



(10) 授权公告号 CN 115996937 B

(45) 授权公告日 2025. 02. 25

(21) 申请号 202180045884.0

刘小海

(22) 申请日 2021.12.16

(74) 专利代理机构 北京市金杜律师事务所

11256

(65) 同一申请的已公布的文献号

申请公布号 CN 115996937 A

专利代理师 陈文平 王北南

(43) 申请公布日 2023.04.21

(51) Int.Cl.

(30) 优先权数据

63/127,044 2020.12.17 US

C09B 57/02 (2006.01)

G01N 21/64 (2006.01)

C12Q 1/686 (2006.01)

(85) PCT国际申请进入国家阶段日

2022.12.27

C09K 11/06 (2006.01)

C07H 21/00 (2006.01)

(86) PCT国际申请的申请数据

PCT/GB2021/053343 2021.12.16

C07H 19/14 (2006.01)

C07H 19/10 (2006.01)

(87) PCT国际申请的公布数据

W02022/129930 EN 2022.06.23

(56) 对比文件

US 2004260093 A1, 2004.12.23

US 2018201981 A1, 2018.07.19

(73) 专利权人 伊鲁米纳剑桥有限公司

地址 英国剑桥

审查员 杜姣

(72) 发明人 E·克雷西娜 A·弗朗卡斯

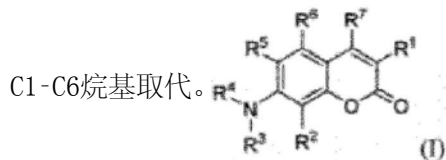
权利要求书4页 说明书46页 附图9页

(54) 发明名称

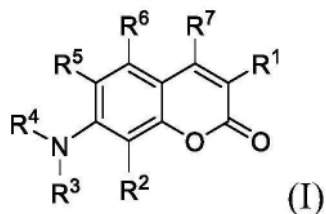
烷基吡啶正离子香豆素染料和在测序应用中的用途

(57) 摘要

本申请涉及式(I)的烷基吡啶正离子取代的香豆素染料以及它们作为荧光标记的用途。例如,这些染料可以用于标记核苷酸以用于核酸测序应用。其中R1是或并且其中R1被一个或多个



1. 一种式 (I) 的化合物、其盐或内消旋形式:



其中R¹是 并且其中R¹被一个或多个C₁烷基

取代;

R²、R⁵、R⁶和R⁷中的每一者独立地是H;

R³和R⁴中的每一者独立地是H、C₁-C₆烷基或经取代的C₁-C₆烷基,所述经取代的C₁-C₆烷基是被选自由以下组成的组的一个或多个取代基取代的C₁-C₆烷基:羧基(-C(O)OH)、羧酸根(-C(O)O⁻)、磺基(-SO₃H)、磺酸根(-SO₃⁻);

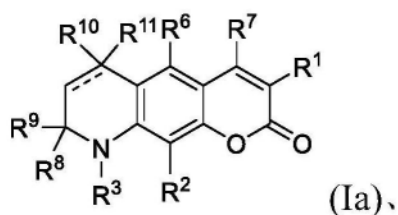
或者,R⁴和R⁵与它们所连接的原子一起形成环或环系,所述环或环系选自由以下组成的组:6元杂环基,其中所述6元杂环基是未取代的6元杂环基或经C₁-C₆烷基取代的6元杂环基;并且

R^a、R^b和R^c中的每一者独立地是C₁-C₆烷基或经取代的C₁-C₆烷基;其中R^a和R^b中的每一者独立地是C₁-C₆烷基或经选自由以下组成的组的一个或多个取代基取代的C₁-C₆烷基:羧基(-C(O)OH)、羧酸根(-C(O)O⁻)、磺基(-SO₃H)、磺酸根(-SO₃⁻);其中R^c是C₁-C₆烷基或经选自由以下组成的组的一个或多个取代基取代的C₁-C₆烷基:羧基(-C(O)OH)、磺基(-SO₃H)、磺酸根(-SO₃⁻)。

2. 根据权利要求1所述的化合物,其中R³是H,并且R⁴是C₁-C₆烷基或经取代的C₁-C₆烷基。

3. 根据权利要求1所述的化合物,其中R³和R⁴中的每一者独立地是C₁-C₆烷基。

4. 根据权利要求1所述的化合物,其中所述式 (I) 的化合物也由式 (Ia):

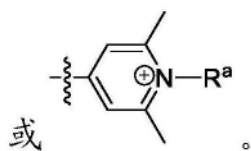


其盐或内消旋形式表示,其中:

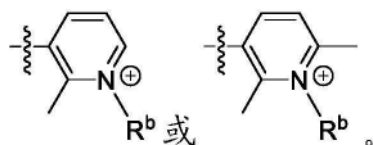
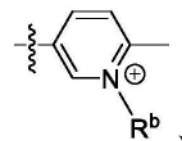
每个R⁸、R⁹、R¹⁰和R¹¹独立地是H、C₁-C₆烷基;并且

由实线和虚线 表示的键选自由单键和双键组成的组,前提条件是当 是双键时,则R¹¹不存在。

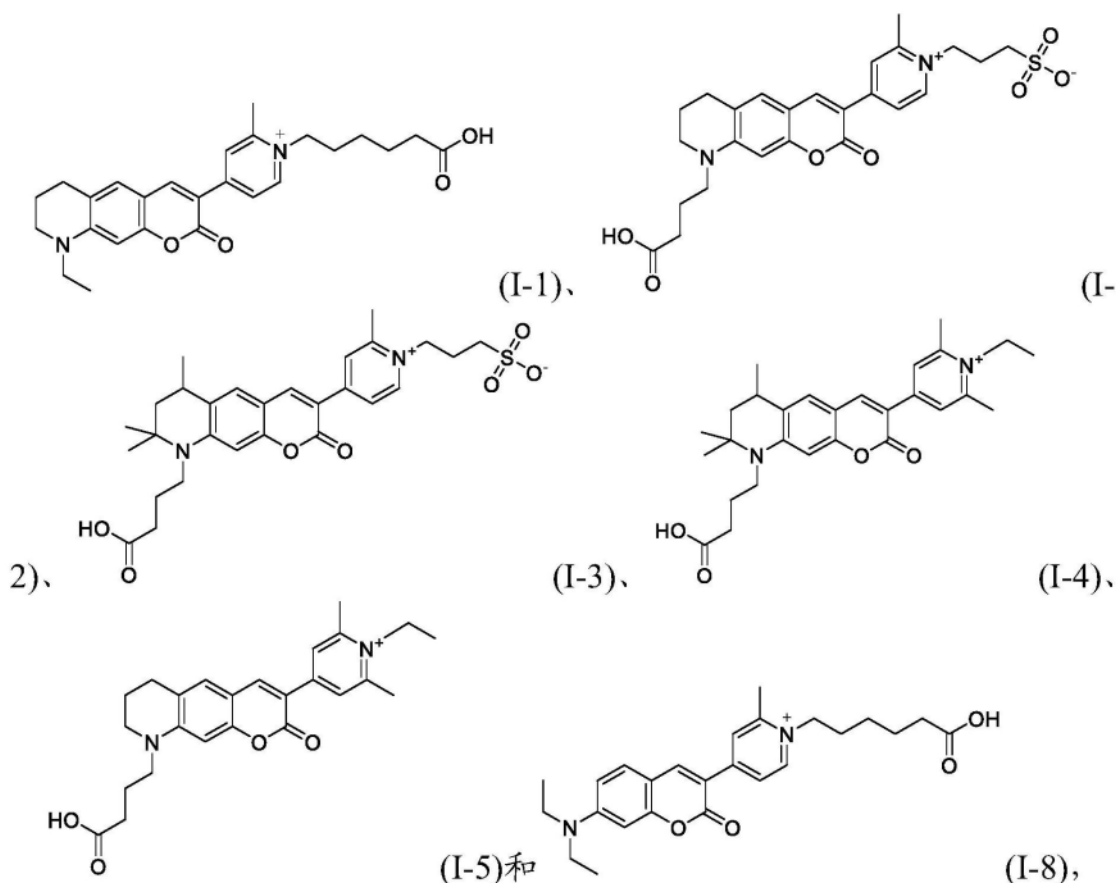
5. 根据权利要求1至4中任一项所述的化合物,其中R¹是



6. 根据权利要求1至4中任一项所述的化合物, 其中 R^1 是



7. 根据权利要求5或6所述的化合物, 其中每个 R^a 和 R^b 独立地是 C_1 - C_6 烷基。
8. 根据权利要求5或6所述的化合物, 其中每个 R^a 和 R^b 独立地是被羧基 ($-C(O)OH$)、羧酸根 ($-C(O)O^-$)、磺基 ($-SO_3H$) 或磺酸根 ($-SO_3^-$) 取代的 C_1 - C_6 烷基。
9. 根据权利要求4至8中任一项所述的化合物, 其中由实线和虚线 ----- 表示的键是双键。
10. 根据权利要求9所述的化合物, 其中 R^{10} 是 H 或 C_1 - C_6 烷基。
11. 根据权利要求10所述的化合物, 其中 R^{10} 是甲基。
12. 根据权利要求4至8中任一项所述的化合物, 其中由实线和虚线 ----- 表示的键是单键。
13. 根据权利要求12所述的化合物, 其中 R^{10} 是 H, 并且 R^{11} 是 C_1 - C_6 烷基。
14. 根据权利要求12所述的化合物, 其中 R^{10} 和 R^{11} 中的每一者是 H。
15. 根据权利要求4至14中任一项所述的化合物, 其中 R^8 和 R^9 中的每一者是 H。
16. 根据权利要求4至14中任一项所述的化合物, 其中 R^8 和 R^9 中的至少一者是 C_1 - C_6 烷基。
17. 根据权利要求16所述的化合物, 其中 R^8 和 R^9 中的每一者是 C_1 - C_6 烷基。
18. 根据权利要求17所述的化合物, 其中 R^8 和 R^9 中的每一者是甲基。
19. 根据权利要求4至18中任一项所述的化合物, 其中 R^3 是 C_1 - C_6 烷基。
20. 根据权利要求4至18中任一项所述的化合物, 其中 R^3 是经取代的 C_1 - C_6 烷基; 所述经取代的 C_1 - C_6 烷基是被选自由以下组成的组的一个或多个取代基取代的 C_1 - C_6 烷基: 羧基 ($-C(O)OH$)、羧酸根 ($-C(O)O^-$)、磺基 ($-SO_3H$)、磺酸根 ($-SO_3^-$)。
21. 根据权利要求20所述的化合物, 其中 R^3 是被选自由以下组成的组的一个或多个取代基取代的 C_1 - C_6 烷基: 羧基 ($-C(O)OH$)、羧酸根 ($-C(O)O^-$)、磺基 ($-SO_3H$)。
22. 根据权利要求21所述的化合物, 其中 R^3 是被羧基取代的 C_1 - C_6 烷基。
23. 根据权利要求1所述的化合物, 所述化合物选自由以下组成的组:



以及它们的盐和内消旋形式。

24. 一种核苷酸或寡核苷酸, 所述核苷酸或寡核苷酸用根据权利要求1至23中任一项所述的化合物进行标记。

25. 根据权利要求24所述的核苷酸或寡核苷酸, 其中所述化合物通过式 (Ia) 的 R^{8a} 、 R^{8b} 或 R^{8c} 的羧基基团连接到所述核苷酸或寡核苷酸。

26. 根据权利要求24所述的核苷酸或寡核苷酸, 其中所述化合物通过式 (I) 的 R^3 或 R^4 的羧基基团连接到所述核苷酸或寡核苷酸。

27. 根据权利要求24至26中任一项所述的核苷酸或寡核苷酸, 其中所述化合物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。

28. 根据权利要求24至27中任一项所述的核苷酸或寡核苷酸, 其中标记的核苷酸或寡核苷酸还包含共价连接到所述核苷酸的核糖或脱氧核糖的3' OH封闭基团。

29. 根据权利要求24所述的核苷酸或寡核苷酸, 其中所述核苷酸或寡核苷酸是与靶多核苷酸的至少一部分杂交的寡核苷酸。

30. 根据权利要求29所述的寡核苷酸, 其中所述靶多核苷酸固定在固体载体上。

31. 根据权利要求30所述的寡核苷酸, 其中所述固体载体包含多个固定的靶多核苷酸的阵列。

32. 一种试剂盒, 所述试剂盒包括用根据权利要求1至23中任一项的第一化合物标记的第一核苷酸。

33. 根据权利要求32所述的试剂盒, 所述试剂盒还包括用第二化合物标记的第二核苷酸, 其中所述第二化合物不同于所述第一标记的核苷酸的所述第一化合物。

34. 根据权利要求33所述的试剂盒,其中所述第一标记的核苷酸和所述第二标记的核苷酸能够使用第一光源波长激发。

35. 根据权利要求33或34所述的试剂盒,所述试剂盒还包括第三核苷酸,其中所述第三核苷酸用不同于所述第一化合物和所述第二化合物的第三化合物标记,并且其中所述第一标记的核苷酸和所述第三标记的核苷酸能够使用第二光源波长激发。

36. 根据权利要求35所述的试剂盒,所述试剂盒还包括第四核苷酸,并且其中所述第四核苷酸未标记。

37. 根据权利要求32至36中任一项所述的试剂盒,其中所述第一标记的核苷酸、所述第二标记的核苷酸和所述第三标记的核苷酸中的每种标记的核苷酸具有能够在单个检测通道中检测的发射光谱。

38. 根据权利要求32至37中任一项所述的试剂盒,所述试剂盒还包括DNA聚合酶和一种或多种缓冲液组合物。

39. 一种确定靶多核苷酸的序列的方法,所述方法包括:

(a) 使引物多核苷酸/靶多核苷酸复合物与一种或多种标记的核苷酸接触,其中所述标记的核苷酸中的至少一种标记的核苷酸是根据权利要求24至28中任一项所述的核苷酸,并且其中所述引物多核苷酸与所述靶多核苷酸的至少一部分互补;

