(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117295751 A
(43)申请公布日 2023.12.26

(21)申请号 202280032974.0
(22)申请日 2022.05.02
(30)优先权数据
63/184,429 2021.05.05 US
(85)PCT国际申请进入国家阶段日
2023.11.03
(86)PCT国际申请的申请数据
PCT/EP2022/061723 2022.05.02
(87)PCT国际申请的公布数据
WO2022/233795 EN 2022.11.10
(71)申请人 伊鲁米纳剑桥有限公司
地址 英国剑桥
(72)发明人 M·卡灵厄姆 N·N·罗曼诺夫 刘小海 J·皮科斯
(74)专利代理机构 北京市金杜律师事务所 11256
专利代理师 陈文平 王北南
(51)Int.Cl.
C07F 5/02(2006.01)

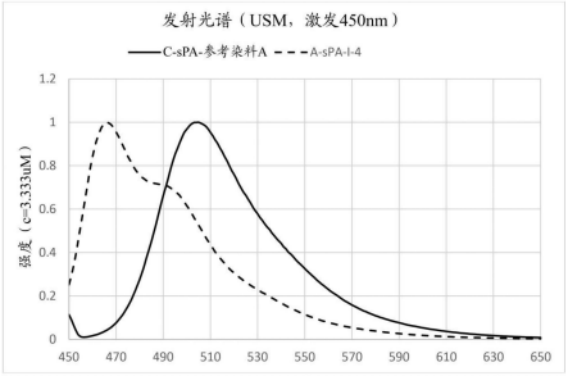
权利要求书7页 说明书43页 附图4页

(54)发明名称

含有双硼稠合杂环的荧光染料及其在测序中的用途

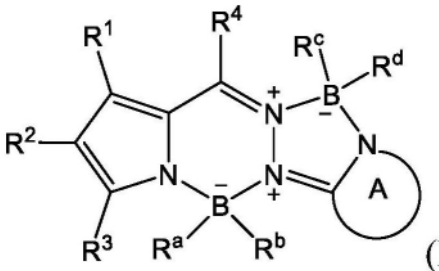
(57)摘要

本申请涉及含有双硼稠合杂环的取代的染料以及其作为荧光标记的用途。这些化合物可以用作用于核酸测序应用中的核苷酸的荧光标记。



CN 117295751 A

1.一种式(I)的化合物：



其盐或内消旋形式，

(I)、

其中R^1、R^2、R^3和R^4中的每一者独立地是H、未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、未取代或取代的氨基、卤基、氰基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基；

R^a、R^b、R^c和R^d中的每一者独立地是卤基、氰基、C_1-C_6烷基，C_1-C_6卤代烷基、C_1-C_6烷氧基、C_1-C_6卤代烷氧基、C_6-C_{10}芳基、C_6-C_{10}芳氧基或-O-C(=O)R^5；

另选地，当R^a和R^b都是-O-C(=O)R^5时，两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基；当R^c和R^d都是-O-C(=O)R^5时，两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基；

R^5是未取代或取代的C_1-C_6烷基；

环A是任选地被一个或多个R^6取代的6元至10元杂芳基；

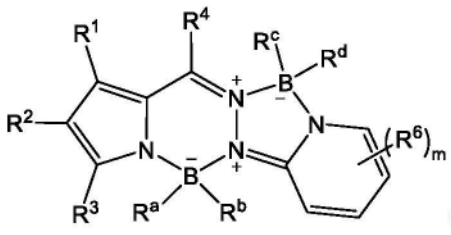
每个R^6独立地是未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、-NR^7R^8、卤基、氰基、羧基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基；

R^7和R^8中的每一者独立地是H、未取代或取代的C_1-C_6烷基，或者R^7和R^8连同它们所附接的氮原子一起形成未取代或取代的3元至10元杂环基；

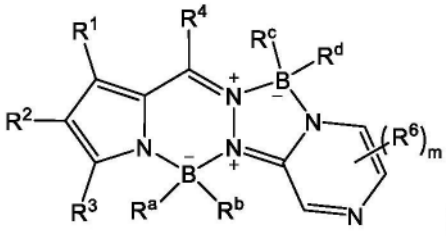
前提条件是R^1、R^2、R^3、R^4和环A中的至少一者包含羧基基团。

2.根据权利要求1所述的化合物，其中环A是任选地被一个或多个R^6取代的六元杂芳基。

3.根据权利要求1或2所述的化合物，所述化合物具有式(Ia)或(Ib)的结构：



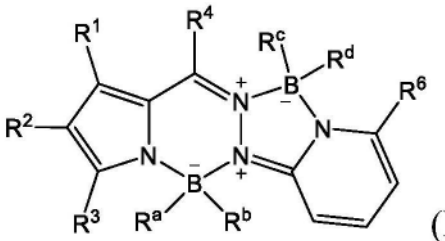
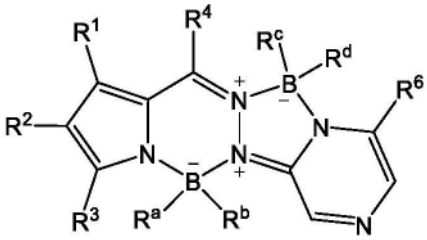
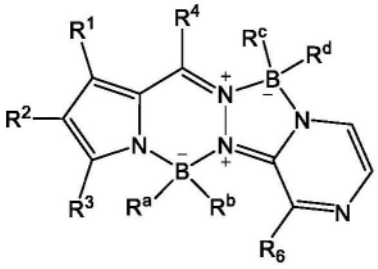
(Ia)、



(Ib)，

或其盐或内消旋形式，其中m是0、1、2或3。

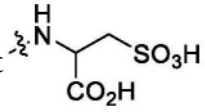
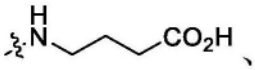
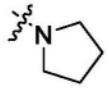
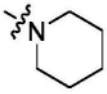
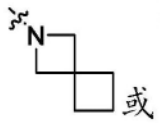
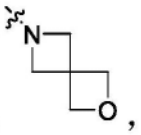
4.根据权利要求3所述的化合物，所述化合物具有式(Ic)、(Id)或(Ie)的结构：


(Ic)、

(Id)、

(Ie)，

或其盐或内消旋形式。

5.根据权利要求1至4中任一项所述的化合物，其中每个R^6独立地是卤基、氰基、羧基、未取代或取代的C_1-C_6烷基、未取代的苯基、被羧基取代的苯基、未取代的5元杂芳基、被羧基取代的5元杂芳基，或-NR^7R^8。

6.根据权利要求5所述的化合物，其中R^6是-NR^7R^8，其中R^7是H并且R^8是被一个或多个选自由羧基、磺基和磺酸根组成的组的取代基取代的C_1-C_6烷基，或者R^7和R^8连同它们所附接的氮原子一起形成任选地被羧基取代的3元至10元杂环基。

7.根据权利要求6所述的化合物，其中R^6是、、、、、或，其中所述环结构中的每一者任选地被羧基取代。

8.根据权利要求1至7中任一项所述的化合物，其中R^1、R^2和R^3中的每一者是H。

9.根据权利要求1至7中任一项所述的化合物，其中R^1、R^2和R^3中的每一者独立地是未取代的C_1-C_6烷基。

10.根据权利要求9所述的化合物，其中R^1和R^3中的每一者是甲基并且R^2是乙基。

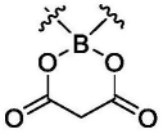
11.根据权利要求1至7中任一项所述的化合物，其中R^1、R^2和R^3中的两者是H或未取代的C_1-C_6烷基，并且R^1、R^2和R^3中的一者是卤基、羧基或被羧基取代的C_1-C_6烷基。

12.根据权利要求1至11中任一项所述的化合物，其中R^4是H或未取代的C_1-C_6烷基。

13.根据权利要求1至10中任一项所述的化合物，其中R^4是各自被羧基取代的C_1-C_6烷基或苯基。

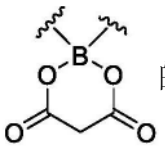
14.根据权利要求1至13中任一项所述的化合物，其中R^a和R^b中的每一者独立地是氟、氰基、甲基、三氟甲基、甲氧基或-O-酰基(-OC(=O)CH_3)。

15.根据权利要求14所述的化合物，其中R^a和R^b都是氟、甲基、三氟甲基、甲氧基或-O-酰基。

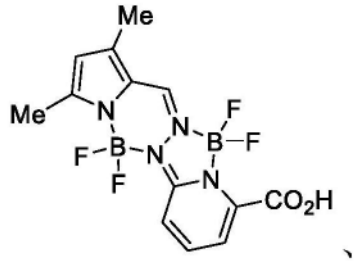
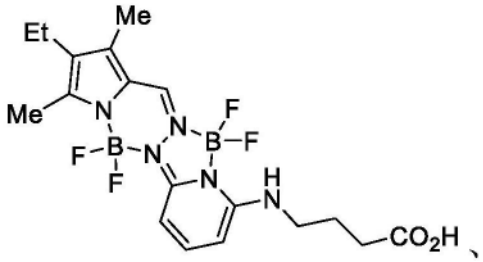
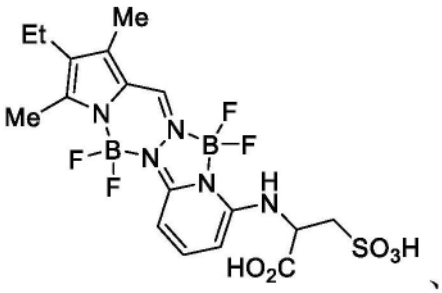
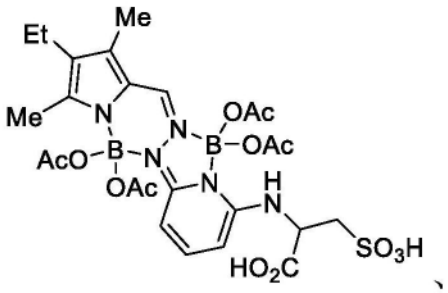
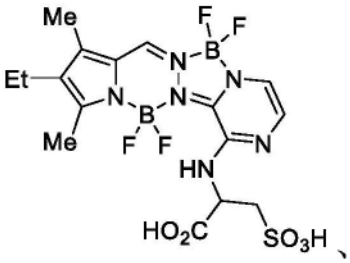
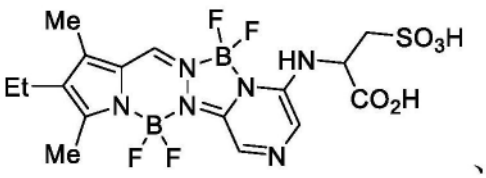
16.根据权利要求1至13中任一项所述的化合物,其中R^a和R^b都是-OC(=O)R^5,并且两个R^5连同它们所附接的原子一起形成具有结构  的6元杂环基。

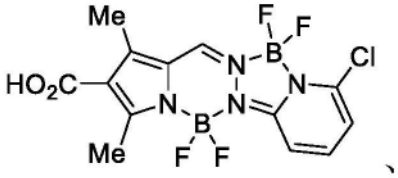
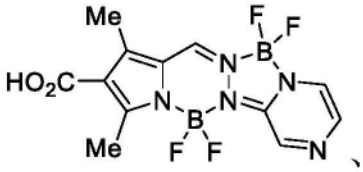
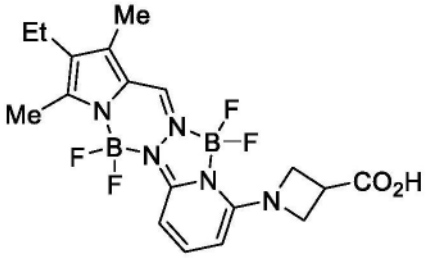
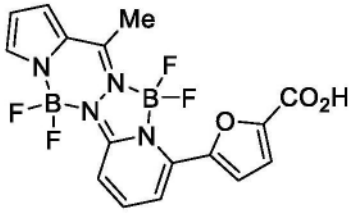
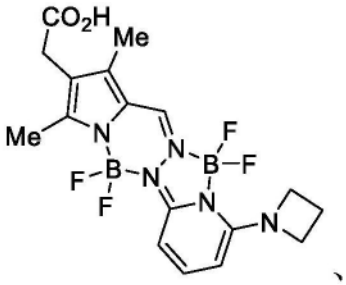
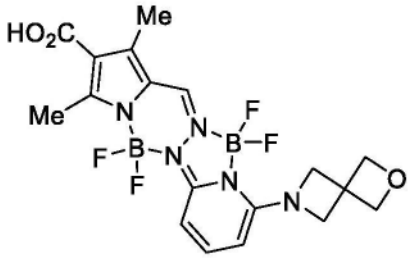
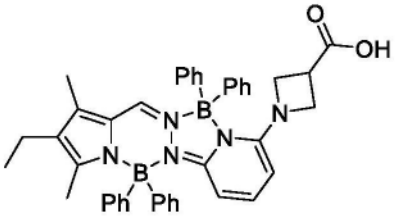
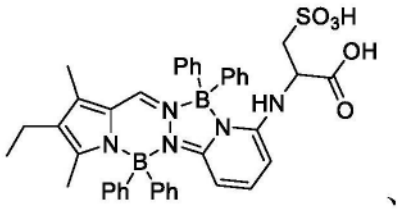
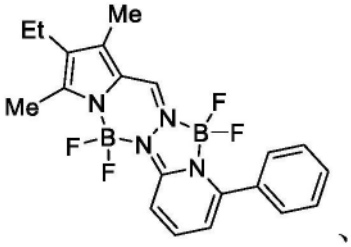
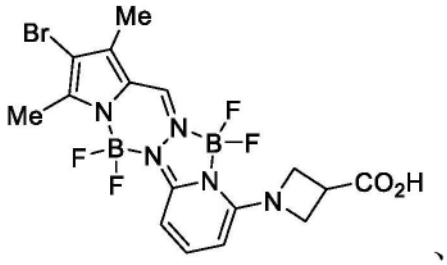
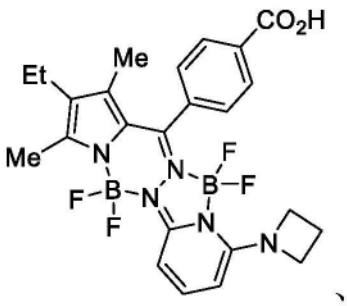
17.根据权利要求1至16中任一项所述的化合物,其中R^c和R^d中的每一者独立地是氟、氰基、甲基、三氟甲基、甲氧基或-O-酰基(-OC(=O)CH_3)。

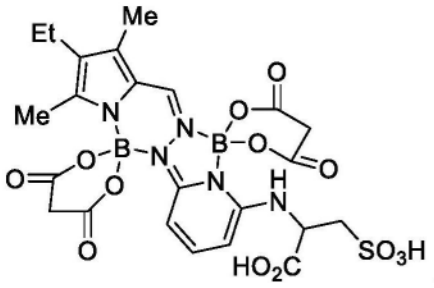
18.根据权利要求17所述的化合物,其中R^c和R^d都是氟、甲基、三氟甲基、甲氧基或-O-酰基。

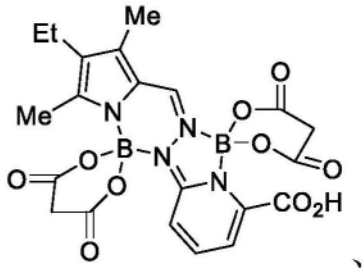
19.根据权利要求1至16中任一项所述的化合物,其中R^c和R^d都是-OC(=O)R^5,并且两个R^5连同它们所附接的原子一起形成具有结构  的6元杂环基。

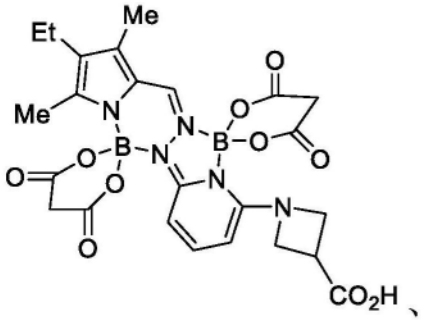
20.根据权利要求1所述的化合物,所述化合物选自由以下项组成的组:

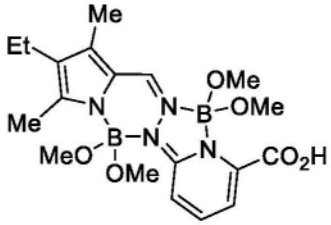

、

、

、

、

、

、

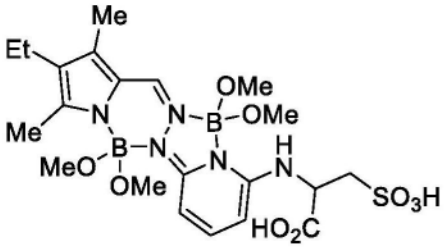

、

、

、

、

、

、

、

、

、

、

、

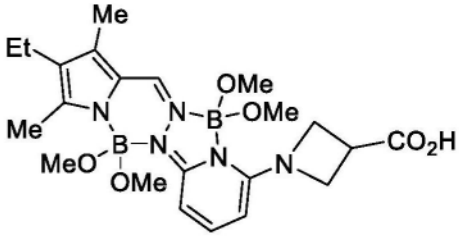


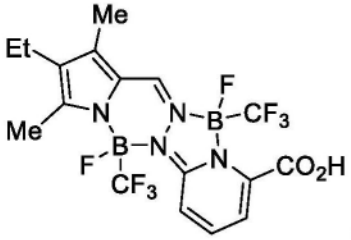


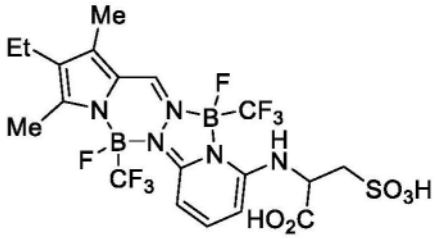


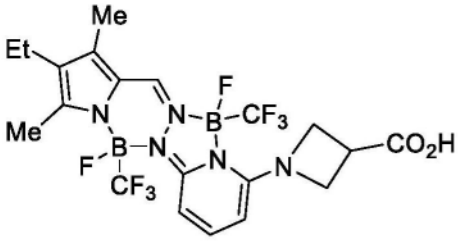


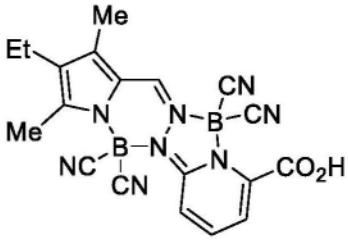


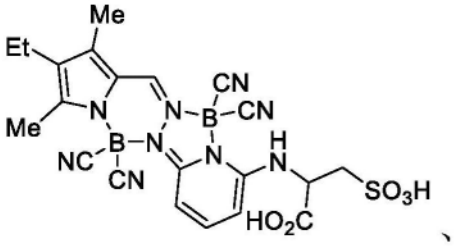


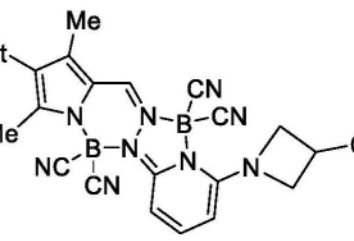


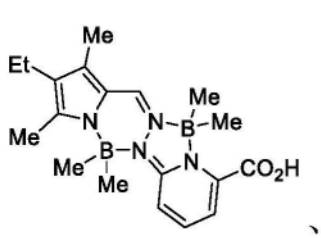
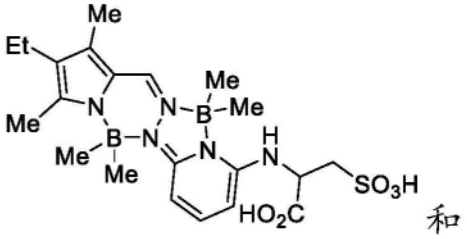
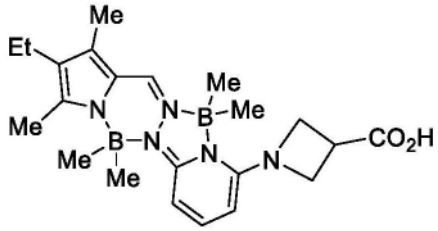











、

和

以及其盐和内消旋形式。

，

21.一种核苷酸,所述核苷酸用根据权利要求1至20中任一项所述的化合物标记。

22.根据权利要求21所述的标记的核苷酸,其中所述化合物经由式(I)的羧基基团附接到所述核苷酸。

23.根据权利要求21或22所述的标记的核苷酸,其中所述化合物通过连接基部分附接到嘧啶碱基的C5位置或7-脱氮嘌呤碱基的C7位置。

24.根据权利要求21至23中任一项所述的标记的核苷酸,所述标记的核苷酸还包含共价附接到所述核苷酸的核糖或脱氧核糖的3'羟基封端基团。

25.一种寡核苷酸或多核苷酸,所述寡核苷酸或多核苷酸包含掺入其中的根据权利要求21至24中任一项所述的标记的核苷酸。

26.根据权利要求25所述的寡核苷酸或多核苷酸,其中所述寡核苷酸或多核苷酸与固定在固体载体的表面上的靶多核苷酸至少部分地互补并杂交。

27.根据权利要求26所述的寡核苷酸或多核苷酸,其中所述固体载体包括多个固定在其上的靶多核苷酸的阵列。

28.一种试剂盒,所述试剂盒包含第一类型的根据权利要求21至24中任一项所述的标记的核苷酸。

29.根据权利要求28所述的试剂盒,所述试剂盒还包含第二类型的标记的核苷酸,其中所述第二类型的标记的核苷酸用与所述第一类型的标记的核苷酸不同的化合物标记。

30.根据权利要求29所述的试剂盒,所述试剂盒还包含第三类型的核苷酸和第四类型的核苷酸,其中所述第二类型的核苷酸、所述第三类型的核苷酸和所述第四类型的核苷酸中的每一者用不同的化合物标记,其中每种标记具有能够与其他标记区分的不同的吸光度最大值。

31.根据权利要求28所述的试剂盒,其中所述试剂盒包含四种类型的核苷酸,其中第一类型的核苷酸是根据权利要求21至24中任一项所述的标记的核苷酸,第二类型的核苷酸携带第二标记,第三类型的核苷酸携带第三标记,并且第四类型的核苷酸未标记(暗色)。

32.根据权利要求28所述的试剂盒,其中所述试剂盒包含四种类型的核苷酸,其中第一类型的核苷酸是根据权利要求21至24中任一项所述的标记的核苷酸,第二类型的核苷酸携带第二标记,第三类型的核苷酸包含携带两种标记的第三类型的核苷酸的混合物,并且第四类型的核苷酸未标记(暗色)。

33.根据权利要求28至32中任一项所述的试剂盒,所述试剂盒还包含DNA聚合酶和一种或多种缓冲液组合物。

34.一种确定多个靶多核苷酸的序列的方法,所述方法包括:

(a)使固体载体在杂交条件下与包含测序引物的溶液接触,其中所述固体载体包含多个固定在其上的不同靶多核苷酸;并且所述测序引物与所述靶多核苷酸的至少一部分互补;

(b)使所述固体载体在适用于DNA聚合酶介导的引物延伸的条件下与包含DNA聚合酶和四种不同类型的核苷酸中的一者或多者的水溶液接触,其中所述标记的核苷酸中的至少一者是根据权利要求21至24中任一项所述的核苷酸;

(c)将一种类型的核苷酸掺入到所述测序引物中以产生延伸的拷贝多核苷酸;以及

(d)对所述延伸的拷贝多核苷酸进行一个或多个荧光测量以确定所掺入的核苷酸的身份。

35.根据权利要求34所述的方法,所述方法还包括(e)从掺入到所述延伸的拷贝多核苷酸中的所述核苷酸中去除所述标记和所述3'封端基团。

36.根据权利要求35所述的方法,所述方法还包括(f)在所述从所掺入的核苷酸中去除所述标记和所述3'封端基团之后对所述固体载体进行洗涤。

37.根据权利要求36所述的方法,所述方法还包括重复步骤(b)至(f),直到确定所述靶多核苷酸的至少一部分的序列。

38.根据权利要求37所述的方法,其中步骤(b)至(f)重复至少50、100、150、200、250或300个循环。

39.根据权利要求35至38中任一项所述的方法,其中在单一化学反应中去除来自掺入到所述延伸的拷贝多核苷酸中的所述核苷酸的所述标记和所述3'封端基团。

40.根据权利要求34至39中任一项所述的方法,其中所述方法在自动化测序仪器上执行,并且其中所述自动化测序仪器包含在不同波长处操作的两个光源。

41.根据权利要求40所述的方法,其中一个光源在约450nm至约460nm处操作。

42.根据权利要求34至41中任一项所述的方法,其中所述四种类型的核苷酸包括dATP、dCTP、dGTP和dTTP或dUTP或它们的非天然核苷酸类似物。

含有双硼稠合杂环的荧光染料及其在测序中的用途

背景技术

技术领域

[0001] 本公开涉及含有双硼稠合杂环的荧光染料及其在核酸测序应用中作为核苷酸的荧光标记的用途。

背景技术

[0002]

[0003] 带有荧光标记的核酸的非放射性检测是分子生物学中的重要技术。在重组DNA技术中采用的许多程序以前依赖于使用带有放射性标记的核苷酸或多核苷酸,例如^{32}P。放射性化合物允许灵敏地检测核酸和其他感兴趣的分子。然而,在放射性同位素的使用中存在严重的局限性,诸如它们的费用、有限的保质期、不足的灵敏度和更重要的安全考虑因素。消除对放射性标记的需求同时降低了安全性风险以及与例如试剂处理相关的环境影响和成本。适合于非放射性荧光检测的方法包括以非限制性示例的方式进行的自动化DNA测序、杂交方法、聚合酶链反应产物的实时检测和免疫测定。

[0004] 对于许多应用而言,希望采用多个光谱可区分的荧光标记,以便实现对多个空间重叠分析物的独立检测。在此类多重方法中,可以减少反应容器的数量,从而简化实验方案并且促进专用试剂盒的生产。例如,在多色自动化DNA测序系统中,多重荧光检测允许在单个电泳通道中分析多个核苷酸碱基,从而通过单色方法提高吞吐量,并且降低与通道间电泳迁移率变化相关的不确定性。

