



(12)发明专利申请

(10)申请公布号 CN 110257493 A

(43)申请公布日 2019.09.20

(21)申请号 201910257744.4

(51)Int.Cl.

(22)申请日 2013.07.12

C12Q 1/6869(2018.01)

(30)优先权数据

61/671,406 2012.07.13 US

(62)分案原申请数据

201380047931.0 2013.07.12

(71)申请人 X-化学有限公司

地址 美国马萨诸塞州

(72)发明人 安东尼·D·基夫

亚历山大·利托夫奇克

马修·A·克拉克

(74)专利代理机构 北京康信知识产权代理有限
责任公司 11240

代理人 沈敬亭

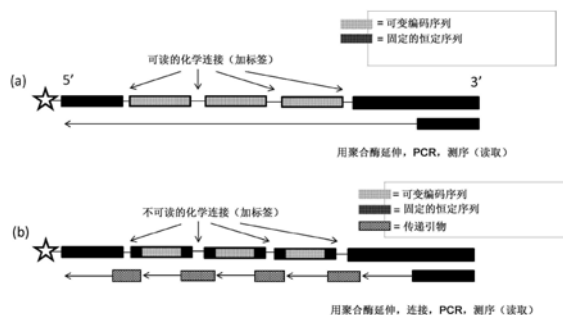
权利要求书3页 说明书44页 附图42页

(54)发明名称

具有聚合酶不可读的编码性寡核苷酸连接的DNA编码文库

(57)摘要

本发明涉及具有聚合酶不可读的编码性寡核苷酸连接的DNA编码文库。本发明涉及寡核苷酸编码文库的复合物以及使此类文库加标签和使用此类文库的方法。特别是,寡核苷酸和方法可以包括具有至少一个连接的复合物,对于该连接,聚合酶具有降低的通读或移位通过的能力。



1. 一种确定复合物的核苷酸序列的方法,所述方法包括:

(a) 提供包括化学连接的所述复合物,其中,所述化学连接是交联寡核苷酸,所述交联寡核苷酸跨越两个相邻标签之间的结点,并且所述交联寡核苷酸经由一个或多个可逆共反应性基团与所述两个相邻标签操作地关联;

(b) 释放所述交联寡核苷酸以产生模板;

(c) 可选地通过聚合酶链式反应扩增所述模板以产生扩增的混合物;和

(d) 测序所述可选扩增的混合物以确定所述复合物的序列,

其中所述复合物包括:

(i) 化学实体,其包括一个或多个支架或者一个或多个结构单元;

(ii) 第一寡核苷酸标签,其编码所述一个或多个支架或者结构单元中的至少一个的特性;和

(iii) 帽部分,其具有第一官能团和第二官能团,其中所述第一官能团与所述化学实体操作地关联且所述第二官能团经由第一连接与所述第一寡核苷酸标签操作地关联,对于所述第一连接,聚合酶具有降低的通读或移位通过的能力;或者

其中所述复合物包括:

(i) 化学实体,其包括一个或多个支架或者一个或多个结构单元;

(ii) 具有 $n-1$ 连接的 n 个数量的寡核苷酸标签,其中 n 是1和10之间的整数,并且其中每个所述连接是在两个相邻标签之间,并且每个标签编码所述一个或多个支架或者结构单元中的至少一个的特性;和

(iii) 帽部分,其具有与所述化学实体操作地关联的第一官能团和与所述 n 个数量的标签中的至少一个经由第一连接操作地关联的第二官能团,

其中,聚合酶具有降低的通读或移位通过所述第一连接和 $n-1$ 连接中的至少一个的能力。

2. 根据权利要求1所述的方法,其中,所述交联寡核苷酸的一部分杂交至两个相邻标签的末端,从而产生所述两个相邻标签之间的缺口或空位。

3. 根据权利要求2所述的方法,进一步包括在步骤 (b) 之前使用化学处理或酶处理连接所述标签。

4. 根据权利要求3所述的方法,其中,所述化学处理或酶处理包括5'-磷酸化。

5. 根据权利要求1至4中任一项所述的方法,进一步包括在所述可选扩增步骤或所述测序步骤之前修复所述模板。

6. 根据权利要求5所述的方法,其中,所述修复步骤包括将所述第一连接和/或 $n-1$ 连接中的至少一个修饰成为能够由聚合酶通读或移位通过的修复连接。

7. 根据权利要求6所述的方法,其中,所述聚合酶能够通读或移位通过至少90%的所述修复连接。

8. 根据权利要求5至7中任一项所述的方法,其中,所述修复步骤包括为所述模板提供修复酶。

9. 根据权利要求8所述的方法,其中,所述修复酶是光裂合酶、糖基化酶、核酸内切酶、Flap核酸内切酶、脱嘌呤/脱嘧啶 (AP) 核酸内切酶、聚ADP核糖聚合酶或甲基转移酶。

10. 根据权利要求9所述的方法,其中,所述聚合酶具有降低的通读或移位通过所述第

一连接和n-1连接的至少约10%的能力。

11. 根据权利要求1至10中任一项所述的方法,其中,一个或多个标签包括在所述一个或多个标签的5'-末端处的5'-连接子和在所述一个或多个标签的3'-末端处的3'-连接子。

12. 根据权利要求11所述的方法,其中,每个5'-连接子包括相同的序列。

13. 根据权利要求11或12所述的方法,其中,每个3'-连接子包括相同的序列。

14. 根据权利要求11所述的方法,其中,每个5'-连接子和/或每个3'-连接子包括不同的序列。

15. 根据权利要求1至14中任一项所述的方法,其中,所述第一连接和/或n-1连接的长度为至少约3埃。

16. 根据权利要求1至15中任一项所述的方法,其中,所述第一连接和/或n-1连接的长度小于约30埃。

17. 根据权利要求1至16中任一项所述的方法,其中,DNA聚合酶和/或RNA聚合酶具有降低的通读或移位通过所述第一连接和/或n-1连接的能力。

18. 根据权利要求1至17中任一项所述的方法,其中,少于约50%的所述第一连接和n-1连接包括酶连接。

19. 根据权利要求1至18中任一项所述的方法,其中,所述第一连接和/或n-1连接包括化学连接。

20. 根据权利要求19所述的方法,其中,所述化学连接包括化学反应性基团、光反应性基团、嵌入部分或交联寡核苷酸。

21. 根据权利要求20所述的方法,其中,至少一种化学反应性基团、光反应性基团或嵌入部分存在于在所述标签的5'-末端处或接近于所述标签的5'-末端处的5'-连接子中和/或存在于在所述标签的3'-末端处或接近于所述标签的3'-末端处的3'-连接子中。

22. 根据权利要求21所述的方法,其中,至少一个所述5'-连接子的序列与相邻3'-连接子的序列互补或相同或足够相似以允许杂交至在分离的杂交事件中的结点的任一侧上的相同的互补寡核苷酸。

23. 根据权利要求20至22中任一项所述的方法,其中,所述化学反应性基团选自一对可取代的炔基基团和可取代的叠氮基基团;一对具有4 π -电子体系的可取代的二烯和具有2 π -电子体系的可取代的亲二烯体或可取代的杂亲二烯体;一对亲核试剂和紧张的杂环亲电子体;一对可取代的氨基基团和醛或酮基团;一对可取代的氨基基团和羧酸基团;一对可取代的胍和醛或酮基团;一对可取代的羟胺和醛或酮基团;一对亲核试剂和可取代的烷基卤化物;铂复合物;烷基化试剂;或呋喃修饰的核苷酸。

24. 根据权利要求20至22中任一项所述的方法,其中,所述光反应性基团包括嵌入部分、补骨脂素衍生物、可取代的氰乙烯基呋唑基团、可取代的乙烯基呋唑基团、可取代的氰乙烯基基团、可取代的丙烯酰胺基团、可取代的双吡丙啶基团、可取代的二苯甲酮、可取代的5-(羧基)乙烯基-尿苷基团或可取代的叠氮基团。

25. 根据权利要求20至22中任一项所述的方法,其中,所述嵌入部分是补骨脂素衍生物、生物碱衍生物、乙吡啶阳离子、吡啶衍生物、蒽环类衍生物或沙利度胺。

26. 根据权利要求25所述的方法,其中,所述补骨脂素衍生物是补骨脂素、8-甲氧基补骨脂素或4-羟甲基-4,5,8-三甲基-补骨脂素。

27. 根据权利要求20所述的方法, 其中, 所述化学连接包括所述交联寡核苷酸, 其中, 在所述交联寡核苷酸的5'-末端处的至少5个核苷酸的序列与在一个或多个标签的3'-末端处的至少5个核苷酸的序列互补, 并且其中, 在所述交联寡核苷酸的3'-末端处的至少5个核苷酸的序列与在一个或多个标签的5'-末端处的至少5个核苷酸的序列互补。

28. 根据权利要求27所述的方法, 其中, 所述一个或多个标签的3'-末端包括3'-连接子。

29. 根据权利要求27或28所述的方法, 其中, 所述一个或多个标签的5'-末端包括5'-连接子。

30. 根据权利要求27至29中任一项所述的方法, 其中, 所述交联寡核苷酸的所述5'-末端和/或3'-末端包括可逆共反应性基团。

31. 根据权利要求28或29所述的方法, 其中, 所述3'-连接子和/或5'-连接子包括可逆共反应性基团。

32. 根据权利要求30或31所述的方法, 其中, 所述可逆共反应性基团是氰乙烯基咪唑基团、氰乙烯基基团、丙烯酰胺基团、硫醇基团或乙烯基砜基团。

33. 根据权利要求1至32中任一项所述的方法, 其中, 所述化学实体经由双官能间隔子操作地关联到所述帽部分。

34. 根据权利要求1至33中任一项所述的方法, 其中, 所述化学实体共价连接至所述帽部分。

35. 根据权利要求1至34中任一项所述的方法, 其中, 所述帽部分包括选自自由双链寡核苷酸、单链寡核苷酸或发夹寡核苷酸组成的组中的寡核苷酸。

36. 根据权利要求35所述的方法, 其中, 所述帽部分包括引物结合区。

37. 根据权利要求1至36中任一项所述的方法, 进一步包括一个或多个第一文库识别标签、用途标签和/或来源标签。

38. 根据权利要求1至37中任一项所述的方法, 其中, 所述复合物包括2至20之间个标签。

39. 根据权利要求1至38中任一项所述的方法, 其中, 每个所述标签包括从5至75个核苷酸。

40. 根据权利要求1至39中任一项所述的方法, 其中, 个体标签组内的每个所述标签包括大约相同的质量。

41. 根据权利要求1至40中任一项所述的方法, 其中, 所述复合物包括RNA、DNA、修饰的DNA和/或修饰的RNA。

42. 根据权利要求41所述的方法, 其中, 所述修饰的DNA或修饰的RNA是相同寡核苷酸内的PNA、LNA、GNA、TNA或它们的混合物。

43. 根据权利要求1至42中任一项所述的方法, 进一步包括尾部分。

具有聚合酶不可读的编码性寡核苷酸连接的DNA编码文库

[0001] 本申请是申请日为2013年7月12日的题为“具有聚合酶不可读的编码性寡核苷酸连接的DNA编码文库”的中国专利申请第201380047931.0号的分案申请。

[0002] 相关申请的参考

[0003] 本申请要求于2012年7月13日提交的美国临时申请号US 61/671,406的权益,其通过参考方式以其全部内容并入本文。

技术领域

[0004] 本发明总体上涉及化合物的DNA编码文库(DNA编码库,DNA编码分子库,DNA-encoded libraries)和使用并创建此类文库的方法。本发明也涉及用于在此类文库中使用的组合物。

背景技术

[0005] DNA编码组合文库提供用于药物发现的许多好处。这些文库可以提供大量的不同化合物,其可被快速筛选和询问。为了进一步增加复杂性,可以编程和自动化进行发现过程的各个步骤。这些步骤包括使用多步、拆分和合并(split-and-pool)合成以将结构单元(building block)添加至原子或多原子支架和使用酶学和/或化学连接以添加编码合成步骤和结构单元的寡核苷酸标签(oligonucleotide tags)。

[0006] 尽管有这些好处,但当必须合成且去卷积非常大或复杂的文库时,可出现许多问题。随着文库的大小增加,可能需要改善的方法以提供使用强大且快速的方法的高产量的编码寡核苷酸标签连接。为了在不同反应条件下创建文库,稳定的寡核苷酸构建体将是有益的,如在高pH和高温条件下稳定的构建体。为简化标签的去卷积,可以由DNA或RNA依赖性聚合酶来识别标签的序列,使得可以通过模板依赖性聚合和序列测定来确定标签群体(标签数量,tag population)统计。当创建具有所有这些有益属性的文库时,可能出现困难。因此,在寡核苷酸编码文库中,存在对于改善的、更强大的筛选和识别小化合物的方法的需要。

发明内容

[0007] 本发明的特征在于用于在DNA编码文库中使用的复合物,其中这些复合物具有一个或多个连接(键,linkage),对于其聚合酶具有降低的通读或移位通过(read or translocate through)的能力。为了发现编码区(例如,帽部分(帽结构,headpiece)、一个或多个标签和/或尾部分(尾结构,tailpiece)的特性(identity),该连接在它暴露于聚合酶之前是反向的或可替代地被聚合酶绕过。该连接可以任何有用的方式被绕过,该方式不可捕获标签的所有序列信息(例如,如在5'-或3'-连接子中,如本文所描述的),但捕获标签的编码性序列(encoding sequence)信息。此类连接可以获得错误加标签(mis-tagging)的较低发生率、扩展将要在制备复合物和文库中使用的连接的数量和种类并提供用于创建和筛选DNA编码的复合物的非酶方法。在本文中描述这些复合物和方法的另外的优点。

[0008] 因此,在一个方面,本发明的特征在于复合物,包括:化学实体,其包括一个或多个支架或者一个或多个结构单元;第一寡核苷酸标签,其编码一个或多个支架或者结构单元中的至少一个的特性;和帽部分,其具有第一官能团和第二官能团,其中第一官能团与化学实体操作地关联且第二官能团经由第一连接与第一标签操作地关联,对于该连接,聚合酶具有降低的通读或移位通过的能力。

[0009] 在另一个方面,本发明的特征在于复合物,包括:化学实体,其包括一个或多个支架或者一个或多个结构单元;具有 $n-1$ 连接的 n 个数量寡核苷酸标签,其中 n 是1和约10之间的整数,并且其中每个连接在两个相邻标签之间且每个标签编码一个或多个支架或者结构单元中的至少一个的特性;和帽部分,其具有与化学实体操作地关联的第一官能团和经由第一连接与 n 个数量的标签中的至少一个操作地关联的第二官能团,其中聚合酶具有降低的通读或移位通过第一连接和 $n-1$ 连接中的至少一个的能力。

[0010] 在一些实施方式中, n 在1和2、1和3、1和4、1和5、1和6、1和7、1和8、1和9、1和10、1和12、1和15、1和18、1和20、2和3、2和4、2和5、2和6、2和7、2和8、2和9、2和10、2和12、2和15、2和18、2和20、3和4、3和5、3和6、3和7、3和8、3和9、3和10、3和12、3和15、3和18、3和20、4和5、4和6、4和7、4和8、4和9、4和10、4和12、4和15、4和18或4和20之间。

[0011] 在一些实施方式中,聚合酶具有降低的通读或移位通过至少约10% (例如,与对照相比,约20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、或甚至100%) 的第一连接和 $n-1$ 连接的能力。在特定的实施方式中,聚合酶具有降低的通读或移位通过约10%至约100%的第一连接和 $n-1$ 连接的能力(例如,与对照相比(例如,与缺少该连接的对照寡核苷酸相比),20%至100%、25%至100%、50%至100%、75%至100%、90%至100%、95%至100%、10%至95%、20%至95%、25%至95%、50%至95%、75%至95%、90%至95%、10%至90%、20%至90%、25%至90%、50%至90%、或75%至90%)。

[0012] 在一些实施方式中,一个或多个标签包括在一个或多个标签的5'-末端处的5'-连接子(connector)和在一个或多个标签的3'-末端处的3'-连接子。在特定的实施方式中,每个5'-连接子和/或每个3'-连接子包括相同的序列。在其他实施方式中,每个5'-连接子和/或每个3'-连接子包括不同的序列。

[0013] 在一些实施方式中,第一连接和/或 $n-1$ 连接为至少约3埃的长度(例如,至少约5、8、10、15、20、30、35、40、50、60、65或70埃)。在一些实施方式中,第一连接和/或 $n-1$ 连接小于约30埃的长度(例如,小于约25、20、15或10埃)。

[0014] 在一些实施方式中,DNA聚合酶和/或RNA聚合酶(例如,如本文所描述的)具有降低的通读或移位通过第一连接和/或 $n-1$ 连接的能力。

[0015] 在一些实施方式中,小于约10% (例如,约10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%或95%) 的第一连接和 $n-1$ 连接包括酶连接。在一些实施方式中,第一连接和 $n-1$ 连接包括0%至90%之间的酶连接(例如,约0%至40%、0%至45%、0%至50%、0%至55%、0%至60%、0%至65%、0%至70%、0%至75%、0%至80%、0%至85%、0%至90%、0%至95%、0%至96%、0%至97%、0%至98%、0%至99%、5%至40%、5%至45%、5%至50%、5%至55%、5%至60%、5%至65%、5%至70%、5%至75%、5%至80%、5%至85%、5%至90%、5%至95%、5%至96%、5%至

97%、5%至98%、5%至99%、10%至40%、10%至45%、10%至50%、10%至55%、10%至60%、10%至65%、10%至70%、10%至75%、10%至80%、10%至85%、10%至90%、10%至95%、10%至96%、10%至97%、10%至98%、10%至99%、15%至40%、15%至45%、15%至50%、15%至55%、15%至60%、15%至65%、15%至70%、15%至75%、15%至80%、15%至85%、15%至90%、15%至95%、15%至96%、15%至97%、15%至98%、15%至99%、20%至40%、20%至45%、20%至50%、20%至55%、20%至60%、20%至65%、20%至70%、20%至75%、20%至80%、20%至85%、20%至90%、20%至95%、20%至96%、20%至97%、20%至98%或20%至99%)。

[0016] 在一些实施方式中,第一连接和/或n-1连接包括化学连接(例如,化学反应性基团、光反应性基团、嵌入(intercalate)部分或交联寡核苷酸)。在特定的实施方式中,至少一个(例如,两个、三个、四个、五个或更多个)化学反应性基团、光反应性基团或嵌入部分存在于在标签的5'-末端处或接近于标签的5'-末端的5'-连接子中和/或在标签3'-末端处或接近于标签3'-末端的3'-连接子中。在其他实施方式中,至少一个5'-连接子的序列与相邻3'-连接子的序列互补或相同或足够相似以允许杂交至互补寡核苷酸。在一些实施方式中、至少10%(例如,约20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或甚至100%)的第一连接和/或n-1连接是化学连接。在其他实施方式中,约10%至约100%的第一连接和n-1连接(例如,20%至100%、25%至100%、50%至100%、75%至100%、90%至100%、95%至100%、10%至95%、20%至95%、25%至95%、50%至95%、75%至95%、90%至95%、10%至90%、20%至90%、25%至90%、50%至90%或75%至90%)是化学连接。

[0017] 在一些实施方式中,化学反应性基团选自一对可选取代的炔基基团和可选取代的叠氮基基团;一对具有4 π -电子体系的可选取代的二烯和具有2 π -电子体系的可选取代的亲二烯体或可选取代的杂亲二烯体;一对亲核试剂和紧张的杂环亲电子体;一对可选取代的氨基基团和醛或酮基团;一对可选取代的氨基基团和羧酸基团;一对可选取代的肼和醛或酮基团;一对可选取代的羟胺和醛或酮基团;一对亲核试剂和可选取代的烷基卤化物;铂复合物;烷基化试剂;或呋喃修饰的核苷酸(例如,任何本文所描述的)。

[0018] 在一些实施方式中,光反应性基团包括嵌入部分、补骨脂素衍生物、可选取代的氰乙烯基呋唑基团(例如,3-氰乙烯基呋唑基团,如3-氰乙烯基呋唑-1'- β -脱氧核苷-5'-三磷酸酯)、可选取代的乙烯基呋唑基团(例如,氨基乙烯基呋唑基团、羧基乙烯基呋唑基团或C₂₋₇烷氧羰基乙烯基呋唑基团,如本文所描述的)、可选取代的氰乙烯基基团、可选取代的丙烯酰胺基团、可选取代的双吡丙啶(diazirine)基团、可选取代的二苯甲酮或可选取代的叠氮基团(例如,任何本文所描述的)。

[0019] 在一些实施方式中,嵌入部分是补骨脂素衍生物(例如,补骨脂素、8-甲氧基补骨脂素或4-羟甲基-4,5,8-三甲基-补骨脂素(HMT-补骨脂素))、生物碱衍生物(例如,黄连素、巴马汀、甲氧槲因、血根碱(例如,其亚胺或烷醇胺形式或马兜铃内酰胺- β -D-葡萄糖苷)、乙吡啶阳离子(例如,溴化乙锭)、吡啶衍生物(例如,原黄素、吡啶黄或安吡啶)、蒽环类衍生物(例如,多柔比星、表柔比星、柔红霉素(道诺霉素)、伊达比星和阿柔比星)或沙利度胺。

[0020] 在一些实施方式中,化学连接包括交联寡核苷酸,其中在交联寡核苷酸的5'-末端处的至少5个核苷酸的序列与在一个或多个标签的3'-末端处的至少5个核苷酸的序列互补

或相同或足够相似以允许杂交至互补寡核苷酸,并且其中在交联寡核苷酸的3'-末端处的至少5个核苷酸的序列与在一个或多个标签的5'-末端处的至少5个核苷酸的序列互补或相同或足够相似以允许杂交至互补寡核苷酸。在特定的实施方式中,一个或多个标签的3'-末端包括3'-连接子。在特定的实施方式中,一个或多个标签的5'-末端包括5'-连接子。

[0021] 在一些实施方式中,交联寡核苷酸的5'-末端和/或3'-末端包括可逆的共反应性基团(例如,氰乙烯基咪唑基团、氰乙烯基基团、丙烯酰胺基团、硫醇基团或乙烯基砜基团,如本文所描述的)。

[0022] 在一些实施方式中,3'-连接子和/或5'-连接子包括可逆的共反应性基团(例如,氰乙烯基咪唑基团、氰乙烯基基团、丙烯酰胺基团、硫醇基团或乙烯基砜基团,如本文所描述的)。

[0023] 在一些实施方式中,化学实体经由双官能间隔子(spacer)(例如,任何本文所描述的)操作地关联到帽部分。在其他实施方式中,化学实体共价地连接至帽部分。在特定的实施方式中,帽部分包括选自自由双链寡核苷酸、单链寡核苷酸或发夹寡核苷酸组成组的寡核苷酸。在一些实施方式中,帽部分包括引物结合区。

[0024] 在本文所描述的任何实施方式中,复合物或方法进一步包括一个或多个第一文库识别标签、用途标签和/或来源标签。在一些实施方式中,复合物或方法包括2至20之间个标签(例如,2至10之间个结构单元或支架标签、一个第一文库识别标签、一个可选的用途标签和一个来源标签)。在一些实施方式中,复合物或方法包括5至75之间个核苷酸(例如,如本文所描述的,如约40或50个核苷酸)。

[0025] 在本文所描述的任何实施方式中,在个体标签集合(set)内的每个标签包括大约相同的质量。

[0026] 在本文所描述的任何实施方式中,复合物包括RNA、DNA、修饰的DNA和/或修饰的RNA(例如,相同寡核苷酸内的PNA、LNA、GNA、TNA或它们的混合物,如本文所描述的)。

[0027] 在本文所描述的任何实施方式中,复合物或方法进一步包括尾部分。

[0028] 在另一个方面,本发明的特征在于文库,其包括本文描述的一种或多种复合物。在一些实施方式中,文库包括多个帽部分。在其他实施方式中,多个帽部分的每个帽部分包括相同的序列区(例如,引物结合区)和不同的编码区(例如,第一标签,其编码文库的用途、文库的来源、文库的特性、文库的历史,连接,间隔子,或另外的第一组分或寡核苷酸序列,其促进杂交、扩增或测序技术)。在特定的实施方式中,文库包括约 10^2 至 10^{20} 之间个复合物(例如,约 10^2 至 10^3 、 10^2 至 10^4 、 10^2 至 10^5 、 10^2 至 10^6 、 10^2 至 10^7 、 10^2 至 10^8 、 10^2 至 10^9 、 10^2 至 10^{10} 、 10^2 至 10^{11} 、 10^2 至 10^{12} 、 10^2 至 10^{13} 、 10^2 至 10^{14} 、 10^2 至 10^{15} 、 10^2 至 10^{16} 、 10^2 至 10^{17} 、 10^2 至 10^{18} 、 10^2 至 10^{19} 、 10^4 至 10^5 、 10^4 至 10^6 、 10^4 至 10^7 、 10^4 至 10^8 、 10^4 至 10^9 、 10^4 至 10^{10} 、 10^4 至 10^{11} 、 10^4 至 10^{12} 、 10^4 至 10^{13} 、 10^4 至 10^{14} 、 10^4 至 10^{15} 、 10^4 至 10^{16} 、 10^4 至 10^{17} 、 10^4 至 10^{18} 、 10^4 至 10^{19} 、 10^4 至 10^{20} 、 10^5 至 10^6 、 10^5 至 10^7 、 10^5 至 10^8 、 10^5 至 10^9 、 10^5 至 10^{10} 、 10^5 至 10^{11} 、 10^5 至 10^{12} 、 10^5 至 10^{13} 、 10^5 至 10^{14} 、 10^5 至 10^{15} 、 10^5 至 10^{16} 、 10^5 至 10^{17} 、 10^5 至 10^{18} 、 10^5 至 10^{19} 或 10^5 至 10^{20} 之间个复合物)。在一些实施方式中,每个复合物是不同的。

[0029] 在另一个方面,本发明的特征在于为包括编码化学实体的第一文库加标签(tagging)的方法,该方法包括:(a)提供具有第一官能团和第二官能团的帽部分(例如,其中帽部分可选地编码信息);(b)将帽部分的第一官能团结合至化学实体的第一组分,其中

帽部分直接连接至第一组分或帽部分由双官能间隔子间接地连接至第一组分；和(c)经由第一连接将帽部分的第二官能团结合至第一寡核苷酸标签以形成复合物，其中聚合酶具有降低的通读或移位通过第一连接的能力；其中可以任何顺序进行步骤(b)和(c)，并且其中第一标签编码步骤(b)的结合反应，从而提供标记的文库。

[0030] 在另一个方面，本发明的特征在于为包括编码化学实体的第一文库加标签的方法，该方法包括：(a)提供具有第一官能团和第二官能团的帽部分(例如，其中帽部分可选地编码信息)；(b)将帽部分的第一官能团结合至化学实体的第一组分，其中帽部分直接连接至第一组分或帽部分由双官能间隔子间接地连接至第一组分；(c)经由第一连接将帽部分的第二官能团结合至第一寡核苷酸标签；(d)结合化学实体的 n_c 个数量的另外的组分，其中 n_c 是1和10之间的整数；和(e)结合具有 n_t 连接的 n_t 个数量的另外的寡核苷酸标签以形成复合物，其中 n_t 为1和10之间的整数，其中每个连接在两个相邻标签之间，并且其中每个标签编码至少一个组分的特性；其中聚合酶具有降低的通读或移位通过第一连接和 n_t 连接中的至少一个的能力；并且其中可以任何顺序进行步骤(b)和(c)，并且其中第一标签为步骤(b)的结合反应编码；其中可以任何顺序进行步骤(d)和(e)，并且其中每个另外的标签为步骤(d)的每个另外的组分的结合反应编码，从而提供标记的文库。

[0031] 在一些实施方式中， n_c 和 n_t 每个独立地为1和2、1和3、1和4、1和5、1和6、1和7、1和8、1和9、1和10、1和12、1和15、1和18、1和20、2和3、2和4、2和5、2和6、2和7、2和8、2和9、2和10、2和12、2和15、2和18、2和20、3和4、3和5、3和6、3和7、3和8、3和9、3和10、3和12、3和15、3和18、3和20、4和5、4和6、4和7、4和8、4和9、4和10、4和12、4和15、4和18或4和20之间的整数。

[0032] 在一些实施方式中，聚合酶具有降低的通读或移位通过至少约10%的第一连接和 n_t 连接的能力(例如，与对照相比约20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或甚至100%的第一连接和 n_t 连接)。在特定的实施方式中，降低的通读或移位通过第一连接和 n_t 连接的能力在约10%至约100%之间(例如，与对照相比(例如，与缺少该连接的对照寡核苷酸相比)，20%至100%、25%至100%、50%至100%、75%至100%、90%至100%、95%至100%、10%至95%、20%至95%、25%至95%、50%至95%、75%至95%、90%至95%、10%至90%、20%至90%、25%至90%、50%至90%或75%至90%)。

[0033] 在一些实施方式中，第一组分和/或另外的组分包括支架或结构单元。

[0034] 在一些实施方式中，步骤(b)包括经由双官能间隔子将帽部分间接地结合至第一组分。

[0035] 在一些实施方式中，该方法进一步包括将用途标签、来源标签和/或第一文库识别标签结合至复合物。在特定的实施方式中，该方法进一步包括提供第二文库并且将第一文库与第二文库结合。在一些实施方式中，该方法进一步包括将尾部分结合至复合物。

[0036] 在一些实施方式中，结合步骤包括化学反应性基团、光反应性基团、嵌入部分或交联寡核苷酸(例如，如本文所描述的，用于本文的任何复合物)。在特定的实施方式中，该方法进一步包括在一个或多个步骤中进行拆分和合并合成、分离一种或多种复合物和/或纯化一种或多种复合物。

[0037] 在本文所描述的任何方法中，该方法获得本文所描述的任何复合物或本文所描述的任何文库。

[0038] 在另一个方面,本发明的特征在于筛选多个化学实体的方法,该方法包括:(a)使靶与本文所描述的任何复合物和/或本文所描述的文库接触;和(b)与对照相比,选择对于靶具有预定特性的一种或多种复合物,从而筛选化学实体。

[0039] 在一些实施方式中,预定特性包括与对照相对齐于靶(例如,如本文所描述的生物靶)的增加了的结合。

[0040] 在一些实施方式中,在步骤(a)之前,进行选自以下组成的列表的一个或多个操作:使一个或多个传递引物(relay primer)与所述复合物退火,其中所述一个或多个传递引物跨越所述第一连接和/或n-1连接;使用聚合酶延伸所述传递引物以产生寡核苷酸片段;和连接所述寡核苷酸片段以产生模板。

[0041] 在另一个方面,本发明的特征在于确定本文所描述的任何复合物的核苷酸序列的方法,该方法包括:(a)使一个或多个传递引物与复合物退火,其中一个或多个传递引物跨越所述第一连接和/或n-1连接;(b)使用聚合酶延伸传递引物以产生寡核苷酸片段;(c)连接寡核苷酸片段以产生模板;(d)可选地通过聚合酶链式反应扩增模板以产生扩增的混合物;和(e)测序可选扩增的混合物以确定复合物的序列。

[0042] 在一些实施方式中,复合物包括在一个或多个标签的5'-末端处的5'-连接子和在一个或多个标签的3'-末端处的3'-连接子,并且其中一个或多个传递引物的每个杂交至相邻的5'-连接子和3'-连接子。在特定的实施方式中,每个5'-连接子包括相同的序列和/或每个3'-连接子包括相同的序列。在一些实施方式中,至少一个5'-连接子的序列与相邻3'-连接子的序列互补(例如,以在杂交条件下形成5'-连接子和3'-连接子之间的双链体)或相同或足够相似以允许杂交至互补寡核苷酸。在进一步实施方式中,该方法包括使5'-连接子杂交至相邻的3'-连接子。在一些实施方式中,用于5'-连接子和3'-连接子之间的特异性结点(junction)(例如,从而形成三螺旋结点)的每个传递引物包括相同的序列。

[0043] 在另一个方面,本发明的特征在于确定本文所描述的任何复合物的核苷酸序列的方法,该方法包括:(a)提供包括化学连接的复合物,其中化学连接是交联寡核苷酸,其跨越两个相邻标签之间的结点,并且其经由一个或多个可逆的共反应性基团与两个相邻标签操作地关联;(b)释放交联寡核苷酸(例如,任何本文所描述的)以产生模板;(c)可选地通过聚合酶链式反应扩增模板以产生扩增的混合物;和(d)测序可选扩增的混合物以确定复合物的序列。在一些实施方式中,部分交联寡核苷酸杂交至两个相邻标签的末端,从而产生两个相邻标签之间的缺口(nick)或空位(gap)。

[0044] 在进一步实施方式中,该方法包括在步骤(b)之前使用化学处理或酶处理(例如,5'-磷酸化)连接标签。

[0045] 在一些实施方式中,该方法进一步包括在可选扩增步骤或测序步骤之前修复模板。在特定的实施方式中,修复步骤包括将第一连接和/或n-1连接中的至少一个修饰成为能够由聚合酶通读或移位通过修复连接(repaired linkage)。在一些实施方式中,聚合酶可以通读或移位通过至少50%(例如,至少60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或甚至100%)的修复连接(例如,与对照相比,如具有没有被修复的第一连接和n-1连接的对照寡核苷酸)。在特定的实施方式中,聚合酶可以通读或移位通过约10%至约100%的修复连接(例如,与对照相比(例如,与具有没有被修复的第一连接和n-1连接的对照寡核苷酸相比),20%至100%、25%至100%、50%至100%、75%至100%、90%

至100%、95%至100%、10%至95%、20%至95%、25%至95%、50%至95%、75%至95%、90%至95%、10%至90%、20%至90%、25%至90%、50%至90%和75%至90%)。在特定的实施方式中,修复步骤包括为模板提供修复酶(例如,如本文所描述的光裂合酶、糖基化酶、核酸内切酶、Flap核酸内切酶、脱嘌呤/脱嘧啶(AP)核酸内切酶、聚ADP核糖聚合酶或甲基转移酶)。

[0046] 在任何上述实施方式中,方法或复合物可以仅包括单链分子,其中帽部分、第一标签和/或一个或多个另外的标签是单链的。

[0047] 在任何上述实施方式中,如本文所描述的,方法进一步包括一个或多个可选步骤来多样话文库或询问文库的成员。在一些实施方式中,方法进一步包括识别小药物样文库成员,其结合或失活治疗关注的蛋白质。在其他实施方式中,方法进一步包括在适用于至少一个文库成员的条件下使文库成员与生物靶接触以结合至靶、去除不结合至靶的一个或多个文库成员并且分析与靶关联的一个或多个寡核苷酸标签。

[0048] 如本文所描述的,使用单链分子可以具有许多好处。因此,在本文所描述的任何实施方式中,与包括一种或多种双链分子(例如,双链帽部分或双链标签)的方法相比,方法和复合物包括帽部分、一个或多个标签、复合物、化学实体、分子或标记的文库的任何成员,具有降低的质量、增加的溶解度(例如,在有机溶剂中)、降低的成本、增加的反应性、增加的靶可达性、降低的流体力学半径和/或增加的分析评估的准确度。在一些实施方式中,在一组标签内的每个标签(例如,第一标签或随后的标签,如果存在的话)具有大约相同的质量(例如,每个标签具有离两个或多个标签之间平均质量约 $\pm 10\%$ 的质量)。在特定的实施方式中,与双链标签(例如,具有质量为约15000道尔顿、约14000道尔顿、约13000道尔顿或约12000道尔顿的双链标签)相比,标签具有降低的质量(例如,小于约15000道尔顿、约14000道尔顿、约13000道尔顿、约12000道尔顿、约11000道尔顿、约10000道尔顿、约9000道尔顿、约8000道尔顿、约7500道尔顿、约7000道尔顿、约6000道尔顿、约6500道尔顿、约5000道尔顿、约5500道尔顿、约4000道尔顿、约4500道尔顿或约3000道尔顿)。在其他实施方式中,与双链标签相比(例如,具有长度小于约20个核苷酸、小于约19个核苷酸、少于约18个核苷酸、小于约17个核苷酸、小于约16个核苷酸、小于约15个核苷酸、小于约14个核苷酸、小于约13个核苷酸、小于约12个核苷酸、小于约11个核苷酸、小于约10个核苷酸、小于约9个核苷酸、小于约8个核苷酸或小于约7个核苷酸的双链标签),标签具有降低的长度。在一些实施方式中,一个或多个标签或文库成员缺少引物结合区和/或恒定区(例如,在选择步骤中,如使用尺寸排阻色谱法的选择)。在一些实施方式中,一个或多个标签或文库成员具有缩减的恒定区(例如,长度小于约30个核苷酸、小于约25个核苷酸、小于约20个核苷酸、小于约19个核苷酸、小于约18个核苷酸、小于约17个核苷酸、小于约16个核苷酸、小于约15个核苷酸、小于约14个核苷酸、小于约13个核苷酸、小于约12个核苷酸、小于约11个核苷酸、小于约10个核苷酸、小于约9个核苷酸、小于约8个核苷酸或小于约7个核苷酸)。在其他实施方式中,方法包括为分子编码的帽部分、部分化学实体、步骤的结合反应(例如,化学或酶连接)或文库的特性,其中编码性帽部分消除对另外的标签以编码此类信息的需要。

[0049] 在任何上述实施方式中,寡核苷酸(例如,帽部分、第一标签和/或一个或多个另外的标签,如果存在的话)编码文库的特性。在一些实施方式中,寡核苷酸(例如,帽部分、第一标签和/或一个或多个另外的标签,如果存在的话)包括第一文库识别序列,其中序列编码

第一文库的特性。在特定的实施方式中,寡核苷酸是第一文库识别标签。在一些实施方式中,方法包括提供第一文库识别标签(其中该标签包括编码第一文库的序列)和/或将第一文库识别标签结合至复合物。在一些实施方式中,方法包括提供第二文库并且将第一文库与第二文库结合。在进一步实施方式中,方法包括提供第二文库识别标签,其中该标签包括编码第二文库的序列。在一些实施方式中,组合多于两个文库(例如,三个、四个、五个、六个、七个、八个、九个、十个或更多个文库)。