(b) 将标记的核苷酸掺入所述引物多核苷酸中以产生延伸的引物多核苷酸/靶多核苷酸复合物;以及

(c) 对所述延伸的引物多核苷酸/靶多核苷酸复合物进行一个或多个荧光测量以确定所掺入的核苷酸的身份。

40. 根据权利要求39所述的方法,其中所述引物多核苷酸/靶多核苷酸复合物通过使所述靶多核苷酸和与所述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。

41. 根据权利要求39或40所述的方法,其中所述方法还包括(d) 从掺入所述引物多核苷酸中的所述核苷酸去除标记和3'封闭基团。

42. 根据权利要求40所述的方法,其中所述方法还包括(e) 从所述延伸的引物多核苷酸洗掉所去除的标记和所述3'封闭基团。

43. 根据权利要求42所述的方法,其中所述方法还包括重复步骤(a)至(e),直到确定模板多核苷酸链的至少一部分的序列。

44. 根据权利要求43所述的方法,其中所述步骤(a)至(e)重复至少50次。

45. 根据权利要求39至44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所述引物多核苷酸中的所述核苷酸的所述标记和所述3'封闭基团。

46. 根据权利要求39至45中任一项所述的方法,其中在自动化测序仪器上进行所述方法,并且其中所述自动化测序仪器包括以不同波长操作的两个光源。

烷基吡啶正离子香豆素染料和在测序应用中的用途

技术领域

[0001] 本公开涉及烷基吡啶正离子取代的香豆素衍生物以及它们作为荧光标记的用途。具体地,这些化合物可以作为核苷酸标记用于核酸测序应用。

背景技术

[0002] 对携带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³²P。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,如它们的费用、有限的保质期、灵敏度不足和更重要的安全考虑因素。消除对放射性标记的需求降低了安全风险以及例如试剂处理相关的环境影响和成本。以非限制性示例的方式,适合于非放射性荧光检测的方法包括自动化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。

[0003] 对于许多应用而言,希望采用多个光谱可区分的荧光标记,以实现多种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。

[0004] 然而,多重荧光检测可能存在问题,并且存在许多限制适当荧光标记的选择的重要因素。首先,在给定的应用中可能难以找到具有可显著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分隔很广,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,并且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终考虑因素是荧光染料必须与试剂化学性质相容的程度,例如DNA合成溶剂和试剂、缓冲液、聚合酶和连接酶。

[0005] 随着测序技术的进展,需要满足所有上述限制并且特别适合于高通量分子方法,如固相测序等的另外的荧光染料化合物、它们的核酸缀合物和多个染料组。

[0006] 具有改进的荧光特性(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。

[0007] 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 $\Delta\lambda$ 约100nm,则分辨率将更好(约 $\Delta 50\text{nm}$ /(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%。

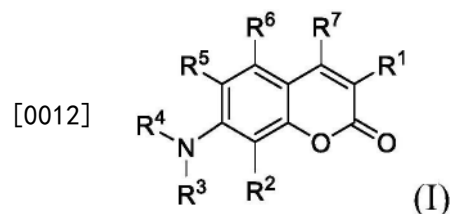
[0008] 某些核酸测序方法采用激光来激发染料标记的核苷酸并对其进行检测。这些仪器使用较长波长的光,诸如红色激光,连同能够在660nm处激发的适当染料。为了在保持可用分辨率的同时检测更密集堆积的核酸测序簇,可以使用较短波长的蓝色光源(450nm至460nm)。在这种情况下,光学分辨率将不受较长波长红色荧光染料的发射波长的限制,而是受能够由次长波长光源(例如,由532nm的“绿色激光”)激发的染料的发射的限制。因此,需要用于测序应用中的荧光检测的蓝色染料标记。

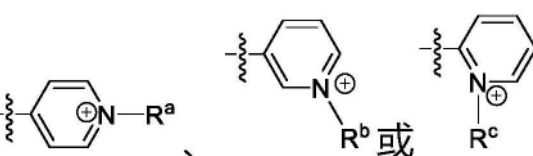
[0009] 香豆素染料家族由于它们显著的光谱特性而引起化学家的注意。然而,只有几种具有大斯托克斯位移(large Stokes shift, LSS)的光稳定的荧光染料是可商购获得的。这些染料中的大部分染料还含有香豆素片段作为支架。因此,设计具有定制的吸附波长和荧光斯托克斯位移、具有良好的稳定性的染料仍然是染料开发中的关键挑战。

发明内容

[0010] 本文描述了适合于核苷酸标记的具有长斯托克斯位移和改进的荧光强度和化学稳定性的烷基吡啶正离子取代的香豆素染料。这些香豆素染料在蓝光和绿光激发下都具有强荧光(例如,这些香豆素染料可以具有约450nm至约530nm、约460nm至约520nm、约475nm至约510nm或约490nm至约500nm的吸收波长)。

[0011] 本公开的一些方面涉及式(I)的化合物或其盐或内消旋形式:



[0013] 其中R¹是  并且其中R¹被一个或多个C₁-C₆

烷基取代;

[0014] 每个R²、R⁵和R⁷独立地是H、C₁-C₆烷基、经取代的C₁-C₆烷基、C₁-C₆烷氧基、C₂-C₆烯基、C₂-C₆炔基、C₁-C₆卤代烷基、C₁-C₆卤代烷氧基、(C₁-C₆烷氧基)C₁-C₆烷基、任选地经取代的氨基、氨基(C₁-C₆烷基)、卤基、氰基、羟基、羟基(C₁-C₆烷基)、硝基、磺酰基、磺基、亚磺酰基、磺酸根、S-亚磺酰氨基或N-亚磺酰氨基;

[0015] R³和R⁴中的每一者独立地是H、C₁-C₆烷基或经取代的C₁-C₆烷基;

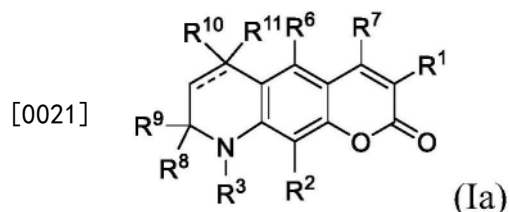
[0016] 可替代地,R²和R³与它们所连接的原子一起形成环或环系,该环或环系选自由以下组成的组:任选地经取代的5元至10元杂芳基或任选地经取代的5元至10元杂环基;

[0017] 可替代地,R⁴和R⁵与它们所连接的原子一起形成环或环系,该环或环系选自由以下组成的组:任选地经取代的5元至10元杂芳基或任选地经取代的5元至10元杂环基;

[0018] R^6 是 H、 C_1-C_6 烷基、经取代的 C_1-C_6 烷基或任选地经取代的 C_6-C_{10} 芳基；并且

[0019] R^a 、 R^b 和 R^c 中的每一者独立地是 C_1-C_6 烷基或经取代的 C_1-C_6 烷基。

[0020] 在一些实施方案中，该式 (I) 的化合物也由式 (Ia) 或其盐或内消旋形式表示：



[0022] 其中每个 R^8 、 R^9 、 R^{10} 和 R^{11} 独立地是 H、 C_1-C_6 烷基、经取代的 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 烷氧基) C_1-C_6 烷基、任选地经取代的氨基、氨基 (C_1-C_6 烷基)、卤基、氰基、羟基、羟基 (C_1-C_6 烷基)、硝基、磺酰基、磺基、亚磺酰基、磺酸根、S-亚磺酰氨基或 N-亚磺酰氨基；并且由实线和虚线-----表示的键选自由单键和双键组成的组，前提条件是当-----是双键时，则 R^{11} 不存在。

[0023] 在式 (I) 或式 (Ia) 的化合物的一些实施方案中， R^1 (例如， R^a 、 R^b 或 R^c) 包含羧基基团 ($-C(O)OH$)。在其他实施方案中， R^3 或 R^4 包含羧基基团。

[0024] 在一些方面，本公开的化合物被如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。标记或缀合可以通过羧基基团进行，该羧基基团可以使用本领域已知的方法与部分 (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。

[0025] 本公开的一些其他方面涉及包含连接基基团以实现例如与底物部分共价连接的染料化合物。可以在染料的任何位置处进行连接，包括在任何 R 基团处。在一些实施方案中，可以通过式 (I) 的 R^1 (例如， R^a 、 R^b 或 R^c) 或者 R^3 或 R^4 进行连接。在一些另外的实施方案中，可以通过式 (Ia) 的 R^1 (例如， R^a 、 R^b 或 R^c) 或 R^3 进行连接。

[0026] 本公开的一些另外的方面提供了由下式定义的标记的核苷或核苷酸化合物：

[0027] N-L-染料

[0028] 其中 N 是核苷或核苷酸；

[0029] L 是任选的连接基部分；并且

[0030] 染料是根据本公开的式 (I) 或式 (Ia) 的荧光化合物的部分，其中该式 (I) 或式 (Ia) 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合。

[0031] 本公开的一些另外的方面涉及用式 (I) 或式 (Ia) 的化合物标记的核苷酸或寡核苷酸。

[0032] 本公开的一些另外的方面涉及一种试剂盒，该试剂盒包括可以用于各种免疫学测定、寡核苷酸或核酸标记或用于 DNA 边合成边测序的染料化合物 (游离或呈标记形式)。在又另一方面，本公开提供了包括染料“组”的试剂盒，这些染料“组”特别适合于在自动化仪器平台上的边合成边测序的循环。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。

[0033] 本公开的另外的方面是一种确定靶多核苷酸的序列的方法，该方法包括：

[0034] (a) 使引物多核苷酸/靶多核苷酸复合物与一种或多种标记的核苷酸 (例如，A、G、C

和T或dATP、dGTP、dCTP和dTTP)接触,其中所述标记的核苷酸中的至少一种标记的核苷酸是本文所述的用式(I)或式(Ia)的烷基吡啶正离子取代的香豆素染料标记的核苷酸,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;

[0035] (b) 将标记的核苷酸掺入该引物多核苷酸/靶核苷酸复合物中以产生延伸的引物多核苷酸/靶核苷酸复合物;以及

[0036] (c) 对该延伸的引物多核苷酸/靶核苷酸复合物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

[0037] 图1A和图1B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与参考染料A相比,香豆素染料I-1的荧光发射光谱。

[0038] 图2A和图2B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与参考染料C相比,香豆素染料I-5的荧光发射光谱。

[0039] 图3A和图3B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与参考染料B相比,香豆素染料I-8的荧光发射光谱。

[0040] 图4A和图4B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与用参考染料A标记的完全官能化A核苷酸(ffA)相比,用香豆素染料I-1标记的ffA的荧光发射光谱。

[0041] 图5A和图5B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与用参考染料C标记的ffA相比,用香豆素染料I-5标记的ffA的荧光发射光谱。

[0042] 图6A和图6B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与用参考染料B标记的ffA相比,用香豆素染料I-8标记的ffA的荧光发射光谱。

[0043] 图7A和图7B展示了分别在450nm和520nm激发波长下在缓冲溶液中,与用参考染料D标记的ffA相比,用香豆素染料I-3标记的ffA的荧光发射光谱。

[0044] 图8A和图8B分别示出了针对含有用香豆素染料I-1标记的ffA的掺入混合物获得的散点图以及从含有用参考染料A标记的ffA的掺入混合物获得的散点图。

[0045] 图8C和图8D分别示出了针对含有用香豆素染料I-3标记的ffA的掺入混合物获得的散点图以及从含有用参考染料D标记的ffA的掺入混合物获得的散点图。

具体实施方式

[0046] 本公开的实施方案涉及具有增强的荧光强度和长斯托克斯位移的烷基吡啶正离子取代的香豆素染料。这些香豆素染料还具有宽的激发波长,并且可以由蓝色光源和绿色光源两者激发。在一些实施方案中,本文所述的烷基吡啶正离子香豆素染料可以用于具有两通道CMOS检测(绿光激发和蓝光激发)的因美纳公司(Illumina)的iSeq™平台中。

[0047] 定义

[0048] 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。

[0049] 应注意,如本说明书和所附权利要求中所使用的,单数形式“一个(a)”、“一种(an)”和“该(the)”包括复数指示物,除非明确并且确切地限于一个指示物。对于本领域技术人员显而易见的是,在不脱离本发明教导的精神或范围的情况下,可以对本文所述的各种实施方案进行各种修改和变化。因此,旨在本文所述的各种实施方案覆盖在所附权利要

求书及其等效物的范围内的其他修改和变化。

[0050] 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,但是也可以包括另外的步骤。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。

[0051] 如本文所用,常见的有机缩写定义如下:

[0052] °C 以摄氏度为单位的温度

[0053] dATP 脱氧腺苷三磷酸

[0054] dCTP 脱氧胞苷三磷酸

[0055] dGTP 脱氧鸟苷三磷酸

[0056] dTTP 脱氧胸苷三磷酸

[0057] ddNTP 双脱氧核苷三磷酸

[0058] ffA 完全官能化A核苷酸

[0059] ffC 完全官能化C核苷酸

[0060] ffG 完全官能化G核苷酸

[0061] ffN 完全官能化核苷酸

[0062] ffT 完全官能化T核苷酸

[0063] h 小时

[0064] RT 室温

[0065] SBS 边合成边测序

[0066] 如本文所用,术语“阵列”是指一群不同的探针分子,该一群不同的探针分子连接到一种或多种底物,使得不同的探针分子可以根据相对位置彼此区分。阵列可以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。可替代地或另外地,阵列可以包括各自带有不同探针分子的单独的底物,其中可以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。其中分离的底物位于表面上的示例性阵列包括但不限于在孔中包括珠粒的阵列,如例如在美国专利第6,355,431B1号、US 2002/0102578和PCT公开第W0 00/63437号中所述。例如,在美国专利第6,524,793号中描述了可以本发明中用于例如使用微流体装置,如荧光激活细胞分选器(FACS)区分液体阵列中的珠粒的示例性形式。可以用于本发明的阵列的另外的示例包括但不限于美国专利第5,429,807号;第5,436,327号;第5,561,071号;第5,583,211号;第5,658,734号;第5,837,858号;第5,874,219号;第5,919,523号;第6,136,269号;第6,287,768号;第6,287,776号;第6,288,220号;第6,297,006号;第6,291,193号;第6,346,413号;第6,416,949号;第6,482,591号;第6,514,751以及第6,610,482号;以及W0 93/17126;W0 95/11995;W0 95/35505;EP 742287;以及EP 799897中所描述的那些阵列。