[0005] 然而,多重荧光检测可能存在问题,并且存在的许多重要因素会限制适当荧光标记的选择。首先,在给定的应用中可能难以找到具有基本上分辨的吸收和发射光谱的染料化合物。另外,当若干荧光染料一起使用时,通过同时激发在可区分光谱区域中产生荧光信号可能是复杂的,因为这些染料的吸收带通常很分开,所以即使是两种染料也难以达到相当的荧光激发效率。许多激发方法使用高功率光源,如激光器,因此染料必须具有足够的光稳定性以承受此类激发。对分子生物学方法特别重要的最后一个考虑因素是荧光染料必须与试剂化学性质相容的程度,诸如DNA合成溶剂和试剂、缓冲液、聚合酶和连接酶。

[0006] 随着测序技术进展,还需要开发满足所有上述限制并且特别适用于高通量分子方法(诸如固相测序等)的另外的荧光染料化合物、其核酸缀合物和多个染料组。

[0007] 具有改进的荧光特性(诸如合适的荧光强度、形状和荧光带的波长最大值)的荧光染料分子可以提高核酸测序的速度和准确性。当在水基生物缓冲液中并且在较高温度下进行测量时,强荧光信号尤其重要,因为大多数有机染料的荧光强度在此类条件下显著较低。此外,染料所附接的碱基的性质也影响荧光最大值、荧光强度和染料的其他光谱特性。核碱基与荧光染料之间的序列特异性相互作用可以通过荧光染料的特异性设计来定制。荧光染料结构的优化可以提高核苷酸掺入的效率、降低测序误差的水平,并且减少核酸测序中试剂的使用,从而降低核酸测序的成本。

[0008] 一些光学和技术的开发已引起图像质量的极大改善,但是最终受到光学分辨率差

的限制。一般来讲,光学显微镜法的光学分辨率限于以所使用的光的波长的大约一半间隔开的对象。实际上,只有相距相当远(至少200nm至350nm)的对象才可以通过光学显微镜法分辨。提高图像分辨率并且增加每单位表面积的可分辨对象数量的一种方法是使用较短波长的激发光。例如,如果用相同的光学器件将光波长缩短Δλ约100nm,则分辨率将更好(约Δ50nm/(约15%)),将记录较少失真的图像,并且可识别区域上的对象的密度将增加约35%。

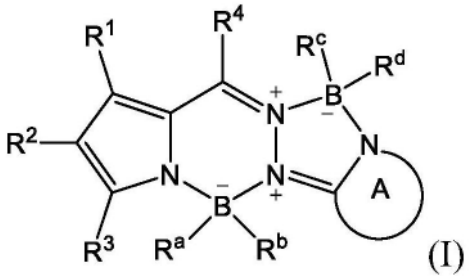
[0009] 某些核酸测序方法采用激光来激发染料标记的核苷酸并对其进行检测。这些仪器使用较长波长的光,诸如红色激光,连同能够在660nm处激发的适当染料。为了在保持可用分辨率的同时检测更密集堆积的核酸测序簇,可以使用较短波长的蓝色光源(450nm至460nm)。在这种情况下,光学分辨率将不受较长波长红色荧光染料的发射波长的限制,而是受能够由次长波长光源(例如,由532nm处的"绿色激光")激发的染料的发射的限制。因此,需要用于测序应用中的荧光检测的蓝色染料标记。

[0010] 尽管蓝色染料化学和相关的激光技术已经得到改进,但用于核苷酸标记的具有强荧光的合适的市售蓝色染料仍然相当罕见。然而,某些染料,特别是香豆素染料在延长的时间段内在水性环境中不稳定。例如,在碱性条件下,某些香豆素染料可能很容易被亲核试剂攻击,从而导致染料的干扰或劣化。含硼荧光染料诸如BODIPY、BOPHY、BOPPY、BOPYPY、BOAHY、BOPAHY已经在多篇科学文献和专利出版物,例如J Am Chem Soc 2014,136(15):5623-5626、Organic Letters 2014,16(11):3048-3051、Organic Letters 2018,20(15):4462-4466、Chinese Chemical Letters 2019,30:2271-2273、Organic Letters 2020,22(12):4588-4592、Chem Communications 2020,56(43):5791-5794、Chemistry A Eur J 2020,26(4):863-872、WO 2015/77427和CN108516985A中报道。然而,设计具有适当吸附、良好化学稳定性和斯托克斯位移的新的含硼染料作为用于测序应用的核酸标记仍然具有挑战性。

发明内容

[0011] 本文描述了一类新的含有双硼稠合杂环的染料,其具有改善的化学稳定性和在蓝光激发(例如,蓝色LED或激光,例如,在约450nm至约460nm处)下的强荧光。这些染料还具有适用于核酸标记的高度可调的吸收和发射特性。

[0012] 本公开的一个方面涉及式(I)的化合物或其盐或内消旋形式:

[0013] 

(I)

[0014] 其中R^1、R^2、R^3和R^4中的每一者独立地是H、未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、未取代或取代的氨基、卤基、氰基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取

代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基；

[0015] R^a、R^b、R^c和R^d中的每一者独立地是卤基、氰基、C_1-C_6烷基，C_1-C_6卤代烷基、C_1-C_6烷氧基、C_1-C_6卤代烷氧基、C_6-C_{10}芳基、C_6-C_{10}芳氧基或-O-C(=O)R^5；

[0016] 另选地，当R^a和R^b都是-O-C(=O)R^5时，两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基；当R^c和R^d都是-O-C(=O)R^5时，两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基；

[0017] R^5是未取代或取代的C_1-C_6烷基；

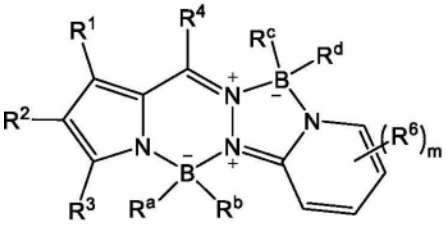
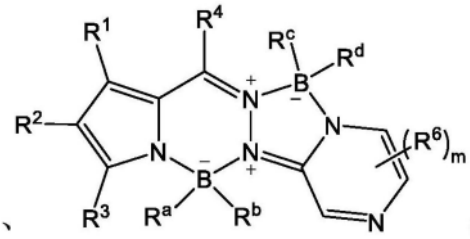
[0018] 环A是任选地被一个或多个R^6取代的6元至10元杂芳基；

[0019] 每个R^6独立地是未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、-NR^7R^8、卤基、氰基、羧基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基；

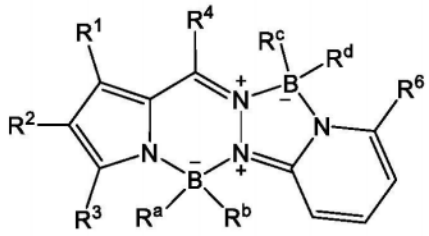
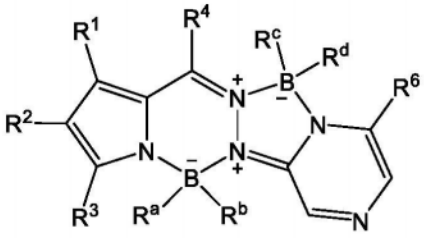
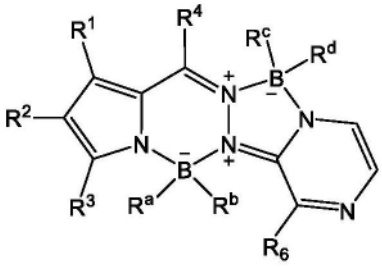
[0020] R^7和R^8中的每一者独立地是H、未取代或取代的C_1-C_6烷基，或者R^7和R^8连同它们所附接的氮原子一起形成未取代或取代的3元至10元杂环基；

[0021] 前提条件是R^1、R^2、R^3、R^4和环A中的至少一者包含羧基基团。

[0022] 在一些实施方案中，式(I)的化合物还可以具有式(Ia)或(Ib)的结构：

[0023]  (Ia)、 (Ib)，

或其盐或内消旋形式，其中m是0、1、2或3。在另外的实施方案中，该化合物可以具有式(Ic)、(Id)或(Ie)的结构：

[0024]  (Ic)、 (Id)、 (Ie)，

或其盐或内消旋形式。

[0025] 在一些方面，本公开的化合物被如以下等底物部分标记或与该底物部分缀合：核苷、核苷酸、多核苷酸、多肽、碳水化合物、配体、颗粒、细胞、半固体表面(例如，凝胶)或固体表面。标记或缀合可经由羧基基团进行，该羧基基团可以使用本领域已知的方法与部分(诸

如核苷酸)或结合到其上的连接基上的氨基或羟基基团进行反应以形成酰胺或酯。

[0026] 本公开的另一个方面涉及包含连接基基团的染料化合物,以实现例如与底物部分(诸如核苷酸)共价附接。连接可以在染料的任何位置处进行。在一些实施方案中,连接可经由R^1、R^2、R^3、R^4和式(I)的环A中的一者进行。

[0027] 本公开的另一方面提供由下式定义的标记的核苷或核苷酸化合物:

[0028] N-L-染料

[0029] 其中N是核苷或核苷酸;

[0030] L是任选的连接基部分;并且

[0031] 染料是根据本公开的式(I)的荧光化合物的部分,其中式(I)(例如,(Ia)、(Ib)、(Ic)、(Id)或(Ie))的化合物的官能团(例如,羧基基团)与连接基部分或核苷/核苷酸的氨基或羟基基团进行反应以形成共价键。

[0032] 本公开的一些另外的方面涉及用式(I)(例如,(Ia)、(Ib)、(Ic)、(Id)或(Ie))的化合物标记的寡核苷酸或多核苷酸。

[0033] 本公开的一些另外的方面涉及一种试剂盒,该试剂盒包含可以用于各种免疫学测定、寡核苷酸或核酸标记或用于DNA合成测序的染料化合物(游离或标记形式)。在又一方面,本公开提供了包含染料“组”的试剂盒,该染料“组”特别适合于在自动化仪器平台上的合成测序循环。在一些方面是包括一种或多种核苷酸的试剂盒,其中至少一种核苷酸是本文所述的标记的核苷酸。

[0034] 本公开的另一方面涉及一种确定多个靶多核苷酸的序列的方法,所述方法包括:

[0035] (a)使固体载体在杂交条件下与包含测序引物的溶液接触,其中所述固体载体包含多个固定在其上的不同靶多核苷酸;并且所述测序引物与所述靶多核苷酸的至少一部分互补;

[0036] (b)使所述固体载体在适用于DNA聚合酶介导的引物延伸的条件下与包含DNA聚合酶和四种不同类型的核苷酸中的一者或多者的水溶液接触,其中至少一种类型的核苷酸是本文所述的标记的核苷酸;

[0037] (c)将一种类型的核苷酸掺入到所述测序引物中以产生延伸的拷贝多核苷酸;以及

[0038] (d)对所述延伸的拷贝多核苷酸进行一个或多个荧光测量以确定所掺入的核苷酸的身份。

附图说明

[0039] 图1示出了当被蓝光(450nm)激发时用参考染料A标记的ffA-spA-I-4和ffC的发射光谱。

[0040] 图2示出了在相同条件下与用参考染料A标记的完全C相比,染料I-1和I-3的随时间变化的剩余荧光信号百分比。

[0041] 图3示出了在MiSeq™上含有ffA-spA-I-4的掺入混合物与两种参考掺入混合物相比的定相百分比。

[0042] 图4A和图4B是当使用蓝光时以1X和5X剂量,在第26个循环时对于含有ffA-spA-I-3的掺入混合物获得的散点图。

[0043] 图4C和图4D是当使用蓝光时以1X和5X剂量,在第26个循环时对于含有ffA-spA-I-

4的掺入混合物获得的散点图。

具体实施方式

[0044] 本公开的实施方案涉及含有双硼稠合杂环的染料,其具有增强的荧光强度、可调的斯托克斯位移和改善的化学稳定性。在一些实施方案中,本文所述染料的斯托克斯位移在约15nm至50nm(例如,约20nm)的范围内。本文所述的含双硼染料可以用于具有两通道检测(绿光激发和蓝光激发)的因美纳公司(Illumina)的测序平台,例如MiSeq™。

[0045] 定义

[0046] 本文所用的章节标题仅用于组织目的,而不应理解为限制所述主题。

[0047] 应注意,如在本说明书和所附权利要求书中所用的,单数形式“一个”、“一种”和“所述”包括复数指代物,除非清楚和明确地限于一指代物。对于本领域技术人员而言显而易见的是,在不脱离本教导的精神或范围的情况下,可以对本文描述的各种实施方案进行各种修改和变化。因此,本文描述的各种实施方案旨在覆盖在所附权利要求书及其等效物的范围内的其他修改和变化。

[0048] 除非另有定义,否则本文所用的所有技术和科学术语的含义与本领域的普通技术人员通常理解的含义相同。术语“包括”以及其他形式(诸如“包括(include/includes/included)”)的使用不是限制性的。术语“具有”以及其他形式(诸如“具有(have/has/had)”)的使用不是限制性的。如本说明书中所用,无论是在过渡短语中还是在权利要求的正文中,术语“包含(comprise(s))”和“包含(comprising)”都将被解释为具有开放式含义。即,上述术语应与短语“至少具有”或“至少包括”同义地解释。例如,当在过程的上下文中使用时,术语“包含”表示该过程至少包括所列举的步骤,但是也可包括额外步骤。当在化合物、组合物或设备的上下文中使用时,术语“包含”是指该化合物、组合物或设备至少包含所列举的特征或组分,但是也可包含额外特征或组分。

[0049] 如本文所用,常见的有机缩写定义如下:

[0050]	℃	以摄氏度计的温度
[0051]	dATP	三磷酸脱氧腺苷
[0052]	dCTP	三磷酸脱氧胞苷
[0053]	dGTP	三磷酸脱氧鸟苷
[0054]	dTTP	三磷酸脱氧胸苷
[0055]	ddNTP	三磷酸双脱氧核苷酸
[0056]	ffA	完全官能化A核苷酸
[0057]	ffC	完全官能化C核苷酸
[0058]	ffG	完全官能化G核苷酸
[0059]	ffN	完全官能化核苷酸
[0060]	ffT	完全官能化T核苷酸
[0061]	h	小时
[0062]	RT	室温
[0063]	SBS	合成测序
[0064]	USM	通用扫描混合物

[0065] 如本文所用,术语“阵列”是指连接到一个或多个底物的一组不同探针分子,使得这些不同探针分子可根据相对位置彼此区分。阵列可包括各自位于底物上的不同可寻址位置处的不同探针分子。替代性地或另外地,阵列可包括各自带有不同探针分子的单独底物,其中可根据底物在底物所连接到的表面上的位置或根据底物在液体中的位置来识别不同的探针分子。其中单独底物位于表面上的示例性阵列包括但不限于在孔中包含珠粒的那些,如例如美国专利第6,355,431B1号、US2002/0102578和PCT公布第WO 00/63437号中所述的。例如,在美国专利第6,524,793号中描述了在本发明中可用于例如使用微流体装置,诸如荧光活化细胞分选器(FACS)区分液体阵列中的珠粒的示例性形式。可用于本发明的阵列的另外的示例包括但不限于美国专利5,429,807;5,436,327;5,561,071;5,583,211;5,658,734;5,837,858;5,874,219;5,919,523;6,136,269;6,287,768;6,287,776;6,288,220;6,297,006;6,291,193;6,346,413;6,416,949;6,482,591;6,514,751和6,610,482;和WO 93/17126;WO 95/11995;WO 95/35505;EP 742 287;和EP 799 897中描述的那些。

[0066] 如本文所用,术语“共价连接的”或“共价键合的”是指形成特征在于原子之间共用电子对的化学键合。例如,共价连接的聚合物涂层是指与底物的官能化表面形成化学键的聚合物涂层,这与经由其他方式(例如,粘附或静电相互作用)粘附到该表面形成比较。应当理解,共价连接到表面的聚合物也可以经由除共价连接之外的方式键合。

[0067] 如本文所用,术语“卤素”或“卤基”意味着元素周期表第7列的放射性稳定原子中的任一种,例如,氟、氯、溴或碘,其中氟和氯是优选的。

[0068] 如本文所用,其中“a”和“b”是整数的“C_a至C_b”是指烷基、烯基或炔基基团中的碳原子数,或环烷基或芳基基团的环原子数。即,烷基、烯基、炔基、环烷基的环和芳基的环可以含有“a”至“b”个(包括端值在内)的碳原子。例如,“C_1至C_4烷基”基团是指具有1个至4个碳的所有烷基基团,即CH_3-、CH_3CH_2-、$CH_3CH_2CH_2$-、$(CH_3)_2CH$-、$CH_3CH_2CH_2CH_2$-、$CH_3CH_2CH(CH_3)$-和$(CH_3)_3C$-;C_3至C_4环烷基基团是指具有3至4个碳原子的所有环烷基基团,即环丙基和环丁基。类似地,“4至6元杂环基”基团是指具有4至6个总环原子的所有杂环基基团,例如氮杂环丁烷、氧杂环丁烷、噁唑啉、吡咯烷、哌啶、哌嗪、吗啉等。如果对于烷基、烯基、炔基、环烷基或芳基基团没有指定“a”和“b”,则将假设这些定义中描述的最广泛范围。如本文所用,术语“C_1-C_6”包括C_1、C_2、C_3、C_4、C_5和C_6,以及由这两个数字中的任一者定义的范围。例如,C_1-C_6烷基包括C_1烷基、C_2烷基、C_3烷基、C_4烷基、C_5烷基和C_6烷基、C_2-C_6烷基、C_1-C_3烷基等。类似地,C_2-C_6烯基包括C_2烯基、C_3烯基、C_4烯基、C_5烯基和C_6烯基、C_2-C_5烯基、C_3-C_4烯基等;并且C_2-C_6炔基包括C_2炔基、C_3炔基、C_4炔基、C_5炔基和C_6炔基、C_2-C_5炔基、C_3-C_4炔基等。C_3-C_8环烷基各自包括含有3、4、5、6、7和8个碳原子或由任何两个数字限定的范围的烃环,诸如C_3-C_7环烷基或C_5-C_6环烷基。

[0069] 如本文所用,“烷基”是指完全饱和(即,不包含双键和三键)的直链或支链烃链。烷基基团可以具有1至20个碳原子(每当在本文中出现时,诸如“1至20”的数值范围是指给定范围内的每个整数;例如,“1至20个碳原子”意味着烷基基团可以由1个碳原子、2个碳原子、3个碳原子等,最多至并包括20个碳原子组成,但本定义还涵盖出现其中没有指定数值范围的术语“烷基”)。烷基基团还可以是具有1至9个碳原子的中等大小烷基。烷基基团还可以是具有1至6个碳原子的低级烷基。仅以举例的方式,“C_{1-6}烷基”或“C_1-C_6烷基”表示烷基链中存在一个至六个碳原子,即,烷基链选自由甲基、乙基、丙基、异丙基、正丁基、异丁基、仲丁基

和叔丁基组成的组。典型的烷基基团包括但绝不限于甲基、乙基、丙基、异丙基、丁基、异丁基、叔丁基、戊基、己基等。

[0070] 如本文所用，“烷氧基”是指式-OR(其中R为如上文所定义的烷基)，诸如“C_{1-9}烷氧基”或“C_1-C_9烷氧基”，包括但不限于甲氧基、乙氧基、正丙氧基、1-甲基乙氧基(异丙氧基)、正丁氧基、异丁氧基、仲丁氧基和叔丁氧基等。

[0071] 如本文所用，“-OAc”或“-O-酰基”是指具有结构-O-C(=O)CH_3的乙酰氧基。

[0072] 如本文所用，“烯基”是指包含一个或多个双键的直链或支链烃链。烯基基团可以具有2至20个碳原子，但本定义还涵盖出现其中没有指定数值范围的术语“烯基”。烯基基团还可以是具有2至9个碳原子的中等大小烯基。烯基基团还可以是具有2至6个碳原子的低级烯基。仅以举例的方式，“C_2-C_6烯基”或“C_{2-6}烯基”表示烯基链中存在两个至六个碳原子，即，烯基链选自由乙烯基、丙烯-1-基、丙烯-2-基、丙烯-3-基、丁烯-1-基、丁烯-2-基、丁烯-3-基、丁烯-4-基、1-甲基-丙烯-1-基、2-甲基-丙烯-1-基、1-乙基-乙烯-1-基、2-甲基-丙烯-3-基、丁-1,3-二烯基、丁-1,2-二烯基和丁-1,2-二烯-4-基组成的组。典型的烯基基团包括但绝不限于乙烯基、丙烯基、丁烯基、戊烯基和己烯基等。

[0073] 如本文所用，“炔基”是指包含一个或多个三键的直链或支链烃链。炔基基团可以具有2至20个碳原子，但本定义还涵盖出现其中没有指定数值范围的术语“炔基”。炔基基团还可以是具有2至9个碳原子的中等大小炔基。炔基基团还可以是具有2至6个碳原子的低级炔基。仅以举例的方式，“C_{2-6}炔基”或“C_2-C_6炔基”表示炔基链中存在两个至六个碳原子，即，炔基链选自由乙炔基、丙炔-1-基、丙炔-2-基、丁炔-1-基、丁炔-3-基、丁炔-4-基和2-丁炔基组成的组。典型的炔基基团包括但绝不限于乙炔基、丙炔基、丁炔基、戊炔基和己炔基等。

[0074] 术语“芳族”是指具有共轭π电子体系的环或环系，并且包括碳环芳族基团(例如，苯基)和杂环芳族基团(例如，吡啶)这两者。该术语包括单环基团或稠环多环(即，共用相邻原子对的环)基团，前提条件是整个环系是芳族的。

[0075] 如本文所用，“芳基”是指在环骨架中仅包含碳的芳族环或环系(即，共用两个相邻碳原子的两个或更多个稠环)。当芳基为环系时，该环系中的每个环均为芳族的。芳基基团可以具有6至18个碳原子，但本定义还涵盖出现其中没有指定数值范围的术语“芳基”。在一些实施方案中，芳基基团具有6至10个碳原子。芳基基团可以被命名为“C_6-C_{10}芳基”、“C_6或C_{10}芳基”或类似的名称。芳基基团的示例包括但不限于苯基、萘基、蒽基和葸基。

[0076] “芳烷基”或“芳基烷基”为作为取代基经由亚烷基基团连接的芳基基团，诸如“C_{7-14}芳烷基”等，包括但不限于苄基、2-苯基乙基、3-苯基丙基和萘基烷基。在一些情况下，亚烷基基团为低级亚烷基基团(即，C_{1-6}亚烷基基团)。

[0077] 如本文所用，“杂芳基”是指在环骨架中含有一个或多个杂原子(即，除碳之外的元素，包括但不限于氮、氧和硫)的芳族环或环系(即，共用两个相邻原子的两个或更多个稠环)。当杂芳基是环系时，该环系中的每个环均是芳族的。杂芳基基团可以具有5至18个环成员(即，构成环骨架的原子(包括碳原子和杂原子)的数目)，但本定义还涵盖出现其中没有指定数值范围的术语“杂芳基”。在一些实施方案中，杂芳基基团具有5至10个环成员或者5至7个环成员。杂芳基基团可以被命名为“5至7元杂芳基”、“5至10元杂芳基”或类似的名称。杂芳基环的示例包括但不限于呋喃基、噻吩基、酞嗪基、吡咯基、噁唑基、噻唑基、咪唑基、吡唑基、异噁唑基、异噻唑基、三唑基、噻二唑基、吡啶基、哒嗪基、嘧啶基、吡嗪基、三嗪基、喹

啉基、异喹啉基、苯并咪唑基、苯并噁唑基、苯并噻唑基、吲哚基、异吲哚基和苯并噻吩基。

[0078] “杂芳烷基”或“杂芳基烷基”为作为取代基经由亚烷基基团连接的杂芳基基团。示例包括但不限于2-噻吩基甲基、3-噻吩基甲基、呋喃基甲基、噻吩基乙基、吡咯基烷基、吡啶基烷基、异噁唑基烷基和咪唑基烷基。在一些情况下,亚烷基基团为低级亚烷基基团(即,C_{1-6}亚烷基基团)。

[0079] 如本文所用,“碳环基”意味着在环系骨架中仅含有碳原子的非芳族环状环或环系。当碳环基为环系时,两个或更多个环可以以稠合、桥接或螺接的方式接合在一起。碳环基可以具有任何饱和度,前提条件是环系中的至少一个环不是芳族的。因此,碳环基包括环烷基、环烯基和环炔基。碳环基基团可以具有3至20个碳原子,但本定义还涵盖出现其中没有指定数值范围的术语“碳环基”。碳环基基团也可以是具有3至10个碳原子的中等大小碳环基。碳环基基团也可以是具有3至6个碳原子的碳环基。碳环基基团可以被命名为“C_{3-6}碳环基”、“C_3-C_6碳环基”或类似的名称。碳环基环的示例包括但不限于环丙基、环丁基、环戊基、环己基、环己烯基、2,3-二氢-茚、双环[2.2.2]辛烷基、金刚烷基和螺[4.4]壬烷基。

[0080] 如本文所用,“环烷基”意味着完全饱和的碳环基环或环系。示例包括环丙基、环丁基、环戊基和环己基。

[0081] 如本文所用,“杂环基”是指在环骨架中含有至少一个杂原子的非芳族环状环或环系。杂环基可以稠合、桥接或螺接的方式接合在一起。杂环基可具有任何饱和度,前提条件是环系中的至少一个环不是芳族的。杂原子可以存在于环系中的非芳族环或芳族环中。杂环基基团可以具有3至20个环元(即,构成环骨架的原子(包括碳原子和杂原子)的数目),但本定义还涵盖出现其中没有指定数值范围的术语“杂环基”。杂环基基团也可以是具有3至10个环元的中等大小杂环基。杂环基基团也可以是具有3至6个环元的杂环基。杂环基基团可以被命名为“3至6元杂环基”或类似的名称。在优选的六元单环杂环基中,杂原子选自O、N或S中的一个至最多三个,并且在优选的五元单环杂环基中,杂原子选自一个或两个选自O、N或S的杂原子。杂环基环的示例包括但不限于吖庚因基、吖啶基、咔唑基、噌啉基、二氧戊环基、咪唑啉基、咪唑烷基、吗啉基、环氧乙烷基、氧杂环庚烷基、硫杂环庚烷基、哌啶基、哌嗪基、二氧代哌嗪基、吡咯烷基、吡咯烷酮基、吡咯烷二酮基、4-哌啶酮基、吡唑啉基、吡唑烷基、1,3-二噁英基、1,3-二噁烷基、1,4-二噁英基、1,4-二噁烷基、1,3-氧硫杂环己烷基、1,4-氧硫杂环己烯基、1,4-氧硫杂环己烷基、2H-1,2-噁嗪基、三噁烷基、六氢-1,3,5-三嗪基、1,3-间二氧杂环戊烯基、1,3-二氧戊环基、1,3-二噻吩基、1,3-二噻烷基、异噁唑啉基、异噁唑烷基、噁唑啉基、噁唑烷基、噁唑烷酮基、噻唑啉基、噻唑烷基、1,3-氧硫杂环戊烷基、二氢吲哚基、异二氢吲哚基、四氢呋喃基、四氢吡喃基、四氢噻吩基、四氢噻喃基、四氢-1,4-噻嗪基、硫代吗啉基、二氢苯并呋喃基、苯并咪唑烷基和四氢喹啉。