[0050] 在任何上述实施方式中,在一个或多个标签中或在多于一个标签的组合中提供所编码的信息。在一些实施方式中,由多于一个标签表示所编码的信息(例如,两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个标签)。在一些实施方式中,由多于一个标签表示所编码的信息,其中所有编码性标签都包含在编码性序列内(例如,通过使用特异性标签组合以编码信息)。在一些实施方式中,由多于一个标签表示所编码的信息,其中少于所有编码性标签都包含在编码性序列内(例如,通过使用来自一组多于一个个体标签的一个标签以编码个体内的编码性序列)。在一些实施方式中,正交地表示所编码的信息,其中由多于一个标签与少于所有编码性信息(其包含在个体文库成员内)的组合表示所编码的信息,使得多于一个相应文库成员需要进行测序以便去卷积所编码的信息。在一些实施方式中,由单个标签表示多于一个化学结构单元(例如,对于外消旋结构单元,如二个、三个、四个、五个、六个、七个、八个、九个、十个或更多个由单个标签表示的结构单元)。

[0051] 在任何上述实施方式中,寡核苷酸(例如,帽部分和/或一个或多个结构单元)编码文库成员的用途(例如,如本文所描述的在选择步骤或结合步骤中使用)。在一些实施方式中,寡核苷酸(例如,帽部分、第一标签和/或一个或多个另外的标签,如果存在的话)包括用途序列,其中序列编码文库中成员子组(subset)在一个或多个步骤中(例如,选择步骤和/或结合步骤)的用途。在特定的实施方式中,寡核苷酸是包括用途顺序的用途标签。在一些实施方式中,寡核苷酸(例如,帽部分和/或一个或多个寡核苷酸标签)编码文库成员的来源(例如,在文库的特定部分中)。在一些实施方式中,寡核苷酸(例如,帽部分、第一标签和/或一个或多个另外的标签,如果存在的话)包括来源序列(例如,具有长度为约10、9、8、7或6个核苷酸的随机或简并序列),其中,与文库的其他成员相比,序列编码文库中成员的来源。在特定的实施方式中,寡核苷酸是包括来源序列的来源标签。在一些实施方式中,该方法进一步包括将用途标签和/或来源标签接合(join)、结合(bind)或操作地关联到复合物。

[0052] 在本文的任何实施方式中,如本文所描述的,方法、组合物和复合物可选地包括尾部分,其中尾部分包括一个或多个文库识别序列、用途序列或来源序列。在特定的实施方式中,该方法进一步包括将尾部分(例如,包括一个或多个文库识别序列、用途顺序或来源序列)接合、结合或操作地关联到复合物。

[0053] 在任何上述实施方式中,方法、组合物和复合物或它们的一部分(例如,帽部分、第一标签和/或一个或多个另外的标签,如果存在的话)包括在半水性、降低的水性或非水性(例如,有机)条件下支持溶解度的修饰。在一些实施方式中,修饰双官能间隔子、帽部分或一个或多个标签以增加所述DNA编码化学文库成员在有机条件下的溶解度。在一些实施方式中,修饰是一个或多个烷基链、聚乙二醇单元、具有正电荷的分支种类或疏水环结构。在一些实施方式中,修饰包括一个或多个修饰的核苷酸,其具有疏水部分(例如,在具有脂族链的T或C碱基的C5位置处修饰,如在以下中:5'-二甲氧基三苯甲基-N4-二异丁基氨基亚甲

基-5-(1-丙炔)-2'-脱氧胞苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;5'-二甲氧基三苯甲基-5-(1-丙炔)-2'-脱氧尿苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;5'-二甲氧基三苯甲基-5-氟-2'-脱氧尿苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;和5'-二甲氧基三苯甲基-5-(苊-1-基-乙炔基)-2'-脱氧尿苷或3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺)或具有疏水部分的嵌入物(例如,偶氮苯)。在一些实施方式中,文库的成员具有从约1.0至约2.5的辛醇:水系数(例如,约1.0至约1.5、约1.0至约2.0、约1.3至约1.5、约1.3至约2.0、约1.3至约2.5、约1.5至约2.0、约1.5至约2.5或约2.0至约2.5)。

[0054] 在任何上述实施方式中,帽部分、尾部分、第一标签、一个或多个另外的标签、文库识别标签、用途标签、和/或来源标签,如果存在的话,可以包括约5至约75个核苷酸(例如,从5至7个核苷酸、5至8个核苷酸、5至9个核苷酸、5至10个核苷酸、5至11个核苷酸、5至12个核苷酸、5至13个核苷酸、5至14个核苷酸、5至15个核苷酸、5至16个核苷酸、5至17个核苷酸、5至18个核苷酸、5至19个核苷酸、5至20个核苷酸、5至30个核苷酸、5至40个核苷酸、5至50个核苷酸、5至60个核苷酸、5至70个核苷酸、6至7个核苷酸、6至8个核苷酸、6至9个核苷酸、6至10个核苷酸、6至11个核苷酸、6至12个核苷酸、6至13个核苷酸、6至14个核苷酸、6至15个核苷酸、6至16个核苷酸、6至17个核苷酸、6至18个核苷酸、6至19个核苷酸、6至20个核苷酸、7至8个核苷酸、7至9个核苷酸、7至10个核苷酸、7至11个核苷酸、7至12个核苷酸、7至13个核苷酸、7至14个核苷酸、7至15个核苷酸、7至16个核苷酸、7至17个核苷酸、7至18个核苷酸、7至19个核苷酸、7至20个核苷酸、8至9个核苷酸、8至10个核苷酸、8至11个核苷酸、8至12个核苷酸、8至13个核苷酸、8至14个核苷酸、8至15个核苷酸、8至16个核苷酸、8至17个核苷酸、8至18个核苷酸、8至19个核苷酸、8至20个核苷酸、9至10个核苷酸、9至11个核苷酸、9至12个核苷酸、9至13个核苷酸、9至14个核苷酸、9至15个核苷酸、9至16个核苷酸、9至17个核苷酸、9至18个核苷酸、9至19个核苷酸、9至20个核苷酸、10至11个核苷酸、10至12个核苷酸、10至13个核苷酸、10至14个核苷酸、10至15个核苷酸、10至16个核苷酸、10至17个核苷酸、10至18个核苷酸、10至19个核苷酸、10至20个核苷酸、10至30个核苷酸、10至40个核苷酸、10至50个核苷酸、10至60个核苷酸、10至70个核苷酸、10至75个核苷酸、11至12个核苷酸、11至13个核苷酸、11至14个核苷酸、11至15个核苷酸、11至16个核苷酸、11至17个核苷酸、11至18个核苷酸、11至19个核苷酸、11至20个核苷酸、12至13个核苷酸、12至14个核苷酸、12至15个核苷酸、12至16个核苷酸、12至17个核苷酸、12至18个核苷酸、12至19个核苷酸、12至20个核苷酸、13至14个核苷酸、13至15个核苷酸、13至16个核苷酸、13至17个核苷酸、13至18个核苷酸、13至19个核苷酸、13至20个核苷酸、14至15个核苷酸、14至16个核苷酸、14至17个核苷酸、14至18个核苷酸、14至19个核苷酸、14至20个核苷酸、15至16个核苷酸、15至17个核苷酸、15至18个核苷酸、15至19个核苷酸、15至20个核苷酸、16至17个核苷酸、16至18个核苷酸、16至19个核苷酸、16至20个核苷酸、17至18个核苷酸、17至19个核苷酸、17至20个核苷酸、18至19个核苷酸、18至20个核苷酸、19至20个核苷酸、20至30个核苷酸、20至40个核苷酸、20至50个核苷酸、20至60个核苷酸、20至70个核苷酸、20至75个核苷酸、30至40个核苷酸、30至50个核苷酸、30至60个核苷酸、30至70个核苷酸、30至75个核苷酸、40至50个核苷酸、40至60个核苷酸、40至70个核苷酸、40至75个核苷酸、50至60个核苷酸、50至70个核苷酸、50至75个核苷酸、60至70个核苷酸、60至75个核苷酸和70至75个核苷酸)。在特定的实施方式中,帽部分、第一标签、第二标签、一个或多个另外的标签、文库识别标签、用途标签和/或来源标签,如果存在

的话,具有小于20个核苷酸的长度(例如,小于19个核苷酸、小于18个核苷酸、小于17个核苷酸、小于16个核苷酸、小于15个核苷酸、小于14个核苷酸、小于13个核苷酸、小于12个核苷酸、小于11个核苷酸、小于10个核苷酸、小于9个核苷酸、小于8个核苷酸或小于7个核苷酸)。

[0055] 在任何上述实施方式中,编码序列(例如,帽部分、尾部分、第一标签、一个或多个另外的标签、文库识别标签、用途标签和/或来源标签,如果存在的话)可以包括多于20个核苷酸(例如,多于25、30、35、40、45、50、55、60、65、70或75个核苷酸)。

[0056] ■ 定义

[0057] “2’-取代的核苷酸”是指在核糖的2’-位置处具有取代的核苷酸碱基。

[0058] “约”是指引用值的 $\pm 10\%$ 。

[0059] “双官能”是指具有两个反应性基团,其允许结合两个化学部分。

[0060] “双官能间隔子”是指具有两个反应性基团的间隔部分,其允许结合化学实体并且编码复合物的信息。在一个非限制性实例中,在化学实体和标签之间提供双官能间隔子。在另一个非限制性实例中,在化学实体与帽部分之间提供双官能间隔子。本文提供示例性双官能间隔子。

[0061] “结合”是指通过共价键或非共价键连接。非共价键包括通过范德华力、氢键、离子键、包埋(entrapment)或物理封装(encapsulation)、吸收、吸附和/或其他分子间力形成的那些连接。可以通过任何有用的方法完成结合,如通过酶促结合(例如,以提供酶促连接(linkage)的酶促连接(ligation))或通过化学结合(例如,以提供化学连接(linkage)的化学连接(ligation))。

[0062] “结构单元”是指化学实体的结构单元,其中单元直接连接至其他化学结构单元或通过支架间接地连接。当化学实体是聚合的或低聚的化学实体时,结构单元是聚合物或低聚物的单体单元。结构单元可以具有一个或多个多样性节点,其允许添加一个或多个其他结构单元或支架。在大多数情况下,每个多样性节点是能够与一个或多个结构单元或支架反应以形成化学实体的官能团。通常,结构单元具有至少两个多样性节点(或反应性官能团),但一些结构单元可以具有一个多样性节点(或反应性官能团)。可替代地,所编码的化学或结合步骤可以包括若干化学组分(例如,多组分缩合反应或多步骤方法)。在两个不同结构单元上的反应性基团应是互补的,即能够一起反应以形成共价或非共价键。

[0063] “化学实体”是指包括一个或多个结构单元和可选的一个或多个支架的化合物。化学实体可以是任何小分子或肽药物或候选药物,其被设计或建造以具有一个或多个期望特性,例如,结合生物靶的能力、溶解度、氢键供体和受体的可用性、键的旋转自由度、正电荷、负电荷等。在某些实施方式中,化学实体可以进一步反应为双官能或三官能(或更大)的实体。

[0064] “化学反应性基团”是指参与模块化反应因此产生连接的反应性基团。示例性反应和反应性基团包括选自以下的那些:具有一对可选取代的炔基基团和可选取代的叠氮基团的惠斯更(Huisgen)1,3-偶极环加成反应;具有一对具有 4π -电子体系的可选取代的二烯和具有 2π -电子体系的可选取代的亲二烯体或可选取代的杂亲二烯体的狄尔斯-阿尔德(Diels-Alder)反应;具有亲核试剂和紧张的(strained)杂环亲电子体的开环反应;硫代磷酸酯基团和碘基团的夹板连接反应;和醛基基团和氨基基团的还原胺化反应,如本文所描述的。

[0065] “互补”是指如本文所定义的,序列能够杂交以形成二级结构(核酸分子的双链体或双链部分)。互补性不需要是完美的而是可以包括在一个、两个、三个或更多个核苷酸处的一个或多个错配。例如,互补序列可以包含核酸碱基,其可以形成根据沃森-克里克(Watson-Crick)碱基配对原则的氢键(例如,G与C、A与T或A与U)或其他氢键基序(例如,二氨基嘌呤与T、5-甲基C与G、2-硫代胸苷与A、肌苷与C、假异胞嘧啶(pseudoisocytosine)与G)。序列和它的互补序列可存在于相同寡核苷酸或不同寡核苷酸中。

[0066] “复合物”或“连接的复合物”是指通过共价键或非共价键与化学实体和/或一个或多个寡核苷酸标签操作地关联的帽部分。该复合物可以可选地包括化学实体和帽部分之间的双官能间隔子。

[0067] 化学实体的“组分”是指支架或结构单元。

[0068] 寡核苷酸标签的“连接子”是指在或靠近具有固定序列的5'-或3'-末端处的部分标签。5'-连接子位于或靠近寡核苷酸的5'-末端处,而3'-连接子位于或靠近寡核苷酸的3'-末端处。当存在于复合物中时,每个5'-连接子可以是相同或不同的,并且每个3'-连接子可以是相同或不同的。在具有多于一个标签的示例性、非限制性复合物中,每个标签可以包括5'-连接子和3'-连接子,其中每个5'-连接子具有相同的序列并且每个3'-连接子具有相同的序列(例如,其中5'-连接子的序列可与3'-连接子的序列相同或不同)。在另一个示例性、非限制性复合物中,5'-连接子的序列被设计成与3'-连接子的序列互补(例如,以允许5'-连接子和3'-连接子之间的杂交),如本文所定义。连接子可以可选地包括一个或多个允许连接的基团(例如,对于其聚合酶具有降低的通读或移位通过的能力的连接,如化学连接)。

[0069] “恒定”或“固定恒定”序列是指不编码信息的寡核苷酸的序列。具有恒定序列的复合物的非限制性、示例性的部分包括引物结合区、5'-连接子或3'-连接子。本发明的帽部分可以编码信息(因此,标签)或可替代地不编码信息(因此,恒定序列)。类似地,本发明的尾部分可以编码或不编码信息。

[0070] “交联寡核苷酸”是指寡核苷酸,如本文所定义,其操作地关联在复合物中两个相邻标签之间的具体结点处。在非限制性实例中,交联寡核苷酸的一个末端杂交至第一标签的3'-连接子,并且交联寡核苷酸的另一个末端杂交至相邻于第一标签的第二标签的5'-连接子。交联寡核苷酸的示例性、非限制性实施方式包括具有一个或多个反应性基团的那些(例如,化学反应性基团、光反应性基团、嵌入部分或可逆共反应性基团或任何本文所描述的),其操作地与相邻标签或相邻标签的连接子关联。

[0071] “多样性节点”是指在支架或结构单元中的位置处的官能团,其允许增加另一个结构单元。

[0072] “帽部分”是指用于文库合成的化学结构,其操作地连接至化学实体的组分和标签,例如,起始寡核苷酸。可选地,帽部分可以包含很少或无核苷酸,但可以提供点,在该点处它们可以被操作地关联。可选地,双官能间隔子将帽部分连接至组分。

[0073] “杂交”是指在严格度的各种条件下配对以形成互补寡核苷酸之间的双链分子或其部分。(例如参见,Wahl,G.M.and S.L.Berger(1987)Methods Enzymol.152:399;Kimmel,A.R.(1987)Methods Enzymol.152:507。)例如,可以以下盐浓度来获得高严格度杂交,其通常小于约750mM NaCl和75mM柠檬酸三钠、小于约500mM NaCl和50mM柠檬酸三钠或小于约

250mM NaCl和25mM柠檬酸三钠。可以在不存在有机溶剂例如甲酰胺下来获得低严格度杂交,同时可以在至少约35%甲酰胺或至少约50%甲酰胺的存在下来获得高严格度杂交。高严格度杂交温度条件将通常包括至少约30℃、37℃或42℃的温度。变化另外的参数,如杂交时间、去污剂例如十二烷基硫酸钠(SDS)的浓度和载体DNA的包含或排除是本领域技术人员所熟知的。通过根据需要结合这些不同条件来完成不同水平的严格度。在一个实施方式中,杂交将发生在30℃,在750mM NaCl、75mM柠檬酸三钠和1%SDS中。在可替代的实施方式中,杂交将发生在37℃,在500mM NaCl、50mM柠檬酸三钠、1%SDS、35%甲酰胺和100μg/ml变性鲑精DNA(ssDNA)中。在进一步的可替代的实施方式中,杂交将发生在42℃,在250mM NaCl、25mM柠檬酸三钠、1%SDS、50%甲酰胺和200μg/ml ssDNA中。关于这些条件的有用变化将是本领域技术人员容易显而易见的。

[0074] 对于大多数应用来说,跟随杂交的洗涤步骤也将在严格度方面变化。可以通过盐浓度且通过温度来限定洗涤严格度条件。如上,可以通过降低盐浓度或通过增加温度来增加洗涤严格度。例如,用于洗涤步骤的高严格度盐浓度可以是,例如,小于约30mM NaCl和3mM柠檬酸三钠或小于约15mM NaCl和1.5mM柠檬酸三钠。用于洗涤步骤的高严格度温度条件将通常包括,例如,至少约25℃、42℃或68℃的温度。在一个实施方式中,洗涤步骤将发生在25℃,在30mM NaCl、3mM柠檬酸三钠和0.1%SDS中。在可替代的实施方式中,洗涤步骤将发生在42℃,在15mM NaCl、1.5mM柠檬酸三钠和0.1%SDS中。在进一步的可替代的实施方式中,洗涤步骤将发生在68℃,在15mM NaCl、1.5mM柠檬酸三钠和0.1%SDS中。关于这些条件的另外的变化将是本领域技术人员容易显而易见的。杂交技术是本领域技术人员所熟知的且描述于,例如,Benton和Davis(*Science* 196:180,1977);Grunstein和Hogness(*Proc.Natl.Acad.Sci.,USA* 72:3961,1975);Ausubel等人(*Current Protocols in Molecular Biology*,Wiley Interscience,New York,2001);Berger和Kimmel(*Guide to Molecular Cloning Techniques*,1987,Academic Press,New York);和Sambrook等人,*Molecular Cloning:A Laboratory Manual*,Cold Spring Harbor Laboratory Press,New York。

[0075] “嵌入部分”是指产生包含两个或多个核苷酸之间部分的反应性基团。在非限制性实例中,嵌入部分与一个或多个核苷酸反应以形成双链体或三链体寡核苷酸之间的链间或链内(inter-or intra-strand)交联。在本文中描述示例性、非限制性嵌入部分。

[0076] “结点”是指复合物中两个相邻标签之间的缺口(缺少核苷酸间键)或空位(缺少一个或多个核苷酸)。结点也可以在存在于两个相邻标签中的两个相邻连接子之间(例如,在第一标签的3'-连接子和相邻于第一标签的第二标签的5'-连接子之间)。

[0077] “文库”是指分子或化学实体的集合。可选地,分子或化学实体结合至一个或多个寡核苷酸,其编码化学实体的分子或部分。

[0078] “连接”是指化学连接实体,其允许操作地关联两个或多个化学结构,其中连接存在于帽部分和一个或多个标签之间、两个标签之间或标签和尾部分之间。化学连接实体可以是非共价键(例如,如本文所描述的)、共价键或两个官能团之间的反应产物。“化学连接”是指由两个官能团之间的非酶、化学反应形成的连接。示例性、非限制性官能团包括化学反应性基团、光反应性基团、嵌入部分或交联寡核苷酸(例如,如本文所描述的)。“酶连接”是指由酶形成的核苷酸间或核苷间连接(键)。示例性、非限制性酶包括激酶、聚合酶、连接酶

或它们的组合。“对于其聚合酶具有降低的通读或移位通过的能力”的连接是指这样的连接,当存在于寡核苷酸模板中时,与缺少该连接的对照寡核苷酸相比,其提供降低量的由聚合酶延长和/或扩增产物。用于确定此类连接的示例性、非限制性方法包括如由以下评估的引物延伸:PCR分析(例如,定量PCR)、RT-PCR分析、液相色谱法-质谱法、序列群体统计或其他方法。示例性、非限制性的聚合酶包括DNA聚合酶和RNA聚合酶,如DNA聚合酶I、DNA聚合酶II、DNA聚合酶III、DNA聚合酶VI、Taq DNA聚合酶、Deep VentR™ DNA聚合酶(高保真嗜热性DNA聚合酶,可从New England Biolabs购得)、T7DNA聚合酶、T4DNA聚合酶、RNA聚合酶I、RNA聚合酶II、RNA聚合酶III或T7RNA聚合酶。

[0079] “多价阳离子”是指能够与多于一个配位体或阴离子形成多于一个键的阳离子。多价阳离子可以形成离子复合物或配位复合物。示例性多价阳离子包括来自碱土金属(例如,镁)和过渡金属(例如,锰(II)或钴(III))的那些和可选结合至一种或多种阴离子和/或一个或多个一价或多配位基的配位体如氯离子、胺和/或乙二胺的那些。

[0080] “寡核苷酸”是指具有5′-末端、3′-末端和在5′-和3′-末端之间的内部位置处的一个或多个核苷酸的核苷酸的聚合物。寡核苷酸可以包括本领域中已知的DNA、RNA或它们的任何衍生物,其可合成并用于碱基对识别。寡核苷酸不是必需具有连续碱基而是可以穿插有接头部分。寡核苷酸聚合物和核苷酸(例如,修饰的DNA或RNA)可以包括天然碱基(例如,腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷、脱氧胞苷、肌苷或二氨基嘌呤)、碱基类似物(例如,2-氨基腺苷、2-硫代胸苷、肌苷、吡咯-嘧啶、3-甲基腺苷、C 5-丙炔基胞苷、C5-丙炔基尿苷、C5-溴尿苷、C5-氟代尿苷、C5-碘尿苷、C5-甲基胞苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、0(6)-甲基鸟嘌呤和2-硫代胞苷)、修饰的碱基(例如,2′-取代的核苷酸、如2′-O-甲基化的碱基和2′-氟代碱基)、嵌入碱基、修饰的糖(例如,2′-氟代核糖;核糖;2′-脱氧核糖;阿拉伯糖;己糖;脱水己糖醇;阿卓糖醇;甘露醇;环己基(cyclohexanyl);环己烯基;也具有氨基磷酸酯(phosphoramidate)主链的吗啉;锁核酸(LNA,例如,其中核糖的2′-羟基由C₁₋₆亚烷基或C₁₋₆杂烷基桥连接至相同核糖糖的4′-碳,其中示例性桥接包括亚甲基、亚丙基、醚或氨基桥);乙二醇核酸(GNA,例如,R-GNA或S-GNA,其中核糖由连接至磷酸二酯键的乙二醇单元取代);苏糖核酸(TNA,其中用α-L-苏呋喃糖基-(3′→2′)取代);和/或核糖中氧的取代(例如,以S、Se或亚烷基如亚甲基或亚乙基))、修饰的主链(例如,肽核酸(PNA),其中2-氨基-乙基-甘氨酸连接取代核糖和磷酸二酯主链)和/或修饰的磷酸酯基团(例如,硫代磷酸酯(phosphorothioate)、5′-N-亚磷酰胺、磷硒酸酯(phosphoroselenate)、硼代磷酸酯(boranophosphate)、硼磷酸酯(boranophosphate esters)、氢磷酸酯、氨基磷酸酯、二氨基磷酸酯、烷基或芳基磷酸酯、磷酸三酯、桥接氨基磷酸酯、桥接硫代磷酸酯和桥接亚甲基-磷酸酯)。寡核苷酸可以是单链(例如,发夹结构)、双链的或具有其他二级或三级结构(例如,茎-环结构、双螺旋、三链体、四链体(quadruplexe)等)。

[0081] “操作地连接”或“操作地关联”是指,两个或多个化学结构直接或间接地连接在一起,以此类作为要保持连接的方式,通过它们预期经历的不同操作。典型地,化学实体和帽部分操作地以间接方式关联(例如,经由适当间隔子共价地)。例如,间隔子可以是具有用于化学实体的连接的位点和用于帽部分的连接位点的双官能部分。

[0082] “光反应性基团”是指反应性基团,其参与由紫外线、可见光或红外线辐射的吸收

引起的反应,因此产生连接。在本文中描述示例性、非限制性光反应性基团。

[0083] “保护基团”是指在制备、加标签或使用寡核苷酸编码文库的一个或多个结合步骤中意在保护寡核苷酸的3'-末端或5'-末端或保护化学实体、支架或结构单元的一个或多个官能团免于不期望的反应的基团。常用的保护基团在Greene, “Protective Groups in Organic Synthesis,” 第4版 (John Wiley&Sons, New York, 2007) 中公开, 其通过参考方式并入本文。用于寡核苷酸的示例性保护基团包括不可逆的保护基团, 如双脱氧核苷酸和双脱氧核苷 (ddNTP或ddN), 并且更优选地, 用于羟基基团的可逆的保护基团, 如酯基团 (例如, 0-(α -甲氧基乙基) 酯、0-异戊酰酯和0-乙酰丙基酯 (0-levulinyl ester)), 三苯甲基基团 (例如, 二甲氧基三苯甲基和单甲氧基三苯甲基)、咕吨基基团 (例如, 9-苯基咕吨-9-基和9-(对-甲氧基苯基) 咕吨-9-基)、酰基基团 (例如, 苯氧基乙酰基和乙酰基) 和甲硅烷基基团 (例如, 叔丁基二甲基甲硅烷基)。用于化学实体、支架和结构单元的示例性、非限制性保护基团包括N-保护基团, 其在合成过程中保护氨基基团免于不期望的反应 (例如, 酰基; 芳酰基; 氨基甲酰基团, 如甲酰基、乙酰基、丙酰基、新戊酰基、叔丁基乙酰基、2-氯乙酰基、2-溴乙酰基、三氟乙酰基、三氯乙酰基、邻苯二甲酰基、邻-硝基苯氧乙酰基、 α -氯丁酰基、苯甲酰基、4-氯苯甲酰基、4-溴苯甲酰基、4-硝基苯甲酰基和手性助剂, 如保护或未受保护的D、L或D, L-氨基酸, 如丙氨酸、亮氨酸、苯丙氨酸等; 含磺酰基基团, 如苯磺酰基、对-甲苯磺酰基等; 氨基甲酸酯形成基团, 如苄氧羰基、对-氯苄氧羰基、对-甲氧基苄氧羰基、对-硝基苄氧羰基、2-硝基苄氧羰基、对-溴苄氧羰基、3,4-二甲氧基苄氧羰基、3,5-二甲氧基苄氧羰基、2,4-二甲氧基苄氧羰基、4-甲氧基苄氧羰基、2-硝基-4,5-二甲氧基苄氧羰基、3,4,5-三甲氧基苄氧羰基、1-(对-联苯基)-1-甲基乙氧羰基、 α,α -二甲基-3,5-二甲氧基苄氧羰基、二苯甲氧基羰基 (benzhydryloxy carbonyl)、叔丁氧羰基、二异丙基甲氧羰基、异丙氧基羰基、乙氧羰基、甲氧羰基、烯丙氧基羰基、2,2,2-三氯乙氧羰基、苯氧羰基、4-硝基苯氧羰基、苄基-9-甲氧羰基、环戊氧基羰基、金刚烷氧羰基、环己氧基羰基、苯硫基羰基等; 烷芳基基团, 如苄基、三苯甲基、苄氧甲基等; 和甲硅烷基基团, 如三甲基甲硅烷基等; 其中优选的N-保护基团是甲酰基、乙酰基、苯甲酰基、新戊酰基、叔丁基乙酰基、丙氨酰基、苯基磺酰基、苄基、叔丁氧羰基 (Boc) 和苄氧羰基 (Cbz)); 在合成过程中保护羟基基团免于不期望的反应的O-保护基团 (例如, 烷基羰基基团, 如酰基、乙酰基、新戊酰基等; 可选取代的芳基羰基基团, 如苯甲酰基; 甲硅烷基基团, 如三甲基甲硅烷基 (TMS)、叔丁基二甲基甲硅烷基 (TBDMS)、三异丙基甲硅烷氧基甲基 (TOM)、三异丙基甲硅烷基 (TIPS) 等; 与羟基形成醚的基团, 如甲基、甲氧基甲基、四氢吡喃基、苄基、对-甲氧基苄基、三苯甲基等; 烷氧羰基, 如甲氧羰基、乙氧羰基、异丙氧羰基、正-异丙氧羰基、正-丁氧羰基、异丁氧羰基、仲丁氧羰基、叔丁氧羰基、2-乙基己氧羰基、环己氧羰基、甲氧羰基等; 烷氧烷氧羰基基团, 如甲氧甲氧羰基、乙氧甲氧羰基、2-甲氧乙氧羰基、2-乙氧乙氧羰基、2-丁氧乙氧羰基、2-甲氧乙氧甲氧羰基、烯丙氧羰基、炔丙氧羰基、2-丁烯氧羰基 (2-butenoxycarbonyl)、3-甲基-2-丁烯氧羰基等; 卤代烷氧羰基, 如2-氯乙氧羰基、2-氯乙氧羰基、2,2,2-三氯乙氧羰基等; 可选取代的芳基烷氧羰基基团, 如苄氧羰基、对-甲基苄氧羰基、对-甲氧基苄氧羰基、对-硝基苄氧羰基、2,4-二硝基苄氧羰基、3,5-二甲基苄氧羰基、对-氯苄氧羰基、对-溴苄氧羰基等; 和可选取代的芳氧羰基基团, 如苯氧羰基、对-硝基苯氧羰基、邻-硝基苯氧羰基、2,4-二硝基苯氧羰基、对-甲基苯氧羰基、间-甲基苯氧羰基、邻-溴苯氧羰基、3,5-二甲基苯氧羰基、对-氯苯氧羰基、2-氯-

4-硝基苯氧羰基等)；羰基保护基团(例如,缩醛和缩酮基团,如二甲基缩醛、1,3-二氧戊环等;缩羧酯基团;和二噻烷基团,如1,3-二噻烷、1,3-二硫戊环等);羧酸保护基团(例如,酯基团,如甲酯、苄酯、叔丁酯、原酸酯等;甲硅烷基基团,如三甲基甲硅烷基、以及任何本文所描述的;和噁唑啉基团);和磷酸酯保护基团(例如,可选取代的酯基团,如甲酯、异丙酯、2-氰乙酯、烯丙酯、叔丁酯、苄酯、苄甲酯、2-(三甲基甲硅烷基)乙酯、2-(甲磺酰基)乙酯、2,2,2-三氯乙酯、3',5'-二甲氧基苯偶姻酯、对-羟基苯甲酰酯等)。

[0084] 对寡核苷酸的末端的“接近度(proximity)”或“接近(in proximity)”是指比其他剩余末端更邻近或更靠近于所述末端。例如,接近于寡核苷酸的3'-末端的部分或基团比5'-末端更邻近或更靠近于3'-末端。在特定的实施方式中,接近于寡核苷酸的3'-末端的部分或基团距离3'-末端一个、二个、三个、四个、五个、六个、七个、八个、九个、十个、十五个或更多个核苷酸。在其他实施方式中,接近于寡核苷酸的5'-末端的部分或基团距离5'-末端一个、二个、三个、四个、五个、六个、七个、八个、九个、十个、十五个或更多个核苷酸。

[0085] “纯化”是指去除存在于反应混合物中的任何未反应的产物或任何试剂,其可能降低将要在连续步骤中使用的化学或生物试剂的活性。纯化可以包括要去除的未反应产物或试剂的一个或多个色谱分离、电泳分离和沉淀。

[0086] “可逆共反应性基团”是指参与可逆反应的反应性基团。示例性、非限制性反应性基团包括光反应性基团,其中暴露于特定的吸收辐射引起光反应性基团之间的连接,并且暴露于不同的、特定的吸收辐射引起所形成的连接的断裂(例如,氰乙烯基呋唑基团、氰乙烯基基团和丙烯酰胺基团)。另一个示例性、非限制性反应性基团包括氧化还原反应性基团,其中此类基团可被可逆地还原或氧化(例如,硫醇基团)。

[0087] “支架”是指以特定的特殊的几何形状显示一个或多个多样性节点的化学部分。在文库合成期间,多样性节点典型地连接至支架,但在某些情况下,一个多样性节点可以在文库合成之前连接至支架(例如,添加一个或多个结构单元和/或一个或多个标签)。在一些实施方式中,支架被衍生,使得它可以在文库合成期间正交地去保护并且随后与不同的多样性节点反应。

[0088] “小分子”药物或“小分子”候选药物是指具有低于约1000道尔顿分子量的分子。小分子可以是有机或无机的、分离的(例如,从化合物文库或天然来源)或通过已知化合物的衍生作用获得的。

[0089] “基本特性”或“基本相同”是指多肽或多核苷酸序列,其分别与参考序列具有相同的多肽或多核苷酸序列,或分别具有氨基酸残基或核苷酸的具体的百分比,其当两个序列最佳对齐时在参考序列内的相应位置处是相同的。例如,氨基酸序列,其“基本相同”于与参考氨基酸序列具有至少50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%特性的参考序列。对于多肽,比较序列的长度一般为至少5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个连续的(邻接的,contiguous)氨基酸,更优选地至少25、50、75、90、100、150、200、250、300或350个连续的氨基酸,并且最优选全长氨基酸序列。对于核酸,比较序列的长度一般为至少5个连续的核苷酸,优选至少10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个连续的核苷酸,并且最优选全长核苷酸序列。可以使用序列分析软件以默认设置来测量序列特性(例如,Sequence Analysis Software Package of the Genetics Computer Group(遗传学计算机组的序列分析软件包),University of

Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705)。此类软件可以通过对各种取代、缺失和其他修饰分配同源性程度匹配类似的序列。

[0090] “标签”或“寡核苷酸标签”是指文库的寡核苷酸部分,其至少部分编码信息。此类信息的非限制性实例包括组分(即,支架或结构单元,如分别在支架标签或结构单元标签中)的添加(例如,通过结合反应)、文库中的帽部分、文库的特性(即,如在特性标签中)、文库的用途(即,如在用途标签中)和/或文库成员的来源(即,如在来源标签中)。

[0091] “尾部分”是指文库的寡核苷酸部分,其在添加所有前述标签的后连接至复合物并且编码文库的特性、文库的用途和/或文库成员的来源。

[0092] “引物”是指寡核苷酸,其能够退火到寡核苷酸模板并且然后以模板依赖方式由聚合酶延伸。

[0093] “传递引物”是指寡核苷酸,其能够退火到在引物与其杂交的模板区域中包含至少一个核苷酸间连接(其降低聚合酶的能力以通读或移位通过通过)的寡核苷酸模板。杂交之后,一个或多个传递引物允许以模板依赖方式由聚合酶延伸。

[0094] 如本文所用的“重组”是指作为至少两个不同杂交事件的结果的聚合酶产物的产生。

[0095] 从以下详细描述和权利要求书,本发明的其他特征和优点将显而易见。

附图说明

[0096] 图1示出示例性复合物。图1的(a)示出复合物(从5'至3'方向),其具有化学实体(星标)、固定的恒定序列(例如,帽部分)、接着是三个可变编码序列(例如,三个标签)和另一个固定的恒定序列(例如,尾部分),其中固定的恒定序列和可变编码序列之间的连接是可读的。可以使用聚合酶用引物(杂交至复合物的3'-末端)进一步延伸复合物,可选地用PCR扩增并测序。图1的(b)示出具有不可读的连接复合物。该复合物(从5'至3'方向)具有化学实体(星标)、固定的恒定序列(例如,帽部分)、接着是三个可变编码序列(例如,三个标签)和另一个固定的恒定序列(例如,尾部分),其中固定的恒定序列和可变编码序列之间的连接是不可读的。每个可变编码序列包括5'-连接子和3'-连接子,其是固定的恒定序列。因为连接是不可读的,传递引物用于跨越该连接并允许用聚合酶延伸,因此形成寡核苷酸片段。然后使用连接酶连接片段,可选地用PCR扩增并测序。可以可选地在选择后完全进行读取过程(例如,可选的延伸、连接、可选的扩增和测序)。可替代地,可以在选择之前进行读取过程的可选延伸和连接部分。

[0097] 图2示出具有由补骨脂素形成的不可读的连接示例性复合物。该示意图提供(从5'至3'方向)可变编码序列、3'-连接子、5'-连接子和另一个可变编码序列。3'-连接子和5'-连接子杂交以形成双链体,并且补骨脂素(嵌入的、光反应部分)形成5'-连接子和3'-连接子之间的不可读的连接。传递引物杂交至5'-连接子和3'-连接子并跨越不可读的连接。用聚合酶延伸复合物的可读部分(例如,可变编码序列)以形成寡核苷酸片段。使用连接酶连接片段和传递引物,可选地用PCR扩增并测序。可以可选地在选择后完全进行读取过程(例如,延伸、连接、可选的扩增和测序)。可替代地,可以在选择之前进行读取过程的延伸和连接部分。

[0098] 图3示出具有由交联寡核苷酸形成的不可读的连接示例性复合物。在顶部示意

图中,复合物包括(从5'至3'方向)化学实体(星标)、固定的恒定序列(例如,帽部分)、经由可逆共反应性基团(例如,在约366nm处交联的氰乙烯基吡唑基团,用X标记交联)交联的交联寡核苷酸,并且接着是在5'-和3'-末端处具有固定的恒定序列(例如,具有5'-连接子和3'-连接子的标签)的可变编码序列。交联寡核苷酸和可变编码序列的组合重复3次以上。然后,接着是最终交联寡核苷酸和另一个固定的恒定序列(例如,尾部分)。用X标记交联。接着,用化学处理和/或酶处理(例如,连接酶和可选的激酶)反应复合物以连接标签。接着,释放交联寡核苷酸以形成模板(例如,使用在约312nm处的吸收)。最后,可选地用PCR扩增并测序模板。可以可选地在选择后完全进行读取过程(例如,可选的激酶步骤、连接、可选的扩增和测序)。可替代地,可以在选择之前进行读取处理的可选激酶步骤和连接步骤。

