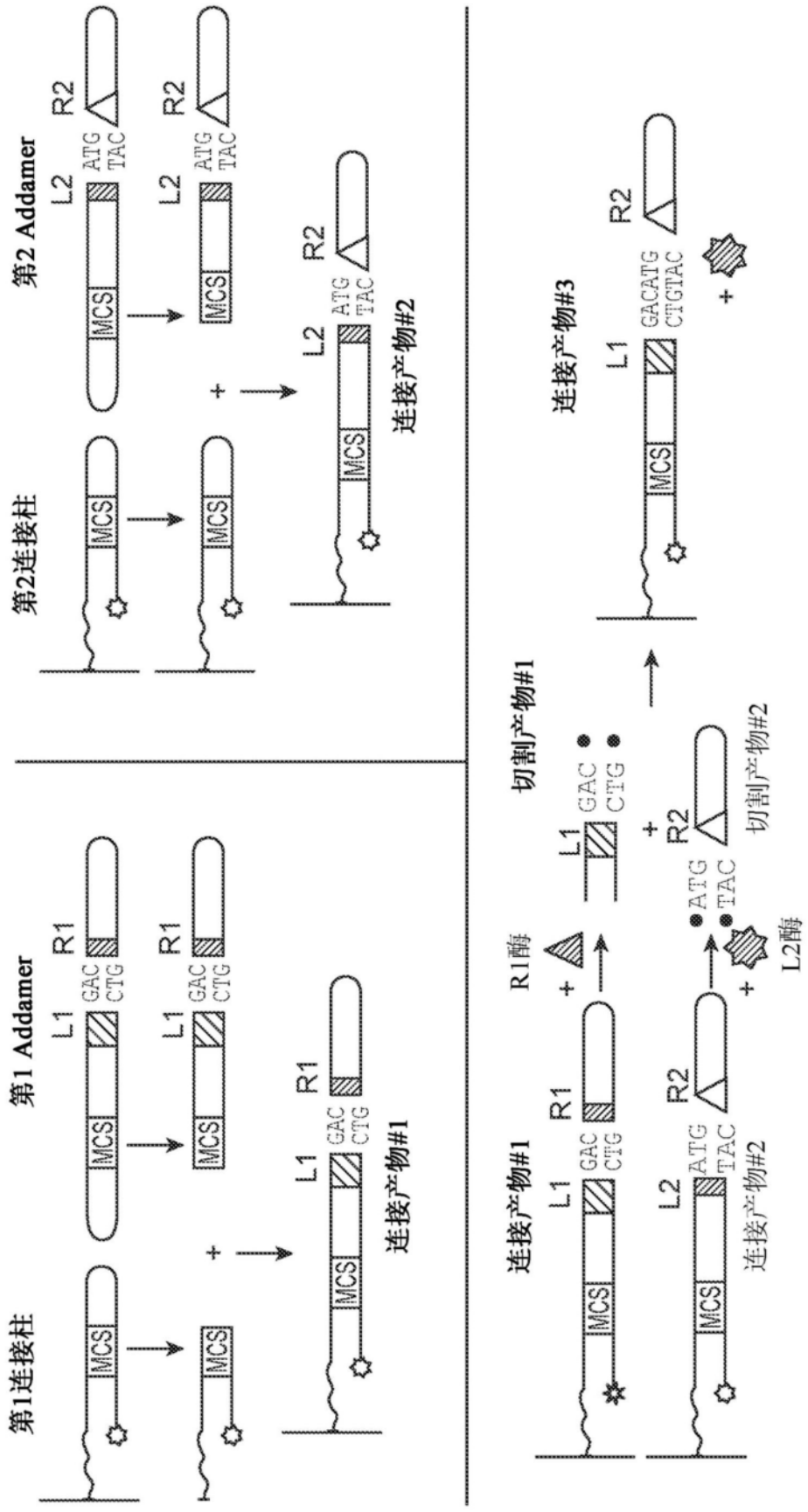

第1连接柱
第1 Addamer
第2连接柱
第2 Addamer
MCS
L1
L2
R1
R2
GAC
CTG
ATG
TAC
连接产物#1
连接产物#2
连接产物#3
切割产物#1
切割产物#2
R1酶
L2酶
GACATG
CTGTAC

图10C

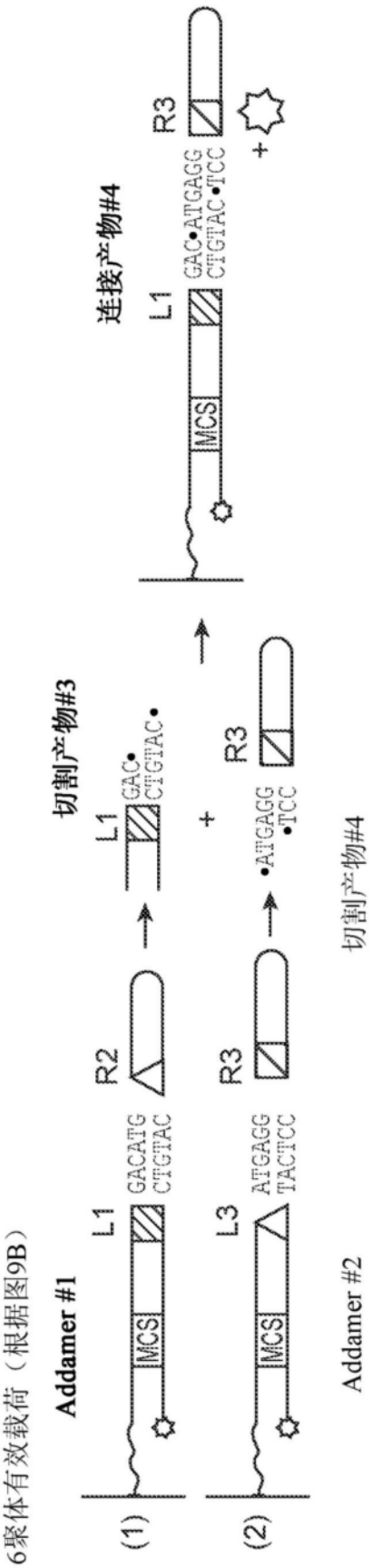

6聚体有效载荷（根据图9B）
Addamer #1
L1
MCS
GACATG
CTGTAC
(1)
L3
MCS
ATGAGG
TACTCC
(2)
R2
R3
切割产物#3
L1
GAC•
CTGTAC•
+
R3
•ATGAGG
•TCC
切割产物#4
Addamer #2
连接产物#4
L1
MCS
GAC•ATGAGG
CTGTAC•TCC
R3
+