[0082] 如本文所用,“烷氧基烷基”或“(烷氧基)烷基”是指通过亚烷基基团连接的烷氧基基团,诸如C_2-C_8烷氧基烷基或(C_1-C_6烷氧基)C_1-C_6烷基,例如-$(CH_2)_{1-3}$-OCH_3。

[0083] “O-羧基”基团是指其中R选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“-OC(=O)R”基团,如本文所定义的。

[0084] “C-羧基”基团是指其中R选自由氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基组成的组的“-C(=O)OR”基团,如本文所定义的。非限制性示例包括羧基(即,-C(=O)OH)。

[0085] “磺酰基”基团是指其中R选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-SO_2R$”基团，如本文所定义的。

[0086] “亚磺酸基”基团是指“$-S(=O)OH$”基团。

[0087] “磺基”基团是指“$-S(=O)_2OH$”或“$-SO_3H$”基团。

[0088] “磺酸根”基团是指“$-SO_3^-$”基团。

[0089] “硫酸根”基团是指“$-SO_4^-$”基团。

[0090] “S-亚磺酰氨基”基团是指其中R_A和R_B各自独立地选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-SO_2NR_AR_B$”基团，如本文所定义的。

[0091] “N-亚磺酰氨基”基团是指其中R_A和R_b各自独立地选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-N(R_A)SO_2R_B$”基团，如本文所定义的。

[0092] “C-酰氨基”基团是指其中R_A和R_B各自独立地选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-C(=O)NR_AR_B$”基团，如本文所定义的。

[0093] “N-酰氨基”基团是指其中R_A和R_B各自独立地选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-N(R_A)C(=O)R_B$”基团，如本文所定义的。

[0094] “氨基”基团是指其中R_A和R_B各自独立地选自氢、C_{1-6}烷基、C_{2-6}烯基、C_{2-6}炔基、C_{3-7}碳环基、C_{6-10}芳基、5至10元杂芳基和3至10元杂环基的“$-NR_AR_B$”基团，如本文所定义的。非限制性示例包括游离氨基(即，$-NH_2$)。

[0095] “氨基烷基”基团是指经由亚烷基基团连接的氨基基团。

[0096] “烷氧基烷基”基团是指经由亚烷基基团连接的烷氧基基团，诸如“C_2-C_8烷氧基烷基”等。

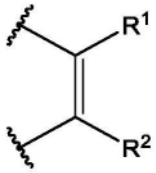
[0097] 当基团被描述为“任选地取代的”时，其可以是未取代的或取代的。同样，当基团被描述为“取代的”时，取代基可以选自所指示的取代基中的一者或多者。如本文所用，取代的基团衍生自未取代的母体基团，其中已存在一个或多个氢原子与另一个原子或基团的交换。除非另外指明，否则当基团被认为是“经取代的”时，这意味着该基团被一个或多个取代基取代，该取代基独立地选自：C_1-C_6烷基、C_1-C_6烯基、C_1-C_6炔基、C_3-C_7碳环基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、C_3-C_7碳环基-C_1-C_6-烷基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、3至10元杂环基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、3至10元杂环基-C_1-C_6-烷基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、芳基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、芳基(C_1-C_6)烷基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、5至10元杂芳基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、5至10元杂芳基(C_1-C_6)烷基(任选地被卤基、C_1-C_6烷基、C_1-C_6烷氧基、C_1-C_6卤代烷基和C_1-C_6卤代烷氧基取代)、卤基、-CN、羟基、C_1-C_6烷氧基、C_1-C_6烷氧基(C_1-C_6)烷基(即，醚)、芳氧基、硫氢基(巯基)、卤基(C_1-C_6)烷基(例如，$-CF_3$)、卤基(C_1-C_6)烷

氧基(例如,-OCF_3)、C_1-C_6烷硫基、芳硫基、氨基、氨基(C_1-C_6)烷基、硝基、O-氨甲酰基、N-氨甲酰基、O-硫代氨甲酰基、N-硫代氨甲酰基、C-酰氨基、N-酰氨基、S-亚磺酰氨基、N-亚磺酰氨基、C-羧基、O-羧基、酰基、氰酸根、异氰酸根、硫氰酸根、异硫氰酸根、亚磺酰基、磺酰基、-SO_3H、磺酸根、硫酸根、亚磺酸基、-OSO_2C_1-C_4烷基和氧代基(=O)。在基团被描述为"任选地取代的"的任何地方,该基团可以被上述取代基取代。在一些实施方案中,当烷基、烯基、炔基、芳基、杂芳基、碳环基或杂环基基团被取代时,它们各自独立地被选自由以下项组成的组的一个或多个取代基取代:卤基、-CN、-SO_3^-、-OSO_3^-、-SO_3H、-SR^A、-OR^A、-NR^BR^C、氧代基、-$CONR^BR^C$、-$SO_2NR^BR^C$、-COOH和-$COOR^B$,其中R^A、R^B和R^C各自独立地选自H、烷基、取代的烷基、烯基、取代的烯基、炔基、取代的炔基、芳基和取代的芳基。

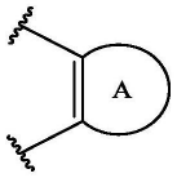
[0098] 如本领域的普通技术人员所理解的,本文所述的化合物可以以离子化形式存在,例如,-CO_2^-、-SO_3^-或-O-SO_3^-。如果化合物包含带正电荷或带负电荷的取代基基团,例如,-SO_3^-,则其还可以包含带负电荷或带正电荷的抗衡离子,使得该化合物整体为中性的。在其他方面,该化合物可以以盐形式存在,其中抗衡离子由共轭酸或碱提供。

[0099] 应当理解,根据上下文,某些基团命名惯例可以包括单基或二基。例如,当取代基需要两个与分子其余部分的连接点时,应当理解,该取代基为二基。例如,被识别为需要两个连接点的烷基的取代基包括二基,诸如-CH_2-、-CH_2CH_2-、-$CH_2CH(CH_3)CH_2$-等。其他基团命名惯例清楚地指示该基团为二基,诸如"亚烷基"或"亚烯基"。

[0100] 当两个"相邻的"R基团据称"连同它们所附接的原子一起"形成环时,这意味着原子、居间键和两个R基团的集合体为所述的环。例如,当存在以下子结构时:

[0101] 

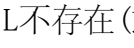
[0102] 并且R^1和R^2被限定为选自由氢和烷基组成的组,或者R^1和R^2连同它们所附接的原子一起形成芳基或碳环基,这意味着R^1和R^2可以选自氢或烷基,或者替代性地,该子结构具有结构:

[0103] 

[0104] 其中A为含有所描绘的双键的芳环或碳环基。

[0105] 当取代基被描述为二基(即,具有与分子的其余部分的两个连接点)时,应理解,除非另外指明,否则取代基可以任何定向构型连接。因此,例如,描绘为-AE-或

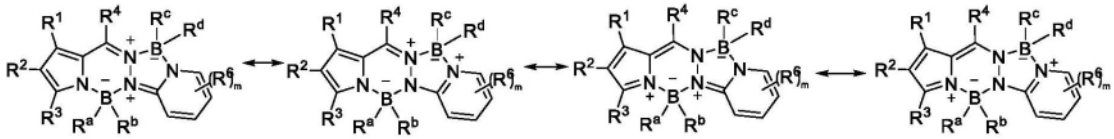
的取代基包括被定向成使得A连接在分子的最左侧连接点的取代基,以及其中A连接在所述分子的最右侧连接点处的取代基。另外,如果基团或取代基被描绘为

,则L定义为任选地存在的连接基部分;当L不存在(或没有)时,此类基团或取代基等同于

。

[0106] 本文所述的化合物可以表示为若干内消旋形式。在绘制单一结构的情况下,预期

存在任何相关的内消旋形式。本文所述的含双硼染料由单一结构表示,但同样可以被显示为任何相关的内消旋形式。示例性内消旋结构在下面针对式(Ia)示出:

[0107] 

[0108] 在示出本文所述化合物的单一内消旋形式的每种情况下,同样设想了替代性的内消旋形式。此外,为了简单起见,化合物的氮原子上的正电荷(当存在四个连接到氮原子的键时)和硼原子上的负电荷(当存在四个连接到硼原子的键时)可能不在某些化合物结构中显示。

[0109] 如本文所用,“核苷酸”包括含氮杂环碱基、糖以及一个或多个磷酸基团。它们是核酸序列的单体单元。在RNA中,糖为核糖,并且在DNA中为脱氧核糖,即,缺乏存在于核糖中的羟基基团的糖。含氮杂环碱基可为嘌呤、脱氮嘌呤或嘧啶碱基。嘌呤碱基包括腺嘌呤(A)和鸟嘌呤(G)以及它们的修饰衍生物或类似物,诸如7-脱氮腺嘌呤或7-脱氮鸟嘌呤。嘧啶碱基包括胞嘧啶(C)、胸腺嘧啶(T)和尿嘧啶(U)以及它们的经修饰的衍生物或类似物。脱氧核糖的C-1原子与嘧啶的N-1或嘌呤的N-9键合。

[0110] 如本文所用,“核苷”在结构上类似于核苷酸,但缺少磷酸部分。核苷类似物的一个示例是其中标记与碱基连接并且没有磷酸基团连接到糖分子的核苷类似物。术语“核苷”在本文中以本领域技术人员所理解的常规含义使用。示例包括但不限于包含核糖部分的核糖核苷和包含脱氧核糖部分的脱氧核糖核苷。经修饰的戊糖部分是其中氧原子已被碳取代和/或碳已被硫或氧原子取代的戊糖部分。“核苷”是可以具有取代的碱基和/或糖部分的单体。另外,核苷可以掺入较大的DNA和/或RNA聚合物和低聚物中。

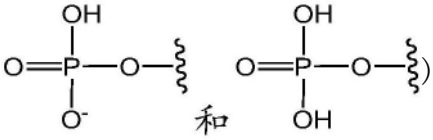
[0111] 术语“嘌呤碱基”在本文中以本领域技术人员所理解的普通含义使用,并且包括其互变异构体。类似地,术语“嘧啶碱基”在本文中以本领域技术人员所理解的普通含义使用,并且包括其互变异构体。任选地取代的嘌呤碱基的非限制性列表包括嘌呤、腺嘌呤、鸟嘌呤、脱氮嘌呤、7-脱氮腺嘌呤、7-脱氮鸟嘌呤、次黄嘌呤、黄嘌呤、别黄嘌呤、7-烷基鸟嘌呤(例如7-甲基鸟嘌呤)、可可碱、咖啡因、尿酸和异鸟嘌呤。嘧啶碱基的示例包括但不限于胞嘧啶、胸腺嘧啶、尿嘧啶、5,6-二氢尿嘧啶和5-烷基胞嘧啶(例如5-甲基胞嘧啶)。

[0112] 如本文所用,当寡核苷酸或多核苷酸被描述为“包含”本文所述的核苷或核苷酸时,这意味着本文所述的核苷或核苷酸与寡核苷酸或多核苷酸形成共价键。类似地,当核苷或核苷酸被描述为寡核苷酸或多核苷酸的一部分,诸如“掺入到”寡核苷酸或多核苷酸中时,这意味着本文所述的核苷或核苷酸与寡核苷酸或多核苷酸形成共价键。在一些此类实施方案中,共价键在寡核苷酸或多核苷酸的3’羟基与在本文中描述为寡核苷酸或多核苷酸的3’碳原子与核苷酸的5’碳原子之间的磷酸二酯键的核苷酸的5’磷酸基团之间形成。

[0113] 如本文所用,术语“可裂解的连接基”并不意在暗示需要除去整个连接基。裂解位点可以位于连接基上确保该连接基的一部分在裂解后保持连接到可检测标记和/或核苷或核苷酸部分的某个位置处。

[0114] 如本文所用,“衍生物”或“类似物”意指具有修饰的碱基部分和/或修饰的糖部分的合成核苷酸或核苷衍生物。此类衍生物和类似物在例如Scheit,Nucleotide Analogs(John Wiley&Son,1980)和Uhlman等人,Chemical Reviews 90:543-584,1990中进行了讨

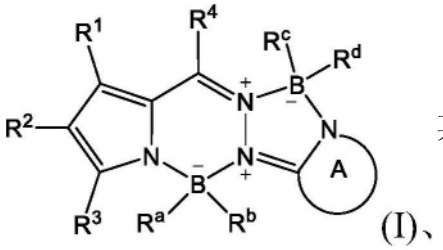
论。核苷酸类似物还可以包括经修饰的磷酸二酯键,包括硫代磷酸酯键、二硫代磷酸酯键、烷基膦酸酯键、苯胺磷酸酯键和氨基磷酸酯键。如本文所用,“衍生物”、“类似物”和“修饰的”可以互换使用,并且由本文定义的术语“核苷酸”和“核苷”涵盖。

[0115] 如本文所用,术语“磷酸酯”以本领域技术人员所理解的普通含义使用,并且包括其质子化形式(例如,和)。如本文所用,术语“单磷酸”、“二磷酸”和“三磷酸”以本领域技术人员所理解的其普通含义使用,并且包括质子化形式。

[0116] 如本文所用,术语“定相”是指SBS中的现象,该现象是由3’终止子和荧光团的不完全去除和/或未能通过在给定测序循环下通过聚合酶完成簇内DNA链的一部分掺入所引起的。预定相是由掺入不具有有效3’终止子的核苷酸引起的,其中掺入事件由于终止失败而提前1个周期。定相和预定相导致特定循环的测量信号强度由来自当前循环的信号以及来自前一个和后一个循环的噪声组成。随着循环的数量增加,受到定相和预定相影响的每个簇的序列分数增加,妨碍对正确碱基的识别。预定相可以由在通过合成测序(SBS)期间存在痕量的不受保护或未封端的3’-OH核苷酸引起。不受保护的3’-OH核苷酸可以在制造过程期间或可能在储存和试剂处理过程期间产生。因此,降低预定相发生率的核苷酸类似物的发现是令人惊讶的,并且相对于现有核苷酸类似物,在SBS应用中提供了巨大优势。例如,所提供的核苷酸类似物可以产生更快的SBS循环时间、较低的定相和预定相值以及较长的测序读段长度。

[0117] 含有式(I)的双硼稠合杂环的染料

[0118] 本公开的一些方面涉及式(I)的含双硼染料以及其盐和内消旋形式:

[0119] 其盐或内消旋形式,

(I)、

[0120] 其中R^1、R^2、R^3和R^4中的每一者独立地是H、未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、未取代或取代的氨基、卤基、氰基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基;

[0121] R^a、R^b、R^c和R^d中的每一者独立地是卤基、氰基、C_1-C_6烷基,C_1-C_6卤代烷基、C_1-C_6烷氧基、C_1-C_6卤代烷氧基、C_6-C_{10}芳基、C_6-C_{10}芳氧基或-O-C(=O)R^5;

[0122] 另选地,当R^a和R^b都是-O-C(=O)R^5时,两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基;当R^c和R^d都是-O-C(=O)R^5时,两个R^5连同它们所附接的原子一起形成未取代或取代的6元至10元杂环基;

[0123] R^5是未取代或取代的C_1-C_6烷基;

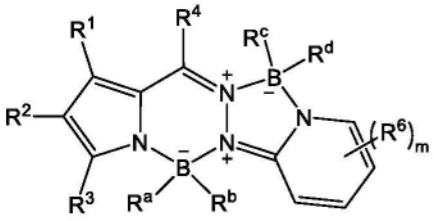
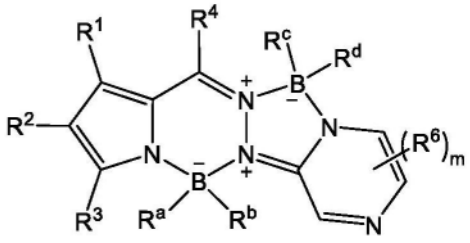
[0124] 环A是任选地被一个或多个R^6取代的六元、七元、八元、九元或十元杂芳基;

[0125] 每个R^6独立地是未取代或取代的C_1-C_6烷基、C_1-C_6烷氧基、C_2-C_6烯基、C_2-C_6炔基、C_1-C_6卤代烷基、C_1-C_6卤代烷氧基、C_1-C_6羟烷基、(C_1-C_6烷氧基)(C_1-C_6烷基)、-NR^7R^8、卤基、氰基、羧基、羟基、硝基、磺酰基、亚磺酸基、磺基、磺酸根、S-亚磺酰氨基、N-亚磺酰氨基、未取代或取代的C_3-C_{10}碳环基、未取代或取代的C_6-C_{10}芳基、未取代或取代的5元至10元杂芳基或未取代或取代的3元至10元杂环基;

[0126] R^7和R^8中的每一者独立地是H、未取代或取代的C_1-C_6烷基,或者R^7和R^8连同它们所附接的氮原子一起形成未取代或取代的3元至10元杂环基;

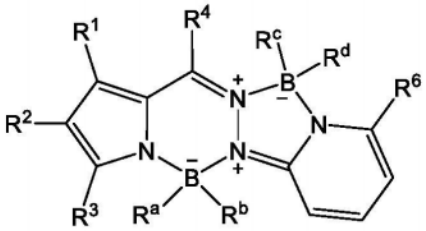
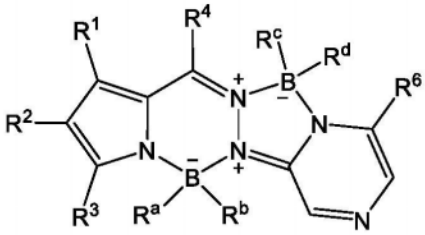
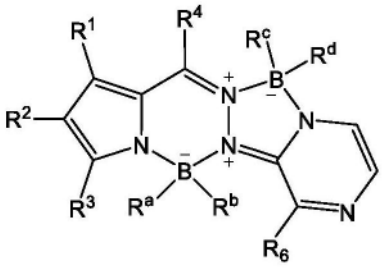
[0127] 前提条件是当R^a、R^b、R^c和R^d中的每一者是氟时,R^1、R^2、R^3、R^4和环A中的至少一者包含羧基基团。在一些另外的实施方案中,R^1、R^2、R^3、R^4和环A中的至少一者包含羧基基团。在一些另外的实施方案中,R^1、R^2、R^3、R^4和环A中的一者包含羧基基团。

[0128] 在式(I)的化合物的一些实施方案中,环A是任选地被一个或多个R^6取代的六元杂芳基。在一些实施方案中,六元杂芳基包含一个或两个氮原子。在另外的实施方案中,六元杂芳基是吡啶基、嘧啶基或吡嗪基。在另外的实施方案中,式(I)的化合物也由式(Ia)或(Ib)表示:

[0129]  (Ia)、 (Ib),

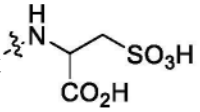
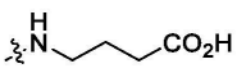
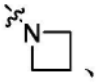
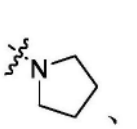
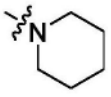
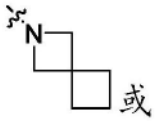
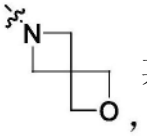
或其盐或内消旋形式,其中m是0、1、2或3。

[0130] 在另外的实施方案中,式(Ia)的化合物也由式(Ic)表示,并且式(Ib)的化合物也由式(Id)或(Ie)表示:

[0131]  (Ic)、 (Id)、 (Ie),

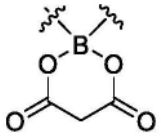
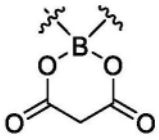
或其盐或内消旋形式。

[0132] 在式(I)和(Ia)-(Ie)的化合物的一些实施方案中,每个R^6独立地是卤基、氰基、羧基、未取代或取代的C_1-C_6烷基、未取代的苯基、被羧基取代的苯基、未取代的5元杂芳基、被羧基取代的5元杂芳基,或-NR^7R^8。在一些实施方案中,R^6是卤基(例如,氟、氯或溴)。在一些实施方案中,R^6是未取代的呋喃或被羧基取代的呋喃。在一些另外的实施方案中,R^6是取代的C_1-C_6烷基(例如,甲基、乙基、正丙基、异丙基、正丁基、异丁基、正戊基、异戊基、正己基或

异己基),其独立地被一个或多个选自由卤基、-CN、$-SO_3^-$、$-SO_3H$、$-NH_2$、-NH(C_1-C_6烷基)、-N(C_1-C_6烷基)$_2$、-C(=O)OH和-C(=O)O(C_1-C_6烷基)组成的组的取代基取代。在一些另外的实施方案中,R^6是$-NR^7R^8$,其中R^7是H并且R^8是被一个或多个选自由羧基、磺基和磺酸根组成的组的取代基取代的C_1-C_6烷基,或者R^7和R^8连同它们所附接的氮原子一起形成被羧基取代的3元至10元杂环基(例如,包含氮原子,或一个氮和第二杂原子(诸如氧或硫)的4元、5元、6元或7元杂环基)。在一些另外的实施方案中,R^6是、、、、、或,其中环结构中的每一者任选地被羧基取代。

[0133] 在式(I)和(Ia)-(Ie)的化合物的一些实施方案中,R^1、R^2和R^3中的每一者是H。在一些其他实施方案中,R^1、R^2和R^3中的每一者独立地是未取代的C_1-C_6烷基(例如,甲基、乙基、正丙基、异丙基、正丁基、异丁基、正戊基、异戊基、正己基或异己基)。在一个实施方案中,R^1和R^3中的每一者是甲基并且R^2是乙基。在一些其他实施方案中,R^1、R^2和R^3中的两者是H或未取代的C_1-C_6烷基,并且R^1、R^2和R^3中的一者是卤基、苯基、5元或6元杂芳基、羧基或被羧基取代的C_1-C_6烷基。在另外的实施方案中,R^1和R^3中的每一者是甲基,并且R^2是溴、氯、氟、苯基、羧基或$-CH_2$-COOH。

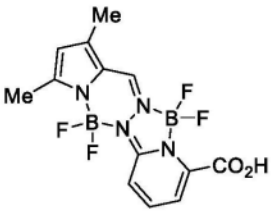
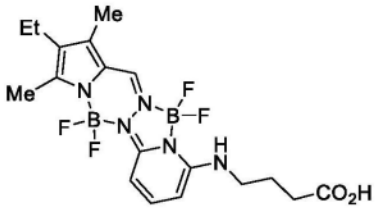
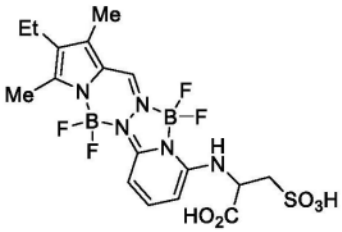
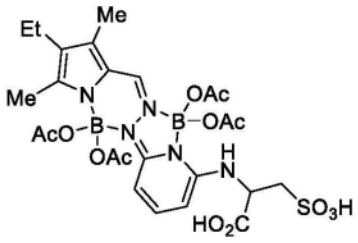
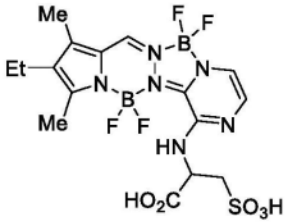
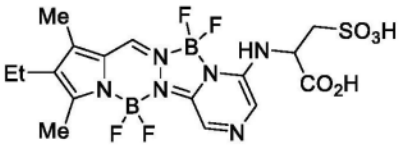
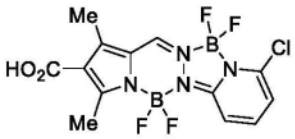
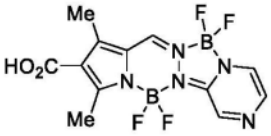
[0134] 在式(I)和(Ia)-(Ie)的化合物的一些实施方案中,R^4是H或未取代的C_1-C_6烷基。在一些其他实施方案中,R^4是各自被羧基取代的C_1-C_6烷基或苯基。

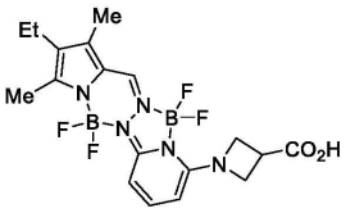
[0135] 在式(I)和(Ia)-(Ie)的化合物的一些实施方案中,R^a和R^b中的每一者独立地是氟、氰基、甲基、三氟甲基、甲氧基、苯基、苯氧基或-O-酰基(-OC(=O)CH_3或-OAc)。在另外的实施方案中,R^a和R^b都是氟、甲基、三氟甲基、甲氧基或-O-酰基。在其他实施方案中,R^a和R^b都是$-OC(=O)R^5$,并且两个R^5连同它们所附接的原子一起形成具有结构的6元杂环基,其中该结构的亚甲基部分可以任选地被一个或两个选自氟、甲基、三氟甲基、甲氧基、苯基或苯氧基的取代基取代。在一些实施方案中,R^c和R^d中的每一者独立地是氟、氰基、甲基、三氟甲基、甲氧基、苯基、苯氧基或-O-酰基。在另外的实施方案中,R^c和R^d都是氟、甲基、三氟甲基、甲氧基或-O-酰基。在其他实施方案中,R^c和R^d都是$-OC(=O)R^5$,并且两个R^5连同它们所附接的原子一起形成具有结构的6元杂环基,其中该结构的亚甲基部分可以任选地被一个或两个选自氟、甲基、三氟甲基、甲氧基、苯基或苯氧基的取代基取代。