[0099] 图4示出具有可逆共反应性基团(一对氰乙烯基吡唑基团和胸苷)的示例性可逆反应,并且这些基团的使用允许形成交联寡核苷酸和寡核苷酸标签的固定的恒定序列之间的交联。可逆共反应性基团(由X和X'标记)存在于3'-连接子、5'-连接子和交联寡核苷酸中。

[0100] 图5示出寡核苷酸CNVKJ,其包含两个氰乙烯基吡唑修饰;CNVK2_P_TagB;CNVK2_NP_TagB和CNVK2_TagA。

[0101] 图6示出寡核苷酸CNVKJ、CNVK2_P_TagB和CNVK2_TagA的光化学交联产物的电泳结果。

[0102] 图7示出纯化的CNVK2_P_TagB和CNVK2_TagA的LCMS。

[0103] 图8示出回收的起始Double_CNVK_template的LCMS。

[0104] 图9示出用T4DNA连接酶处理CNVK2_P_TagB、CNVK2_TagA和CNVKJ的复合物的LCMS。

[0105] 图10示出光化学交联的非磷酸化的模拟标签对CNVK2_NP_TagB与CNVK2_TagA和CNVKJ的磷酸化-连接反应的LCMS结果。

[0106] 图11示出寡核苷酸5PS02_A9_TA,其在5'-末端处用C2-补骨脂素修饰,以及寡核苷酸PS0_HP_A9_TCT。

[0107] 图12示出在茎(stem)内使用补骨脂素的光化学寡核苷酸交联以模拟寡核苷酸5PS02_A9_TA(其在5'-末端处用C2-补骨脂素修饰)和寡核苷酸PS0_HP_A9_TCT的一个加标签事件的LCMS。

[0108] 图13示出寡核苷酸Tag1_PsoCVU和SplintC_PsoC2和寡核苷酸5PS0C2_A9_GA(在5'-末端处用C2补骨脂素修饰)。

[0109] 图14示出补骨脂素光连接的示意图。

[0110] 图15示出补骨脂素光连接的时间-进程的凝胶。

[0111] 图16示出5PsoC2_A9_GA和Tag1_PsoCVU的光连接结合物的LCMS。

[0112] 图17示出5'-生物素化的寡核苷酸5Bio_Tag_PsoCVU和5'-补骨脂素寡核苷酸5PsoC2_A9_GA与夹板寡核苷酸SplintC_PsoC2。

[0113] 图18示出补骨脂素-胸苷光化学连接的寡核苷酸结合物(conjugate) (Bio_Pso)的LCMS。

[0114] 图19示出使用末端和非末端引物延伸和连接从Bio_Pso产生的cDNA的LCMS。

[0115] 图20示出使用末端和非末端引物延伸,从Bio_Pso产生的cDNA的无连接对照的LCMS。

[0116] 图21示出寡核苷酸CVU_G和CVU_A和寡核苷酸Tag1_PsoCVU,SplintC_CVU和

SplintA_CVU。

[0117] 图22示出已照射的5'-5-(羧基)乙烯基-2'-脱氧尿苷和3'-胸苷之间形成的光化学连接结点的结构。

[0118] 图23示出在365nm处CVU光连接的时间-进程的凝胶。

[0119] 图24示出CVU_G与Tag1_PsoCVU的光化学连接产物的LCMS。

[0120] 图25示出羧基乙烯基尿苷-胸苷光化学连接的寡核苷酸结合物 (Bio_CVU) 的LCMS。

[0121] 图26示出Bio_CVU模板减去单个dA核苷酸的LCMS。

[0122] 图27示出寡核苷酸S_IDTN₃和S_IDTalkyne。

[0123] 图28示出S_IDTN₃和S_IDTalkyne的凝胶纯化的结合物的LCMS分析。

[0124] 图29示出TKR_Central和S_IDTN₃寡聚物和5'-二苯并环辛炔 (DBCO) 修饰的寡聚物TKR_DBCO_S。

[0125] 图30示出TKR_DBCO_S和TKR_Central之间的无铜点击反应的LCMS时间进程。

[0126] 图31示出2cl_s结合物的LCMS。

[0127] 图32示出Conjugate_Click_S和Conjugate_Click_L。

[0128] 图33示出在ePsplint的存在和不存在下,通过Deep Vent (外切) 聚合酶, Conjugate_Click_L和Conjugate_Click_S的夹板依赖重组介导的cDNA产生/PCR扩增。24个循环,200pM起始结合物浓度。1-分子量标记;2-无模板;3-无模板+ePsplint;4-Conjugate_Click_S;5-Conjugate_Click_S+ePsplint;6-Conjugate_Click_L;5-Conjugate_Click_L+ePsplint。

[0129] 图34示出提出的夹板依赖重组介导的cDNA产生/PCR扩增的机制。

[0130] 图35示出在自由溶液 (free solution) 中,通过Deep Vent外切聚合酶单独和混合在一起的Conjugate_Click_S和Conjugate_Click_L的夹板依赖重组介导的cDNA产生/PCR扩增。所有反应辅以ePsplint。1-分子量标记;2-无模板;3-Conjugate_Click_S;4-Conjugate_Click_L;5-Conjugate_Click_S和Conjugate_Click_L。

[0131] 图36示出用标签关联 (编码信息) 的结果拖曳 (拖尾, shuffling) (损失), 在自由溶液中夹板依赖重组介导的cDNA产生/PCR扩增产物的重组的可能机制。

[0132] 图37示出在自由溶液中且在乳化水性区室中用如下条件通过Deep Vent (外切) 聚合酶的Conjugate_Click_S和Conjugate_Click_L混合物的重组介导的cDNA产生/PCR扩增: 在94℃30秒、在52℃30秒和在72℃60秒, 进行35个循环和5pM的初始结合物浓度;1-分子量标记;2-自由溶液, 3-乳化水性区室。

[0133] 图38示出TKR_S结合物和TKR_L结合物以及它们的LCMS迹线。

[0134] 图39示出TKR_S和TKR_L的Deep Vent (外切) 聚合酶重复依赖重组介导的cDNA产生/PCR扩增。1-分子量标记, 2-无模板, 3-TKR_S, 4-TKR_L, 5-TKR_L和TKR_S。条件: 5uL的10x Thermopol缓冲液; 2.5uL的每个正向和反向引物的10uM溶液; 2.5uL的10mM dNTP混合物 (NEB); 40pM终浓度的结合物或它们的1:1混合物, 2.5uL的Deep Vent (外切) 聚合酶且加水补至50uL。如下循环反应21次: 在94℃30秒、在52℃30秒和在72℃60秒。

[0135] 图40示出衍生自自由溶液中混合的TKR_S和TKR_L的重复依赖重组介导cDNA产生/PCR扩增的克隆扩增产物的序列数据。

[0136] 图41示出TKR_2_click_S结合物和TKR_2_click_L结合物和它们的LCMS迹线。

[0137] 图42示出在自由溶液中2_click_S和2_click_L的Deep Vent (外切) 聚合酶重复依赖重组介导的cDNA产生/PCR扩增。1-分子量标记;2-无模板控制;3-2_click_S;4-2_click_L;5-2_click_S和2_click_L的1:1混合物。条件:5uL的10x Thermopol缓冲液;2.5uL的每个正向和反向引物的10uM溶液;2.5uL的10mM dNTP混合物 (NEB);40pM终浓度的结合物或它们的1:1混合物,2.5uL的Deep Vent (外切) 聚合酶且加水补至50uL。如下循环反应22次:在94℃30秒、在52℃30秒和在72℃60秒。

[0138] 图43示出衍生自自由溶液中TKR_2_click_S和TKR_2_click_L混合物的重复依赖重组介导cDNA产生/PCR扩增产物的克隆扩增产物的序列数据。

具体实施方式

[0139] 我们已开发了具有至少一个连接(例如,化学连接)的复合物,对于其聚合酶具有降低的通读或移位通过的能力。不受聚合酶可读性限制的标签文库的能力极大地扩展可以使用的加标签方法的范围。可能的、非限制性的优点包括标签和/或文库成员的降低的质量;在不同反应条件下(例如,在适用于化学实体和标签合成的水性和/或有机条件下)文库成员的增加的溶解度;增加的稳定性(例如,在水性和/或有机反应条件下,与此类连接的标签相比,其中此类寡核苷酸标签和连接可以更耐特定的反应条件);降低的成本(例如,减少昂贵酶试剂的使用);增加的易用性(例如,氰乙烯基唑交联的使用发生在5.5-9.5的宽pH范围上,其在适用于形成复合物的化学实体部分的反应条件下允许交联);降低的反应步骤数量和减少的一种或多种试剂的使用(例如,减少缓冲液交换步骤的使用,如在组合拆分反应期间,其中数千个个体、小等分试样需要独立处理;沉淀步骤或pH改变步骤);增加的保真度(例如,当将杂交介导的方法用于降低错误加标签事件的发生频率时);和与用于形成化学实体的反应条件的增加的相容性(例如,允许正交官能度的使用,其中可以使用唯一的、加标签的化学作用(例如,紫外线照射的使用),其在后拆分/混合步骤期间不会发生,其也将降低错误加标签事件的频率)。

[0140] 因为此类连接可以具有降低的由聚合酶的可读性,我们也已开发允许测序此类复合物的方法。这些方法包括使用传递引物以跨越连接和/或使用可释放的交联寡核苷酸。这些复合物和方法可以用于通过建立特定标签和特定化学反应或结构单元之间的编码关系来创建可以选择化学实体的不同文库。为了识别一个或多个化学实体,寡核苷酸标签可被扩增、克隆、测序并且通过使用已建立的关系关联。在下文中详细描述创建这些复合物文库和使这些复合物文库加标签的方法。

[0141] ■ 复合物

[0142] 本发明的特征在于复合物,其包括化学实体、一个或多个标签和与化学实体和一个或多个标签操作地关联的帽部分。帽部分和标签之间或两个标签之间的至少一个连接是对于其聚合酶具有降低的通读或移位通过的能力的连接。进一步在下文中描述化学实体、帽部分、标签、连接和双官能间隔子。

[0143] ■ 化学实体

[0144] 本发明的化学实体或成员(例如,小分子或肽)可以包括一个或多个结构单元并且可选地包括一个或多个支架。

[0145] 支架S可以是单原子或分子支架。示例性单原子支架包括碳原子、硼原子、氮原子

或磷原子等。示例性多原子支架包括环烷基基团、环烯基基团、杂环烷基基团、杂环烯基基团、芳基基团或杂芳基基团。杂芳基支架的具体实施方式包括三嗪,如1,3,5-三嗪、1,2,3-三嗪或1,2,4-三嗪;嘧啶;吡嗪;哒嗪;呋喃;吡咯;吡咯啉;吡咯烷;噁唑;吡唑;异噁唑;吡喃;吡啶;吡啶;吡啶;或嘌呤。

[0146] 通过任何有用的方法将支架S操作地连接至标签。在一个实例中,S是直接连接至帽部分的三嗪。为了获得该示例性支架,将三氯三嗪(即具有三个氯的三嗪的氯化前体)与帽部分的亲核基团反应。使用这种方法,S具有有着可以用于取代的氯的三个位置,其中两个位置可用多样性节点且一个位置连接至帽部分。接着,将结构单元 A_n 添加至支架的多样性节点,并且将编码结构单元 A_n 的标签 A_n (“标签 A_n ”)连接至帽部分,其中可以任何顺序进行这两个步骤。然后,将结构单元 B_n 添加至剩余多样性节点,并且将编码结构单元 B_n 的标签 B_n 连接至标签 A_n 的末端。在另一个实例中,S是操作地连接至标签的三嗪,其中三氯三嗪与PEG的亲核基团(例如,氨基基团)、标签的脂族或芳族接头反应。如上所描述的,可以添加结构单元和关联标签。

[0147] 在另一个实例中,S是操作地连接至结构单元 A_n 的三嗪。为了获得该支架,具有两个多样性节点的结构单元 A_n (例如,亲电基团和亲核基团,如Fmoc-氨基酸)与接头的亲核基团反应(例如,PEG的末端基团、脂族或芳族接头,其连接至帽部分)。然后,三氯三嗪与结构单元 A_n 的亲核基团反应。使用这种方法,S的所有三个氯位置被用作用于结构单元的多样性节点。如本文所描述的,可以添加另外的结构单元和标签,并且可以添加另外的支架 S_n 。

[0148] 示例性结构单元 A_n 包括,例如,氨基酸(例如, α -、 β -、 γ -、 δ -和 ϵ -氨基酸,以及天然和非天然氨基酸的衍生物)、与胺的化学反应的反应物(例如,叠氮或炔链)或硫醇反应物或它们的组合。结构单元 A_n 的选择取决于例如在接头中使用的反应性基团的性质、用于化学合成的支架部分和溶剂的性质。

[0149] 示例性结构单元 B_n 和 C_n 包括化学实体的任何有用结构单元,如可选取代的芳族基团(例如,可选取代的苯基或苄基)、可选取代的杂环基团(例如,可选取代的喹啉基、异喹啉基、吲哚基、异吲哚基、氮杂吲哚基、苯并咪唑基、氮杂苯并咪唑基、苯并异噁唑基、吡啶基、哌啶基或吡咯烷基)、可选取代的烷基基团(例如,可选取代的直链或支链 C_{1-6} 烷基基团或可选取代的 C_{1-6} 氨基烷基基团)或可选取代的碳环基团(例如,可选取代的环丙基、环己基或环己烯基)。特别有用的结构单元 B_n 和 C_n 包括具有一个或多个反应性基团的那些,如具有为反应性基团或可被化学修饰以形成反应性基团的一个或可选取代基的可选取代的基团(例如,任何本文所描述的)。示例性反应性基团包括一种或多种胺($-NR_2$,其中每个R独立地为H或可选取代的 C_{1-6} 烷基)、羟基、烷氧基($-OR$,其中R是可选取代的 C_{1-6} 烷基,如甲氧基)、羧基($-COOH$)、酰胺或化学反应的取代基。例如在标签 B_n 或 C_n 中可以引入限制位点,其中可以通过进行PCR和用相应限制酶之一的限制消化来识别复合物。

[0150] ■ 帽部分

[0151] 在文库中,帽部分将每个化学实体操作地连接至它的编码寡核苷酸标签。通常,帽部分是具有两个可以进一步衍生的官能团的起始寡核苷酸,其中第一官能团将化学实体(或其组分)操作地连接至帽部分,并且第二官能团将一个或多个标签操作地连接至帽部分。双官能间隔子可以可选地用作帽部分和化学实体之间的间隔部分。

[0152] 帽部分的官能团可以用于形成与化学实体组分的共价键和与标签的另一个共价

键。该组分可以是小分子的任何部分,如具有多样性节点或结构单元的支架。可替代地,帽部分可以被衍生以提供终止于官能团(例如,羟基、胺、羧基、巯基、炔基、叠氮基或磷酸基团)中的间隔子(即,将帽部分与将要在文库中形成的小分子分离开的间隔部分),其用于形成与化学实体组分的共价连接。间隔子可连接至帽部分的5'-末端、在内部位置之一处或3'-末端。当间隔子连接至内部位置之一时,间隔子可以使用本领域中已知的标准技术被操作地连接至衍生的碱基(例如,尿苷的C5位置)或内部地放置在寡核苷酸内。在本文中描述示例性间隔子。

[0153] 帽部分可以具有任何有用的结构。帽部分可以是,例如,1至100个核苷酸的长度,优选5至20个核苷酸的长度,并且最优选5至15个核苷酸的长度。如本文所描述的,帽部分可以是单链或双链的,并且可由天然或修饰的核苷酸组成。例如,化学部分操作地连接至帽部分的3'-末端或5'-末端。在特定的实施方式中,帽部分包括由序列内的互补碱基形成的发夹结构。例如,化学部分可以被操作地连接至帽部分的内部位置、3'-末端或5'-末端。

[0154] 通常,帽部分包括5'-或3'-末端上的非自身互补序列,其允许通过聚合、酶连接或化学反应结合寡核苷酸标签。帽部分可以允许寡核苷酸标签的连接和可选纯化和磷酸化步骤。在添加最后标签之后,另外的衔接子(adapter)序列可添加至最后标签的5'-末端。示例性衔接子序列包括引物结合序列或具有标记(例如,生物素)的序列。在其中使用许多结构单元和相应标签的情况下(例如,100),可以在寡核苷酸合成步骤期间采用混合-和-拆分策略以创建所需数量的标签。用于DNA合成的此类混合-和-拆分策略是本领域中已知的。可以通过PCR扩增所得文库成员接着是相对于感兴趣的一个或多个靶标来选择结合实体。

[0155] 帽部分或复合物可以可选地包括一个或多个引物结合序列。例如,帽部分具有发夹环区中的序列,其用作用于扩增的引物结合区,其中引物结合区具有比帽部分中的序列更高的用于其互补引物(例如,其可以包括侧接的标识符区域)的熔融温度。在其他实施方式中,复合物包括在一个或多个标签(其编码一个或多个结构单元)两侧(either side)上的两个引物结合序列(例如,使得可以发生PCR反应)。可替代地,帽部分可以包含在5'-或3'-末端上的一个引物结合序列。在其他实施方式中,帽部分是发夹,并且环区形成引物结合位点或通过寡核苷酸与环的3'侧上的帽部分的杂交来引入引物结合位点。引物寡核苷酸,其包含与帽部分的3'-末端的同源区域并且在其5'-末端上带有引物结合区(例如,为了使得可以发生PCR反应)可以杂交至帽部分且可以包含编码结构单元的标签或另外的结构单元。引物寡核苷酸可以包含另外的信息,如随机核苷酸的区域,例如,2至16个核苷酸的长度,包括其用于生物信息学分析。

[0156] 帽部分可以可选地包括发夹结构,其中可以通过任何有用的方法来获得该结构。例如,帽部分可以包括形成分子间碱基配对伴侣的互补碱基,如通过沃森-克里克DNA碱基配对(例如,腺嘌呤-胸腺嘧啶和鸟嘌呤-胞嘧啶)和/或通过摆动碱基配对(例如,鸟嘌呤-尿嘧啶、肌苷-尿嘧啶、肌苷-腺嘌呤和肌苷-胞嘧啶)。在另一个实例中,帽部分可以包括修饰的或取代的核苷酸,与未修饰的核苷酸相比,修饰的或取代的核苷酸可以形成更高的亲和力双链体形成,此类修饰或取代的核苷酸是本领域中已知的。在另一个实例中,帽部分包括一个或多个交联碱基以形成发夹结构。例如,可以例如通过使用补骨脂素来交联单链内的碱基或不同双链中的碱基。

[0157] 帽部分或复合物可以可选地包括一个或多个标记(label),其允许检测。例如,帽

部分、一个或多个寡核苷酸标签和/或一个或多个引物序列可以包括同位素、放射成像试剂、标志物、示踪剂、荧光标记(例如,若丹明或荧光素)、化学发光标记、量子点和报告分子(例如,生物素或his-标签)。

[0158] 在其他实施方式中,帽部分或标签可被修饰以支持,在半水性、降低的水性或非水性(如,有机)条件下的溶解度。通过修饰,帽部分或标签的核苷酸碱基可呈现更疏水,修饰例如,具有脂肪链的T或C碱基的C5位置而不显著破坏与它们的互补碱基形成氢键的能力。示例性的修饰或取代的核苷酸是5'-二甲氧基三苯甲基-N4-二异丁基氨基亚甲基-5-(1-丙炔基)-2'-脱氧胞苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;5'-二甲氧基三苯甲基-5-(1-丙炔基)-2'-脱氧尿苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;5'-二甲氧基三苯甲基-5-氟-2'-脱氧尿苷、3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;和5'-二甲氧基三苯甲基-5-(苊-1-基-乙炔基)-2'-脱氧尿苷或3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺。

[0159] 此外,帽部分寡核苷酸可以穿插(intersperse)有促进有机溶剂中的溶解度的修饰。例如,偶氮苯亚磷酰胺可以将疏水部分引入进帽部分设计。疏水亚酰胺进帽部分的此类插入可以发生在分子内的任何地方。然而,如果用于标签去卷积,该插入不能干扰随后的在文库合成中使用另外的DNA标签的加标签或一旦选择完成的后续PCR或微阵列分析。本文所描述的此类添加至帽部分设计将呈现帽部分溶解于,例如,15%、25%、30%、50%、75%、90%、95%、98%、99%或100%有机溶剂。因此,添加疏水残基进帽部分设计允许改善在半水性或非水性(例如,有机)条件下的溶解度,同时使帽部分呈现胜任寡核苷酸加标签。此外,随后引入到文库的DNA标签也可以在T或C碱基的C5位置处修饰,使得它们也赋予文库更疏水和可溶解于用于文库合成的后续步骤的有机溶剂。

[0160] 在特定的实施方式中,帽部分和第一标签可以是相同的实体,即,多个帽部分-标签实体可构造为所有享有公共部分(例如,引物结合区)和所有不同的另一个部分(例如,编码区)。这些可以在“拆分”步骤中利用且在它们编码的事件发生后合并。

[0161] 在特定的实施方式中,帽部分可以编码信息,例如,通过包括编码第一拆分步骤的序列或编码文库特性的序列,如通过使用涉及特异性文库的特定序列。

[0162] ■寡核苷酸标签

[0163] 本文所描述的寡核苷酸标签(例如,标签或帽部分的部分或尾部分的部分)可以用于编码任何有用的信息,如分子、化学实体的部分、组分(例如、支架或结构单元)的添加、文库中的帽部分、文库的特性、一个或多个文库成员的用途(例如,文库等分试样中的成员的用途)和/或文库成员的来源(例如,通过使用来源序列)。

[0164] 寡核苷酸中的任何序列可以用于编码任何信息。因此,一个寡核苷酸序列可用作多于一个目的,如编码两个或多个类型的信息或提供起始寡核苷酸,其也编码一个或多个类型的信息。例如,第一标签可以编码第一结构单元的添加以及文库的识别。在另一个实例中,帽部分可以用于提供将化学实体操作地连接至标签的起始寡核苷酸,其中帽部分另外地包括编码文库特性的序列(即,文库识别序列)。因此,本文所描述的任何信息可以编码在分离的寡核苷酸标签中或可以组合且编码在相同寡核苷酸序列中(例如,寡核苷酸标签,如标签或帽部分)。

[0165] 结构单元序列编码结构单元的特性和/或以结构单元引导的结合反应的类型。该

结构单元序列包括在标签中,其中该标签可以可选地包括下文所描述的一个或多个类型的序列(例如,文库识别序列、用途序列和/或来源序列)。

[0166] 文库识别序列编码特定文库的特性。为了允许两个或多个文库的混合,文库成员可以包含一个或多个文库识别序列,如在文库识别标签(即,包括文库识别序列的寡核苷酸)中、在连接的标签中、在帽部分序列的部分中或在尾部分序列中。这些文库识别序列可以用于推断编码关系,其中标签的序列被翻译并且与化学(合成)历史信息关联。因此,这些文库识别序列允许两个或多个文库的混合在一起用于选择、扩增、纯化、测序等。

[0167] 用途序列编码在文库的个体等分试样中一个或多个文库成员的历史(即,用途)。例如,可以用不同的反应条件、结构单元和/或选择步骤处理单独的等分试样。具体地,该序列可以用于识别此类等分试样且推断它们的历史(用途)从而允许将具有不同的历史(用途)(例如,不同的选择实验)的相同文库的等分试样混合在一起用于混合在一起的样品一起用于选择、扩增、纯化、测序等的目的。这些用途序列可以包括在帽部分、尾部分、标签、用途标签(即,包括用途序列的寡核苷酸)或本文所描述的任何其他标签中(例如,文库识别标签或来源标签)。

[0168] 来源序列是任何有用长度(例如,约6个寡核苷酸)的简并的(随机的,随机产生的)寡核苷酸序列,其编码文库成员的来源。该序列用于随机细分文库成员,其另外在所有方面等同于由序列信息区分的实体,使得衍生自独特的祖先模板(例如,选定的文库成员)的扩增产物的观测可以与衍生自相同祖先模板(例如,选定的文库成员)的多个扩增产物的观测区分开来。例如,在文库形成之后且在选择步骤之前,每个文库成员可以包括不同的来源序列,如在来源标签中。选择后,所选定的文库成员可以被扩增以产生扩增产物,并且预期可以观测到包括来源序列(例如,在来源标签中)的文库成员的部分,并与每个其他文库成员中的来源序列相比。由于来源序列是简并的,每个文库成员的每个扩增产物应具有不同的来源序列。然而,在扩增产物中相同来源序列的观测可指示衍生自相同模板分子的多个扩增子。当期望在扩增前而不是扩增后确定编码标签的群体的统计数字(statistics)和统计数据(demographics)时,可以使用来源标签。这些来源序列可以包括在帽部分、尾部分、标签、来源标签(即,包括来源序列的寡核苷酸)或本文所描述的任何其他标签中(例如,文库识别标签或用途标签)。

[0169] 本文所描述的任何类型的序列可以包括在帽部分中。例如,帽部分可以包括一个或多个结构单元序列、文库识别序列、用途序列或来源序列。

[0170] 本文所描述的任何这些序列可以包含在尾部分中。例如,尾部分可以包括一个或多个文库识别序列、用途顺序或来源序列。

[0171] 本文所描述的任何标签可以包括在具有固定序列的5'-或3'-末端处或接近于具有固定序列的5'-或3'-末端的连接子。通过提供反应性基团(例如,化学反应性基团或光反应性基团)或通过提供用于允许连接的试剂(例如,在一个或多个连接子或交联寡核苷酸中嵌入部分或可逆反应性基团的试剂)的位点,连接子促进连接(例如,化学连接)的形成。每个5'-连接子可以是相同的或不同的,并且每个3'-连接子可以是相同的或不同的。在具有多于一个标签的示例性、非限制性复合物中,每个标签可以包括5'-连接子和3'-连接子,其中每个5'-连接子具有相同的序列并且每个3'-连接子具有相同的序列(例如,其中5'-连接子的序列可与3'-连接子的序列相同或不同)。连接子提供可以用于一个或多个连接的序

列。为了允许结合传递引物或杂交交联寡核苷酸,连接子可以包括允许连接的一个或多个官能团(例如,对于其聚合酶具有降低的通读或移位通过的能力的连接,如化学连接)。

[0172] 这些序列可以包括本文所描述的用于寡核苷酸的任何修饰,如一个或多个修饰,其促进在有机溶剂中的溶解度(例如,任何本文所描述的,如用于帽部分)、其提供天然磷酸二酯键的类似物(例如,硫代磷酸酯类似物)或其提供一个或多个非天然寡核苷酸(例如,2'-取代的核苷酸,如2'-O-甲基化的核苷酸和2'-氟核苷酸,或任何本文所描述的)。

[0173] 这些序列可以包括本文所描述的用于寡核苷酸的任何特征。例如,这些序列可以包括在小于20个核苷酸的标签中(例如,如本文所描述的)。在其他实例中,包括一个或多个这些序列的标签具有大约相同的质量(例如,每个标签都有离标签的特定组(其编码特定的变量)内之间平均质量约 $\pm 10\%$ 的质量);缺少引物结合(例如,恒定)区;缺少恒定区;或具有降低长度的恒定区(例如,长度小于30个核苷酸、小于25个核苷酸、小于20个核苷酸、小于19个核苷酸、小于18个核苷酸、小于17个核苷酸、小于16个核苷酸、小于15个核苷酸、小于14个核苷酸、小于13个核苷酸、小于12个核苷酸、小于11个核苷酸、小于10个核苷酸、小于9个核苷酸、小于8个核苷酸或小于7个核苷酸)。

[0174] 用于该长度的文库和寡核苷酸的测序策略可以可选地包括分别增加读取保真度或测序深度的串接(concatenation)或连锁(catenation)策略。特别是,已在对于SELEX的文献中描述选择缺少引物结合区的编码文库,如Jarosch等人,Nucleic Acids Res.34:e86(2006)中所描述的,其通过参考并入本文。例如,可以修饰(例如,在选择步骤之后)文库成员以包括在复合物的5'-末端上的第一衔接子序列和在复合物的3'-末端上的第二衔接子序列,其中第一序列基本上互补于第二序列且引起形成双链体。为了进一步提高产量,将两个固定悬挂核苷酸(例如,CC)添加至5'-末端。在特定的实施方式中,第一衔接子序列是5'-GTGCTGC-3'(SEQ ID NO:1),并且第二衔接子序列是5'-GCAGCACCC-3'(SEQ ID NO:2)。

[0175] ■ 连接

[0176] 本发明的连接存在于编码信息的寡核苷酸之间(例如,如帽部分和标签之间、两个标签之间或标签和尾部分之间),其中此类连接包括对于其聚合酶具有降低的通读或移位通过的能力的任何连接。示例性连接包括化学连接,其包括一个或多个化学反应性基团、光反应性基团、嵌入部分、交联寡核苷酸或可逆共反应性基团。

[0177] 可以测试连接以确定聚合酶是否具有降低的通读或移位通过该连接的能力。可以通过任何有用的方法测试这种能力,如液相色谱法-质谱法、RT-PCR分析、序列群体统计和/或PCR分析。

[0178] 在特定的实施方式中,化学连接包括使用一个或多个化学反应对以提供连接。示例性化学反应对是这样的对:其包括可选取代的炔基基团和可选取代的叠氮基团、经由惠斯更1,3-偶极环加成反应以形成三唑;具有 4π -电子体系的可选取代的二烯(例如,可选取代的1,3-不饱和化合物,如可选取代的1,3-丁二烯、1-甲氧基-3-三甲基甲硅烷基-1,3-丁二烯、环戊二烯、环己二烯或呋喃)和具有 2π -电子体系的可选取代的亲二烯体或可选取代的杂亲二烯体(例如,可选取代的烯基基团或可选取代的炔基基团)、经由迪尔斯-阿尔德反应以形成环烯;具有紧张的杂环亲电子体(例如,可选取代的环氧化物、氮丙啶、吡丙啶鎓离子(aziridinium ion)或表硫鎓离子(episulfonium ion))的亲核试剂(例如,可选取代的胺或可选取代的硫醇)、经由开环反应以形成杂烷基;具有碘基团的硫代磷酸酯基团,如在

包含5'-碘dT的寡核苷酸与3'-硫代磷酸酯寡核苷酸的夹板连接中;可选取代的氨基基团与醛基团或酮基团的反应,如3'-醛修饰的寡核苷酸(其可以可选地通过氧化市售获得的3'-甘油基修饰的寡核苷酸来获得)与5'-氨基寡核苷酸(即,在还原胺化反应中)或5'-胍寡核苷酸的反应;可选取代的氨基基团和羧酸基团或硫醇基团(例如,有或没有使用反式-4-(马来酰亚胺甲基)环己烷-1-羧酸琥珀酰亚胺酯(SMCC)或1-乙基-3-(3-二甲氨基丙基)碳二亚胺(EDAC)的对;可选取代的胍和醛或酮基团的对;可选取代的羟胺和醛或酮基团的对;或亲核试剂和可选取代的烷基卤化物的对。

[0179] 铂复合物、烷基化试剂或呋喃修饰的核苷酸也可用作化学反应性基团以形成链间或链内连接。可以在两种寡核苷酸之间使用此类试剂,并且其可以可选地存在于交联寡核苷酸中。

[0180] 示例性、非限制性铂复合物包括顺铂(顺式-二胺二氯铂(II)),例如,以形成GG链内连接),反铂(反式-二胺二氯铂(II)),例如,以形成GXG链间连接,其中X可以是任何核苷酸)、卡铂、吡铂(ZD0473)、奥马铂或奥沙利铂以形成例如GC、CG、AG或GG连接。任何这些连接可以是链间或链内连接。

[0181] 示例性、非限制性烷基化试剂包括氮芥(二氯甲二乙胺(mechlorethamine)),例如,以形成GG连接)、苯丁酸氮芥、美法仑、环磷酰胺、环磷酰胺的前药形式(例如,4-氢过氧基环磷酰胺和异环磷酰胺)、1,3-双(2-氯乙基)-1-亚硝基脲(BCNU、卡莫司汀)、氮丙啶(例如,丝裂霉素C、三乙撑蜜胺或三乙撑硫代磷酰胺(硫代-tepa)以形成GG或AG连接)、六甲蜜胺、烷基磺酸酯(例如,白消安以形成GG连接)或亚硝基脲(例如,2-氯亚硝基脲以形成GG或CG连接,如卡莫司汀(BCNU)、氯脲霉素(chlorozotocin)、洛莫司汀(CCNU)和司莫司汀(甲基-CCNU))。任何这些连接可以是链间或链内连接。

[0182] 呋喃修饰的核苷酸也可以用于形成连接。一旦原位氧化(例如,用N-溴代琥珀酰亚胺(NBS)),呋喃部分形成反应性氧代烯醛(oxo-enal)衍生物,其与互补碱基反应以形成链间连接。在一些实施方式中,呋喃修饰的核苷酸与互补A或C核苷酸形成连接。示例性、非限制性呋喃修饰的核苷酸包括任何2'-(呋喃-2-基)丙酰基氨基-修饰的核苷酸;或2-(呋喃-2-基)乙基乙二醇核酸的无环、修饰的核苷酸。

[0183] 光反应性基团也可用作反应性基团。示例性、非限制性光反应性基团包括嵌入部分、补骨脂素衍生物(例如,补骨脂素、HMT-补骨脂素或8-甲氧基补骨脂素)、可选取代的氰乙烯基咪唑基团、可选取代的乙烯基咪唑基团、可选取代的氰乙烯基基团、可选取代的丙烯酰胺基团、可选取代的双吡丙啶基团、可选取代的二苯甲酮(例如,4-苯甲酰基苯甲酸的琥珀酰亚胺酯或二苯甲酮异硫氰酸酯)、可选取代的5-(羧基)乙烯基尿苷基团(例如,5-(羧基)乙烯基-2'-脱氧尿苷)或可选取代的叠氮基团(例如,芳基叠氮或卤代芳基叠氮,如4-叠氮基-2,3,5,6-四氟苯甲酸(ATFB)的琥珀酰亚胺酯)。

[0184] 嵌入部分也可用作反应性基团。示例性、非限制性嵌入部分包括补骨脂素衍生物、生物碱衍生物(例如,黄连素、巴马汀、甲氧槲因、血根碱(例如,其亚铵或烷醇胺形式)或马兜铃内酰胺- β -D-葡萄糖苷(aristololactam- β -D-glucoside))、乙吡啶阳离子(例如,溴化乙锭)、吡啶衍生物(例如,原黄素、吡啶黄或安吡啶)、蒽环类衍生物(例如,多柔比星、表柔比星、柔红霉素(道诺霉素)、伊达比星和阿柔比星)或沙利度胺。

[0185] 对于交联寡核苷酸,可以使用任何有用的反应性基团(例如,本文所描述的)以形

成链间或链内连接。示例性反应性基团包括化学反应性基团、光反应性基团、嵌入部分和可逆共反应性基团。用于与交联寡核苷酸一起使用的交联试剂包括但不限于烷基化试剂(例如,如本文所描述的)、顺铂(顺式-二胺二氯铂(II))、反式-二胺二氯铂(II)、补骨脂素、HMT-补骨脂素、8-甲氧基补骨脂素、呋喃修饰的核苷酸、2-氟-脱氧肌苷(2-F-dI)、5-溴-脱氧胞苷(5-Br-dC)、5-溴-脱氧尿苷(5-Br-dU)、5-碘-脱氧胞苷(5-I-dC)、5-碘-脱氧尿苷(5-I-dU)、反式-4-(马来酰亚胺甲基)环己烷-1-羧酸琥珀酰亚胺酯、SMCC、EDAC或硫代乙酸琥珀酰亚胺基酯(SATA)。

[0186] 寡核苷酸也可被修饰以包含硫醇部分,其可与各种硫醇反应性基团如马来酰亚胺、卤素、碘乙酰胺反应,因此可以用于使两个寡核苷酸交联。硫醇基团可连接至寡核苷酸的5'-或3'-末端。

[0187] 对于在嘧啶(例如,胸苷)位置处的双链寡核苷酸之间的链间交联,可以选择嵌入、光反应部分补骨脂素。在用紫外光(约254nm)的照射之后,补骨脂素嵌入到双链体中并且与嘧啶形成共价链间交联,优选在5'-TpA位点处。补骨脂素部分可以共价连接至修饰的寡核苷酸(例如,通过烷烃链,如C₁₋₁₀烷基或聚乙二醇基团,如-(CH₂CH₂O)_nCH₂CH₂-,其中n是1至50的整数)。也可以使用示例性补骨脂素衍生物,其中非限制性衍生物包括4'-(羟基乙氧基甲基)-4,5',8-三甲基补骨脂素(HMT-补骨脂素)和8-甲氧基补骨脂素。

[0188] 可以修饰交联寡核苷酸的各个部分以引入连接。例如,寡核苷酸中的末端硫代磷酸酯也可以用于使两个相邻的寡核苷酸连接。卤代尿嘧啶/胞嘧啶也可用作寡核苷酸中的交联剂修饰。例如,2-氟-脱氧肌苷(2-F-dI)修饰的寡核苷酸可与含二硫化物的二胺或硫丙基胺反应以形成二硫化物连接。

[0189] 如下所述,可逆共反应性基团包括选自以下的那些:氰乙基呋唑基团、氰乙基基团、丙烯酰胺基团、硫醇基团或磺酰乙基硫醚。可选取代的氰乙基呋唑(CNV)基团也可用在寡核苷酸中以在互补链中交联到嘧啶碱基(例如,胞嘧啶、胸腺嘧啶和尿嘧啶、以及它们的修饰的碱基)。在366nm下照射后,CNV基团促进与相邻嘧啶碱基的[2+2]环加成,其引起链间交联。在312nm下照射逆转交联并由此提供用于寡核苷酸链的可逆交联的方法。非限制性CNV基团是3-氰乙基呋唑,其可以包括为羧基乙烯基呋唑核苷酸(例如,作为3-羧基乙烯基呋唑-1'-β-脱氧核苷-5'-三磷酸酯)。