图10D

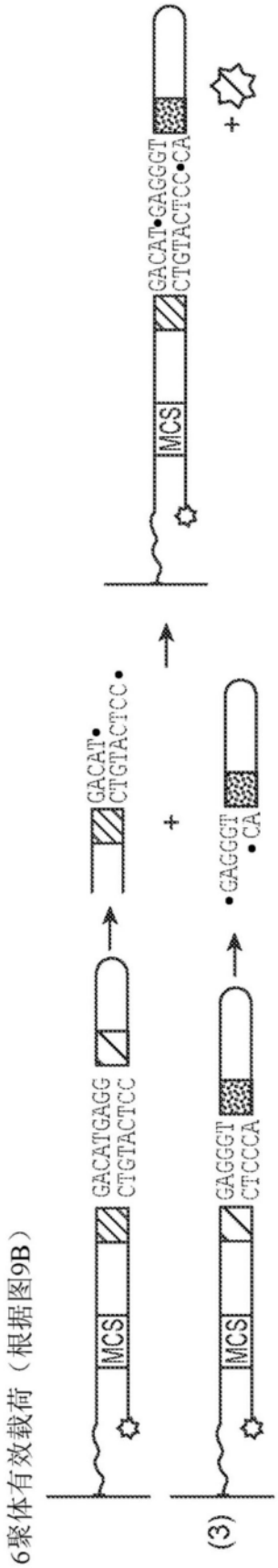

6聚体有效载荷（根据图9B）
MCS
GACATGAGG
CTGTACTCC
(3)
MCS
GAGGGT
CTCCCA
GACAT•
CTGTACTCC•
+
•GAGGGT
•CA
MCS
GACAT•GAGGGT
CTGTACTCC•CA
+

图10E

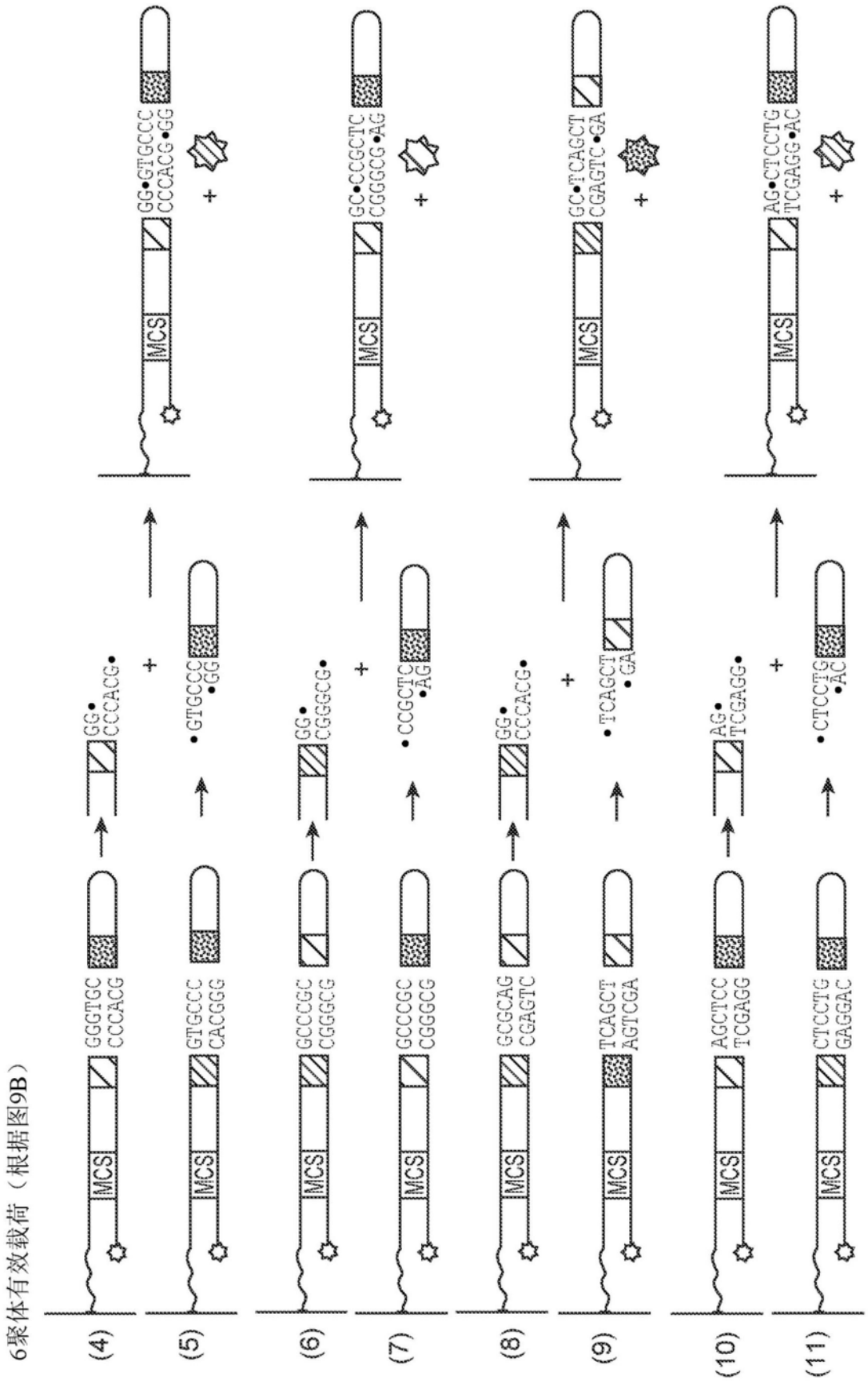

6聚体有效载荷（根据图9B）
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)
MCS
GGGTGC CCCACG
GG• •CCCACG
GTGCCC CACGGG
• GTGCCC •GG
GCCCGC CGGGCG
GG• CGGGCG•
• CCGCTC •AG
GCGCAG CGAGTC
GG• CCCACG•
TCAGCT AGTCGA
• TCAGCT • GA
AGCTCC TCGAGG
AG• TCGAGG•
CTCCTG GAGGAC
• CTCCTG •AC
GG•GTGCCC CCCACG•GG
GC•CCGCTC CGGGCG•AG
GC•TCAGCT CGAGTC•GA
AG•CTCCTG TCGAGG•AC

