



(12) 发明专利申请

(10) 申请公布号 CN 117716030 A

(43) 申请公布日 2024.03.15

(21) 申请号 202280053115.X

(22) 申请日 2022.08.02

(30) 优先权数据

63/228,448 2021.08.02 US

(85) PCT国际申请进入国家阶段日

2024.01.29

(86) PCT国际申请的申请数据

PCT/US2022/039193 2022.08.02

(87) PCT国际申请的公布数据

W02023/014728 EN 2023.02.09

(71) 申请人 安捷伦科技有限公司

地址 美国加利福尼亚州

(72) 发明人 D·陶西格 N·M·桑帕斯

(74) 专利代理机构 北京坤瑞律师事务所 11494

专利代理师 封新琴

(51) Int.Cl.

C12N 15/10 (2006.01)

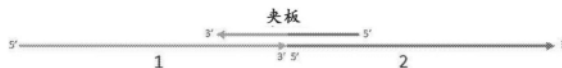
权利要求书3页 说明书26页 附图13页

(54) 发明名称

使用索引夹板寡核苷酸通过连接组装长核酸

(57) 摘要

本公开文本涉及一种通过酶促连接与索引夹板寡核苷酸分子杂交的寡核苷酸分子来组装长核酸的方法。还公开了可用于进行所公开的方法的包含索引夹板寡核苷酸的寡核苷酸结构。本发明的试剂盒包括包含所公开的寡核苷酸结构的核酸文库。寡核苷酸以高度特异性的顺序连接以构建所希望的选择的连接产物。



1. 一种用于组装两个核酸组分以产生多组分连接产物的方法,所述方法包括:

(A) 使包含第一DNA序列组分的第一单链寡核苷酸和包含第二DNA序列组分的第二单链寡核苷酸与索引夹板杂交,其中所述第一单链寡核苷酸的5'末端和3'末端两端均与所述索引夹板杂交,并且所述第二单链寡核苷酸的5'末端和3'末端两端均与所述索引夹板杂交;以及

(B) 将所述第一单链寡核苷酸的3'末端与所述第二单链寡核苷酸的5'末端连接,由此产生第一多组分连接产物。

2. 根据权利要求1所述的方法,其中:

(a) 所述索引夹板进一步包括夹板序列,以及在所述夹板序列侧翼的第一索引序列和第二索引序列;

(b) 所述夹板序列与所述第一单链寡核苷酸的3'末端和所述第二单链寡核苷酸的5'末端二者杂交;以及

(c) 所述第一索引序列与所述第一单链寡核苷酸的5'末端杂交以形成第一双链索引元件,并且所述第二索引序列与所述第二单链寡核苷酸的3'末端杂交以形成第二双链索引元件。

3. 根据权利要求2所述的方法,其中所述索引夹板的夹板序列含有:

(a) 5'元件,其与所述第二单链寡核苷酸的5'末端杂交形成双链5'夹板元件,以及

(b) 3'元件,其与所述第一单链寡核苷酸的3'末端杂交形成双链3'夹板元件,

其中所述夹板序列的5'元件和3'元件是连续的,即没有插入核苷酸。

4. 根据权利要求2或3所述的方法,其中:

(1) 所述第一索引序列位于所述索引夹板的3'末端,以及

(2) 所述第二索引序列位于所述索引夹板的5'末端。

5. 根据权利要求4所述的方法,其中:

(a) 所述第一单链寡核苷酸从5'至3'包含:

(1) 构成所述第一单链寡核苷酸的5'末端序列的第三索引序列,和

(2) 待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸的3'末端序列;以及

(b) 所述第二单链寡核苷酸从5'至3'包含:

(1) 待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和

(2) 构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及

(c) 所述索引夹板从5'至3'包含:

(1) 5'末端序列,所述5'末端序列是所述第二索引序列,

(2) 任选的第一间隔序列,

(3) 所述夹板序列,

其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交,并且其中所述夹板序列的3'元件与所述第一单链寡核苷酸的3'末端序列杂交,

(4) 任选的第二间隔序列,以及

(5) 3'末端序列,所述3'末端序列是所述第一索引序列。

6. 根据权利要求5所述的方法,其中每个间隔序列独立地由1至30个核苷酸组成。
7. 根据权利要求2或3所述的方法,其中:
  - (1) 所述第一索引序列位于所述索引夹板的5'末端,以及
  - (2) 所述第二索引序列位于所述索引夹板的3'末端。
8. 根据权利要求7所述的方法,其中:
  - (a) 所述第一单链寡核苷酸从5'至3'包含:
    - (1) 构成所述第一单链寡核苷酸的5'末端序列的第三索引序列,和
    - (2) 待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸的3'末端序列;以及
  - (b) 所述第二单链寡核苷酸从5'至3'包含:
    - (1) 待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和
    - (2) 构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及
  - (c) 所述索引夹板从5'至3'包含:
    - (1) 5'末端序列,所述5'末端序列是所述第一索引序列,
    - (2) 任选的第一间隔序列,
    - (3) 所述夹板序列,其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交,并且其中所述夹板序列的3'元件与所述第一单链寡核苷酸的3'末端序列杂交,
  - (4) 任选的第二间隔序列,以及
  - (5) 3'末端序列,所述3'末端序列是所述第二索引序列。
9. 根据权利要求7或8所述的方法,其中每个间隔序列独立地由1至30个核苷酸组成。
10. 根据权利要求2至9中任一项所述的方法,其中每个双链索引元件包含10至40个碱基对。
11. 根据权利要求1至10中任一项所述的方法,所述方法进一步包括:
  - (C) 将来自所述多组分连接产物的索引夹板变性,并且随后迭代地进行步骤(A)、(B)和(C)。
12. 根据权利要求5至11中任一项所述的方法,其中将所述第四索引序列可去除地连接至所述第二DNA序列组分。
13. 根据权利要求12所述的方法,其中所述第二单链寡核苷酸进一步包含在第二DNA序列组分的3'处的IIIs型限制性酶位点,其中所述限制性酶切割位点紧邻所述第二DNA序列组分的3'。
14. 根据权利要求12所述的方法,其中所述第四索引序列通过可切割接头连接至所述第二DNA序列组分,其中切割位点紧邻所述第二DNA序列组分的3'。
15. 根据权利要求12至14中任一项所述的方法,所述方法进一步包括:
  - 从所述第一连接产物中的第二DNA组分上去除所述第四索引序列,
  - 提供具有所述第四索引序列的另一个第二单链寡核苷酸和包含第二夹板序列的另一个索引夹板,
  - 随后进行步骤(A)、(B)和(C),

其中所述第一连接产物是步骤(A)的第一单链寡核苷酸,并且

其中所述另一个索引夹板的第二夹板序列含有与所述另一个第二单链寡核苷酸中的5'末端序列杂交的5'元件和与所述第一连接产物的3'末端序列杂交的3'元件。

16. 根据权利要求12至14中任一项所述的方法,其中所述索引夹板进一步包含与第五索引序列相邻的第二夹板序列,其中所述第二夹板序列的5'元件可以与所述第二DNA序列组分的3'末端序列杂交。

17. 根据权利要求5至11中任一项所述的方法,其中所述第三索引序列可去除地连接至所述第一DNA序列组分。

18. 根据权利要求17所述的方法,其中所述第一单链寡核苷酸进一步包含在所述第一DNA序列组分的5'的IIIs型限制性酶位点,其中所述限制性酶切割位点紧邻所述第一DNA序列组分的5'。

19. 根据权利要求17所述的方法,其中所述第三索引序列通过可切割接头连接至所述第一DNA序列组分,其中切割位点紧邻所述第一DNA序列组分的5'。

20. 根据权利要求17至19中任一项所述的方法,所述方法进一步包括:

从所述第一连接产物中的第一DNA序列组分上去除所述第三索引序列,提供具有所述第三索引序列的另一个第一单链寡核苷酸和另一个索引夹板,随后进行步骤(A)、(B)和(C),

其中所述第一连接产物是步骤(A)的第二单链寡核苷酸,并且

其中所述另一个索引夹板的夹板序列含有与所述另一个第一单链寡核苷酸中的3'末端序列杂交的3'元件和与所述第一连接产物中的5'末端序列杂交的5'元件。

21. 根据权利要求1至20中任一项所述的方法,其中:

所述杂交步骤在基于所述索引元件的熔解温度( $T_m$ )的温度范围内进行,并且

所述连接步骤在高于所述夹板序列的5'和3'元件的熔解温度( $T_m$ )并且低于所述索引元件的熔解温度( $T_m$ )的温度下进行。

22. 根据权利要求1至21中任一项所述的方法,其中所述连接步骤的温度为约45°C至70°C,并且所述连接酶是热稳定的。

23. 根据权利要求1至22中任一项所述的方法,其中所述索引元件的熔解温度为约60°C至90°C。

24. 根据权利要求1至23中任一项所述的方法,其中所述5'夹板元件和所述3'夹板元件的熔解温度为约20°C至42°C。

25. 一种用于组装两个核酸组分以产生多组分连接产物的寡核苷酸结构,所述寡核苷酸结构包含第一单链寡核苷酸、第二单链寡核苷酸和索引夹板,其中所述第一单链寡核苷酸和所述第二单链寡核苷酸与所述索引夹板部分地杂交。

26. 一种寡核苷酸结构的集合或文库,每个寡核苷酸结构具有根据权利要求25所述的寡核苷酸结构。

27. 一种试剂盒,所述试剂盒包含根据权利要求25所述的集合或文库。

28. 根据权利要求27所述的试剂盒,所述试剂盒进一步包含DNA连接酶。

## 使用索引夹板寡核苷酸通过连接组装长核酸

### 相关申请的交叉引用

[0001] 本申请要求2021年8月2日提交的美国临时申请号63/228,448的优先权权益,将其全部内容通过引用以其整体并入本文。

### 技术领域

[0002] 本公开文本涉及分子生物学的技术领域。特别地,本公开文本涉及用于产生长核酸,尤其是长DNA文库的方法。

### 背景技术

[0003] DNA文库广泛应用于生物学研究以及药物发现。目前,DNA文库最经常通过寡核苷酸(“oligo”)的化学合成(在固体表面上以大规模平行的方式产生或通过合并单独化学合成的寡核苷酸)产生。这些方法的一个主要限制是DNA分子的长度,所述DNA分子可以通过单链化学合成制备。最长的可商购的寡核苷酸文库目前具有350个核苷酸的最大长度。然而,许多研究应用(诸如基因合成、蛋白质工程、基因组工程和突变饱和扫描)将受益于更长的DNA文库。

[0004] 一种产生较长长度的文库的方法涉及通过DNA连接酶的作用将两个或更多个化学合成的寡核苷酸连接成较长构建体而酶促组装文库成员。本技术领域已知的方法依赖于特定双链序列上的单链相容突出端以实现序列的特异性退火,以便以预定顺序连接序列。参见例如美国专利号6,495,318。所述‘318专利还披露了具有末端单链部分的两个特定序列,所述末端单链部分与具有与每个碎片互补的区域的单个夹板寡核苷酸杂交。在理想条件下,寡核苷酸适当地与夹板杂交,并且连接酶连接两个碎片以形成更长的构建体。与此类似的方法目前被多种研究小组采用,并且有成功也有失败。此方法的主要缺点是杂交特异性,这可能导致质量差的文库。

[0005] 用于长寡核苷酸的酶促组装的可替代方法涉及将文库碎片与固体表面DNA阵列杂交,以便在连接之前在空间上分隔构建体。参见例如美国专利号10,538,796。

[0006] 在一些方面,本发明解决了对于可以在溶液中以对目的序列高的度特异性实践的方法的需要。

### 附图说明

[0007] 图1描绘了用于将寡核苷酸组装成多组分DNA构建体的寡核苷酸结构的类型。寡核苷酸结构由两个寡核苷酸碎片(1是组装组分1,上游组装组分,并且2是组装组分2,下游组装组分)组成,所述两个寡核苷酸碎片与具有与每个寡核苷酸碎片互补的相邻序列区域的单个夹板寡核苷酸(夹板)杂交。箭头指示寡核苷酸的3’末端;平端是寡核苷酸的5’末端。预期的连接点是组装组分1和2的连接处。

[0008] 图2描绘了用于将寡核苷酸组装成多组分DNA构建体的寡核苷酸结构的另一个类型。寡核苷酸结构由与索引夹板结构杂交的两个寡核苷酸碎片(组装组分1和组装组分2)组

成。索引夹板结构具有一对夹板寡核苷酸成员。所述对的每个成员具有索引序列,所述索引序列与另一个成员的另一个索引序列杂交以形成双链索引元件(索引),因此用于使两个组装组分彼此紧密靠近。一个夹板寡核苷酸成员具有与组装组分1寡核苷酸互补的夹板序列区域(L-夹板)。另一个夹板寡核苷酸成员具有与组装组分1的3'末端部分和组装组分2寡核苷酸的5'末端部分互补的夹板序列区域(R-夹板)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。预期的连接点是组装组分的连接处。

[0009] 图3描绘了用于将寡核苷酸组装成多组分DNA构建体的寡核苷酸结构的另一个类型。寡核苷酸结构由与索引夹板结构(类似于图2中的结构)杂交的两个寡核苷酸碎片(组装组分1和组装组分2)组成。索引夹板结构具有一对夹板寡核苷酸成员。所述对的每个成员具有索引序列,所述索引序列与另一个成员的另一个索引序列杂交以形成双链索引元件。在此结构中,夹板寡核苷酸成员具有与组装组分1寡核苷酸中的序列互补的夹板序列区域(L-夹板),所述组装组分1寡核苷酸中的序列不接近组装组分1寡核苷酸的3'末端,从而导致组装组分1的环状部分。另一个夹板寡核苷酸成员具有与组装组分1的3'末端部分和组装组分2寡核苷酸的5'末端部分互补的夹板序列区域(R-夹板)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。预期的连接点是组装组分的连接处。

[0010] 图4描绘了用于将寡核苷酸组装成多组分DNA构建体的寡核苷酸结构的另一个类型。寡核苷酸结构由与单个索引夹板分子杂交的两个寡核苷酸碎片(组装组分1和组装组分2)组成。在此结构中,组装组分1寡核苷酸具有在组装序列的5'的5'索引序列。所述索引序列与索引夹板分子的索引序列杂交以形成双链索引元件。夹板寡核苷酸成员具有这样的序列区域(R-夹板),所述序列区域与组装组分1寡核苷酸中的在组装组分1寡核苷酸的3'末端的序列互补并且与组装组分2寡核苷酸的5'末端部分互补,从而导致组装组分1的成环(有点类似于图3)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。预期的连接点是组装组分的连接处。

[0011] 图5A-图5C描绘了用于将寡核苷酸组装成多组分DNA构建体的寡核苷酸结构的另一个类型。寡核苷酸结构包含与两个寡核苷酸碎片(组装组分1和组装组分2)杂交的索引夹板分子(描绘于图5A中)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。虚线表示任意的间隔序列。如图5A所示,在此类型的寡核苷酸结构中,索引夹板分子包含夹板序列(夹板)和两个索引序列A和B(索引序列B和索引序列A)。索引夹板分子的两个索引序列位于夹板序列的侧翼,具有插入的间隔序列(间隔物1和间隔物2)(虚线)。夹板序列具有5'元件和3'元件,如在夹板的放大图中所描绘的。夹板序列的5'元件与组装组分2寡核苷酸的5'末端杂交。夹板序列的3'元件与组装组分1寡核苷酸的3'末端杂交。索引夹板分子的一个索引序列与一个组装组分寡核苷酸的末端处的索引序列杂交。索引夹板分子的另一个索引序列与另一个组装组分寡核苷酸的末端处的索引序列杂交。图5A和图5B展示了索引夹板分子结合寡核苷酸结构中的两个组装组分的两种示例性取向。图5B描绘了这样的取向,其中通过索引序列A(索引夹板分子的5'末端处的索引序列)与组装组分2的3'末端处的索引序列的杂交形成双链索引元件(索引2)。双链索引元件(索引1)通过索引序列B(索引夹板分子的3'末端处的索引序列)与组装组分1的5'末端处的索引序列杂交形成。夹板序列的5'元件与组装组分2寡核苷酸的5'末端杂交。夹板序列的3'元件与组装组分1寡核苷酸的3'末端杂交。预期的连接点是组装组分的连接处。图5C描绘了第二取向,其中通过索引序列A(索引夹板

分子的5'末端处的索引序列)与组装组分1的5'末端处的索引序列(区段C)的杂交形成双链索引元件(索引1)。双链索引元件(索引2)通过索引序列B(索引夹板分子的3'末端处的索引序列)与组装组分1的3'末端处的索引序列(区段F)杂交形成。夹板序列的5'元件与组装组分2寡核苷酸的5'末端杂交。夹板序列的3'元件与组装组分1寡核苷酸的3'末端杂交。预期的连接点是组装组分的连接处。

[0012] 图6展示了图2的变型,其中寡核苷酸结构包含至少一个阻断寡核苷酸(例如,阻断物1),其可以与组装序列的不与夹板杂交的区域杂交。两个阻断寡核苷酸(阻断物1和阻断物2)描绘于图6中。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。预期的连接点是组装组分的连接处。一个或多个阻断核苷酸可以用于图5B和图5C的寡核苷酸结构中。

[0013] 图7描绘了在4部分组装过程的第一连接步骤中待连接的两个寡核苷酸结构。(还参见图8和图9)。第一连接步骤以类似于2部分组装的机制在单独的反应中组装组装组分1+2(顶部结构)和组装组分3+4(底部结构)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。指示了每个寡核苷酸结构中的预期连接点。除了第一夹板结构(夹板1)外,顶部寡核苷酸结构的索引夹板还包含在间隔节段中的第二夹板序列(夹板2)和在5'索引序列(索引2)与第一夹板结构(夹板1)之间的另一个索引序列(索引3)。底部寡核苷酸结构的索引夹板包含在第一夹板结构(夹板3)与5'索引序列(索引4)之间的索引3的互补序列(索引3互补序列)。