[0190] 可以修饰CNV基团以用另一个反应性基团取代反应氰基团以提供可选取代的乙烯基呋唑基团。对于乙烯基呋唑基团的示例性非限制性反应性基团包括-CONR_{N1}R_{N2}的酰胺基团,其中每个R_{N1}和R_{N2}可以是相同或不同的,并且独立地为H和C₁₋₆烷基,例如,-CONH₂;-CO₂H的羧基基团;或C₂₋₇烷氧基羰基基团(例如,甲氧基羰基)。此外,反应性基团可位于乙烯基基团的α或β碳上。示例性乙烯基呋唑基团包括氰乙基呋唑基团,如本文所描述的;氨基乙烯基呋唑基团(例如,氨基乙烯基呋唑核苷酸,如3-氨基乙烯基呋唑-1'-β-脱氧核苷-5'-三磷酸酯);羧基乙烯基呋唑基团(例如,羧基乙烯基呋唑核苷酸,如3-羧基乙烯基呋唑-1'-β-脱氧核苷-5'-三磷酸酯);和C₂₋₇烷氧基羰基乙烯基呋唑基团(例如,烷氧基羰基乙烯基呋唑核苷酸,如3-甲氧基羰基乙烯基呋唑-1'-β-脱氧核苷-5'-三磷酸酯)。具有此类基团的另外的可选取代的乙烯基呋唑基团和核苷酸提供在美国专利号US 7,972,792和Yoshimura and Fujimoto,Org.Lett.10:3227-3230(2008)的化学式中,其都通过参考将其全部内容并入本文。

[0191] 其他可逆反应性基团包括硫醇基基团和另一个硫醇基团以形成二硫化物,以及硫醇基团和乙烯基砜基团以形成磺酰乙基硫醚。硫醇-硫醇基团可以可选地包括由具有双-(N-碘乙酰基)哌嗪基磺酰罗丹明的反应形成的连接。其他可逆反应性基团(例如,如某些光反应性基团)包括可选项代的二苯甲酮基团。非限制性实例是二苯甲酮尿嘧啶(BPU),其可以用于含BPU寡核苷酸双链体的链间交联的位点选择性形成和序列选择性形成。在加热后可以逆转该交联,提供用于两个寡聚核苷酸链的可逆交联的方法。

[0192] 在其他实施方式中,化学连接包括引入磷酸二酯键的类似物,例如,用于选择后PCR分析和测序。磷酸二酯的示例性类似物包括硫代磷酸酯连接(例如,如通过使用硫代磷酸酯基团和离去基团如碘基团所引入的)、磷酸胺连接或二硫代磷酸酯连接(例如,如通过使用二硫代磷酸酯基团和离去基团如碘基团所引入的)。

[0193] 对于本文所描述的任何基团(例如,化学反应性基团、光反应性基团、嵌入部分、交联寡核苷酸或可逆共反应性基团),该基团可以在寡核苷酸的末端处或接近于寡核苷酸的末端或在5'-和3'-末端之间并入。此外,一个或多个基团可存在于每个寡核苷酸中。当需要反应性基团对时,则寡核苷酸可以被设计成促进基团对之间的反应。在与嘧啶碱基共反应的氰乙基呋唑基团的非限制性实例中,第一寡核苷酸可以被设计成包括在5'-末端处或接近于5'-末端的氰乙基呋唑基团。在该实例中,第二寡核苷酸可以被设计成与第一寡核苷酸互补并包括在当第一和第二寡核苷酸杂交时与氰乙基呋唑基团对齐的位置处的共反应嘧啶碱基。本文的任何基团和具有一个或多个基团的任何寡核苷酸可以被设计成促进基团之间的反应以形成一个或多个连接。

[0194] ■ 双官能间隔子

[0195] 可以改变帽部分和化学实体之间的双官能间隔子以提供适当的间隔部分和/或增加帽部分在有机溶剂中的溶解度。多种间隔子是可市售获得的,其可以将帽部分与小分子文库结合。间隔子通常由直链或支链组成,并且可以包括C₁₋₁₀烷基、1至10个原子的杂烷基、C₂₋₁₀烯基、C₂₋₁₀炔基、C₅₋₁₀芳基、3至20个原子的环或多环体系、磷酸二酯、肽、寡糖、寡核苷酸、低聚物、聚合物或聚烷二醇(例如,聚乙二醇,如-(CH₂CH₂O)_nCH₂CH₂-,其中n是1至50的整数)或它们的组合。

[0196] 双官能间隔子可以提供文库的帽部分和化学实体之间的适当间隔部分。在某些实施方式中,双官能间隔子包括三个部分。部分1可以是反应性基团,其与DNA形成共价键,如,例如,羧酸,优选通过N-羟基琥珀酰亚胺(NHS)酯活化以与DNA上的氨基基团(例如,氨基修饰的dT)反应;亚酰胺,以修饰单链帽部分的5'或3'-末端(通过标准寡核苷酸化学作用的方法实现);化学反应对(例如,在Cu(I)催化剂存在下的叠氮基-炔烃环加成或任何本文所描述的);或硫醇反应性基团。部分2也可以是反应性基团,其与化学实体、结构单元A_n或支架形成共价键。此类反应性基团可以是,例如,胺、硫醇、叠氮或炔烃。部分3可以是可变长度的化学惰性间隔部分,在部分1和2之间引入。此类间隔部分可以是乙二醇单元(例如,不同长度的PEG)的链、烷烃、烯烃、多烯链或肽链。间隔子可以包含具有疏水部分(如,例如,苯环)的分支或嵌入以改善有机溶剂中帽部分的溶解度,以及用于文库检测目的荧光部分(例如,荧光素或Cy-3)。帽部分设计中的疏水残基可随着间隔子设计而变化以促进在有机溶剂中的文库合成。例如,设计帽部分和间隔子组合以具有适当的残基,其中辛醇:水系数(P_{oct})为例如1.0至2.5。

[0197] 可凭经验选择间隔子用于给定的小分子文库设计,使得可以在有机溶剂例如在15%、25%、30%、50%、75%、90%、95%、98%、99%或100%有机溶剂中合成文库。可以在文库合成之前使用模拟反应来改变间隔子以选择适当的链长度,其溶解有机溶剂中的帽部分。示例性间隔子包括以下那些:具有增加的烷基链长度、增加的聚乙二醇单元、具有正电荷(以中和帽部分上的负磷酸电荷)的分支种类或增加的疏水性的量(例如,添加苯环结构)。

[0198] 市售获得的间隔子的实例包括氨基-羧酸间隔子,如为肽的那些(例如,Z-Gly-Gly-Gly-OSu(N- α -苄氧羰基-(甘氨酸)₃-N-琥珀酰亚胺基酯)或Z-Gly-Gly-Gly-Gly-Gly-Gly-OSu(N- α -苄氧羰基-(甘氨酸)₆-N-琥珀酰亚胺基酯,SEQ ID NO:3)、PEG(例如,Fmoc-氨基PEG2000-NHS或氨基-PEG(12-24)-NHS)或烷烃酸链(例如,Boc- ϵ -氨基己酸-OSu);化学反应对间隔子,如结合肽部分的本文所描述的那些化学反应对(例如,叠氮高丙氨酸-Gly-Gly-Gly-OSu(SEQ ID NO:4)或炔丙基甘氨酸-Gly-Gly-Gly-OSu(SEQ ID NO:5))、PEG(例如,叠氮基-PEG-NHS)或烷烃酸链部分(例如,5-叠氮戊酸,(S)-2-(叠氮甲基)-1-Boc-吡咯烷-4-叠氮苯胺或4-叠氮基-丁-1-酸N-羟基琥珀酰亚胺酯);硫醇反应性间隔子,如为PEG的那些(例如,SM(PEG)_n-NHS-PEG-马来酰亚胺)、烷烃链(例如,3-(吡啶-2-基二磺酰基)-丙酸-OSu(3-(pyridin-2-yl-disulfanyl)-propionic acid-OSu)或6-(3'-[2-吡啶基二硫]-丙酰胺基)己酸磺基琥珀酰亚胺基酯);和用于寡核苷酸合成的亚酰胺,如氨基修饰剂(例如,6-(三氟乙酰氨基)-己基-(2-氰乙基)-(N,N-二异丙基)-亚磷酰胺)、硫醇修饰剂(例如,S-三苯甲基-6-巯基己基-1-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺或化学反应对修饰剂(例如,6-己炔-1-基-(2-氰乙基)-(N,N-二异丙基)-亚磷酰胺、3-二甲氧基三苯甲基氧基-2-(3-(3-炔丙基氧基丙酰氨基)丙酰氨基)丙基-1-O-琥珀酰基(3-dimethoxytrityloxy-2-(3-(3-propargyloxy propanamido)propanamido)propyl-1-O-succinoyl)、长链烷基氨基CPG或4-叠氮基-丁-1-酸N-羟基琥珀酰亚胺酯)。另外的间隔子是本领域中已知的,并且可以在文库合成期间使用的那些包括但不限于5'-O-二甲氧基三苯甲基-1',2'-双脱氧核糖-3'-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;9-O-二甲氧基三苯甲基-三甘醇,1-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;3-(4,4'-二甲氧基三苯甲基氧基)丙基-1-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺;和18-O-二甲氧基三苯甲基六乙二醇,1-[(2-氰乙基)-(N,N-二异丙基)]-亚磷酰胺。本文的任何间隔子可以串联地添加至不同的另一组合以产生不同期望长度的间隔子。

[0199] 间隔子也可以是分支的,其中分支的间隔子是本领域中公知的,并且实例可由对称或不对称的双体(doublers)或对称的三体(trebler)组成。例如参见,Newcome等人,Dendritic Molecules:Concepts,Synthesis,Perspectives,VCH Publishers(1996);Boussif等人,Proc.Natl.Acad.Sci.USA 92:7297-7301(1995);and Jansen et al.,Science 266:1226(1994)。

[0200] ■酶连接和化学连接技术

[0201] 各种连接技术可以用于添加支架、结构单元、间隔子、连接、标签和/或帽部分以产生复合物。因此,本文所描述的任何结合步骤可以包括任何有用的连接技术,如酶连接和/或化学连接。这些结合步骤可以包括将一个或多个标签添加至帽部分或复合物;将间隔子添加至帽部分;并将一个或多个支架或结构单元添加至帽部分或复合物。在特定的实施方

式中,用于任何寡核苷酸的连接技术提供所得产物,其可被转录和/或反转录以允许解码文库或用于用一个或多个DNA或RNA聚合酶进行模板依赖性聚合。

[0202] 通常,酶连接产生具有天然磷酸二酯键的寡核苷酸,其可被转录和/或逆转录。在本文中提供酶连接的示例性方法且包括使用一种或多种RNA或DNA连接酶,如T4RNA连接酶、T4DNA连接酶、CircLigase™ ssDNA连接酶、CircLigase™ II ssDNA连接酶和ThermoPhage™ ssDNA连接酶(Prokaryote Ltd., Reykjavik, Iceland)。

[0203] 化学连接也可以用于产生能够被转录或反转录的寡核苷酸。可能需要测试提供能够被转录或反转录的寡核苷酸的化学连接技术的效力。可以通过任何有用的方法测试这种效力,如液相色谱法-质谱法、RT-PCR分析和/或PCR分析。在特定的实施方式中,化学连接包括使用一个或多个化学反应以提供可以被转录或反转录的间隔部分。具体地,适用于化学反应的反应是用于连接方法的优选候选(Kolb等人, Angew. Chem. Int. Ed., 40:2004-2021 (2001); Van der Eycken等人, QSAR Comb. Sci., 26:1115-1326 (2007))。在一个实施方式中,连接的寡核苷酸包含聚合酶具有降低的通读或移位通过的能力的连接,例如,“不可读”连接。

[0204] ■ 以促进酶连接或化学连接的反应条件

[0205] 本文所描述的方法可以包括一个或多个反应条件,其促进帽部分和标签之间或两个标签之间的酶连接或化学连接。这些反应条件包括使用标签内的修饰核苷酸,如本文所描述的;使用具有不同长度的供体标签和受体标签且改变标签浓度;使用不同类型的连接酶,以及它们的组合(例如,CircLigase™ DNA连接酶和/或T4RNA连接酶),并改变它们的浓度;使用具有不同分子量的聚乙烯乙二醇(PEG)并改变它们的浓度;使用非PEG聚集(crowding)试剂(例如,甜菜碱或牛血清白蛋白);改变用于连接的温度和持续时间;改变各种试剂的浓度,其包括ATP、Co(NH₃)₆Cl₃和酵母无机焦磷酸盐;使用酶或化学磷酸化的寡核苷酸标签;使用3'-保护的标签;并使用预腺苷化标签。这些反应条件也包括化学连接。

[0206] 帽部分和/或标签可以包括一个或多个修饰的或取代的核苷酸。在优选实施方式中,帽部分和/或标签包括促进酶连接的一个或多个修饰的或取代的核苷酸,如2'-O-甲基核苷酸(例如,2'-O-甲基鸟嘌呤或2'-O-甲基尿嘧啶),2'-氟核苷酸或用作连接底物利用的任何其他修饰的核苷酸。可替代地,修饰帽部分和/或标签以包括一个或多个化学反应性基团来支持化学连接(例如,可选取代的炔基基团和可选取代的叠氮基基团)。可选地,标签寡核苷酸在末端处用化学反应性基团官能化,并且可选地,保护这些末端之一,使得该基团可独立地被处理并且可减少副反应(例如,减少的聚合副反应)。

[0207] 酶连接可以包括一个或多个连接酶。示例性连接酶包括CircLigase™ ssDNA连接酶(EPICENTRE Biotechnologies, Madison, WI)、CircLigase™ II ssDNA连接酶(也来自EPICENTRE Biotechnologies)、ThermoPhage™ ssDNA连接酶(Prokaryote Ltd., Reykjavik, Iceland)、T4RNA连接酶和T4DNA连接酶。在优选实施方式中,连接包括使用RNA连接酶或RNA连接酶和DNA连接酶的组合。连接可以进一步包括一种或多种可溶性多价阳离子,如Co(NH₃)₆Cl₃,结合一种或多种连接酶。

[0208] 在连接步骤之前或之后,可以出于三个原因而纯化复合物。首先,可以纯化复合物以去除未反应的帽部分或标签,其可导致交叉反应并将“噪音”引入到编码过程。第二,可以纯化复合物以去除任何试剂或未反应的原料,其可以抑制或降低连接酶的连接活性。例如,

磷酸可导致降低的连接活性。第三,可能需要去除引入到化学或连接步骤的实体以使得能够进行随后的化学或连接步骤。在本文中描述纯化复合物的方法。

[0209] 酶和化学连接可以包括聚乙二醇,其具有超过300道尔顿的平均分子量(例如,超过600道尔顿、3,000道尔顿、4,000道尔顿或4500道尔顿)。在特定的实施方式中,聚乙二醇具有从约3,000道尔顿至9,000道尔顿的平均分子量(例如,从3,000道尔顿至8,000道尔顿、从3,000道尔顿至7,000道尔顿、从3,000道尔顿至6,000道尔顿和从3,000道尔顿至5,000道尔顿)。在优选实施方式中,聚乙二醇具有从约3,000道尔顿至约6,000道尔顿的平均分子量(例如,从3,300道尔顿至4,500道尔顿、从3,300道尔顿至5,000道尔顿、从3,300道尔顿至5,500道尔顿、从3,300道尔顿至6,000道尔顿、从3,500道尔顿至4,500道尔顿、从3,500道尔顿至5,000道尔顿、从3,500道尔顿至5,500道尔顿和从3,500道尔顿至6,000道尔顿,如4,600道尔顿)。聚乙二醇可以任何有用的量存在,如从约25% (w/v) 至约35% (w/v),如30% (w/v)。

[0210] 在本发明的优选实施方式中,使用概述如下的连接方案,通过单链寡核苷酸与单链寡核苷酸的连接安装标签:帽部分:25 μ M (5'末端:5'-单磷酸/2'-OMe G,居间核苷酸:2'-脱氧,以及3'末端:2'-封闭/3'-封闭);标签:25 μ M (5'-末端:2'-OMe/5'-OH G,居间核苷酸:2'-脱氧,以及3'-末端:3'-OH/2'-OMe);Co (NH₃)₆Cl₃:1mM;PEG 4600:30% (w/v);T4RNA连接酶(Promega):1.5单位/ μ l;酵母无机焦磷酸酶:0.0025单位/ μ l;Tris:50mM;MgCl₂:10mM;ATP:1mM;pH:7.5;和水:余量。

[0211] 在进一步实施方式中,方案包括在37℃培养20小时。为了实际文库构建的目的,可以使用较高浓度的帽部分、标签和/或连接酶,并且以这些浓度的此类修饰将是本领域技术人员显而易见的。化学连接方法可以包括任何化学方法,通过操作地关联的连接,使得能够操作地关联两个编码标签而无论聚合酶是否能够通读或移位通过。

[0212] ■ 用于确定复合物的核苷酸序列的方法

[0213] 本发明的特征在于用于确定复合物的核苷酸序列的方法,使得可以在标签的序列和化学实体的结构单元(structural units)(或结构单元(building blocks))之间建立编码关系。特别是,可以由寡核苷酸中的碱基序列推断化学实体的特性和/或历史。使用这种方法,可以用特定的标签序列处理包括不同化学实体或成员(例如,小分子或肽)的文库。

[0214] 图1-图4提供用于确定复合物的核苷酸序列的各种示例性方法。这些图提供示例性方法以跨越不可读的连接和/或释放交联寡核苷酸,其引起寡核苷酸模板的形成。

[0215] 图1b提供使用传递引物的示例性方法。特别是,传递引物杂交至固定的恒定序列部分(例如,5'-连接子和3'-连接子),其中用于传递引物的每个序列可以是相同或不同的。传递引物跨越不可读的连接,并且延伸沿复合物的可读部分(即,编码序列,例如,一个或多个标签)进行以产生寡核苷酸片段。然后,可以连接传递引物和片段以形成模板,其可以可选地被扩增和测序。

[0216] 图2-图4提供示例性的不可读的连接。图2提供由补骨脂素形成的连接,其在由相邻5'-连接子和3'-连接子形成的双链体中反应。图3提供由交联寡核苷酸形成的连接,其杂交至相邻的5'-连接子和3'-连接子。

[0217] 本文所描述的任何连接可以是可逆或不可逆的。可逆连接包括光反应性连接(例如,氰乙烯基咪唑基团和胸苷,如在图4中)和氧化还原连接。在本文中描述另外的连接。

[0218] 在可替代的实施方式中,可以酶修复“不可读的”连接以产生可读或至少可移位的连接。酶修复过程是本领域技术人员公知的,并且包括但不限于嘧啶(例如,胸苷)二聚体修复机制(例如,使用光裂合酶或糖基化酶(例如,T4嘧啶二聚体糖基化酶(PDG)))、碱基切除修复机制(例如,使用糖基化酶、脱嘌呤/脱嘧啶(AP)核酸内切酶、Flap核酸内切酶或聚ADP核糖聚合酶(例如,人类脱嘌呤/脱嘧啶(AP)核酸内切酶,APE 1;核酸内切酶III(Nth)蛋白质;核酸内切酶IV;核酸内切酶V;甲酰氨基嘧啶[fapy]-DNA糖基化酶(Fpg);人类8-氧代鸟嘌呤糖基化酶1(α 同种型)(hOGG1);人类核酸内切酶VIII样1(hNEIL1);尿嘧啶-DNA糖基化酶(UDG);人类单链选择性单官能尿嘧啶DNA糖基化酶(SMUG1);和人类烷基腺嘌呤DNA糖基化酶(hAAG)),其可以可选地与用于修复的一种或多种核酸内切酶、DNA或RNA聚合酶和/或连接酶组合)、甲基化修复机制(例如,使用甲基鸟嘌呤甲基转移酶)、AP修复机制(例如,使用脱嘌呤/脱嘧啶(AP)核酸内切酶(例如,APE1;核酸内切酶III;核酸内切酶IV;核酸内切酶V;Fpg;hOGG1;和hNEIL1),其可以可选地与用于修复的一种或多种核酸内切酶、DNA或RNA聚合酶和/或连接酶组合)、核苷酸切除修复机制(例如,使用切除修复交叉互补蛋白质或切除核酸酶,其可以可选地与用于修复的一种或多种核酸内切酶,DNA或RNA聚合酶和/或连接酶组合)和错配修复机制(例如,使用核酸内切酶(例如,T7核酸内切酶I;MutS、MutH和/或MutL),其可以可选地与用于修复的一种或多种外切核酸酶、核酸内切酶、解旋酶、DNA或RNA聚合酶和/或连接酶组合)。商业酶混合物可以用于容易地提供这些类型的修复机制,例如,PreCR®Repair Mix(New England Biolabs Inc., Ipswich MA),其包括Taq DNA连接酶、核酸内切酶IV、Bst DNA聚合酶、Fpg、尿嘧啶-DNA糖基化酶(UDG)、T4PDG(T4核酸内切酶V)和核酸内切酶VIII。

[0219] ■ 用于标记编码文库的方法

[0220] 本发明的特征在于用于使寡核苷酸标签与化学实体操作地关联的方法,使得可以在标签的序列和化学实体的结构单元(structural units)(或结构单元(building blocks))之间建立编码关系。特别是,可以从寡核苷酸中的碱基序列推断化学实体的特性和/或历史。使用这种方法,可以用特定的标签序列编码包括不同化学实体或成员(例如,小分子或肽)的文库。

[0221] 通常,这些方法包括使用帽部分,其具有至少一个可化学阐述的官能团和至少一个单链寡核苷酸可以与其结合(或连接)的官能团。可以通过任何有用的方式如通过酶结合(例如,用一个或多个RNA连接酶和/或DNA连接酶连接)或通过化学结合(例如,通过两个官能团,如亲核试剂和离去基团之间的取代反应)实现结合。

[0222] 为了创建文库中的许多化学实体,包含帽部分的溶液可分为多个等分试样,然后放入多个物理上分离的隔室中,如多孔板的孔。通常,这是“拆分”步骤。在每个隔室或孔内,用每个等分试样内的单链标签进行连续化学反应和连接步骤。记录化学反应条件和单链标签的序列之间的关系。可以以任何顺序进行反应和连接步骤。然后,组合或“合并”经反应和连接的等分试样,并且可选地可以在此时进行纯化。可以可选地重复这些拆分和合并步骤。

[0223] 接着,如本文所描述的,可对于具体特征或功能测试和/或选择文库。例如,带标签的化学实体的混合物可分成至少两个群体,其中第一群体结合至特定的生物靶并且第二群体不结合至具体的生物靶(例如,通过负选择或正选择)。然后可以选择地捕获第一群体(例如,通过提供目标靶的柱上的洗脱或通过用目标靶的等分试样孵育),并且可选地进一步分

析或测试,如以可选的洗涤、纯化、负选择、正选择或分离步骤。

[0224] 最后,所选群体中的一个或多个成员(或化学实体)的化学历史可以通过操作地连接寡核苷酸的序列来确定。将序列与特定结构单元关联之后,该方法可以识别具有所选特征的文库的个体成员(例如,增加的结合至靶蛋白的倾向从而引起治疗效果)。对于进一步测试和优化,可以随后通过合成具有或不具有其关联寡核苷酸标签的识别文库成员来制备候选治疗化合物。

[0225] 本文所描述的方法可以包括任何数量的可选步骤以多样化文库或询问文库成员。对于本文所描述的任何加标签方法,可以用另外的“n”个数量的连接、分离和/或磷酸化步骤添加连续的“n”个数量的标签。示例性可选步骤包括使用一种或多种限制性核酸内切酶限制文库成员关联的编码寡核苷酸;例如,用任何修复酶、如本文所描述的那些酶修复关联的编码寡核苷酸;将一个或多个衔接子序列连接至用于文库成员关联的编码寡核苷酸(例如,如一个或多个衔接子序列)的末端的一端或两端以提供用于扩增和测序的引物序列或提供用于序列的固定化的标记,如生物素;使用反转录酶、转录酶或另一种模板依赖性聚合酶反转录或转录、可选地接着是反转录复合物中装配标签;使用例如PCR扩增复合物中装配标签;例如,通过使用细菌转化、乳液形成、稀释、表面捕获技术等产生复合物中装配标签的一个或多个群体的克隆分离株(clonal isolates);例如,通过使用克隆分离株做为用于核苷酸的模板依赖性聚合的模板扩增复合物中装配标签的一个或多个群体的克隆分离株;和例如通过使用克隆分离株做为用于用荧光标记的核苷酸的模板依赖性聚合的模板序列测定复合物中装配标签的一个或多个群体的克隆分离株。在本文中描述用于扩增和测序寡核苷酸标签的另外的方法。

[0226] 这些方法可以用于识别和发现任何数量的具有特定特征或功能的化学实体,例如,在选择步骤中。所期望的特征或功能可作用于将文库划分成至少两个部分的基础,该部分具有有着所期望功能的文库中至少一个成员或相关成员的伴随富集。在特定的实施方式中,方法包括识别小药物样文库成员,其结合或失活治疗目标的蛋白质。在另一个实施方式中,设计化学反应的序列,并且选择一组结构单元,使得在所定义的化学条件下,所选的结构单元的反应将产生组合的多种分子(或分子的文库),其中一种或多种分子可以具有作为用于特定蛋白质的治疗试剂的效用。例如,选择化学反应和结构单元以创建具有通常存在于激酶抑制剂中的结构基团的文库。在任何这些情况下,寡核苷酸标签编码文库成员的化学历史,并且在每种情况下,可由任何具体的标签组合来表示化学可能性的集合。

[0227] 在一个实施方式中,化学实体的文库或其部分与生物靶联系,在适合于文库的至少一个成员结合至靶的条件下,接着是去除不结合至靶的文库成员,并分析与靶关联的一个或多个寡核苷酸标签。该方法可以可选地包括通过本领域中已知的方法扩增标签。示例性生物靶包括酶(例如,激酶、磷酸酶、甲基化酶、去甲基化酶、蛋白酶和DNA修复酶)、蛋白质,涉及以下蛋白质:蛋白质相互作用(例如,用于受体的配体)、受体靶(例如,GPCR和RTK)、离子通道、细菌、病毒、寄生虫、DNA、RNA、朊病毒(prion)和碳水化合物。

[0228] 在另一个实施方式中,结合至靶的化学实体不进行扩增而直接分析。示例性的分析方法包括微阵列分析,其包括渐逝共振光子晶体分析;用于去卷积标签的基于珠粒的方法(例如,通过使用his-标签);无标记的光子晶体生物传感器分析(例如,来自SRU Biosystems, Inc., Woburn, MA的 **BIND®** Reader);或基于杂交的方法(例如,通过使用固定

化寡核苷酸的阵列,其互补于存在于标签的文库中的序列)。

[0229] 此外,化学反应对(或官能团)可容易地包括在固相寡核苷酸合成方案中,并且将支持寡核苷酸的有效化学连接。此外,所得连接的寡核苷酸可作为用于使用一种或多种聚合酶的模板依赖性聚合的模板。因此,可以修饰本文所描述的用于使编码文库加标签的任何结合步骤以包括一种或多种酶连接和/或化学连接技术。示例性连接技术包括酶连接,如使用一种或多种RNA连接酶和/或DNA连接酶;和化学连接,如使用化学反应对(例如,包括可选取代的炔基和叠氮基官能团的对)。

[0230] 此外,可以在拆分-混合步骤中组合一个或多个文库。为了允许两个或多个文库的混合,文库成员可以包含一个或多个文库识别序列,如在文库识别标签中、在连接标签中或作为帽部分序列的一部分,如本文所描述的。

[0231] 具有降低质量的方法

[0232] 当与双链标签相比时,从单链标签的降低质量引出对于单链编码策略的许多动机。降低的质量潜在地赋予若干好处,包括增加的溶解度、降低的成本、增加的反应性、增加的靶可达性、降低的流体力学半径、增加的分析评估的精度等。除了使用单链加标签方法,可以通过包括一个或多个下列项的使用来实现进一步降低质量:具有降低长度的一个或多个标签、恒定质量标签组、编码帽部分、缺少引物结合区和/或恒定区的一个或多个文库成员、具有降低恒定区的一个或多个文库成员或本文所描述的任何其他方法。

[0233] 为了最小化文库中成员的质量,可以降低一个或多个标签的长度,如降低至尽可能短以编码每个拆分尺寸的长度。特别是,标签可小于20个核苷酸(例如,小于19个核苷酸、小于18个核苷酸、小于17个核苷酸、小于16个核苷酸、小于15个核苷酸、小于14个核苷酸、小于13个核苷酸、小于12个核苷酸、小于11个核苷酸、小于10个核苷酸、小于9个核苷酸、小于8个核苷酸或小于7个核苷酸)。如下文实例中所述,较短的标签(例如,约10个核苷酸或更短)可以用于标签连接。

[0234] 也可以使用恒定质量策略,其可有助于在文库合成期间分析。此外,恒定质量标签组可以允许识别所有单个错误发生(例如,源自误读序列或源自标签的化学或酶连接的错误)和大部分多个错误发生。恒定质量单链标签组的长度和编码能力之间的关系(例如,支持特异性结构单元拆分尺寸或文库特性等的最小长度)概述于下表1中。因此,使用恒定质量标签组可以用于提供有益的编码能力,同时维持文库形成期间的错误识别。

[0235] 表1

[0236]

长度	碱基 #1	碱基 #2	碱基 #3	碱基 #4	组合
1	1	0	0	0	1
2	1	1	0	0	2
3	1	1	1	0	6
4	1	1	1	1	24
5	2	1	1	1	60
6	2	2	1	1	180
7	2	2	2	1	630
8	2	2	2	2	2,520
9	3	2	2	2	7,560
10	3	3	2	2	25,200
11	3	3	3	2	92,400
12	3	3	3	3	369,600
13	4	3	3	3	1,201,200
14	4	4	3	3	4,204,200
15	4	4	4	3	15,765,750
16	4	4	4	4	63,063,000
17	5	4	4	4	214,414,200
18	5	5	4	4	771,891,120
19	5	5	5	4	2,933,186,256
20	5	5	5	5	11,732,745,024

[0237] 为了最小化文库中的质量,帽部分可以用于不仅连接化学部分和标签,而且编码特定文库的特性或特定步骤。例如,帽部分可以编码信息,例如,编码第一拆分或文库的特性的多个帽部分,如通过使用涉及特异性文库的特定序列。

[0238] 此外,可以在一个或多个选择步骤中排除来自DNA编码化学实体的文库的引物结合(例如,恒定)区。然后,可以在选择后通过例如单链连接添加这些区域。一个示例性策略将包括提供在编码寡核苷酸的5'-末端处的化学实体,基于任何有用的特定特征或功能选择特定的化学实体,并且将尾部分寡核苷酸连接至编码寡核苷酸的3'-末端,该编码寡核苷酸包括引物结合序列并且可以可选地包含一个或多个标签,例如“用途”标签、“来源”标签等,如本文所描述的。该引物结合序列然后可以用于启动模板依赖性聚合以产生cDNA(或cRNA),其互补于所选择的文库成员。cDNA或cRNA随后将在其3'-末端处连接至寡核苷酸,该寡核苷酸包含引物结合序列,并且目前其编码信息通过引物结合序列侧接在两侧上,可以使用已建立的方法测序和/或扩增寡核苷酸,如任何本文所描述的。

[0239] 可以通过省略或降低一个或多个恒定序列(其分离编码标签)的尺寸来进一步最小化质量。单链连接无需待连接的末端之间或这些末端和夹板之间的互补关系。因此,不需要固定的序列以支持酶连接。标签之间的短固定区域可以有用于标签的信息学剖析(parsing)或其他在硅中(in silico)去卷积过程。

[0240] 用于编码文库中化学实体的方法

[0241] 本发明的方法可以用于合成具有不同数量的由寡核苷酸标签编码化学实体的文库。结构单元和编码DNA标签的实例可见于美国专利申请公开号US 2007/0224607,通过参考并入本文。

[0242] 从一个或多个结构单元和可选的支架形成每个化学实体。支架用于提供在特定几何形状中的一个或多个多样性节点(例如,提供三个节点的三嗪,其空间上设置在杂芳基环

或线性几何形状周围)。

[0243] 可以将结构单元及其编码标签直接或间接地(例如,经由间隔子)添加至帽部分以形成复合物。当该帽部分包括间隔子时,将结构单元或支架添加至间隔子的末端。当不存在间隔子时,可以将结构单元直接添加至帽部分或结构单元本身可以包括与帽部分的官能团反应的间隔子。在本文中描述示例性间隔子和帽部分。

[0244] 可以以任何有用的方式添加支架。例如,支架可以被添加至间隔子或帽部分的末端,并且可以将连续结构单元添加至支架的可用的多样性节点。在另一个实例中,首先将结构单元 A_n 添加至间隔子或帽部分,然后使支架 S 的多样性节点与结构单元 A_n 中的官能团反应。编码特定支架的寡核苷酸标签可以可选地添加至帽部分或复合物。例如,将 S_n 添加至 n 个反应容器中的复合物,其中 n 是大于1的整数,并且标签 S_n (即,标签 $S_1, S_2, \dots, S_{n-1}, S_n$)结合至复合物的官能团。

[0245] 可以在多重合成步骤中添加结构单元。例如,将可选具有另外的间隔子的帽部分的等分试样分到 n 个反应容器中,其中 n 是2或更大的整数。在第一步骤中,将结构单元 A_n 添加至每个 n 反应容器中(即,将结构单元 $A_1, A_2, \dots, A_{n-1}, A_n$ 添加至反应容器 $1, 2, \dots, n-1, n$ 中),其中 n 是整数,并且每个结构单元 A_n 是唯一的。在第二步骤中,将支架 S 添加至每个反应容器中,以形成 A_n-S 复合物。可选地,可以将支架 S_n 添加至每个反应容器中以形成 A_n-S_n 复合物,其中 n 是大于2的整数,并且每个支架 S_n 可以是唯一的。在第三步骤中,将结构单元 B_n 添加至含有 A_n-S 复合物的每个 n 反应容器中(即,将结构单元 $B_1, B_2, \dots, B_{n-1}, B_n$ 添加至含有 $A_1-S, A_2-S, \dots, A_{n-1}-S, A_n-S$ 复合物的反应容器 $1, 2, \dots, n-1, n$ 中),其中每个结构单元 B_n 是唯一的。在进一步的步骤中,可以将结构单元 C_n 添加至含有 B_n-A_n-S 复合物的每个 n 反应容器中(即,结构单元 $C_1, C_2, \dots, C_{n-1}, C_n$ 添加至含有 $B_1-A_1-S \dots B_n-A_n-S$ 复合物的反应容器 $1, 2, \dots, n-1, n$ 中),其中每个结构单元 C_n 是唯一的。所得文库将具有 n^3 个数量的具有 n^3 标签的复合物。以这种方式,可以使用另外的合成步骤来结合另外的结构单元以进一步多样性文库。

[0246] 在形成文库之后,所得复合物可以可选地纯化且经受聚合或连接反应,例如,到尾部分。该一般策略可以被扩展至包括另外的多样性节点和结构单元(例如, D, E, F 等)。例如,第一多样性节点与结构单元和/或 S 反应并且由寡核苷酸标签编码。然后,另外的结构单元与所得复合物反应,并且由另外的结构单元衍生随后的多样性节点,其由用于聚合或连接反应的引物编码。

[0247] 为了形成编码文库,在每个合成步骤之后或之前将寡核苷酸标签添加至复合物。例如,在将结构单元 A_n 添加至每个反应容器之前或之后,将标签 A_n 结合至帽部分的官能团(即,将标签 $A_1, A_2, \dots, A_{n-1}, A_n$ 添加至含有帽部分的反应容器 $1, 2, \dots, n-1, n$ 中)。每个标签 A_n 具有不同的序列,其与每个唯一的结构单元 A_n 关联,并且确定标签 A_n 的序列提供结构单元 A_n 的化学结构。以这种方式,另外的标签用于编码另外的结构单元或另外的支架。

[0248] 此外,添加至复合物的最后标签可以包括引物结合序列或提供官能团以允许结合(例如,通过连接)引物结合序列。引物结合序列可以用于扩增和/或测序复合物的寡核苷酸标签。用于扩增且用于测序的示例性方法包括聚合酶链式反应(PCR)、直链扩增(LCR)、滚环扩增(RCA)或本领域中已知的任何其他方法以扩增或测定核酸序列。

[0249] 使用这些方法,可以形成大文库,其具有大量的编码化学实体。例如,帽部分与间隔子和结构单元 A_n 反应,其包括1000种不同的变体(即 $n=1,000$)。对于每个结构单元 A_n ,将

DNA标签 A_n 连接或引物延伸至帽部分。可以在1,000孔板或 10×100 孔板中进行这些反应。可以合并所有反应物、可选地纯化并且拆分成第二组的板。接着,可以用结构单元 B_n 进行相同程序,其也包括1,000种不同的变体。可以将DNA标签 B_n 连接至 A_n -帽部分复合物,并且可以合并所有反应。所得文库包括 $A_n \times B_n$ 的 $1,000 \times 1,000$ 个组合(即1,000,000种化合物),其由1,000,000种不同的标签组合标记。相同方法可扩展为添加结构单元 C_n 、 D_n 、 E_n 等。所产生的文库然后可以用于识别结合至靶的化合物。可以可选地通过DNA标签的PCR和测序来评估结合至文库的化学实体的结构以识别所富集的化合物。