图10F

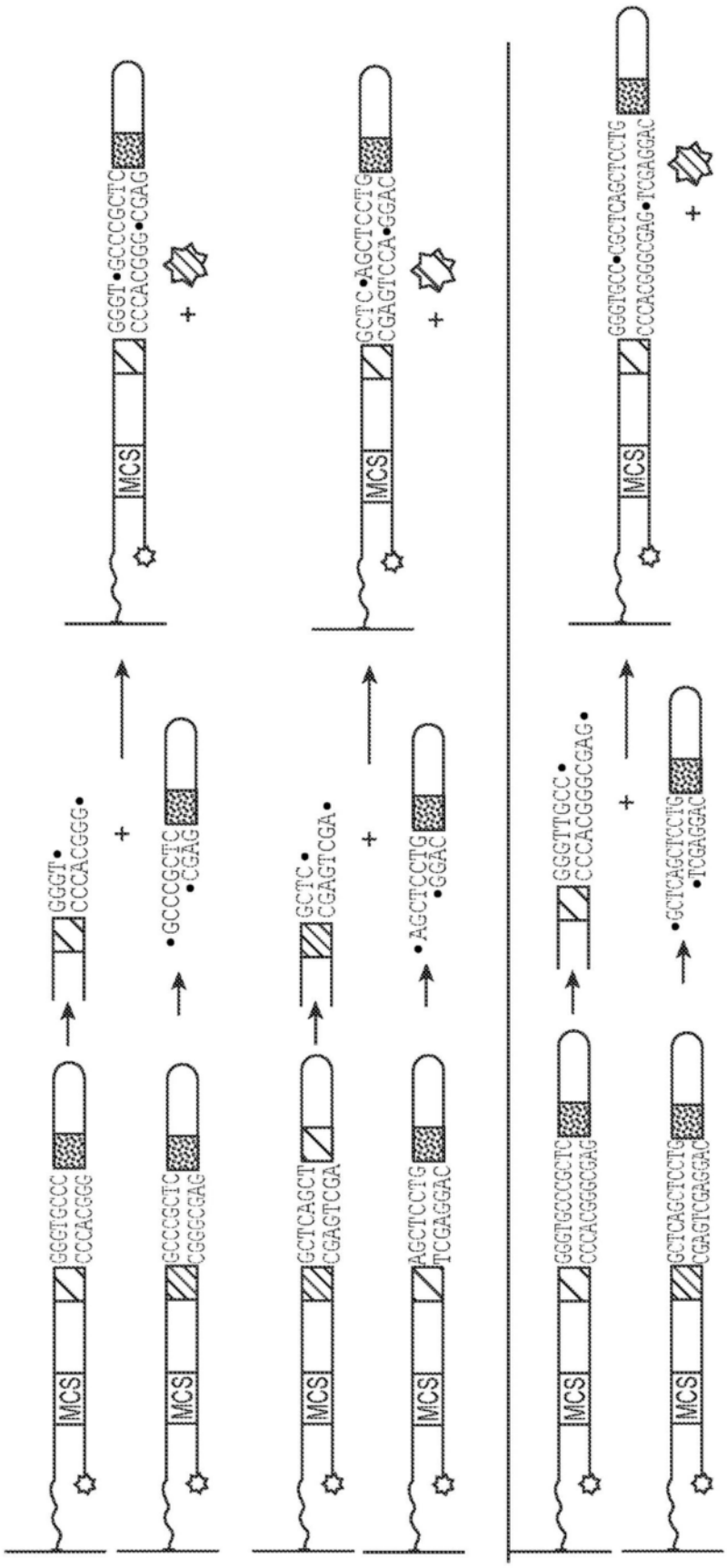

MCS
GGGTGCCC
CCCACGGG
GGGT•
CCCACGGG
GCCCGCTC
CGGGCGAG
•GCCCGCTC
•CGAG
GCTCAGCT
CGAGTCGA
GCTC•
CGAGTCGA
AGCTCCTG
TCGAGGAC
•AGCTCCTG
•GGAC
GGGTGCCGCTC
CCCACGGGCGAG
GGGTTGCC•
CCCACGGGCGAG•
GCTCAGCTCCTG
CGAGTCGAGGAC
•GCTCAGCTCCTG
•TCGAGGAC
GGGT•GCCCGCTC
CCCACGGG•CGAG
GCTC•AGCTCCTG
CGAGTCCA•GGAC
GGGTGCC•CGCTCAGCTCCTG
CCCACGGGCGAG•TCGAGGAC
+

图10G

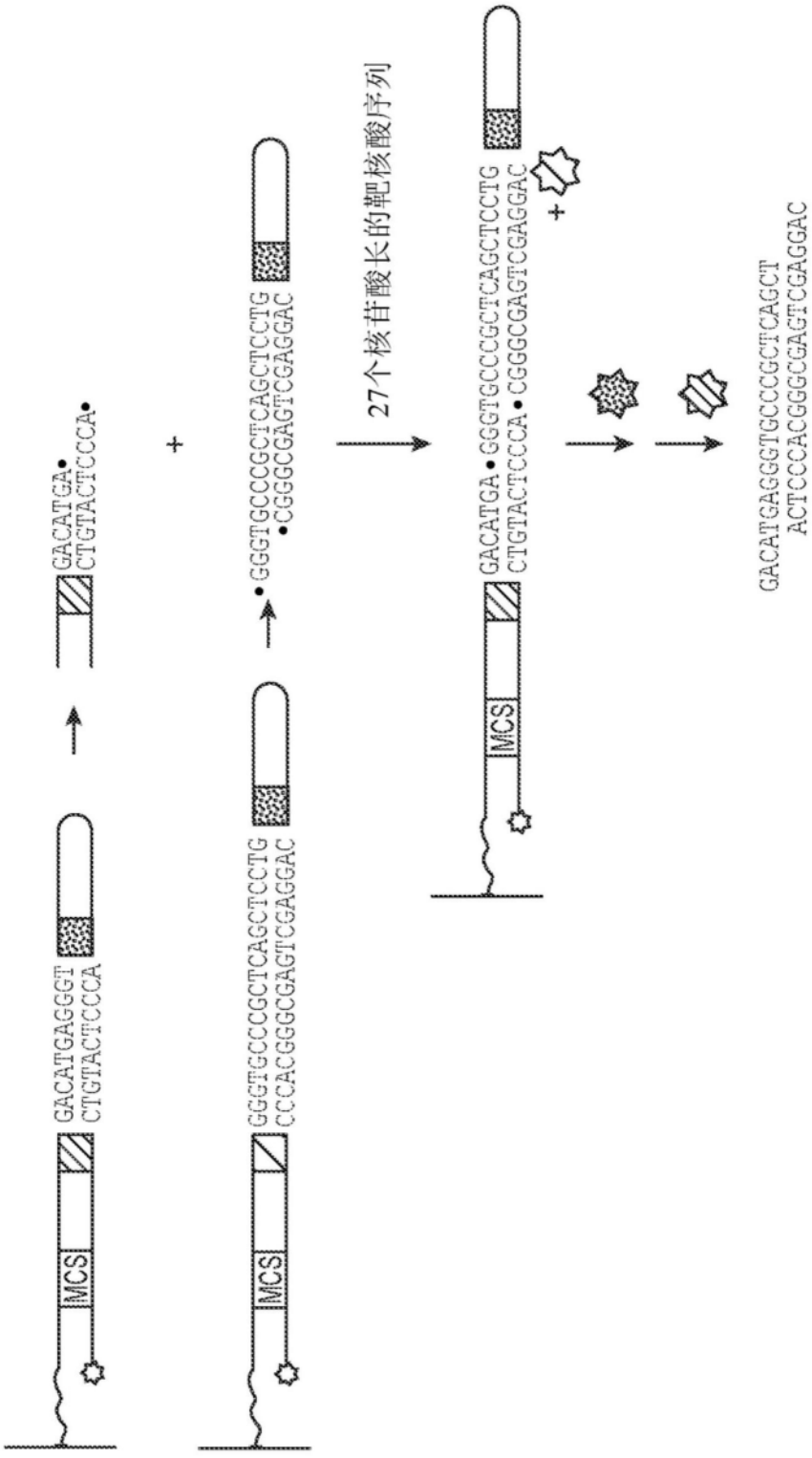

MCS
GACATGAGGGT
CTGTACTCCCA
GACATGA
CTGTACTCCCA
MCS
GGGTGCCCGCTCAGCTCCTG
CCCACGGGCGAGTCGAGGAC
GGGTGCCCGCTCAGCTCCTG
CGGGCGAGTCGAGGAC
27个核苷酸长的靶核酸序列
MCS
GACATGA
CTGTACTCCCA
GGGTGCCCGCTCAGCTCCTG
CGGGCGAGTCGAGGAC
GACATGAGGGTGCCCGCTCAGCT
ACTCCCACGGGCGAGTCGAGGAC