[0014] 图8描绘了4部分组装过程的下一步骤。图7中描绘的组装步骤的连接产物展示于图8中。连接产生两个多组分连接产物:组装产物(1-2)(描绘于(a)中)和组装产物(3-4)(描绘于(c)中)。还描绘了对于组装产物(1-2)和组装产物(3-4),位于索引序列与组装组分之间的IIS型限制性酶切割位点(RE切割位点)。使图7中所描绘的组装步骤的连接产物各自经受用IIS型限制性酶的消化,以在RE切割位点处切割(图7中未示出)。在(a)中的连接产物中,切割去除索引2,仅在组装产物(1-2)的3'末端留下组装组分序列,如(b)所展示。在(c)中的连接产物中,切割去除索引4,仅在组装产物(3-4)的5'末端留下组装组分序列,如(d)所展示。所得产物例如通过SPRI纯化来纯化。

[0015] 图9描绘了4部分组装过程的第二连接步骤。然后将限制性消化的纯化产物(参见图8中的(b)和(d))如示意图(参见索引3)所描绘的杂交在一起并且连接以产生组装产物(1-2-3-4)。此产物可以例如通过SPRI纯化并且经受另外的连接或经由PCR扩增。例如,引物序列可以存在于索引1与组装组分1的组装序列(正向引物)之间以及在索引5与组装组分4的组装序列(反向引物)之间。

[0016] 图10-图13一起示意性地展示了所公开的方法的变体方面。图10展示了寡核苷酸结构,其中两个组装组分的索引序列与单个夹板(“索引夹板”)杂交,所述单个夹板与两个索引序列杂交形成双链索引1和索引2。

[0017] 图11描绘了图10的寡核苷酸结构的连接产物。图11还展示了任选的方面,其中索引夹板被生物素化以附接至固体支持物或索引阵列。

[0018] 图12A和图12B描绘了寡核苷酸结构,其中第二夹板分子(“构建体夹板”)与组装组分的3'和5'末端杂交。图12A描绘了寡核苷酸结构,其中索引夹板经由间隔物(“任选的系链”)连接至构建体夹板。这代表所要求保护的索引夹板分子的另一个实施方案,所述索引夹板分子从5'至3'包含第一索引序列(在寡核苷酸1的5'末端处结合索引以形成索引1);第

二索引序列(在寡核苷酸2的3'末端处结合索引以形成索引2);间隔物(任选的系链):夹板,其中5'元件与5'末端寡核苷酸2杂交,并且3'元件与寡核苷酸链1的3'末端杂交。图12B描绘了寡核苷酸结构,其中未连接至索引夹板的第二夹板分子(“构建体夹板”)与组装组分的3'和5'末端杂交。还描绘了任选的方面,其中索引夹板被生物素化以附接至固体支持物或索引阵列。

[0019] 图13示出了图12B的连接产物。图13还示出了连接产物的组装组分上的任选的另外的结构特征。在索引夹板的5'末端处描绘了任选的限制性酶位点。PBS1和BPS1代表可用于扩增组装产物的(任选的)引物结合位点。

[0020] 图14描绘了实验的数据图,所述实验评价了使用如图1、图2和图3中所展示的寡核苷酸结构(其各自进一步包含组装组分1和组装组分2)中描绘的三种类型的测试夹板分子进行低同源性文库的溶液组装。对于每种类型的测试夹板分子,测试了三种夹板与构建体的比率(1:1、2:1和4:1)。图2和图3的测试夹板分子是用(+)和不用(-)3个核苷酸(nt)的支柱(stilt)制备的。通过PCR扩增的组装构建体的下一代测序(NGS)表征组装构建体。图中的Y轴是15个集合中的每一个中正确组装构建体的读段的平均数。更多信息参见实施例1。

[0021] 图15描绘了图5B的测试夹板分子的示意图(上图)和来自概念验证实验的数据(下图)。箭头指示寡核苷酸的3'末端;平端是寡核苷酸的5'末端。引物F-用于PCR的正向引物的位置。引物R-用于PCR的反向引物的位置。 $T_m$ 是与左和右构建体寡核苷酸(分别为组装组分1和组装组分2)杂交的夹板的每个节段的熔解温度(即,5'元件和3'元件的熔解温度)。BC1和BC2指示配对的条形码序列的位置。索引1和索引2指示寡核苷酸结构中的索引元件。索引夹板分子的索引序列与夹板序列之间的虚线指示间隔序列,所述间隔序列在此测试中是20个碱基的聚T核苷酸序列。更多信息参见实施例2。

[0022] 图16示意性地展示了三种类型的代表性寡核苷酸文库。实心框代表在组装的构建体侧翼的独特条形码。

[0023] 图17描绘了在三种类型的代表性寡核苷酸文库(参见图16)中利用三种类型的测试夹板分子(参见图1、图2和图5B的寡核苷酸结构)测试组装的代表性数据。更多信息参见实施例3。

[0024] 图18描绘了测试夹板分子(图5B的寡核苷酸结构)的示意图和使用具有不同夹板长度和连接温度的所述分子进行组装的代表性数据。在实验中,设计了每个夹板-构建体杂交元件(5'元件和3'元件)的熔解温度( $T_m$ )范围,并且测试了在八(8)种不同温度下的连接。图中描绘了65.9°C连接数据的中值读段计数的数量的数据。对于每种夹板温度,正确读段和错误读段的分数以中值读段计数表示。更多信息参见实施例4。

[0025] 在任何附图中或在说明书中使用“半”,例如“半构建体左”,不旨在限制。例如,“半构建体左”和“半构建体右”不旨在指示组装组分1和组装组分2具有相同的长度。组装组分的尺寸可以不同。在此上下文中,“半”相当于“部分”(“portion”)、“部分”(“part”)或“寡核苷酸分子”。

### 具体实施方式

[0026] 本公开文本尤其涉及一种用于通过酶促连接与索引夹板寡核苷酸分子杂交的寡核苷酸而在溶液中产生DNA文库的方法。在一些方面,寡核苷酸是化学合成的。所述方法

总体上可用于通过连接较短的组分来产生生长核酸,包括DNA或RNA。

#### A. 定义

[0027] 如本文所用的,单数形式“一个/种(a)”、“一个/种(an)”和“所述(the)”包括复数种/个指示物,除非上下文另外明确说明。

[0028] 如本文所用,术语“核酸组装”是指将至少两个核酸片段或寡核苷酸以特定顺序物理连接在一起以产生所希望的核酸序列(多组分构建体)或此过程所得核酸的方法。每个片段或寡核苷酸是“组装组分”(在本文中有时通俗地是指“半构建体”)。当连接两个组装组分时,上游组分是“上游组装组分”(或组装组分1或左半构建体或半构建体左),并且下游组分是“下游组装组分”(或组装组分2或右半构建体或半构建体右)。上游组装组分(组装组分1)的3'末端与下游组装组分(组装组分2)的5'末端连接。组装组分包含组装序列和至少一个非组装序列。“组装序列”是目的序列,其包含在最终的多组分构建体中。非组装序列是指与组装方法相关的序列,诸如索引序列、引物序列、限制性酶序列、条形码序列等。

[0029] 如本文所用,术语“5'末端序列”是指构成核酸的5'末端的寡核苷酸序列。如本文所用,术语“3'末端序列”是指构成核酸的3'末端的寡核苷酸序列。末端序列可以含有2个或更多个连续核苷酸、3个或更多个连续核苷酸、5个或更多个连续核苷酸、或10个或更多个连续核苷酸、30个或更多个连续核苷酸、50个或更多个连续核苷酸、或100个或更多个连续核苷酸。3'或5'“基本上末端的序列”是指邻近但不包括3'或5'末端的序列。例如,3'或5'基本上末端的序列的3'或5'末端可以分别在3'或5'末端核苷酸的约100个核苷酸内、约50个核苷酸内、约30个核苷酸内、约25个核苷酸内、约20个核苷酸内、约15个核苷酸内、约10个核苷酸内、约5个核苷酸内、约4个核苷酸内、约3个核苷酸内、约2个核苷酸内或1个核苷酸内。

[0030] 术语“核酸”、“多核苷酸”或“寡核苷酸”是指DNA分子、RNA分子或其类似物。如本文所用,术语“核酸”、“多核苷酸”和“寡核苷酸”包括但不限于DNA分子,诸如cDNA、基因组DNA或合成DNA;和RNA分子,诸如引导RNA、信使RNA或合成RNA。此外,如本文所用,所述术语包括单链和双链形式。

[0031] 术语“杂交”(“hybridization”)或“杂交”(“hybridizing”)是指这样的过程,其中完全或部分互补的多核苷酸链在合适的杂交条件下聚集在一起以形成双链结构或区域,其中两个组成链通过氢键连接。如本文所用,术语“部分杂交”包括双链结构或区域含有一个或多个凸起或错配的情况。尽管典型地在腺嘌呤与胸腺嘧啶或腺嘌呤与尿嘧啶(分别为A与T或A与U)或胞嘧啶与鸟嘌呤(C与G)之间形成氢键,但也可以形成其他非规范碱基对(参见例如,Adams等人,“The Biochemistry of the Nucleic Acids,”第11版,1992)。考虑了经修饰的核苷酸可以形成允许或促进以非规范方式杂交的氢键。

[0032] 如本文所用,序列的术语“部分”、“区段”、“元件”或“片段”是指所述序列的小于完整序列的任何部分(例如,核苷酸子序列或氨基酸子序列)。多核苷酸的部分、区段、元件或片段可以具有大于1的任何长度,例如,长度为至少5、10、15、20、25、30、40、50、75、100、150、200、300或500个或更多个核苷酸。

[0033] 如本文所用,术语“索引夹板”是指包含以下的寡核苷酸:(1)可以与至少一个靶序列杂交的夹板序列,和(2)可以与靶分子中的索引序列杂交的索引序列。参见例如,图5A。“靶标”是组装组分。典型地,索引夹板被设计成与两个靶标杂交,以便于促进两个靶标连接成多组分构建体。参见例如图5B。为此,将夹板序列典型地与两个靶标杂交,并且使一个靶

标的3'末端和另一个靶标的5'末端接近。靶标中的索引序列是为了与索引夹板中的索引序列杂交而添加到靶标中的额外序列。

[0034] 如本文所用,术语“索引夹板特异性有序连接”是指这样的方案,其中使用连接酶将基本上单链的片段彼此连接以产生多组分核酸连接产物,其中连接产物中片段的顺序由索引夹板分子来引导,所述索引夹板分子包含与待连接的两个片段的3'末端序列和5'末端序列杂交的夹板。“基本上单链的片段”是指包含部分双链部分的片段,诸如与索引夹板分子中的夹板序列杂交的3'或5'末端序列和/或与索引夹板分子中的索引序列杂交的3'或5'末端的索引序列(或3'或5'基本上末端的索引序列)。

[0035] “索引元件”是指由具有彼此互为反向互补序列的序列的两条链形成(构成)的双链DNA区段。“索引序列”是指这样的序列,其是彼此互为反向互补序列并且通过彼此杂交形成索引元件的序列的两条链之一。在一些实施方案中,索引序列不是回文的(例如,没有茎环或其他二级结构)。单个索引夹板上的两个索引序列彼此不杂交。

[0036] 本公开文本的方法和产物至少部分地基于出乎意料的发现,即利用如图5图A中示意性描绘的索引夹板分子导致寡核苷酸文库中多组分构建体的组装改善。通过使用索引夹板特异性有序连接,其中待连接的每对组装组分具有相应的索引夹板分子,所述相应的索引夹板分子具有夹板序列和对这对组装组分特异性的索引序列,在寡核苷酸文库中、甚至在同源寡核苷酸文库(其中寡核苷酸共享相当程度的序列同源性的寡核苷酸文库)中组装多个正确构建体是可能的。索引夹板分子的索引序列不是构建体序列的一部分,因此不受文库成员之间的同源性影响。与之前的多路组装的高通量方法相比,这是出乎意料的益处和改善。

[0037] 如本文所用,“单链寡核苷酸的混合物”是指含有溶解在其中的多种不同单链寡核苷酸的水溶液。混合物可以包含至少50个、至少100个、至少500个、至少1,000个、至少5,000个、至少10,000个或至少50,000个或更多个寡核苷酸,包括在50与例如50,000之间的任何值。寡核苷酸的混合物可以通过以下方式制备:原位合成寡核苷酸,即在阵列中在适当的位置合成寡核苷酸并且然后在寡核苷酸已经合成后从阵列表面上切割所述寡核苷酸。参见,例如,Geary等人(Nature Methods 2004 1:241-248)和LeProust等人(Nucleic Acids Research 2010 38:2522-2540)。

[0038] 如本文所用的术语“寡核苷酸”表示核苷酸的多聚体。例如,寡核苷酸可以具有长度为约2至约200个核苷酸、至多约50个核苷酸、至多约100个核苷酸、至多约500个核苷酸、或核苷酸数量在2与500之间的任何整数值。在一些实施方案中,寡核苷酸的长度可以在30至300个核苷酸或30至400个核苷酸的范围内。寡核苷酸可以含有核糖核苷酸单体(即,可以是寡核糖核苷酸)和/或脱氧核糖核苷酸单体。寡核苷酸的长度可以是例如10至20、21至30、31至40、41至50、51至60、61至70、71至80、80至100、100至150、150至200、200至250、250至300、300至350、或350至400个核苷酸,以及这些范围之间的任何整数值。

[0039] 如本文所用,术语“可以组装以产生多组分连接产物的单链寡核苷酸的集合”及其语法等效物是指可以酶促组装成更长序列的寡核苷酸的集合,所述更长序列在本文中称为“多组分构建体”或“多组分连接产物”,其含有来自限定顺序的各寡核苷酸的序列。如从本公开文本可以理解,集合的单链寡核苷酸可以含有:(i) 3'或5'末端(或基本上末端)序列,所述序列是与索引夹板分子中的互补序列杂交以形成双链索引元件的索引序列,和(ii)包

含与索引夹板分子中的夹板序列杂交的5'或3'末端序列的组装序列。“组装序列”是目的序列,所述目的序列包含在最终构建体中,并且不包括与组装方法相关的序列,诸如索引序列、引物序列、限制性酶序列、条形码序列等。

[0040] 如本文所公开的,提供了多个数值范围。应理解,还具体考虑的在该范围的上限与下限之间的每个中间值。还具体考虑了由所陈述的范围涵盖的每个较小范围或中间值。术语“约”通常是指所指示数字的正或负10%。例如,“约10%”可以指示9%至11%的范围,并且“约20”可以意指18-22。“约”的其他含义可以从上下文中变得清楚,诸如四舍五入,因此,例如“约1”也可以意指0.5至1.4。

[0041] 在整个说明书中可以出现术语的其他定义。

#### B. 方法和产物的描述

[0042] 图1中描绘了一种用于组装特异性寡核苷酸用于酶促连接以形成所希望的构建体的方法。图1描绘了待连接的两个寡核苷酸,一个是上游寡核苷酸(“组装组分1”),并且另一个是下游寡核苷酸(“组装组分2”)。上游构建体寡核苷酸在连接点的上游(以5'至3'为基础),并且下游构建体寡核苷酸在连接点的下游。上游和下游构建体寡核苷酸与单个夹板寡核苷酸杂交。夹板寡核苷酸具有两个互补序列的区域,一个与上游构建体寡核苷酸杂交,并且另一个与下游构建体寡核苷酸杂交,使得左构建体的3'末端紧邻右构建体的5'末端以形成预期的连接位点。在杂交条件下,两个组装组分与单个夹板寡核苷酸杂交,并且连接酶连接两个碎片以形成更长的多组分构建体。

[0043] 在一些实施方案中,所公开的方法采用包含夹板区域和两个索引序列A和B的索引夹板分子(参见例如图5A)以组装用于连接的组装组分。索引夹板分子的两个索引序列位于夹板序列的侧翼,具有插入的间隔序列(虚线)。夹板区域具有5'元件和3'元件,如在夹板的放大图中所描绘的。夹板序列的5'元件与组装组分2寡核苷酸的5'末端杂交(参见例如图5B)。夹板序列的3'元件与组装组分1寡核苷酸的3'末端杂交。在一些方面,索引序列A和B不是回文的(例如,没有茎环或其他二级结构)并且也不彼此杂交。索引夹板分子的一个索引序列与一个半构建体末端处的索引序列杂交。索引夹板分子的另一个索引序列与另一个半构建体末端处的索引序列杂交。因此,在夹板序列和索引序列二者都与两个半构建体杂交的情况下,索引夹板使两个半构建体接近,并且将它们的位置对齐以进行连接。

[0044] 索引序列不是目的序列的一部分。它们是添加至两个组装组分末端的额外序列。因此,它们可以被设计为与反应混合物中的其他序列尽可能不同,从而减少非预期的组装组分对的机会,特别是当多种不同的组装组分在相同反应(即,多路)中连接时。

[0045] 夹板的3'元件和5'元件各自具有约5至20个核苷酸的长度并且任选地具有小于索引元件溶解温度的溶解温度。在一些实施方案中,夹板元件不应在单独的情况下(即,在不存在索引元件的情况下)足以在反应条件下驱动杂交,但是当只有组装组分通过索引夹板的索引序列与两个组装组分的杂交而保持在适当的位置时才会瞬时杂交,以便基本上减少或消除两个组装组分的非预期的配对。当反应中的组装组分共享相当大程度的同源性(例如,同源寡核苷酸文库)时,这是特别重要的。

[0046] 如图5B和图5C所展示的,索引夹板分子结合寡核苷酸结构中的两种组装组分存在两种示例性取向。

[0047] 图5B描绘了一种取向,其中索引序列A(索引夹板分子的5'末端处的索引序列)与

组装组分2的3'末端处的索引序列杂交,形成双链索引元件(索引2)。索引序列B(索引夹板分子的3'末端处的索引序列)与组装组分1的5'末端处的索引序列杂交,形成双链索引元件(索引1)。