[0250] 该方法可以被修改以避免在添加每个结构单元之后加标签或避免合并(或混合)。例如,可以通过将结构单元 A_n 添加至 n 个反应容器,其中 n 是大于1的整数,并且将相同结构单元 B_1 添加至每个反应孔来修改该方法。这里,对于每个化学实体 B_1 是相同的,并且因此,不需要编码该结构单元的寡核苷酸标签。在添加结构单元之后,可合并或不合并复合物。例如,在结构单元添加的最后步骤后,不合并文库,并且分别筛选各池(pool)以识别结合至靶的一种或多种化合物。为了避免在合成之后合并所有反应物,例如可以使用**BIND**[®] Reader (来自SRU Biosystems, Inc.) 来以高通量形式监测传感器表面上的结合(例如,384孔板和1,536孔板)。例如,可以用DNA标签 A_n 编码结构单元 A_n ,并且可以通过其在孔板内的位置来编码结构单元 B_n 。然后可以通过使用结合试验(例如,使用**BIND**[®] Biosensor,也可以由SRU Biosystems, Inc. 获得或使用ELISA试验)、并且通过由测序、微阵列分析和/或限制消化分析进行分析 A_n 标签来识别候选化合物。该分析允许识别产生所期望的分子的结构单元 A_n 和 B_n 的组合。

[0251] 扩增方法可以可选地包括形成油包水乳液以创建多个水性微反应器。可以调整反应条件(例如,复合物的浓度和微反应器的尺寸)以提供(平均来说)具有化合物文库的至少一个成员的微反应器。每个微反应器也可以包含靶、能够结合至复合物或复合物的一部分(例如,一个或多个标签)和/或结合靶的单个珠粒和具有一个或多个必需试剂以进行核酸扩增的扩增反应溶液。在扩增微反应器中的标签之后,标签的扩增拷贝将结合至微反应器中的珠粒,并且可以通过任何有用的方法来识别涂覆的珠粒。

[0252] 一旦识别了来自结合至目标靶的第一文库的结构单元,可以以迭代的方式来制备第二文库。例如,可以添加一个或两个另外的多样性节点,并且创建并取样第二文库,如本文所描述的。可以根据需要重复该过程多次以创建具有所期望分子和药学性能分子。

[0253] 各种连接技术可以用于添加支架、结构单元、间隔子、连接和标签。因此,本文所描述的任何结合步骤可以包括任何有用的一种或多种连接术。示例性连接技术包括酶连接,如使用一种或多种RNA连接酶和/或DNA连接酶,如本文所描述的;和化学连接,如使用化学反应对,如本文所描述的。

[0254] 实施例1

[0255] 对于复合物的一般策略

[0256] 可以形成本发明的复合物以包括一个或多个连接,聚合酶具有降低的通读或移位通过其的能力。例如,可以形成复合物以包括(从5'至3'方向)化学实体(星标)、固定的恒定序列(例如,帽部分),接着是三个可变编码序列(例如,三个标签)和另一个固定的恒定序列(例如,尾部分),其中固定的恒定序列和可变编码序列之间的连接可以是不可读的(图1B)。

[0257] 可以通过任何有用的部分或官能团来形成不可读的连接。非限制性实例是补骨脂

素(图2),其可以形成5'-连接子和3'-连接子之间的不可读的连接。

[0258] 可以通过使寡核苷酸与氰乙烯基吡唑基团的可逆共反应性基团交联而形成不可读的连接。复合物可以包括(从5'至3'方向)化学实体(星标)、固定的恒定序列(例如,帽部分)、经由氰乙烯基吡唑基团的可逆共反应性基团交联(在约366nm处交联,以X标记)的交联寡核苷酸,接着是具有在5'-和3'-末端处的固定的恒定序列(例如,具有5'-连接子和3'-连接子的标签)的可变编码序列(图3)。具有可逆共反应性基团的示例性可逆反应(氰乙烯基吡唑基团和胸苷的对),其可存在于3'-连接子、5'-连接子和交联寡核苷酸中(图4)。其他不可读的连接可以通过使用修复酶来呈现去卷积。

[0259] 作为读取或去卷积处理的一部分,复合物可与化学处理和/或酶处理(例如,连接酶和可选的激酶)反应以连接标签。当相邻标签之间的结点为缺口时,则连接酶可以用于连接标签。可选地,激酶可以用于将末端5'-羟基基团转化为磷酸基团。当相邻标签之间的结点是需用一个或多个核苷酸延伸的空位时,则在连接标签(可选包括激酶)之前,可以将聚合酶用于延伸跨越空位。可选地,聚合酶具有降低的链位移(strand-displacement)活性。接着,释放交联寡核苷酸以形成模板(例如,使用约312nm下照射)。最后,可选地用PCR扩增模板且测序。

[0260] 对于任何这些复合物,每个可变编码序列可以可选地包括5'-连接子和3'-连接子,其都是固定的恒定序列。因为连接是不可读的,传递引物可以用于跨越连接并允许用聚合酶延伸,因此形成寡核苷酸片段。然后可以使用连接酶连接该片段,可选地用PCR扩增并测序。

[0261] 可替代地,可以使用重组介导的方法将不可读的连接的一侧上的编码序列复制为单个cDNA序列,其中重组事件有利地发生于源自相同祖先模板序列的序列。该机制保留标签关联信息并呈现从衍生的序列数据推断出的这些关联。产生不可读的连接(其可成为用于重组多发的cDNA的模板)的一种可能方法是在连接的任一侧建立重复同源区。

[0262] 实施例2

[0263] 使用氰乙烯基吡唑的光化学寡核苷酸交联以模拟一个加标签事件

[0264] 从Nihon Techno Service(日本)获得寡核苷酸CNVKJ,其含有两个氰乙烯基吡唑修饰(图5)。从Integrated DNA Technologies (IDT) (IA) 获得寡核苷酸CNVK2_P_TagB、CNVK2_NP_TagB和CNVK2_TagA。每个寡核苷酸溶解于水中至1mM的浓度。

[0265] 光化学交联条件如下:寡核苷酸CNVKJ、CNVK2_P_TagB和CNVK2_TagA在1.5mL天然色聚丙烯微量离心管(Fisher Scientific, 02-682-550)中以20uL或50uL等分试样,在500mM pH 7.0磷酸盐缓冲液中以各100uM以等摩尔比混合。然后使用UVL-21(4瓦)灯(95-0018-02)(Ultra Violet Products, CA)在4℃在365nm处照射每个反应2小时。然后通过10%变性PAAG上的电泳以及通过LCMS(Thermo Fisher Scientific)来分析反应产物(图6)。超过70%的起始原料转化成光化学交联寡核苷酸结合物,其具有16,894Da观测到的MW(计算出为16,890Da)。当在500mM pH 5.5磷酸盐缓冲液或500mM pH 9.5硼酸盐缓冲液中进行相同交联反应时,观测到类似的结果和产量。

[0266] 在10%变性PAAG上纯化CNVK2_P_TagB、CNVK2_TagA和CNVKJ的光化学交联寡核苷酸结合物,并且通过使用Spectroline E-Series EB-160C(6瓦)灯(Spectronics Corp., NY)在312nm处(在80℃)照射1小时、在1.5mL天然色聚丙烯微量离心管中(Fisher

Scientific, 02-682-550) 以50uL等分试样在80℃加热而去交联。通过LCMS分析产物。反应定量地产生不变的起始寡核苷酸模拟标签的观测到MW为6,578Da (计算出为6,579Da) 的CNVK2_P_TagB和观测到MW为5,558Da (计算出为5,558Da) 的CNVK2_TagA (图7)。

[0267] 为了研究多个含CNVK的交联寡核苷酸的解离,我们改编上述方案以制备具有两个CNVK寡聚物的复合物,CNVKJ和Splint_14_CNVK (Splint_14_CNVK是CGAXCGTGTCA XCG,其中X是CNVK,由Biosearch Technologies, CA合成),光化学交联到单个寡核苷酸,Double_CNVK_Template

[0268] (AAAAAAGTCGTGACACGTCGGAAAAA AAAAAACGGTGACACGGTCGAAAAA, IDT (IA))。由LCMS分析该复合物:观测到MW为25,079Da (计算出为25,055Da)。然后其溶解于水中至100uM浓度,并且由UV 312nm在80℃照射1小时,再次如上所述。通过LCMS分析产物以发现起始Double_CNVK_template的恢复:观测到MW为16,492Da (计算出为16,492Da) (图8)。

[0269] 实施例3

[0270] 氰乙烯基咪唑光化学交联的寡核苷酸的连接和光化学去交联,以模拟恢复来自光化学交联标签的序列信息

[0271] 如例2中所描述的,制备、光化学交联并凝胶纯化CNVK2_P_TagB, CNVK2_TagA和CNVKJ的复合物。

[0272] 将光化学交联的模拟标签对溶解于1×T4DNA连接酶缓冲液 (NEB, MA),并用T4DNA连接酶 (NEB, MA) 在37℃孵育1小时。如例2中所述,该反应的产物然后被去交联,并通过变性凝胶电泳和LCMS来分析 (图9)。以高产量观测到连接反应的产物,观测到MW为12,119Da (计算出为12,119Da)。

[0273] 也制备相应的光化学交联的非磷酸化的模拟标签对CNVK2_NP_TagB与CNVK2_TagA和CNVKJ,如上所述。该复合物的LCMS分析显示观测到的MW为16,503Da (计算出为16,499Da,图10)。将光化学交联复合物溶解于1×T4连接酶缓冲液 (NEB),并且用T4多核苷酸激酶和T4DNA连接酶 (NEB) 的混合物在37℃孵育1小时。该连接的磷酸化-连接反应的产物然后通过用UV在312nm处在80℃照射1小时进行光化学去交联,如例2中所述。通过变性凝胶电泳和LCMS观测该磷酸化-连接反应的产物,其显示观测到的MW为11,894Da (计算出的MW为11,896Da (在连接产物的5'末端处的另外的磷酸化)) (图10)。

[0274] 实施例4

[0275] 茎内使用补骨脂素的光化学寡核苷酸交联以模拟一个加标签事件

[0276] 从Biosearch, CA获得寡核苷酸5PS02_A9_TA,其在5'-末端处用C2-补骨脂素修饰 (图11)。从IDT (IA) 获得寡核苷酸PS0_HP_A9_TCT (图11)。在500mM pH 7.0磷酸盐缓冲液中,两种寡核苷酸溶解至各至0.5mM浓度,并通过加热至95℃退火,并且随后冷却至室温。然后使用UVL-21紧凑UV灯 (UVP) 以UV光在365nm处在4℃照射该反应物30分钟。通过LCMS分析产物。我们观测到光化学交联产物的迅速形成,其具有17,643Da的观测到的MW (计算出的MW为17,643.4Da,参见图12)。

[0277] 实施例5

[0278] 使用补骨脂素的光化学寡核苷酸连接以模拟一个加标签事件

[0279] 从IDT (IA) 获得寡核苷酸Tag1_PsoCVU和SplintC_PsoC2 (图13)。从Biosearch (CA) 获得寡核苷酸5PSOC2_A9_GA (在5'-末端处用C2补骨脂素修饰) (图13)。

[0280] 如下制备Tag1_PsoCVU和5PsoC2_A9_GA的光化学连接寡核苷酸结合物。通过一个寡核苷酸的5'-补骨脂素吡喃酮部分与第二寡核苷酸的3'胸苷的光化学反应、在设计为杂交至第一和第二寡核苷酸两者且共定位它们的反应末端的第三寡核苷酸的存在下完成光连接(图14)。

[0281] 光化学连接条件如下:在1.5mL天然色聚丙烯微量离心管(Fisher Scientific, 02-682-550)中、在500mM pH 7.0磷酸盐缓冲液中、在10uL光连接反应中组合寡核苷酸5PsoC2_A9_GA(最终为1mM)和Tag1_PsoCVU(最终为1.1mM)和SplintC_PsoC2(最终为1.1mM)。先将反应混合物加热至95℃接着是缓慢冷却至室温。然后使用UVL-21紧凑UV灯(UVP)用UV光在365nm处在4℃照射反应物。随着时间-进程取1uL的等分试样,并且在15%TBE/8M尿素变性聚丙烯酰胺凝胶上分析(图15)。使用荧光TLC板的UV遮蔽可视化且拍照凝胶。以高产量形成5PsoC2_A9_GA和Tag1_PsoCVU的光连接结合物。24小时后,产量约为90%(图15)。

[0282] 通过LCMS分析24小时样品(图16)。主要的LC峰对应于5PsoC2_A9_GA和Tag1_PsoCVU的光连接结合物,其具有15,252.4Da的观测到的MW(计算出的MW为15,243.1Da)。

[0283] 实施例6

[0284] 使用末端引物延伸、非末端引物延伸和连接,从非聚合酶可读的光化学连接的补骨脂素结合的模拟标签产生cDNA,以模拟单个加标签事件的读取

[0285] 使用类似于实施例5中描述的方法产生补骨脂素-胸苷光化学连接的寡核苷酸结合物(Bio_Pso),其具有以下两个模拟标签寡核苷酸:5'-生物素化的寡核苷酸5Bio_Tag_PsoCVU和5'-补骨脂素寡核苷酸5PsoC2_A9_GA以及夹板寡核苷酸SplintC_PsoC2,从IDT(IA)获得所有三个寡核苷酸,并示于图17。24小时后,在10%TBE尿素变性PAAG上纯化并通过LCMS分析反应物Bio_Pso,显示具有观测到的MW为15,834.0Da(计算出的MW为15,812.7Da)的产物(图18)。

[0286] 将补骨脂素-胸苷光化学连接的寡核苷酸结合物Bio_Pso用作模板,以用T4DNA聚合酶(NEB,MA)、通过末端寡核苷酸引物(FAMprimer)和非末端寡核苷酸引物(Phos-SplintC_PsoC2)二者的酶延伸产生cDNA。使用T4PNK(NEB,MA)通过SplintC_PsoC2的5'-磷酸化来产生Phos-SplintC_PsoC2。在1×T4DNA连接酶缓冲液中(NEB,MA)在含有各10uM的FAMprimer、Phos-SplintC_PsoC2和Bio_Pso的100uL反应中进行cDNA产生,该反应中补充有1mM的每种dNTP和10uL的T4DNA聚合酶(NEB,MA)。在37℃孵育反应混合物1小时,然后补充0.5mM ATP和5uL的T4DNA连接酶(2,000u/uL,NEB,MA),然后在37℃孵育另外1小时。

[0287] 然后用200uL的链霉亲和素涂覆的DynaBeads M280(Invitrogen,在室温下用PBS预洗涤1小时)孵育反应混合物,用1mL的PBS洗涤珠粒,并且用35uL的100mM NaOH洗脱产物。洗脱液然后通过添加5uL的1M Tris HCl pH 7.0中和,并且通过LCMS分析。

[0288] 所得产物的LCMS分析,用在495nm处检测,表明约50%的FAMprimer已延伸至与Bio_Pso模板减去单个dA核苷酸(观测到的MW为15,170Da,预期全长MW为15,458Da)的全长互补序列。大部分其余的FAMprimer被延伸直到光化学连接结点(观测到的MW为11,154Da,计算出的MW为11,154Da)(图19)。

[0289] 当进行上述cDNA-产生而没有添加连接酶时,未观测到全长互补序列(减去单个dA)(图20)。

[0290] 实施例7

[0291] 使用羧乙烯基尿苷的光化学寡核苷酸连接以模拟一个加标签事件

[0292] 由Trilink (CA) 合成寡核苷酸CVU_G和CVU_A (图21)。两种寡核苷酸用5-(羧基) 乙烯基-2'-脱氧尿苷来5'-修饰 (图21)。从IDT DNA (IA) 获得寡核苷酸Tag1_PsoCVU、SplintC_CVU和SplintA_CVU (图21)。

[0293] 在照射的5'-5-(羧基) 乙烯基-2'-脱氧尿苷和3'-胸苷之间形成的光化学连接结点的结构示于图22。

[0294] 在1.5mL天然色聚丙烯微量离心管 (Fisher Scientific, 02-682-550) 中用溶解于500mM的pH 7.0磷酸钠缓冲液中的1mM CVU_G、1.1mM Tag1_PsoCVU和1.1mM SplintC_CVU来产生10uL光化学连接反应混合物。将反应混合物加热至95℃, 接着是缓慢冷却至室温。然后使用UVL-21紧凑UV灯 (UVP) 由UV光在365nm处在4℃照射反应物24小时。随着时间-进程取1uL的等分试样, 并且在15%TBE/8M尿素变性聚丙烯酰胺凝胶上电泳 (图23)。通过荧光TLC板的UV遮蔽可视化且拍照凝胶。照射24小时后, 超过90%的CVU_G寡核苷酸已光化学连接。

[0295] 通过LCMS分析24小时的样品 (图24)。主峰对应于CVU_G与Tag1_PsoCVU的光化学连接产物。观测到的MW为15,260.6Da (计算出的MW为15,238.7Da)。

[0296] 实施例8

[0297] 使用末端引物延伸、非末端引物延伸和连接, 从非聚合酶可读的光化学连接的羧乙烯基尿苷结合的模拟标签产生cDNA以模拟单个加标签事件的读取

[0298] 使用类似于在实施例7中描述的方法产生羧基乙烯基尿苷-胸苷光化学连接的寡核苷酸结合物 (Bio_CVU), 其具有以下两个模拟标签寡核苷酸: 5'-5-羧基乙烯基尿苷寡核苷酸5'-CVU_A和5'-生物素化的寡核苷酸5Bio_Tag_PsoCVU, 以及夹板寡核苷酸SplintA_CVU, 从IDT (IA) 获得所有三个寡核苷酸, 并示于图21和图25。如在实施例6中所描述的, 凝胶纯化反应产物, 并且LCMS分析显示具有观测到的MW为15,811.8Da (计算出的MW为15,792.3Da) 的产物 (图25)。

[0299] 将羧基乙烯基尿苷-胸苷光化学连接的寡核苷酸结合物Bio_CVU用作模板, 以用T4DNA聚合酶 (NEB, MA) 通过末端寡核苷酸引物 (FAMprimer) 和非末端寡核苷酸引物 (Phos-SplintA_CVU) 的酶延伸产生cDNA。使用T4PNK (NEB, MA) 通过SplintA_CVU的5'-磷酸化来产生Phos-SplintA_CVU。在1×T4DNA连接酶缓冲液 (NEB, MA) 中在含有10uM每FAMprimer、Phos-SplintA_CVU和Bio_CVU的100uL反应中进行cDNA产生, 该反应补充有1mM的每种dNTP和10uL的T4DNA聚合酶 (NEB, MA)。在37℃孵育反应混合物1小时, 然后补充0.5mM ATP和5uL的T4DNA连接酶 (2,000u/uL, NEB, MA), 然后在37℃孵育另外1小时。

[0300] 然后用200uL的链霉亲和素涂覆的DynaBeads M280 (Invitrogen, 在室温下用PBS预洗涤1小时) 培养反应混合物, 用1mL的PBS洗涤珠粒, 并且用35uL的100mM NaOH洗脱产物。然后通过添加5uL的1M Tris HCl pH 7.0中和洗脱液, 并且通过LCMS分析。

[0301] 所得产物的LCMS分析, 用在495nm处检测, 表明约80%的FAMprimer已延伸至与Bio_CVU模板减去单个dA核苷酸的全长互补序列 (观测到的MW为15,176Da, 计算出的全长MW为15,458Da)。大部分其余的FAMprimer延伸直到光化学连接结点之前的三个核苷酸 (观测到的MW为10,560Da, 计算出的MW为10,559Da) (图26)。

[0302] 当进行上述cDNA-产生而没有添加连接酶时, 未观测到全长互补序列 (减去单个

dA)。

[0303] 实施例9

[0304] 使用非聚合酶可读点击结合的寡核苷酸连接以模拟一个加标签事件

[0305] 从IDT (IA) 获得寡核苷酸S_IDTN₃和S_IDTalkyne (图27)。

[0306] 寡核苷酸S_IDTN₃的3'-叠氮修饰是IDT专有的修饰 (3' IDT叠氮修饰/3叠氮N/ (MW 350))。

[0307] HPLC-纯化含/3叠氮N/修饰的寡核苷酸。

[0308] 将含己炔基修饰的寡核苷酸脱盐并使用而无需进一步纯化。

[0309] 通过Cu (I) 催化的3+1叠氮基炔烃环加成来完成这些模拟标签的点击结合,如下:

[0310] 如下制备储备溶液:Cu (OAc)₂·H₂O (FW 181.6), 10mM在DMF中;抗坏血酸钠 (FW 198.1), 20mM在H₂O中和三-(苄基三唑基甲基)胺 (TBTA) (FW 530.6), 10mM在DMF中。向微量离心管中添加10uL pH7.0磷酸盐缓冲液 (500mM水溶液), 10uL的S_IDTalkyne (1mM水溶液, 10nmol, 1当量) 和10uL S-IDTN₃ (1mM水溶液, 10nmol, 1摩尔当量)。向该溶液中添加5uL的Cu_Pre-mix (2摩尔当量的Cu (OAc)₂, 4摩尔当量的抗坏血酸钠, 1摩尔当量的TBTA)。反应在室温下孵育过夜。通过LCMS (Thermo) 分析产物 (图28), 并且发现MW为26,304.8Da (计算值: 26,311.2Da)。通过使用10%PAAG/8M尿素的电泳进一步纯化结合物。该结合产物被称为Conjugate_Click_S。

[0311] 实施例10

[0312] 使用一对正交的非聚合酶可读的点击结合物的寡核苷酸连接以模拟一对正交加标签事件

[0313] 从IDT获得TKR_Central和S_IDTN₃寡聚物 (图29)。从Biosearch Technologies (CA) 获得5'-二苯并-环辛炔 (DBCO) 修饰的寡聚物TKR_DBCO_S (图29)。

[0314] 首先,使用环辛炔类无铜点击反应应用TKR_Central和TKR_DBCO_S进行点击结合。在0.2M pH 7.0磷酸钠缓冲液中将等摩尔量的每个寡核苷酸 (10uL的1mM水溶液) 混合在一起。通过LCMS监测反应进程。反应在30至60min内完全,产生超过90%的纯结合物。MS表明MW为25,336Da (计算出的25,341.8Da) (图30)。

[0315] 其次,使用在实施例9中描述的Cu (I) 催化的点击程序,在TKR_Central和TKR_DBCO_S的结合产物与S_IDTN₃之间进行点击结合。在10%TBE/8M尿素聚丙烯酰胺凝胶上纯化所得最终结合物2c1_S,并且通过LCMS (Thermo) 分析,其显示对于2c1_S的预期分子量为38,578.1Da (计算出的38,560.4Da) (图31)。

[0316] 例11

[0317] 在自由溶液中并且在乳化的水性液滴中使用非末端引物介导的重组,PCR扩增从非聚合酶可读的点击结合的模拟标签衍生的cDNA,以模拟从化学结合标签读取序列信息

[0318] 如在实施例9中所描述的,通过S_IDTN₃和S_IDTalkyne的结合来制备Conjugate_Click_S (图32)。

[0319] 类似于在实施例10中所描述的结合方案,通过L_IDTN₃:TGCGGTCTAACTGTCTAGGCACTTGTTTCGTTTGCCAGTGTGAGGAATGAACAGG/3叠氮N/和L_IDTalkyne:/5己炔基/TACCGAATTGCTTCCTCGTACAGTTCTAAGGCGCTTGGACACCACTTCAATCGGGTGCTATGCTT (IA, IDT) 的结合来制备Conjugate_Click_L (图32)。凝胶纯化后,LCMS证实MW为37,582.6Da (计算出的37,

584.4Da)。

[0320] 从(IA, IDT)获得寡脱氧核苷酸引物,包括正向引物: TGCGGTCTAACTGTCTA, 反向引物: AAG CATAGCACCCGATT和ePsplint: GGCAATTCGGTACCTGTTTCATTCC。

[0321] 将结合物Conjugate_Click_L和Conjugate_Click_S分别稀释到1nM,并且用作用于使用Deep Vent (外切) DNA聚合酶 (NEB, MA) 的夹板依赖性重组介导的cDNA产生/PCR扩增的模板。如下制备50uL Deep Vent (外切) 反应物: 5uL的10×Thermopol缓冲液; 2.5uL的每个正向和反向引物的10uM溶液, 1uL的100nM ePsplint引物; 2.5uL 10mM dNTP混合物 (NEB); 5uL的2nM结合物或它们的1:1混合物的稀释液, 2.5uL Deep Vent (外切) 并补加水至50uL。如下循环反应24次: 在94℃30秒、在52℃30秒和在72℃60秒。

[0322] 当在末端引物和夹板引物“ePsplint”的存在下在自由溶液中分别扩增Conjugate_Click_L或Conjugate_Click_S结合物、并且然后通过电泳评价时,每个引起具有与模板的长度关联的迁移率的单个产物(图33中的泳道5和泳道7)。缺少夹板引物“ePsplint”的类似扩增没有给出产物(图33中的泳道4和泳道6)。这表明通过Deep Vent (外切) 聚合酶的结合物的夹板依赖性重组介导cDNA产生/PCR扩增依赖于ePsplint引物的存在(图33)。对于该重组和扩增的可能机制示于图34。

[0323] 当在末端引物和夹板引物“ePsplint”的存在下在自由溶液中将Conjugate_Click_L和Conjugate_Click_S结合物二者混合并在一起扩增、并且然后通过电泳评价时,产生不同迁移率的产物的阶梯,其具有与每个模板的长度关联的迁移率、以及具有与每个模板之间的中间长度关联的迁移率(图35中的泳道5)。对于该重组和扩增的可能机制示于图36。

[0324] 当以相对于乳液液滴的低浓度结合物、在末端引物和夹板引物“ePsplint”二者的存在下、在油包水乳液中混合并在一起扩增Conjugate_Click_L结合物和Conjugate_Click_S结合物二者、并且然后通过电泳评价时,观测到一对产物,具有与到每个模板的长度相关联的迁移率(图37中的泳道3)。乳液的液滴用作个体微反应器,在其内扩增分离的个体结合物。通过在轻质矿物油 (Sigma, cat#M-3516) 中混合4.5% (vol/vol) 的Span 80 (Fluka, cat#85548)、0.4% (vol/vol) 的Tween 80 (Sigma, cat#S-8074) 和0.05% (vol/vol) 的Triton X-100 (Sigma, cat#T-9284) 来制备油/表面活性剂相。通过混合5uL的10×Thermopol缓冲液; 2.5uL的每个引物的10uM溶液、1uL的100nM ePsplint引物; 2.5uL 10mM dNTP混合物 (NEB); 2.5uL DVexo- (NEB)、2.5uL的BSA (NEB) 并补加水至50uL来制备水相(50uL)。将结合物Conjugate_Click_L和结合物Conjugate_Click_S的1:1混合物添加至该反应,使最终浓度为5pM。然后向50uL的水相添加150uL的非水相,并且通过涡旋5分钟进行乳化。如下进行重组/扩增过程: 在94℃30秒、在52℃30秒和在72℃60秒,进行32个循环。

[0325] 在夹板依赖性重组介导的cDNA产生/PCR扩增之后,通过以21,000g微量离心20分钟来破坏乳液,然后去除非水相,并且将水相和中间相萃取到氯仿中。然后通过非变性电泳(4%琼脂糖微凝胶, Invitrogen CA) 评价水相。

[0326] 乳液划区的夹板依赖性重组介导的cDNA产生/PCR扩增的结果表明,拖曳的(shuffled)PCR产物的形成的程度显著降低或不存在(泳道3, 图37)。

[0327] 例12

[0328] 在自由溶液中使用重复同源介导的重组,PCR扩增从非聚合酶可读点击结合的模

拟标签衍生的cDNA,以模拟单个加标签事件的读取

[0329] 两个结合物,TKR_S和TKR_L(图38),每个包含使用Cu(I)催化的环加成和在实施例9中所描述的方案、在3'-叠氮基修饰的寡核苷酸和5'-己炔基修饰的寡核苷酸之间形成的单个非聚合酶可读的点击结合连接。TKR_S和TKR_L各自包含重复的12聚体GGAATGAACAGG序列,其侧接非聚合酶可读的连接。通过使以下两个IDT DNA寡核苷酸结合来制备TKR_S:TKR_S_N₃:TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG/3叠氮N/(也称为S_IDTN₃,在例9中)和TKR_S_Hex:/5'-己炔基/GGAATGAACAGGGCCTGATTGAAGCAATCGGGTGCTATGCTT,而通过使以下两个IDT DNA寡核苷酸结合来制备TKR_L:TKR_L_N₃:TGCGGTCTAACTGTCTAGGCACTTGTTTCGTTTGCCAGTGTGATGGACACCACTTGAATGAACAGG/3叠氮N/和TKR_L_Hex:/5'-己炔基/GGAATGAACAGGGGAATGAACAGGTTCTCTCGTACAGTTCTAAGGCGCTCAATCGGGTGCTATGCTT。在变性条件下PAGE-纯化且通过LCMS分析两个结合物以显示TKR_S的分子量:26,447.5Da(计算出的26,449.2Da),以及TKR_L的分子量:37,582.6Da(计算出的37,584.4Da)。

[0330] 将两个结合物各自稀释到1nM,并且用于在具有正向引物TGCGGTCTAACTGTCTA和反向引物AAGCATAGCACCCGATT(IDT)的自由溶液中、用Deep Vent(外切)DNA聚合酶(NEB)的重复依赖性重组介导的cDNA产生/PCR扩增。

[0331] 使用Deep Vent(外切)DNA聚合酶(NEB)进行重复依赖性重组介导的cDNA产生/PCR扩增反应。5uL的10×Thermopol缓冲液;2.5uL的每个正向和反向引物的10uM溶液;2.5uL 10mM dNTP混合物(NEB);40pM最终浓度的结合物或它们的1:1混合物,2.5uL Deep Vent(外切)聚合酶并补加水至50uL的最终体积。如下热循环反应22次:在94℃30秒、在52℃30秒和在72℃60秒。

[0332] 在4%琼脂糖微凝胶(Invitrogen)上可视化产物(图39)。

[0333] Deep Vent(外切)聚合酶重复依赖性重组介导的cDNA产生/PCR扩增反应产生大量扩增子,其衍生自各自分离的扩增的结合物(泳道3和泳道4,图39)。在相同条件下,结合物的混合物的扩增未引起任何中间长度扩增产物的产生(泳道5,图39)。为了进一步证实在这类扩增产物中没有标签关联(编码信息)的混乱(损失),我们克隆且测序了从TKR_S和TKR_L的共扩增混合物衍生的扩增产物。这些序列数据证实在从共扩增结合物衍生的扩增产物中完全没有模拟标签对的混乱(图40)。

[0334] 实施例13

[0335] 在自由溶液中使用重复同源介导的重组,PCR扩增从不可读两次点击结合的标签衍生的cDNA,以模拟一对加标签事件的读取

[0336] 在类似于在实施例10中所描述的两步程序中制备两个结合物,2_c1_S和2_c1_L(图41),每个包含两个不可读点击结合连接,每个分别由不同的12聚体重复序列(GAAGGCGTTATG-点击-GAAGGCGTTATG)和(GGAATGAACAGG-点击-GGAATGAACAGG)侧接。

[0337] 首先,使用在TKR_central:/5'-己炔基/GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG/3叠氮N/(IDT)和TKR_DBCO_S:(DBCO)GAAGGCGTTATGTCCGTACTCTGCAATCGGGTGCTATGCTT或TKR_DBCO_L(DBCO)GAAGGCGTTATGTGATATCCGTGGTGCTGAGTTCCAATCGGGTGCTATGCTT之间的环辛炔基无铜点击反应进行结合。从Biosearch(CA)获得两个含DBCO寡核苷酸,DBCO是5'-二苯并-环辛炔修饰,对于其从Glen Research(VA)获得亚磷酸胺试剂。这些结合的产物分别命名为TKR_S_N₃和TKR_L_N₃。

[0338] 其次,使用Cu(I)催化的点击反应,在第一结合产物TKR_S_N₃和TKR_L_N₃之间进行结合。所得到的结合物,TKR_2_click_S和TKR_2_click_L(图41)使用电泳在10%TBE/8M尿素聚丙烯酰胺凝胶上纯化,并且通过LCMS(Thermo)分析,其显示对于TKR_2_click_S的预期分子量为38,578.1Da(计算出的38,560.4Da),并且对于TKR_2_click_L的预期分子量为49,913.3Da(计算出的49,916.7Da)(图41)。

[0339] 将两个结合物各自稀释到1nM,并且用于使用Deep Vent(外切)聚合酶(NEB)的重复依赖性重组介导的cDNA产生/PCR扩增。对于扩增,连同5uL的10×Thermopol缓冲液、使用正向引物TGCGGTCTAACTGTCTA和反向引物AAGCATAGCACCCGATT(IDT DNA);2.5uL的每个正向和反向引物的10uM溶液;2.5uL 10mM dNTP混合物(NEB);40pM最终浓度的结合物或它们的1:1混合物,2.5uL Deep Vent(外切)聚合酶并补加水至50uL。反应如下热循环22次:在94℃30秒、在52℃30秒和在72℃60秒。在4%琼脂糖凝胶(Invitrogen)上可视化扩增产物。

[0340] Deep Vent(外切)聚合酶重复依赖性重组介导的cDNA产生/PCR扩增反应产生衍生自每个分离扩增的结合物的大量扩增子(泳道3和泳道4,图42)。在相同条件下,结合物的混合物的扩增未产生任何中间长度扩增产物的产生(泳道5,图42)。为了进一步证实在这些扩增产物中缺少标签关联(编码信息)的混乱(损失),我们克隆并测序了从TKR_2_click_S和TKR_2_click_L的共扩增混合物衍生的扩增产物。这些序列数据证实在从共扩增结合物衍生的扩增产物中完全缺少模拟标签对的混乱(图43)。

[0341] ■其他实施方式

[0342] 在本说明书中提到的所有文献、专利申请和专利通过参考并入本文。

[0343] 本发明所述的方法和体系的各种修改和变型将对本领域技术人员显而易见只要不脱离本发明的范围和精神即可。尽管已结合具体期望实施方式进行了描述本发明,但应当理解,如所要求保护的本发明不应被过度地限于此类具体实施方式。事实上,用于实施本发明的所描述方法的各种修改对医学、药学或相关领域的本领域技术人员而言是显而易见的,都旨在处于本发明的范围内。