[0136] 在式(I)和(Ia)-(Ie)的化合物的任何实施方案中,当C_3-C_{10}碳环基(例如,C_3-C_{10}环烷基)、C_6-C_{10}芳基、5元至10元杂芳基或3元至10元杂环基被取代时,其可以被一个或多个R^6取代。当基团被定义为取代的C_1-C_6烷基时,其可以是被羧基、羧酸根、磺基、磺酸根、-C(O)O(C_1-C_6烷基)或$-C(O)NR^eR^f$取代的C_1、C_2、C_3、C_4、C_5或C_6烷基(例如,甲基、乙基、正丙基、异丙基、正丁基、异丁基、正戊基、异戊基、正己基或异己基),并且其中每个R^e和R^f独立地是H或被

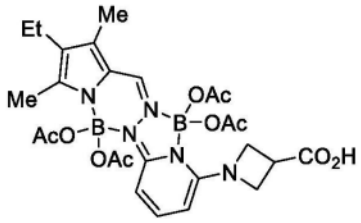
羧基、羧酸根、磺基或磺酸根取代的C_1-C_6烷基。

[0137] 式(I)的化合物的另外的实施方案包括但不限于：

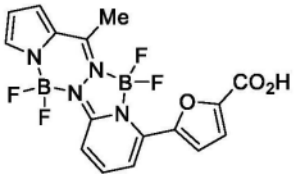
[0138]  (I-1) 、  (I-2) 、  (I-3) 、  (I-4) 、  (I-5) 、  (I-6) 、  (I-7) 、  (I-8) 、



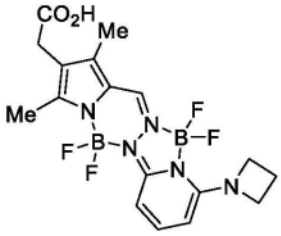
(I-9) 、



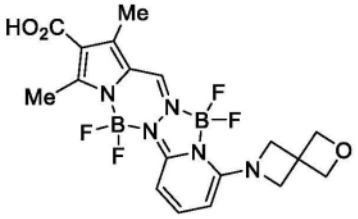
(I-10) 、



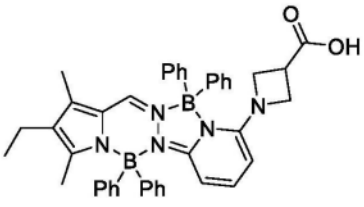
(I-11) 、



(I-12) 、

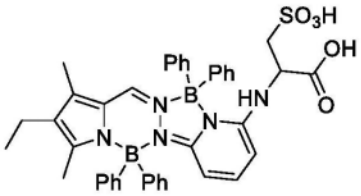


(I-13) 、

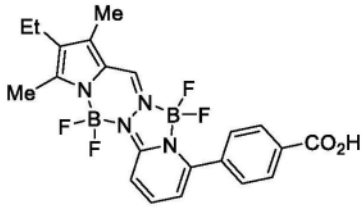


(I-14) 、

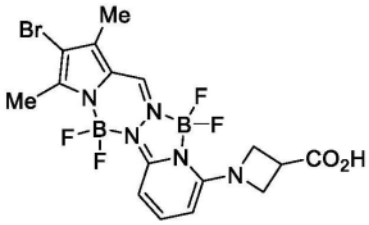
[0139]



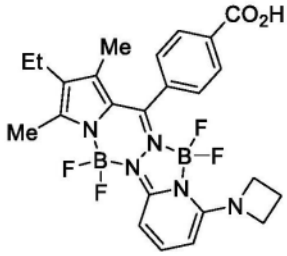
(I-15) 、



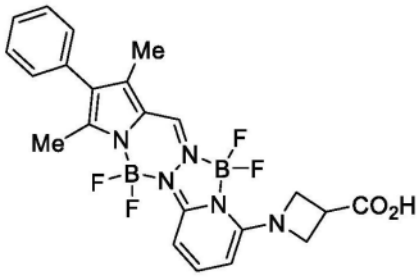
(I-16) 、



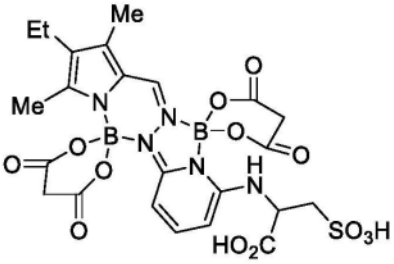
(I-17) 、



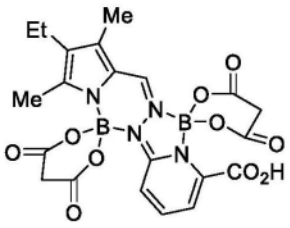
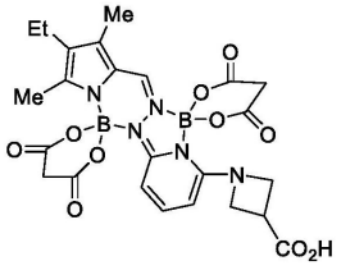
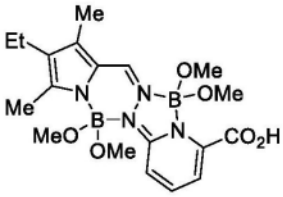
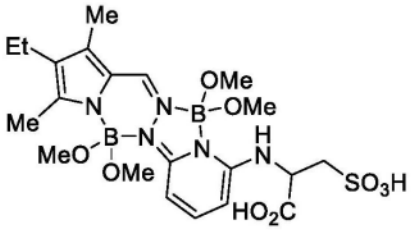
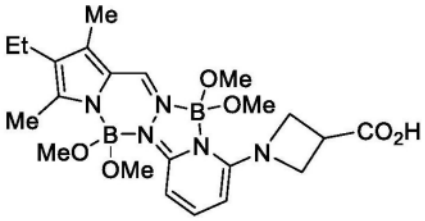
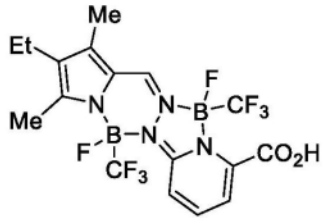
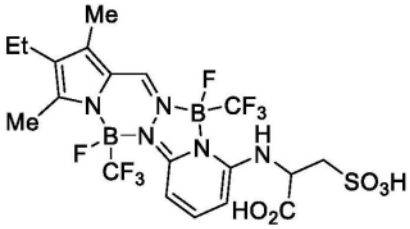
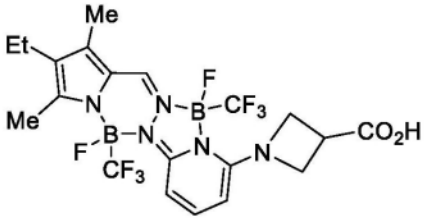
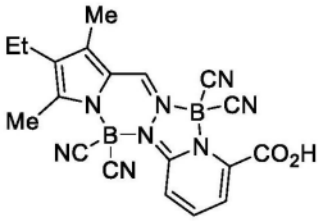
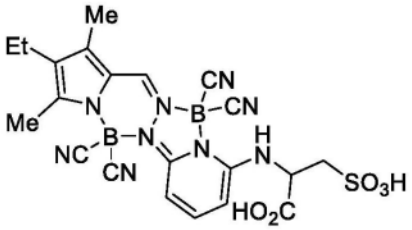
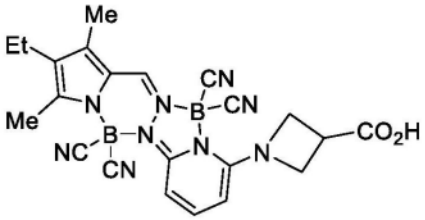
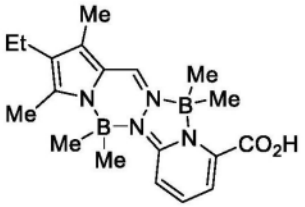
(I-18) 、

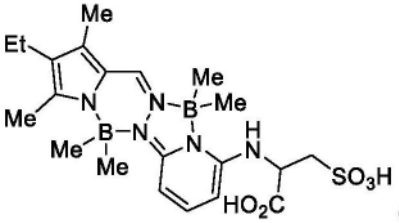
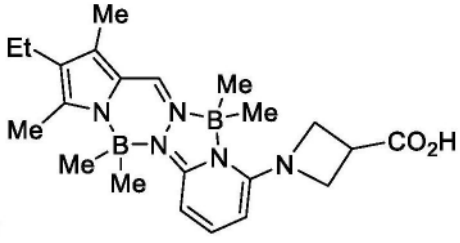


(I-19) 、



(I-20) 、


(I-21)、

(I-22)、

(I-23)、

(I-24)、

(I-25)、

(I-26)、

(I-27)、

(I-28)、

(I-29)、

(I-30)、

(I-31)、

(I-32)、

[0141]

(I-33)和

(I-34),

以及其盐和内消旋体形式。非限制性示例为对应的C_1-C_6烷基羧酸酯(如由化合物的羧基基团形成的甲酯、乙酯、异丙酯和叔丁酯)。

[0142] 在本文所述的双硼稠合杂环化合物的任何实施方案中,可以进一步修饰该化合物以引入与其共价键合的光保护部分,例如环辛四烯部分。

[0143] 标记的核苷酸

[0144] 根据本公开的一个方面,本文所述的染料化合物适用于附接到底物部分,特别是包含连接基基团以使得能够附接到底物部分的染料化合物。底物部分事实上可以是本公开的染料可以与之缀合的任何分子或物质,并且以非限制性示例的方式,可以包括核苷、核苷酸、多核苷酸、碳水化合物、配体、颗粒、固体表面、有机聚合物和无机聚合物、染色体、核、活细胞,以及它们的组合或装配物。染料可以通过任选的连接基通过多种方式(包括疏水吸引、离子吸引和共价连接)缀合。在一些方面,染料通过共价附接缀合至底物。更具体地讲,共价连接借助于连接基基团实现。在一些情况下,此类标记的核苷酸也被称为"经修饰的核苷酸"。

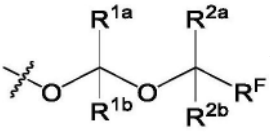
[0145] 本公开的一些方面涉及用如本文所述的式(I)(包括(Ia)-(Ie))的染料或其内消旋形式的盐,或含有本文所述的光保护部分COT的其衍生物标记的核苷酸。标记的核苷酸或寡核苷酸可以经由羧基(-CO_2H)或烷基-羧基基团附接到本文所公开的染料化合物以形成酰胺或烷基-酰胺键。在一些另外的实施方案中,羧基基团可以呈羧基基团的活化形式的形式,例如,酰胺或酯,这样的形式可以用于附接到核苷酸或寡核苷酸的氨基或羟基基团。如本文所用,术语"活化酯"是指能够在温和条件下反应例如与含有氨基基团的化合物反应的羧基基团衍生物。活化酯的非限制性示例包括但不限于对硝基苯基、五氟苯基和琥珀酰亚胺酯。

[0146] 例如,式(I)(包括(Ia)-(Ie))的染料化合物可以经由式(I)的R^1、R^2、R^3、R^4和环A中的一者附接到核苷酸。在一些此类实施方案中,式(I)的R^1、R^2、R^3、R^4和环A中的一者包含羧基基团,并且该附接在式(I)的化合物的羧基官能团与核苷酸或核苷酸连接基的氨基官能团之间形成酰胺部分。

[0147] 在一些实施方案中,染料化合物可以经由核苷酸碱基共价附接到核苷酸。在一些此类实施方案中,标记的核苷酸可以具有任选地通过连接基部分连接到嘧啶碱基的C5位置或7-脱氮嘌呤碱基的C7位置的染料。例如,核碱基可以是7-脱氮腺嘌呤,并且染料任选地通过连接基在C7位置处附接到7-脱氮腺嘌呤。核碱基可以是7-脱氮鸟嘌呤,并且染料任选地通过连接基在C7位置处附接到7-脱氮鸟嘌呤。核碱基可以是胞嘧啶,并且染料任选地通过连接基在C5位置处附接到胞嘧啶。作为另一示例,核碱基可以是胸腺嘧啶或尿嘧啶,并且染料任选地通过连接基连接到胸腺嘧啶或尿嘧啶的C5位置处。

[0148] 3'封端基团

[0149] 标记的核苷酸或寡核苷酸还可以具有共价连接到核苷酸的核糖或脱氧核糖的封端基团。该封端基团可以连接在核糖或脱氧核糖上的任何位置处。在特定实施方案中,该封端基团位于核苷酸的核糖或脱氧核糖的3′位置处。各种3'封端基团公开于WO2004/018497和WO2014/139596中,这些申请特此以引用方式并入。例如,封端基团可以是叠氮甲基(-CH_2N_3)或取代的叠氮甲基(例如,-$CH(CHF_2)N_3$或$CH(CH_2F)N_3$)、或连接至核糖或脱氧核糖部分的3'氧原子的烯丙基。在一些实施方案中,3'封端基团为叠氮甲基,与核糖或脱氧核糖的3'碳形成3′-OCH_2N_3。

[0150] 在一些其他实施方案中,3'封端基团和3'氧原子形成共价附接到核糖或脱氧核糖的3'碳的结构的缩醛基团,其中:

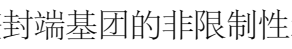
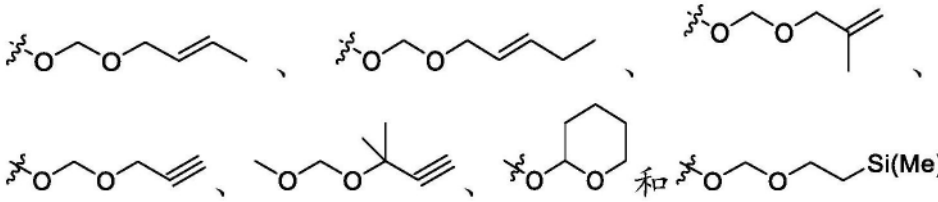
[0151] R^{1a}和R^{1b}各种独立地为H、C_1-C_6烷基、C_1-C_6卤代烷基、C_1-C_6烷氧基、C_1-C_6卤代烷氧基、氰基、卤素、任选地取代的苯基或任选地取代的芳烷基；

[0152] R^{2a}和R^{2b}各种独立地为H、C_1-C_6烷基、C_1-C_6卤代烷基、氰基或卤素；

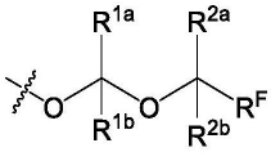
[0153] 替代地，R^{1a}和R^{2a}与其所连接的原子一起形成任选地取代的五至八元杂环基基团；

[0154] R^F为H、任选地取代的C_2-C_6烯基、任选地取代的C_3-C_7环烯基、任选地取代的C_2-C_6炔基或任选地取代的(C_1-C_6亚烷基)Si(R^{3a})$_3$；并且

[0155] 每个R^{3a}独立地为H、C_1-C_6烷基或任选地取代的C_6-C_{10}芳基。

[0156] 附加的3′羟基封端基团公开于美国公布2020/0216891A1中，该文献全文以引用方式并入。缩醛封端基团的非限制性示例为(AOM)、各自共价连接到核糖或脱氧核糖的3'碳。

[0157] 3'封端基团的脱保护

[0158] 在一些实施方案中，可以通过使用水溶性膦试剂去除或脱保护叠氮甲基3'羟基保护基团。非限制性示例包括三(羟甲基)膦(THMP)、三(羟乙基)膦(THEP)或三(羟丙基)膦(THP或THPP)。本文所述的3'封端基团可以在各种化学条件下去除或裂解。对于含有乙烯基或烯基部分的缩醛封端基团非限制性裂解条件包括在膦配体，例如三(羟甲基)膦(THMP)或三(羟丙基)膦(THP或THPP)的存在下的Pd(II)络合物，诸如Pd(OAc)$_2$或烯丙基氯化Pd(II)二聚体。对于含有炔基基团(例如，乙炔基)的那些封端基团，其也可以通过在膦配体(例如，THP或THMP)的存在下的Pd(II)络合物(例如，Na_2PdCl_4、K_2PdCl_4、Pd(OAc)$_2$或烯丙基氯化Pd(II)二聚体)去除。

[0159] 钯裂解试剂

[0160] 在一些实施方案中，本文所述的3'封端基团可以被钯催化剂裂解。在一些此类实施方案中，Pd催化剂是水溶性的。在一些此类实施方案中，为Pd(0)络合物(例如，三(3,3′,3″-次膦基三(苯磺酸基)钯(0)九钠盐九水合物)。在一些情况下，Pd(0)络合物可以通过试剂诸如烯烃、醇、胺、膦或金属氢化物还原Pd(II)络合物而原位生成。合适的钯源包括Na_2PdCl_4、K_2PdCl_4、Pd(CH_3CN)$_2Cl_2$、(PdCl(C_3H_5))$_2$、[Pd(C_3H_5)(THP)]Cl、[Pd(C_3H_5)(THP)$_2$]Cl、Pd(OAc)$_2$、Pd(Ph_3)$_4$、Pd(dba)$_2$、Pd(Acac)$_2$、$PdCl_2$(COD)和Pd(TFA)$_2$。在一个此类实施方案中，Pd(0)络合物由Na_2PdCl_4原位产生。在另一个实施方案中，钯源是烯丙基氯化钯(II)二聚体[(PdCl(C_3H_5))$_2$]。在一些实施方案中，Pd(0)络合物通过将Pd(II)络合物与膦混合而在水溶液中产生。合适的膦包括水溶性膦，诸如三(羟丙基)膦(THP)、三(羟甲基)膦(THMP)、1,3,5-三氮杂-7-磷杂金刚烷(PTA)、双(对磺酸基苯基)苯基膦二水合物钾盐、三(羧乙基)膦(TCEP)和三苯基膦-3,3',3'-三磺酸三钠盐。

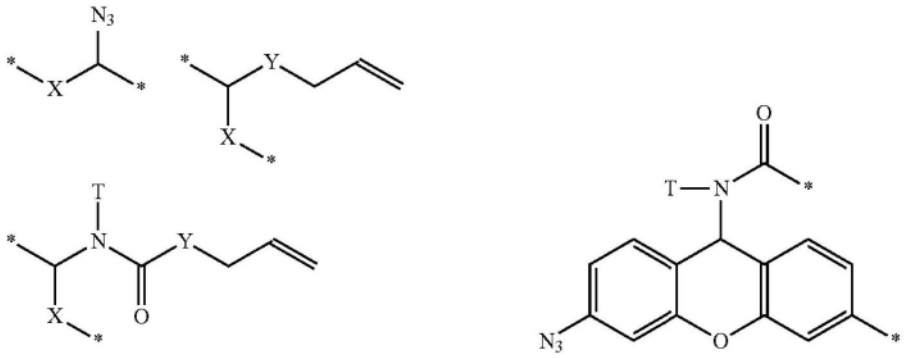
[0161] 在一些实施例中,Pd(0)通过将Pd(II)络合物[(PdCl(C_3H_5))$_2$]与THP原位混合来制备。Pd(II)络合物和THP的摩尔比可以为约1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9或1:10。在一些另外的实施方案中,可以添加一种或多种还原剂,诸如抗坏血酸或其盐(例如,抗坏血酸钠)。在一些实施方案中,裂解混合物可以含有另外的缓冲液试剂,诸如伯胺、仲胺、叔胺、碳酸盐、磷酸盐或硼酸盐,或它们的组合。在一些另外的实施方案中,缓冲液试剂包括乙醇胺(EA)、三(羟甲基)氨基甲烷(Tris)、甘氨酸、碳酸钠、磷酸钠、硼酸钠、2-二甲基乙醇胺(DMEA)、2-二乙基乙醇胺(DEEA)、N,N,N′,N′-四甲基乙二胺(TEMED)或N,N,N′,N′-四乙基乙二胺(TEEDA)或它们的组合。在一个实施方案中,缓冲剂是DEEA。在另一个实施方案中,缓冲剂含有一种或多种无机盐,诸如碳酸盐、磷酸盐或硼酸盐,或它们的组合。在一个实施方案中,无机盐为钠盐。

[0162] 连接基

[0163] 如本文所公开的染料化合物可以在取代基位置之一处包括反应性连接基基团,用于将该化合物共价连接至底物或另一分子。反应性连接基团是能够形成键(例如,共价键或非共价键)、特别是共价键的部分。在一个特定实施方案中,连接基可以是可裂解的连接基。使用术语"可裂解的连接基"并不意在暗示需要除去整个连接基。裂解位点可以位于连接基上确保该连接基的一部分在裂解后保持连接到染料和/或底物部分的某个位置处。以非限制性示例的方式,可裂解的连接基可以是亲电可裂解的连接基、亲核可裂解的连接基、可光裂解的连接基、在还原条件下可裂解的(例如含有二硫化物或叠氮化物的连接基)、在氧化条件下可裂解的、通过使用安全捕获连接基可裂解的,以及通过消除机制可裂解的。使用可裂解的连接基将染料化合物连接到底物部分确保了如果需要的话,可以在检测后移除标记,从而避免下游步骤中的任何干扰信号。

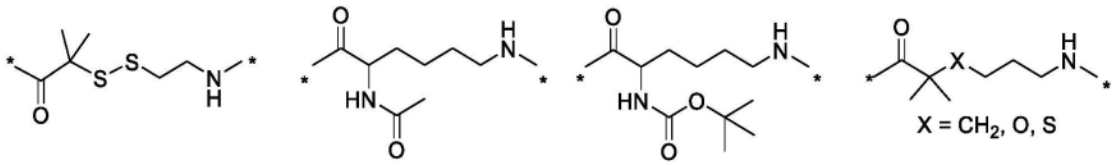
[0164] 可用的连接基基团可以见于PCT公开WO2004/018493(以引用方式并入本文)中,其示例包括可以使用水溶性膦或由过渡金属和至少部分水溶性的配体形成的水溶性过渡金属催化剂裂解的连接基。在水性溶液中,后者形成至少部分水溶性的过渡金属络合物。此类可裂解的连接基可以用于将核苷酸的碱基连接到标记,诸如本文所示的染料。

[0165] 特定的连接基包括PCT公开WO2004/018493(以引用方式并入本文)中所公开的那些,诸如包括下式的部分的那些:

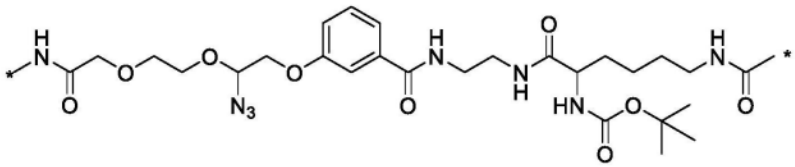
[0166] 

[0167] (其中X选自包括O、S、NH和NQ的组,其中Q为C1-10取代或未取代的烷基基团,Y选自包括O、S、NH和N(烯丙基)的组,T为氢或C_1-C_{10}取代或未取代的烷基基团,并且*指示所述部分与核苷酸或核苷的其余部分连接的位置)。在一些方面,连接基将核苷酸的碱基连接到标记,诸如,本文所述的染料化合物。

[0168] 连接基的附加示例包括美国公开2016/0040225(以引用方式并入本文)中所公开的那些,诸如包括下式的部分的那些:

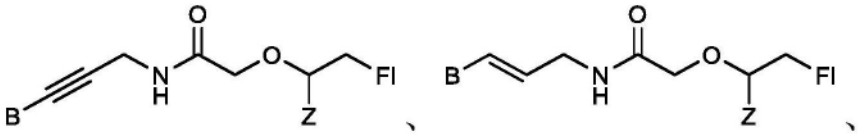


[0169]

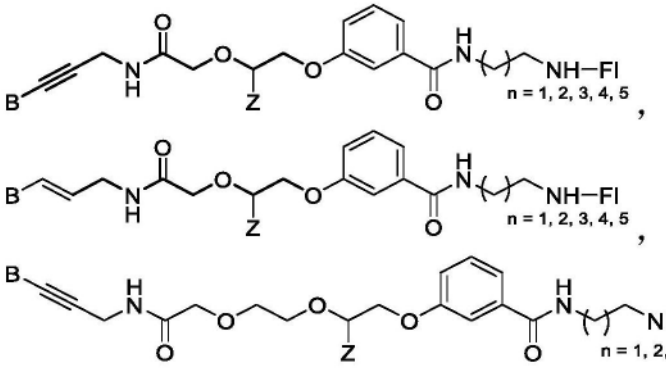


[0170] (其中*指示所述部分与核苷酸或核苷的其余部分连接的位置)。本文所展示的连接基部分可以包括核苷酸/核苷与标记之间的全部或部分连接基结构。本文所展示的连接基部分可以包括核苷酸/核苷与标记之间的全部或部分连接基结构。

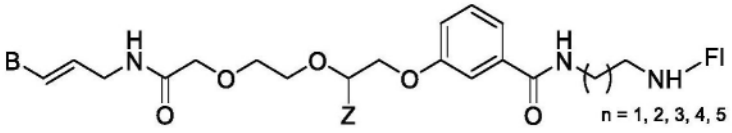
[0171] 连接基的附加示例包括下式的部分:

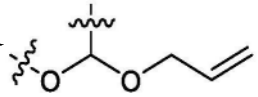


[0172]



[0173]



其中B是核碱基;Z是-N_3(叠氮基)、-O-C_1-C_6烷基、-O-C_2-C_6烯基或-O-C_2-C_6炔基;并且Fl包括染料部分,所述染料部分可以含有另外的连接基结构。本领域普通技术人员理解,本文所述的染料化合物通过使染料化合物的官能团(例如,羧基)与连接基的官能团(例如,氨基)反应而共价键合到连接基。在一个实施方案中,可裂解连接基包含  ("AOL"连接基部分),其中Z为-O-烯丙基。

[0174] 在特定实施方案中,荧光染料(荧光团)和鸟嘌呤碱基之间的连接基的长度可以改变,例如,通过引入聚乙二醇间隔区基团,从而与通过本领域已知的其他键附接到鸟嘌呤碱基的相同荧光团相比增加了荧光强度。示例性连接基及其特性在PCT公开WO2007020457(以引用方式并入本文)中示出。连接基的设计,尤其是其增加的长度,可以允许当掺入到多核苷酸(诸如DNA)中时,改善连接到鸟苷核苷酸的鸟嘌呤碱基的荧光团的亮度。因此,当染料用于需要检测附接到含有鸟嘌呤的核苷酸的荧光染料标记的任何分析方法中时,有利的是