[0048] 图5C描绘了第二取向,其中索引序列A(索引夹板分子的5'末端处的索引序列)与组装组分1的5'处的索引序列杂交,形成双链索引元件(索引2)。索引序列B(索引夹板分子的3'末端处的索引序列)与组装组分2的3'处的索引序列杂交,形成双链索引元件(索引1)。注意,索引夹板分子展示有180°的扭转,以适应夹板的5'元件与组装组分2寡核苷酸的5'末端杂交并且夹板序列的3'元件与组装组分1寡核苷酸的3'末端结合杂交。扭转仅用于展示的目的并且不旨在限制。

[0049] 本发明在将多核苷酸组装成多组分构建体中的示例性优点和区别性特征包括以下。本发明不需要用于组装方法的阵列或固体支持物。所述方法可以用于组装单链寡核苷酸。在一个方面,所述方法使用合成寡核苷酸进行。在一个方面,所述方法使用包含单链寡核苷酸混合物的寡核苷酸文库进行。混合物可以包含至少50个、至少100个、至少500个、至少1,000个、至少5,000个、至少10,000个或至少50,000个或更多个寡核苷酸。另外,在一些方面,所述方法可以作为迭代过程进行,其中非组装序列可以在下一个循环之前干净地去除,像II型限制(即,片段连接的位点是“无疤的”)。所述方法可以用于组装复杂文库、低同源性文库和高同源性文库,诸如具有基本上相同的序列(除了小的变型)的文库。

[0050] 索引夹板可以以其他方式使用。在图2所示的一种变型中,代表预期构建体的左半部分和右半部分的两个寡核苷酸与一对夹板寡核苷酸成员杂交。这对夹板寡核苷酸成员的每个成员进一步含有索引序列,所述索引序列是所述对的另一个成员中的索引序列的反向互补序列,并且因此能够彼此杂交并且形成在本文中称为索引元件(在图2中标记为“索引”)的双链DNA部分。两个索引序列间接地用于将两个组装组分集合在一起。索引序列不是组装序列本身的一部分。在图2中,右索引夹板含有与左构建体寡核苷酸的3'末端同源的夹板序列(R-夹板)和与右构建体寡核苷酸的5'末端同源的节段。在图2中,R-夹板含有约10个核苷酸(nt),但它可以在例如约5nt至约20nt的范围内。

[0051] 所有夹板序列不必须与靶标的末端序列杂交。索引夹板的另一个变型展示于图3中。此寡核苷酸结构是图2寡核苷酸结构的变体。诸位发明人考虑了此结构可以减少或消除由缺少设计的组装组分的5'末端的截短合成产物引起的干扰。在此结构中,夹板寡核苷酸成员具有与组装组分1寡核苷酸中的序列互补的夹板序列区域(L-夹板),所述组装组分1寡核苷酸中的序列不接近组装组分1寡核苷酸的3'末端,从而导致组装组分1的环状部分。另一个夹板寡核苷酸成员具有与组装组分1的3'末端部分和组装组分2寡核苷酸的5'末端部分互补的夹板序列区域(R-夹板)。

[0052] 在图4中,组装组分1寡核苷酸在5'末端组装组分1处具有5'索引序列。所述索引序列与索引夹板分子的索引序列杂交以形成双链索引元件。夹板分子具有与组装组分1组装组分1的3'末端部分并且与组装组分2寡核苷酸的5'末端部分互补的序列,导致组装组分1的成环(类似于图3,但仅使用一个索引夹板)。

[0053] 如本文所示,实施方案(诸如在图5A-图5C中所展示的实施方案)提供了值得注意的优点。使用图5索引夹板分子组装低同源性文库或复杂文库产生高准确度(即,相对于读段的总数,正确读段的总数)。参见图17。“复杂文库”是指除了外围引物结合位点之外具有

非常小序列相似性的文库序列。另外,没有辍失(dropout)构建体。因此,数据表明使用图5的索引夹板分子将组装组分组装能够实现高得多的辨别。因此,本发明能够正确组装构建体,否则由于夹板分子(诸如图1中的夹板分子)的混杂杂交而不可能正确组装构建体。

[0054] 因此,本公开文本提供了一种利用特异性寡核苷酸采用用于酶促连接的索引夹板组装多组分构建体的方法。除了其他方面,本公开文本提供了一种用于通过索引夹板特异性有序连接来组装两个或更多个特定组分以形成具有预定顺序的多组分连接产物的方法。所述方法可以在溶液中并且在不存在固体支持物的情况下进行。所述方法可以用于通过来自以限定顺序连接两个、三个或四个或更多个单链寡核苷酸的序列来产生合成多组分构建体。所述方法可以是多路的,即可以在相同溶液中产生多种不同的合成多组分构建体(例如,至少2种、至少10种、至少100种、至少1,000种或至少10,000种、或至少100,000种多组分构建体)。此过程也可以重复几次以产生更长的核酸。

[0055] 在某些实施方案中,使用合成有机化学领域熟知的方法通过化学合成产生寡核苷酸分子。本文所述的一些寡核苷酸分子可以可替代地使用本领域已知的方法(包括体外转录和基于细胞的表达)来制备。

[0056] 在本公开文本的方法中使用的各种寡核苷酸分子以及所述方法在以下本公开文本中更详细地描述。

#### 索引夹板

[0057] 在一些实施方案中,本公开文本的方法采用如图5A中所展示的索引夹板。索引夹板包含夹板区域、第一索引序列和第二索引序列。在一些实施方案中,第一索引序列和第二索引序列位于夹板区域的侧翼。第一索引序列和第二索引序列可以位于索引夹板的5'和3'末端。在一个方面,第一索引序列和第二索引序列中的一个或两个包含索引夹板的末端序列(例如,5'末端序列和/或3'末端序列)。在一些实施方案中,第一索引序列和第二索引序列中的一个或两个包含索引夹板的基本上末端的序列(例如,5'基本上末端的序列和/或3'基本上末端的序列)但不包含末端核苷酸。夹板区域具有5'序列元件和3'序列元件,所述5'序列元件和3'序列元件分别与两个组装组分中的每一个杂交,并且使组装组分的末端接近。如图5A-图5C所展示的,夹板序列的5'元件与下游组装组分(组装组分2)的5'末端杂交,而夹板序列的3'元件与上游组装组分(组装组分1)的3'末端杂交。结果,下游组装组分的5'末端和上游组装组分的3'末端集合在一起,并且可以通过连接酶连接。

[0058] 索引夹板分子结合寡核苷酸结构中的两个组装组分有两种示例性取向,分别示出于图5B和图5C中。

[0059] 索引序列不是目的序列的一部分,并且不会出现在最终组装体中。它们是添加至两个组装组分末端的额外序列。因此,它们可以被设计为与反应混合物中的其他序列尽可能不同,从而减少非预期的组装组分对的机会,特别是当多种不同的组装组分在相同反应(即,多路)中连接时。

[0060] 索引元件的特征包括以下。索引序列被设计为仅与其互补序列杂交。因此,对于任何给定的索引元件,两个序列(互补序列或反向互补序列)都不存在于最终构建体分子中。类似地,对于包含一个或多个间隔序列的索引夹板分子,二个索引序列(互补序列或反向互补序列)都不存在于间隔序列中。索引序列优选地被设计为成本身不具有稳定的二级结构(诸如茎-环结构),并且因此不是回文的。此外,索引元件的核苷酸组成优选地被设计为具

有用于在组装方法中使用的合适熔解温度 ( $T_m$ )。在一些实施方案中,索引元件被设计为在夹板元件仅能与组装组分瞬时杂交的温度范围内保持杂交,除了夹板元件与组装组分具有100%的同一性。夹板元件序列中与构建体元件的错配使杂交不稳定。不受理论的束缚,认为这种动态杂交允许有利于形成100%同一性的夹板元件与组装组分杂交的采样。认为这有助于在本公开文本的方法中形成的多组分构建体的高准确度。在一些方面,组装准确度(即,相对于读段的总数,正确读段的总数)是至少80%、85%或至少90%,诸如至少91%、92%、93%、94%、95%或至少96%。不受理论的束缚,还认为有助于预期构建体的低至无的辍失,并且因此提供所代表的预期构建体的高百分比。在一些方面,预期构建体的百分比是预期构建体总数的至少80%、85%或至少90%,诸如至少91%、92%、93%、94%、95%、96%、97%或至少98%。

[0061] 任选地,索引序列是随机的高复杂性序列,其是碱基平衡的 ( $A=T=G=C$ ) 以帮助预测熔解温度。可以包括均聚物重复。在其他实施方案中,索引序列不包括重复元件,诸如均聚物和/或二核苷酸重复,例如GTGTGTGT。高熔解温度在所公开的方法的实践中是有用的。每个索引序列的 $T_m$ 可以是约50°C至约80°C、约65°C至约72°C、约55°C至约75°C、约50°C至约70°C、或60°C至约80°C。在一个方面,每个索引序列的 $T_m$ 是约65°C至约69°C。在一个方面,索引序列可以是20nt至60nt。索引序列可以是20个核苷酸(nt)至50nt、25nt至45nt、25nt至35nt或28nt至32nt。在一个方面,索引序列是约30个核苷酸。保持索引序列较短的原因是最大化组装组分中的构建体序列长度。制备非常短的索引序列的主要障碍是特异性和双链体稳定性降低。索引序列可以制备为序列的文库。针对适当的基因组筛选潜在的索引序列以排除可能杂交的序列。在一个方面,针对人类基因组筛选潜在的索引序列。序列的文库中的索引序列可以被设计为共享一些特征(诸如长度、碱基平衡或熔解温度),同时被设计为不利于不同索引序列之间的杂交。在一个方面,序列的文库可以含有多个索引序列集合,一个集合内的序列共享特征,而不同于其他集合。例如,序列的文库可以包含一个被设计为具有一个预测熔解温度的索引序列集合,而文库内的第二个索引序列集合可以被设计为具有不同的熔解温度。

#### 夹板

[0062] 索引夹板包含夹板序列。夹板序列旨在引导两个组装组分的配对的特异性。如图5A所展示,夹板序列(夹板)具有两个序列元件:5'元件(5'元件)和3'元件(3'元件)。夹板序列的5'元件与下游组装组分的5'末端杂交。更具体地,5'元件与下游组装组分(组装组分2)的5'末端或5'基本上末端的序列杂交(参见例如图5B)。夹板序列的3'元件与上游组装组分(组装组分1)的3'末端杂交。更具体地,3'元件与上游组装组分(组装组分1)的区段D的3'末端或3'基本上末端的序列杂交(参见例如图5B)。

[0063] 夹板典型地是短序列或熔解温度低于索引区域的熔解温度的序列。在一些实施方案中,每个夹板元件被设计为本身不足以在反应条件下驱动杂交。相反,夹板序列被设计为仅当组装组分通过与索引夹板寡核苷酸的索引序列杂交而保持在适当位置时才与组装组分寡核苷酸瞬时杂交。在一个方面,夹板序列的5'元件和3'元件是连续的,没有插入核苷酸。在一个方面,夹板序列的5'元件和3'元件是基本上连续的,例如,在5'元件与3'元件之间存在插入的很少(一个、两个或三个)核苷酸,并且插入的很少核苷酸不旨在与组装组分的任一个组装序列结合。夹板序列的5'元件和3'元件可以独立地是约5个核苷酸至约25个

核苷酸的长度、或约5个核苷酸至约20个核苷酸的长度、或约8个核苷酸至约15个核苷酸的长度、或约10个核苷酸至13个核苷酸的长度。夹板的5'元件和3'元件总体上具有相似的熔解温度。夹板元件的熔解温度小于索引元件的熔解温度。夹板元件的熔解温度的选择受组装步骤中使用的连接酶的影响。在一些方面,夹板元件的熔解温度( $T_m$ ),当通过 $T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7$ )计算时,可以在10°C至80°C、14°C至44°C、20°C至42°C或22°C至40°C的范围内。熟练技术人员将认识到,可以针对不同的连接酶选择不同的夹板熔解温度,例如,可以在较高温度下使用的连接酶(例如,热稳定的连接酶,诸如9°N<sup>TM</sup>DNA连接酶(New England Biolabs,摩洛哥伊普斯威奇))可以与具有较高熔解温度的夹板一起使用。

[0064] 在一些方面,索引夹板包含间隔序列,所述间隔序列是在索引夹板的特定结构特征(例如,索引序列与夹板序列)之间的接头。在一些实施方案中,间隔物的长度可以是1至30个核苷酸。间隔物是任何序列或非核酸组分,其被设计为不与组装序列或索引序列杂交。在一些实施方案中,间隔物可以是简单序列(诸如均聚物序列)或短重复序列(诸如二或三核苷酸重复)。在一些方面,如图5A所展示,索引夹板具有两个间隔序列。在一些方面,索引夹板具有多于两个间隔序列以连接一个或多个另外的结构特征。例如,如本文其他地方所讨论的,用于迭代组装应用的索引夹板可以含有第二夹板元件和第三索引序列,其中间隔序列可以任选地隔开特定结构特征。

#### 组装组分

[0065] 在一些实施方案中,组装组分是包含希望在最终构建体组装体中的序列和索引序列的寡核苷酸分子。索引序列在组装组分中的位置取决于所述组分在预期的连接反应中是左(上游)组分还是右(下游)组分。

[0066] 典型地,在上游组装组分中,索引夹板序列位于组装组分寡核苷酸的5'末端或基本上位于5'末端。组装序列位于组装组分寡核苷酸的3'末端区域。在一个方面,索引序列与组装序列邻接。在其他方面,另外的序列(结构特征)可以位于索引序列与组装序列之间。此类结构特征可以是以下中的一个或多个:至少一个引物序列、至少一个条形码序列和/或至少一个限制性酶切割位点。在一些方面,IIS型限制性酶位点紧邻组装序列的上游。参见例如图8,分子(c)。在此方面,IIS型限制性酶位点的定位使得在切割位点处的酶促切割将所有非组装序列与组装序列分离。此方面对于所述方法的迭代形式是有用的,例如,其中第一组装步骤的产物然后用作随后的组装步骤中的组分。

[0067] 典型地,在下游组装组分中,索引夹板序列位于组装组分寡核苷酸的3'末端或基本上位于3'末端。组装序列位于组装组分寡核苷酸的5'末端序列。在一个方面,索引序列与组装序列(区段E)邻接。在其他方面,另外的序列可以位于组装序列(区段E)与索引序列(区段F)之间。此类序列可以是以下中的一个或多个:至少一个引物序列、至少一个条形码序列和至少一个限制性酶切割位点。在一些方面,IIS型限制性酶位点紧邻组装序列的下游。参见,例如图8,分子(a)。IIS型限制性酶位点的定位使得在切割位点处的酶促切割将所有非组装序列从组装序列上去除。此方面对于所述方法的迭代形式是有用的。此方面对于所述方法的迭代形式是有用的,例如,其中第一组装步骤的产物然后用作随后的组装步骤中的组分。

#### 方法

[0068] 本公开文本提供了一种用于组装两个核酸组装组分以产生连接产物(多组分组装

体)的方法。在一些实施方案中,所述方法包括使上游组装组分和下游组装组分与存在于反应混合物中的索引夹板杂交。在一些实施方案中,在杂交结构中,上游组装组分的5'末端和3'末端二者均与索引夹板杂交,并且下游组装组分的5'末端和3'末端二者均与索引夹板杂交。杂交步骤随后是连接步骤。在连接步骤中,上游组装组分的3'末端与下游组装组分的5'末端连接,由此产生第一连接产物。在一个方面,反应混合物是多路的并且包含单链寡核苷酸的混合物,所述混合物包含多个上游组装组分种类、多个下游组装组分种类和多个索引夹板分子种类。所述多个索引夹板分子种类中的每一种可以与混合物中的预定的上游组装组分和预定的下游组装组分杂交。

[0069] 所述方法在水溶液中进行。对于杂交步骤和连接步骤二者,水溶液可以是相同的。用于连接步骤的水溶液可以是具有适合于在所述方法中使用的连接酶的合适pH和盐浓度的任何水溶液。反应溶液是典型地缓冲的,并且包含一种或多种盐,并且可以任选地包含适合于连接酶的分子拥挤剂,诸如PEG(聚乙二醇)。在相同的水溶液用于杂交和连接步骤二者的一些实施方案中,缓冲液的组分可以选择为或多或少地有利于连接酶或者或多或少地有利于杂交反应,以实现最佳的结果。例如,如果镁的浓度支持更严格的杂交步骤,则溶液中镁的浓度可以选择为略低于或显著高于对于连接酶反应理想的浓度。

[0070] 杂交步骤可以通过在水浴中孵育来进行,所述水浴在足以形成包含索引夹板和两个组装组分的寡核苷酸结构的时间段内将反应从升高的温度冷却至较冷的温度。杂交步骤的温度特征由索引序列的 $T_m$ 和夹板区域的 $T_m$ 指导。例如,杂交步骤可以在较高温度下开始,随后是冷却期。起始温度超过索引序列的 $T_m$ 以溶解任何二级结构。示例性的温度特征是约95°C至约40°C,诸如93°C至44°C。杂交冷却的持续时间可以在约15分钟至数小时的范围内。示例性的持续时间是1.75小时。在一个方面,杂交步骤包括经1.75小时从93°C冷却至44°C。在一些实施方案中,杂交步骤的温度可以随时间变化或经几个循环从较高温度循环到较低温度。

[0071] 连接步骤通过使杂交的寡核苷酸结构与连接酶接触进行。可以将反应混合物的温度改变为用于待使用的连接酶的适当温度,并且在一些实施方案中,温度低于索引元件的溶解温度并且高于夹板元件的溶解温度。例如,杂交后,将反应混合物的温度升高至约45°C至约70°C、约45°C至65°C、约50°C至65°C或约61°C至65°C等的范围,然后使杂交的寡核苷酸结构与热稳定的连接酶接触。较低的温度可以用于非热稳定的连接酶。连接步骤进行合适的时间长度,其范围可以从几分钟至几小时。任选地,通过本领域已知的方法终止连接步骤。例如,连接酶可以通过高温、通过化学变性剂变性,或者可以添加EDTA(以结合连接反应所需的镁)。