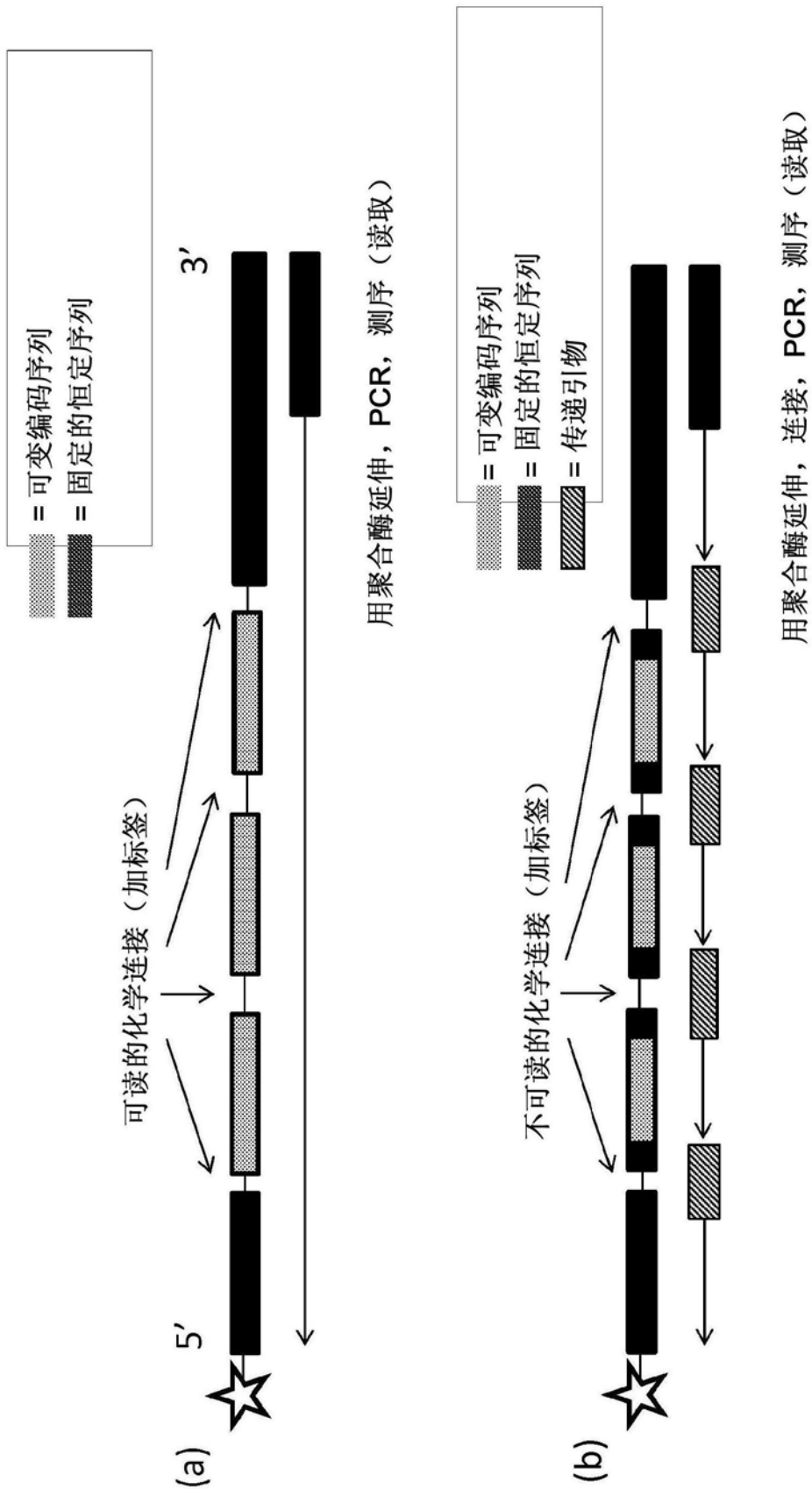


图1

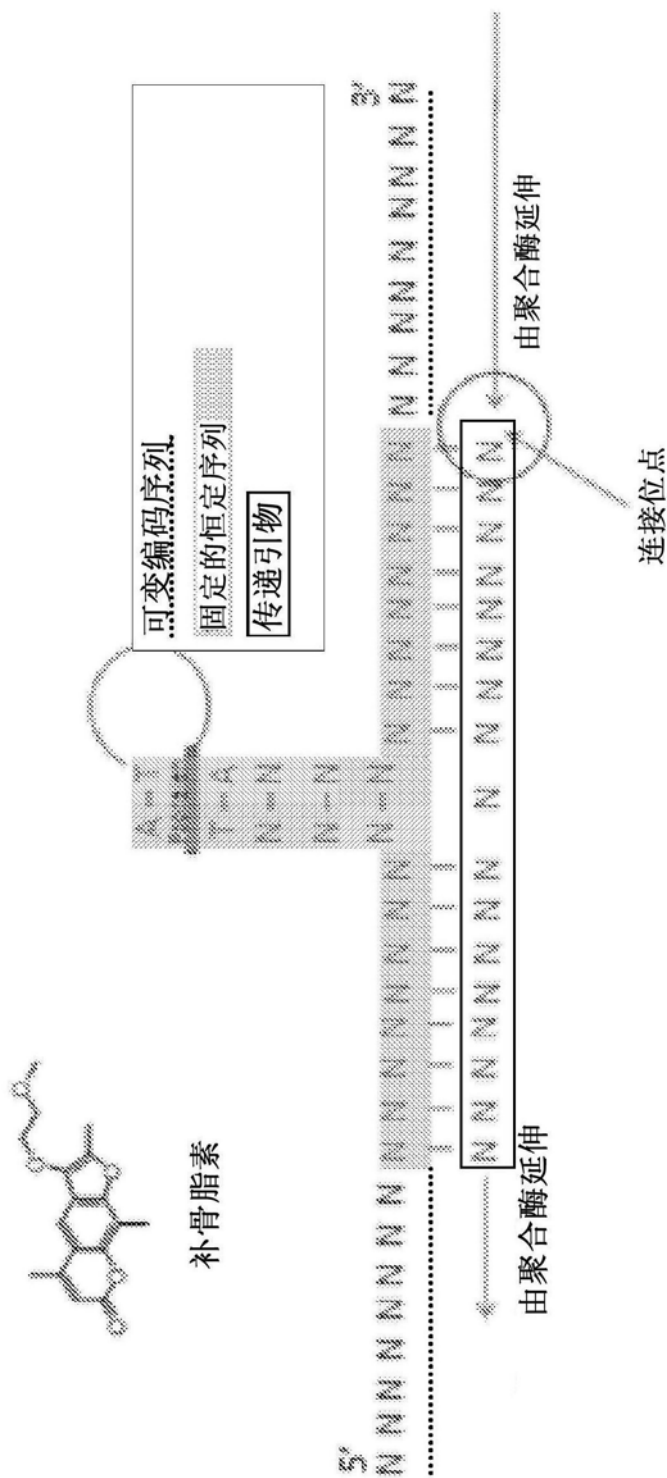


图2

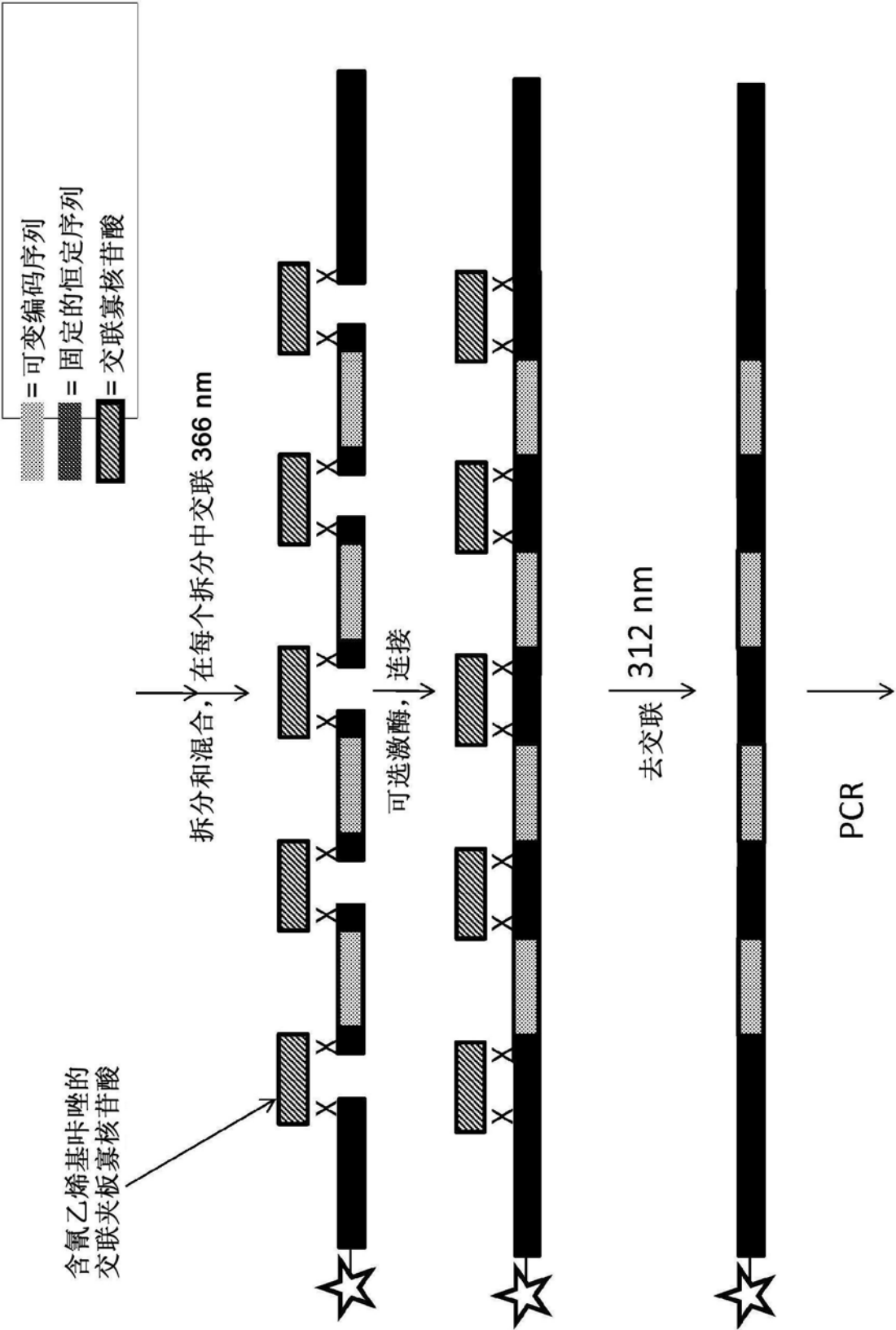


图3

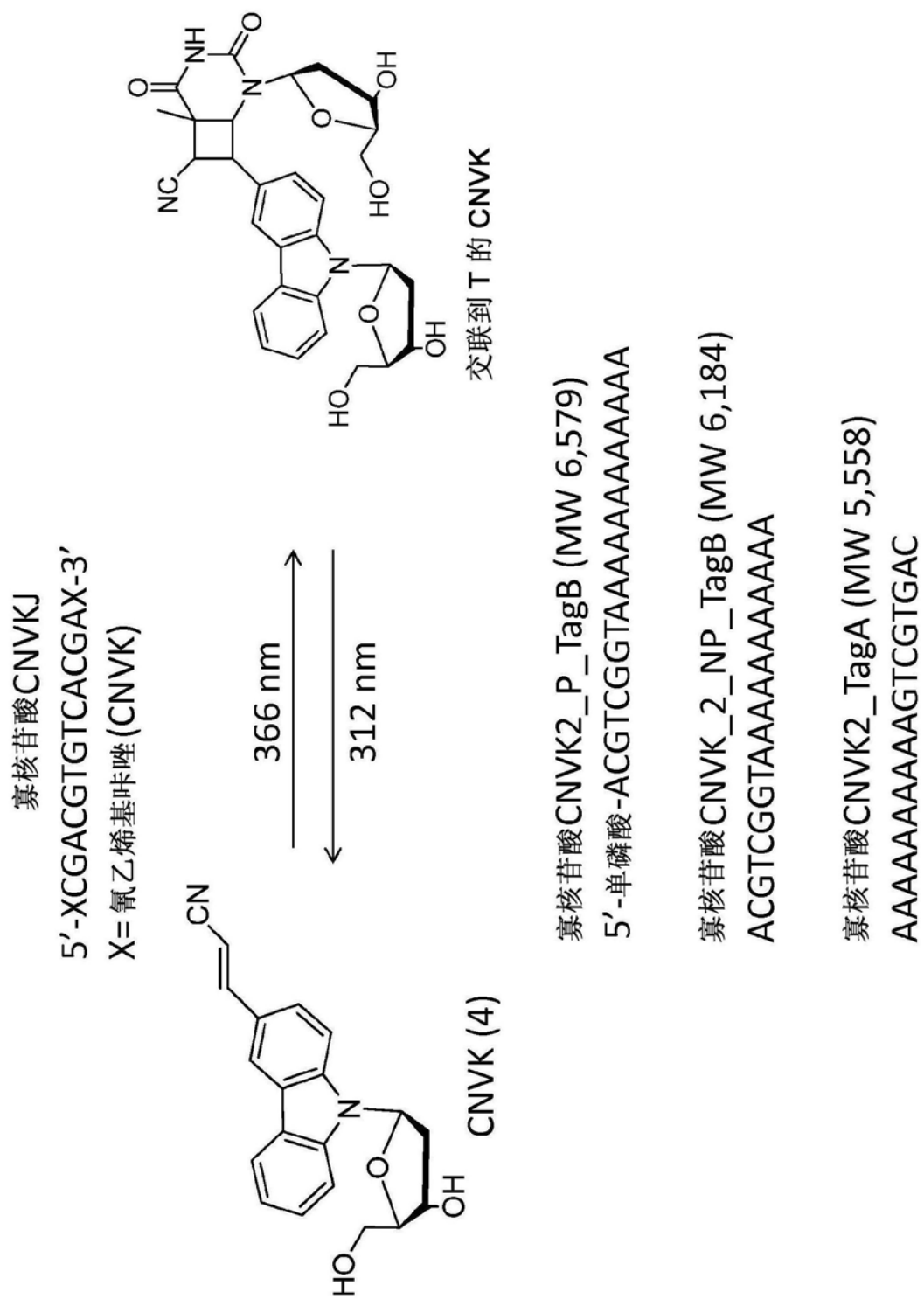


图5

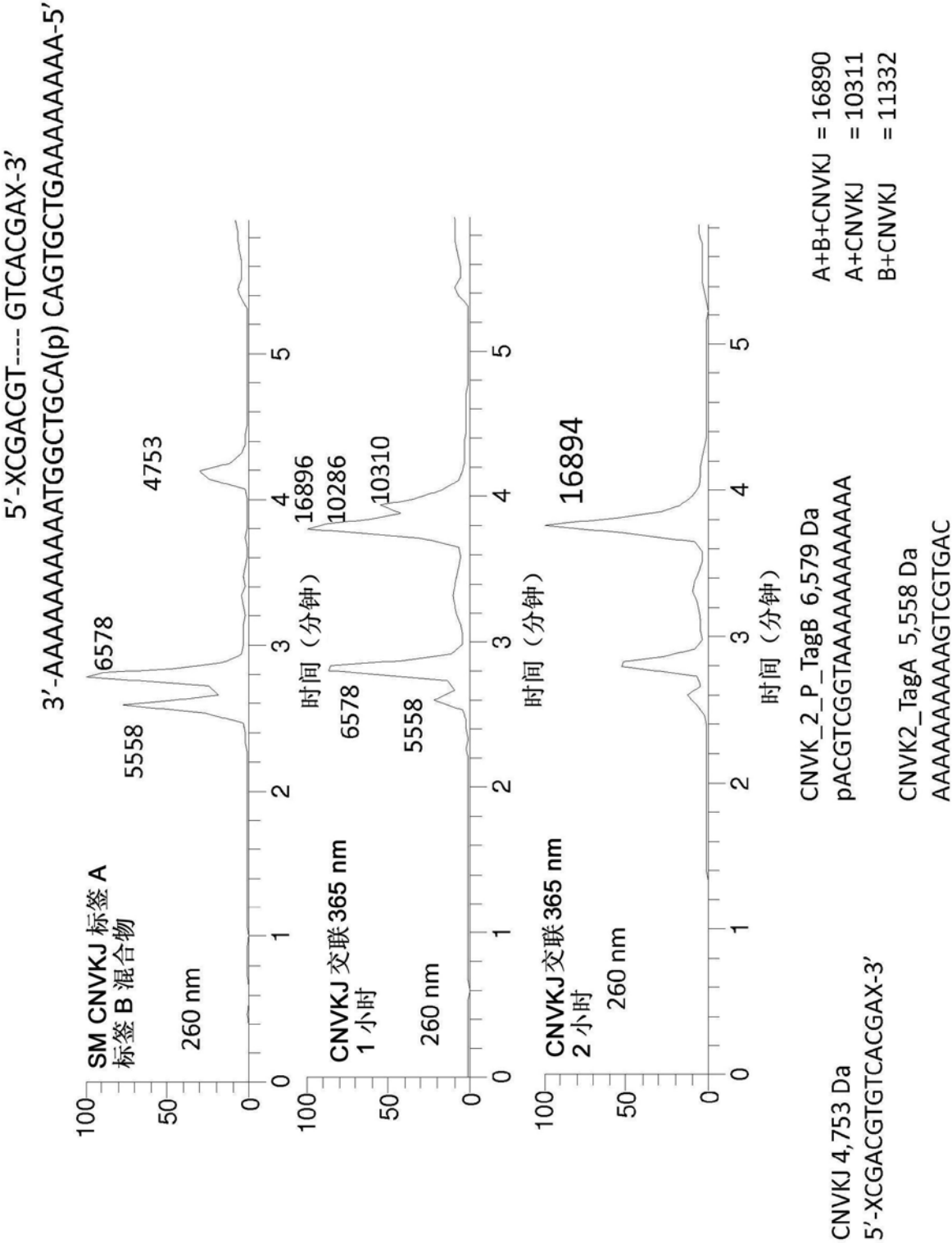


图6

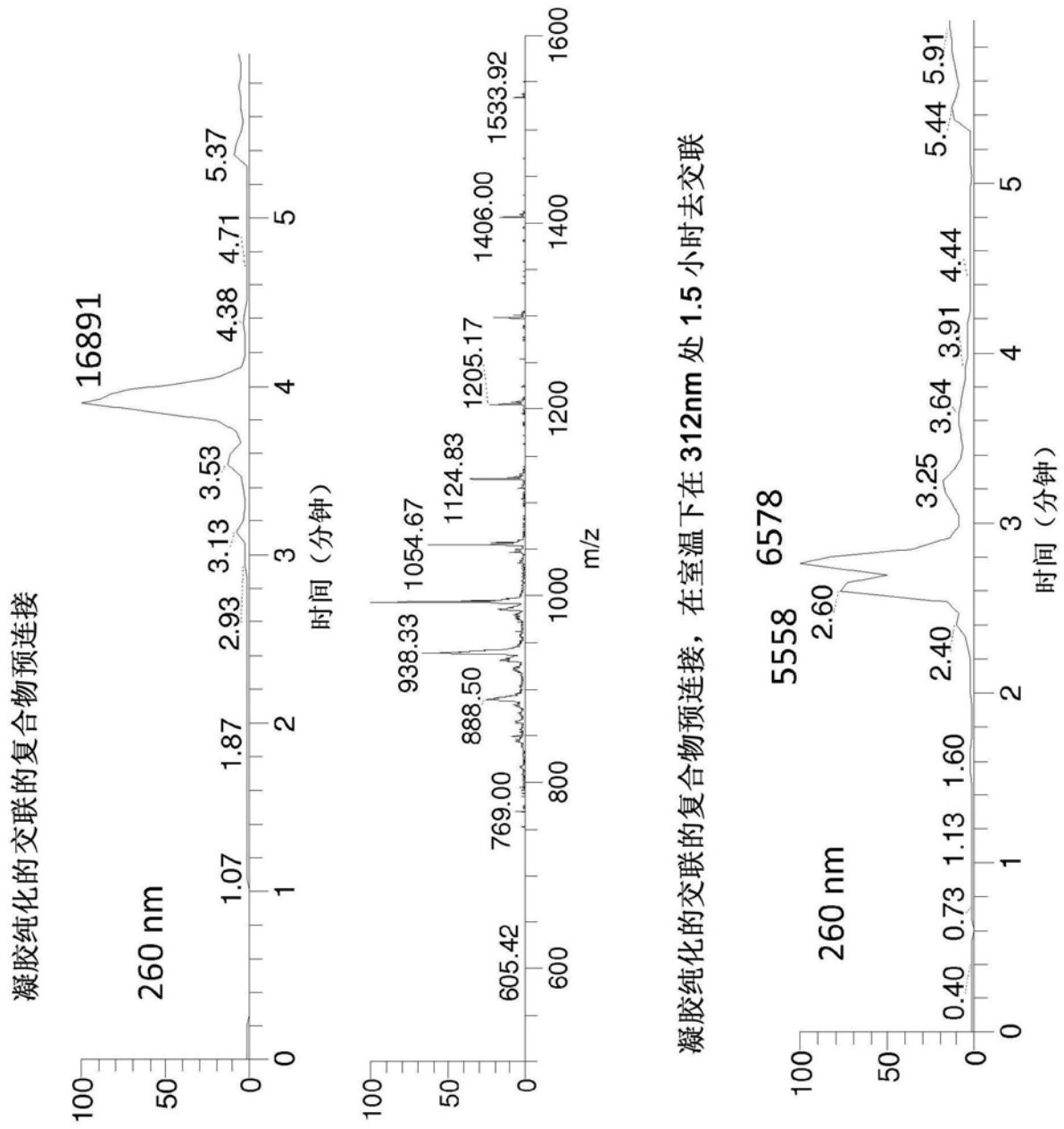


图7

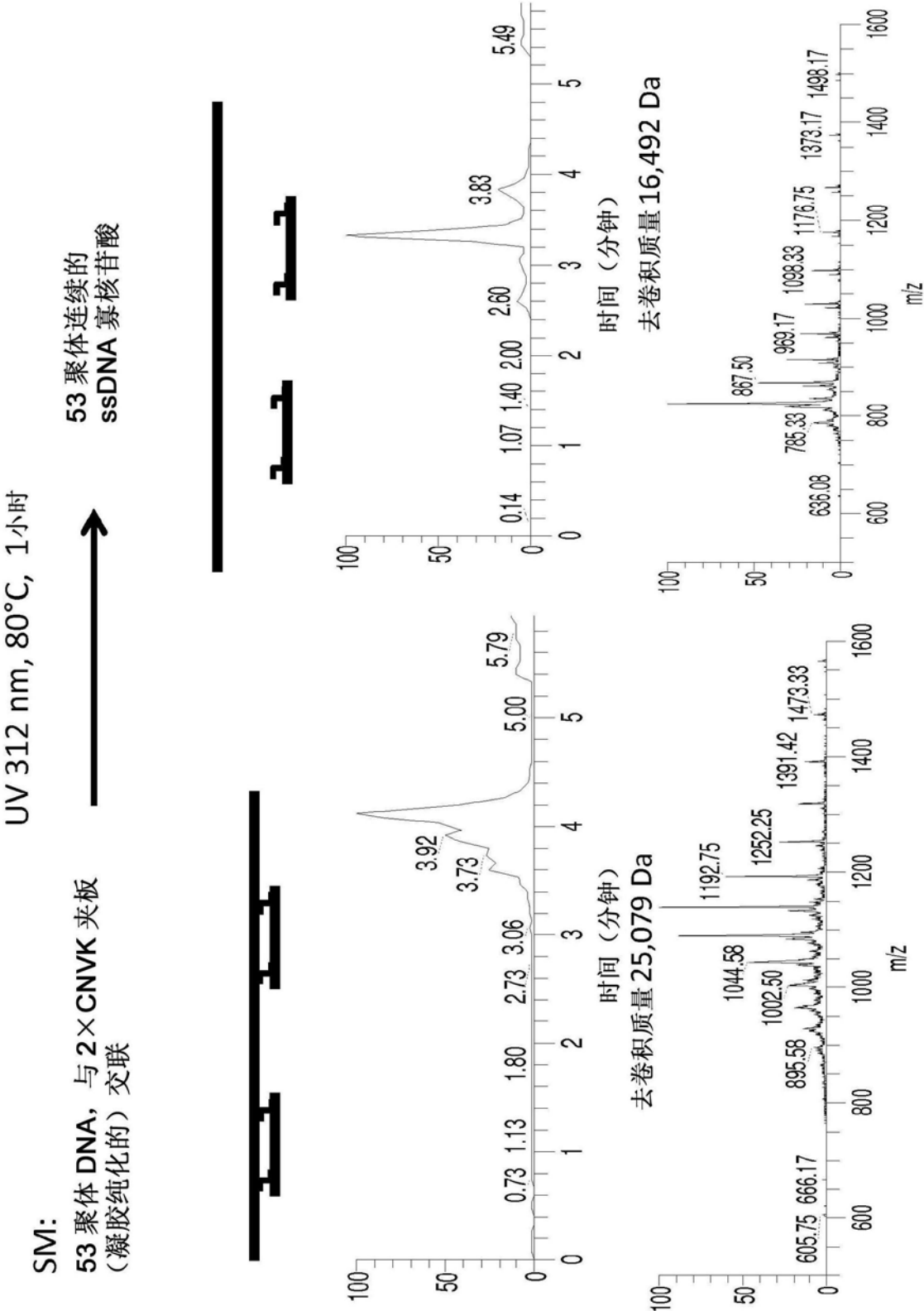


图8

CNVK 交联的复合物的 T4 DNA 连接，
接着是在 80°C 在 312 nm 处去交联持续 1 小时

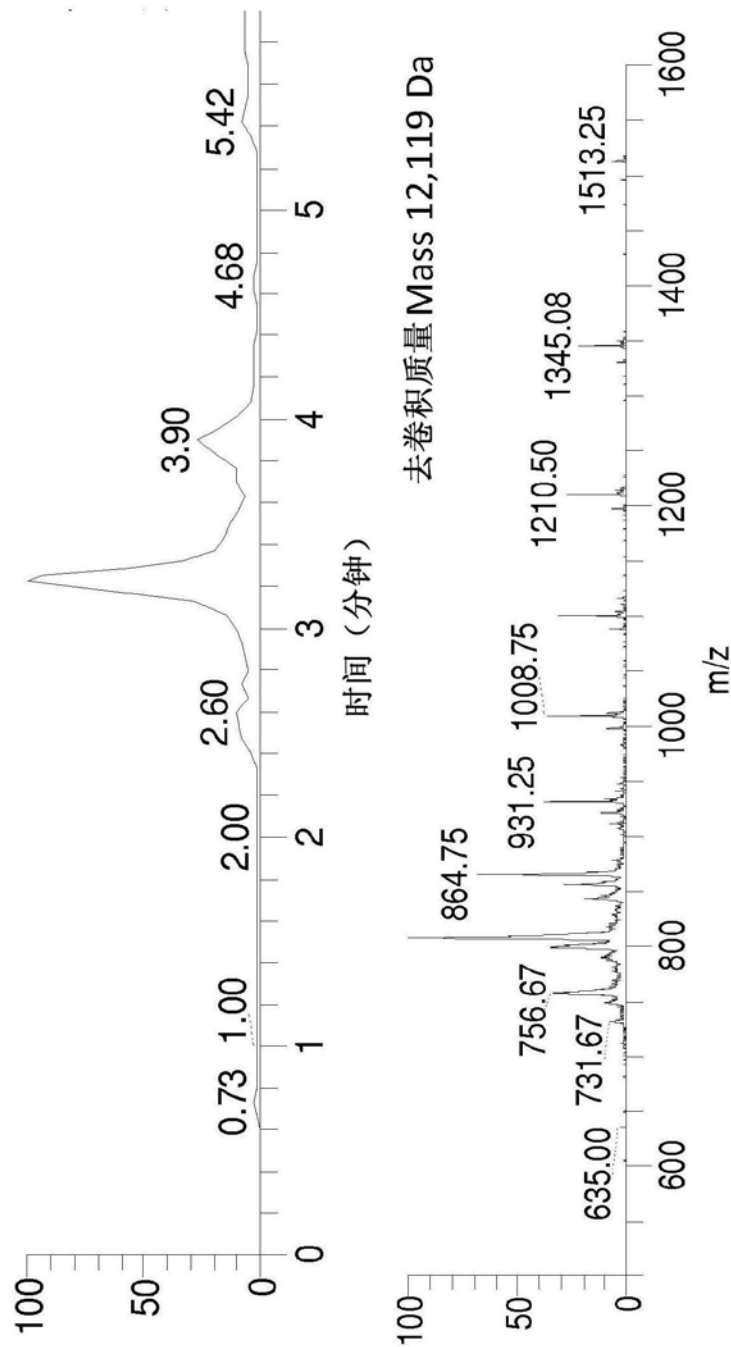


图9

SM:

标签 A+ 标签 B+ 夹板复合物

产物:

连接到标签 B 的标签 A

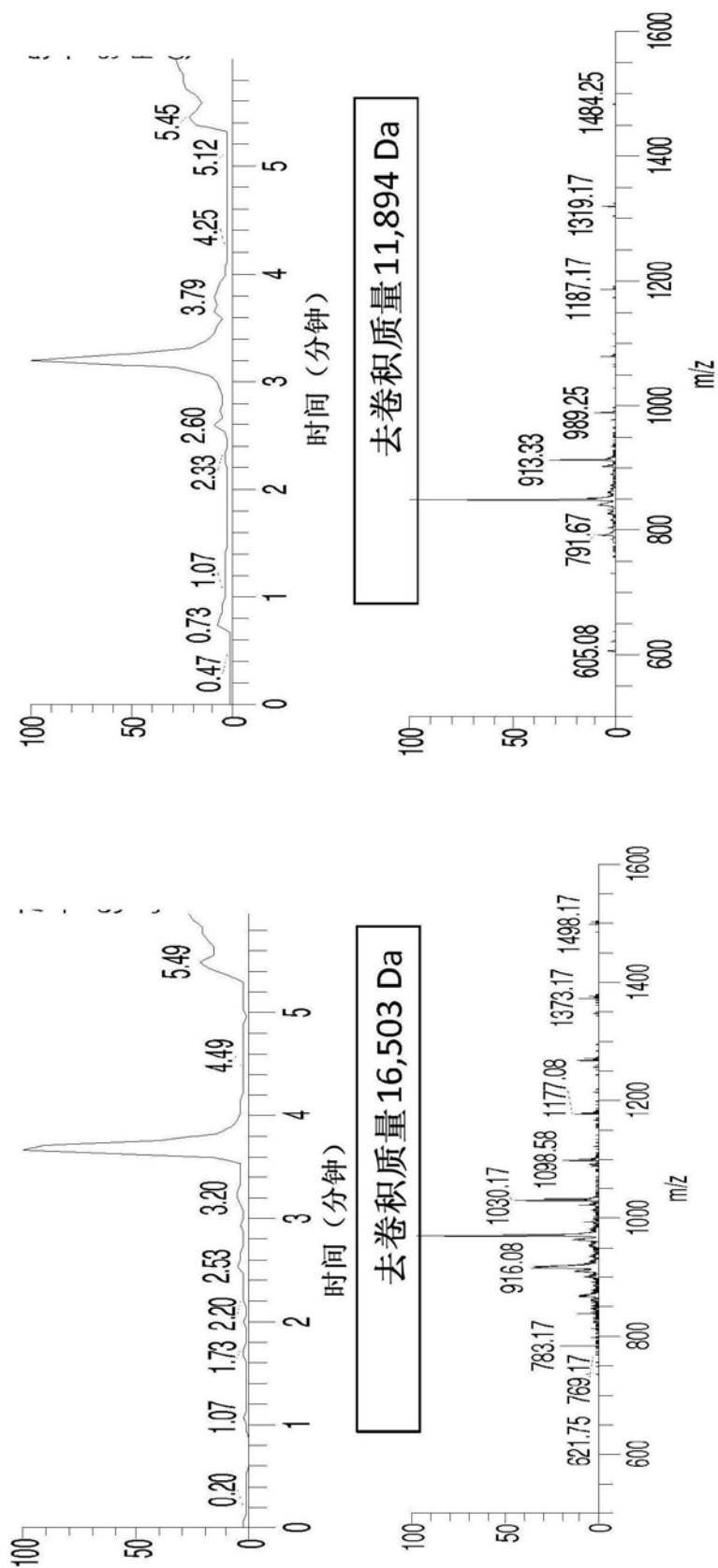


图10

5PSO2_A9_TA (MW 11,556.3 Da)
5'-补骨脂素 (C2- 连接)-TAGCGGATGCACAAAAAAGGCGAGCT TGCGTACTG-3'

PSO_HP_A9_TCT (MW 6,087.1 Da)
5'-AAAAAA AAAGCATCCGCTCT -3'

图11

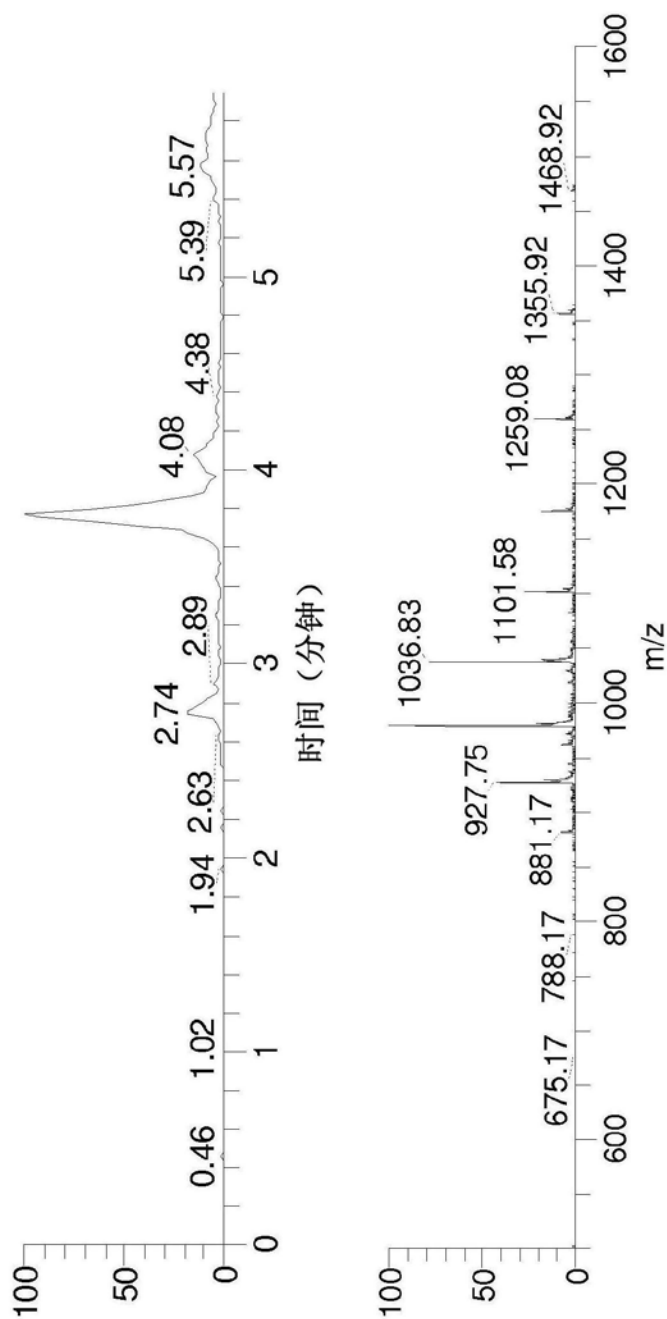
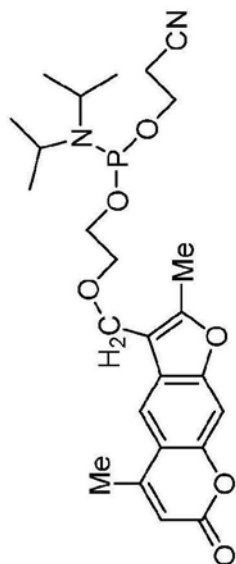


图12



补骨脂素 (C2) 亚磷酸胺

Tag1_PsoCVU (MW 3,662.5 Da)

5'-AAAACTCAGGT-3'

SplintC_PsoC2 (MW 4,529.0 Da)

5'-TCCGCTCGACCTGAG-3'

5PsoC2_A9_GA (MW 11,580.6 Da)

5'-补骨脂素(C2)-GAGCGGATGCACAAAAAGGCGAGCTTGCGTACTG-3'

图13

5'-NNNNNNNNNT PNNNNNNNNNNNNNNNNNNNNNN-3'
3'-NNNAGNNNNNN-5'

P = 5'-补骨脂素

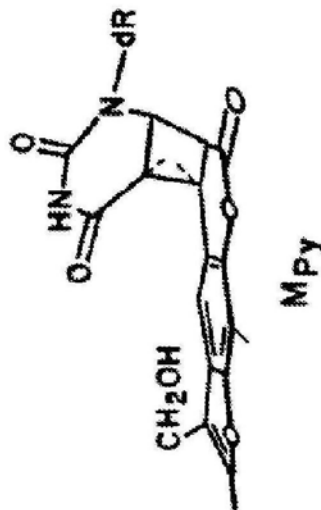


图14

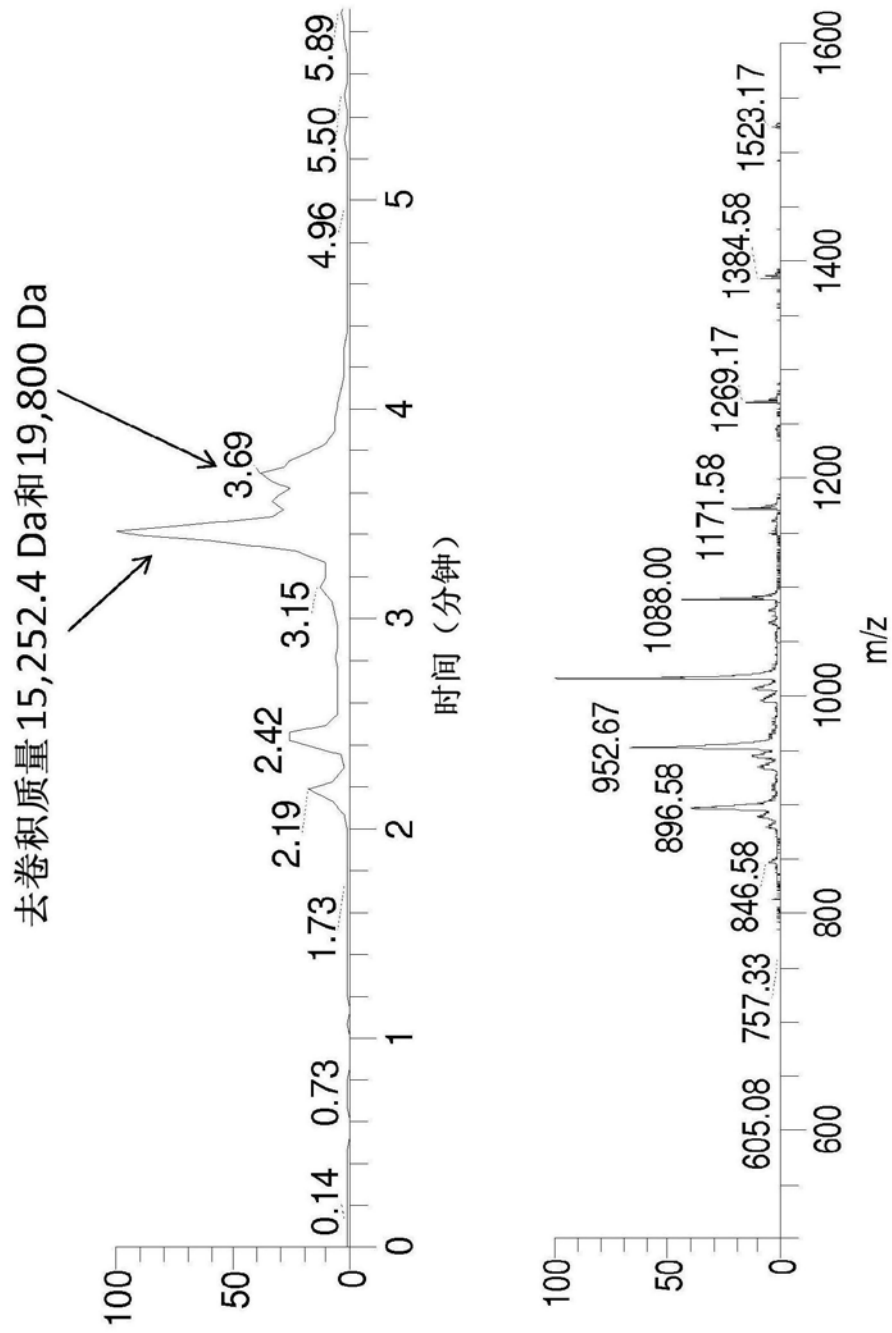


图16

5Bio_Tag_PsoCVU (MW 4,232.1 Da) /5BioTEG/是专有
5'-/5BioTEG/AAAACTCAGGT-3' 生物素修饰

5PsoC2_A9_GA (MW 11,580.6 Da)
5'-补骨脂素 (C2)-GAGCGGATGCAAAAAAGCGAGCTTGCGTACTG-3'

SplintC_PsoC2 (MW 4,529.0 Da) FAM引物 (MW 5,099.5 Da)
5'-TCCGCTCGACCTGAG-3' 5'-/56-FAM/AGTACGCAAGCTCGC-3'

Phos-SplintC_PsoC2
5'-单磷酸酯-TCCGCTCGACCTGAG-3'