图10H

>5050Seq03
CGACGTGTCTATCGAATACGTTGTCTGCCAACATCTGTCGCAACCAATGTAGGCCGGC
AATGTTAAGCTGTCGACGATGCAAGTCAATAAGCGAACGAGTGCTCCGTCTCTCAA
CCTCATAGTGACTAGGAACGCGTAATGCATGAACAATGCCGTCAACCGTCAGGTGATT
AGGCGTGTCGATTCTCCGTCCTTGTACAAGACGTGGTAGAACCGGCAATGTTCGACTC
TGTCGAAGAACGCTGTCAGAGTCAATCAACATCTGCTCGTGTAGTTCTGTGCAGGTAT
GACGGATTCC

图11A

寡聚物对（片段）

片段1：

GTGTCTATCG
GCTGCACAGA

片段2：

AATACGTTGTCTGCCAACATCTGT
TAGCTTATGCAACAGACGGTTGTA

片段3：

CGCAACCAATGTAGGCCGGC
GACAGCGTTGGTTACATCCG

片段4：

AATGTTAAGCTGTCGACGATGCAAGTC
GCCGTTACAATTCGACAGCTGCTACGT

片段5：

AATAAGCGAACGAGTGCTCCGT
TCAGTTATTCGCTTGCTCACGA

片段6：

CTCTCTCAACCTCATAGTGACTAGGAA
GGCAGAGAGAGTTGGAGTATCACTGAT

片段7：

CGCGTAATGCATGAACAATGC
CCTTGCGCATTACGTACTTGT

片段8：

CGTCAACCGTCAGGTGATTAGGC
TACGGCAGTTGGCAGTCCACTAA

片段9：

GTGTCGATTCTCCGTCCTT
TCCGCACAGCTAAGAGGCA

片段10：

GTACAAGACGTGGTAGAACCGGCA
GGAACATGTTCTGCACCATCTTGG

片段11：

ATGTTCGACTCTGTCGAAGAAC
CCGTTACAAGCTGAGACAGCTT

片段12：

GCTGTCAGAGTCAATCAAC
CTTGCGACAGTCTCAGTTA

片段13：

ATCTGCTCGTGTAGTTCTGTG
GTTGTAGACGAGCACATCAAG

片段14：

CAGGTATGACGGATTCC
ACACGTCCATACTGCCT

图11B

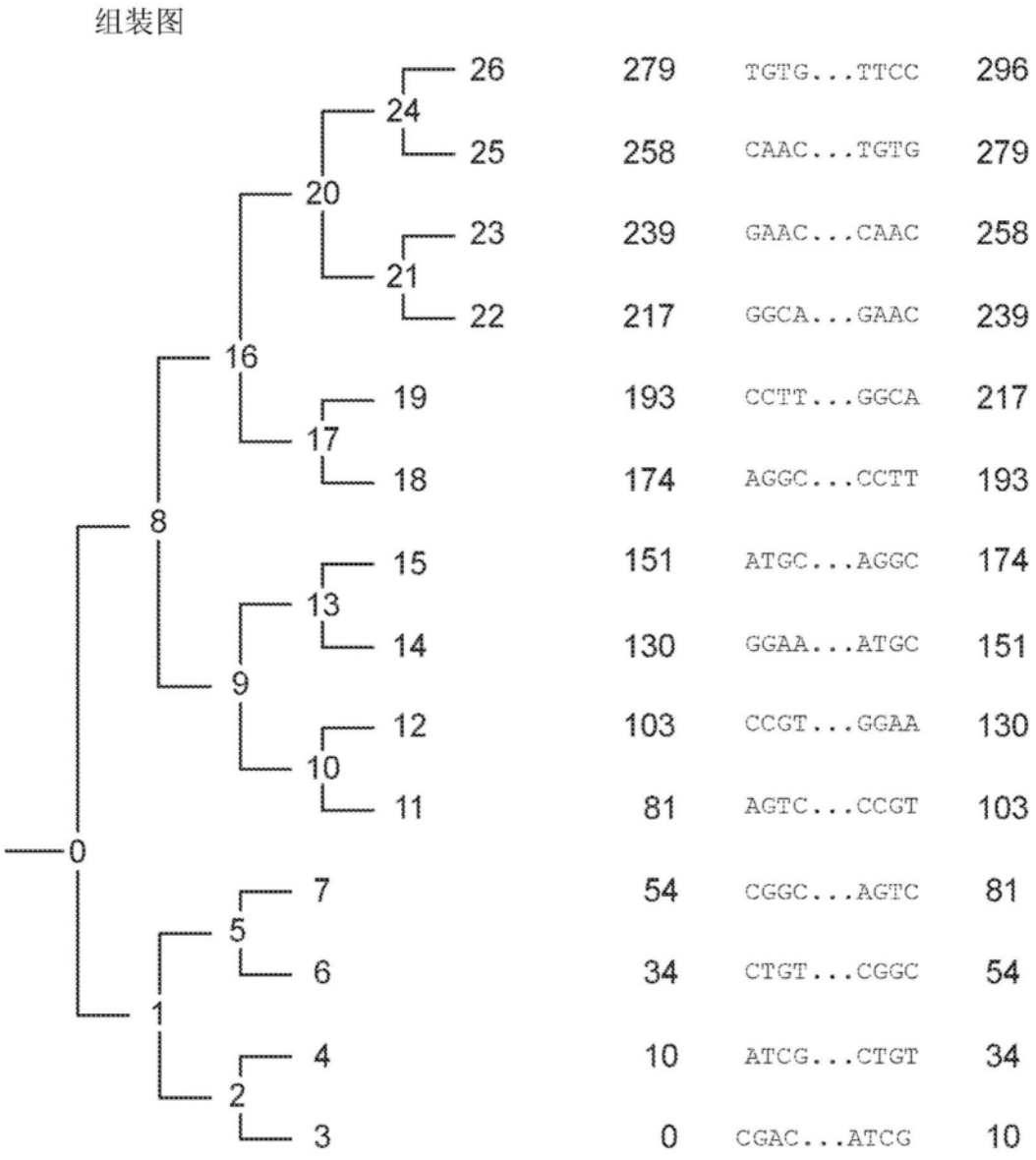

组装图
26 279 TGTG...TTCC 296
24
25 258 CAAC...TGTG 279
20
23 239 GAAC...CAAC 258
21
22 217 GGCA...GAAC 239
16
19 193 CCTT...GGCA 217
17
18 174 AGGC...CCTT 193
8
15 151 ATGC...AGGC 174
13
14 130 GGAA...ATGC 151
9
12 103 CCGT...GGAA 130
10
11 81 AGTC...CCGT 103
0
7 54 CGGC...AGTC 81
5
6 34 CTGT...CGGC 54
1
4 10 ATCG...CTGT 34
2
3 0 CGAC...ATCG 10