[0072] 所述方法可以用任何连接酶进行。DNA和RNA连接酶是可商购的。可替代地,连接酶可以是化学合成的(参见例如,Creighton, *Proteins: Structures and Molecular Principles*, 第2版, W.H. Freeman & Co., NY, 1992) 或可以经由分子生物学方法(诸如在细胞中表达)重组制备。重组连接酶可以制备为融合蛋白,例如,以帮助纯化。在某些方面,连接酶以纯化或分离的形式提供。在某些实施方案中,连接酶以约80%、约90%、约95%或约99%纯度提供。在某些实施方案中,连接酶作为组合物的一部分提供。在某些实施方案中,连接酶提供在水性组合物中,所述水性组合物适合于用作或包含在用于组装多组分构建体的组合物中。本领域技术人员熟知可以包含在此类连接酶反应组合物中的各种物质。

[0073] 连接酶的选择将影响连接步骤的温度。不同的连接酶可以具有不同的效率、在不同温度下的不同活性以及对连接位点处的错配的不同敏感性。有许多在较低温度范围起作用的嗜中温DNA连接酶(mesophilic DNA ligase)的例子。热稳定的连接酶也是可用的。热稳定的连接酶允许在较高温度(例如,高达65°C)下的连接酶反应条件。较高温度反应条件可以有助于更准确的组装构建体。在一个方面,在连接位点处发生连接之前,热稳定的连接酶具有更大的足迹要求(例如,约25个核苷酸的杂交序列)。此特征可以有助于更大的特异性,导致连接产物的纯度更高。在一个方面,可以使用多种连接酶。在一个方面,可以将不同的连接酶依序添加至相同的杂交反应中。在一个方面,可以在相同的反应中包含不同的连接酶。例如,可以在反应中包括嗜中温连接酶和热稳定的连接酶,并且在升高嗜热连接酶的温度之前,可以在较低温度下使用嗜中温连接酶。此类实施方案可以适用于用序列的文库实践所述方法,所述序列的文库包含一个被设计为具有一个预测熔解温度的索引序列集合,而文库内的第二个索引序列集合可以被设计为具有不同的熔解温度。

[0074] 寡核苷酸组分的浓度是影响连接准确度的反应参数。寡核苷酸组分的高浓度增加了错误连接结果的可能性。因此,较低的浓度可以导致连接产物的准确度提高。

[0075] 总体上,可以改变杂交条件、杂交序列的长度和连接条件以实现最高的准确组装构建体的产率。例如,如果条件集合导致特定组分集合的不准确组装,则可以使用更高的连接温度和/或更严格的杂交条件。鉴于本公开文本和本领域的常规知识,优化杂交条件和连接酶条件在熟练技术人员的范围内。参见例如,Sambrook等人,Molecular Cloning A Laboratory Manual,第2版Cold Spring Harbor Press(1989)或Anderson,Nucleic Acid Hybridization,第一版,BIOS Scientific Publishers Limited(1999)。

[0076] 在一个方面,所述方法可以使用图2的索引夹板进行,并且进一步包括至少一个阻断核苷酸。图6展示了寡核苷酸结构,其中所述寡核苷酸结构包含与组装组分的组装序列中的单链区域杂交的至少一个阻断寡核苷酸。此变型被考虑为防止非预期的杂交。可以设计阻断寡核苷酸以阻断单链区域的一部分或全部。在图6所描绘的实施方案中,寡核苷酸结构包含两条另外的寡核苷酸链,一条与组装组分1序列杂交,并且一条与组装组分2序列杂交。考虑了阻断寡核苷酸可以合成为图6中描绘的索引夹板分子的两个寡核苷酸分子的一部分。阻断寡核苷酸也可以用于图5的寡核苷酸结构或本发明的其他实施方案中。考虑了阻断寡核苷酸可以合成为图5中描绘的索引夹板分子的一部分。

[0077] 所述方法可以迭代地进行。在一个方面,下游组装组分包含在3'末端处的可去除的(可拆卸的)索引序列。在杂交和连接步骤以提供包含组装组分的连接产物之后,去除(例如,经由酶促切割或切割可切割接头)索引序列,使得连接产物的3'末端序列是组装序列的3'序列。然后,此第一连接产物可用作所述方法的另一次迭代中的上游组装组分,以产生第二连接产物。反应混合物包含与第一连接产物和另一个下游组装组分杂交的索引夹板分子。

[0078] 在一个方面,可去除的(或可拆卸的)的索引序列包含限制性酶位点。在一个方面,限制性酶位点用于IIS型限制性酶。IIS型限制酶包含特定的一组酶,所述特定的一组酶识别不对称DNA序列并且在其识别序列之外的限定距离处切割,通常在1至20个核苷酸内。IIS型限制性酶的这种特定切割作用模式允许切割,使得保留所希望的组装序列并且去除非组装序列。这允许DNA操作,所述DNA操作不改变最终组装序列并且因此能够在多组分构建体

中产生无疤连接点。

[0079] 在一个方面,可去除的(或可拆卸的)索引序列可以通过可切割接头连接至组装序列的索引序列。光可切割接头化学是本领域已知的并且可从TriLink Biotechnologies、Integrated DNA Technologies和Glen Research商购。

[0080] 在一个方面,所述方法是4部分组装,如图7至图9所展示,其中组装过程在第一组装的连接产物上重复进行。在此方面,有两个连接步骤,由限制性酶消化步骤分开。杂交步骤形成如图7中所描绘的两个寡核苷酸结构。在图7中所描绘的顶部寡核苷酸结构中,寡核苷酸结构呈图5B中所展示的取向(但它可以被设计为呈图5C中所描绘的取向)。索引夹板分子从5'至3'包含与下游组装组分2杂交以形成索引2的第一索引序列、夹板序列(夹板1)和与上游组装组分1杂交以形成索引1的第二索引序列。夹板1与上游组装组分1的3'末端序列和下游组装组分2的5'末端序列杂交。除了第一夹板结构(夹板1)外,顶部寡核苷酸结构的索引夹板进一步包含在间隔节段中的第二夹板序列(夹板2)和在5'索引序列(索引2)与第一夹板结构(夹板1)之间的另一个索引序列(以形成索引3)。下游组装组分2包含另外的结构特征,所述结构特征是在组装序列的3'处的限制性酶序列(描绘于图8中)。

[0081] 在图7所描绘的底部寡核苷酸结构中,寡核苷酸结构呈图5C所展示的取向(但它可以被设计为呈图5B中的取向)。索引夹板分子从5'至3'包含与上游组装组分3杂交以形成索引4的第一索引序列、夹板序列(夹板3)和与下游组装组分4杂交以形成索引5的第二索引序列。夹板3与上游组装组分3的3'末端序列和下游组装组分4的5'末端序列杂交。底部寡核苷酸结构的索引夹板进一步包含在5'索引序列(索引4)与夹板3之间的索引3的互补序列(索引3互补序列)。

[0082] 连接步骤产生组分1和2的连接产物(组装产物(1-2))和组分3和4的第二连接产物(组装产物(3-4)),分别如图8的结构(a)和(c)所展示。如图8所展示,顶部结构(a)的索引夹板包含在组装组分2的组装序列的3'处并且在索引2的5'处的限制性酶位点(RE切割位点)。结构(c)的索引夹板包含在组装组分3的组装序列的5'处并且在索引3的3'处的限制性酶位点。索引夹板包含两个另外的结构特征:位于夹板2与索引4的索引序列之间的索引序列(其为形成索引3的互补序列)以及在索引4的索引序列的5'处的限制性酶位点。使连接产物经受IIS型限制性酶消化,以切割组装产物1-2和组装产物3-4中的限制性酶位点,以分别去除索引2和索引4。因此,在所述方法的另一次迭代中,连接产物1-2的消化产物可以是上游组装组分,并且连接产物3-4的消化产物可以是下游组装组分。可以将消化产物纯化,例如通过SPRI纯化(例如,SPRI代表固相可逆固定),然后经受杂交步骤。杂交步骤产生图9所展示的寡核苷酸结构。如所展示,通过两个索引夹板分子的索引3序列的杂交形成双链索引3。夹板2作用于与组装组分杂交以形成连接位点;5'元件结合组装产物(3-4)的5'末端,并且3'元件结合组装产物(1-2)的3'末端。进行第二连接步骤,产生多组分连接产物1-2-3-4(未示出)。第二连接步骤的产物可以任选地进行SPRI纯化并且经由PCR扩增。

[0083] 如本文所示,基于索引的连接(其中两个组装组分的相对末端与具有两个互补索引序列的共同索引夹板分子杂交)提供了良好的特异性。例如,索引元件的熔解温度( $T_m$ )可以不高两个杂交的夹板元件中的每一个的熔解温度( $T_m$ )。以这种方式,索引结合在杂交条件下是稳定的,而夹板结合是亚稳定的。因此,特别设计的索引序列可以驱动杂交反应并且避免夹板序列与反应混合物中的非预期靶标之间的不希望的结合。

[0084] 对于甚至更大的产物的组装,可以重复相同的过程。

[0085] 进一步考虑了可能存在其中特异性可以通过所公开的方法的变型而改善的情况。下面描述用于增加索引序列的熔解温度并且因此增加特异性的方法。此方面也取决于两步连接过程。简言之,如图10-图13所展示,在第一连接步骤中,两个组装组分寡核苷酸与包含两个彼此相邻的索引序列但不包含夹板区域(参见图10中的“索引夹板”)的分子杂交,并且在索引序列处连接在一起,从而在第一连接步骤中形成长得多的寡核苷酸(参见图11)。在第一连接步骤(连接索引序列的末端)之后,可以将温度升高至高于具有其互补夹板序列的索引序列的熔解温度,并且在此较高温度下添加第二分子(与两个组装靶标杂交的“构建体夹板”) (参见图12B),从而允许更大的严格性,减少同源构建体之间交叉杂交的机会,并且因此潜在地甚至比先前描述的实施方案更好的特异性。每个分子的第二连接事件发生,而第一连接的连接产物的游离末端与夹板寡核苷酸杂交并且随后连接形成单链环(参见图13)。为了稳定环的形成,第二夹板的一个节段可以比另一部分长得多,从而在一端形成稳定的双链体,而另一端仅需要与双链体末端短暂接触以便使其变成连接的,从而形成环。另外,可以通过在第一连接步骤后通过添加更多缓冲溶液将溶液稀释来抑制嵌合体形成。这增加了通过增加溶液的熵而发生分子内连接和经分子间连接的环形成的可能性。

[0086] 如图10所展示,两个组装组分通过在一个组装组分的5'末端处的索引序列和在另一个组装组分的3'末端处的索引序列而与夹板分子杂交。连接后,连接产物包含两个组装组分,其中这两个索引序列在连接产物的中间彼此相邻并且连接。参见图11。连接后,连接产物通过现在连接的两个索引序列(索引1-索引2)保持与索引夹板杂交,并且具有显著增加的熔解温度。例如,在一个方面,其中每个索引序列是30个核苷酸长的并且每个索引元件具有约69°C的熔解温度,双链体中两个索引序列的组合熔解温度比约86°C的平均温度高约17°C。这意味着反应溶液的温度可以升高至大于80度但小于85度(10w 80s)而不使双链体变性的温度。

[0087] 任选地,在所述方法的步骤之间,可以通过例如以下方式从反应溶液中去除未结合的片段或未连接的产物:将杂交溶液洗掉,以清洁反应溶液。任选地,这种清洁可以通过将索引夹板寡核苷酸与固体支持物结合进行。例如,索引夹板寡核苷酸可以具有附接至其3'末端(如图11所示)或位于其5'末端的生物素。然后可以将寡核苷酸双链体与通常用于靶标富集的磁性链霉亲和素珠结合。这些需要在管中混合,孵育几分钟,然后借助强梯度磁场拉到管的侧面或底部,如通常用SPRI珠进行。然后,可以通过以下方式完成洗涤步骤:取出杂交溶液,并且重新填充容器并且用洗涤缓冲液在管中冲洗连接产物一次或多次。这将降低未连接的寡核苷酸的浓度并且降低第二连接步骤中不正确连接事件的可能性。

[0088] 可以通过添加水或更严格的连接缓冲液或通过根据缓冲液视情况而定加热至高于熔解温度而从双链体(并且因此从珠)上释放连接产物。可以将热稳定的连接酶,诸如9° N™连接酶(NEB,摩洛哥伊普斯威奇),添加至连接酶缓冲液中,并且将温度升高至80°C或更高或升高至另一种热稳定的连接酶保持活性的最高温度,此时可以将用于连接产物的游离末端的夹板分子添加至管中。可替代地,可以在加热之前将夹板添加至混合物中,并且在加热后添加连接酶。此步骤可以用在溶液中的寡核苷酸进行或在它们保持与磁珠结合的情况下进行。此第二连接事件描绘于图12和图13中。在连接反应的时间段后,可以通过添加EDTA或简单地通过快速冷却至室温或在室温下使连接反应停止。

[0089] 图13示出了寡核苷酸上任选的另外的结构特征,包括例如引物结合位点(PBS1和BPS1)和任选的限制性酶位点。可以通过PCR或通过滚环扩增进行扩增。通过向管中添加与环状寡核苷酸上的任一引物结合位点互补的单一引物和具有高持续能力(processivity)的DNA聚合酶(诸如噬菌体phi29 DNA聚合酶)来进行滚环扩增。可替代地,可以通过聚合酶、一对引物和dNTP进行PCR,然后在热循环仪中进行热循环。

[0090] 有时在组装产物的PCR扩增过程中,未连接的寡核苷酸可以通过聚合酶延伸并且产生不需要的嵌合体。图12和图13中所示的产生环化DNA的实施方案可以通过添加外切核酸酶步骤来减少此嵌合体的产生。在此另外的步骤中,外切核酸酶可以用于从其5'末端消化单链DNA和双链DNA,同时留下环状DNA组装体。外切核酸酶步骤应在PCT扩增之前进行。可以通过添加EDTA或通过将反应混合物加热至75°C来杀死酶或二者兼有而使消化过程停止。这些环状构建体随后通过PCR或通过使用进行性聚合酶诸如phi29的滚环扩增来扩增。

[0091] 在一个方面,省略了图10中连接两个组装组分的第一连接步骤。第二夹板(“构建体夹板”)两个组装组分的5'和3'游离末端与构建体序列杂交,然后将其连接。

[0092] 在与图10-图13相关的另一个实施方案中,索引夹板和构建体夹板借助接头分子或序列(图12A中的“任选的系链”)连接。所述分子具有图5A的索引夹板的组分,但以不同的5'至3'顺序。

[0093] 在本公开文本的一些实施方案中,多组分构建体可以根据本公开文本在溶液中第一次组装,并且在组装后通过本领域已知的任何方法附接至固体基底。因此,本公开文本可以用于制备多组分构建体的阵列,其中所述多组分构建体在组装后附接至阵列基底。随后,多组分构建体或多组分构建体的一个池或多个池可以任选地且选择性地从阵列基底上切割并且用作一个文库或多个文库。可替代地,多组分构建体可以附接至珠上。构建体可以通过化学手段或通过特异性杂交附接至珠上。取决于附接的方法,可以使用不同的珠来捕获特定的不同构建体,或者一个珠可以捕获多个构建体。文库与珠的附接可以提供储存、洗涤、浓缩或酶促操作文库的优点。

#### 试剂盒

[0094] 本公开文本还提供了可用于进行所公开的方法的试剂盒。在一个方面,提供了含有用于进行上述方法的试剂的试剂盒,所述试剂盒包含寡核苷酸文库和DNA连接酶,所述寡核苷酸文库包含单链寡核苷酸的混合物,所述混合物包含至少一种类型的上游组装组分、一种类型的下游组装组分和用于引导组装组分的连接的索引夹板。在某些实施方案中,所述试剂盒包含一种或多种其他反应组分。在某些实施方案中,将适当量的一种或多种反应组分提供在一个或多个容器中或保持在基底上。试剂盒的另外的组分的例子包括但不限于反应缓冲液、用于检测连接产物的一种或多种试剂(例如,探针或PCR引物)等。所使用的反应组分可以以各种形式提供。例如,可以将组分(例如,酶、寡核苷酸、探针和/或引物)悬浮在水溶液中或与珠结合或作为冷冻干燥或冻干的粉末或粒料。本公开文本的试剂盒可以在任何合适的温度下提供。例如,为了储存含有在液体中的蛋白质组分或其复合物的试剂盒,优选地将它们提供并且维持低于0°C(优选在约-20°C),可能在含有甘油或其他合适的防冻剂的抗冻溶液中。

[0095] 试剂盒或系统可以含有足以用于至少一次测定的量的本文所述组分的任何组合。在一些应用中,一种或多种反应组分可以以预先测量的单次使用量提供在单独的(典型地

是一次性的)管或等效容器中。试剂盒中提供的组分的量可以是任何适当的量并且可以取决于产品所针对的市场。在其中提供组分的一个或多个容器可以是能够保持所提供的形式的任何常规容器,例如,微量离心管、微量滴定板、安瓿、瓶或集成测试装置(诸如流体装置(fluidic device)、柱筒、侧向流动)或其他类似装置。

[0096] 试剂盒还可以包含用于保持容器或容器组合的包装材料。用于此类试剂盒和系统的典型包装材料包括固体基质(例如,玻璃、塑料、纸、箔、微颗粒等),所述固体基质将反应组分或检测探针保持在多种构造中(例如,在小瓶、微量滴定板孔、微阵列等中)的任一种中。试剂盒可以进一步包含以有形形式记录的组分使用说明书。

[0097] 所公开的组装方法取决于索引夹板分子的索引序列与组装组分分子之间稳定且特异性的杂交。相比之下,利用夹板引导组装的现有技术方法取决于夹板分子与两个组装组分之间的稳定且特异性的杂交。不受理论的束缚,认为本公开文本的方法(图5A-图5C)与现有技术方法之间的根本区别在于,在所公开的方法中,不使用希望的组装序列使待组装的碎片处在用于连接的位置。所公开的组装方法中的夹板序列被设计得太短而不能稳定杂交,并且因此不足以允许组装不存在索引杂交。索引夹板分子的索引序列不是构建体序列的一部分,因此不受文库成员之间的同源性影响。与之前的多路组装的高通量方法相比,这是出乎意料的益处和改善。