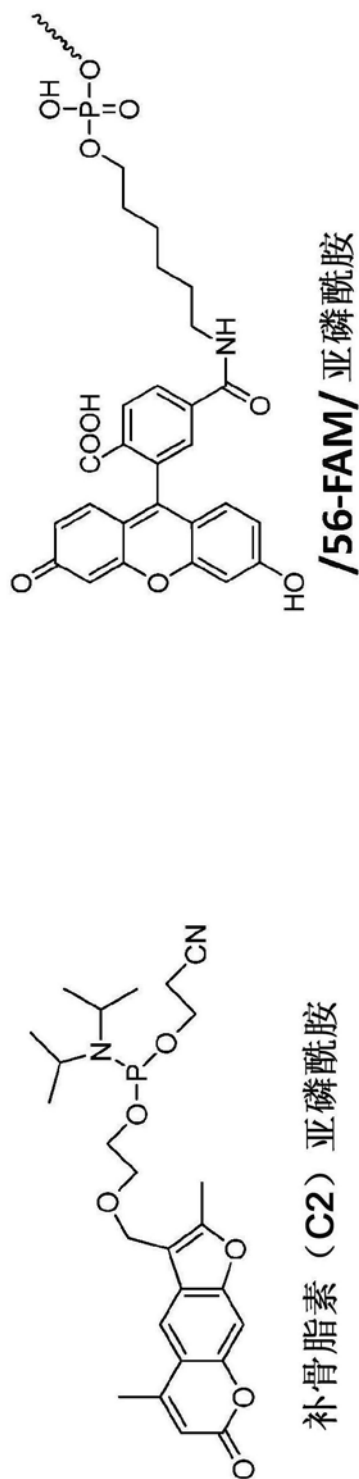


图17

凝胶纯化的光化学结合物 Bio_Pso 的 LCMS

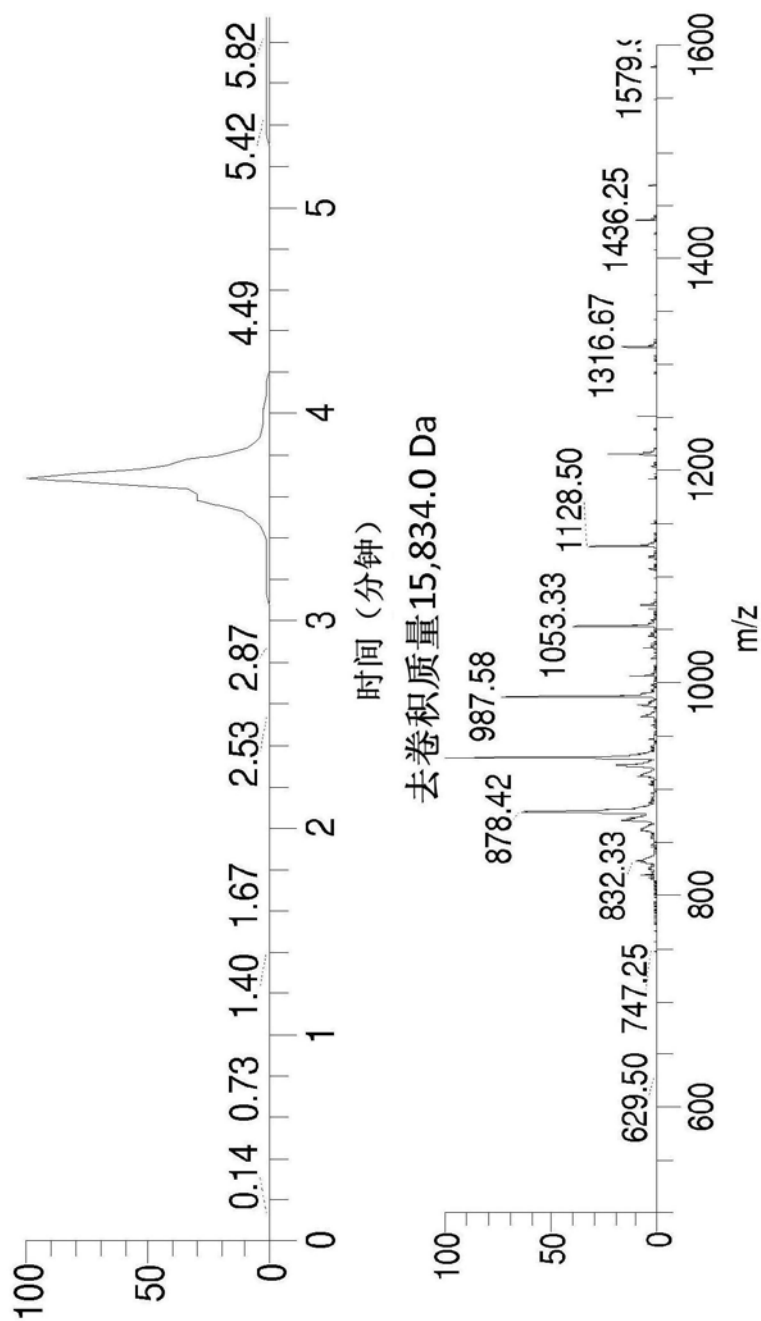


图18

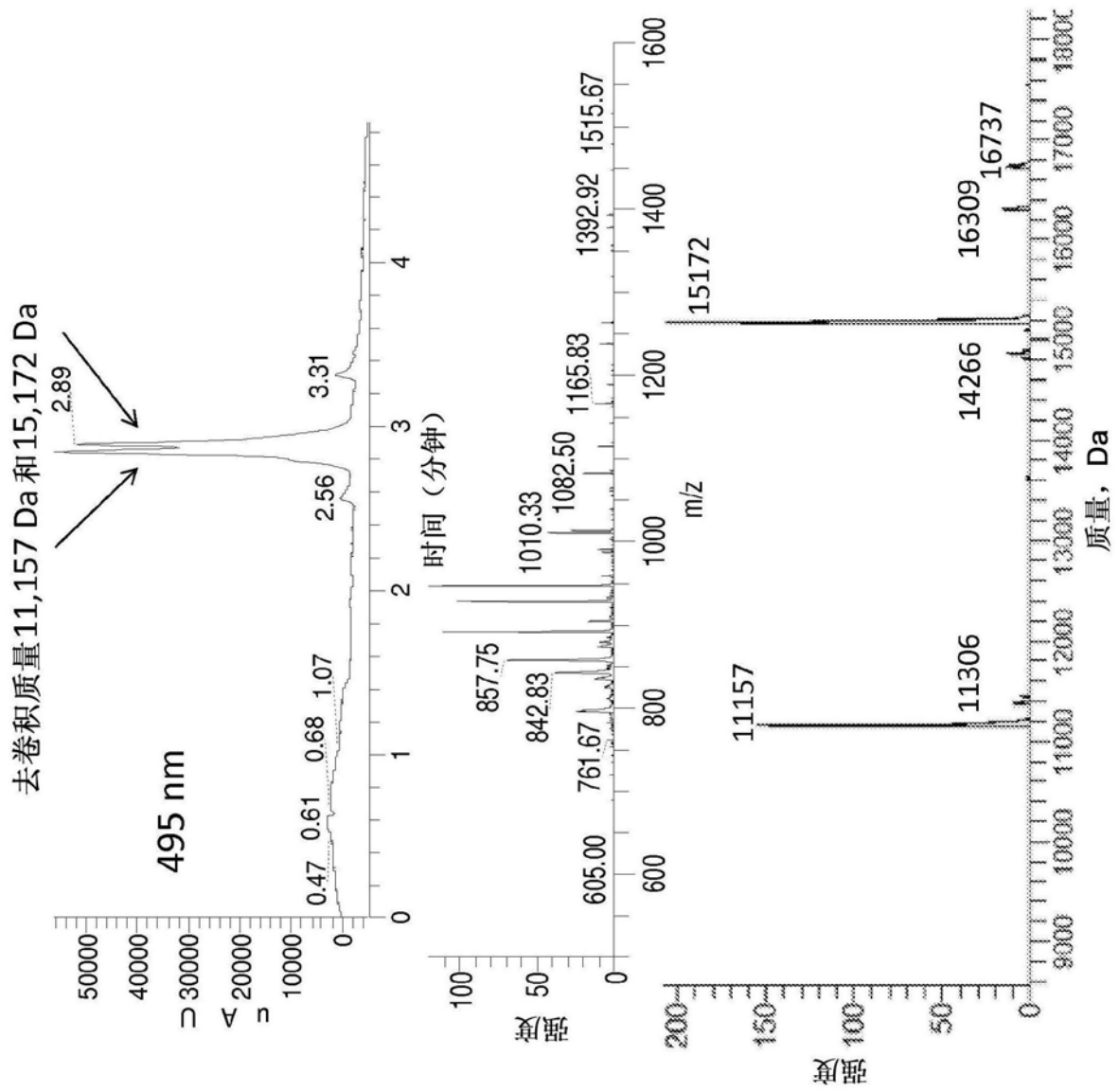


图19

使用末端和非末端引物延伸，
从 Bio_Pso 产生的 cDNA 的无连接对照的 LCMS

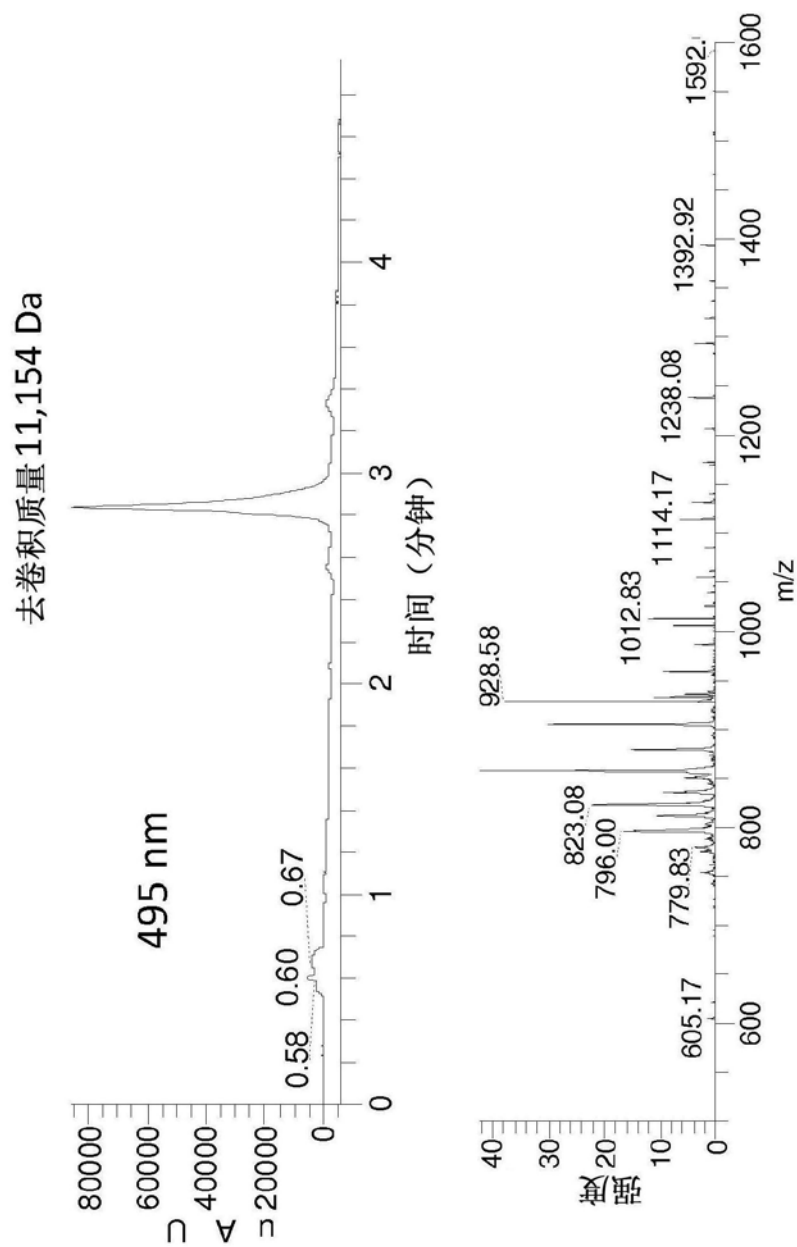
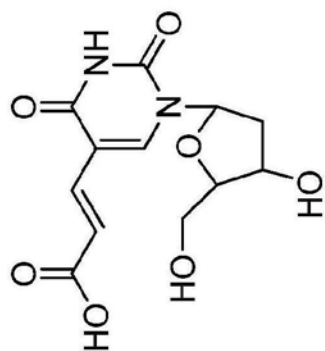


图20



5-(羧基)乙烯基-2'-脱氧尿苷

SplintC_CVU

5'-TCCGCTCAACCTGAG-3'

SplintA_CVU

5'-TCCGCTTAACTGAG-3'

Tag1_PsoCVU (MW 3,662.5 Da)

5'-AAAACTCAGGT-3'

CVU_G (MW 11,576.2 Da)

5'-(5-(羧基)乙烯基-2'-脱氧尿苷)GAGCGGATGCACAAAAAAGCGAGCTTGCGTACTG-3'

CVU_A (MW 11560.2 Da)

5'-(5-(羧基)乙烯基-2'-脱氧尿苷)AAGCGGATGCACAAAAAAGCGAGCTTGCGTACTG-3'

图21

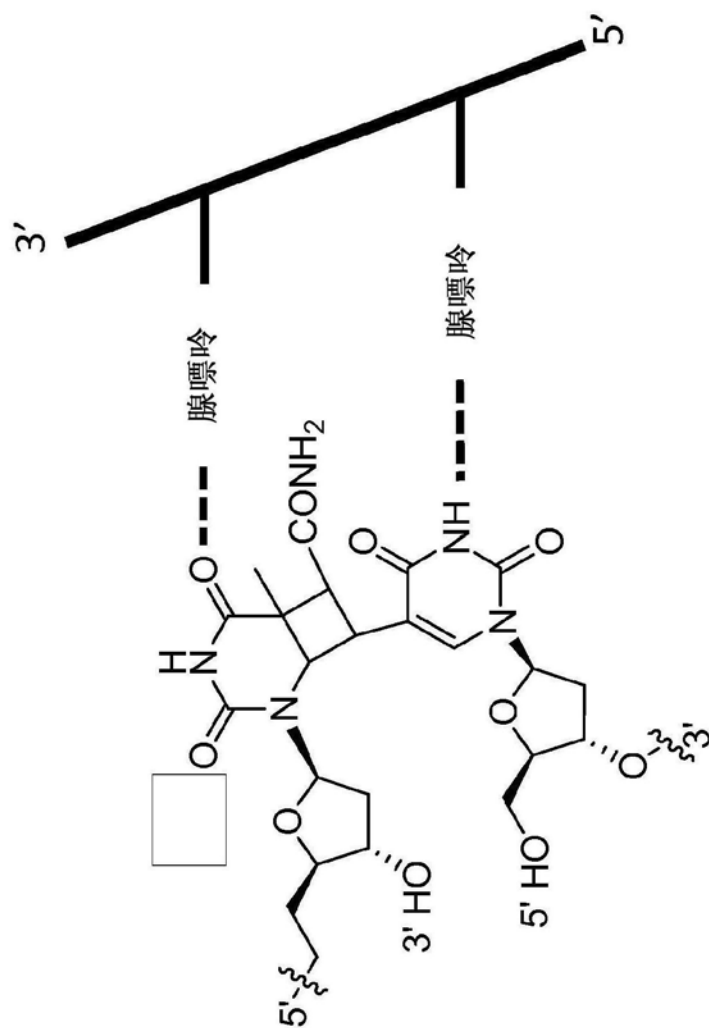


图22

在 365 nm 处 CVU 光连接的时间 - 进程

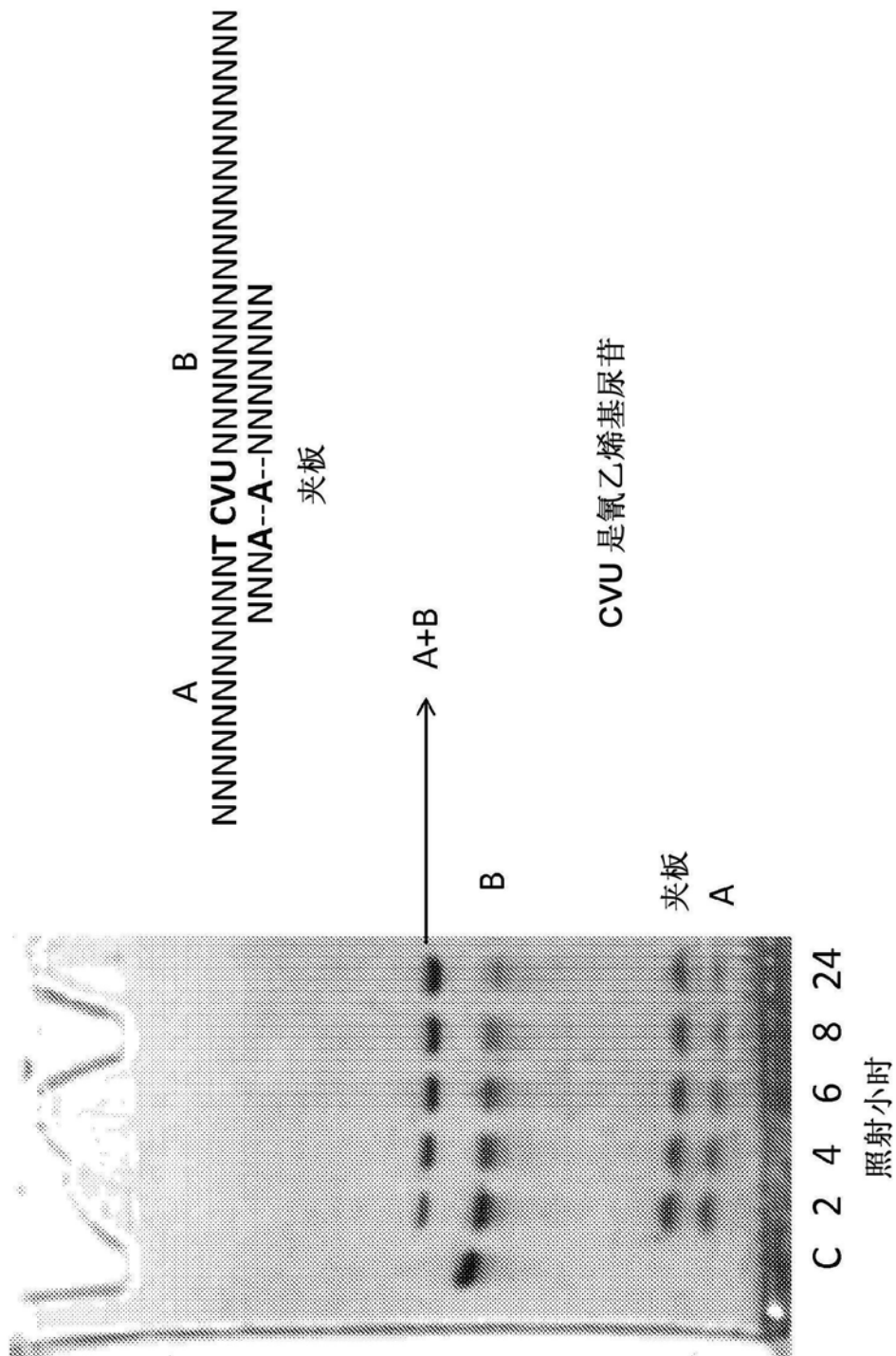


图23

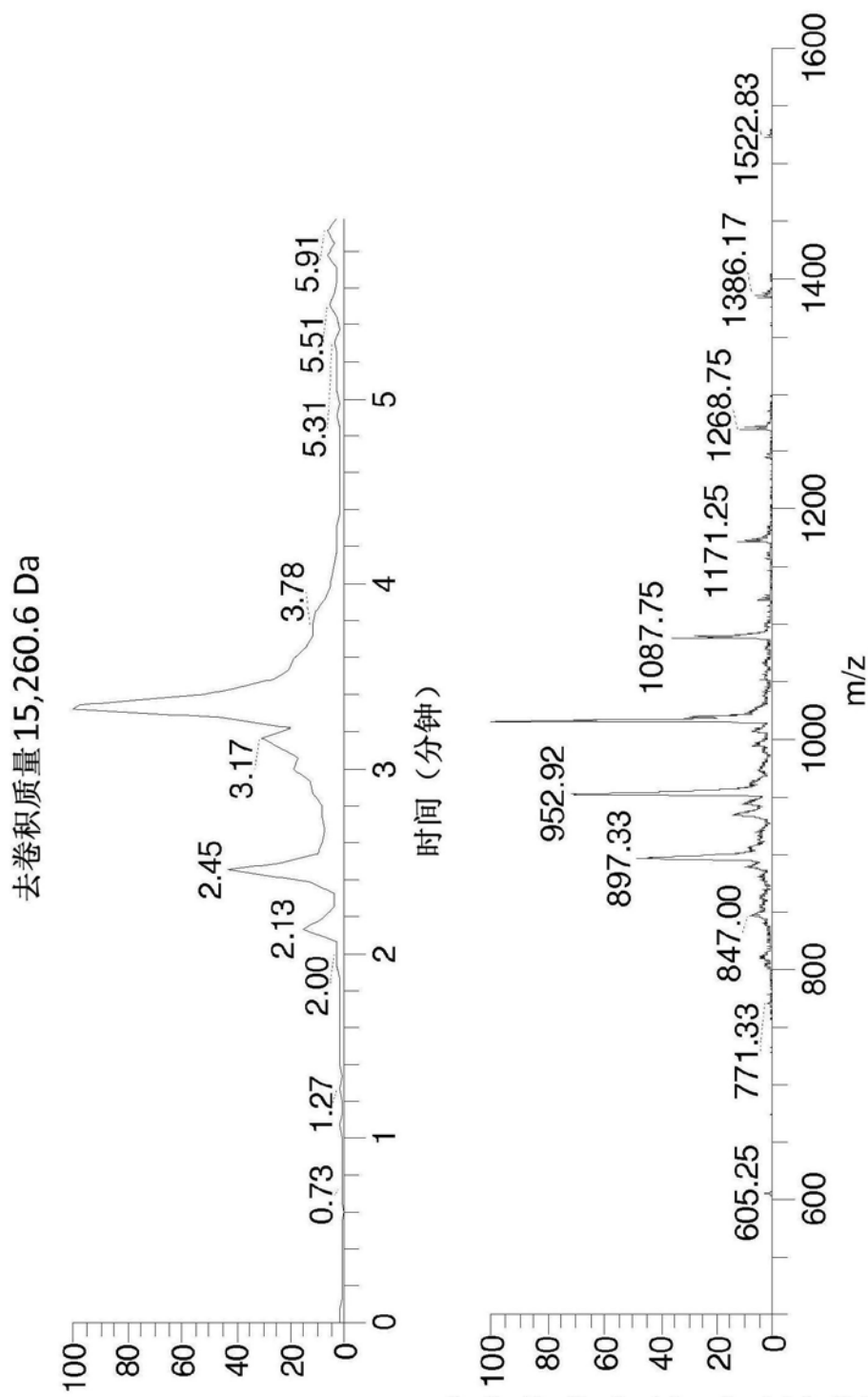


图24

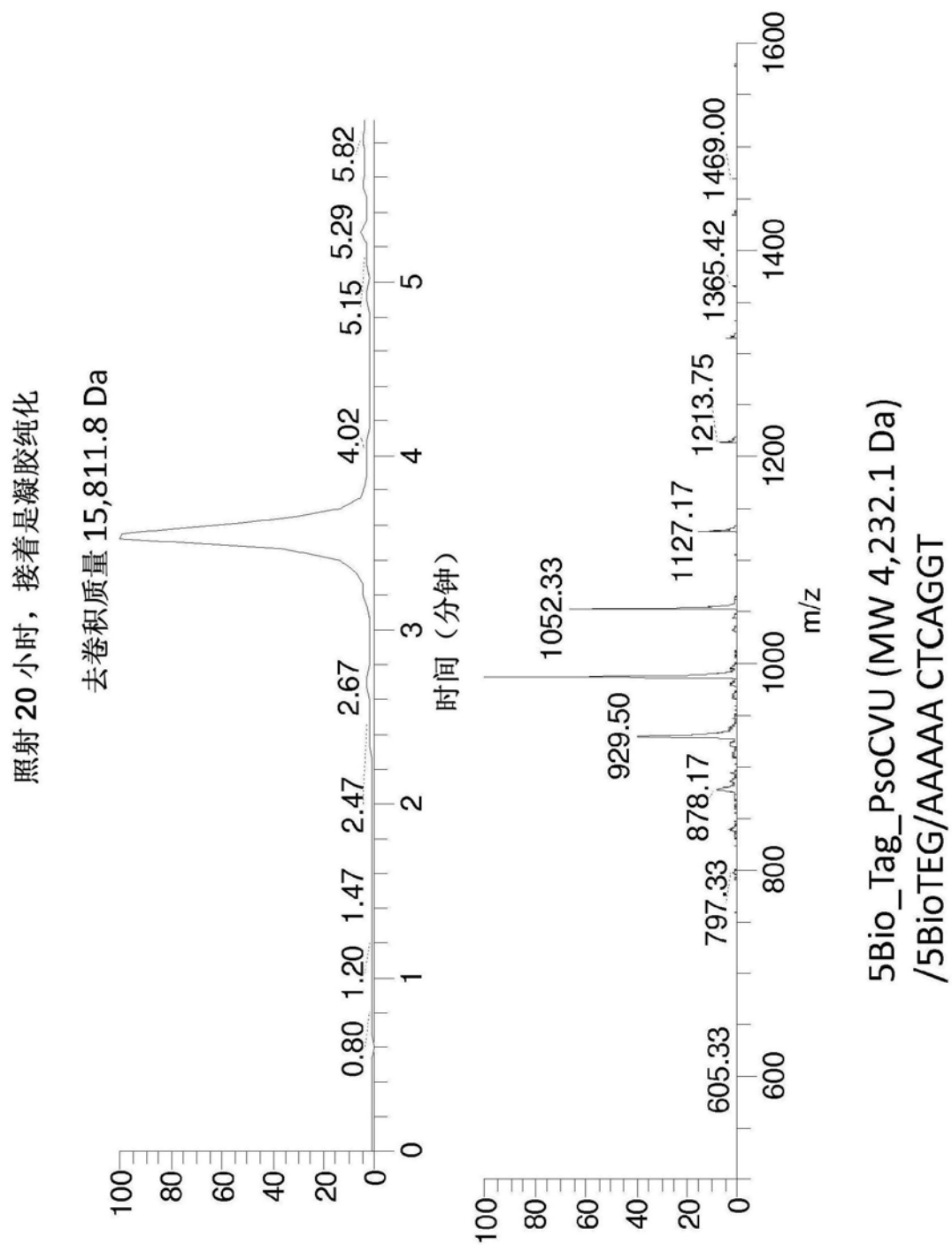


图25

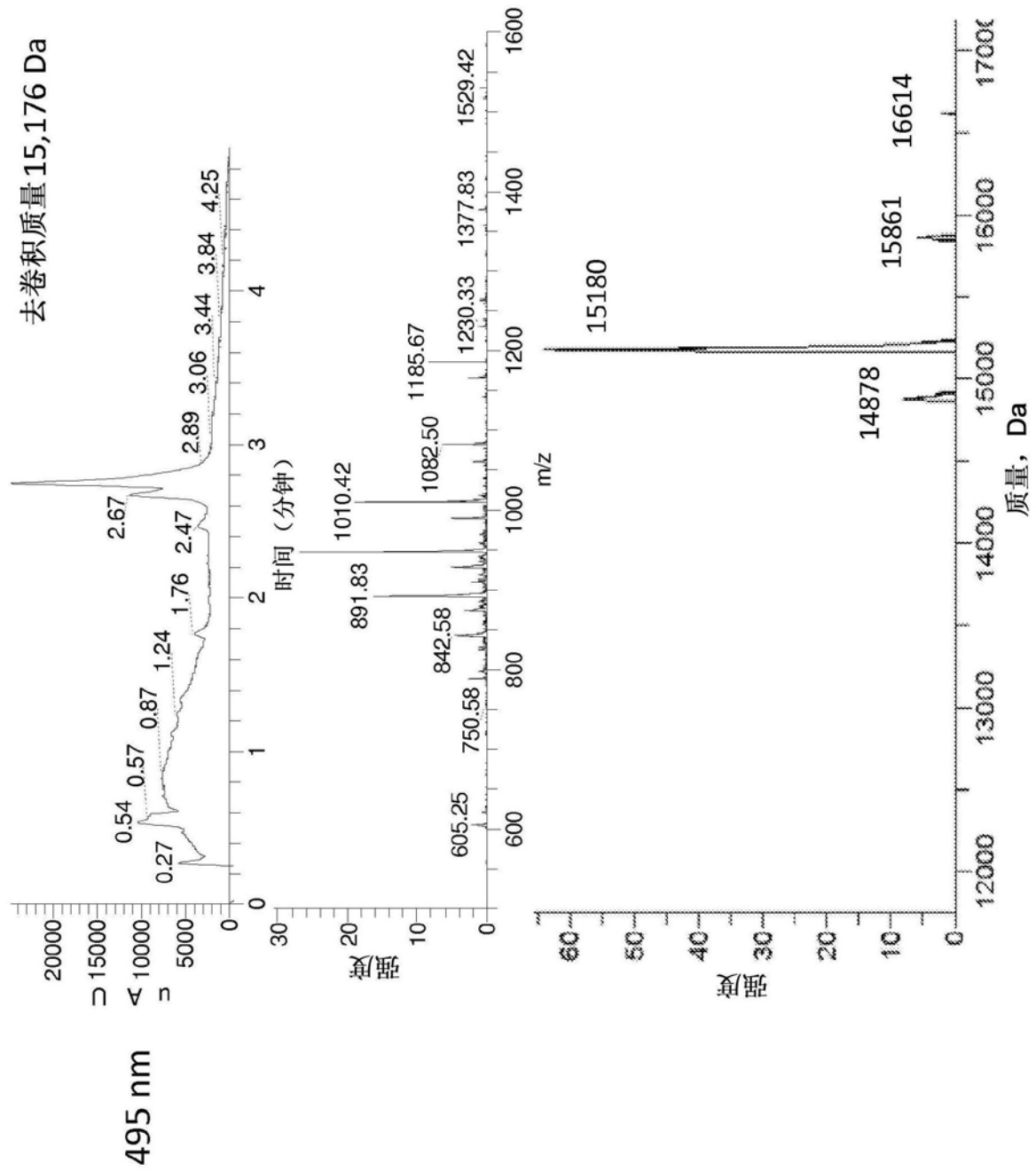


图26

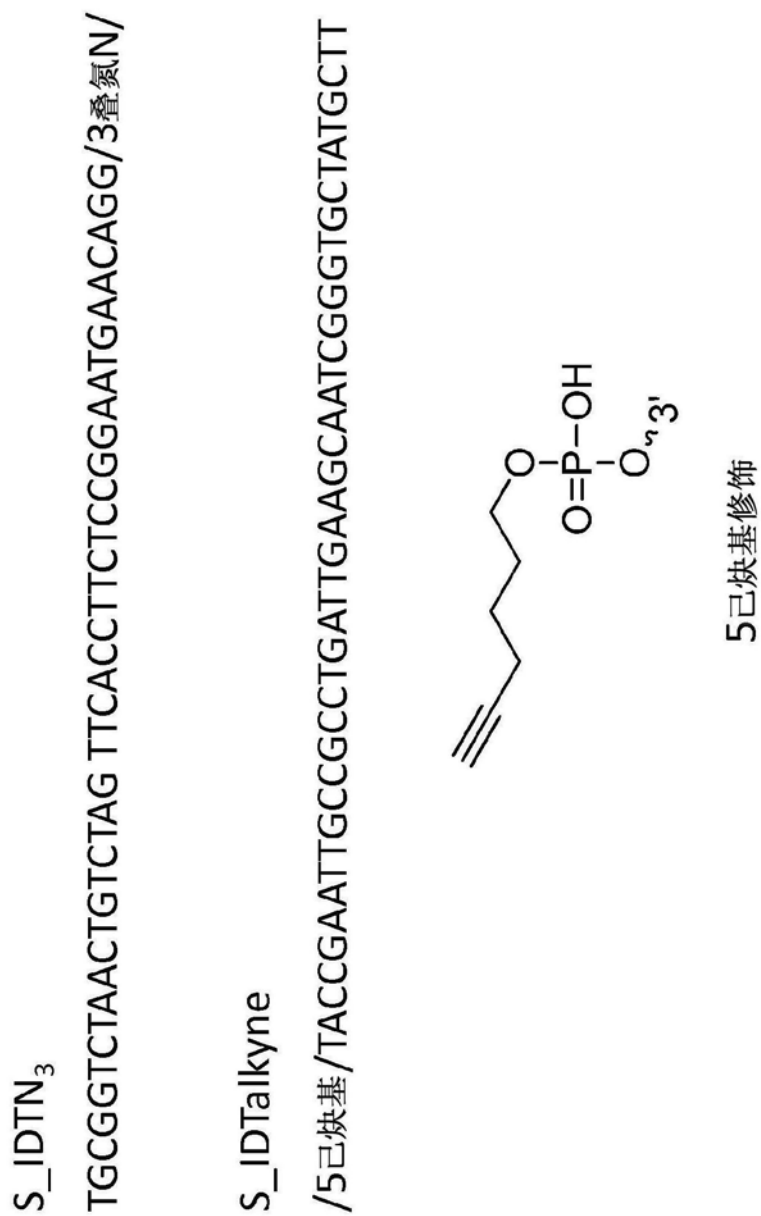


图27

S_IDTN₃ 和 S_IDTalkyne 的凝胶纯化结合物的 LC/MS 分析
观测到的去卷积质量 26,304.8 Da (计算出的 26,311.2 Da)

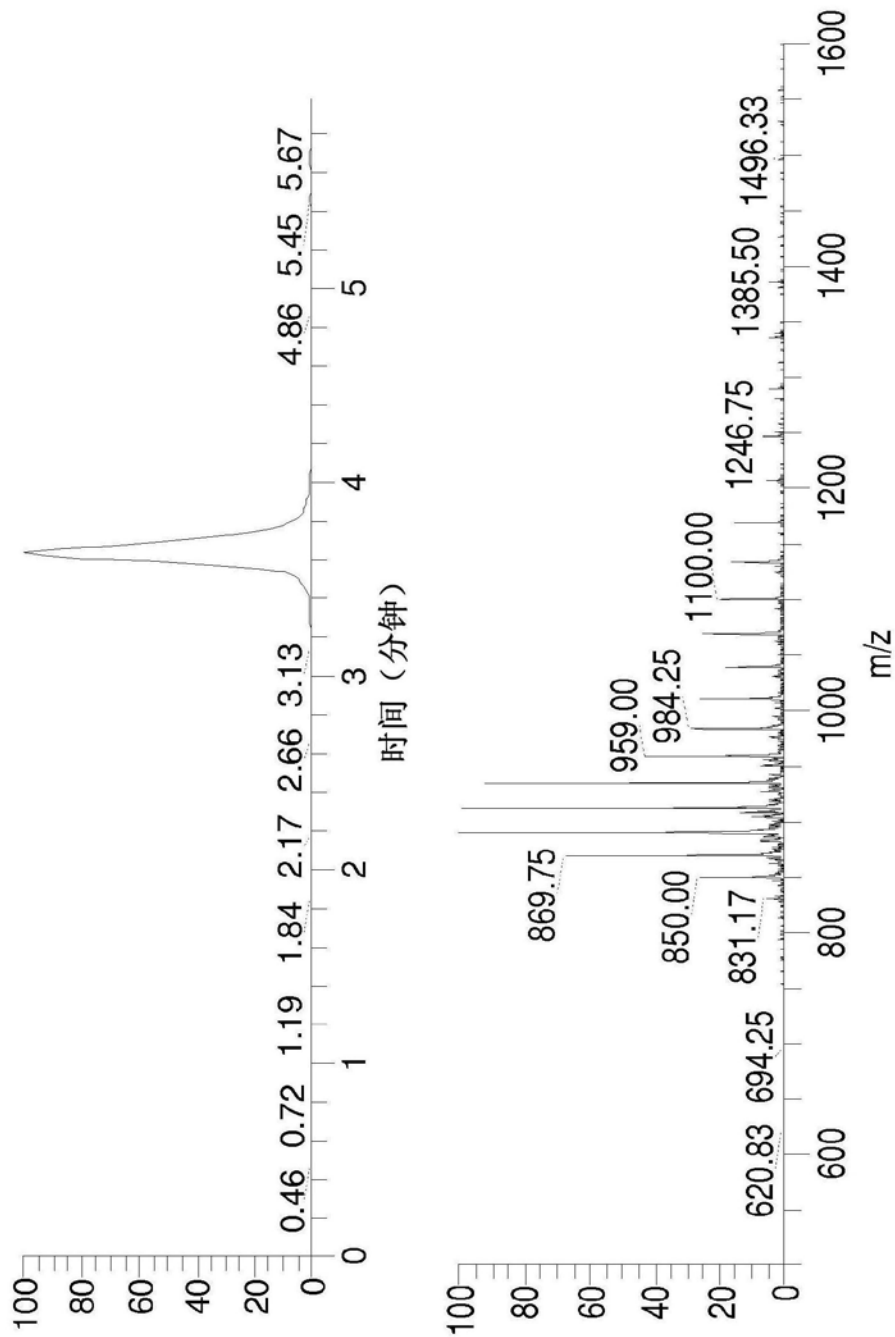
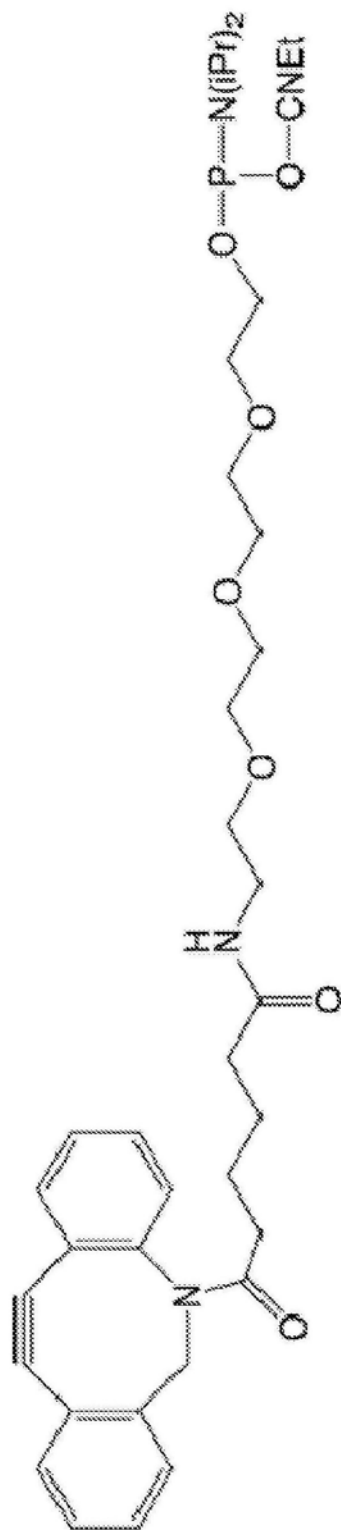