连接基包含式-((CH$_2$)$_2$O)$_n$-的间隔区基团,其中n为介于2和50之间的整数,如WO 2007/020457中所述。

[0175] 核苷和核苷酸可以在糖或核碱基上的位点处标记。如本领域所知的,“核苷酸”由含氮碱基、糖和一个或多个磷酸基团组成。在RNA中,该糖是核糖,而在DNA中是脱氧核糖,即,缺乏存在于核糖中的羟基基团的糖。含氮碱基是嘌呤或嘧啶的衍生物。嘌呤为腺嘌呤(A)和鸟嘌呤(G),并且嘧啶为胞嘧啶(C)和胸腺嘧啶(T),或在RNA的情况下为尿嘧啶(U)。脱氧核糖的C-1原子与嘧啶的N-1或嘌呤的N-9键合。核苷酸也是核苷的磷酸酯,其中酯化发生在附接到糖的C-3或C-5的羟基基团上。核苷酸通常是单磷酸、二磷酸或三磷酸。

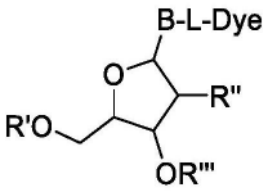
[0176] “核苷”在结构上类似于核苷酸,但缺少磷酸部分。核苷类似物的一个示例是其中标记与碱基连接并且没有磷酸基团连接到糖分子的核苷类似物。

[0177] 虽然碱基通常被称为嘌呤或嘧啶,但是技术人员将理解,可获得不改变核苷酸或核苷经历Watson-Crick碱基配对的能力的衍生物和类似物。“衍生物”或“类似物”意味着这样的化合物或分子:其核心结构与母体化合物的核心结构相同或非常类似,但其具有允许衍生的核苷酸或核苷与另一分子连接的化学或物理修饰,诸如不同的或附加的侧基。例如,碱基可以为脱氮嘌呤。在特定实施方案中,衍生物应当能够经历Watson-Crick配对。“衍生物”和“类似物”还包括例如具有经修饰的碱基部分和/或经修饰的糖部分的合成核苷酸或核苷衍生物。此类衍生物和类似物在例如Scheit,Nucleotide analogs(John Wiley&Son,1980)和Uhlman等人,Chemical Reviews 90:543-584,1990中进行了讨论。核苷酸类似物还可以包括经修饰的磷酸二酯键,包括硫代磷酸酯键、二硫代磷酸酯键、烷基膦酸酯键、苯胺磷酸酯键、氨基磷酸酯键等。

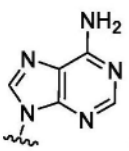
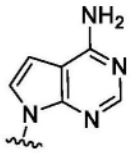
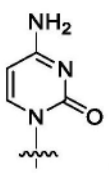
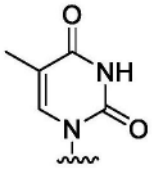
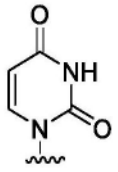
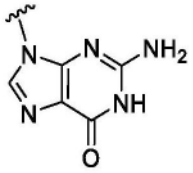
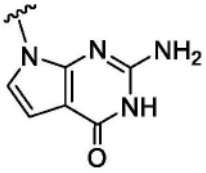
[0178] 染料可以例如通过连接基附接到核苷酸碱基上的任何位置。在特定实施方案中,仍然可以对所得的类似物进行Watson-Crick碱基配对。特定的核碱基标记位点包括嘧啶碱基的C5位置或7-脱氮嘌呤碱基的C7位置。如上所述,连接基基团可以用于将染料共价连接到核苷或核苷酸。

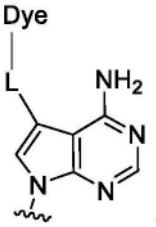
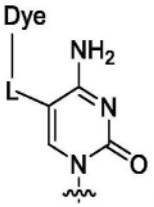
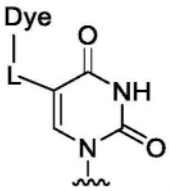
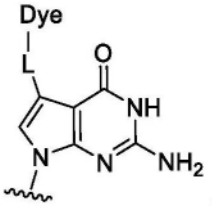
[0179] 在特定实施方案中,标记的核苷酸或寡核苷酸可以是可酶促掺入的和可酶促延伸的。因此,连接基部分可以具有足够的长度以将核苷酸连接到化合物,使得该化合物不显著干扰核酸复制酶对核苷酸的总体结合与识别。因此,连接基还可以包含间隔区单元。间隔区距离为例如核苷酸碱基距裂解位点或标记的距离。

[0180] 用本文所述的染料标记的核苷或核苷酸可以具有下式:

[0181] 

[0182] 其中染料是含有本文所述的稠合双硼杂环(标记)部分的染料(在染料的官能团与连接基“L”的官能团之间共价键合之后);B为核碱基,诸如尿嘧啶、胸腺嘧啶、胞嘧啶、腺嘌呤、7-脱氮腺嘌呤、鸟嘌呤、7-脱氮鸟嘌呤等;L为可以存在或可以不存在的任选的连接基;R'可以为H,或-OR'为单磷酸根、二磷酸根、三磷酸根、硫代磷酸根、磷酸酯类似物、连接至反应性含磷基团的-O-,或者被封端基团保护的-O-;R”是H或OH;并且R”'是H、本文所述的3'羟基封端基团,或-OR”'形成亚磷酰胺。其中-OR”'是亚磷酰胺,R'是酸可裂解的羟基保护基团,其允许随后在

自动化合成条件下进行单体偶联。在一些另外的实施方案中,B包括  、  、  、  、  、  或  ,或它们的任选地取代的衍生物和类似物。在一些另外的实施方案中,标记的核碱基包括结构

 、  、  或  。

[0183] 在一个特定实施方案中,封端基团是与染料化合物分开的并且独立于染料化合物,即,不与后者连接。替代性地,染料可以包含3'-OH封端基团的全部或部分。因此,R"'可以为可以构成或可以不构成染料化合物的3'羟基封端基团。

[0184] 在又一个替代性的实施方案中,戊糖的3'碳上不存在封端基团,并且连接到碱基的染料(或染料和连接基构造)例如可以具有足以充当掺入另外的核苷酸的障碍的尺寸或结构。因此,该障碍可以归因于空间位阻或可以归因于尺寸、电荷和结构的组合,无论染料是否附接到糖的3'位置。

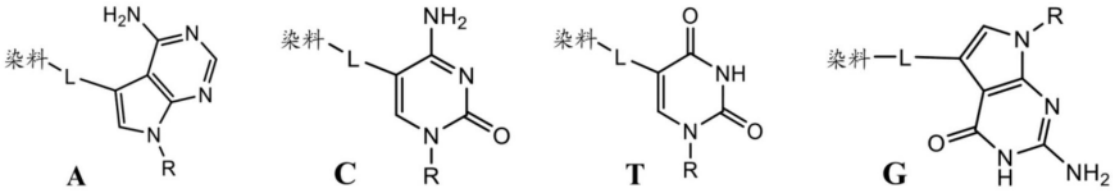
[0185] 在又另一个替代性实施方案中,该封端基团存在于戊糖的2'或4'碳上,并且可以具有足以充当掺入另外的核苷酸的障碍的尺寸或结构。

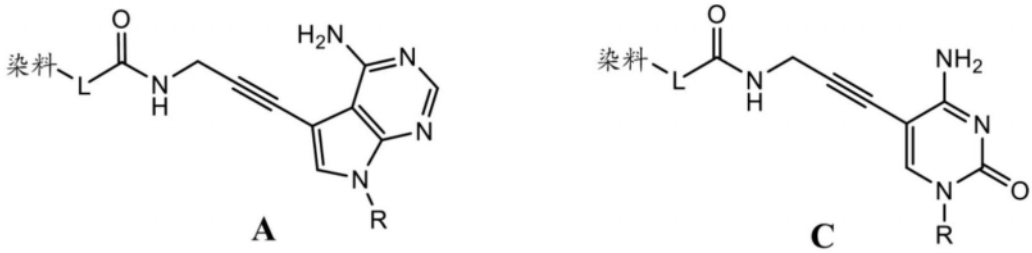
[0186] 使用封端基团允许对聚合进行控制,诸如当掺入标记的核苷酸时通过停止延伸来控制。如果封闭效应是可逆的,例如,以非限制性示例的方式,则通过改变化学条件或通过除去化学障碍,可以在某些点处停止延伸,然后允许其继续。

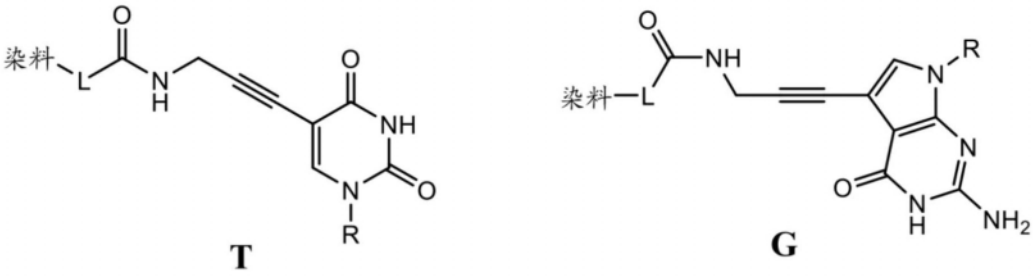
[0187] 在一个特定实施方案中,连接基(在染料和核苷酸之间)和封端基团两者均存在并且为单独的部分。在特定实施方案中,连接基和封端基团在相同或基本上类似的条件下都是可裂解的。因此,脱保护和脱封闭过程可能更有效,因为只需要单一处理就可以除去染料化合物和封闭基团这两者。然而,在一些实施方案中,连接基和封端基团不需要可在类似条件下裂解,而是可在不同条件下单独裂解。

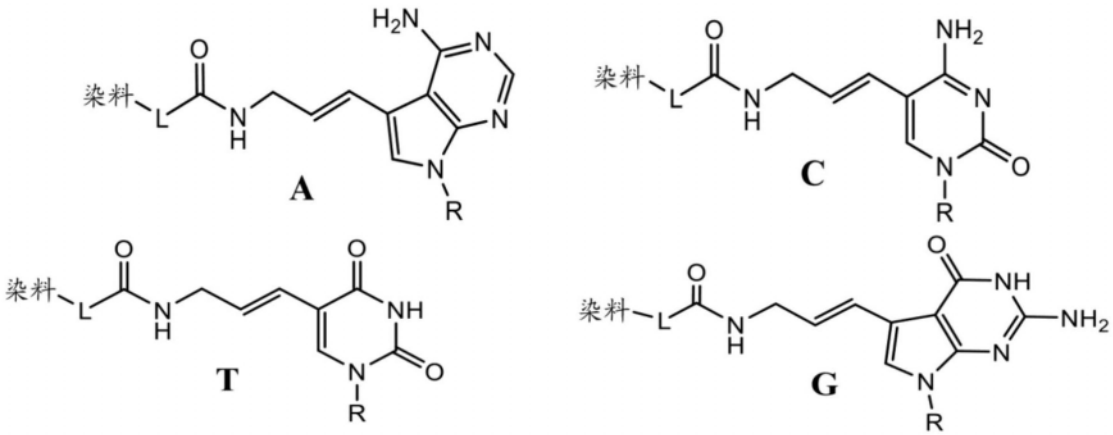
[0188] 本公开还涵盖掺入了染料化合物的多核苷酸。此类多核苷酸可以为分别由以磷酸二酯键接合的脱氧核糖核苷酸或核糖核苷酸组成的DNA或RNA。多核苷酸可以包含与如本文示出的至少一种经修饰的核苷酸(例如,用染料化合物标记)组合的天然存在的核苷酸、除本文所述的标记的核苷酸之外的非天然存在的(或经修饰的)核苷酸或它们的任何组合。根据本公开的多核苷酸还可以包括非天然的骨架键和/或非核苷酸化学修饰。还设想了由包含至少一种标记的核苷酸的核糖核苷酸和脱氧核糖核苷酸的混合物组成的嵌合结构。

[0189] 如本文所述的非限制性示例性标记的核苷酸包括:

[0190] 

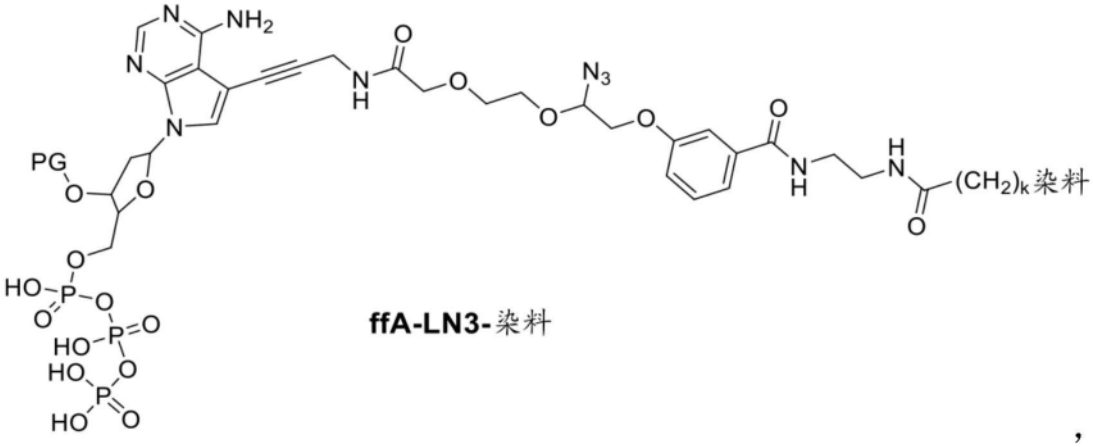


[0191] 



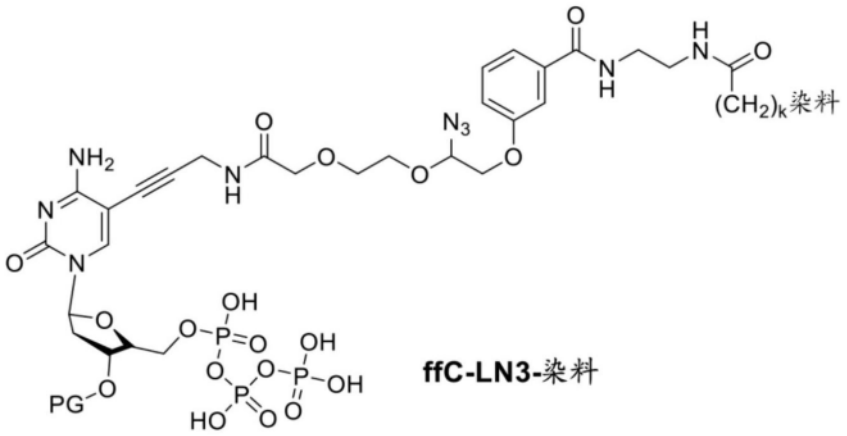
[0192] 其中L表示连接基,并且R表示如上所述的核糖或脱氧核糖部分,或具有被单磷酸、二磷酸或三磷酸取代的5'位置的核糖或脱氧核糖部分。

[0193] 在一些实施方案中,非限制性示例性荧光染料缀合物在下文示出:

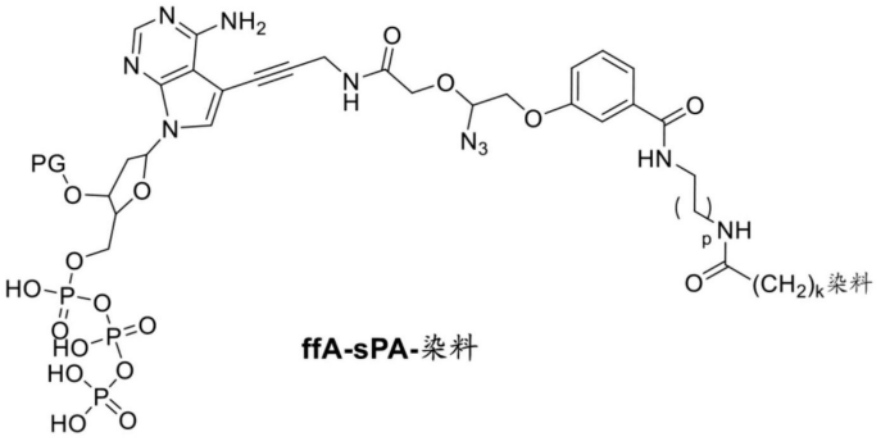
[0194] 

ffA-LN3-染料

,

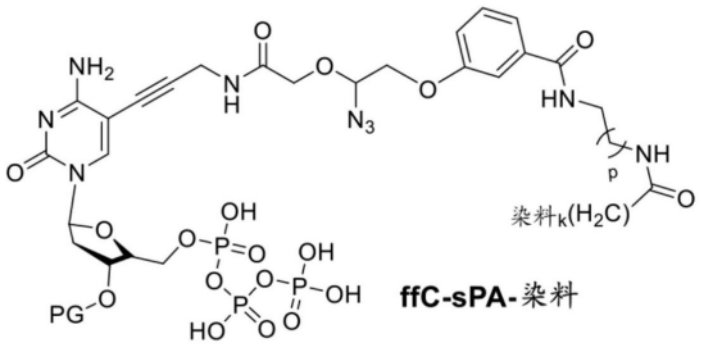


,

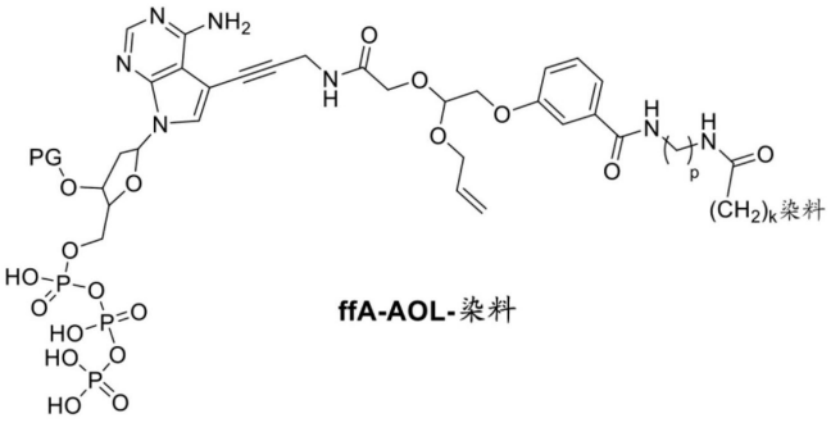


[0195]

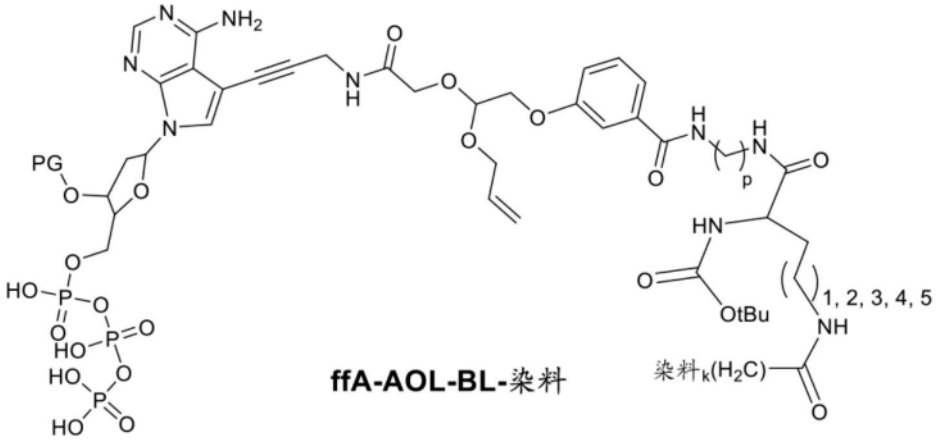
,



,

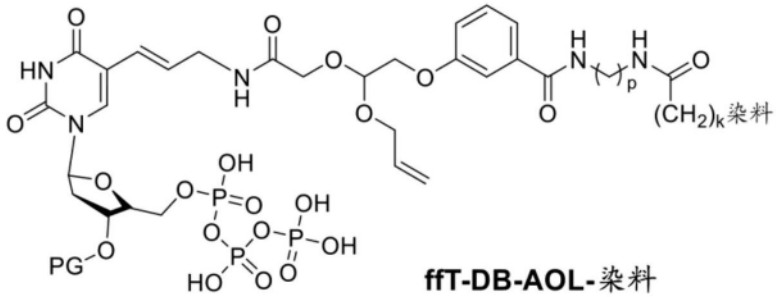


,

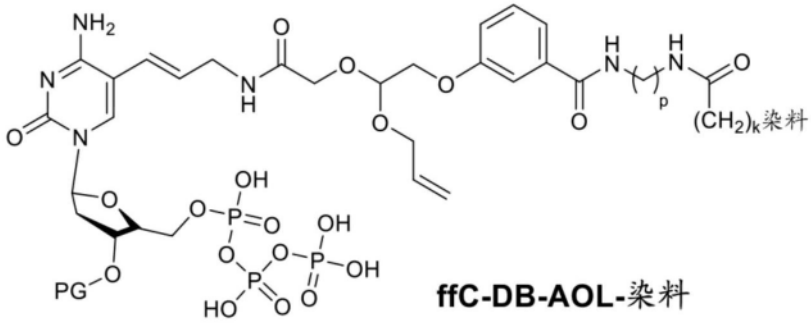


,

[0196]

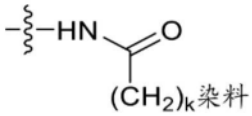


,



,

[0197] 其中PG代表本文所述的3'OH封端基团;p为1、2、3、4、5、6、7、8、9或10的整数;并且k为0、1、2、3、4或5。在一个实施方案中,-O-PG是AOM。在另一个实施方案中,-O-PG为-O-叠氮甲基。在一个实施方案中,k为5。在一些另外的实施方案中,p为1、2或3;并且k为5。



是指作为连接基部分的氨基与染料的羧基基团之间的反应的结果的染料与可裂解连接基的连接点。在本文所述的标记的核苷酸的任何实施方案中,核苷酸为核苷酸三磷酸。

[0198] 本公开的另外的方面涉及包含本文所述的标记的核苷酸的寡核苷酸或多核苷酸。在一些实施方案中,寡核苷酸或多核苷酸与靶多核苷酸的至少一部分杂交和/或互补。在一些实施方案中,靶多核苷酸被固定在固体载体上。在一些另外的实施方案中,固体载体包含多个固定的靶多核苷酸的阵列。在另外的实施方案中,固体载体包括图案化流动池。

[0199] 本公开的另外的方面涉及包含本文所述的一种或多种双硼染料的蛋白质标签或抗体。具体地,蛋白质标签或抗体可以包含相同染料的多个拷贝以增加荧光强度。蛋白质标签或抗体可以用作特异性结合特定类型的未标记的3'封端核苷酸的亲和试剂。

[0200] 试剂盒

[0201] 本文提供了包括用本公开的双硼染料(即,第一标记)标记的第一类型的核苷酸的试剂盒。在一些实施方案中,试剂盒还包括第二类型的标记的核苷酸,其用第二化合物(即,第二标记)标记,该第二化合物不同于第一标记的核苷酸中的双硼染料。在一些另外的实施方案中,试剂盒可以包括第三类型的核苷酸,其中第三类型的核苷酸用不同于第一标记和第二标记的第三化合物(即,第三标记)进行标记。在一些另外的实施方案中,试剂盒还可以包括第四类型的核苷酸。在一些此类实施方案中,第四类型的核苷酸未标记(暗色)。在其他实施方案中,与第一核苷酸、第二核苷酸和第三核苷酸相比,第四类型的核苷酸用不同的化合物进行标记,并且每种标记具有可与其他标记区分开的不同吸光度最大值。在一些实施方案中,核苷酸可以用于涉及使用具有不同波长的两种光源的测序应用中。在一些实施方案中,第一光源具有约500nm至约550nm、约510nm至约540nm或约520nm至约530nm(例如,520nm)的波长。第二光源具有约400nm至约480nm、约420nm至约470nm或450nm至约460nm(例如,450nm)的波长。在另外的实施方案中,第一标记、第二标记和第三标记中的每一者具有可以在两个单独的收集滤光器或通道中收集的发射光谱。

[0202] 在一些实施方案中,试剂盒可以含有四种类型的标记的核苷酸(A、C、G和T或U),其中第一类型的核苷酸用如本文所公开的化合物进行标记。在此类试剂盒中,四种类型的核苷酸中的每一者可以用与其他三种核苷酸上的标记相同或不同的化合物进行标记。另选地,四种类型的核苷酸中的第一类型是本文所述的标记的核苷酸(即,用本文所述的双硼染料标记),第二类型的核苷酸携带第二标记,第三类型的核苷酸携带第三标记,并且第四类型的核苷酸未标记(暗色)。作为另一示例,四种类型的核苷酸中的第一类型是本文所述的标记的核苷酸,第二类型的核苷酸携带第二标记,第三类型的核苷酸包含携带两种标记的第三类型的核苷酸的混合物(例如,携带第一标记的第三类型的核苷酸和携带第二标记的第三类型的核苷酸),并且第四类型的核苷酸未标记(暗色)。在该具体示例中,第三类型核苷酸的两种标记中的一者或两者可以是结构上不同于第一标记或第二标记但可以在相同波长的光源下激发但具有更强发射信号强度的标记(例如,携带第三标记的第三类型的核苷酸和携带第四标记的第三类型的核苷酸,其中第三标记可以在与第一标记相同的波长下激发,第四标记可以在与第二标记相同的波长下激发)。在这些示例中,标记化合物中的一者或多者可以具有不同的吸光度最大值和/或发射最大值,使得所述化合物能够与其他化合物区分开。例如,每种化合物可以具有不同的吸光度最大值和/或发射最大值,使得这些化合物中的每一种化合物在光谱上能够与其他三种化合物(或如果第四核苷酸未标记,则两种化合物)区分开。应当理解,吸收光谱和/或发射光谱中除最大值之外的部分可以不同,并且可以利用这些差异来区分化合物。该试剂盒可以是这样的:即,这些化合物中的两种或更多种化合物具有不同的吸光度最大值。本文所述的双硼染料通常吸收低于500nm的区域中的光。例如,这些双硼染料可以具有约410nm至约480nm、约420nm至约470nm、约440nm至约460nm的吸收波长。