[0067] 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。

[0068] 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。

[0069] 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,“C₁至C₄烷基”基团是指具有1至4个碳的所有烷基基团,即CH₃-、CH₃CH₂-、CH₃CH₂CH₂-、(CH₃)₂CH-、CH₃CH₂CH₂CH₂-、CH₃CH₂CH(CH₃)-和(CH₃)₃C-C₃至C₄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“C₁-C₆”包括C₁、C₂、C₃、C₄、C₅和C₆,以及由这两个数字中任一个数字定义的范围。例如,C₁-C₆烷基包括C₁烷基、C₂烷基、C₃烷基、C₄烷基、C₅烷基和C₆烷基、C₂-C₆烷基、C₁-C₃烷基等。类似地,C₂-C₆烯基包括C₂烯基、C₃烯基、C₄烯基、C₅烯基和C₆烯基、C₂-C₅烯基、C₃-C₄烯基等;并且C₂-C₆炔基包括C₂炔基、C₃炔基、C₄炔基、C₅炔基和C₆炔基、C₂-C₅炔基、C₃-C₄炔基等。C₃-C₈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C₃-C₇环烷基或C₅-C₆环烷基。

[0070] 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,“C₁₋₆烷基”或“C₁-C₆烷基”表示烷基链中存在一个至六个碳原子,即,烷基链选自自由甲基、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。

[0071] 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),如“C₁₋₉烷氧基”或“C₁-C₉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。

[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,“C₂-C₆烯基”或“C₂₋₆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,“C₂₋₆炔基”或“C₂-C₆炔基”表示炔基链中存在两个至六个碳原子,即,炔基链选自乙炔基、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“C₆-C₁₀芳基”、“C₆或C₁₀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、蒽基和蒽基。

[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“C₇₋₁₄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,C₁₋₆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、呋喃基、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,C₁₋₆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“C₃₋₆碳环基”、“C₃-C₆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吡啶基、吡啶基、呋喃基、噻吩基、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-噻嗪基、三噻烷基、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、异噻唑啉基、异噻唑烷基、噻唑啉基、噻唑烷基、噻唑烷酮基、噻唑啉基、噻唑烷基、1,3-氧硫杂环戊烷基、二氢吡啶基、异二氢吡啶基、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,如C₂-C₈烷氧基烷基或(C₁-C₆烷氧基)C₁-C₆烷基,例如, -(CH₂)₁₋₃-OCH₃。

[0083] “O-羧基”基团是指其中R选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀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、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀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、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5至10元杂芳基和3至10元杂环基的“-SO₂R”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“-SO₃H”基团。

[0088] “磺酸根”基团是指“-SO₃⁻”基团。

[0089] “硫酸根”基团是指“-SO₄⁻”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、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5至10元杂芳基和3至10元杂环基的“-SO₂NR_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、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5至10元杂芳基和3至10元杂环基的“-N(R_A)SO₂R_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, -NH₂)。

[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,如“C₂-C₈烷氧基烷基”等。

[0097] 当基团被描述为“任选地经取代的”时,该基团可以是未经取代的或经取代的。同样,当基团被描述为“经取代的”时,取代基可以选自所指示的取代基中的一个或多个取代基。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, -CF₃)、卤基(C_1 - C_6)烷氧基(例如, -OCF₃)、 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、-SO₃H、磺酸根、硫酸根、亚磺酸基、-OSO₂C₁-C₄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,这些基团各自独立地被一个或多个选自由以下组成的组的取代基取代:卤基、-CN、-SO₃⁻、-OSO₃⁻、-SO₃H、-SR^A、-OR^A、-NR^BR^C、氧代、

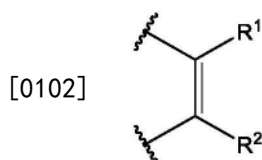
[0098] -CONR^BR^C、-SO₂NR^BR^C、-COOH和-COOR^B,其中R^A、R^B和R^C各自独立地选自H、烷基、经取代的烷基、烯基、经取代的烯基、炔基、经取代的炔基、芳基和经取代的芳基。

[0099] 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电离形式存在,例如, -CO₂⁻、-SO₃⁻或-O-SO₃⁻。如果化合物包含带正电荷或带负电荷的取代基基团,例如,SO₃⁻,

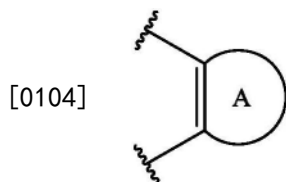
则其还可以包含带负电荷或带正电荷的抗衡离子,使得该化合物整体为中性的。在其他方面,该化合物可以以盐形式存在,其中抗衡离子由共轭酸或碱提供。

[0100] 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 $-\text{CH}_2-$ 、 $-\text{CH}_2\text{CH}_2-$ 、 $-\text{CH}_2\text{CH}(\text{CH}_3)\text{CH}_2-$ 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”。

[0101] 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:



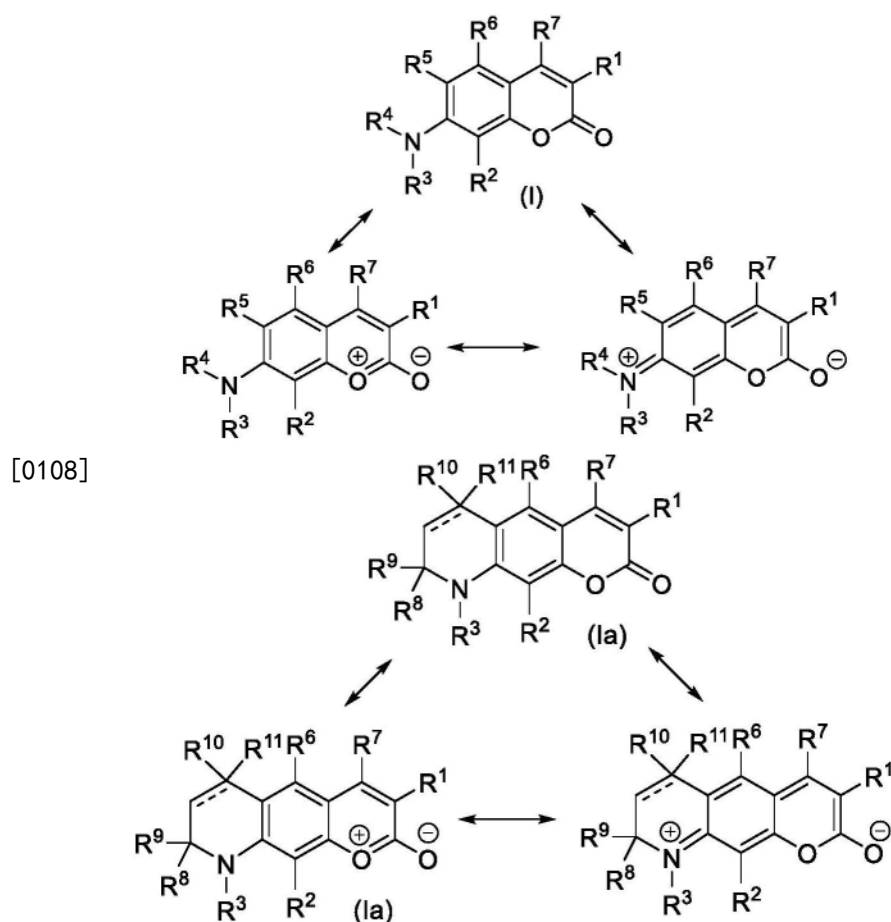
[0103] 并且 R^1 和 R^2 被限定为选自氢和烷基组成的组,或者 R^1 和 R^2 连同它们所附接的原子一起形成芳基或碳环基,这意味着 R^1 和 R^2 可以选自氢或烷基,或者替代性地,该子结构具有结构:



[0105] 其中A为含有所描绘的双键的芳环或碳环基。

[0106] 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 $-\text{AE}-$ 或 $\text{A}-\text{E}$ 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 $\text{A}-\text{L}-\text{E}$ 并且L定义为任选地存在的连接基部分;当L不存在(或缺失)时,此类基团或取代基等于 $\text{A}-\text{E}$ 。

[0107] 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。本文所述的香豆素化合物由单一结构表示,但同样可以被显示为任何相关的内消旋形式。以下分别示出了式(I)和式(Ia)的示例性内消旋结构:



[0109] 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。

[0110] 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。含氮杂环碱基可以为嘌呤、脱氮嘌呤或嘧啶碱基。嘌呤碱基包括腺嘌呤(A)和鸟嘌呤(G)以及它们的经修饰的衍生物或类似物,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)以及它们的修饰衍生物或类似物。脱氧核糖的C-1原子与嘧啶的N-1或嘌呤的N-9键合。

[0111] 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。

[0112] 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。任选地经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)。

[0113] 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。

[0114] 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的标记和/或核苷或核苷酸部分的位置。

[0115] 如本文所用,“衍生物”或“类似物”意指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、烷基磷酸酯键、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。

[0116] 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

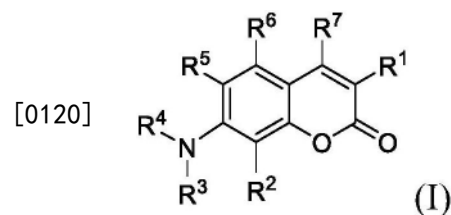
其质子化形式(例如, $\text{O}=\text{P}(\text{OH})(\text{O}-\text{R})$ 和 $\text{O}=\text{P}(\text{OH})_2(\text{O}-\text{R})$)。如本文所用,术语“单磷酸”、“二磷酸”

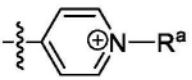
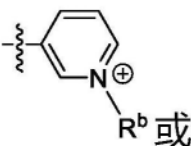
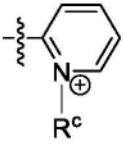
和“三磷酸”以本领域技术人员所理解的其普通含义使用,并且包括质子化形式。

[0117] 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个循环。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边合成边测序(SBS)期间存在痕量的未保护的或未封闭的3'-OH核苷酸引起。未保护的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。

[0118] 式(I)的荧光染料

[0119] 本公开的一些方面涉及式(I)的香豆素染料以及其盐和内消旋形式:



[0121] 其中 R^1 是 、或 ，并且其中 R^1 被一个或多个 C_1-C_6

烷基取代；

[0122] 每个 R^2 、 R^5 和 R^7 独立地是H、 C_1-C_6 烷基、经取代的 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 烷氧基) C_1-C_6 烷基、任选地经取代的氨基、氨基(C_1-C_6 烷基)、卤基、氰基、羟基、羟基(C_1-C_6 烷基)、硝基、磺酰基、磺基、亚磺酸基、磺酸根、S-亚磺酰氨基或N-亚磺酰氨基；

[0123] R^3 和 R^4 中的每一者独立地是H、 C_1-C_6 烷基或经取代的 C_1-C_6 烷基；

[0124] 可替代地， R^2 和 R^3 与它们所连接的原子一起形成环或环系，该环或环系选自由以下组成的组：任选地经取代的5元至10元杂芳基或任选地经取代的5元至10元杂环基；

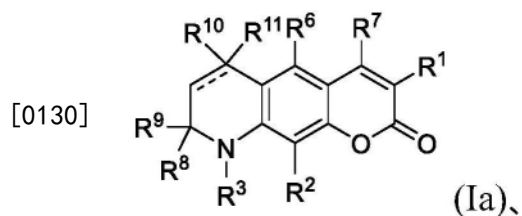
[0125] 可替代地， R^4 和 R^5 与它们所连接的原子一起形成环或环系，该环或环系选自由以下组成的组：任选地经取代的5元至10元杂芳基或任选地经取代的5元至10元杂环基；

[0126] R^6 是H、 C_1-C_6 烷基、经取代的 C_1-C_6 烷基或任选地经取代的 C_6-C_{10} 芳基；并且

[0127] R^a 、 R^b 和 R^c 中的每一者独立地是 C_1-C_6 烷基或经取代的 C_1-C_6 烷基。

[0128] 在式(I)的化合物的一些实施方案中， R^3 和 R^4 中的至少一者是H。在一些另外的实施方案中， R^3 和 R^4 都是H。在其他实施方案中， R^3 是H，并且 R^4 是 C_1-C_6 烷基或经取代的 C_1-C_6 烷基。在其他实施方案中， R^3 和 R^4 中的每一者独立地是 C_1-C_6 烷基或经取代的 C_1-C_6 烷基。经取代的 C_1-C_6 烷基包括但不限于甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基、正己基等，这些烷基被以下如以下等一个或多个取代基取代：羧基、羧酸根($-C(O)O^-$)、磺基($-SO_3H$)、磺酸根($-SO_3^-$)、硫酸根($-O-SO_3^-$)、任选地经取代的氨基(如Boc保护的氨基基团)、 $-C(O)OR^{12}$ 或 $-C(O)NR^{13}R^{14}$ ，其中 R^{12} 是任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基，并且其中 R^{13} 和 R^{14} 中的每一者独立地是H、任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基。在一个实施方案中， R^3 和 R^4 中的每一者是乙基。在另一个实施方案中， R^3 是H，并且 R^4 是被羧基取代的正丙基。

[0129] 式(I)的化合物的一些实施方案也由式(Ia)表示，其中式(I)的 R^4 和 R^5 与它们所连接的原子一起形成具有以下结构的任选地经取代的6元杂环基：