图28

TKR_Central: /5己炔基/GGAATGAACAGG **GTAAGCTGGAGTGAAGCGGTTATG/3叠氮N/**

S_IDTN₃: TCGGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG /3叠氮N/

TKR_DBCO_S: /DBCO/GAAGGCGTTAIGTCCGTACTCTTGCAATCGGGTGCTAIGCTT



DBCO 亚磷酰胺

图29

观测到的去卷积质量 25,336 Da (计算出的 25,341.8 Da)

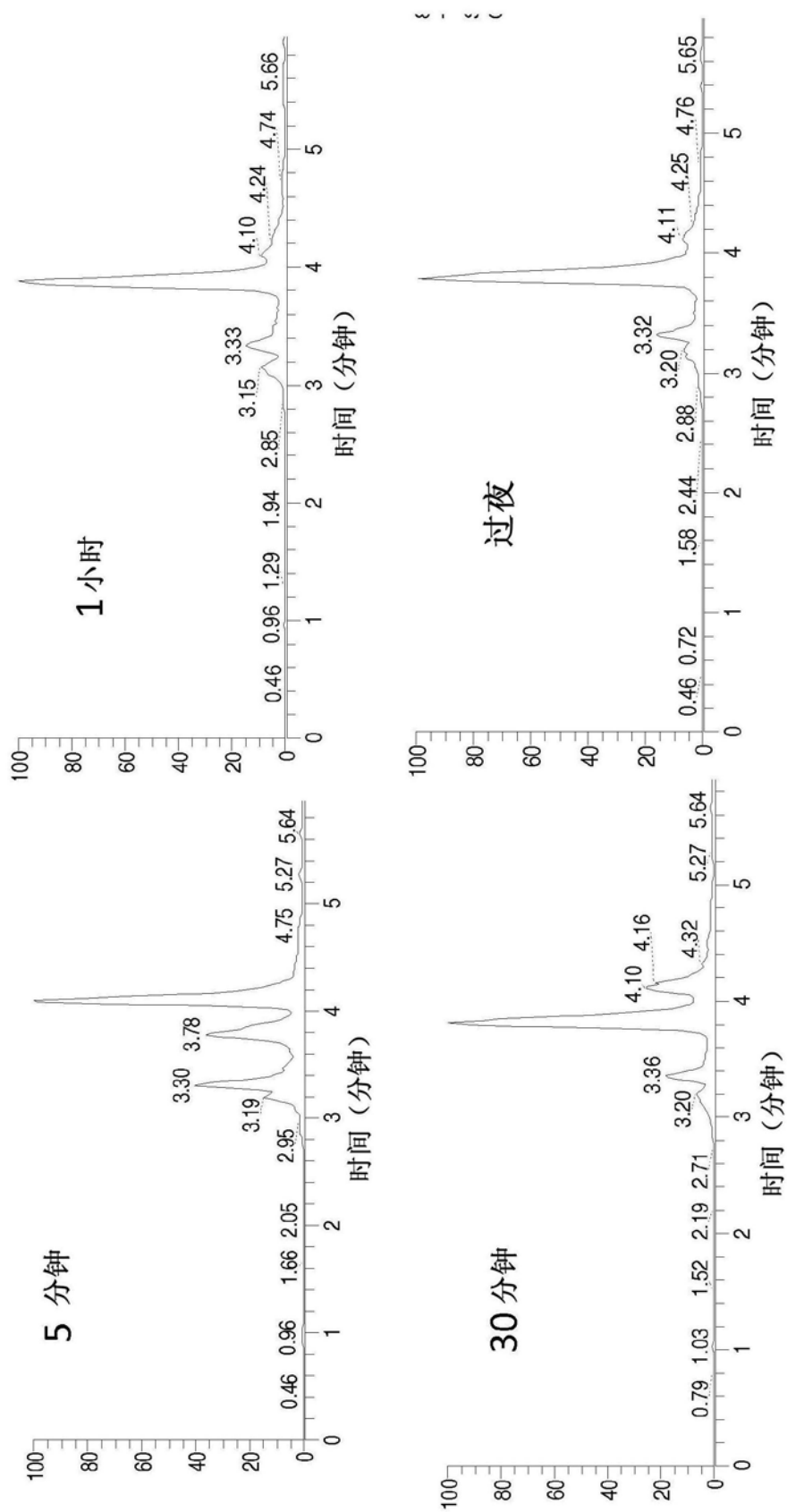


图30

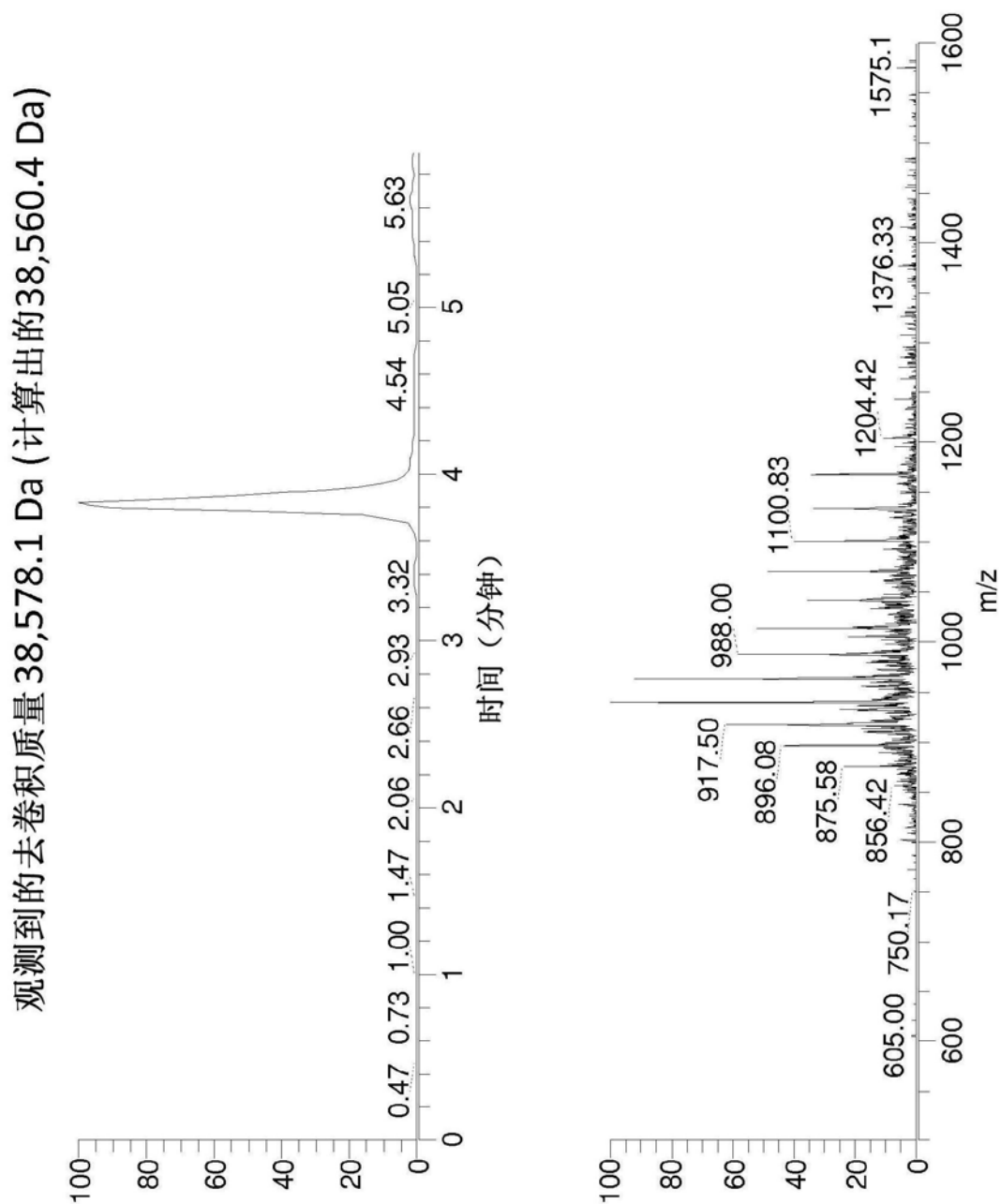


图31

Conjugate_Click_S
TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG(点击)TACCGAATTGCCGCCTGATTGAAG
-CAATCGGGTGCTATGCTT

Conjugate_Click_L
TGCGGTCTAACTGTCTAGGCACTTGTTCGTTTGCCAGTGTGAGGAATGAACAGG(点击)TACCGAATTGC
CTTCCTCG-TACAGTTCTAAGGCGCTTGGACACCACTTCAATCGGGTGCTATGCTT

图32

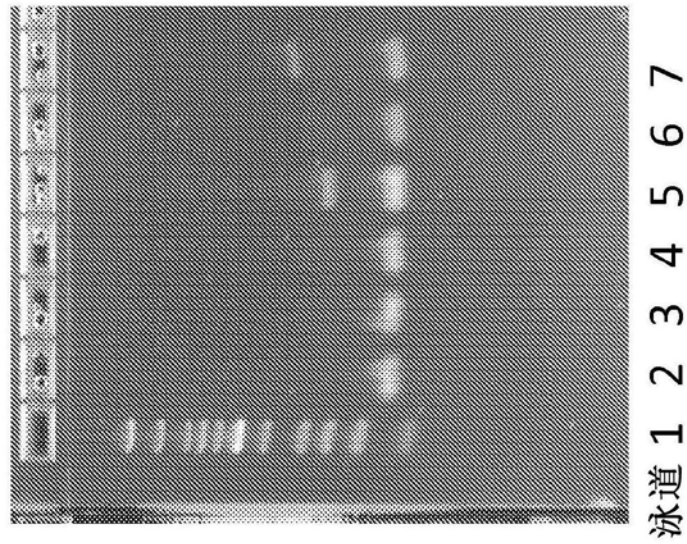


图33

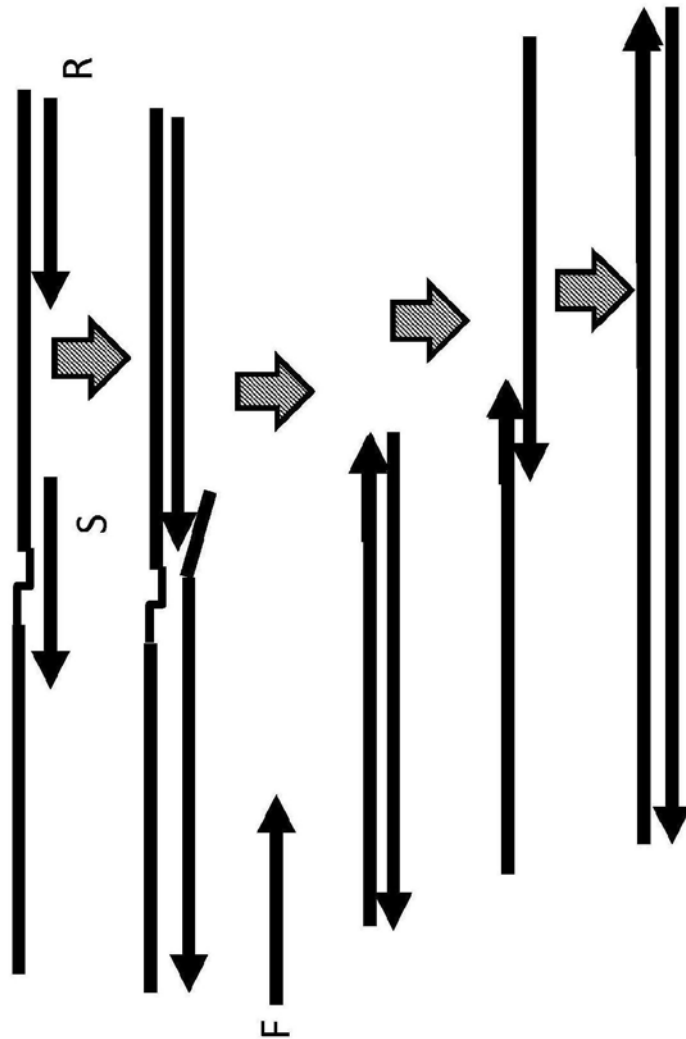


图34

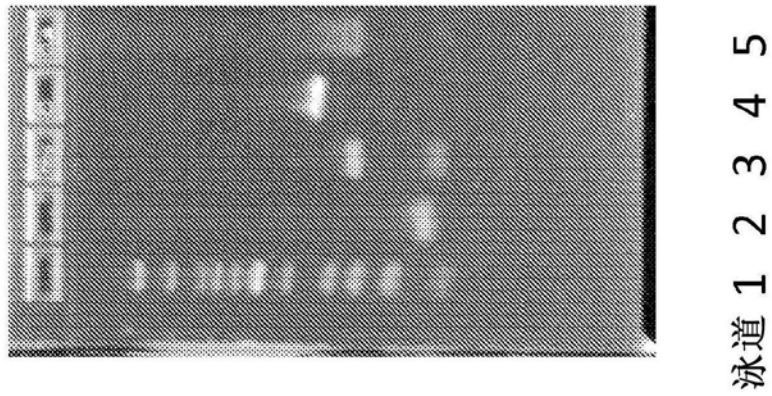


图35

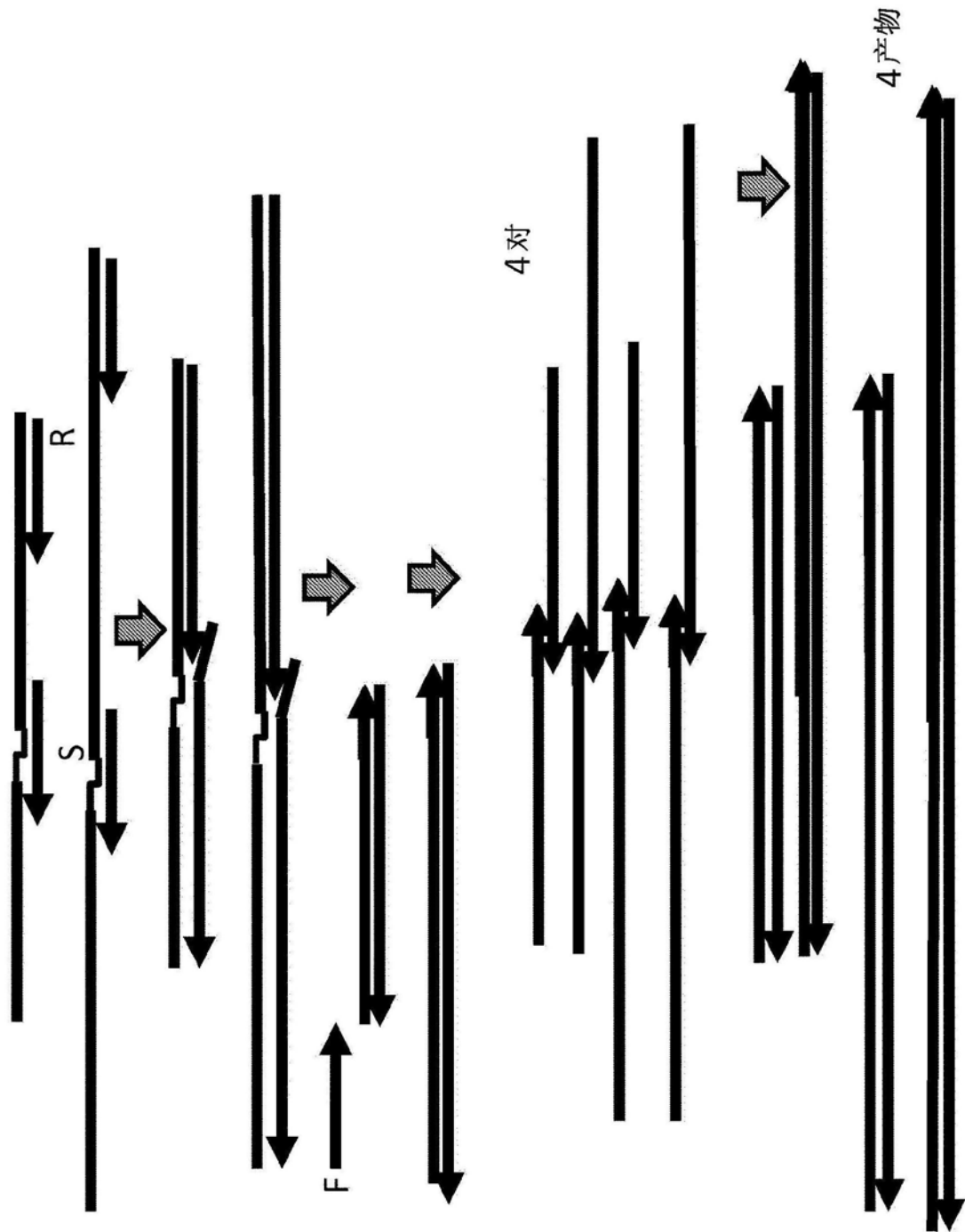


图36

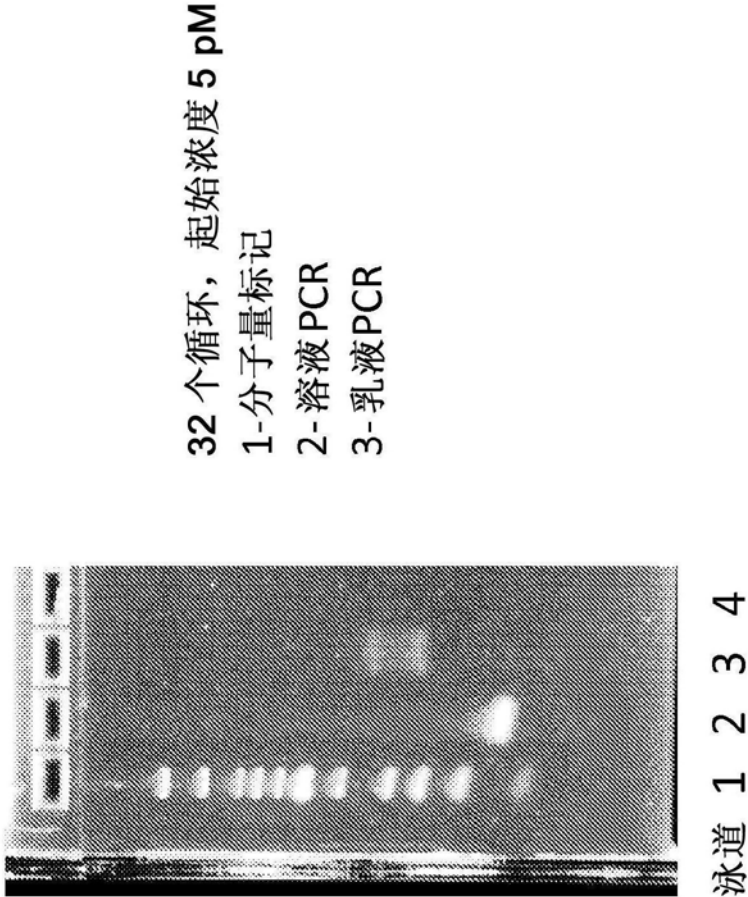


图37

TKR_S:
TGCGGTCTAGTCTACCTTCTCCGGAATGAACAGG(点击)GGAATGAACAGGGCCTGATTGAAGCAATCGGGTGCTATGCTT

TKR_L:
TGCGGTCTAACTGTCTAGGCACCTGTTGGTTGCCAGTGTGATGGACACCACTTGAATGAACAGG(点击)GGAATGAACAGGGGAATGAACAG
GTTCTCGTACAGTTCTAAGGCGCTCAATCGGGTGCTATGCTT

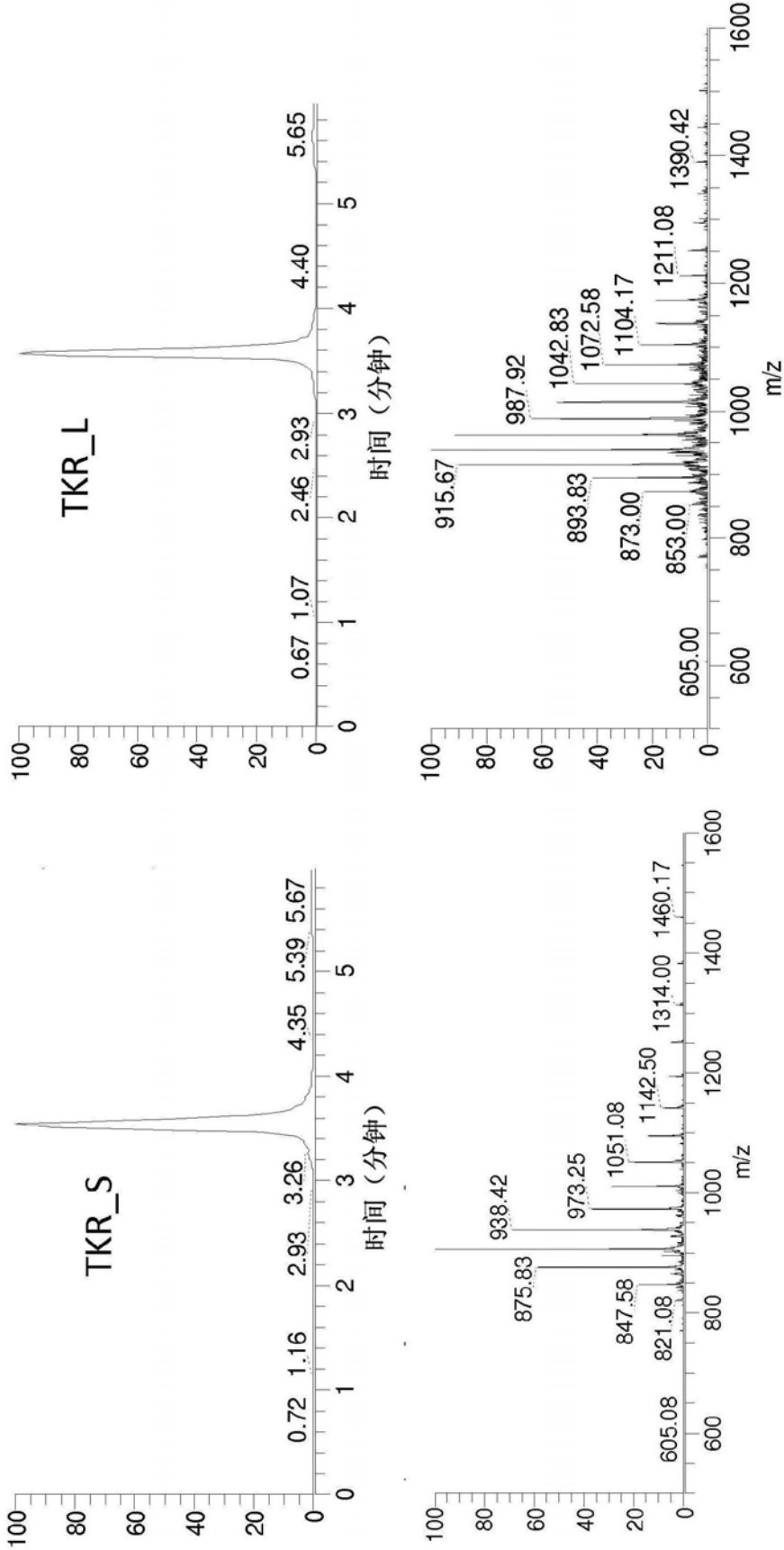


图38

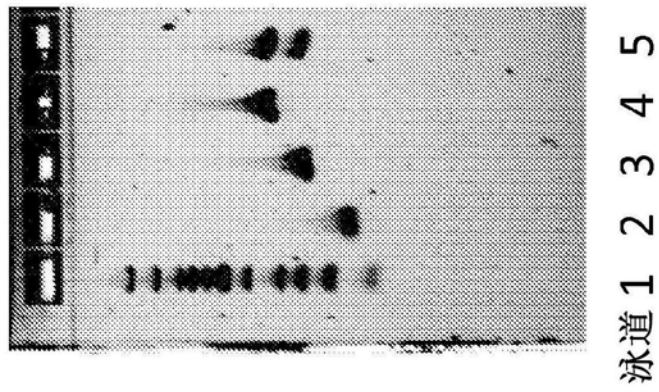


图39

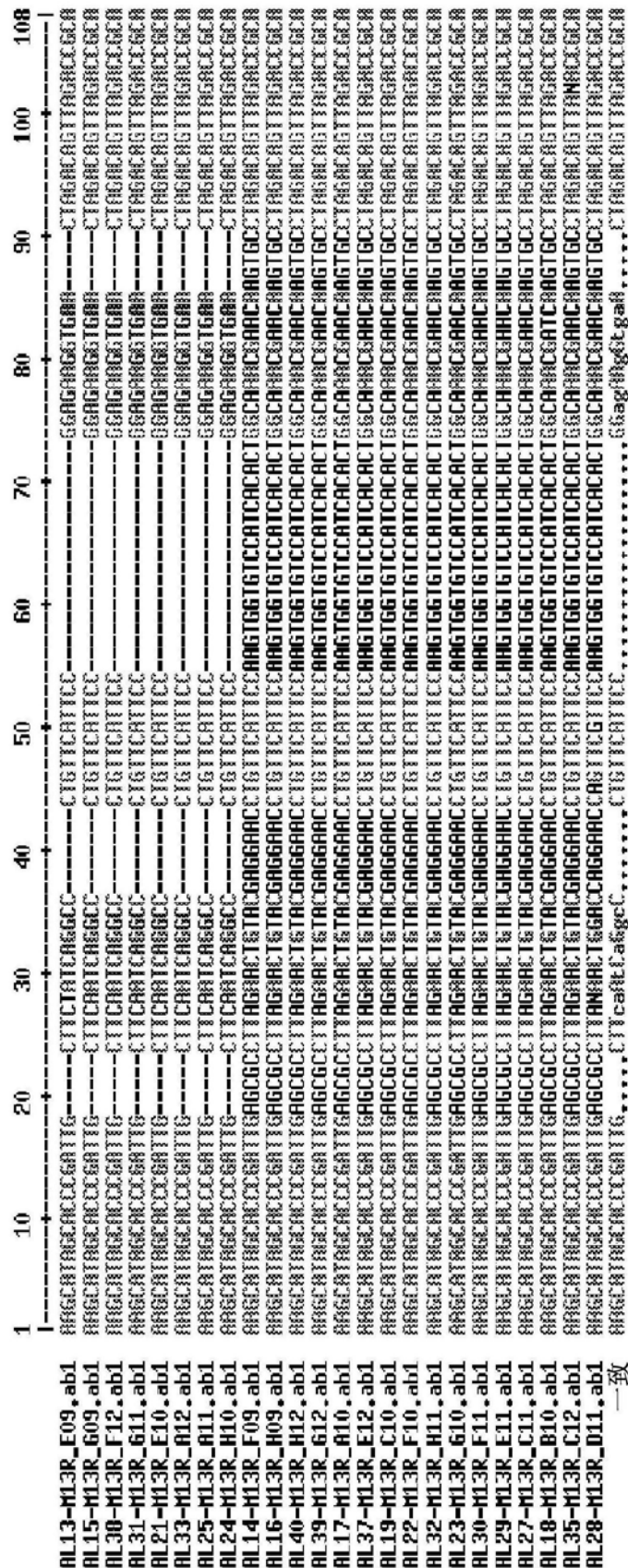


图40

TKR_2_click_S
 TCGGCTCTAACTGTCTAGTTACACTTCTCCGGAATGAACAGG(点击)GGAATGAACAGGGTAAGCTGGAGTGAAGCGGTTATG(DBCO)
 GAAGGCGTTATGTCCGTACTCTTGCAATCGGGTGCTATGCTT

TKR_2_click_L
 TCGGCTCTAACTGTCTAGGCACTTGTCGTTGCCAGTGTGATGGACACCACCTTGAATGAACAGG(点击)GGAATGAACAGGGTAAGC
 TGGAGTGAAGGCGTTATG(DBCO)GAAGGCGTTATGTGATATCCGTGGTGTCTGAGTTCCAATCGGGTGCTATGCTT

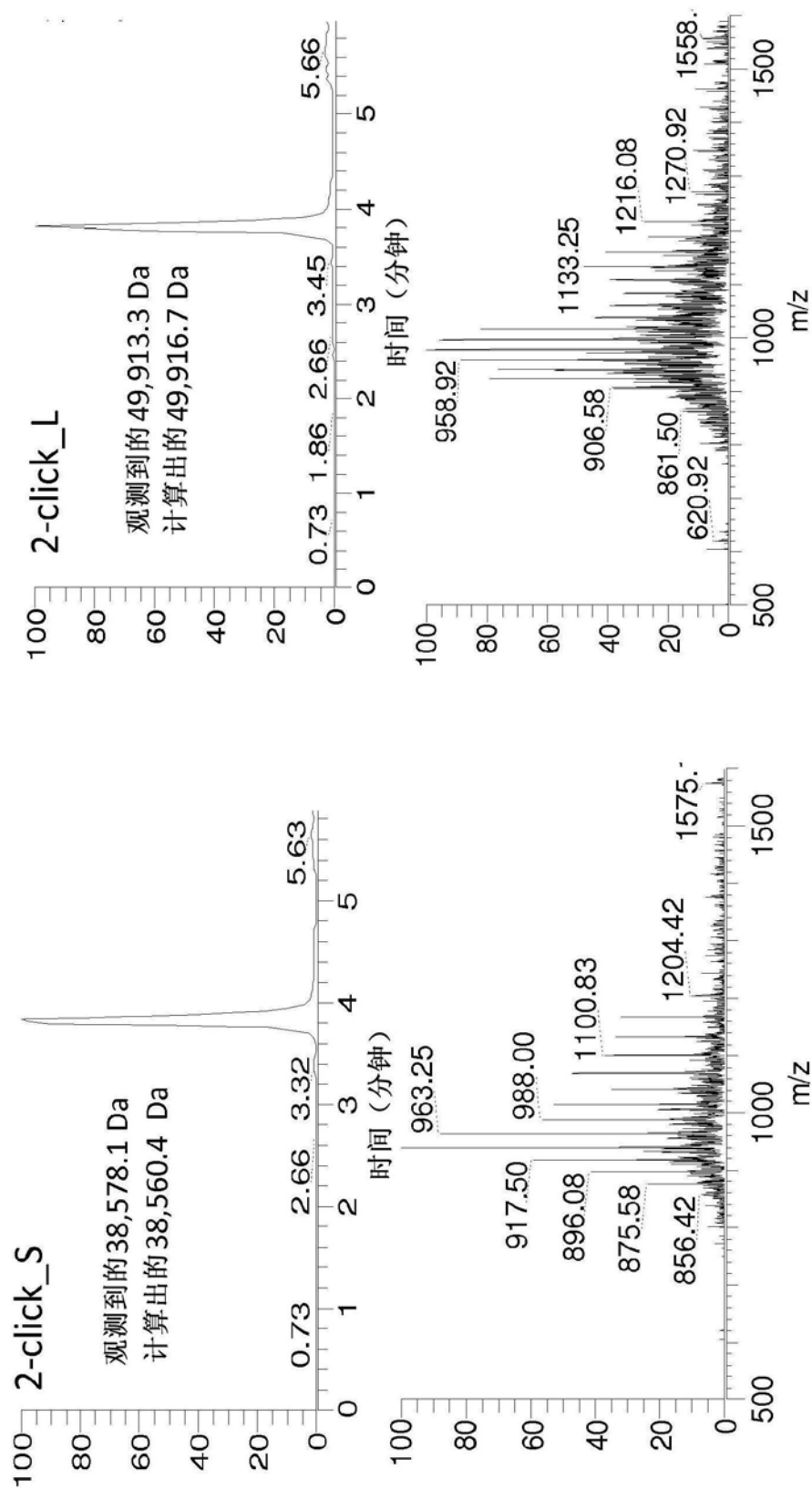


图41

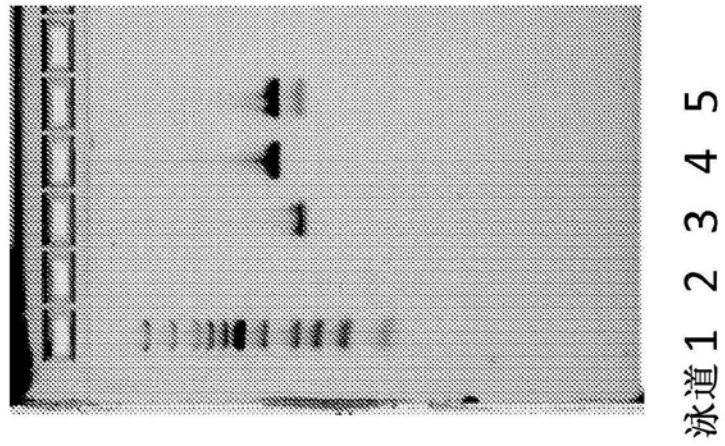


图42

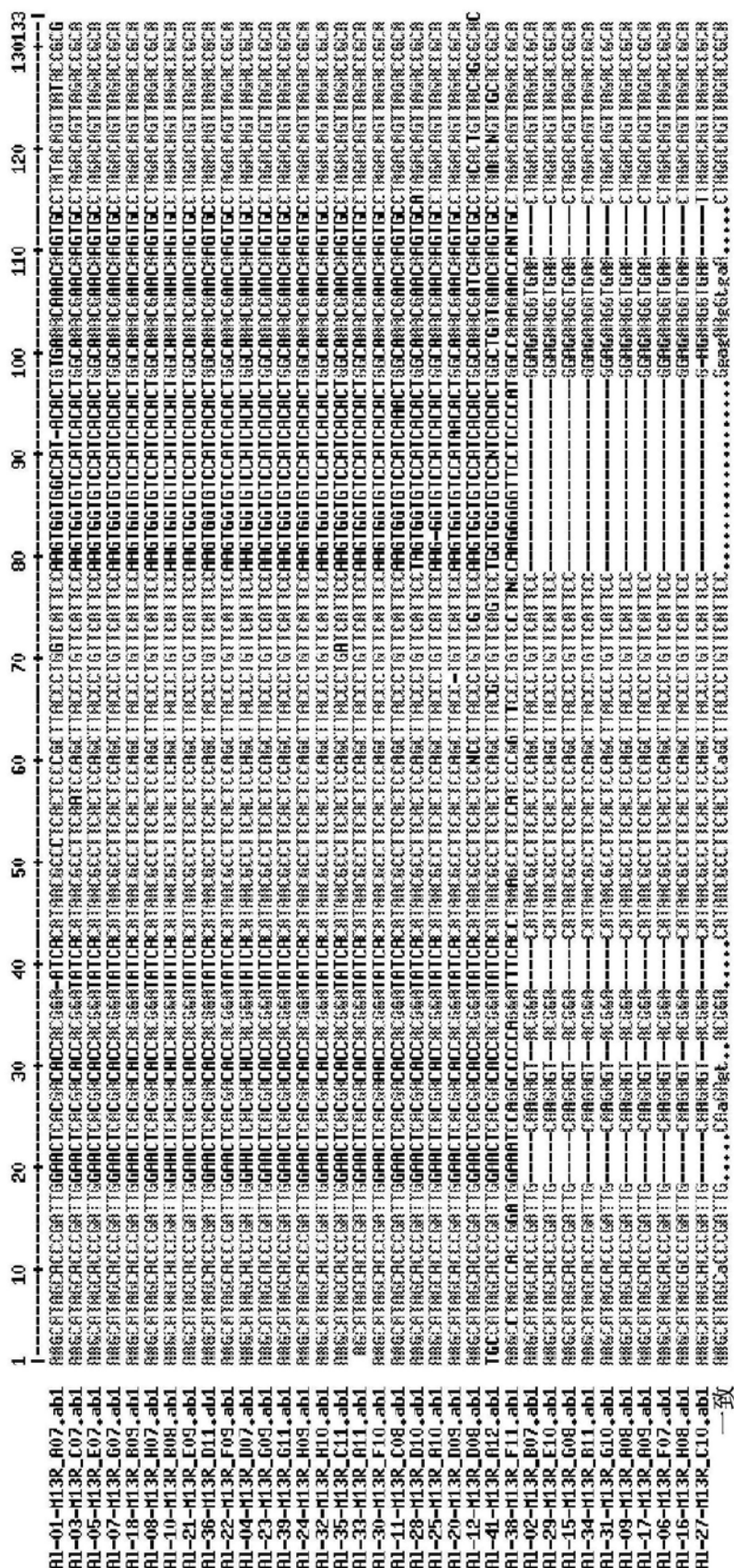


图43