#### 实施例

[0098] 根据以下实施例可以进一步理解本传授内容的方面,这些实施例不应被解释为以任何方式限制本传授内容的范围。

#### 实施例1

[0099] 设计并且进行以下实验以评价使用图1、图2和图3中描绘的寡核苷酸夹板结构在溶液中制备低同源性文库。因此,评价了三种类型的测试夹板分子。术语“测试夹板分子”是指图1中的单个夹板分子。在图2和图3中,“测试夹板分子”是指通过互补索引序列杂交以形成索引元件而形成的两个寡核苷酸结构。在图2和图3中的索引元件是相同的并且具有在60℃至75℃范围内的 $T_m$ 。杂交的索引的高 $T_m$ 有效地使两个核苷酸结构在低于 $T_m$ 的温度下成为单个分子。

[0100] 构建体序列取自低同源性基因组序列UTR(非翻译区域)的集合。组装一百四十(140)种UTR。通过设计测试的夹板分子,正确构建体是连接至相同UTR序列的UTR序列。因此,对于140种不同的UTR,有140种不同的测试夹板分子以及总共140种可能组装的正确构建体和总共19,460种可能的不正确构建体。典型地,140种UTR中的每一种在原始寡核苷酸文库中具有两种形式:一种形式在5'末端具有引物位点,并且另一种形式在3'末端具有引物位点,用于PCR扩增和引入MiSeq(Illumina)所需的衔接子。

[0101] 对于每种类型的测试夹板分子,测试三种测试夹板分子与构建体的比率(1:1、2:1和4:1)。对于图2和图3的测试夹板分子(二者都具有索引),还测试了 $a+/-3$ 个核苷酸的间隔物(称为支柱)。3个核苷酸的支柱位于测试夹板分子的索引部分与夹板部分之间。结果,在此实验中组装了十五个集合。每个集合具有140种正确UTR构建体,其可以从420种不同的UTR序列组装。通过测序可区分来自不同集合的产物,其中在DNA文库材料的产生中,对于每个子集,不同的序列打印不同的分子拷贝数。

[0102] 对于每个集合,制备UTR和测试夹板分子的寡核苷酸文库。将约80ng文库溶解(在

适合于激酶反应和连接酶反应的缓冲液中),并且然后根据制造商的推荐使用T4多核苷酸激酶(NEB)进行激酶反应,以便将DNA寡核苷酸的5'末端磷酸化。通过在水浴中孵育来杂交激酶产物,所述水浴经1.75小时从93°C冷却至44°C,以形成图1、图2和图3所描绘的寡核苷酸结构。然后将杂交产物加热至连接温度并且添加1微升( $\mu$ l)9°N<sup>TM</sup>连接酶(New England Biolabs Inc.,摩洛哥伊普斯威奇)。连接反应进行1小时。将连接产物使用1.5xAMPure® XP珠(Beckman Coulter,加利福尼亚州布瑞亚市)纯化并且通过PCR扩增。通过NGS(下一代测序)对PCR产物进行测序,以表征每个集合中正确组装的程度。

### 结果

[0103] NGS读段数据示出于图14中。数据显示所有三种类型的测试夹板分子(图1、图2或图3中描绘的寡核苷酸结构)在所有15个集合中都起作用。图1的测试夹板分子表现最佳。特别地,其中夹板/构建体分子比率为4:1的集合在所测试的三个比率中表现最好,对于正确组装构建体而言具有接近700的平均读段数。图2和图3的数据彼此相似,其中图2的数据显示正确组装构建体的读段平均数略高。图2测试夹板分子中3个核苷酸的支柱的存在对图2测试夹板分子几乎没有影响。图3测试夹板分子中3个核苷酸的支柱的存在导致相对于没有支柱的测试夹板分子的数据略微减少的平均读段。

### 实施例2

[0104] 为了检查与实施例1中的那些不同的测试夹板分子,设计并且进行以下实验。此实验使用如图5B的寡核苷酸结构中所展示的测试夹板分子。测试夹板分子是索引夹板分子,所述索引夹板分子是单个寡核苷酸序列,其含有位于测试夹板分子的5'和3'末端并且在结合两个组装组分的夹板区域侧翼的索引序列。制备组装组分(包含待连接以形成组装构建体的组装序列)以进一步包括测试夹板分子的索引序列的反向互补序列,并且由此使得测试夹板分子能够在两个位置与组装组分杂交。对于左组装组分,测试夹板分子的结合是:(1)在与左组装组分的3'末端处的特定序列结合的夹板元件处,和(2)在左组装组分的5'末端处的索引序列处(杂交以形成索引元件)。对于右组装组分,测试夹板分子的结合是:(1)在与右组装组分的5'末端处的特定序列结合的夹板元件处,和(2)在右组装组分的5'末端处的索引序列处(杂交以形成另一个索引元件)。左组装组分的3'末端处的索引元件具有与右组装组分的3'末端处的索引元件不同的序列,因此在两个单独的索引元件之间没有交叉杂交。左组装组分的5'末端处的索引元件由30个核苷酸组成并且具有在65°C-72°C的范围内的T<sub>m</sub>,以及右组装组分的3'末端处的索引元件由30个核苷酸组成并且具有在65°C-72°C范围内的T<sub>m</sub>。

[0105] 构建体序列(2250个序列)取自低同源性基因组序列UTR(非翻译区域)的集合。组装序列的长度为200个核苷酸。组装组分包含UTR构建体序列、索引序列、条形码和引物序列。如图15所展示,左组装组分(左)从5'至3'包含索引序列(当与索引夹板杂交时以形成索引1)、引物序列(引物F)、条形码(BC1)和组装序列。右组装组分(右)从5'至3'包含组装序列、条形码(BC2)、引物序列(引物R)和索引序列(当与索引夹板杂交时以形成索引2)。

[0106] 测试夹板分子从5'至3'包含索引序列(与右组装组分杂交以在寡核苷酸结构中形成索引2)、间隔物(虚线)、夹板序列、另一个间隔物(虚线)、和另一个索引序列(杂交以在寡核苷酸结构中形成索引1)。夹板序列含有与右组装组分的组装序列的5'末端序列杂交的5'序列元件和与右组装组分的组装序列的3'末端序列杂交的3'序列。夹板的5'序列元件和3'

序列元件中的每一个是至少12个核苷酸长的,并且各自具有大于或等于28°C的T<sub>m</sub>。(使用公式1计算: $T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7$ )。

[0107] 通过设计实验中利用的夹板分子,正确构建体是与相同UTR连接的UTR,产生400个核苷酸的多组分构建体(不包括非构建体序列)。每个UTR组装组分被设计为进一步包含6个独特的条形码对。正确构建体由在组装的构建体中正确配对的左和右条形码指示。因此,对于每种UTR,有六种可能的正确构建体(和30种可能的不正确构建体)。由于测试了2250种不同的UTR序列和6种独特的条形码对,因此总共有13,500种独特的索引夹板分子(2250x 6)以及可能有13,500种独特的正确构建体。

[0108] 制备含有UTR组分构建体和索引夹板分子的寡核苷酸文库。将约80ng文库溶解,并且然后根据制造商的推荐使用T4多核苷酸激酶(NEB)进行激酶反应,以便将DNA寡核苷酸的5'末端磷酸化。通过在水浴中孵育来杂交激酶产物,所述水浴经1.75小时从93°C冷却至44°C。然后将杂交产物加热至连接温度60.6°C并且添加1微升(μl)9°N™连接酶(New England Biolabs Inc.,摩洛哥伊普斯威奇)。连接反应进行10分钟。将连接产物使用1.5x AMPure® XP珠(Beckman Coulter,加利福尼亚州布瑞亚市)纯化并且通过PCR扩增。通过NGS(下一代测序)对PCR产物进行测序,以表征正确组装的程度。

#### 结果

[0109] 数据示出于图15的表中。在13,500种可能的正确构建体中,通过对少于10,000,000个读段的文库测序而检测到13,465种。基于位于组装体的相对末端的独特条形码,此数据显示在反应中组装了99.70%可能的正确构建体,但不表明条形码之间的每个序列的全长序列准确度。

[0110] 因此,图5B的寡核苷酸结构实现了值得注意地多的数量的正确构建体。这对于寡核苷酸文库的制备是非常重要的数字。

#### 实施例3

[0111] 设计并且进行以下实验以测试使用图1、图2和图5B的寡核苷酸结构的组装。设计实验以比较三(3)种类型的寡核苷酸文库:(a)复杂文库、(b)拼块文库(tiled library)和(c)同源文库。图16中示意性地描绘了3种类型的文库。

[0112] 在此实验的复杂文库中,除外围引物位点外,序列具有很小的序列相似性。复杂文库(a型)具有600种不同的UTR序列(即,总共600种可能的正确组装构建体)。同源文库(c型)利用60种不同的UTR序列,并且每种不同的UTR序列具有10种独特条形码化的构建体(总共600种可能的正确组装构建体)。在此实验的拼块文库中,每个成员与3个其他成员的序列同一性重叠。对于拼块文库(b型),对于150种不同的UTR序列中的每一种有4种拼块构建体(具有82-246个碱基对的重叠序列同一性)(总共600种可能的正确组装构建体)。对于每个文库,组装的构建体是400个碱基长的。

[0113] 用每种类型的测试夹板分子(参见图1、图2和图5B的寡核苷酸结构)测试每种类型的文库,得到如表1所示的9个测试类别。

表1

类别	索引夹板	文库类型
1a	图1	复杂文库 (类型a)
1b		拼块文库 (类型b)
1c		同源文库 (类型c)
2a	图2	复杂文库 (类型a)
2b		拼块文库 (类型b)
2c		同源文库 (类型c)
3a	图5B	复杂文库 (类型a)
3b		拼块文库 (类型b)
3c		同源文库 (类型c)

[0114] 九个类别都在相同的文库中并且组装在一起。通过NGS(下一代测序)对PCR产物进行测序,以表征正确组装的程度。

#### 结果

[0115] 数据示出于图17中。读段总数的数据(正确的和不正确的)描绘在左上图中。这些数据显示,同源文库(c型)中图1和图2的索引夹板导致绝大多数的不正确组装构建体。百分比准确度数据(即,相对于读段的总数,正确读段的总数)描绘在左下图中。所有三种类型的索引夹板分子在复杂文库(a型)中均产生高准确度,并且在拼块文库(b型)中产生中等水平的准确度。仅图5B的夹板分子对于同源文库(c型)具有高准确度。在右上图中描绘了中值正确读段数据。在右下图中描绘了辍失构建体的数量的数据。“辍失构建体”是指没有正确读段的正确构建体(每个文库中有600种可能的正确构建体)。这些数据显示了图1和图2组装的索引夹板对于每种类型文库而言具有辍失构建体。例如,图1索引夹板显示了对于复杂文库(a型)而言超过50个辍失构建体,对于拼块文库(b型)而言超过60个辍失构建体以及对于同源文库(c型)而言接近70个辍失构建体。对于图5B索引夹板,对于任何文库而言都没有辍失构建体。

[0116] 在此数据中观察到图5B数据的两个值得注意的方面。首先,使用图5B索引夹板分子的组装方法是唯一可以有用的准确度水平组装c型(复杂文库)文库类型的方法(参见左下图)。第二,使用图5B索引夹板分子的组装方法没有辍失(参见右下图)。在图5B的组装中组装了所有600种可能的正确组装构建体。相比之下,其他两种方法具有50-80个辍失,即,构建体不是由600种可能的正确构建体构建的。

[0117] 这些数据指示了使用图5B中描绘的索引夹板分子对正确组装的影响,并且证明了构建体的正确组装,这是原本由于夹板分子的混杂杂交而不可能的。

#### 实施例4

[0118] 设计并且进行以下实验以测量两个实验变量在利用如图5B所示的索引夹板分子组装构建体中的作用。

[0119] 构建体序列取自低同源性基因组序列UTR(非翻译区域)的集合。为夹板序列的各一半设计十六(16)个可变长度,按熔解温度( $T_m$ ,如由 $T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7$ 计算)组织。对于夹板的一半,制定14°C至44°C(按2°C变化)的温度阈值。对于每个夹板设计,组装四十(40)种UTR独特序列(16X 40=640种总UTR序列)。将十(10)种独特的条形码用于每一种UTR序列(总共6,400种可能的正确构建体)。如果条形码在组装的构建体中正确配对(通过NGS测量),则称为“正确”对。组装组分进一步包含正向引物(上游组装组分)和反向

引物(下游组装组分)的序列以允许PCT扩增。在8个不同的连接温度下组装文库。图18描绘了用于此实验的图5B测试夹板分子的示意图。

[0120] 制备UTR组装组分和各种测试夹板分子的寡核苷酸文库。将约80ng文库溶解(在适合于激酶反应和连接酶反应的缓冲液中),并且然后根据制造商的推荐使用T4多核苷酸激酶(NEB)进行激酶反应,以便将DNA寡核苷酸的5'末端磷酸化。通过在水浴中孵育来杂交激酶产物,所述水浴经1.75小时从93°C冷却至44°C,以形成图5B中所描绘的寡核苷酸结构。然后将杂交产物加热至一个测试连接温度并且添加1微升( $\mu$ l)9°N™连接酶(New England Biolabs Inc.,摩洛哥伊普斯威奇)。连接反应进行1小时。将连接产物使用1.5x AMPure® XP珠(Beckman Coulter,加利福尼亚州布瑞亚市)纯化并且通过PCR扩增。通过NGS(下一代测序)对PCR产物进行测序,以表征正确组装的程度。

### 结果

[0121] 数据总体上证明,组装的特异性在较高的连接温度下增加。数据总体上还证明,具有短夹板长度(较低 $T_m$ )的组装构建体的效率在高连接温度下降低。代表性数据描绘于图18的图中。数据展示了在两个夹板区段的 $T_m$ 范围内,利用图5B中的索引夹板分子高程度的可能正确组装。

### 示例性实施方案

[0122] 产物

实施方案1.一种用于组装两个核酸组分以产生多组分连接产物的寡核苷酸结构,所述寡核苷酸结构包含第一单链寡核苷酸、第二单链寡核苷酸和索引夹板,其中所述第一单链寡核苷酸和所述第二单链寡核苷酸与所述索引夹板部分地杂交。

实施方案2.根据实施方案1所述的寡核苷酸结构,其中

(a)所述索引夹板包括夹板序列,以及在所述夹板序列侧翼的第一索引序列和第二索引序列;

(b)所述夹板序列与所述第一单链寡核苷酸的3'末端和所述第二单链寡核苷酸的5'末端二者杂交;以及

(c)所述第一索引序列与所述第一单链寡核苷酸的5'末端杂交以形成第一双链索引元件,并且所述第二索引序列与所述第二单链寡核苷酸的3'末端杂交以形成第二双链索引元件。

实施方案3.根据实施方案2所述的寡核苷酸结构,其中所述索引夹板的夹板序列包含:

(a)与所述第二单链寡核苷酸的5'末端杂交(反向互补)的5'元件,以及

(b)与所述第一单链寡核苷酸的3'末端杂交(反向互补)的3'元件,

其中所述夹板序列的5'元件和3'元件是连续的,即没有插入核苷酸。

实施方案4.根据实施方案3所述的寡核苷酸结构,其中(a)的5'杂交夹板元件和(b)的3'杂交夹板元件中的每一个具有约14°C至约44°C的溶解温度( $T_m$ ),其中

$$T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7。$$

实施方案5.根据实施方案2、3和4中任一项所述的寡核苷酸结构,其中:

(1)所述第一索引序列位于所述索引夹板的3'末端,以及

(2)所述第二索引序列位于所述索引序列的5'末端。

实施方案6.根据实施方案5所述的寡核苷酸结构,其中:

(a)所述第一单链寡核苷酸从5'至3'包含:

- (1)构成所述第一单链寡核苷酸(左组装组分)的5'末端序列的第三索引序列,和
- (2)待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸(左组装组分)的3'末端序列;以及

(b)所述第二单链寡核苷酸从5'至3'包含:

- (1)待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和

(2)构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及

(c)所述索引夹板从5'至3'包含:

- (1)5'末端序列,所述5'末端序列是所述第二索引序列,
- (2)任选的第一间隔序列,
- (3)所述夹板序列,

其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交(反向互补),并且

其中所述夹板序列的3'元件与第一单链寡核苷酸的3'末端序列杂交(反向互补),并且

(4)任选的第二间隔序列,以及

(5)3'末端序列,所述3'末端序列是所述第一索引序列。

实施方案7.根据实施方案6所述的寡核苷酸结构,其中每个间隔序列独立地由1至30个核苷酸组成。

实施方案8.根据实施方案2、3和4中任一项所述的寡核苷酸结构,其中:

(1)所述第一索引序列位于所述索引夹板的5'末端,以及

(2)所述第二索引序列位于所述索引序列的3'末端。

实施方案9.根据实施方案8所述的寡核苷酸结构,其中:

(a)所述第一单链寡核苷酸从5'至3'包含:

(1)构成所述第一单链寡核苷酸的5'末端序列的第三索引序列,和

(2)待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸(左组装组分)的3'末端序列;以及

(b)所述第二单链寡核苷酸从5'至3'包含:

(1)待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和

(2)构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及;以及

(c)所述索引夹板从5'至3'包含:

- (1)5'末端序列,所述5'末端序列是所述第一索引序列,
- (2)任选的第一间隔序列,
- (3)所述夹板序列,

其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交(反向互补),并且

其中所述夹板序列的3'元件与第一单链寡核苷酸(左组装组分)的3'末端序列杂交(反向互补),并且

(4)任选的第二间隔序列,以及

(5)3'末端序列,所述3'末端序列是所述第二索引序列。

实施方案10.根据实施方案8或9所述的寡核苷酸结构,其中每个间隔序列独立地由1至30个核苷酸组成。

实施方案11.根据实施方案2至10中任一项所述的寡核苷酸结构,其中每个双链索引元件包含10至40个碱基对。

实施方案12.根据实施方案2至11中任一项所述的寡核苷酸结构,其中每个双链索引元件具有约10°C至约80°C的溶解温度( $T_m$ ),其中

$$T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7.$$

实施方案13.根据实施方案6、7、9、10和11中任一项所述的寡核苷酸结构,其中所述第四索引序列(区段F)可去除地(或可拆卸地)连接至所述第二DNA序列组分。

实施方案14.根据实施方案13所述的寡核苷酸结构,其中所述第二单链寡核苷酸进一步包含在第二DNA序列组分的3'处的II型限制酶位点,其中所述限制性酶切割位点紧邻所述第二DNA序列组分的3'。