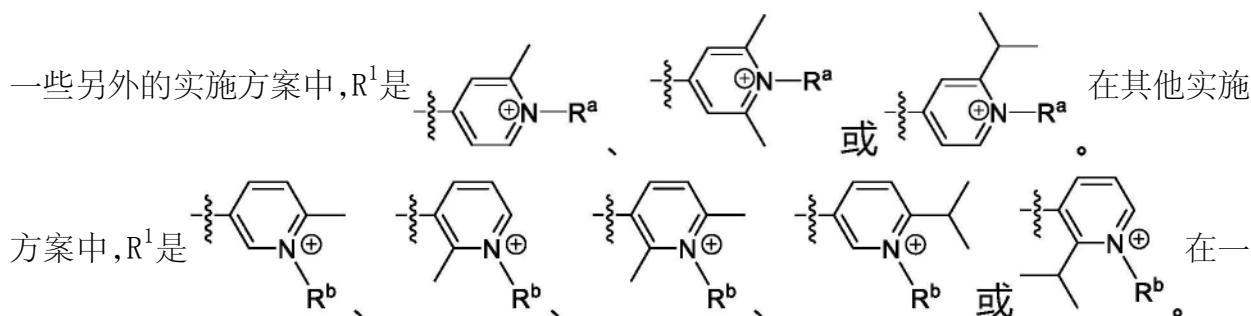
[0131] 其盐或内消旋形式：

[0132] 每个 R^8 、 R^9 、 R^{10} 和 R^{11} 独立地是H、 C_1-C_6 烷基、经取代的 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 烷氧基) C_1-C_6 烷基、任选地经取代的氨基、氨基(C_1-C_6 烷基)、卤基、氰基、羟基、羟基(C_1-C_6 烷基)、硝基、磺酰基、磺基、亚磺酸基、磺酸根、S-亚磺酰氨基或N-亚磺酰氨基；

[0133] 由实线和虚线-----表示的键选自由单键和双键组成的组，前提条件是当-----是

双键时,则 R^{11} 不存在。

[0134] 在式(I)或(Ia)的化合物的一些实施方案中, R^1 被一个 C_1-C_6 烷基(例如,甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基或正己基)取代。在一些其他实施方案中, R^1 独立地被两个 C_1-C_6 烷基取代。在其他实施方案中, R^1 独立地被三个 C_1-C_6 烷基取代。在



些此类实施方案中,每个 R^a 和 R^b 独立地是 C_1-C_6 烷基(例如,甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基或正己基等)。在一些另外的实施方案中,每个 R^a 和 R^b 是独立地经取代的 C_1-C_6 烷基(例如,甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基或正己基,这些烷基被如以下等一个或多个取代基取代:羧基、羧酸根($-C(O)O^-$)、磺基($-SO_3H$)、磺酸根($-SO_3^-$)、硫酸根($-O-SO_3^-$)、任选地经取代的氨基(如Boc保护的氨基基团)、 $-C(O)OR^{12}$ 或 $-C(O)NR^{13}R^{14}$,其中 R^{12} 是任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基,并且其中 R^{13} 和 R^{14} 中的每一者独立地是H、任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基。例如, R^a 和 R^b 中的每一者独立地是被羧基、羧酸根、磺基或磺酸根取代的正丙基、正丁基或正戊基。在一些实施方案中,取代位于直链 C_2 烷基、 C_3 烷基、 C_5 烷基、 C_6 烷基或 C_6 烷基的末端处。

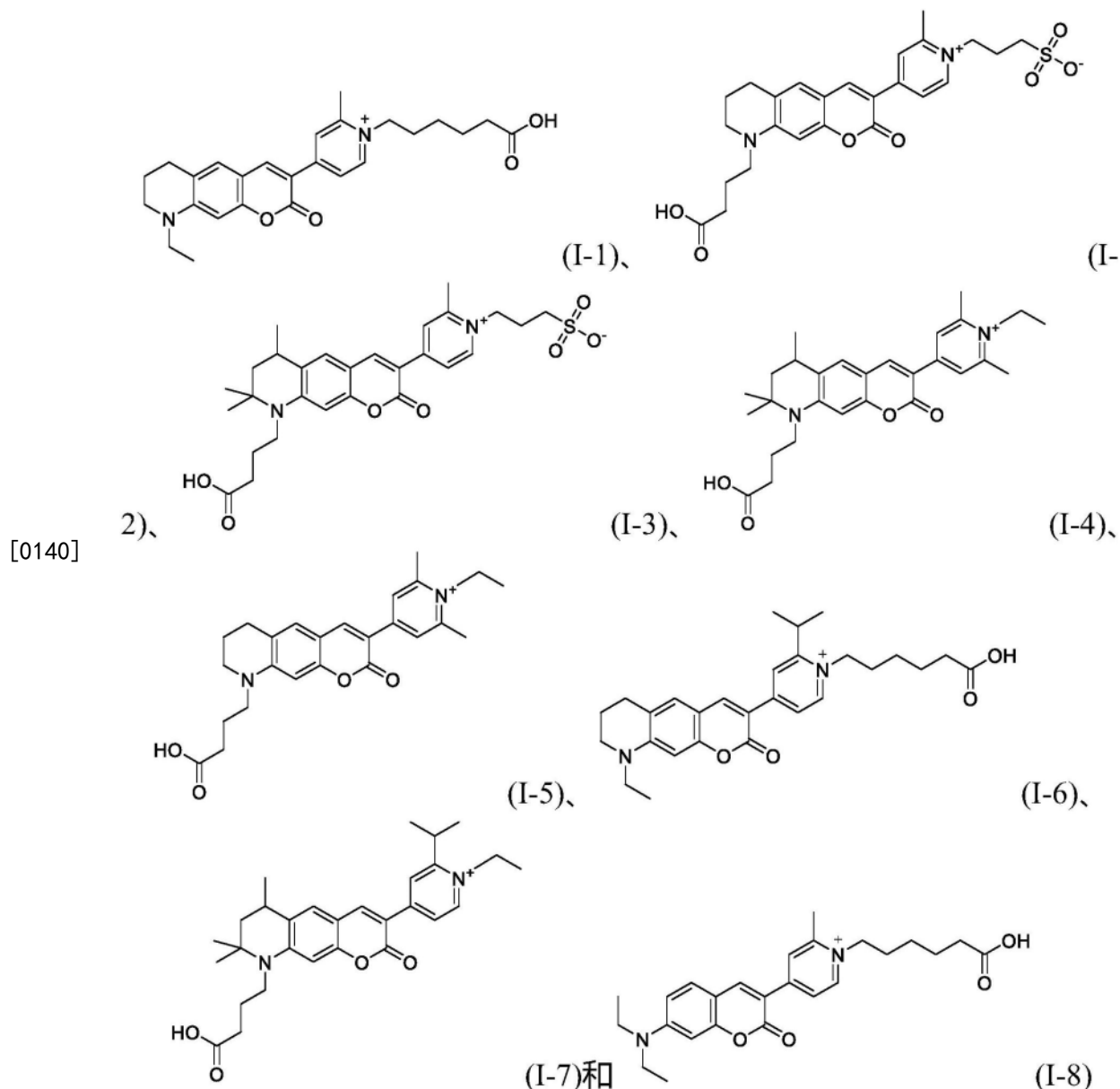
[0135] 在式(Ia)的化合物的一些实施方案中,由实线和虚线-----表示的键是双键。在一些此类实施方案中, R^{10} 是H或 C_1-C_6 烷基。在一个示例中, R^{10} 是甲基。在一些其他实施方案中,由实线和虚线-----表示的键是单键。在一些此类实施方案中, R^{10} 是H,并且 R^{11} 是 C_1-C_6 烷基。在其他实施方案中, R^{10} 和 R^{11} 中的每一者是H。在式(Ia)的化合物的一些实施方案中, R^8 和 R^9 中的每一者是H。在其他实施方案中, R^8 和 R^9 中的至少一者是 C_1-C_6 烷基。在另外的实施方案中, R^8 和 R^9 中的每一者是 C_1-C_6 烷基。在一个示例中, R^8 和 R^9 中的每一者是甲基。在一些实施方案中, R^3 是H。在其他实施方案中, R^3 是 C_1-C_6 烷基(例如,甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基或正己基等)。在另外的实施方案中, R^3 是经取代的 C_1-C_6 烷基(例如,甲基、乙基、异丙基、正丙基、正丁基、2-丁基、正戊基、2-戊基或正己基,这些烷基被如以下等一个或多个取代基取代:羧基、羧酸根($-C(O)O^-$)、磺基($-SO_3H$)、磺酸根($-SO_3^-$)、硫酸根($-O-SO_3^-$)、任选地经取代的氨基、 $-C(O)OR^{12}$ 或 $-C(O)NR^{13}R^{14}$,其中 R^{12} 是任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基,并且其中 R^{13} 和 R^{14} 中的每一者独立地是H、任选地经取代的 C_1-C_6 烷基、任选地经取代的 C_6-C_{10} 芳基、任选地经取代的5元至10元杂芳基或任选地经取代的 C_3-C_7 环烷基。例如, R^3 是各自任选地被羧基、羧酸根、磺基或磺酸根取代的乙基、正丙基、正丁基或正戊基。作为另一示例, R^3 是被 $-C(O)NR^{13}R^{14}$ 取代的乙基、正丙基、正丁基或正戊基,并且其中每个 R^{13} 和 R^{14} 独立地是被羧基、羧酸根、 $-C(O)OR^{12}$ 、磺基或磺酸根取代的 C_1-C_6 烷基。

[0136] 在式(I)或(Ia)的化合物的一些实施方案中, R^2 是H。在其他实施方案中, R^2 和 R^3 与它们所连接的原子连接在一起以形成任选地经取代的5元、6元或7元杂环基。在一些此类实施方案中, R^2 和 R^3 与它们所连接的原子连接在一起以形成被一个或多个 C_1 - C_6 烷基取代的6元杂环基。

[0137] 在式(I)或(Ia)的化合物的一些实施方案中, R^6 是H或任选地经取代的苯基。

[0138] 在式(I)或(Ia)的化合物的一些实施方案中, R^7 是H。

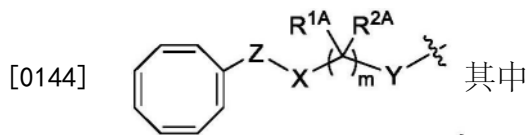
[0139] 式(I)或(Ia)的化合物的另外的实施方案包括以下:



[0141] 和其盐和内消旋形式。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);对应的亚胺类似物(其中代替地,香豆素核-C(=O)部分是-C(=NH)),以及其盐和内消旋形式。

[0142] 环辛四烯(COT)光保护部分

[0143] 在一些实施方案中,可以进一步修饰本文所述的荧光化合物(式(I)或(Ia))以引入与它们共价键合的光保护部分,例如,环辛四烯部分包括结构



[0145] R^{1A} 和 R^{2A} 中的每一者独立地是H、羟基、卤素、叠氮基、硫醇、硝基、氰基、任选地经取代的氨基、羧基、 $-C(O)OR^{5A}$ 、 $-C(O)NR^{6A}R^{7A}$ 、任选地经取代的 C_{1-6} 烷基、任选地经取代的 C_{1-6} 烷氧基、任选地经取代的 C_{1-6} 卤代烷基、任选地经取代的 C_{1-6} 卤代烷氧基、任选地经取代的 C_{2-6} 烯基、任选地经取代的 C_{2-6} 炔基、任选地经取代的 C_{6-10} 芳基、任选地经取代的 C_{7-14} 芳烷基、任选地经取代的 C_{3-7} 碳环基、任选地经取代的5至10元杂芳基或任选地经取代的3至10元杂环基；

[0146] X^1 和 Y^1 各自独立地是键、 $-O-$ 、 $-S-$ 、 $-NR^{3A}-$ 、 $-C(=O)-$ 、 $-C(=O)-O-$ 、

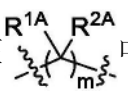
[0147] $-C(=O)-NR^{4A}-$ 、 $-S(O)_2-$ 、 $-NR^{3A}-C(=O)-NR^{4A}-$ 、 $-NR^{3A}-C(=S)-NR^{4A}-$ 、任选地经取代的 C_{1-6} 亚烷基或任选地经取代的亚杂烷基，其中至少一个碳原子被O、S或N替代；

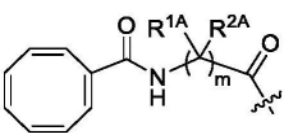
[0148] Z不存在，为任选地取代的 C_{2-6} 亚烯基或任选地取代的 C_{2-6} 亚炔基；

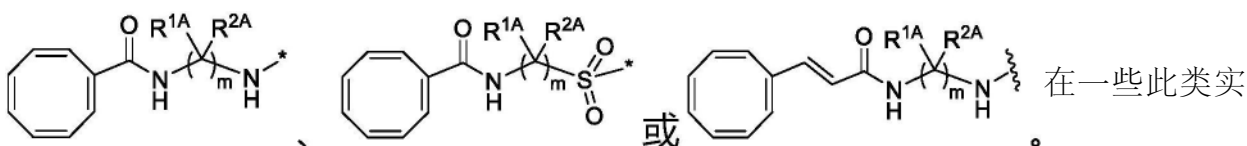
[0149] R^{3A} 和 R^{4A} 中的每一者独立地是H、任选地经取代的 C_{1-6} 烷基或任选地经取代的 C_{6-10} 芳基；

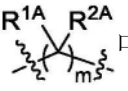
[0150] R^{5A} 是任选地经取代的 C_{1-6} 烷基、任选地经取代的 C_{6-10} 芳基、任选地经取代的 C_{7-14} 芳烷基、任选地经取代的 C_{3-7} 碳环基、任选地经取代的5元至10元杂芳基或任选地经取代的3元至10元杂环基；