[0203] 本文示出的双硼化合物、核苷酸或试剂盒可以用于检测、测量或识别生物系统(包括例如其过程或组分)。可以采用所述化合物、核苷酸或试剂盒的示例性技术包括测序、表达分析、杂交分析、遗传分析、RNA分析、细胞测定(例如,细胞结合或细胞功能分析)或者蛋白质测定(例如,蛋白质结合测定或蛋白质活性测定)。该用途可以是在用于执行特定技术的自动化仪器(诸如自动化测序仪器)上。测序仪器可以包括以不同波长操作的两个光源。

[0204] 在一个特定实施方案中,本文所描述的标记的核苷酸可以与未标记或天然核苷酸或其任何组合进行组合供应。核苷酸的组合可以作为独立的单独组分(例如,每个容器或管一个核苷酸类型)或作为核苷酸混合物(例如,在同一容器或管中混合的两种或更多种核苷酸)提供。

[0205] 在试剂盒包含多种,特别地是两种或三种、或者更特别地是四种核苷酸的情况下,不同的核苷酸可以用不同的染料化合物标记,或者一种可以是暗色的,没有染料化合物。在不同的核苷酸用不同的染料化合物标记的情况下,试剂盒的一个特征是染料化合物是光谱上可区分的荧光染料。如本文所用,术语"光谱上可区分的荧光染料"是指当一个样品中存在两种或更多种此类染料时,以能够通过荧光检测设备(例如,商业的基于毛细管的DNA测序平台)区分的波长发射荧光能量的荧光染料。当用荧光染料化合物标记的两种核苷酸以试剂盒形式提供时,一些实施方案的特征在于光谱上可区分的荧光染料能够以相同的波长激发,例如,通过相同的光源激发。当用荧光染料化合物标记的四种核苷酸以试剂盒形式提供时,一些实施方案的特征在于光谱上可区分的荧光染料中的两种荧光染料均能够以一个波长激发,并且另外两种光谱上可区分的染料均能够以另一个波长激发。染料的特定激发波长在450nm至460nm、490nm至500nm,或520nm或以上(例如,532nm)。

[0206] 在一个替代性的实施方案中,本公开的试剂盒可以包含其中相同的碱基用两种不同的化合物标记的核苷酸。第一核苷酸可以用本公开的化合物标记,例如在小于500nm处吸收的"蓝色"染料。第二核苷酸可以用光谱上不同的化合物标记,例如在小于600nm但高于500nm处吸收的"绿色"染料。第三核苷酸可以被标记为本公开的化合物和光谱上不同的化合物的混合物,并且第四核苷酸可以是"暗色"并且不含标记。因此,简单地讲,核苷酸1至4可以被标记为"蓝色"、"绿色"、"蓝色/绿色"和暗色。为了进一步简化仪器,可以用单个光源所激发的两种染料标记四种核苷酸,并且因此核苷酸1至4的标记可以是"蓝色1"、"蓝色2"、"蓝色1/蓝色2"和暗色。

[0207] 尽管本文以具有用不同染料化合物标记的不同核苷酸的构型为例来说明试剂盒,但是应当理解,试剂盒可以包括具有相同染料化合物的2种、3种、4种或更多种不同核苷酸。

[0208] 除了标记的核苷酸之外,试剂盒还可以一起包含至少一种附加组分。所述另外的组分可以是在本文示出的方法中或在下文的实施例部分中识别的组分中的一种或多种组分。可以结合到本公开的试剂盒中的组分的一些非限制性示例在下文示出。在一些实施方案中,试剂盒还包括DNA聚合酶(诸如9°N聚合酶的突变体,诸如WO 2005/024010中公开的那些)和一种或多种缓冲液组合物。一种缓冲液组合物可以包含抗氧化剂,诸如抗坏血酸或抗坏血酸钠,该抗氧化剂可以用于在检测期间保护染料化合物免受光损伤。另外的缓冲液组合物可以包含可用于裂解3'封端基团和/或可裂解连接基的试剂。例如,由过渡金属和至少部分水溶性配体(诸如钯络合物)形成的水溶性膦或水溶性过渡金属催化剂。试剂盒的各种组分可以浓缩形式提供以在使用前稀释。在此类实施方案中,还可以包括合适的稀释缓冲液。同样,在本文示出的方法中识别的组分中的一种或多种组分可以被包括在本公开的试剂盒中。在本文所述的核苷酸或标记的核苷酸的任何实施方案中,核苷酸包含3'封端基团。

[0209] 另选地,试剂盒可以包括一种或多种不同类型的未标记的3'封端核苷酸和一种或多种亲和试剂(例如,蛋白质标签和抗体),其中至少一种亲和试剂用本文所述的双硼染料的多个拷贝标记。

[0210] 测序方法

[0211] 包含根据本公开的染料化合物的核苷酸可以用于任何分析方法,诸如包括检测附接到此类核苷酸的荧光标记的方法,无论是以其本身使用,还是掺入到较大的分子结构或缀合物中或者与较大的分子结构或缀合物相关联。在该语境中,术语“掺入到多核苷酸中”可以意味着5’磷酸以磷酸二酯键与第二核苷酸的3’羟基基团接合,该第二核苷酸本身可以形成较长多核苷酸链的一部分。本文示出的核苷酸的3’端可以以磷酸二酯键或可以不以磷酸二酯键与另外的核苷酸的5’磷酸接合。因此,在一个非限制性实施方案中,本公开提供了检测掺入到多核苷酸中的标记的核苷酸的方法,该方法包括:(a)将至少一种本公开的标记的核苷酸掺入到多核苷酸中,以及(b)通过检测来自附接到所述核苷酸的染料化合物的荧光信号,来确定掺入到多核苷酸中的核苷酸的身份。

[0212] 该方法可以包括:合成步骤(a),其中将根据本公开的一种或多种标记的核苷酸掺入到多核苷酸中;以及检测步骤(b),其中通过检测或定量测量掺入到多核苷酸中的一种或多种标记的核苷酸的荧光,来检测所述核苷酸。

[0213] 本申请的一些实施方案涉及一种确定靶多核苷酸(例如,单链靶多核苷酸)的序列的方法,该方法包括:(a)使引物多核苷酸(例如,测序引物)与一种或多种标记的核苷酸(诸如核苷三磷酸A、G、C和T)接触,其中所述标记的核苷酸中的至少一者是本文所述的标记的核苷酸,并且其中该引物多核苷酸与该靶多核苷酸的至少一部分互补;(b)将标记的核苷酸掺入到所述引物多核苷酸中;以及(c)执行一次或多次荧光测量以确定所掺入的核苷酸的身份。在一些此类实施方案中,引物多核苷酸/靶多核苷酸复合物通过使靶多核苷酸和与靶多核苷酸的至少一部分互补的引物多核苷酸接触来形成。在一些实施方案中,该方法还包括(d)从掺入到引物多核苷酸中的核苷酸中去除标记部分和3’羟基封端基团。在一些另外的实施方案中,该方法还可以包括(e)将去除的标记部分和3’封端基团从引物多核苷酸链上洗掉。在一些实施方案中,重复步骤(a)至步骤(d)或步骤(a)至步骤(e),直至确定靶多核苷酸链的至少一部分的序列。在一些情况下,步骤(a)至(d)或步骤(a)至(e)重复至少30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250或300个循环。在一些实施方案中,在单一化学反应中去除来自掺入到引物多核苷酸链中的核苷酸的标记部分和3’封端基团。在一些另外的实施方案中,该方法在自动化测序仪器上执行,并且其中该自动化测序仪器包含在不同波长下操作的两个光源。在一些实施方案中,在完成本文所述的测序步骤的重复循环之后进行序列确定。

[0214] 本公开的一些实施方案涉及一种用于确定靶多核苷酸(例如,单链靶多核苷酸)的序列的方法,该方法包括:(a)使引物多核苷酸与包含四种不同类型的核苷酸缀合物中的一种或多种核苷酸缀合物的掺入混合物接触,其中第一类型的核苷酸缀合物包含第一标记,第二类型的核苷酸缀合物包含第二标记,并且第三类型的核苷酸缀合物包含第三标记,其中第一标记、第二标记和第三标记中的每一种标记在光谱上彼此不同,并且其中引物多核苷酸与靶多核苷酸的至少一部分互补;(b)将来自混合物的一种核苷酸缀合物掺入引物多核苷酸以产生延伸的引物多核苷酸;(c)使用第一激发光源执行第一成像事件并且检测来自延伸的多核苷酸的第一发射信号;以及(d)使用第二激发光源执行第二成像事件并且检测来自延伸的多核苷酸的第二发射信号;其中所述第一激发光源和所述第二激发光源具有不同波长;并且其中第一发射信号和第二发射信号在单个发射检测通道中检测或收集。在

一些实施方案中,本文所述的双硼染料可以用作该方法中描述的第一标记、第二标记或第三标记中的任一种标记。在一些实施方案中,该方法不包括在第一成像事件之后并且在第二成像事件之前对混合物中的任何核苷酸缀合物的化学修饰。在一些另外的实施方案中,掺入混合物还包含第四类型的核苷酸,其中第四类型的核苷酸未标记或者用不发射来自第一成像事件或第二成像事件的信号的荧光部分标记。在此测序方法中,基于第一成像事件和第二成像事件的检测模式确定每种所掺入的核苷酸缀合物的身份。例如,第一类型的核苷酸缀合物的掺入由第一成像事件中的信号态和第二成像事件中的暗态来确定。第二类型的核苷酸缀合物的掺入由第一成像事件中的暗态和第二成像事件中的信号态来确定。第三类型的核苷酸缀合物的掺入由第一成像事件和第二成像事件两者中的信号态来确定。第四类型的核苷酸缀合物的掺入由第一成像事件和第二成像事件两者中的暗态来确定。在另外的实施方案中,步骤(a)至(d)以重复循环进行(例如,至少30、50、100、150、200、250、300、400或500次),并且该方法还包括基于每种顺序掺入的核苷酸缀合物的身份依次确定单链靶多核苷酸的至少一部分的序列。在一些实施方案中,第一激发光源相比第二激发光源具有更短的波长。在一些此类实施方案中,第一激发光源具有约400nm至约480nm、约420nm至约470nm或约450nm至约460nm的波长(即"蓝光")。在一个实施方案中,第一激发光源具有约450nm的波长。第二激发光源具有约500nm至约550nm、约510nm至约540nm或约520nm至约535nm(即"绿光")的波长。在一个实施方案中,第二激发光源具有约520nm的波长。在其他实施方案中,第一激发光源相比第二激发光源具有更长的波长。在一些此类实施方案中,第一激发光源具有约500nm至约550nm、约510nm至约540nm或约520nm至约535nm(即"绿光")的波长。在一个实施方案中,第二激发光源具有约520nm的波长。第二激发光源具有约400nm至约480nm、约420nm至约470nm或约450nm至约460nm的波长(即"蓝光")。在一个实施方案中,第二激发光源具有约450nm的波长。

[0215] 本公开的一些实施方案涉及一种确定多个靶多核苷酸(例如,多个不同的靶多核苷酸)的序列的方法,所述方法包括:(a)使固体载体在杂交条件下与包含测序引物的溶液接触,其中所述固体载体包含多个固定在其上的不同靶多核苷酸;并且所述测序引物与所述靶多核苷酸的至少一部分互补;(b)使所述固体载体在适用于DNA聚合酶介导的引物延伸的条件下与包含DNA聚合酶和四种不同类型的核苷酸中的一者或多者的水溶液接触,其中至少一种类型的核苷酸是本文所述的标记的核苷酸;(c)将一种类型的核苷酸掺入到所述测序引物中以产生延伸的拷贝多核苷酸;以及(d)对所述延伸的拷贝多核苷酸进行一个或多个荧光测量以确定所掺入的核苷酸的身份。在一些实施方案中,该方法还包括(e)从掺入到延伸的拷贝多核苷酸中的核苷酸中去除3'封端基团。在一些此类实施方案中,步骤(e)还去除掺入的核苷酸的标记。在一些实施方案中,该方法还包括(f)在所述从掺入的核苷酸中去除标记和3'封端基团之后对固体载体进行洗涤。在另外的实施方案中,该方法包括重复步骤(b)至(f),直到确定靶多核苷酸的至少一部分的序列。在一些此类实施方案中,步骤(b)至(f)重复至少50、100、150、200、250或300个循环。在另外的实施方案中,在单一化学反应中去除来自掺入到延伸的拷贝多核苷酸中的核苷酸的标记和3'封端基团。在一些实施方案中,步骤(d)包括两次成像和荧光测量。在另外的实施方案中,该方法在自动化测序仪器上执行,并且其中该自动化测序仪器包含在不同波长处操作的两个光源。在一些此类实施方案中,一个光源在约400nm至约480nm、约420nm至约470nm或约450nm至约460nm(即"蓝

光”)处操作。在另外的实施方案中,另一个光源在约500nm至约550nm、约510nm至约540nm或约520nm至约535nm(即“绿光”)处操作。在一些实施方案中,四种类型的核苷酸包括dATP、dCTP、dGTP和dTTP或dUTP或它们的非天然核苷酸类似物。在特定实施方案中,包含DNA聚合酶和四种不同类型的核苷酸中的一者或多者的水溶液包含或为具有携带第一标记(用本文所述的双硼染料标记)的第一类型的核苷酸、携带第二标记的第二类型的核苷酸,和携带两种标记的混合物的第三类型的核苷酸,和未标记(暗色)的第四类型的核苷酸的掺入混合物。例如,第三类型的核苷酸可以是携带第一标记的第三类型的核苷酸和携带第二标记的第三类型的核苷酸的混合物。在此类实施方案中,第一类型的核苷酸的掺入可以由第一成像事件/荧光测量中的信号态和第二成像事件/荧光测量中的暗态来确定。第二类型的核苷酸的掺入可以由第一成像事件/荧光测量中的暗态和第二成像事件/荧光测量中的信号态来确定。第三类型的核苷酸的掺入由第一成像事件/荧光测量和第二成像事件/荧光测量两者中的信号态来确定。第四类型的核苷酸缀合物的掺入由第一成像事件/荧光测量和第二成像事件/荧光测量两者中的暗态来确定。在另一个实施方案中,掺入混合物包含携带第一标记(用本文所述的双硼染料标记)的第一类型的核苷酸、携带第二标记的第二类型的核苷酸、携带第三标记的第三类型的核苷酸和未标记的第四类型的核苷酸。在这种情况下,第一标记、第二标记和第三标记中的每一者在光谱上彼此不同,第一标记能够通过第一光源激发,第二标记能够通过第二光源激发,第三标记能够通过第一光源和第二光源两者激发。结果,四种类型的核苷酸的掺入也可以基于本文所述的相同信号模式来区分。

[0216] 在本文所述的测序方法的一些实施方案中,至少一种核苷酸在合成步骤中通过聚合酶的作用被掺入到多核苷酸(诸如本文所述的单链引物多核苷酸)中。然而,可以使用将核苷酸接合到多核苷酸的其他方法,诸如,化学寡核苷酸合成或者将标记的寡核苷酸与未标记的寡核苷酸连接。因此,当术语“掺入”关于核苷酸和多核苷酸使用时,可以涵盖通过化学方法以及酶促方法进行的多核苷酸合成。

[0217] 在一个特定实施方案中,进行了合成步骤并且该合成步骤可以任选地包括将模板多核苷酸链或靶多核苷酸链与包含本公开的荧光标记的核苷酸的反应混合物一起温育。还可以在允许在退火到模板或靶多核苷酸链的多核苷酸链上的游离3’羟基基团和标记的核苷酸上的5’磷酸基团之间形成磷酸二酯键的条件下提供聚合酶。因此,合成步骤可以包括通过核苷酸与模板/靶链的互补碱基配对来指导多核苷酸链的形成。

[0218] 在所述方法的所有实施方案中,检测步骤可以在向其中掺入了标记的核苷酸的多核苷酸链退火到模板/靶链上时进行,或者在其中将两条链分离的变性步骤之后进行。在合成步骤和检测步骤之间可以包括另外的步骤,例如化学反应步骤或酶促反应步骤,或者纯化步骤。具体地讲,可以分离或纯化掺入了标记的核苷酸的多核苷酸链,然后进一步加工或用于后续的分析。以举例的方式,在合成步骤中掺入如本文所述的标记核苷酸的多核苷酸链随后可以用作标记的探针或引物。在其他实施方案中,本文示出的合成步骤的产物可以经受进一步的反应步骤,并且如果需要,这些后续步骤的产物被纯化或分离。

[0219] 用于合成步骤的合适条件将为熟悉标准分子生物学技术的人员所熟知。在一个实施方案中,合成步骤可以类似于标准引物延伸反应,该标准引物延伸反应在合适的聚合酶的存在下使用核苷酸前体(包括如本文所述的标记的核苷酸)形成与模板/靶链互补的延伸的多核苷酸链(引物多核苷酸链)。在其他实施方案中,合成步骤本身可以形成产生标记的

双链扩增产物的扩增反应的一部分,该标记的双链扩增产物由来源于引物多核苷酸链和模板多核苷酸链的复制的退火互补链组成。其他示例性合成步骤包括切口平移、链置换聚合、随机引发的DNA标记等。用于合成步骤的特别有用的聚合酶是能够催化如本文示出的标记的核苷酸的掺入的聚合酶。可以使用多种天然存在的或突变的/经修饰的聚合酶。以举例的方式,热稳定聚合酶可以用于使用热循环条件进行的合成反应,而热稳定聚合酶可能不是等温引物延伸反应所期望的。能够掺入根据本公开的标记的核苷酸的合适的热稳定聚合酶包括WO 2005/024010或WO06120433中描述的那些,这些公开中的每一篇以引用方式并入本文。在较低温度诸如37℃处进行的合成反应中,聚合酶不必是热稳定聚合酶,因此对聚合酶的选择将取决于许多因素,诸如反应温度、pH、链置换活性等。示例性聚合酶包括但不限于Pol 812、Pol 1901、Pol 1558或Pol 963。例如,在美国专利公开2020/0131484 A1和2020/0181587 A1中描述了Pol 812、Pol 1901、Pol 1558或Pol 963 DNA聚合酶的氨基酸序列,这些公开均以引用方式并入本文。

[0220] 在具体的非限制性实施方案中,本公开涵盖以下的方法:核酸测序、重新测序、全基因组测序、单核苷酸多态性评分、涉及当用本文示出的染料标记的经修饰的核苷酸或核苷掺入到多核苷酸中时对其进行检测的任何其他应用。

[0221] 本公开的特定实施方案提供了包含根据本公开的染料部分的标记的核苷酸在多核苷酸边合成边测序反应中的用途。合成测序通常涉及使用聚合酶或连接酶在5'至3'方向上将一个或多个核苷酸或寡核苷酸顺序添加至生长的多核苷酸链,以便形成与待测序的模板/靶核酸互补的延伸多核苷酸链。存在于一个或多个添加的核苷酸中的碱基的身份可以在检测或"成像"步骤中确定。添加的碱基的身份可以在每个核苷酸掺入步骤之后测定。然后可以使用常规的Watson-Crick碱基配对规则来推断模板的序列。使用用本文示出的染料标记的核苷酸来测定单个碱基的身份可能是有用的,例如,在单核苷酸多态性的评分中,并且此类单碱基延伸反应在本公开的范围内。

[0222] 在本公开的一个实施方案中,模板/靶多核苷酸的序列通过检测附接到掺入的核苷酸的荧光标记,检测一个或多个核苷酸掺入到与待测序的模板多核苷酸互补的新生链中来确定。模板多核苷酸的测序可以用合适的引物(或制备成发夹构建体,其将包含引物作为发夹的一部分)引发,并且新生链通过在聚合酶催化的反应中向引物的3'端添加核苷酸而以逐个方式延伸。

[0223] 在特定实施方案中,不同的核苷酸三磷酸(A、T、G和C)中的每一种可以用独特的荧光团进行标记,并且还在3'位置处包含封端基团以防止不受控制的聚合。替代性地,四种核苷酸中的一种可以是未标记的(暗色)。聚合酶将核苷酸掺入到与模板多核苷酸/靶多核苷酸互补的新生链中,并且封端基团防止核苷酸的进一步掺入。任何未掺入的核苷酸都可以被洗掉,并且来自每种掺入的核苷酸的荧光信号可以通过合适的装置(如使用光源激发和合适的发射滤光器的电荷耦合装置)以光学方式"读取"。然后可以(同时或顺序地)除去(脱保护)3'封端基团和荧光染料化合物,以暴露新生链用于进一步掺入核苷酸。通常,将在每个掺入步骤之后确定所掺入的核苷酸的身份,但这不是严格必需的。类似地,美国专利号5,302,509(以引用方式并入本文)公开了对固定在固体载体上的多核苷酸进行测序的方法。

[0224] 如上文所例示,该方法利用了在DNA聚合酶的存在下将荧光标记的3'-封闭的核苷酸A、G、C和T掺入到与固定的多核苷酸互补的生长链中。聚合酶掺入与靶多核苷酸互补的碱

基,但是被3'-封闭基团防止进一步添加。然后可以确定掺入的核苷酸的标记,并且通过化学裂解除去封闭基团以允许发生进一步的聚合。在合成测序反应中待测序的核酸模板可以是期望测序的任何多核苷酸。用于测序反应的核酸模板将通常包含具有游离的3'羟基基团的双链区域,该双链区域充当用于在测序反应中添加另外的核苷酸的引物或起始点。待测序模板的该区域将在互补链上悬垂该游离的3'羟基基团。待测序模板的悬垂区域可以是单链的,但也可以是双链的,前提条件是在与待测序的模板链互补的链上存在"切口",以提供用于引发测序反应的游离3'OH基团。在此类实施方案中,测序可以通过链置换进行。在某些实施方案中,可以添加带有游离的3'羟基基团的引物,作为与待测序模板的单链区域杂交的单独组分(例如,短寡核苷酸)。替代性地,待测序的引物和模板链可以各自形成能够形成分子内双链体(诸如发夹环结构)的部分自身互补核酸链的一部分。发夹多核苷酸及其可以附接到固体载体的方法在PCT公开号WO0157248和WO2005/047301中公开,这些公开中的每一篇以引用方式并入本文。核苷酸可以被连续地添加到生长引物,导致在5'至3'方向上合成多核苷酸链。可以确定已添加的碱基的性质,特别是但不必在每次添加核苷酸之后,从而提供核酸模板的序列信息。因此,将核苷酸掺入核酸链(或多核苷酸)中的方式为,经由与该核苷酸的5'磷酸基团形成磷酸二酯键而将该核苷酸接合到核酸链的游离3'羟基基团。

[0225] 待测序的核酸模板可以是DNA或RNA,或者甚至是由脱氧核苷酸和核糖核苷酸组成的杂合分子。核酸模板可以包含天然存在的和/或非天然存在的核苷酸以及天然或非天然的骨架键,前提条件是这些不阻止测序反应中模板的复制。

[0226] 在某些实施方案中,待测序的核酸模板可以经由本领域已知的任何合适的连接方法(例如经由共价附接)附接到固体载体。在某些实施方案中,模板多核苷酸可以直接附接到固体载体(例如,基于二氧化硅的载体)。然而,在本公开的其他实施方案中,固体载体的表面可以以某种方式修饰,以便允许模板多核苷酸的直接共价附接,或者通过本身可以非共价地附接到固体载体的水凝胶或聚电解质多层来固定模板多核苷酸。

[0227] 其中多核苷酸已直接附接到载体(例如,基于二氧化硅的载体,诸如WO00/06770(以引用方式并入本文)中所公开的那些)的阵列,其中通过玻璃上的环氧侧基与多核苷酸上的内部氨基基团之间的反应将多核苷酸固定在玻璃载体上。此外,可以通过硫基亲核物质与固体载体的反应将多核苷酸附接到固体载体,例如,如WO2005/047301(以引用方式并入本文)中所述。固体支撑的模板多核苷酸的其他另外的示例是模板多核苷酸附接到在基于二氧化硅或其他固体载体上负载的水凝胶,例如如WO00/31148、WO01/01143、WO02/12566、WO03/014392、美国专利6,465,178和WO00/53812中所述,这些专利中的每一篇以引用方式并入本文。

[0228] 模板多核苷酸可以固定于其上的特定表面是聚丙烯酰胺水凝胶。聚丙烯酰胺水凝胶描述于上文引用的参考文献和WO2005/065814中,该专利以引用方式并入本文。可以使用的特定水凝胶包括WO2005/065814和美国公开2014/0079923中所述的那些。在一个实施方案中,水凝胶为PAZAM(聚(N-(5-叠氮基乙酰氨基戊基)丙烯酰胺-共-丙烯酰胺))。

[0229] DNA模板分子可以连接到珠粒或微粒,例如,如美国专利第6,172,218号(以引用方式并入本文)中所述。与珠粒或微粒的附接可以用于测序应用。可以制备珠粒文库,其中每个珠粒包含不同的DNA序列。示例性文库及其产生方法描述于Nature,437,376-380(2005);Science,309,5741,1728-1732(2005),这些文献各自以引用方式并入本文。使用本文示出

的核苷酸对此类珠粒的阵列进行测序在本公开的范围内。

[0230] 待测序的模板可以形成固体载体上的“阵列”的一部分,在这种情况下,阵列可以采取任何方便的形式。因此,本公开的方法适用于所有类型的高密度阵列,包括单分子阵列、成簇阵列和珠粒阵列。用本公开的染料化合物标记的核苷酸可以用于对基本上任何类型的阵列上的模板(包括但不限于通过将核酸分子固定在固体载体上而形成的那些)进行测序。