图11C

第1轮连接：

片段1：

GTGTCTATCG
GCTGCACAGA

片段2：

AATACGTTGTCTGCCAACATCTGT
TAGCTTATGCAACAGACGGTTGTA

片段3：

CGCAACCAATGTAGGCCGGC
GACAGCGTTGGTTACATCCG

片段4：

AATGTTAAGCTGTCGACGATGCAAGTC
GCCGTTACAATTCGACAGCTGCTACGT

片段5：

AATAAGCGAACGAGTGCTCCGT
TCAGTTATTCGCTTGCTCACGA

片段6：

CTCTCTCAACCTCATAGTGACTAGGAA
GGCAGAGAGAGTTGGAGTATCACTGAT

片段7：

CGCGTAATGCATGAACAATGC
CCTTGCGCATTACGTACTTGT

片段8：

CGTCAACCGTCAGGTGATTAGGC
TACGGCAGTTGGCAGTCCACTAA

片段9：

GTGTCGATTCTCCGTCCTT
TCCGCACAGCTAAGAGGCA

片段10：

GTACAAGACGTGGTAGAACCGGCA
GGAACATGTTCTGCACCATCTTGG

片段11：

ATGTTCGACTCTGTCGAAGAAC
CCGTTACAAGCTGAGACAGCTT

片段12：

GCTGTCAGAGTCAATCAAC
CTTGCGACAGTCTCAGTTA

片段13：

ATCTGCTCGTGTAGTTCTGTG
GTTGTAGACGAGCACATCAAG

片段14：

CAGGTATGACGGATTCC
ACACGTCCATACTGCCT

图11D

第2轮连接：

片段1+2：

GTGTCTATCGAATACGTTGTCTGCCAACATCTGT
GCTGCACAGATAGCTTATGCAACAGACGGTTGTA

片段3+4：

CGCAACCAATGTAGGCCGGCAATGTTAAGCTGTCGACGATGCAAGTC
GACAGCGTTGGTTACATCCGGCCGTTACAATTCGACAGCTGCTACGT

片段5+6：

AATAAGCGAACGAGTGCTCCGTCTCTCTCAACCTCATAGTGACTAGGAA
TCAGTTATTCGCTTGCTCACGAGGCAGAGAGAGTTGGAGTATCACTGAT

片段7+8：

CGCGTAATGCATGAACAATGCCGTCAACCGTCAGGTGATTAGGC
CCTTGCGCATTACGTACTTGTTACGGCAGTTGGCAGTCCACTAA

片段9+10：

GTGTCGATTCTCCGTCCTTGTACAAGACGTGGTAGAACCGGCA
TCCGCACAGCTAAGAGGCAGGAACATGTTCTGCACCATCTTGG

片段11+12：

ATGTTCGACTCTGTCGAAGAACGCTGTCAGAGTCAATCAAC
CCGTTACAAGCTGAGACAGCTTCTTGCGACAGTCTCAGTTA

片段13+14：

ATCTGCTCGTGTAGTTCTGTGCAGGTATGACGGATTCC
GTTGTAGACGAGGCACATCAAGACACGTCCATACTGCCT

图11E

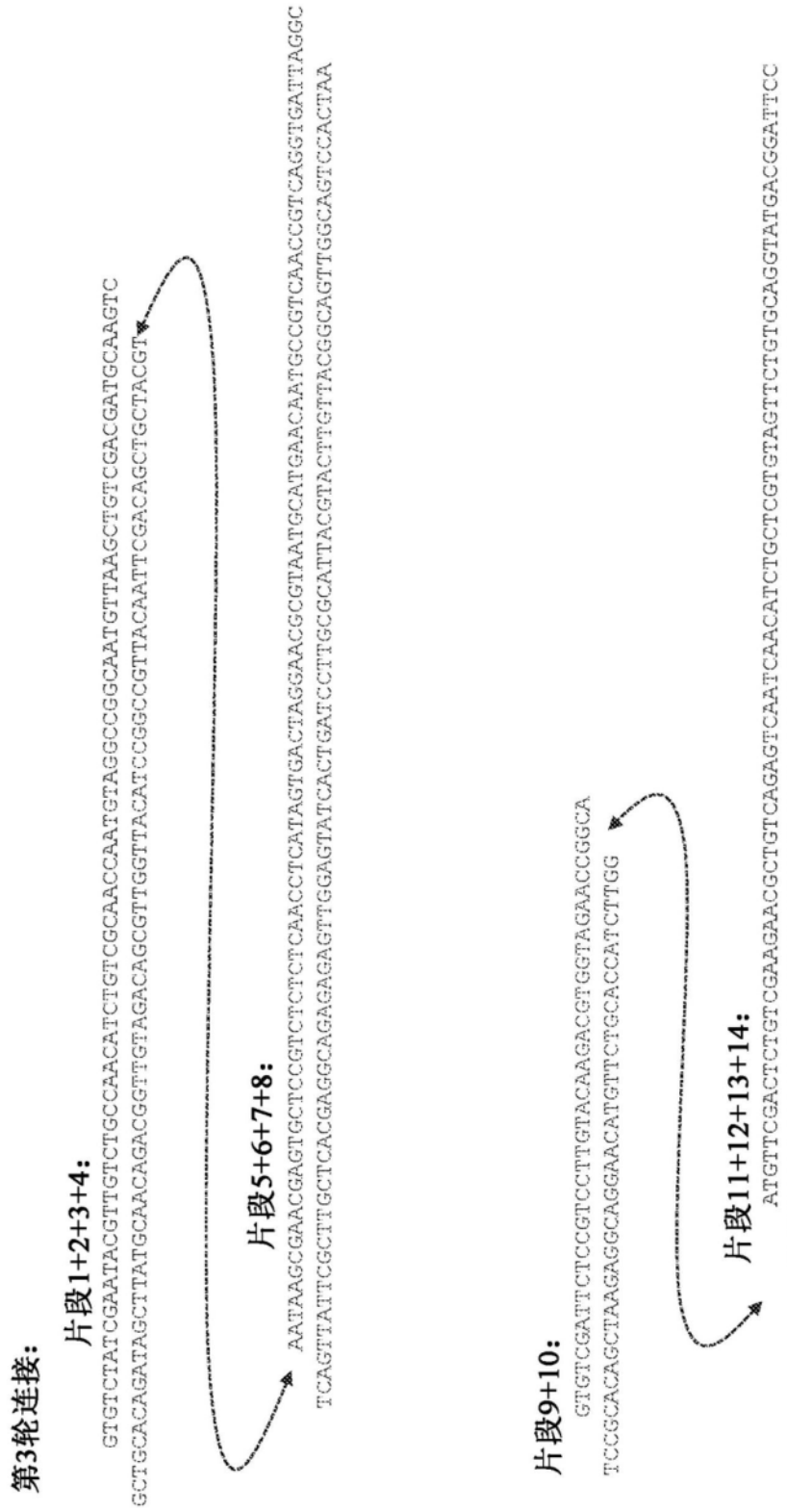

第3轮连接:
片段1+2+3+4:
GTGTCTATCGAATACGTTGTCTGCCAACAACATCTGTCGCAACCAATGTAGGCCGGCAATGTTAAGCTGTCGACGATGCAAGTC
GCTGCACAGATAGCTTATGCAACAGACGGTTGTAGACAGCGTTGGTTACATCCGGCCGTTACAATTCGACAGCTGCTACGT
片段5+6+7+8:
AATAAGCGAACGAGTGCTCCGTCTCTCTCAACCTCATAGTGACTAGGAACGCGTAATGCATGAACAATGCCGTCAACCGTCAGGTGATTAGGC
TCAGTTATTCGCTTGCTCACGAGGCAGAGAGAGTTGGAGTATCACTGATCCTTGCGCATTACGTACTTGTTACGGCAGTTGGCAGTCCACTAA
片段9+10:
GTGTCGAATTCTCCGTCCTTGTACAAGACGTGGTAGAACCGGCA
TCCGCACAGCTAAGAGGCAGGAACATGTTCTGCACCATCTTGG
片段11+12+13+14:
ATGTTCGACTCTGTCGAAGAACGCTGTCAGAGTCAATCAACATCTGCTCGTGTAGTTCTGTGCAGGTATGACGGATTCC
CCGTTACAAGCTGAGACAGCTTCTTGCGACAGTCTCAGTTAGTTGTAGACGAGCACATCAAGACACGTCCATACTGCCT