实施方案15.根据实施方案13所述的寡核苷酸结构,其中所述第二单链寡核苷酸通过可切割接头可去除地(或可拆卸地)连接至所述第二DNA序列组分,其中切割位点紧邻所述第二DNA序列组分的3'。

实施方案16.根据实施方案2至15中任一项所述的寡核苷酸结构,所述寡核苷酸结构进一步包含至少一个与区段D或区段E的序列区段杂交(反向互补)的阻断寡核苷酸组分,其中所述序列区段不包括与所述夹板杂交的序列。

实施方案17.一种根据实施方案1至16中任一项所述的寡核苷酸结构的集合或文库。

实施方案18.一种包含根据实施方案17所述的集合或文库的试剂盒。

实施方案19.根据实施方案18所述的试剂盒,所述试剂盒进一步包含DNA连接酶。

#### [0123] 方法

实施方案20.一种用于组装两个核酸组分以产生多组分连接产物的方法,所述方法包括:

(A)使包含第一DNA序列组分的第一单链寡核苷酸和包含第二DNA序列组分的第二单链寡核苷酸与索引夹板杂交,其中所述第一单链寡核苷酸的5'末端和3'末端两端均与所述索引夹板杂交,并且所述第二单链寡核苷酸的5'末端和3'末端两端均与所述索引夹板杂交;以及

(B)将所述第一单链寡核苷酸的3'末端与所述第二单链寡核苷酸的5'末端连接,由此产生第一多组分连接产物。

实施方案21.根据实施方案20所述的方法,其中:

(a)所述索引夹板进一步包括夹板序列,以及在所述夹板序列侧翼的第一索引序列和第二索引序列;

(b)所述夹板序列与所述第一单链寡核苷酸的3'末端和所述第二单链寡核苷酸的

5'末端二者杂交;以及

(c)所述第一索引序列与所述第一单链寡核苷酸的5'末端杂交以形成第一双链索引元件,并且所述第二索引序列与所述第二单链寡核苷酸的3'末端杂交以形成第二双链索引元件。

实施方案22.根据实施方案21所述的方法,其中所述索引夹板的夹板序列包含:

- (a)5'元件,其与所述第二单链寡核苷酸的5'末端杂交形成双链5'夹板元件,以及
- (b)3'元件,其与所述第一单链寡核苷酸的3'末端杂交形成双链3'夹板元件,其中所述夹板序列的5'元件和3'元件是连续的,即没有插入核苷酸。

实施方案23.根据实施方案22所述的方法,其中所述双链5'夹板元件和所述双链3'夹板元件各自具有约10°C至约80°C的溶解温度(T<sub>m</sub>),其中

$$T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7.$$

实施方案24.根据实施方案21、22或23所述的方法,其中:

- (1)所述第一索引序列位于所述索引夹板的3'末端,以及
- (2)所述第二索引序列位于所述索引夹板的5'末端。

实施方案25.根据实施方案24所述的方法,其中:

- (a)所述第一单链寡核苷酸从5'至3'包含:

(1)构成所述第一单链寡核苷酸的5'末端序列的第三索引序列,和  
(2)待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸的3'末端序列;以及

- (b)所述第二单链寡核苷酸从5'至3'包含:

(1)待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和

- (2)构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及

- (c)所述索引夹板从5'至3'包含:

- (1)5'末端序列,所述5'末端序列是所述第二索引序列,
- (2)任选的第一间隔序列,
- (3)所述夹板序列,

其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交,并且其中所述夹板序列的3'元件与所述第一单链寡核苷酸的3'末端序列杂交,

- (4)任选的第二间隔序列,以及

- (5)3'末端序列,所述3'末端序列是所述第一索引序列。

实施方案26.根据实施方案25所述的方法,其中每个间隔序列独立地由1至30个核苷酸组成。

实施方案27.根据实施方案21、22或23所述的方法,其中:

- (1)所述第一索引序列位于所述索引夹板的5'末端,以及
- (2)所述第二索引序列位于所述索引夹板的3'末端。

实施方案28.根据实施方案27所述的方法,其中:

- (a)所述第一单链寡核苷酸从5'至3'包含:

- (1)构成所述第一单链寡核苷酸的5'末端序列的第三索引序列,和

(2) 待连接到所述多组分连接产物中的第一DNA序列组分,所述第一DNA序列组分构成所述第一单链寡核苷酸的3'末端序列;以及

(b) 所述第二单链寡核苷酸从5'至3'包含:

(1) 待连接到所述多组分连接产物中的第二DNA序列组分,所述第二DNA序列组分构成所述第二单链寡核苷酸的5'末端序列,和

(2) 构成所述第二单链寡核苷酸的3'末端序列的第四索引序列;以及

(c) 所述索引夹板从5'至3'包含:

(1) 5'末端序列,所述5'末端序列是所述第一索引序列,

(2) 任选的第一间隔序列,

(3) 所述夹板序列,

其中所述夹板序列的5'元件与所述第二单链寡核苷酸的5'末端序列杂交,并且

其中所述夹板序列的3'元件与所述第一单链寡核苷酸的3'末端序列杂交,

(4) 任选的第二间隔序列,以及

(5) 3'末端序列,所述3'末端序列是所述第二索引序列。

实施方案29. 根据实施方案27或28所述的方法,其中每个间隔序列独立地由1至30个核苷酸组成。

实施方案30. 根据实施方案21至29中任一项所述的方法,其中每个双链索引元件包含10至40个碱基对。

实施方案31. 根据实施方案21至30中任一项所述的方法,其中每个双链索引元件具有约10°C至约80°C的溶解温度( $T_m$ ),其中

$$T_m = (2 * (A_{\text{计数}} + T_{\text{计数}})) + (4 * (C_{\text{计数}} + G_{\text{计数}})) - 7.$$

实施方案32. 根据实施方案20至31中任一项所述的方法,所述方法进一步包括:

(C) 将来自所述多组分连接产物的索引夹板变性,并且

随后迭代地进行步骤(A)、(B)和(C)。

实施方案33. 根据实施方案20至32中任一项所述的方法,其中所述第四索引序列可去除地(或可拆卸地)连接至所述第二DNA序列组分。

实施方案34. 根据实施方案33所述的方法,其中所述第二单链寡核苷酸进一步包含在第二DNA序列组分的3'处的IIIs型限制性酶位点,其中限制性酶切割的位点紧邻所述第二DNA序列组分的3'。

实施方案35. 根据实施方案33所述的方法,其中所述第四索引序列通过可切割(或可拆卸)接头连接至所述第二DNA序列组分,其中切割位点紧邻所述第二DNA序列组分的3'。

实施方案36. 根据实施方案33至35中任一项所述的方法,所述方法进一步包括:

从所述第一连接产物中的第二DNA组分上去除所述第四索引序列,

提供具有所述第四索引序列的另一个第二单链寡核苷酸和包含第二夹板序列的另一个索引夹板,

随后进行步骤(A)、(B)和(C),

其中所述第一连接产物是步骤(A)的第一单链寡核苷酸,并且

其中所述另一个索引夹板的第二夹板序列含有与所述另一个第二单链寡核苷酸中的5'末端序列杂交的5'元件和与所述第一连接产物的3'末端序列杂交的3'元件。

实施方案37.根据实施方案33至35中任一项所述的方法,其中所述索引夹板进一步包含与第五索引序列相邻的第二夹板序列,其中所述第二夹板序列的5'元件可以与所述第二DNA序列组分的3'末端序列杂交。

实施方案38.根据实施方案20至32中任一项所述的方法,其中所述第三索引序列可去除地(或可拆卸地)连接至所述第一DNA序列组分。

实施方案39.根据实施方案38所述的方法,其中所述第一单链寡核苷酸进一步包含在第一DNA序列组分的5'的II<sub>s</sub>型限制性酶位点,其中所述限制性酶切割位点紧邻所述第一DNA序列组分的5'。

实施方案40.根据实施方案38所述的方法,其中所述第三索引序列通过可切割(或可拆卸)接头连接至所述第一DNA序列组分,其中切割位点紧邻所述第一DNA序列组分的5'。

实施方案41.根据实施方案38至40中任一项所述的方法,所述方法进一步包括:

从所述第一连接产物中的第一DNA序列组分上去除所述第三索引序列,

提供具有所述第三索引序列的另一个第一单链寡核苷酸和另一个索引夹板,

随后进行步骤(A)、(B)和(C),

其中所述第一连接产物是步骤(A)的第二单链寡核苷酸,并且

其中所述另一个索引夹板的夹板序列含有与所述另一个第一单链寡核苷酸中的3'末端序列杂交的3'元件和与在所述第一连接产物中的5'末端序列杂交(反向互补)的5'元件。

实施方案42.根据实施方案20至41中任一项所述的方法,其中:

所述杂交步骤在基于所述索引元件的熔解温度( $T_m$ )的温度范围内进行,并且

所述连接步骤在高于所述夹板序列的5'和3'元件的熔解温度( $T_m$ )并且低于所述索引元件的熔解温度( $T_m$ )的温度下进行。

实施方案43.根据实施方案20至42中任一项所述的方法,其中所述连接步骤的温度为约45°C至70°C,并且所述连接酶是热稳定的。

实施方案44.根据实施方案20至43中任一项所述的方法,其中所述索引元件的熔解温度为约60°C至90°C。

实施方案45.根据实施方案20至44中任一项所述的方法,其中所述5'夹板元件和所述3'夹板元件的熔解温度为约20°C至42°C。

[0124] 对示例性或优选实施方案的前述描述应当被视为是说明性的而不是限制性的,本公开文本应当被视为如权利要求所限定。如将容易理解的,在不脱离如权利要求中所阐述的本公开文本的情况下,可以利用以上阐述的特征的许多变型和组合。这种变型不被视为脱离本公开文本的范围,并且所有这种变型都旨在包括在以下权利要求的范围内。将本文引用的所有参考文献都通过引用以其整体并入。