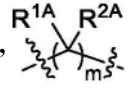
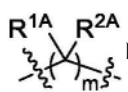
[0151] R^{6A} 和 R^{7A} 中的每一者独立地是H、任选地经取代的 C_{1-6} 烷基、任选地经取代的 C_{6-10} 芳基、任选地经取代的 C_{7-14} 芳烷基、任选地经取代的 C_{3-7} 碳环基、任选地经取代的5元至10元杂芳基或任选地经取代的3元至10元杂环基；

[0152] 在  中 R^{1A} 和 R^{2A} 所连接的碳原子任选地被O、S或N替代，前提条件是当所述碳原子被O或S替代时，则 R^{1A} 和 R^{2A} 均不存在；当所述碳原子被N替代时，则 R^{2A} 不存在；并且m为介于0和10之间的整数。在一些实施方案中，X和Y均不是键。

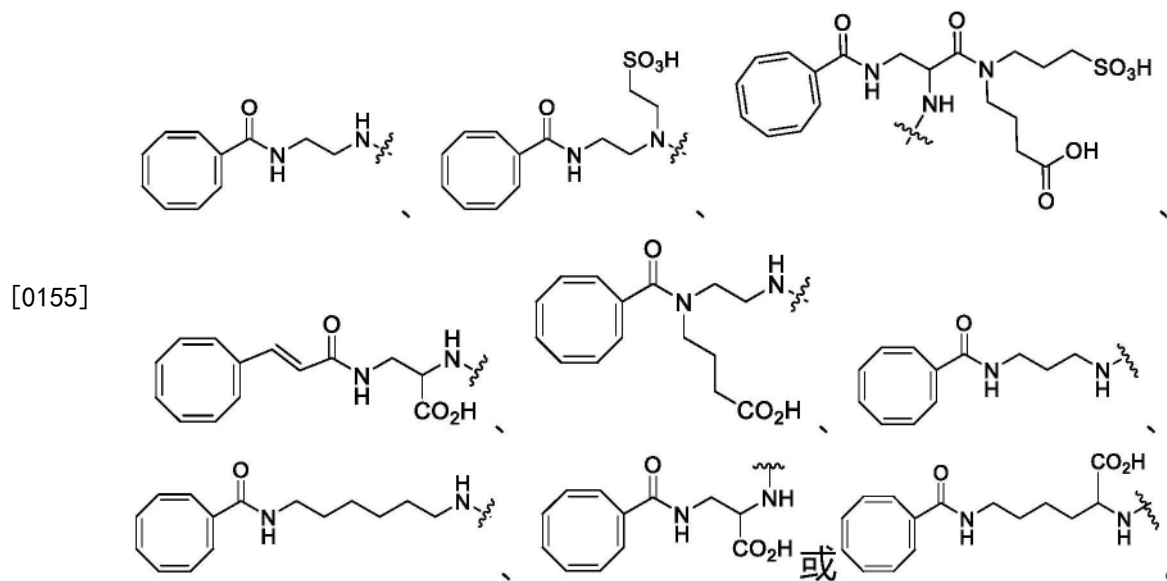
[0153] 在一些实施方案中，环辛四烯部分包括结构 、



施方案中， R^{1A} 和 R^{2A} 中的至少一者是氢。在一些另外的实施方案中， R^{1A} 和 R^{2A} 均为氢。在一些其他实施方案中， R^{1A} 为H并且 R^{2A} 为任选地经取代的氨基、羧基或 $-C(O)NR^{6A}R^{7A}$ 。在一些实施方案中，m为1、2、3、4、5或6，并且 R^{1A} 和 R^{2A} 中的每一者独立地为氢、任选地经取代的氨基、羧基、 $-C(O)NR^{6A}R^{7A}$ 或它们的组合。在一些另外的实施方案中，当m是2、3、4、5或6时，一个 R^{1A} 是氨基、羧基或 $-C(O)NR^{6A}R^{7A}$ ，并且剩余 R^{1A} 和 R^{2A} 是氢。在一些实施方案中， 中 R^{1A} 和 R^{2A} 连接的

至少一个碳原子被O、S或N替代。在一些此类实施方案中，中的一个碳原子被氧原子替代，并且连接到所述替代的碳原子的 R^{1A} 和 R^{2A} 都不存在。在一些其他实施方案中，当中的一个碳原子被氮原子替代时，连接到所述替代的碳原子的 R^{2A} 不存在，并且连接到所述替代的碳原子的 R^{1A} 是氢或 C_{1-6} 烷基。在 R^{1A} 和 R^{2A} 的任何实施方案中，当 R^{1A} 或 R^{2A} 是 $-C(O)NR^{6A}R^{7A}$ 时， R^{6A} 和 R^{7A} 可以独立地是H、 C_{1-6} 烷基或经取代的 C_{1-6} 烷基（例如，被 $-CO_2H$ 、 $-NH_2$ 、 $-SO_3H$ 或 $-SO_3^-$ 取代的 C_{1-6} 烷基）。

[0154] 在一些另外的实施方案中，本文所述的荧光染料包含具有以下结构的环辛四烯部分：



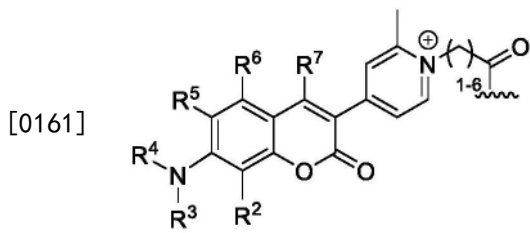
[0156] 本文所述的COT部分可以由本文所述的荧光染料的官能团（例如，羧基基团）与COT衍生物的氨基基团之间形成酰胺键（其中未示出酰胺键的羰基基团）反应产生。

[0157] 标记的核苷酸或寡核苷酸

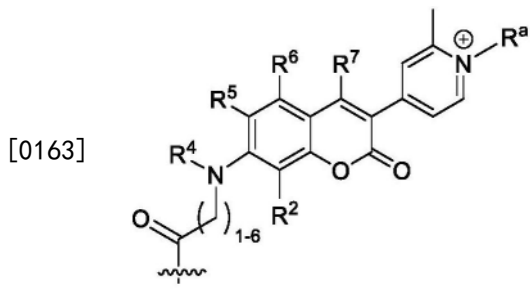
[0158] 根据本公开的一个方面，本文所述的染料化合物适用于连接到底物部分，特别是包含连接基基团以使得能够连接到底物部分。底物部分事实上可以是本公开的染料可以与之缀合的任何分子或物质，并且以非限制性示例的方式，可以包括核苷、核苷酸、多核苷酸、碳水化合物、配体、颗粒、固体表面、有机聚合物和无机聚合物、染色体、核、活细胞，以及它们的组合或装配物。染料可以通过任意的连接基通过多种方式（包括疏水吸引、离子吸引和共价连接）缀合。在一些方面，染料通过共价连接缀合到底物。更具体地讲，共价连接借助于连接基基团实现。在一些情况下，此类标记的核苷酸也被称为“经修饰的核苷酸”。

[0159] 本公开的一些方面涉及用如本文所述的式(I)或式(Ia)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或寡核苷酸。标记的核苷酸或寡核苷酸可以经由羧基($-CO_2H$)或烷基-羧基基团连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，例如酰胺或酯的形式，其可以用于连接到核苷酸或寡核苷酸的氨基或羟基基团。如本文所用的术语“活化酯”，是指能够在温和条件下与例如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。

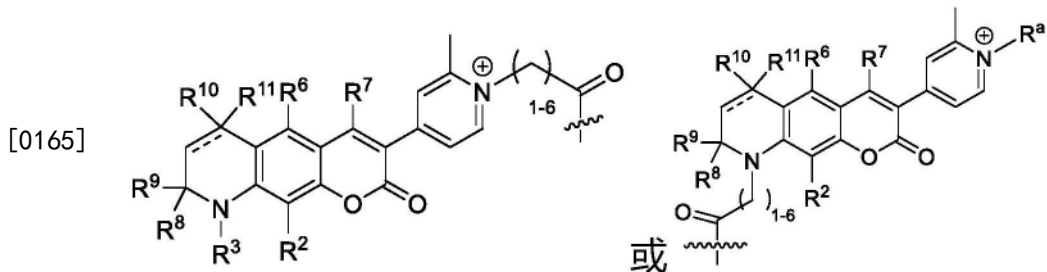
[0160] 例如,式(I)的染料化合物可以通过式(I)的 R^1 (例如, R^a 、 R^b 或 R^c)或 R^3/R^4 之一连接到核苷酸或寡核苷酸。在一些此类实施方案中,式(I)的 R^1 包含 $-CO_2H$ 或 $(CH_2)_{1-6}-CO_2H$,并且该连接形成 R^1 的羧基官能团与核苷酸或核苷酸连接基的氨基官能团之间的酰胺部分。作为一个示例,标记的核苷酸或寡核苷酸可以包含具有以下结构的染料部分:



[0162] 在其他实施方案中,式(I)的 R^3 或 R^4 包括 $-CO_2H$ 或 $(CH_2)_{1-6}-CO_2H$,并且该连接使用 $-CO_2H$ 基团形成酰胺。例如,标记的核苷酸或寡核苷酸可以包含以下染料部分:



[0164] 类似地,式(Ia)的染料化合物可以经由式(Ia)的 R^1 (例如, R^a 、 R^b 或 R^c)或 R^3 ,通过在 R^1 或 R^3 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来连接到核苷酸或寡核苷酸。例如,标记的核苷酸或寡核苷酸可以包含以下染料部分:



[0166] 在其他实施方案中,式(I)或式(Ia)的 R^b 或 R^c 包括 $-CO_2H$ 或 $(CH_2)_{1-6}-CO_2H$,并且该连接使用 $-CO_2H$ 基团形成酰胺。

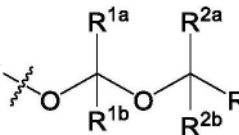
[0167] 在一些实施方案中,染料化合物可以经由核苷酸碱基共价地连接到寡核苷酸或核苷酸。在一些此类实施方案中,标记的核苷酸或寡核苷酸可以具有任选地通过连接基部分连接到嘧啶碱基的C5位置或7-脱氮嘌呤碱基的C7位置的染料。例如,核碱基可以是7-脱氮腺嘌呤,并且染料任选地通过连接基在C7位置处连接到7-脱氮腺嘌呤。核碱基可以是7-脱氮鸟嘌呤,并且染料任选地通过连接基在C7位置处连接到7-脱氮鸟嘌呤。核碱基可以是胞嘧啶,并且染料任选地通过连接基在C5位置处连接到胞嘧啶。作为另一示例,核碱基可以是胸腺嘧啶或尿嘧啶,并且染料任选地通过连接基在C5位置处连接到胸腺嘧啶或尿嘧啶。

[0168] 3'-OH封端基团

[0169] 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

闭基团位于核苷酸的核糖或脱氧核糖的3'-OH位置处。各种3'-OH封闭基团公开于W02004/018497和W02014/139596中,这些申请以引用方式特此并入。例如,封闭基团可以为连接至核糖或脱氧核糖部分的3'氧原子的叠氮甲基(-CH₂N₃)或经取代的叠氮甲基(例如,-CH(CHF₂)N₃或CH(CH₂F)N₃)或烯丙基。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'-OCH₂N₃。

[0170] 在一些其他实施方案中,3'封闭基团和3'氧原子形成共价连接到核糖或脱氧核糖

的3'碳的结构  的缩醛基团,其中:

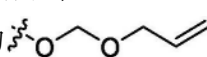
[0171] R^{1a}和R^{1b}各种独立地为H、C₁-C₆烷基、C₁-C₆卤代烷基、C₁-C₆烷氧基、C₁-C₆卤代烷氧基、氰基、卤素、任选地取代的苯基或任选地取代的芳烷基;

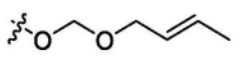
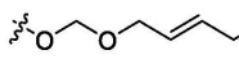
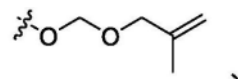
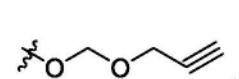
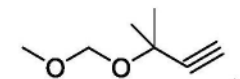
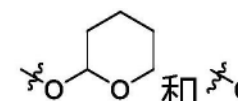
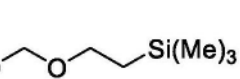
[0172] R^{2a}和R^{2b}各种独立地为H、C₁-C₆烷基、C₁-C₆卤代烷基、氰基或卤素;

[0173] 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;

[0174] R^F为H、任选地经取代的C₂-C₆烯基、任选地经取代的C₃-C₇环烯基、任选地经取代的C₂-C₆炔基或任选地经取代的(C₁-C₆亚烷基)Si(R^{3a})₃;并且

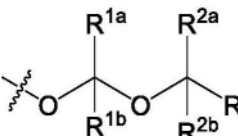
[0175] 每个R^{3a}独立地为H、C₁-C₆烷基或任选地取代的C₆-C₁₀芳基。

[0176] 另外的3'-OH封闭基团公开于美国公开第2020/0216891A1号中,该公开全文以引用方式并入。缩醛封端基团的非限制性示例为 

(AOM)、 、 、 、 、 、  和  Si(Me)₃, 各自共价连接到核糖或脱氧核糖的3'碳。

[0177] 3'-OH封端基团的脱保护

[0178] 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(THPE)或三(羟丙基)膦(THP或THPP)。本文所述的3'-缩醛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,Pd(OAc)₂或烯丙基氯化Pd(II)二聚体)去除。

[0179] 钯裂解试剂

[0180] 在一些实施方案中,本文所述的3'羟基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 、 $\text{Pd}(\text{CH}_3\text{CN})_2\text{Cl}_2$ 、 $(\text{PdCl}(\text{C}_3\text{H}_5))_2$ 、 $[\text{Pd}(\text{C}_3\text{H}_5)(\text{THP})]\text{Cl}$ 、 $[\text{Pd}(\text{C}_3\text{H}_5)(\text{THP})_2]\text{Cl}$ 、 $\text{Pd}(\text{OAc})_2$ 、 $\text{Pd}(\text{Ph}_3)_4$ 、 $\text{Pd}(\text{dba})_2$ 、 $\text{Pd}(\text{Acac})_2$ 、 $\text{PdCl}_2(\text{COD})$ 和 $\text{Pd}(\text{TFA})_2$ 。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 $[(\text{PdCl}(\text{C}_3\text{H}_5))_2]$ 。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。