图11F

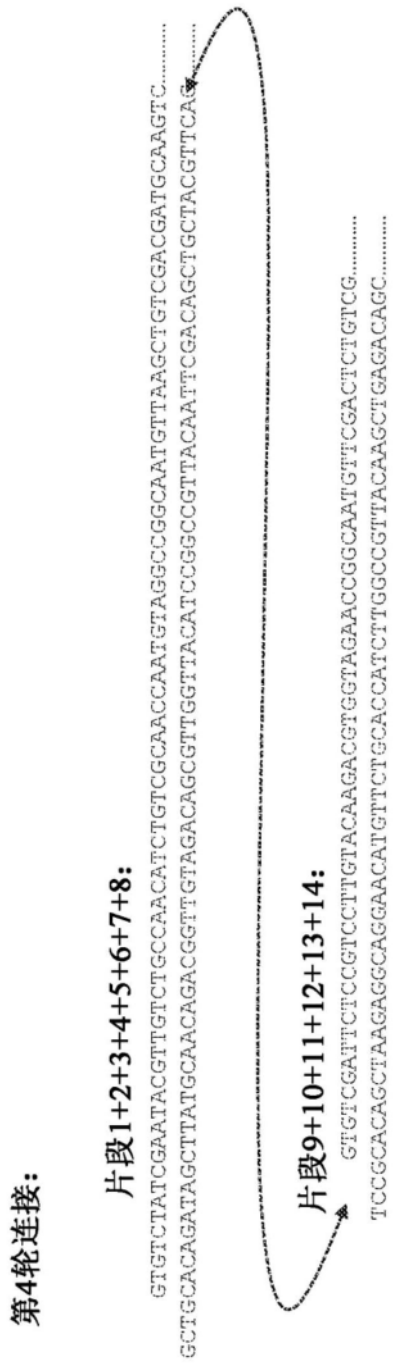

第4轮连接：
片段1+2+3+4+5+6+7+8：
GTGTCTATCGAATACGTTGTCTGCCAACACATCTGTCGCAACCAATGTAGGCCGGCAATGTTAAGCTGTCGACGATGCAAGTC.....
GCTGCACAGATAGCTTATGCAACAGACGGTTGTAGACAGCGTTGGTTACATCCGGCCGTTACAATTCGACAGCTGCTACGTTCAG.....
片段9+10+11+12+13+14：
GTGTCGATTCTCCGTCCTTGTACAAGACGTGGTAGAACCGGCAATGTTCGACTCTGTCG.....
TCCGCACAGCTAAGAGGCAGGAACATGTTCTGCACCATCTTGGCCGTTACAAGCTGAGACAGC.....