[0231] 然而,用本公开的染料化合物标记的核苷酸在对成簇阵列进行测序的背景中是特别有利的。在成簇阵列中,阵列上的不同区域(通常称为位点或特征)包含多个多核苷酸模板分子。一般来讲,所述多个多核苷酸分子不能通过光学手段单独地分辨,而是作为整体被检测到。取决于阵列的形成方式,阵列上的每个位点可以包含一个单独多核苷酸分子的多个拷贝(例如,该位点对于特定的单链核酸种类或双链核酸种类是同质的)或甚至少量不同多核苷酸分子的多个拷贝(例如,两个不同核酸种类的多个拷贝)。核酸分子的成簇阵列可以使用本领域中公知的技术产生。以举例的方式,WO 98/44151和WO00/18957(这些专利中的每一篇以引用方式并入本文)描述了扩增核酸的方法,其中模板和扩增产物均保持固定在固体载体上,以便形成由固定的核酸分子的簇或“集落”组成的阵列。根据这些方法制备的成簇阵列上存在的核酸分子是使用用本公开的染料化合物标记的核苷酸进行测序的合适模板。

[0232] 用本公开的染料化合物标记的核苷酸也可用于对单分子阵列上的模板进行测序。如本文所用的术语“单分子阵列”或“SMA”,是指分布(或排列)在固体载体上的多核苷酸分子群,其中任何单独的多核苷酸与该分子群的所有其他多核苷酸的间距使得有可能单独地分辨单独的多核苷酸分子。因此,在一些实施方案中,固定到固体载体的表面上的靶核酸分子能够通过光学手段分辨。这意味着一个或多个不同的信号(每个信号代表一种多核苷酸)将出现在所用的特定成像装置的可分辨区域内。

[0233] 可以实现单分子检测,其中阵列上的相邻多核苷酸分子之间的间距为至少100nm,更具体地讲至少250nm,还更具体地讲至少300nm,甚至更具体地讲至少350nm。因此,每个分子能够作为单分子荧光点来单独地分辨和检测,并且来自所述单分子荧光点的荧光也表现出单步光漂白。

[0234] 术语“单独分辨的”和“单独分辨”在本文中用于规定,当可视化时,有可能将阵列上的一个分子与其相邻分子区分开。阵列上各个分子之间的间隔将部分地通过用于分辨各个分子的特定技术来确定。通过参考已公布的申请WO00/06770和WO 01/57248将理解单分子阵列的一般特征,这些申请中的每个申请以引用方式并入本文。虽然本公开的标记的核苷酸的一种用途是用于合成测序反应中,但此类核苷酸的效用却不限于此类方法。事实上,本文所述的标记的核苷酸可以有利地用于需要检测连接于掺入到多核苷酸中的核苷酸的荧光标记的任何测序方法。

[0235] 具体地讲,用本公开的染料化合物标记的核苷酸可以用于自动化荧光测序方案,尤其是基于Sanger和同事的链终止测序方法的荧光染料-终止子循环测序。此类方法通常使用酶和循环测序将荧光标记的双脱氧核苷酸掺入引物延伸测序反应中。所谓的Sanger测序方法和相关方案(Sanger型)利用带有标记的双脱氧核苷酸的随机化链终止。

[0236] 因此,本公开还涵盖用染料化合物标记的核苷酸,这些染料化合物是在3’位置和

2'位置处都缺乏羟基基团的双脱氧核苷酸,此类经修饰的双脱氧核苷酸适合在Sanger型测序方法等中使用。

[0237] 应当理解,用掺入3'封端基团的本公开的染料化合物标记的核苷酸也可以用于Sanger方法和相关方案,因为通过使用具有3'羟基封端基团的核苷酸可以实现与通过使用双脱氧核苷酸所实现的效果相同的效果:两者均防止随后的核苷酸的掺入。在根据本公开的具有3'封端基团的核苷酸将用于Sanger型测序方法的情况下,应当理解,附接到核苷酸的染料化合物或可检测标记不需要经由可裂解的连接基连接,因为在掺入本公开的标记的核苷酸的每种情况下;随后不需要掺入核苷酸,因此不需要从核苷酸上除去标记。

[0238] 另选地,还可以使用未标记的核苷酸和含有本文所述的荧光染料的亲和试剂来进行本文所述的测序方法。例如,在步骤(a)的掺入混合物中,四种不同类型的核苷酸(例如,dATP、dCTP、dGTP和dTTP或dUTP)中的一种类型、两种类型、三种类型或每种类型可以未被标记。四种类型的核苷酸(例如,dNTP)中的每一种类型都具有3'羟基封端基团,以确保仅单一碱基可以通过聚合酶添加到引物多核苷酸的3'端。在步骤(b)中掺入未标记的核苷酸后,洗掉剩余未掺入的核苷酸。然后引入亲和试剂,其特异性地识别并结合掺入的dNTP以提供包含掺入的dNTP的标记的延伸产物。在WO 2018/129214和WO 2020/097607中已经公开了未标记的核苷酸和亲和试剂在合成测序中的用途。使用未标记核苷酸的本公开的修饰的测序方法可以包括以下步骤:

[0239] (a')使引物多核苷酸/靶多核苷酸复合物与一种或多种未标记的核苷酸(例如,dATP、dCTP、dGTP和dTTP或dUTP)接触,其中引物多核苷酸与靶多核苷酸的至少一部分互补;

[0240] (b')将核苷酸掺入到引物多核苷酸中以产生延伸的引物多核苷酸;

[0241] (c')使延伸的引物多核苷酸与一组亲和试剂在一种亲和试剂特异性结合掺入的未标记的核苷酸的条件下接触,以提供标记的延伸的引物多核苷酸/靶多核苷酸复合物;

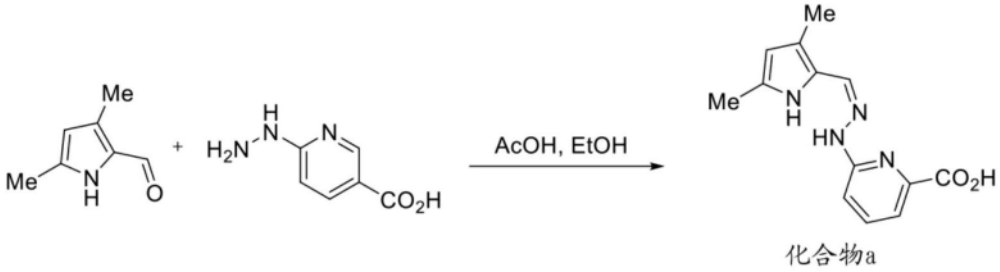
[0242] (d')对标记的延伸的引物多核苷酸/靶多核苷酸复合物执行一次或多次荧光测量,以确定掺入的核苷酸的身份。

[0243] 在本文所述的修改的测序方法的一些实施方案中,掺入混合物中的未标记的核苷酸中的每一种核苷酸含有3'封端基团。在另外的实施方案中,在下一掺入循环之前去除掺入的核苷酸的3'封端基团。在其他另外的实施方案中,该方法还包括从掺入的核苷酸中去除亲和试剂。在其他另外的实施方案中,在相同反应中去除3'羟基封端基团和亲和试剂。在一些实施方案中,亲和试剂组可以包含特异性结合第一类型的核苷酸的第一亲和试剂、特异性结合第二类型的核苷酸的第二亲和试剂和特异性结合第三类型的核苷酸的第三亲和试剂。在一些另外的实施方案中,第一亲和试剂、第二亲和试剂和第三亲和试剂中的每一者均包含一个或多个可检测的、光谱可区分的标记。在一些实施方案中,亲和试剂可以包括蛋白质标签、抗体(包括但不限于抗体的结合片段、单链抗体、双特异性抗体等)、适体、打结素、affimer或以合适的特异性和亲和力结合掺入的核苷酸的任何其他已知的试剂。在一个实施方案中,至少一种亲和试剂是抗体或蛋白质标签。在另一个实施方案中,第一类型的亲和试剂、第二类型的亲和试剂和第三类型的亲和试剂中的至少一种是包含一个或多个可检测标记(例如,相同可检测标记的多个拷贝)的抗体或蛋白质标签,其中可检测标记是或包括本文所述的双硼染料部分。

[0244] 实施例

[0245] 在以下实施例中更详细地公开了附加的实施方案,这些实施例并非旨在以任何方式限制权利要求书的范围。

[0246] 实施例1.含有稠合双硼杂环的染料的合成

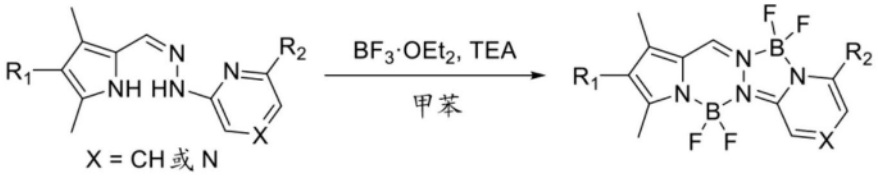
[0247] 

化合物a

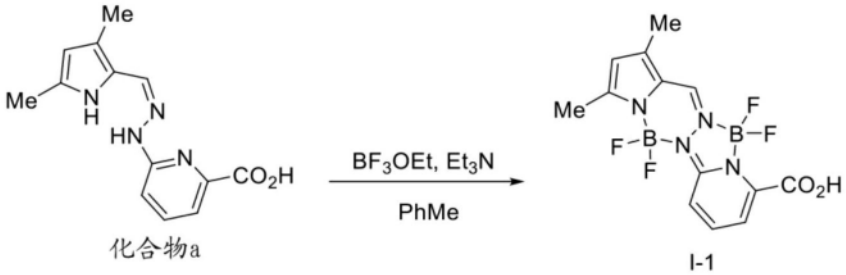
[0248] 将含3,5-二甲基吡咯-2-甲醛(369mg,3.00mmol)和6-肼基烟酸(459mg,3.00mmol)的EtOH(20mL)用AcOH(100μL)处理并在回流下加热5h。将所得沉淀物在真空下过滤,并用EtOH洗涤,得到相应的腙产物(化合物a),为黄色固体(595mg,77%)。^{1}H NMR(400MHz,DMSO-d_6)δ10.91(s,1H),10.73(s,1H),8.59(d,J=2.2Hz,1H),8.07-7.90(m,2H),7.28(d,J=8.9Hz,1H),5.66(d,J=2.5Hz,1H),2.19(s,3H),2.06(s,3H)。

[0249] 式(I)的一些含双硼的稠合吡啶并杂环和吡嗪并杂环的衍生物根据本文所述的一般程序制备。

[0250] 一般程序A

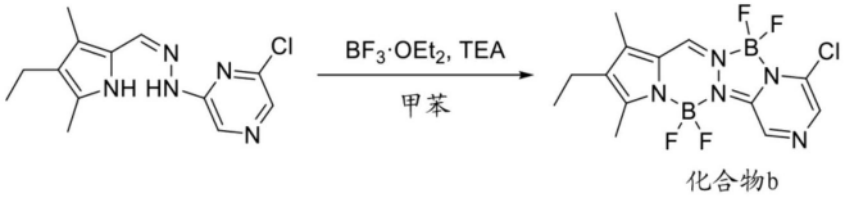
[0251] 

[0252] 将相关取代的{2-[(1H-吡咯-2-基)亚甲基]肼基}吡啶或吡嗪(1.0当量)的甲苯溶液用TEA(18.0当量)处理。将反应混合物回流10分钟,然后滴加$BF_3 \cdot OEt_2$(20.0当量)。将反应混合物在回流下搅拌5h。在真空下除去反应溶剂。将粗物质溶解在DCM中,并且将有机层用H_2O洗涤,然后经无水Na_2SO_4干燥。通过快速色谱法纯化粗产物。

[0253] 

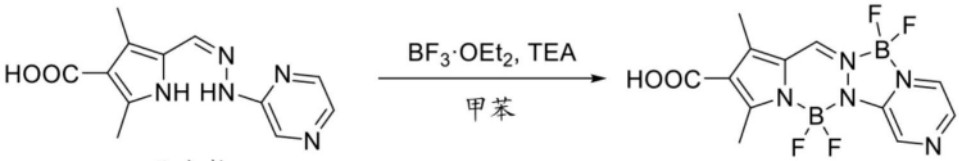
化合物a I-1

[0254] 将含腙化合物a(129mg,0.5mmol)的甲苯(10mL)用Et_3N(1.25mL)处理并在室温处搅拌10分钟。然后滴加BF_3OEt(1.5mL)并且将反应混合物在回流下搅拌18h。将混合物冷却至室温,真空浓缩,然后通过制备型反相HPLC纯化,得到化合物I-1(70μmol,14%)。质谱:[M$^-$]=353

[0255] 

化合物b

[0256] 基于一般程序A,由(Z)-2-氯-6-(2-((4-乙基-3,5-二甲基-1H-吡咯-2-基)亚甲基)肼基)吡嗪a制备化合物b。将粗化合物通过快速色谱法纯化,得到化合物a,为亮黄色固体(收率:61%)。MS[M+H]$^+$=374。

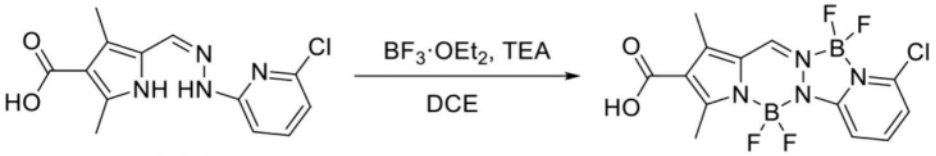
[0257] 

化合物c I-8

[0258] 基于一般程序A由化合物c制备化合物I-8。将粗化合物通过快速色谱法纯化,得到I-8,为亮黄色固体(收率:67%)。MS[M+H]$^+$=356,[M-H]$^-$=354。

[0259] 
化合物d 化合物e

[0260] 基于一般程序A由化合物d制备化合物e。通过NMR和LCMS确认结构和组成。

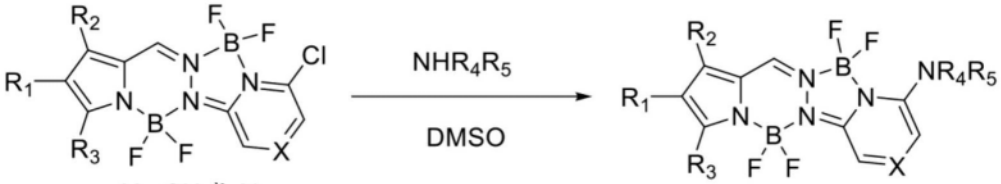
[0261] 

化合物f I-7

[0262] 根据一般程序A,由5-((2-(6-氯吡啶-2-基)亚肼基)甲基)-2,4-二甲基-1H-吡咯-3-甲酸(化合物f)(1.0当量)制备化合物I-7,为亮黄色固体。将粗物质通过快速色谱法纯化并得到最终产物,为亮黄色固体(收率:59%)。MS[M-H]$^-$=387。

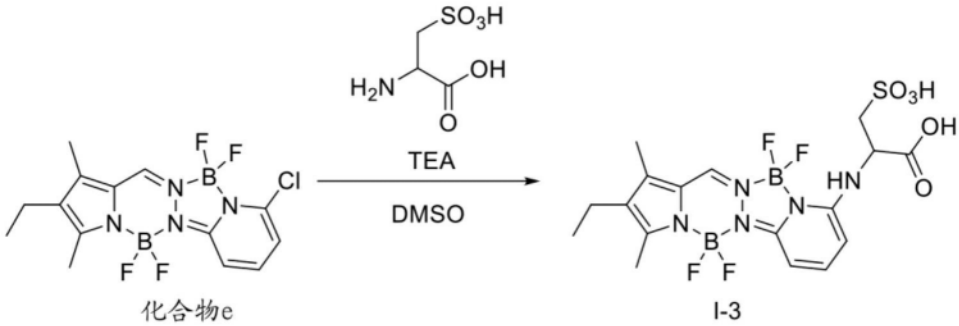
[0263] 式(I)的含双硼的稠合吡啶并杂环和吡嗪并杂环的一些新的官能衍生物还可以例如根据一般程序B通过取代基的修饰,例如取代吖嗪环的2位上的氯原子来制备。

[0264] 一般程序B

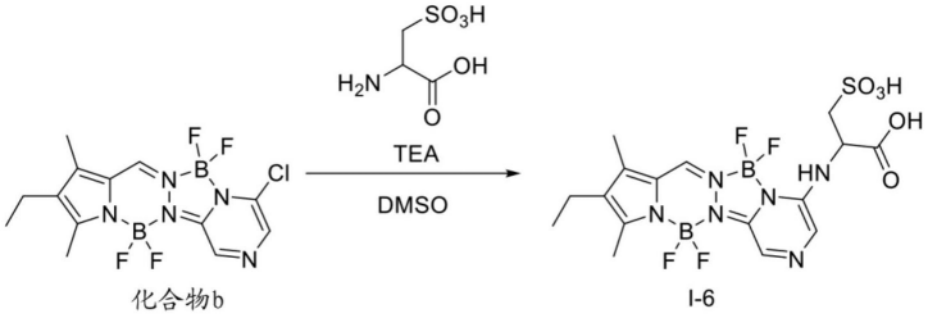
[0265] 

X= CH或 N

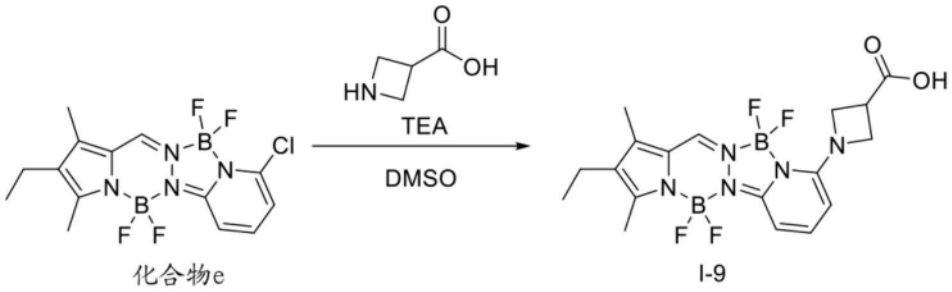
[0266] 将适当氯取代的化合物(1当量)、伯胺、仲胺或氨基酸(1.1当量)和TEA(2当量)在DMSO中的混合物在95℃处搅拌5h。然后将反应混合物用MeCN和0.1M TEAB稀释并通过反相制备型HPLC纯化。

[0267] 

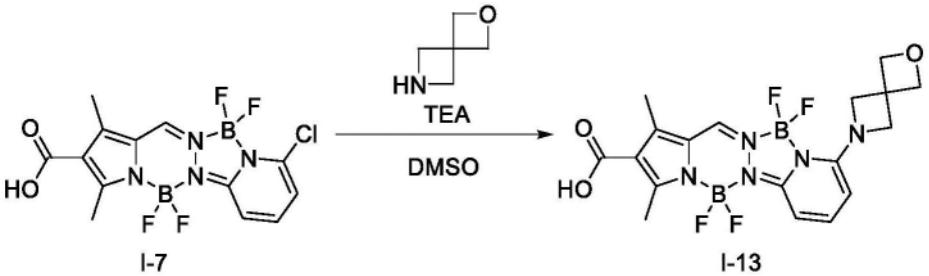
[0268] 按照一般程序B,通过使化合物e与磺基丙氨酸反应来制备化合物I-3。将反应混合物在95℃处加热5h,得到最终产物,为亮黄色固体(收率:79%)。MS[M-H]⁻=504。

[0269] 

[0270] 按照一般程序B,通过使化合物b与磺基丙氨酸反应来制备化合物I-6。将反应混合物在95℃处加热16h,得到最终产物,为亮黄色固体(收率:5%)。MS[M-H]⁻=505。

[0271] 

[0272] 根据一般程序B,通过使化合物e与氮杂环丁烷-3-甲酸反应来制备化合物I-9。将反应混合物在95℃处加热5h,得到最终产物,为亮黄色固体(收率:99%)。MS[M-H]⁻=436。

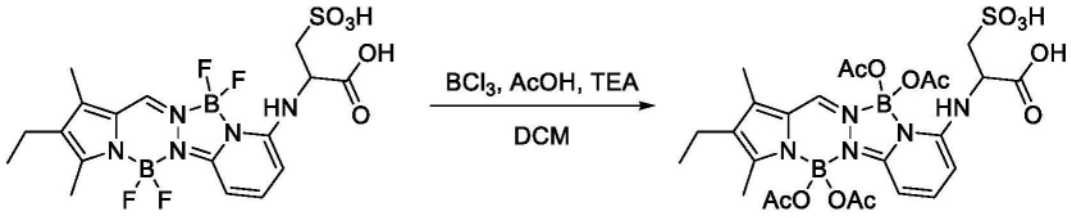
[0273] 

[0274] 根据一般程序B,通过使化合物I-7与2-氧杂-6-氮杂螺[3.3]庚烷反应来制备化合物I-13。分离产物,为亮黄色固体(收率:69%)。MS[M-H]⁻=450。

[0275] 一般程序C

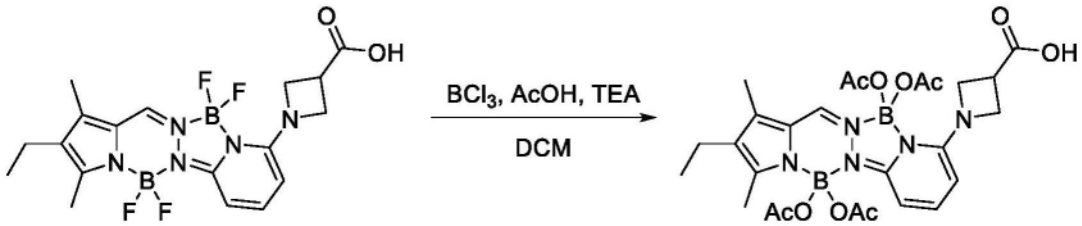
[0276] 

[0277] 向适当的含有双-二氟硼的稠合杂环(1当量)在DCM中的溶液中滴加BCl_3(4.5当量)。将反应混合物搅拌30min,然后添加TEA(12.0当量),随后添加脂族、芳族单碳酸或二碳酸,例如乙酸(8.0当量)或丙二酸(4.0当量)或它们的衍生物。将反应混合物搅拌16h。将粗物质通过硅藻土过滤,并且另外用DCM洗涤硅藻土。在真空下除去溶剂,并且将所得残余物溶解在含有0.1M TEAB的MeCN中,并通过反相制备型HPLC纯化。

[0278] 

I-3 I-4

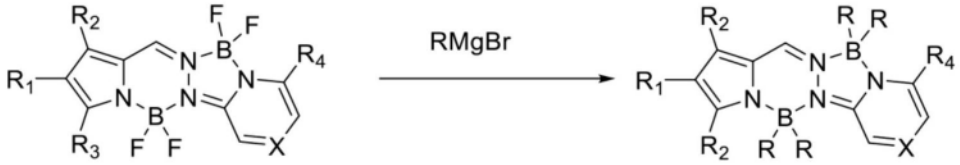
[0279] 基于一般程序C,使用乙酸由I-3制备化合物I-4。将反应混合物在室温处搅拌16h,得到I-3(收率:5%)。MS[M-H]⁻=664。

[0280] 

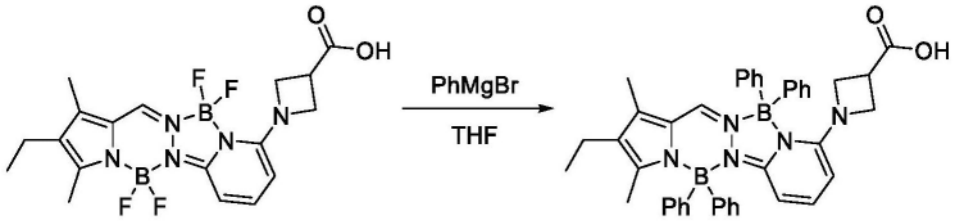
I-9 I-10

[0281] 基于一般程序C,由I-9制备化合物I-10。将反应混合物在室温处搅拌16h,得到I-10(收率:3%)。MS[M-H]⁻=596。

[0282] 一般程序D

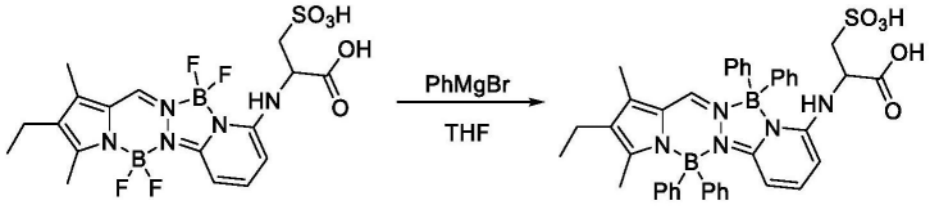
[0283] 

[0284] 在-78℃处,向适当的含有双-二氟硼的稠合杂环(1当量)在THF中的溶液中滴加格氏试剂RMgBr(20.0当量)。将反应混合物搅拌16h。在真空下除去溶剂。将残余物溶解在DCM中,并且将有机层用H_2O/NH_4Cl洗涤,然后经无水Na_2SO_4干燥,并通过反相制备型HPLC纯化。

[0285] 

I-9 I-14

[0286] 基于一般程序D,通过使I-9与PhMgBr反应来制备化合物I-14(收率:12%)。MS[M-H]⁻=668。

[0287] 

I-3 I-15

[0288] 基于一般程序D,通过使I-3与PhMgBr反应来制备化合物I-15(收率:3%)。MS[M-H]⁻=736。

[0289] 本文所公开的一些示例性染料的荧光光谱汇总在下表1中。

[0290] 表1.