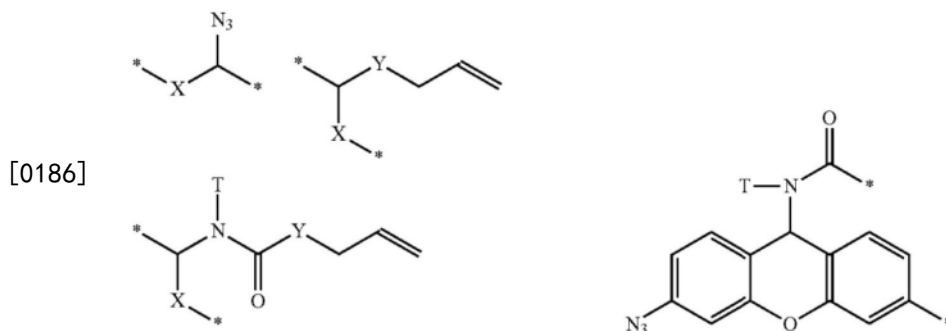
[0181] 在一些实施例中，Pd(0)通过将Pd(II)络合物 $[(\text{PdCl}(\text{C}_3\text{H}_5))_2]$ 与THP原位混合来制备。Pd(II)络合物和THP的摩尔比可以为约1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9或1:10。在一些另外的实施方案中，可以添加一种或多种还原剂，诸如抗坏血酸或其盐(例如，抗坏血酸钠)。在一些实施方案中，裂解混合物可以含有另外的缓冲液试剂，诸如伯胺、仲胺、叔胺、碳酸盐、磷酸盐或硼酸盐，或它们的组合。在一些另外的实施方案中，缓冲液试剂包括乙醇胺(EA)、三(羟甲基)氨基甲烷(Tris)、甘氨酸、碳酸钠、磷酸钠、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。

[0182] 连接基

[0183] 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。

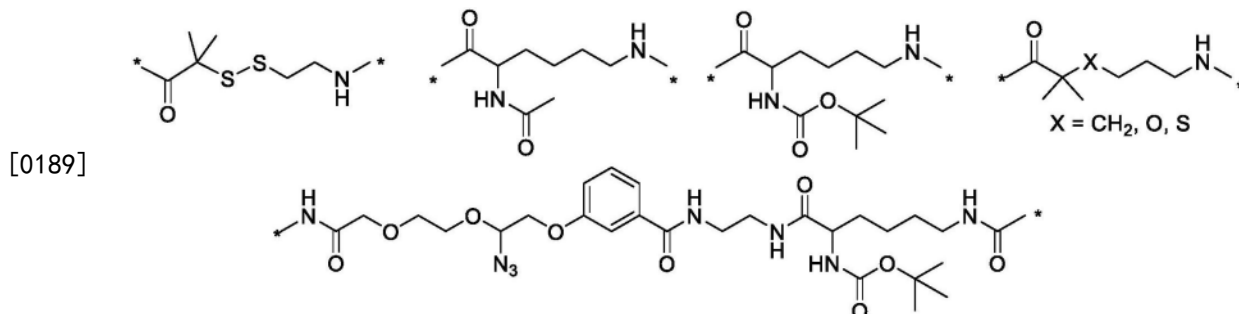
[0184] 可用的连接基基团可以见于PCT公开W02004/018493(以引用方式并入本文)中，其示例包括可以使用水溶性膦或由过渡金属和至少部分水溶性的配体形成的水溶性过渡金属催化剂裂解的连接基。在水性溶液中，后者形成至少部分水溶性的过渡金属络合物。此类可裂解的连接基可以用于将核苷酸的碱基连接到标记，诸如本文所示的染料。

[0185] 特定的连接基包括PCT公开W02004/018493(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：



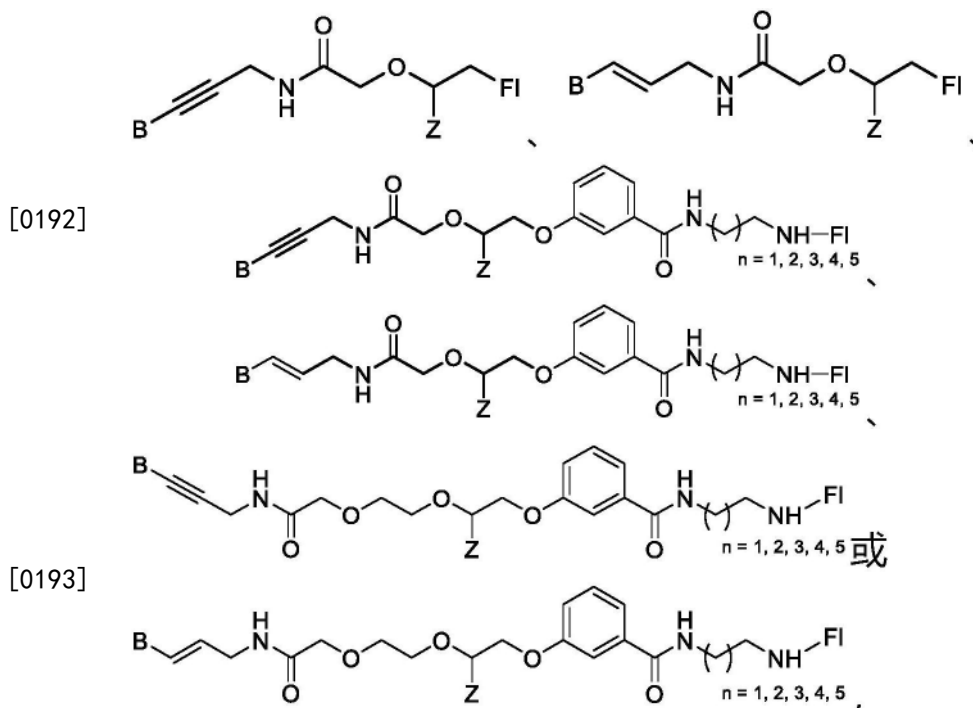
[0187] (其中X选自包括O、S、NH和NQ的组,其中Q为C₁-10取代或未取代的烷基基团,Y选自包括O、S、NH和N(烯丙基)的组,T为氢或C₁-C₁₀取代或未取代的烷基基团,并且*指示所述部分与核苷酸或核苷的其余部分连接的位置)。在一些方面,连接基将核苷酸的碱基连接到标记,例如,本文所述的染料化合物。

[0188] 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:



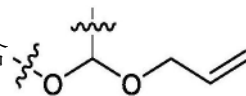
[0190] (其中*指示所述部分与核苷酸或核苷的其余部分连接的位置)。本文所展示的连接基部分可以包括核苷酸/核苷与标记之间的全部或部分连接基结构。本文所展示的连接基部分可以包括核苷酸/核苷与标记之间的全部或部分连接基结构。

[0191] 连接基的其他示例包括下式的部分:



[0194] 其中B是核碱基;Z是-N3(叠氮基)、-O-C₁-C₆烷基、-O-C₂-C₆烯基或-O-C₂-C₆炔基;并且F1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。

[0195] 在特定实施方案中,荧光染料(荧光团)与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连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W02007020457(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连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-(CH₂)₂O_n-的间隔区基团,其中n为介于2和50之间的整数,如W0 2007/020457中所述。

[0196] 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,糖为核糖,而在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,其中酯化发生在连接到糖的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。

[0197] “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。

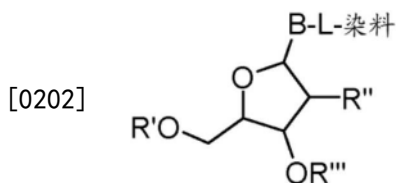
[0198] 虽然碱基通常被称为嘌呤或嘧啶,但是技术人员将理解,可获得不改变核苷酸或核苷经历Watson-Crick碱基配对的能力的衍生物和类似物。“衍生物”或“类似物”意味着这样的化合物或分子:其核心结构与母体化合物的核心结构相同或非常类似,但其具有允许衍生的核苷酸或核苷与另一分子连接的化学或物理修饰,诸如不同的或附加的侧基。例如,碱基可以为脱氮嘌呤。在特定实施方案中,衍生物应当能够经历Watson-Crick配对。“衍生物”和“类似物”还包括例如具有经修饰的碱基部分和/或经修饰的糖部分的合成核苷酸或核苷衍生物。此类衍生物和类似物在例如Scheit, Nucleotide analogs (John 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、烷基磷酸酯键、苯胺磷酸酯键、氨基磷酸酯键等。

[0199] 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。

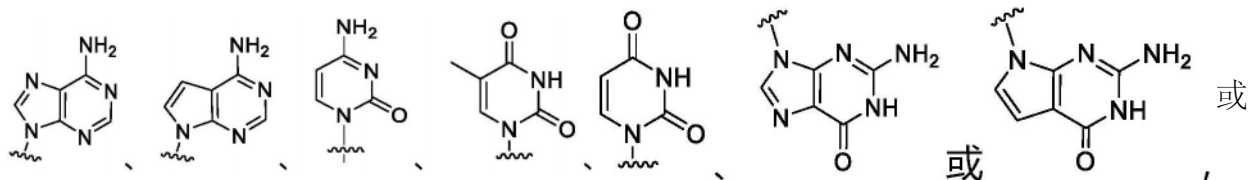
[0200] 在特定实施方案中,标记的核苷酸或寡核苷酸可以是可酶促掺入的和可酶促延伸的。因此,连接基部分可以具有足够的长度以将核苷酸连接到化合物,使得该化合物不显著干扰核酸复制酶对核苷酸的总体结合与识别。因此,连接基还可以包含间隔区单元。间隔区

距离为例如核苷酸碱基距裂解位点或标记的距离。

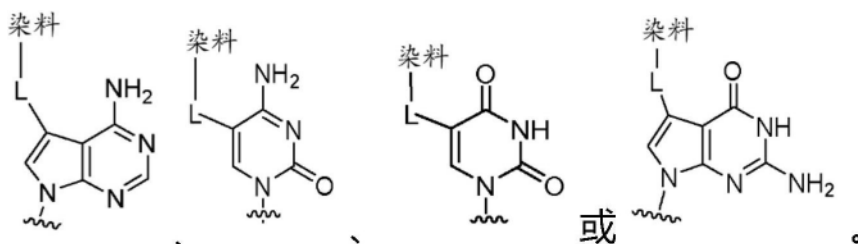
[0201] 用本文所述的染料标记的核苷或核苷酸可以具有下式：



[0203] 其中染料是本文所述的染料化合物(标记)部分(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为可以存在或不存在的任意的连接基;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-或被封闭基团保护的-O-;R''是H或OH;并且R'''是H,本文所述的3'-OH封闭基团或-OR'''形成亚磷酰胺。其中-OR'''为亚磷酰胺,R'为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



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



[0204] 在特定实施方案中,封闭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''可以为可以构成或不构成染料化合物的3'-OH封闭基团。

[0205] 在又一个替代性实施方案中,戊糖的3'碳上不存在封闭基团,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。

[0206] 在仍又一个替代性实施方案中,该封闭基团存在于戊糖的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。

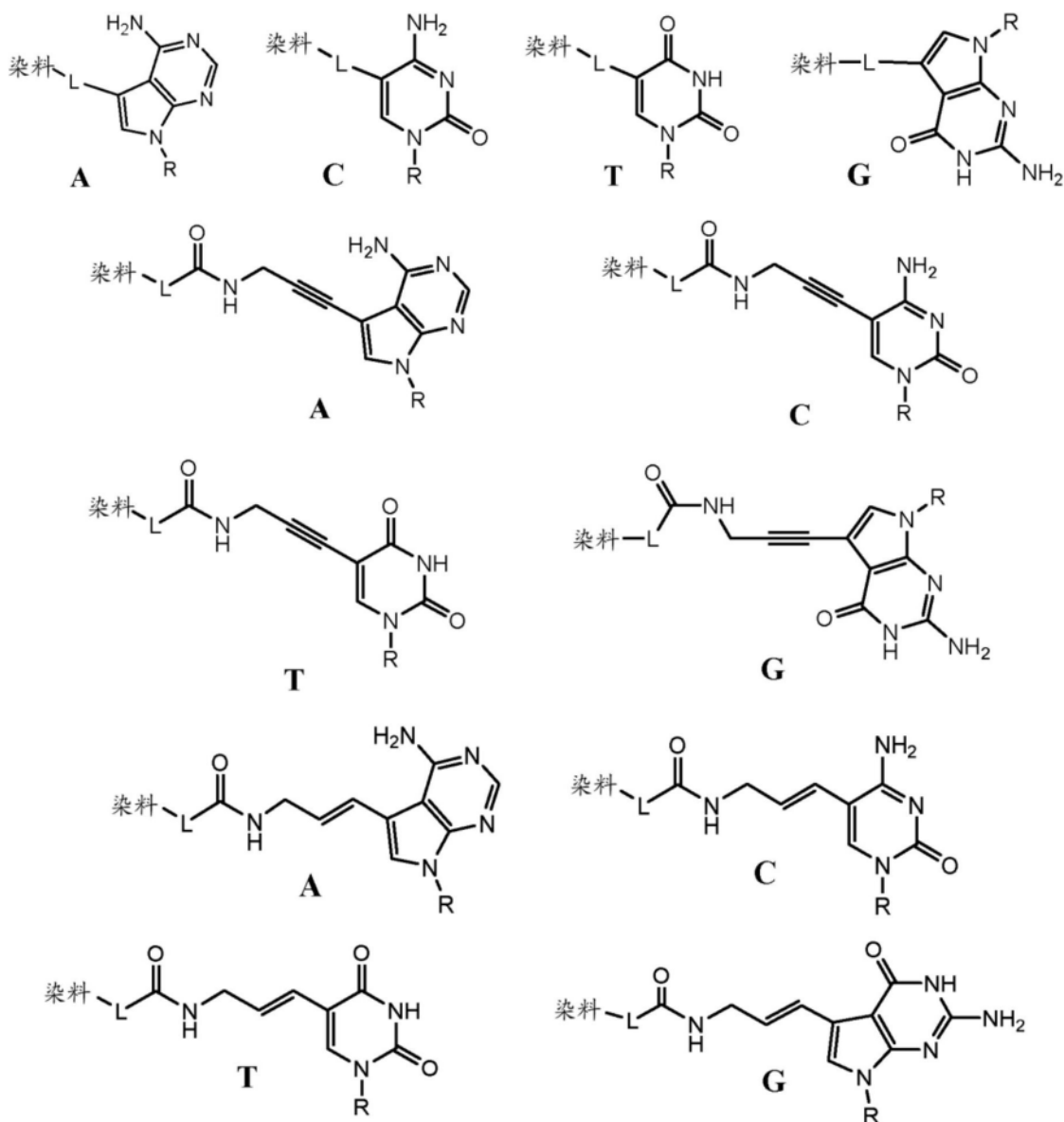
[0207] 使用封端基团允许对聚合进行控制,诸如当掺入标记的核苷酸时通过停止延伸来控制。如果封闭效应是可逆的,例如,以非限制性示例的方式,则通过改变化学条件或通过除去化学障碍,可以在某些点处停止延伸,然后允许其继续。