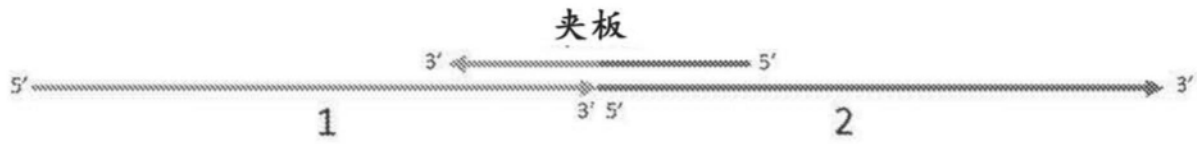


图1

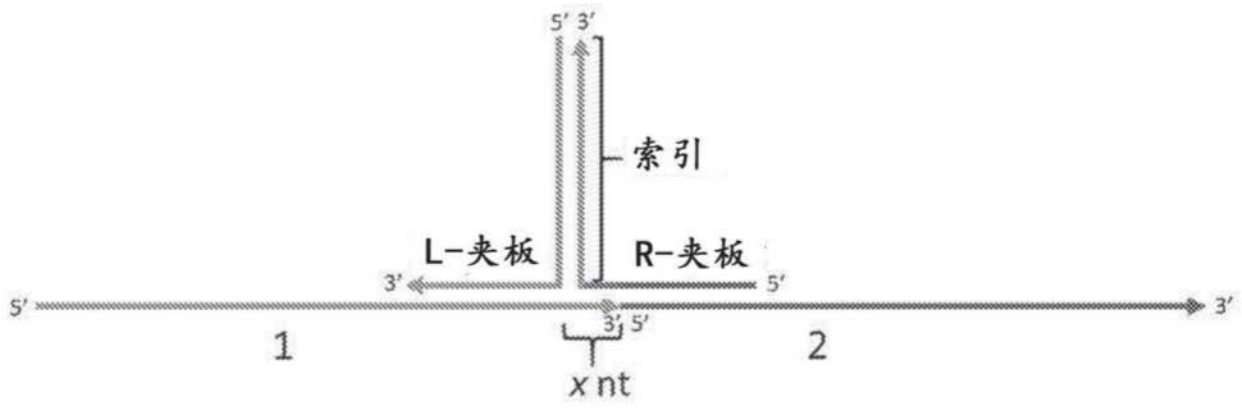


图2

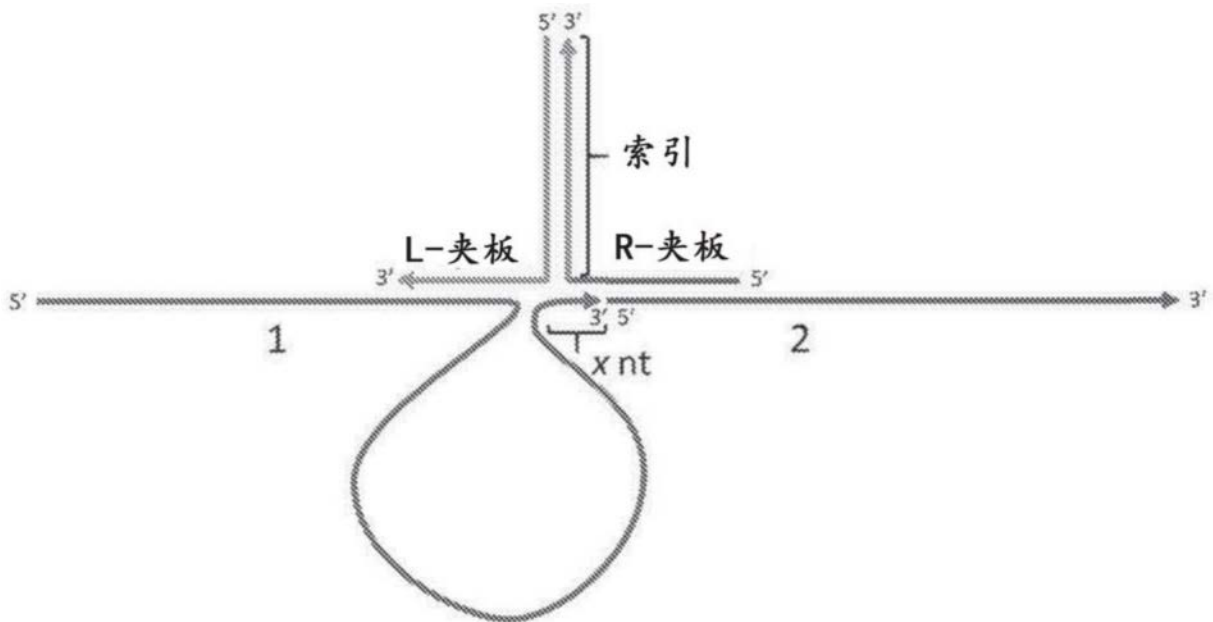


图3

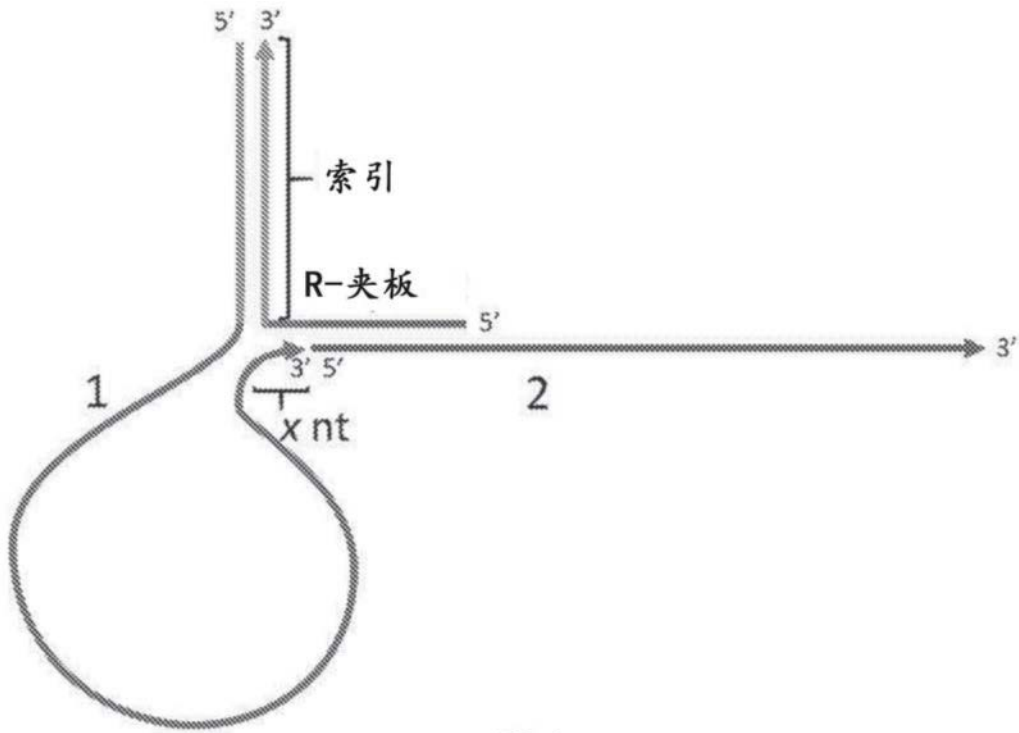


图4

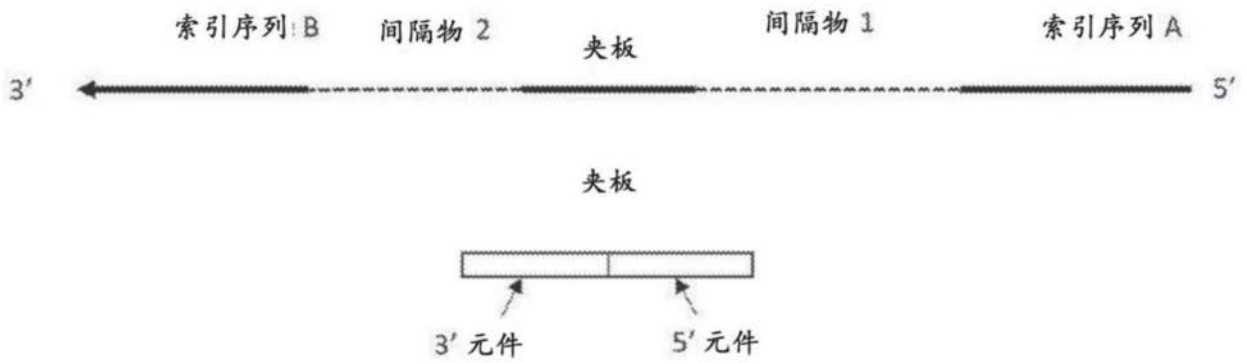


图5A

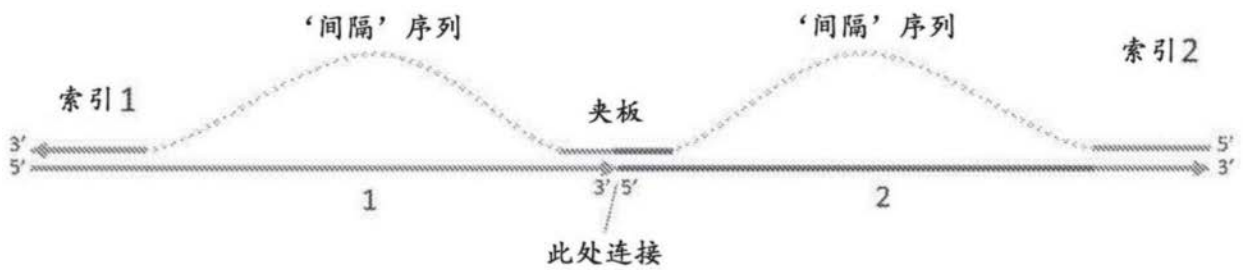


图5B

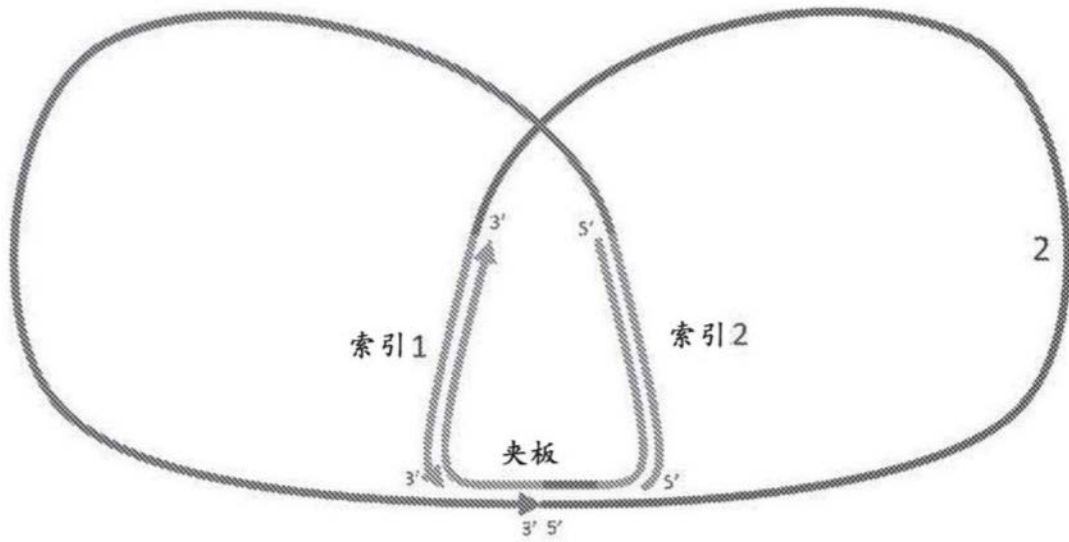


图5C

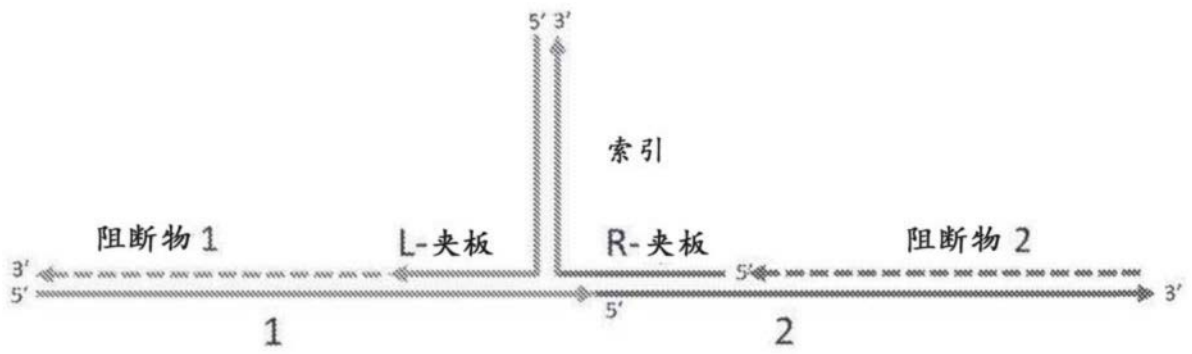


图6

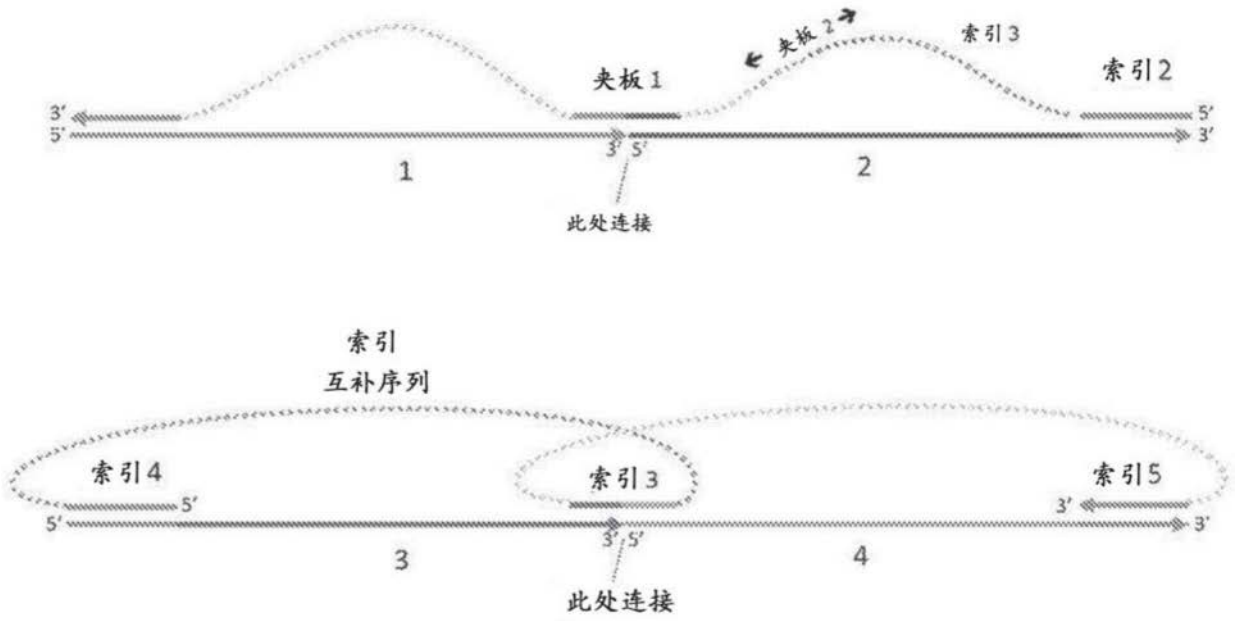


图7

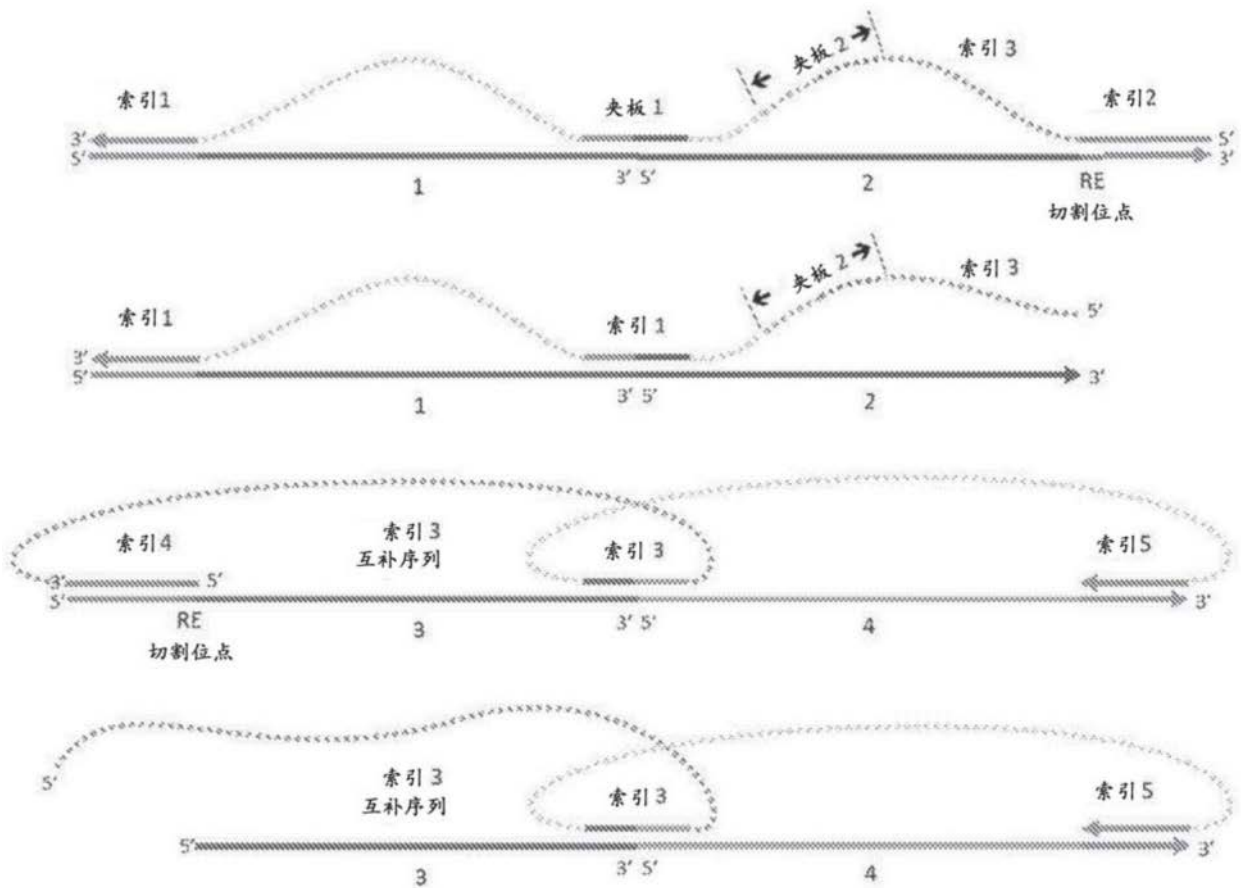


图8

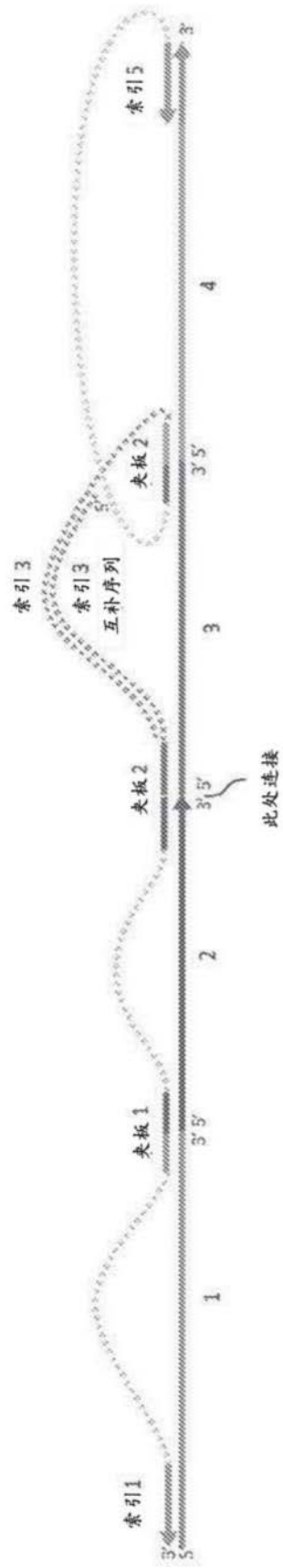


图9

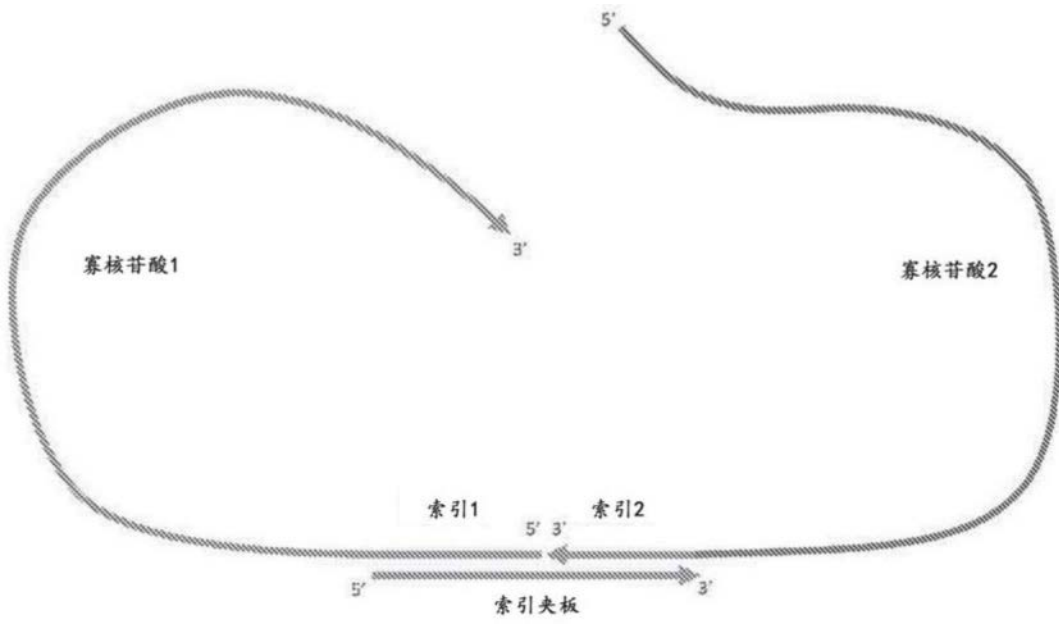


图10

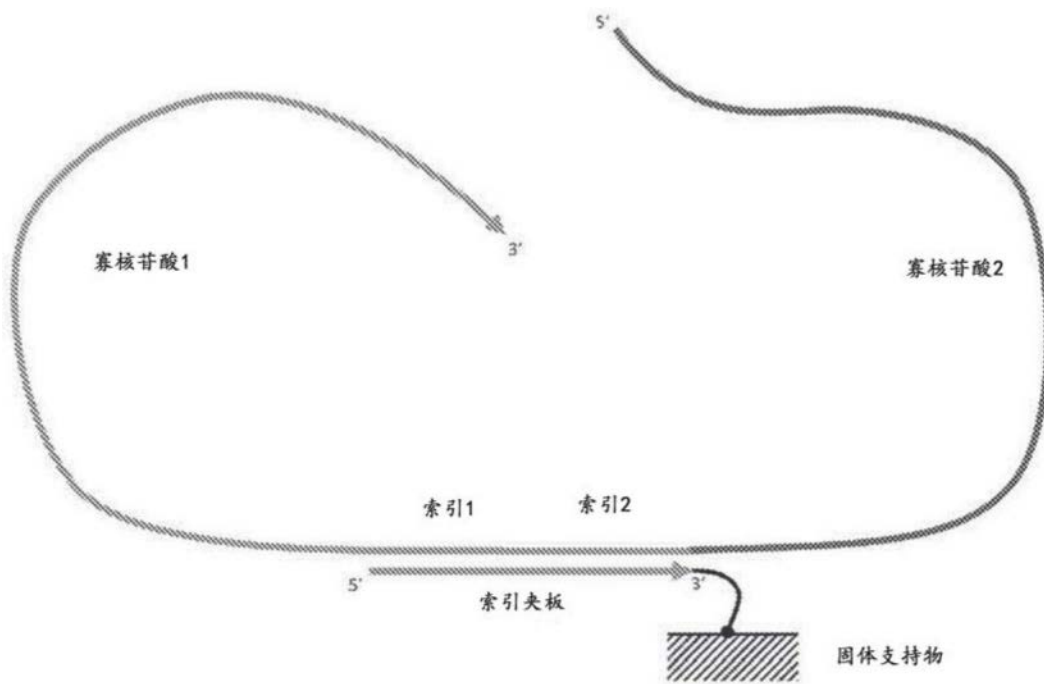


图11

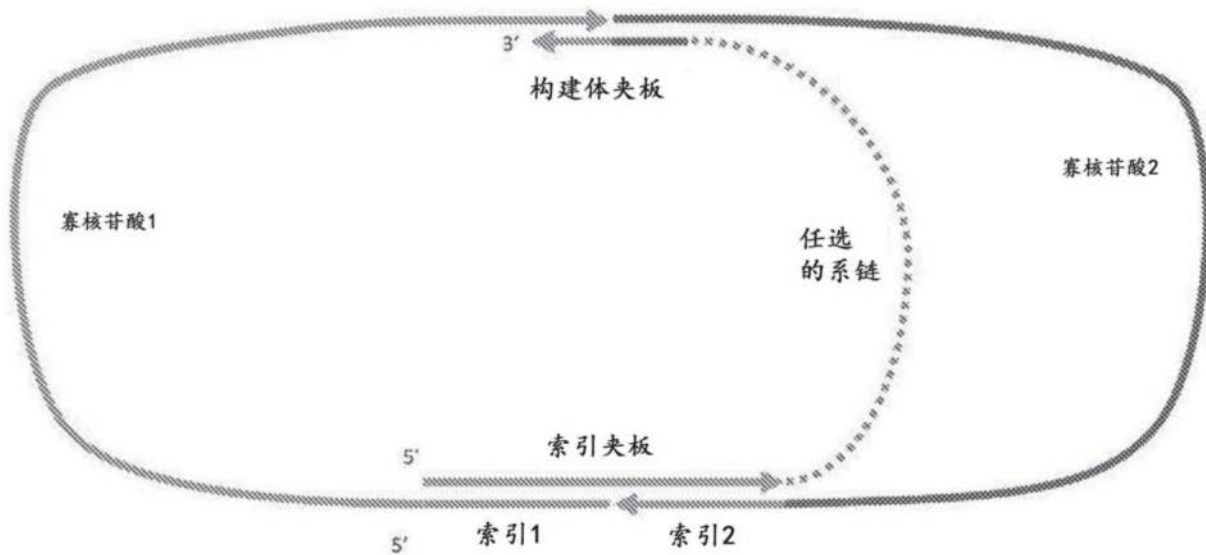