[0291]

染料	吸光度最大值(nm)	发射最大值(nm)	斯托克斯位移(nm)	溶剂
I-1	429	472	43	EtOH
I-3	442	458	16	ACN
I-4	449	471	22	ACN
I-6	467	500	33	ACN
I-7	422	451	29	ACN
I-8	429	557	128	ACN
I-9	450	480	30	ACN
I-10	449	491	42	ACN
I-13	445 449	471 470	26 21	ACN EtOH
I-14	462	481	19	ACN
I-15	470	487	17	ACN

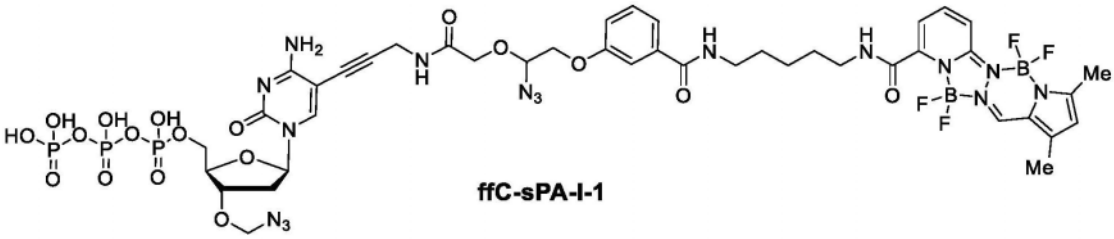
[0292] 实施例2.用含有双硼稠合杂环的染料标记的ffN的合成

[0293] 本文所述的含双硼的稠合杂环化合物可以通过与含有氨基部分的适当官能化核苷酸衍生物的偶联反应而用于核苷酸标记。

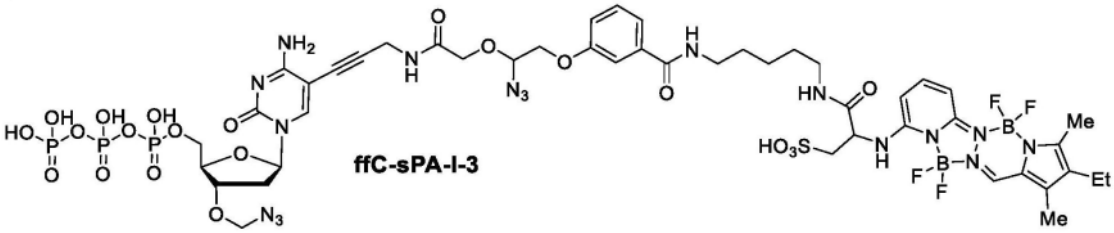
[0294] 一般程序E

[0295] 所述的含双硼的稠合杂环化合物可以通过与它们含有氨基部分的适当官能化核苷酸的偶联反应而用于核苷酸标记。将式(I)的染料溶解在无水N,N'-二甲基乙酰胺(DMA)中。添加N,N-二异丙基乙胺(DIPEA),随后添加TNTU。在室温处,将反应物在氮气下搅拌30分钟。将活化的双硼染料溶液添加到3'封端的2'-脱氧核苷三磷酸-连接基的三乙基碳酸氢铵(TEAB)溶液中,并且将反应物在室温处搅拌18小时。首先在DEAE-Sephadex A25上通过离子交换色谱法纯化粗产物。汇集含有官能化核苷酸的级分,并且将溶剂在减压下蒸发至干燥。

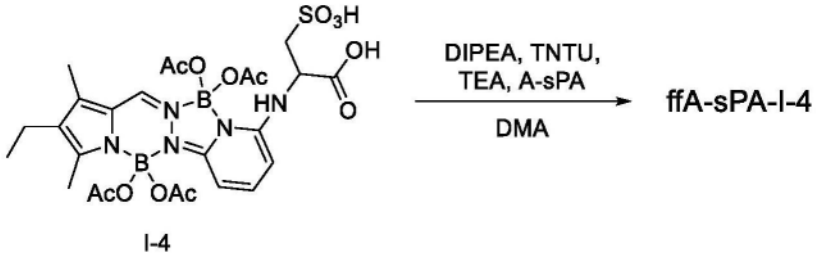
使用YMC-Pack-Pro C18柱通过制备型RP-HPLC进一步纯化粗材料。最终化合物通过LC-MS、分析型RP-HPLC和UV-Vis光谱法表征。

[0296] 

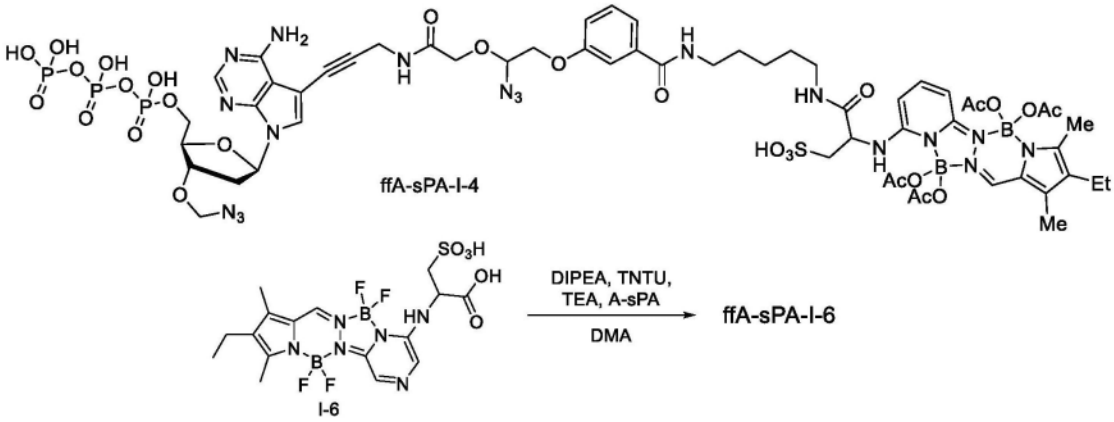
[0297] 基于ffN偶联的一般程序由I-1制备ffC-sPA-I-1(收率:14%)。MS[M⁻]=1257。

[0298] 

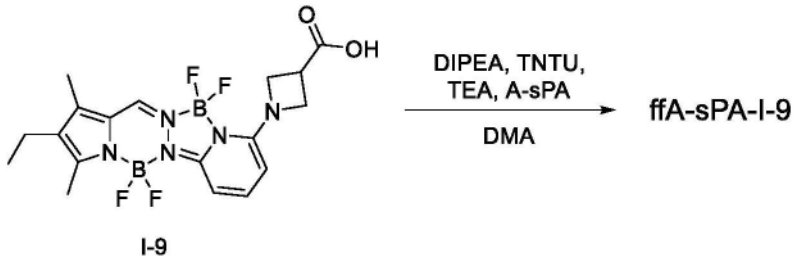
[0299] 基于ffN偶联的一般程序由I-3制备ffC-sPA-I-3(收率:6%)。MS[M⁻]=1408。

[0300] 

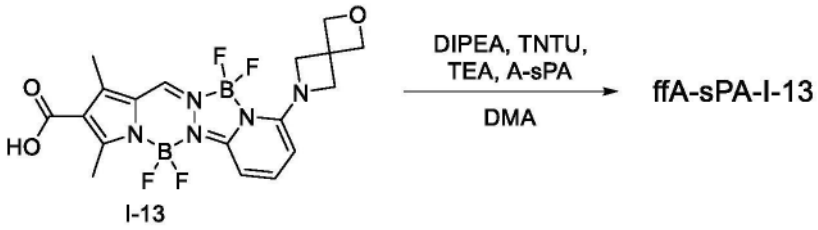
[0301] 基于ffN偶联的一般程序由I-4制备ffA-sPA-I-4(收率:24%)MS[M-2H]$^{2-}$=795。

[0302] 

[0303] 基于ffN偶联的一般程序由I-6制备ffA-sPA-I-6(收率:8%)。MS[M-2H]$^{2-}$=717。

[0304] 

[0305] 基于ffN偶联的一般程序由I-9制备ffA-sPA-I-9(收率:17%)。MS[M-2H]$^{2-}$=681。

[0306] 

[0307] 基于ffN偶联的一般程序由I-13制备ffA-sPA-I-13。将反应混合物在40℃处加热48h,并且由于偶联反应缓慢,A-SpA和DIPEA的最终当量分别为(2.0当量)和(20.0当量)。MS $[M-H]^-$=1378,$[M+H]^+$=1380。

[0308] 本文所公开的示例性ffN的荧光光谱汇总在下表2中。

[0309] 表2.

[0310]

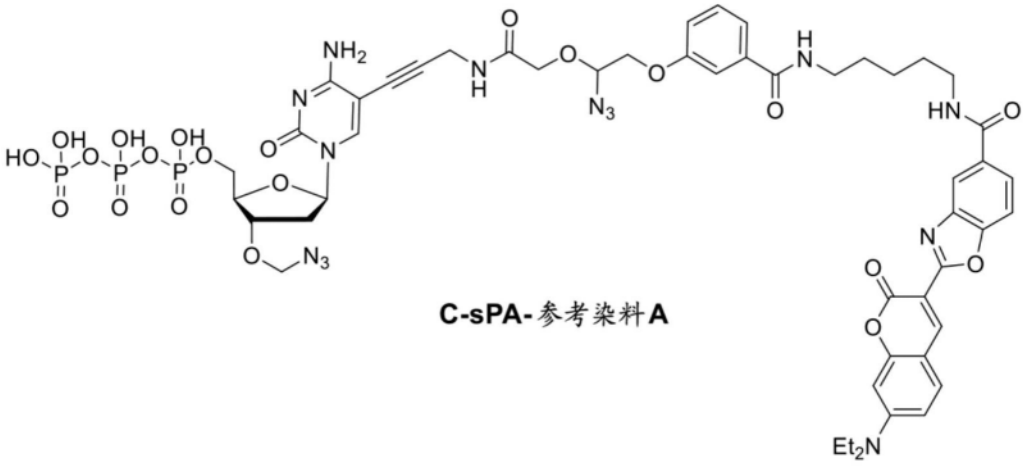
ffN	吸光度最大值 (nm)	荧光最大值 (nm)	斯托克斯位移 (nm)	溶剂
ffC-sPA-I-1	411	482	69	USM
ffC-sPA-I-3	443	462	19	USM
ffA-sPA-I-4	447	463	16	H_2O
ffA-sPA-I-6	464	545	81	H_2O

[0311]

ffA-sPA-I-9	448	476	28	H_2O
	455	474	19	USM
ffA-sPA-I-13	439	503	64	H_2O
	442	497	55	USM

[0312] 实施例3.ffN光谱特性比较

[0313] 在该实施例中,对与双硼染料I-4(A-sPA-I-4)缀合的完全官能化的A核苷酸(ffA)的光谱特性进行表征。图1示出了A-spA-I-4和用参考染料A(C-sPA-参考染料A)标记的市售完全官能化的C核苷酸(ffC)在通用扫描混合物(USM,1M Tris pH 7.5,0.05% TWEEN,20mM抗坏血酸钠,10mM没食子酸乙酯)中的发射光谱。使用石英或塑料比色皿在Agilent Cary 100UV-Vis分光光度计上和在Cary Eclipse荧光分光光度计上获取光谱。据观察,A-sPA-I-4与参考染料A相比具有更短的斯托克斯位移。

[0314] 

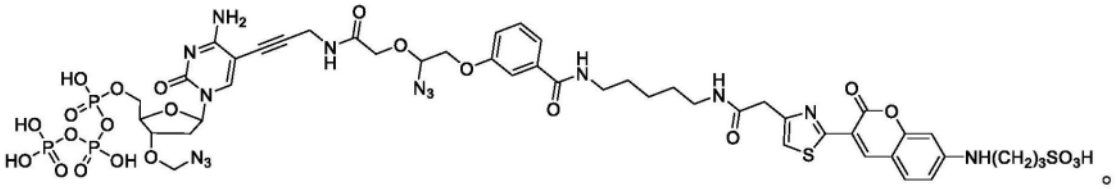
[0315] 实施例4.双硼染料的稳定性

[0316] 通过将化合物在含有50mM乙醇胺的掺入缓冲液中在37℃处在黑暗中温育2天,对化合物I-1和I-3的稳定性进行评估并与用参考染料A标记的市售ffC进行比较。使用石英比

色皿在Agilent Cary 100 UV-Vis分光光度计上和在Cary Eclipse荧光分光光度计上测量溶液的荧光强度。另外,取等分溶液并通过分析型HPLC进行分析。图2表明,与C-sPA-参考染料A相比,I-1和I-3的荧光强度随时间推移非常缓慢地降低,表明在相同条件下双硼染料I-1和I-3与参考染料A相比更稳定。

[0317] 实施例5.因美纳MiSeq™平台上的测序实验

[0318] 在因美纳MiSeq™仪器上测试了用双硼染料I-4标记的ffA,该仪器设置为在蓝色激发光(约450nm)的情况下拍摄第一图像并且在绿色激发光(约520nm)的情况下拍摄第二图像。实验使用的掺入混合物包括以下五种ffN:A-sPA-I-4、用已知的聚甲炔绿色染料NR550S0标记的ffA(A-sPA-NR550S0)、用蓝色香豆素染料标记的ffC(C-sPA-参考染料B)、用绿色染料NR550S0标记的ffT(T-sPA-NR550S0)和在50mM乙醇胺缓冲液pH 9.6、50mM NaCl、1mM EDTA、0.2% CHAPS、4mM $MgSO_4$和DNA聚合酶中的未标记的ffG(暗色G)。图3示出了含有ffA-spA-I-4的ffN组与市售参考1ffN组和参考2ffN组相比的测序矩阵定相百分比。参考1ffN组包括以下ffN:Dark G、T-LN3-AF550POPOS0、C-sPA-参考染料A、C-LN3-S07181、A-BL-参考染料A、A-BL-NR550S0。参考2ffN组包括以下ffN:Dark G、T-LN3-AF550POPOS0、C-sPA-参考染料B、C-LN3-S07181、A-sPA-BL-参考染料B、A-sPA-BL-NR550S0。C-sPA-参考染料B的结构是:

[0319] 

[0320] 据观察,在26个循环之后,包括双硼染料标记的ffA的ffN组的定相百分比小于0.1%。然而,当光剂量增加时,定相值也增加。

[0321] 图4A和图4B是在第26个循环时对于含有ffA-spA-I-3的掺入混合物获得的散点图。图4C和图4D是在第26个循环时对于含有ffA-spA-I-4的掺入混合物获得的散点图。据观察,对于用I-3标记的ffA,5倍的光剂量(5×)引起云状物的光漂白(参见图4B,右上象限)。然而,当氟基团被-OAc替代时,用I-4标记的ffA的光稳定性大大提高,如图4D,右上象限所示。

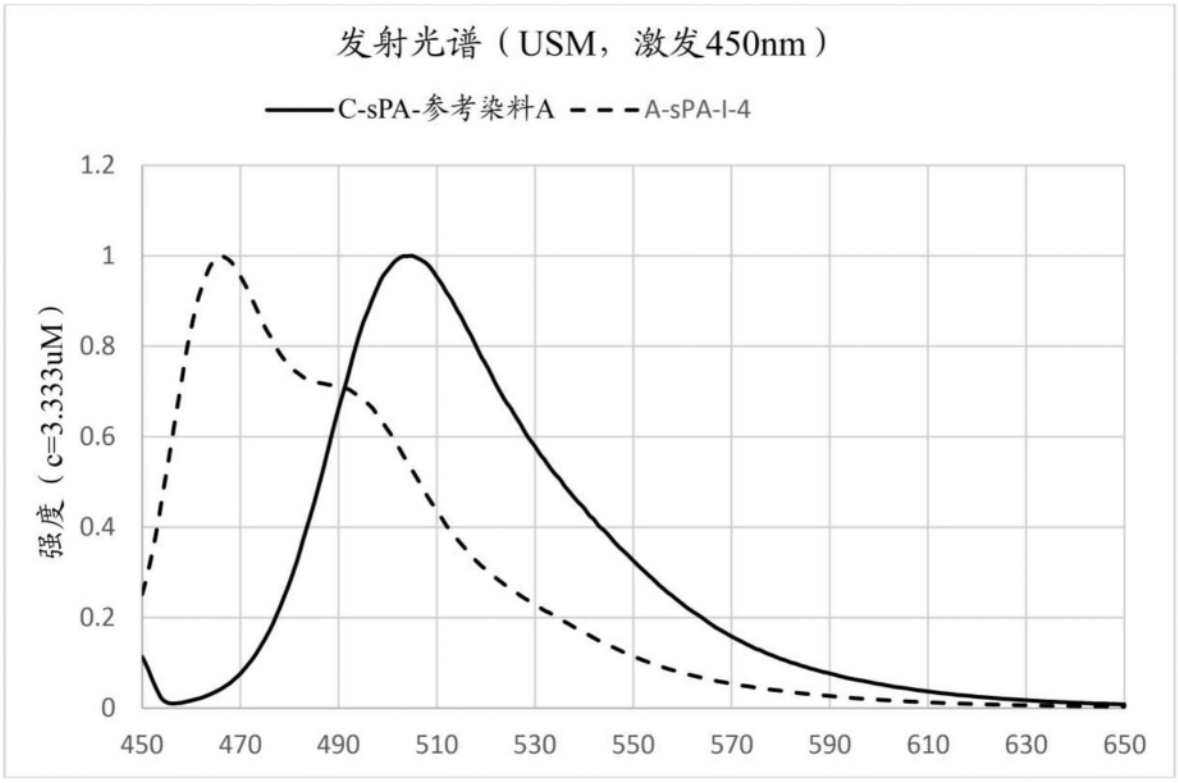

发射光谱（USM，激发450nm）
C-sPA-参考染料A
A-sPA-I-4
强度（c=3.333uM）
1.2
1
0.8
0.6
0.4
0.2
0
450
470
490
510
530
550
570
590
610
630
650

图1

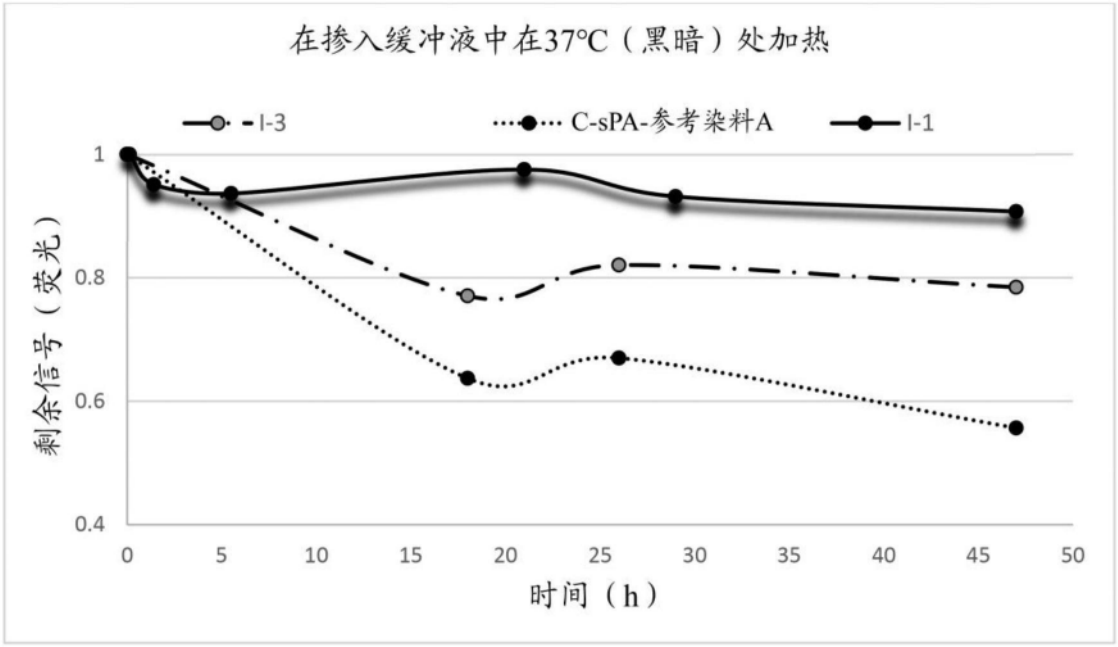

在掺入缓冲液中在37°C（黑暗）处加热
I-3
C-sPA-参考染料A
I-1
剩余信号（荧光）
1
0.8
0.6
0.4
0
5
10
15
20
25
30
35
40
45
50
时间（h）

图2

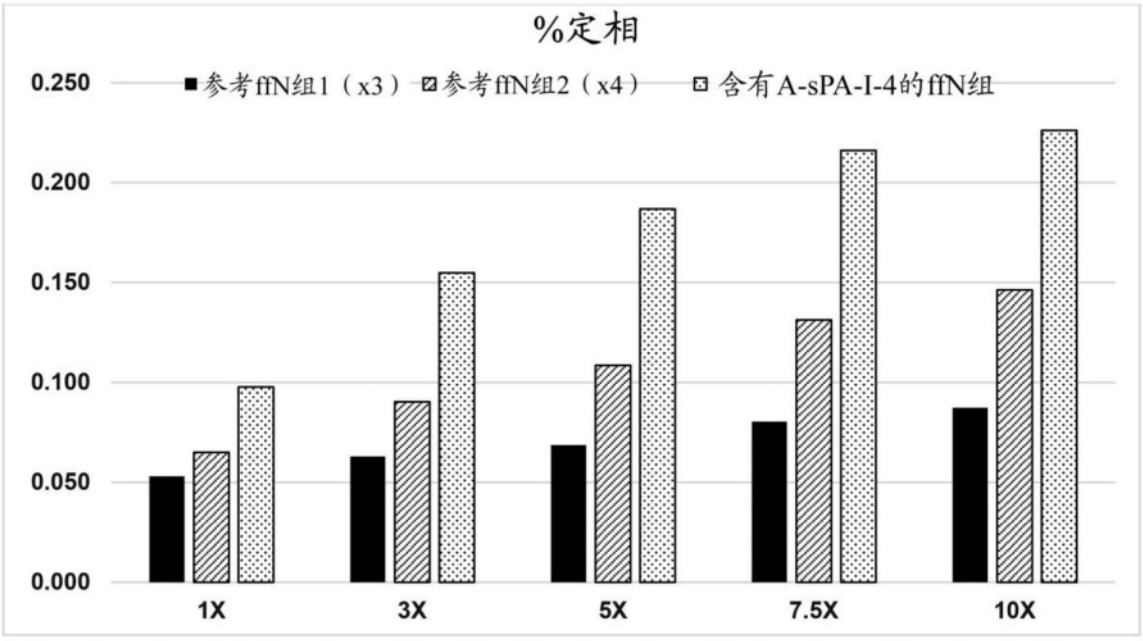

%定相
0.250
0.200
0.150
0.100
0.050
0.000
参考ffN组1（x3）
参考ffN组2（x4）
含有A-sPA-I-4的ffN组
1X
3X
5X
7.5X
10X

图3

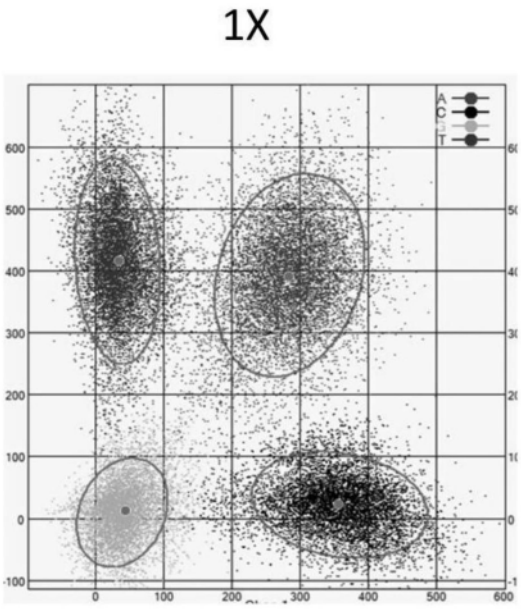

1X
A
C
T
600
500
400
300
200
100
0
-100
0
100
200
300
400
500
600

图4A

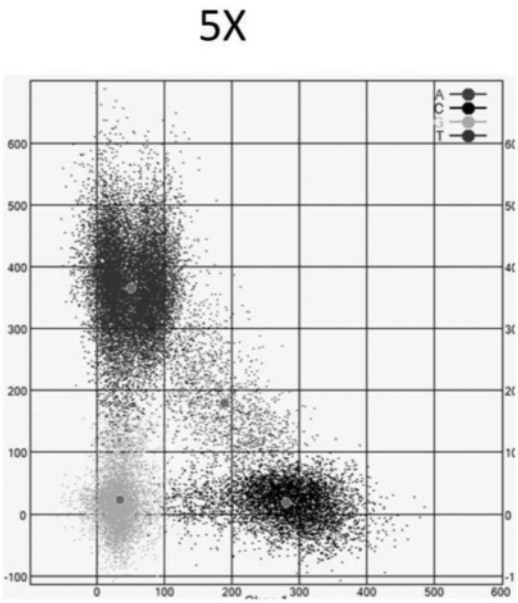

5X
A
C
T
600
500
400
300
200
100
0
-100
0
100
200
300
400
500
600

图4B

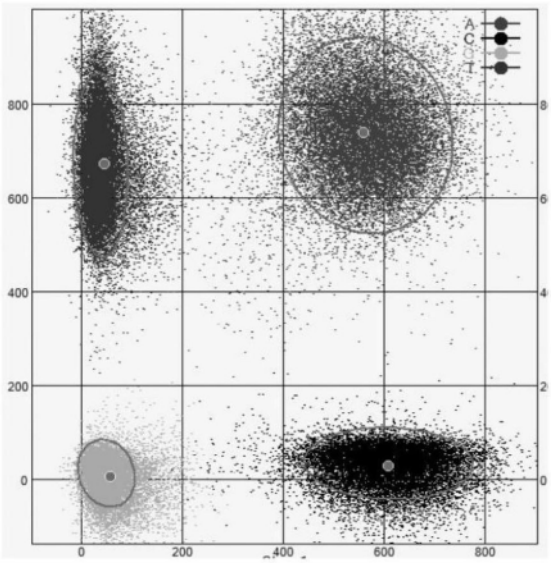

A
C
T
800
600
400
200
0
0
200
400
600
800

图4C

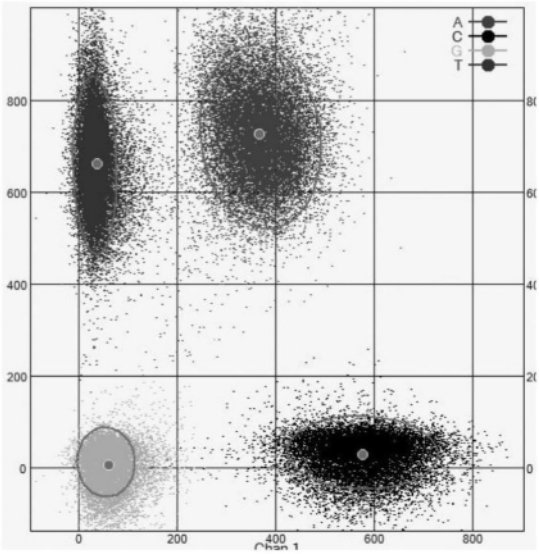


图4D