[0208] 在特定实施方案中,连接基(在染料和核苷酸之间)和封闭基团均存在并且为单独的部分。在特定实施方案中,连接基和封闭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,连接基和封闭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。

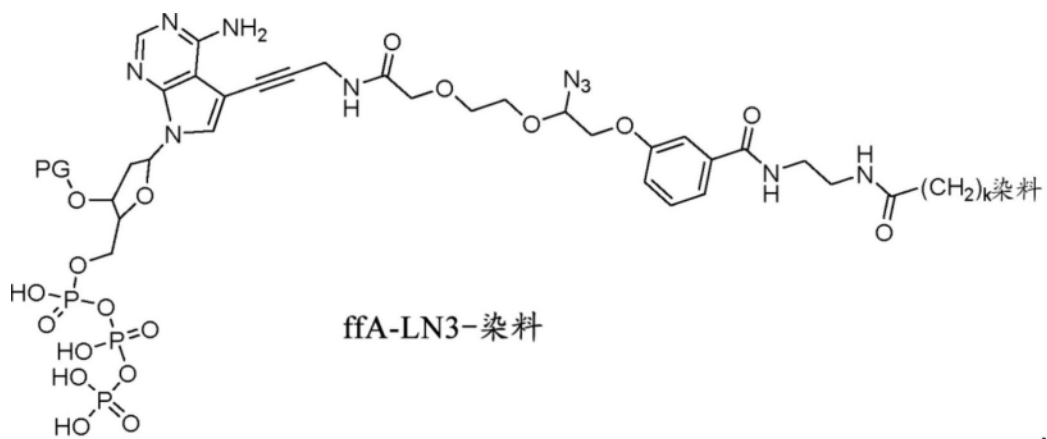
[0209] 本公开还涵盖掺入了染料化合物的多核苷酸。此类多核苷酸可以为分别由以磷酸二酯键接合的脱氧核糖核苷酸或核糖核苷酸组成的DNA或RNA。多核苷酸可以包含与如本文示出的至少一种经修饰的核苷酸(例如,用染料化合物标记)组合的天然存在的核苷酸、除本文所述的标记的核苷酸之外的非天然存在的(或经修饰的)核苷酸或它们的任何组合。根据本公开的多核苷酸还可以包括非天然的骨架键和/或非核苷酸化学修饰。还设想了由包含至少一种标记的核苷酸的核糖核苷酸和脱氧核糖核苷酸的混合物组成的嵌合结构。

[0210] 如本文所述的非限制性示例性标记的核苷酸包括:

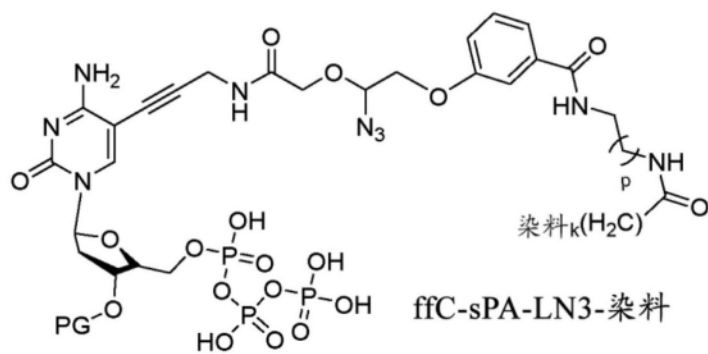
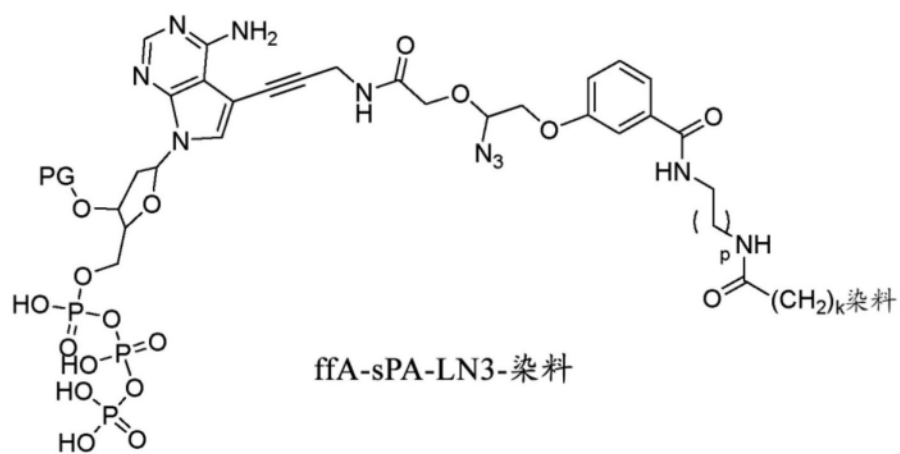
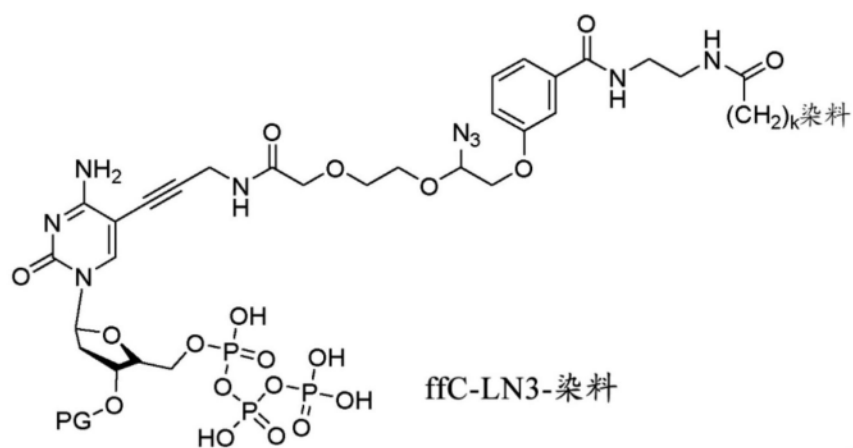


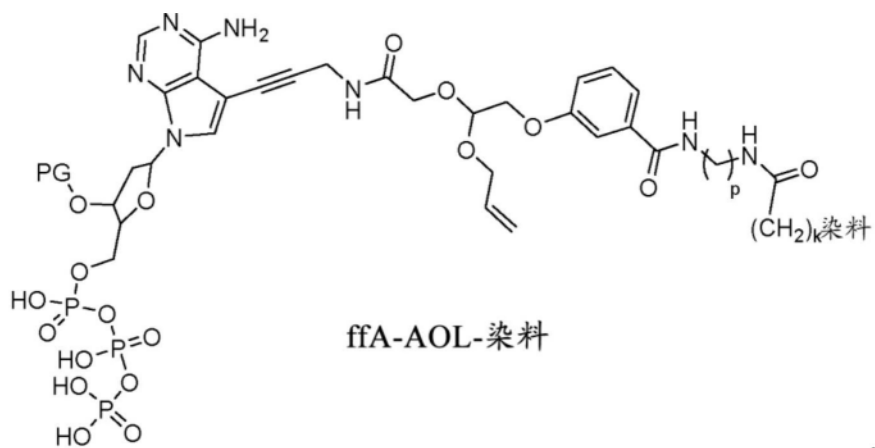
[0212] 其中L表示连接基,并且R表示如上所述的核糖或脱氧核糖部分,或具有被单磷酸、二磷酸或三磷酸取代的5'位置的核糖或脱氧核糖部分。

[0213] 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:

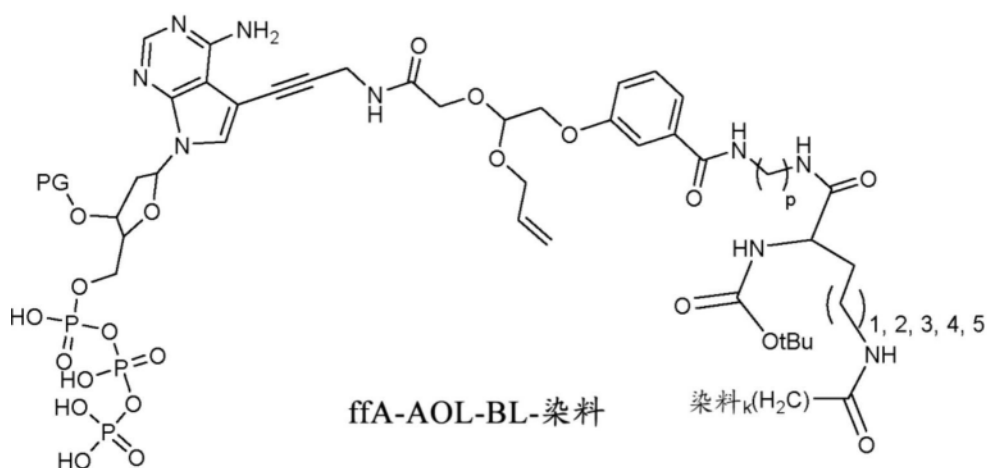


[0214]

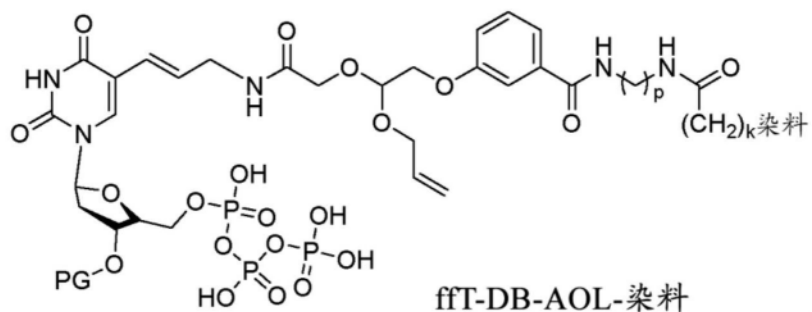




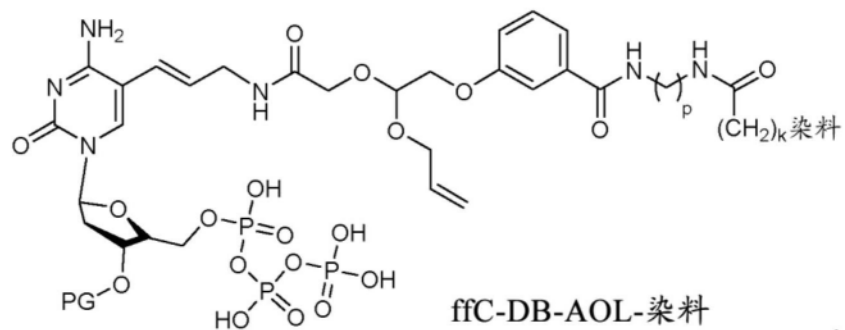
ffA-AOL-染料



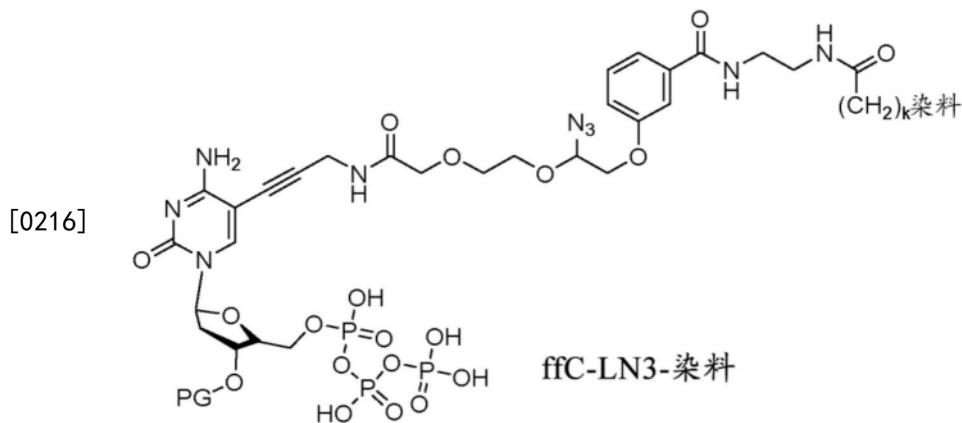
ffA-AOL-BL-染料



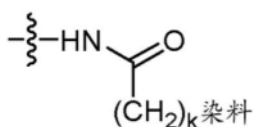
ffT-DB-AOL-染料



ffC-DB-AOL-染料



[0217] 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。

[0218] 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。在一些实施方案中，寡核苷酸与靶多核苷酸的至少一部分杂交。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。在另外的实施方案中，图案化流动池被制造在CMOS芯片上。在另外的实施方案中，图案化流动池包括多个纳米孔。在仍另外的实施方案中，多个纳米孔直接在每个CMOS光电二极管(像素)上对准。