图11G

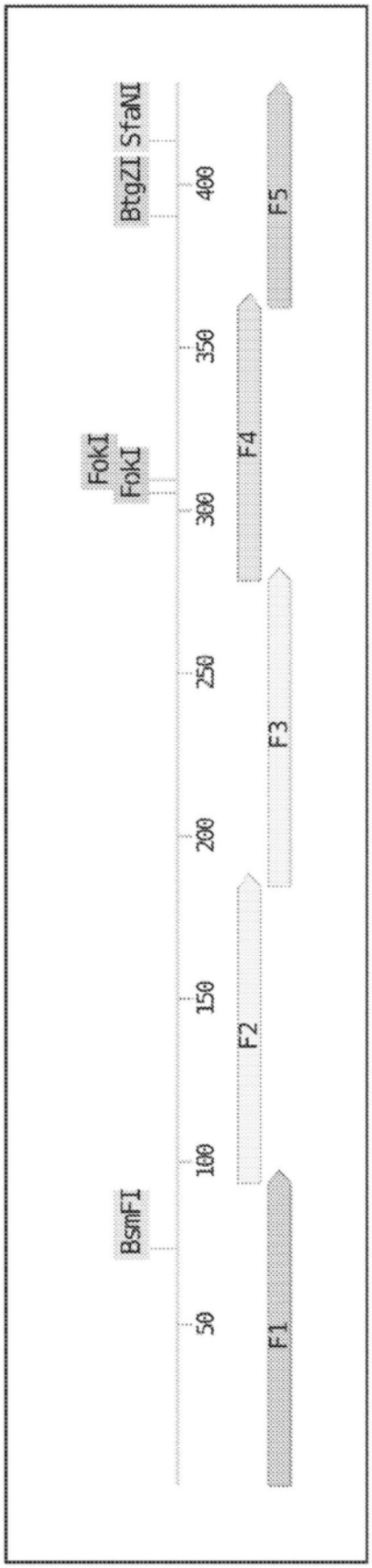

BtgZI
SfaNI
FokI
FokI
BsmFI
F1
F2
F3
F4
F5
50
100
150
200
250
300
350
400

图12

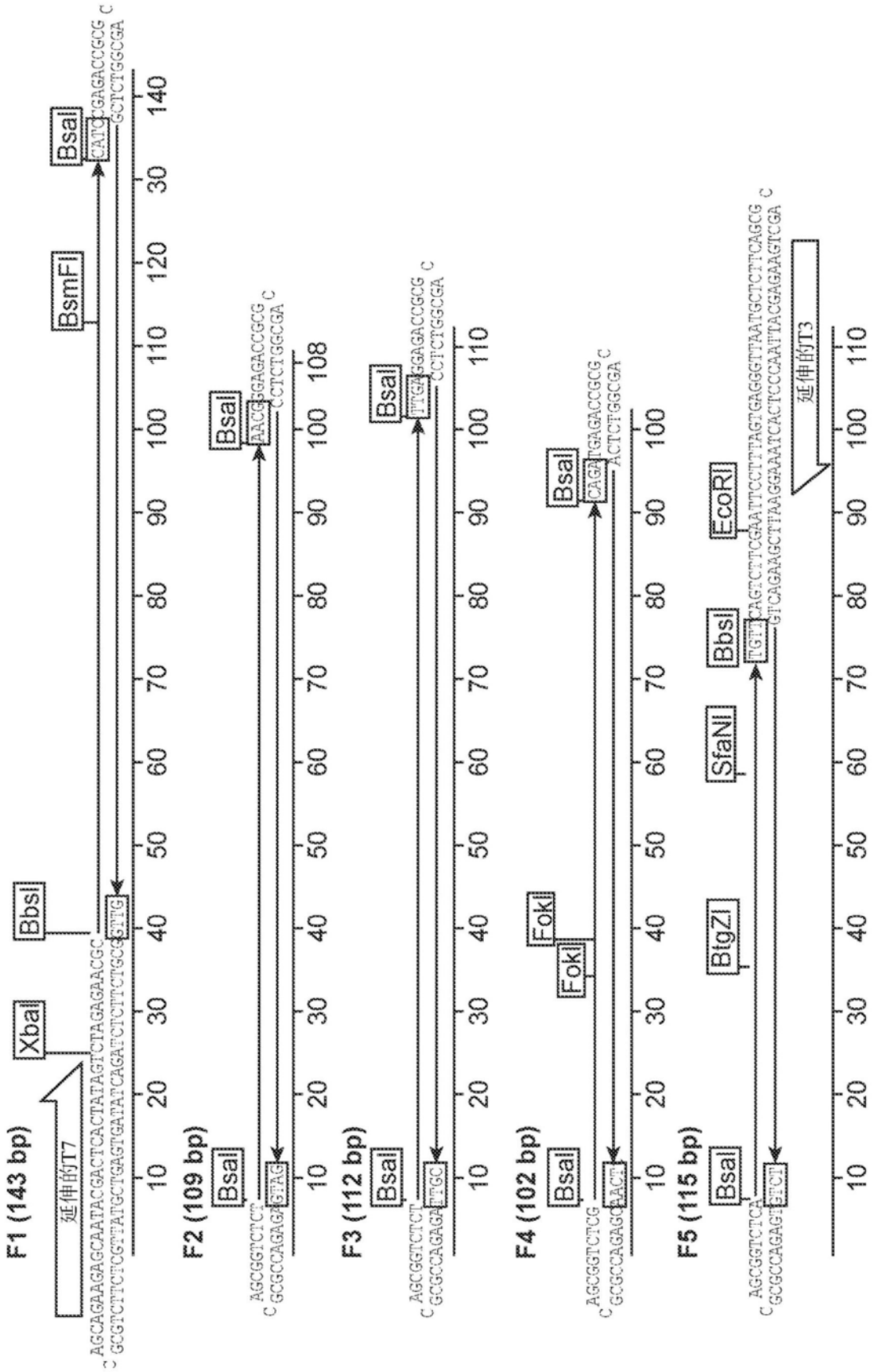

F1 (143 bp)
延伸的T7
XbaI
BbsI
BsmFI
BsaI
F2 (109 bp)
BsaI
F3 (112 bp)
BsaI
F4 (102 bp)
BsaI
FokI
FokI
BtgZI
SfaNI
BbsI
EcoRI
F5 (115 bp)
BsaI
延伸的T3