图12A

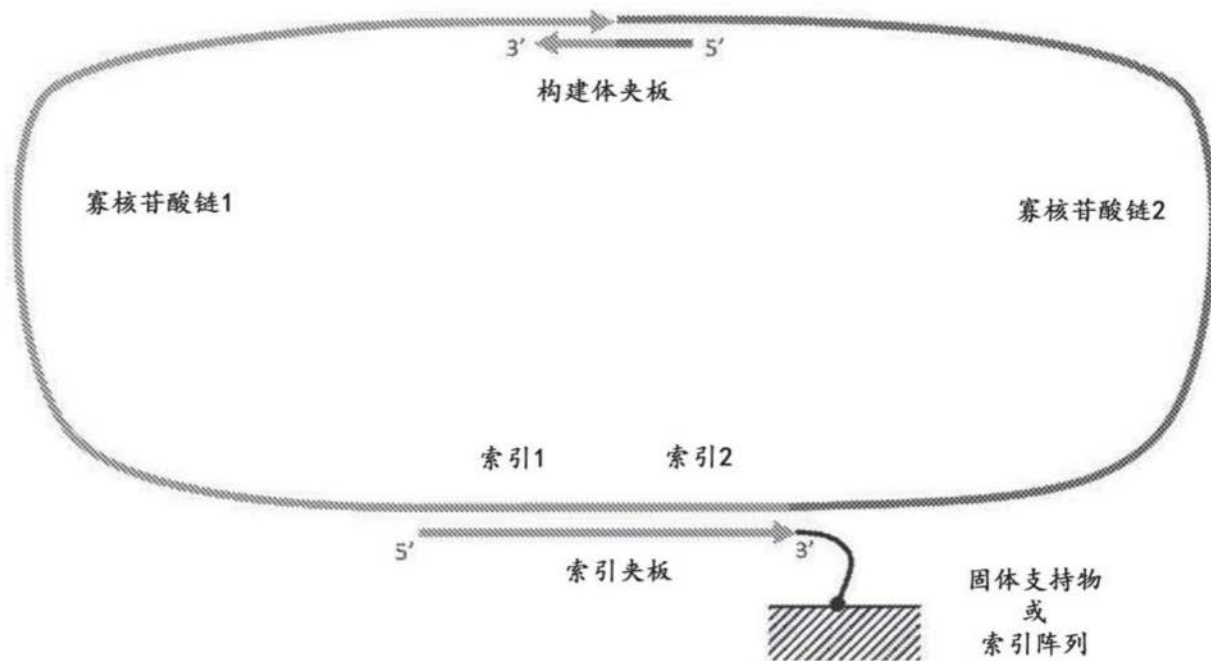


图12B

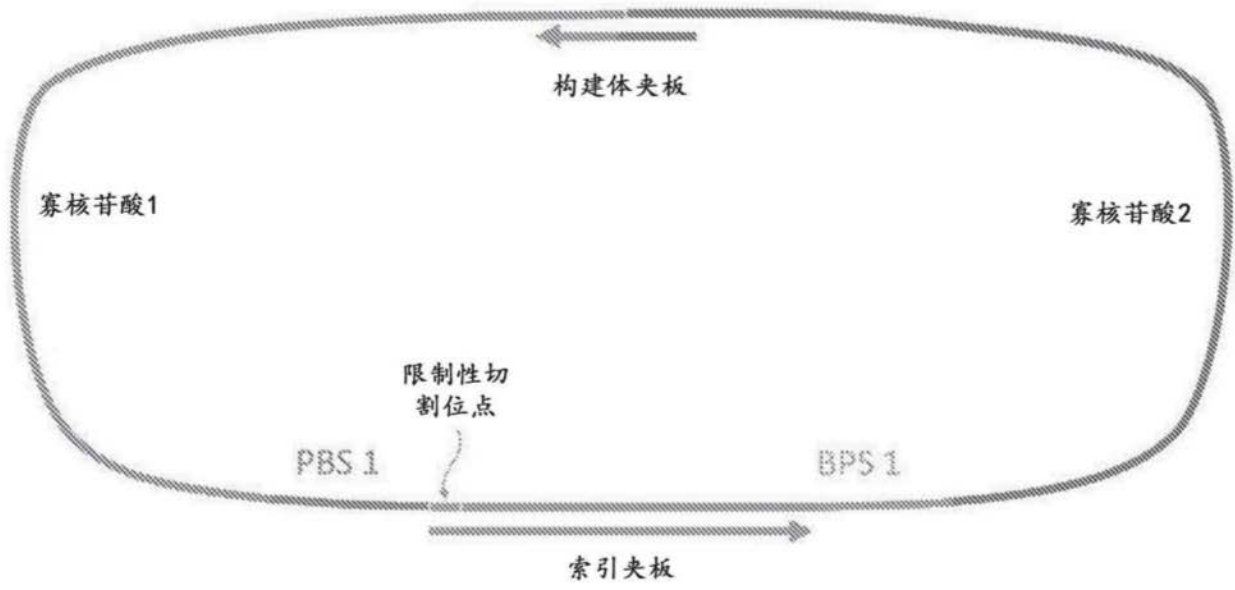


图13



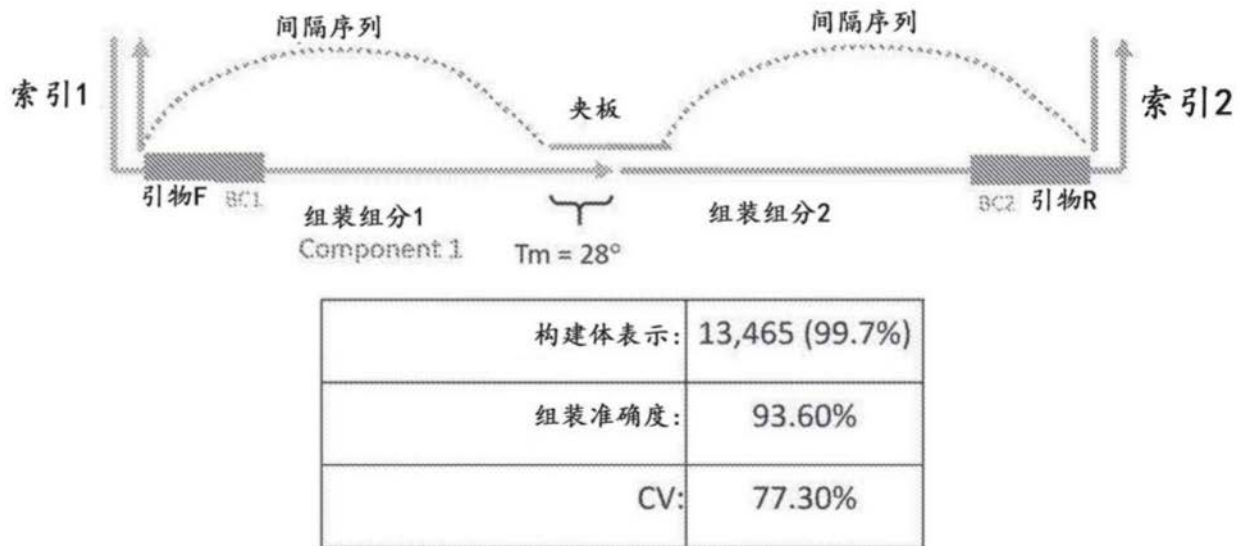
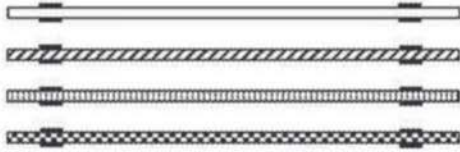


图15

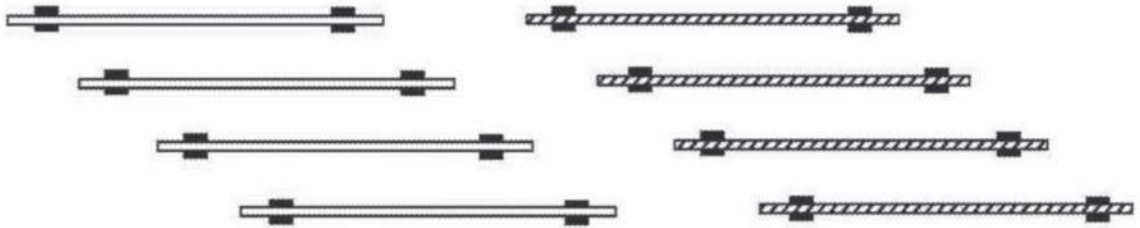
**类型a: 复杂文库**

文库序列除外围引物结合位点外具有很小的序列相似性  
(600种不同的UTR序列)。



**类型b: 拼块文库**

每个成员与3个其他成员重叠的序列同一性为82-246个碱基对。  
在此例子中对于150种不同的UTR序列中的每一种有4种平铺构建体。



**类型c: 同源文库**

在此例子中对于60种不同的UTR序列中的每一种有10种独特条形码化的构建体。

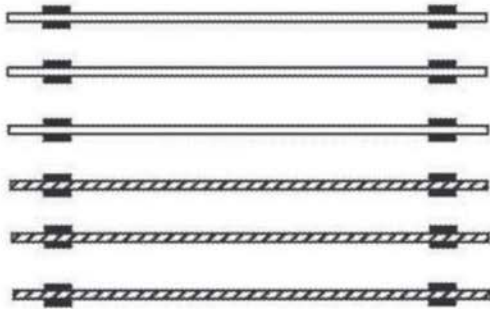


图16

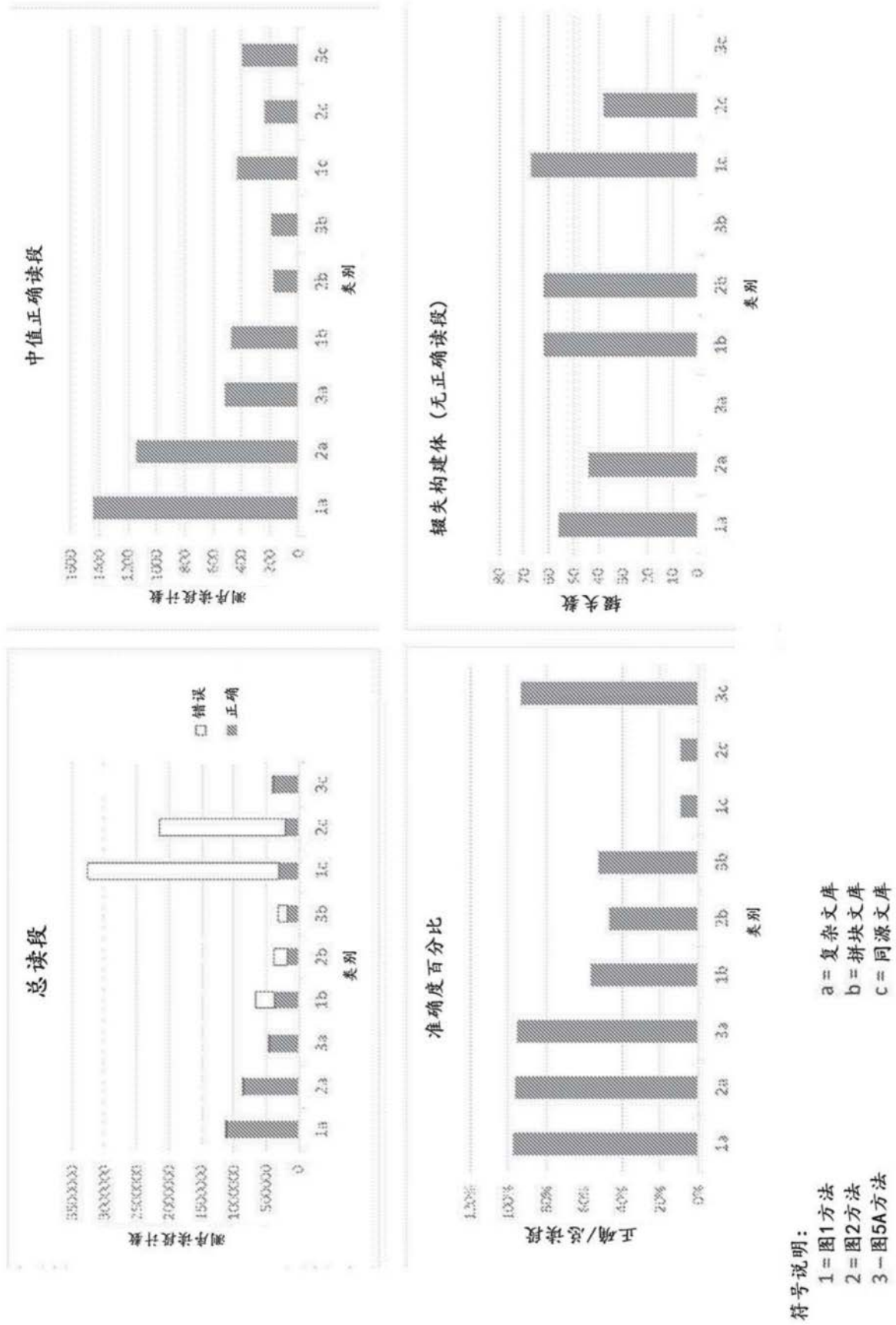


图17

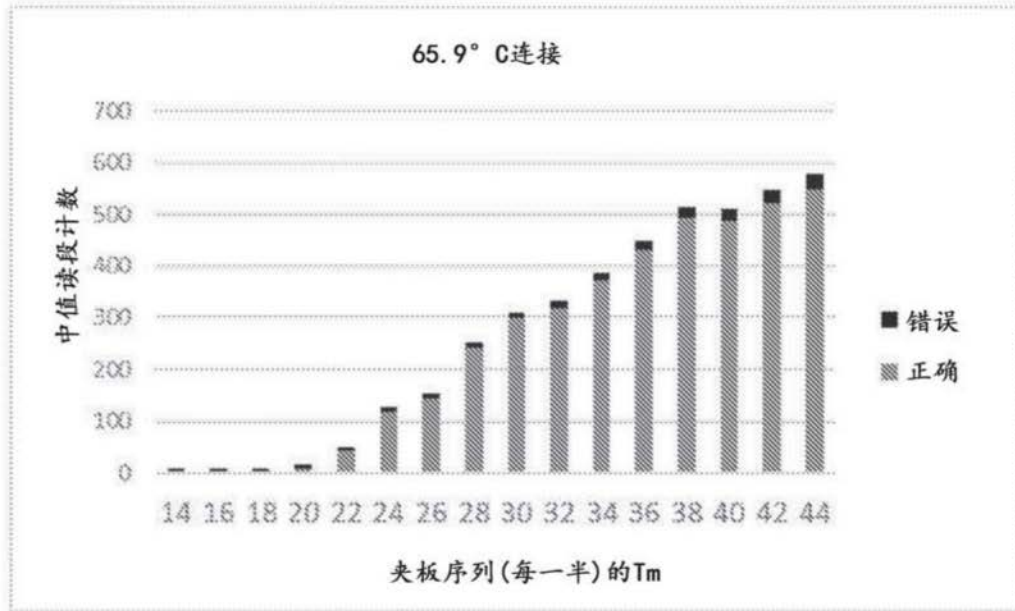
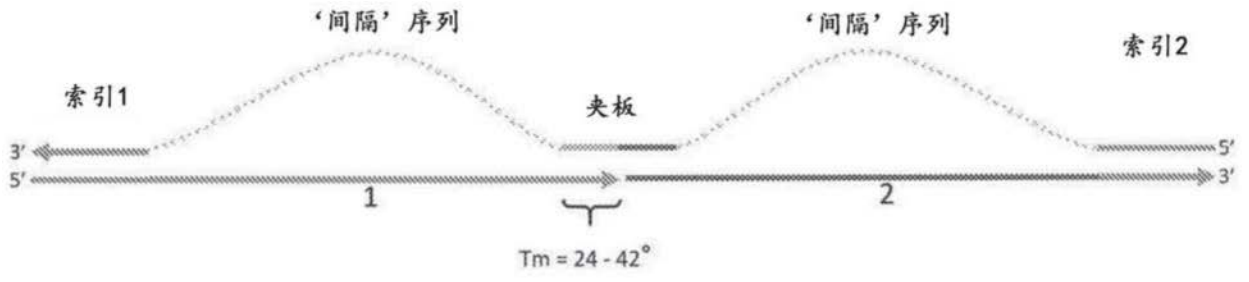


图18