图13

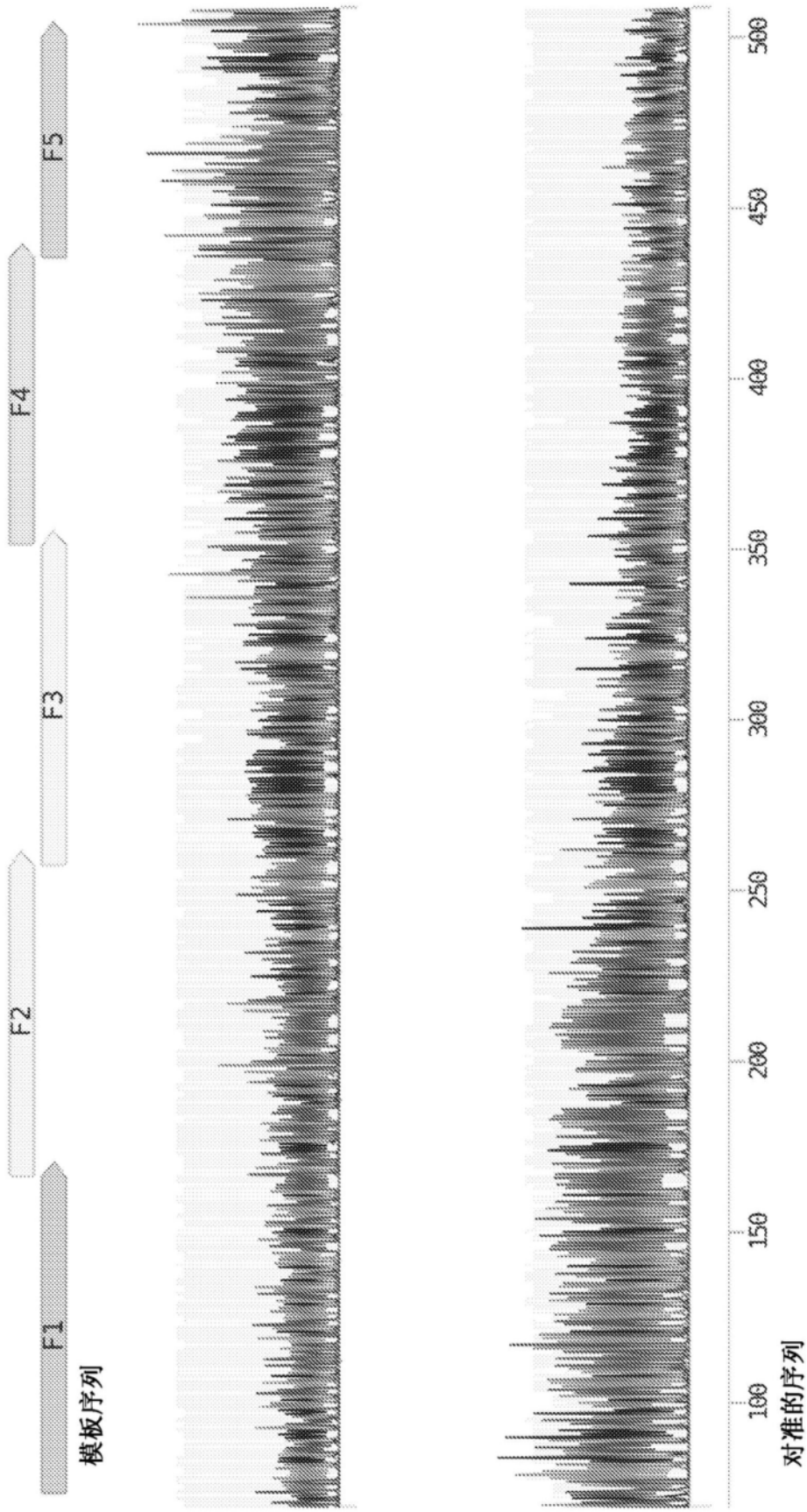
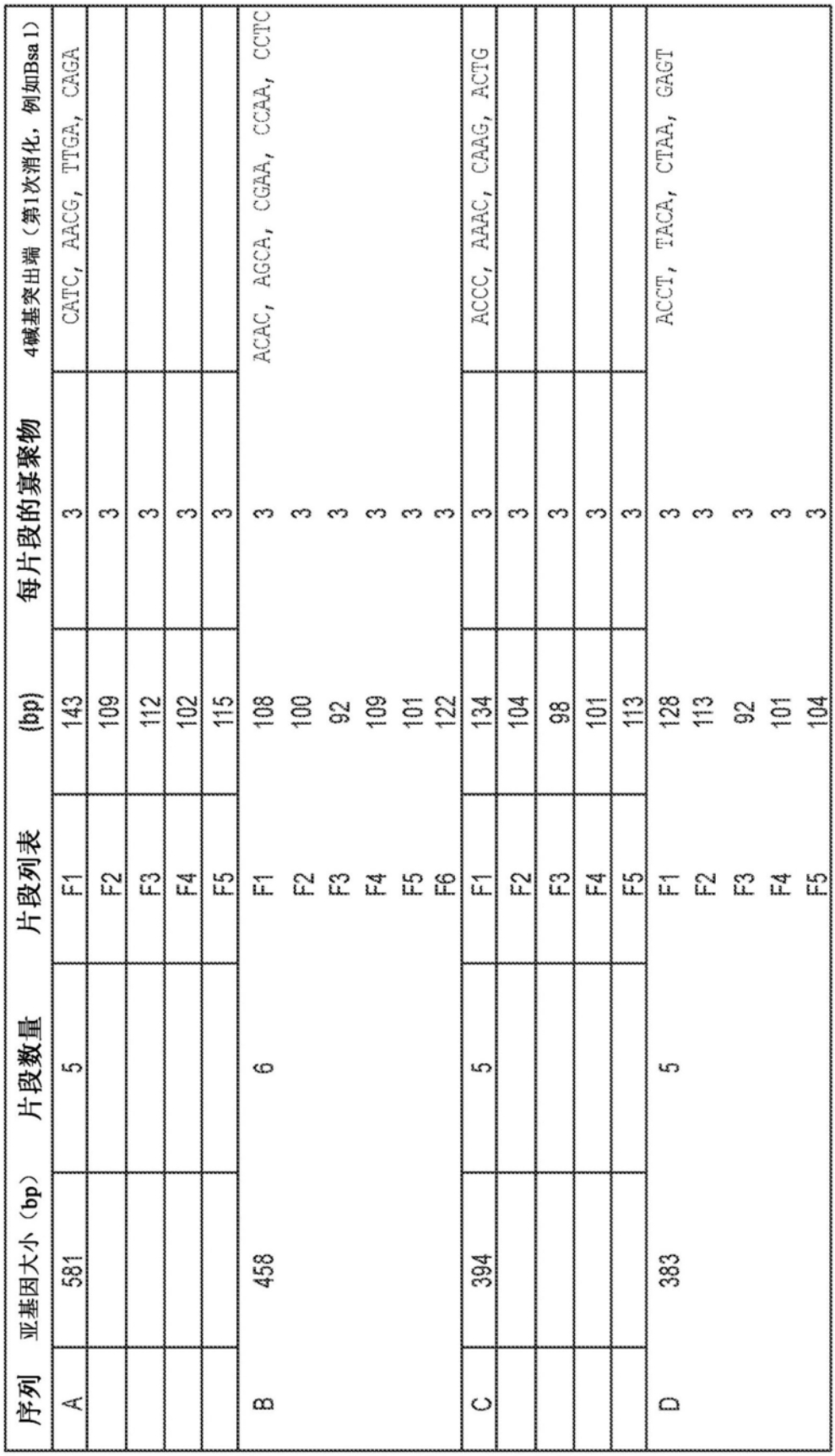
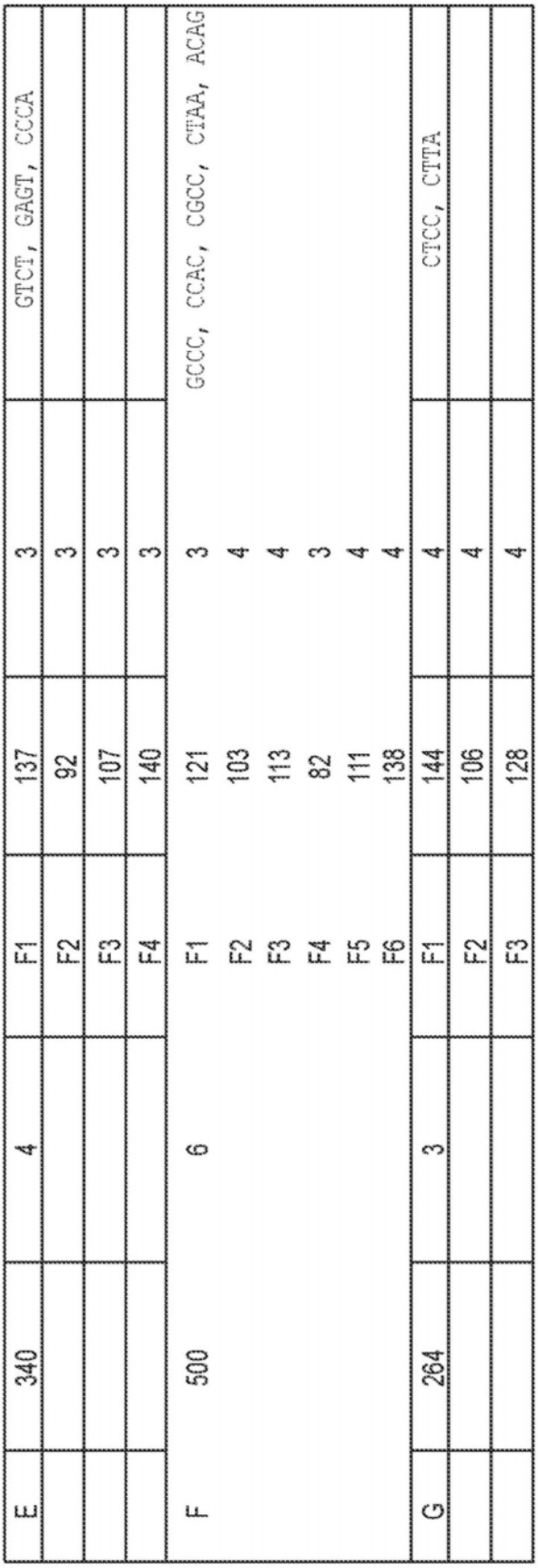


图14



序列	亚基因大小（bp）	片段数量	片段列表	(bp)	每片段的寡聚物	4碱基突出端（第1次消化，例如Bsa I）
A	581	5	F1	143	3	CATC, AACG, TTGA, CAGA
			F2	109	3	
			F3	112	3	
			F4	102	3	
			F5	115	3	
B	458	6	F1	108	3	ACAC, AGCA, CGAA, CCAA, CCTC
			F2	100	3	
			F3	92	3	
			F4	109	3	
			F5	101	3	
			F6	122	3	
C	394	5	F1	134	3	ACCC, AAAC, CAAG, ACTG
			F2	104	3	
			F3	98	3	
			F4	101	3	
			F5	113	3	
D	383	5	F1	128	3	ACCT, TACA, CTAA, GAGT
			F2	113	3	
			F3	92	3	
			F4	101	3	
			F5	104	3	

图15



E	340	4	F1	137	3	GTCT, GAGT, CCCA
			F2	92	3	
			F3	107	3	
			F4	140	3	
F	500	6	F1	121	3	GCCC, CCAC, CGCC, CTAA, ACAG
			F2	103	4	
			F3	113	4	
			F4	82	3	
			F5	111	4	
			F6	138	4	
G	264	3	F1	144	4	CTCC, CTTA
			F2	106	4	
			F3	128	4	

图15(续1)