[0219] 试剂盒

[0220] 本文提供了包括用本公开的烷基吡啶正离子香豆素化合物(即，第一标记)标记的第一核苷酸的试剂盒。在一些实施方案中，试剂盒还包括第二标记的核苷酸，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烷基吡啶正离子香豆素化合物。在一些实施方案中，可使用单个激发源激发第一标记的核苷酸和第二标记的核苷酸，该单个激发源可以是具有第一激发波长的第一光源。例如，第一标记和第二标记的激发带可以至少部分地重叠，使得在光谱重叠区域中的激发引起这两种标记发射荧光。在一些另外的实施方案中，试剂盒可以包括第三核苷酸，其中第三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。在一些此类实施方案中，第一标记的核苷酸和第三标记的核苷酸可使用第二激发源激发，该第二激发源可以是具有与第一激发波长不同的第二激发波长的第二光源。例如，第一标记和第三标记的激发带可以至少部分地重叠，使得在光谱重叠区域中的激发引起这两种标记发射荧光。在一些另外的实施方案中，试剂盒还可以包括第四核苷酸。在一些此类实施方案中，第四核苷酸未标记(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。在仍其他实施方案中，第四核苷酸未标记。在一些实施方案中，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)的激发波长。在替代性实施方案中,第一光源具有约400nm至约480nm、约420nm至约470nm或450nm至约460nm(例如450nm)的激发波长。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)的激发波长。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种标记具有可以在单个发射收集滤光器或通道中收集的发射光谱。

[0221] 在一些实施方案中,该试剂盒可以包括四种标记的核苷酸(A、C、G和T或U),其中这四种核苷酸中的第一核苷酸用如本文所公开的化合物进行标记。在这种试剂盒中,这四种核苷酸中的每一种核苷酸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。可替代地,这四种核苷酸中的第一核苷酸是本文所描述的标记的核苷酸,这四种核苷酸中的第二核苷酸携带第二标记,第三核苷酸携带第三标记,并且第四核苷酸未标记(暗色)。作为另一示例,这四种核苷酸中的第一核苷酸是本文所述的标记的核苷酸,这四种核苷酸中的第二核苷酸携带第二标记,第三核苷酸携带两个标记的混合物,并且第四核苷酸未标记(暗色)。因此,标记化合物中的一种或多种化合物可以具有不同的吸光度最大值和/或发射最大值,使得该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。本文所公开的烷基吡啶正离子香豆素染料通常吸收低于500nm的区域中的光。例如,这些香豆素染料可以具有约450nm至约530nm、约460nm至约520nm、约475nm至约510nm或约490nm至约500nm的吸收波长。

[0222] 本文示出的化合物、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)。可以采用这些化合物、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。

[0223] 在具体实施方案中,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它们的任何组合组合供应。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。

[0224] 在试剂盒包括多种,具体地两种或三种、或者更具体地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,或者一种核苷酸可以是暗色的,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介于450nm-460nm、490nm-500nm之间或者为520nm或更高(例如,532nm)。

[0225] 在一些实施方案中,试剂盒包括用本公开的化合物标记的第一核苷酸和用第二染料标记的第二核苷酸,其中染料具有至少10nm,具体地20nm至50nm或30nm至40nm的吸光度最大值差值。更具体地,第一标记可以具有高于40nm、高于50nm或高于60nm的斯托克斯位移。第二标记可以具有约80nm、高于90nm或高于100nm的斯托克斯位移(其中“斯托克斯位移”是峰吸收波长和峰发射波长之间的距离)。此外,第一标记可以具有约460nm至约520nm、约475nm至约510nm或约490nm至约500nm的吸收最大值。第二标记可以具有约400nm至约470nm或约450nm至约460nm的吸收最大值。在另外的实施方案中,试剂盒还可以第三标记的核苷酸,其中第三标记具有高于520nm的吸收最大值。第三标记可以具有高于20nm、高于30nm或高于40nm的斯托克斯位移,或介于20nm-40nm之间的斯托克斯位移。试剂盒可以还包括未标记的第四核苷酸。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种标记具有大于540nm、大于550nm、大于560nm、大于570nm、大于580nm、大于590nm或大于600nm的发射最大值。在一些实施方案中,可以在单个发射收集通道或滤光器(例如,从约580nm至约700nm的收集区域)中检测或收集第一标记、第二标记和第三标记的发射光谱。

[0226] 在替代性实施方案中,本公开的试剂盒可以包括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。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”并且不合标记。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。

[0227] 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。

[0228] 除了标记的核苷酸之外,试剂盒还可以一起包括至少一种另外的组分。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(如突变DNA聚合酶)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,如抗坏血酸或抗坏血酸钠,这些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以用于裂解3'封闭基团和/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形成的水溶性磷或水溶性过渡金属催化剂(如钨络合物)。试剂盒的各种组分可以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,核苷酸或标记的核苷酸包含3'羟基封闭基团。

[0229] 测序方法

[0230] 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,如包括检测连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缔合。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'羟基基团连接,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'磷酸连接。因此,在一个非限制性实施方案中,本公开提供了检测掺入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多核苷酸中,以及(b)通过检测来自连接到所述核苷酸的染料化合物的荧光信号来确定掺入多核苷酸中的核苷酸的身份。

[0231] 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多核苷酸中的一种或多种标记的核苷酸的荧光来检测该一种或多种标记的核苷酸。

[0232] 本申请的一些实施方案涉及一种确定靶多核苷酸的序列的方法,该方法包括:(a)使引物多核苷酸/靶多核苷酸复合物与一种或多种标记的核苷酸(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种标记的核苷酸是本文所述的标记的核苷酸,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)将标记的核苷酸掺入引物多核苷酸中以产生延伸的引物多核苷酸;和(c)执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)从掺入引物多核苷酸中的核苷酸去除标记部分和3'封闭基团。在一些另外的实施方案中,该方法还可以包括(e)从引物多核苷酸链洗掉去除的标记部分和3'封闭基团。在一些实施方案中,重复步骤(a)至(d)或步骤(a)至(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引物多核苷酸链中的核苷酸的标记部分和3'封闭基团。在一些另外的实施方案中,在自动化测序仪器上进行该方法,并且其中自动化测序仪器包括以不同波长操作的两个光源。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。

[0233] 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单链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;其中第一激发光源和第二激发光源具有不同波长;并且其中在单个发射检测通道中检测或收集第一发射信号和第二发射信号。在一些实施方案中,本文所述的色烯并喹啉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和第二成像事件之前对混合物中的任何核苷酸缀合物进行化学修饰。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状态和第二成像事件中的暗色状态确定。第二类型的核苷酸缀合物的掺入由第一成像事件中的暗色状态和第二成像事件中的信号状态确定。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中的信号状态确定。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中的暗色状态确定。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,第一激发光源的波长比第二激发光源的波长短。在一些此类实施方案中,第一激发光源具有约400nm到约480nm、约420nm到约470nm或约450nm到约460nm(即“蓝光”)的波长。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530nm(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,第一激发光源的波长比第二激发光源的波长长。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530nm(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(即“蓝光”)的波长。在一个实施方案中,第二激发光源具有约450nm的波长。

[0234] 在一个实施方案中,在合成步骤中通过聚合酶的作用将至少一种核苷酸掺入多核苷酸(如本文所述的单链引物多核苷酸)中。然而,可以使用将核苷酸连接到多核苷酸的其他方法,例如,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。

[0235] 在特定实施方案中,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'羟基基团与标记的核苷酸上的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。

[0236] 在这些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,可以分离或纯化掺入了标记的核苷酸的多核苷酸链,并且然后进一步加工或用于后续的分析。以举例的方式,在合成步骤中掺入了如本文所述的标记的核苷酸的多核苷酸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。

[0237] 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构成。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/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W0 2005/024010或W006120433中描述的那些热稳定聚合酶,这些文献中的每个文献通过引用并入本文。在较低温度如37℃下进行的合成反应中,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,如反应温度、pH、链置换活性等。

[0238] 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经修饰的核苷酸或核苷进行检测的任何其他应用。

[0239] 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。边合成边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确定单个碱基的身份的用途可能是有用的,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之内。

[0240] 在本公开的实施方案中,模板/靶多核苷酸的序列通过检测连接到掺入的核苷酸的荧光标记,通过检测一种或多种核苷酸掺入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。

[0241] 在特定实施方案中,不同的核苷三磷酸(A、T、G和C)中的每一种核苷三磷酸可以用独特的荧光团进行标记,并且还在3'位置处包含封闭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/靶多核苷酸互补的新生链中,并且封闭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)去除(脱保护)3'封闭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,美国专利第5,302,509号(该专利以引用方式并入本文)公开了用于对固定在固体载体上的多核苷酸进行测序的方法。

[0242] 如上文所例示,该方法利用了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。发夹多核苷酸及它们可以连接到固体载体的方法在PCT公开第W00157248号和第W02005/047301号中公开,这些文献中的每个文献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,通过经由与核苷酸的5'磷酸基团形成磷酸二酯键而将核苷酸接合到核酸链的游离的3'羟基基团,从而将该核苷酸掺入到核酸链(或多核苷酸)中。

[0243] 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。

[0244] 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(例如经由共价连接)连接到固体载体。在某些实施方案中,模板多核苷酸可以直接连接到固体载体(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一些方式修饰,以便允许模板多核苷酸的直接共价连接,或者通过水凝胶或聚电解质多层来固定模板多核苷酸,该多层本身可以非共价地连接到固体载体。

[0245] 在多核苷酸已直接连接到载体(例如,基于二氧化硅的载体,如在W000/06770(以引用方式并入本文)中所公开的那些载体)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。另外,可以通过巯基亲核物质与固体载体的反应将多核苷酸连接到固体载体,例如,如W02005/047301(以引用方式并入本文)中所述。固体负载的模板多核苷酸的仍另外的示例是模板多核苷酸连接到负载在基于二氧化硅或其他固体载体上的水凝胶,例如,如W000/31148、W001/01143、W002/12566、W003/014392、美国专利第6,465,178号和W000/53812中所述,这些文献中的每个文献以引用方式并入本文。

[0246] 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所述的参考文献和W02005/065814中,所述文献以引用方式并入本文。可以使用的特定水凝胶包括W02005/065814和美国公开第2014/0079923号中所述的那些水凝胶。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))。

[0247] 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号中所述(该专利以引用方式并入本文)。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。示例性文库及它们的产生方法描述于Nature, 437, 376-

380(2005);Science,309,5741,1728-1732(2005)中,这些文献中的每个文献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內。

[0248] 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模板)进行测序。

[0249] 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W000/18957(这些文献中的每个文献并入本文)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”构成的阵列。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。

[0250] 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。

[0251] 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。

[0252] 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W000/06770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边合成边测序反应,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。

[0253] 具体地,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。

[0254] 因此,本公开还涵盖用染料化合物标记的核苷酸,这些核苷酸是在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。

[0255] 应当理解,用掺入3' 封闭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' OH封闭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。在根据本公开的并且具有3' 封闭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。

[0256] 可替代地,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,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、两种、三种或每一种核苷酸可以是未标记的。四种类型的核苷酸(例如,dNTP)中的每一种核苷酸具有3' 羟基封闭基团,以确保仅单个碱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,该亲和试剂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中公开了未标记的核苷酸和亲和试剂在边合成边测序中的用途。使用未标记的核苷酸的本公开的经修改的测序方法可以包括以下步骤:

[0257] (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;

[0258] (b') 将核苷酸掺入引物多核苷酸中以产生延伸的引物多核苷酸(即,延伸的引物多核苷酸/靶多核苷酸复合物);

[0259] (c') 在一种亲和试剂与掺入的未标记的核苷酸特异性结合的条件下使延伸的引物多核苷酸与一组亲和试剂接触以提供标记的延伸引物多核苷酸(即,标记的延伸引物多核苷酸/靶多核苷酸复合物);

[0260] (d') 对标记的延伸引物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。

[0261] 在本文所述的经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未标记的核苷酸含有3' 羟基封闭基团。在另外的实施方案中,在下一掺入循环之前去除所掺入的核苷酸的3' 羟基封闭基团。在仍另外的实施方案中,该方法还包括从所掺入的核苷酸中去除亲和试剂。在仍另外的实施方案中,在同一反应中去除3' 羟基封闭基团和亲和试剂。在一些实施方案中,该一组亲和试剂可以包含与第一类型的核苷酸特异性结合的第一亲和试剂、与第二类型的核苷酸特异性结合的第二亲和试剂和与第三类型的核苷酸特异性结合的第三亲和试剂。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种亲和试剂包含光谱可区分的可检测标记。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、单链抗体、双特异性抗体等的结合片段)、适体、打结素(knottin)、affimer或以合适的特异性和亲和力与掺入的核苷酸结合的任何其他已知的试剂。在一个实施方案中,该组中的一种或多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亲和试剂是包含一种或多种可检测标记(例如,相同可检测标记

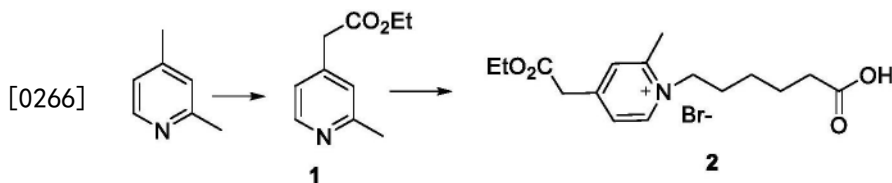
的多个拷贝)的抗体或蛋白质标签,并且可检测标记包含或为本文所述的烷基吡啶正离子香豆素染料部分。

[0262] 实施例

[0263] 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。

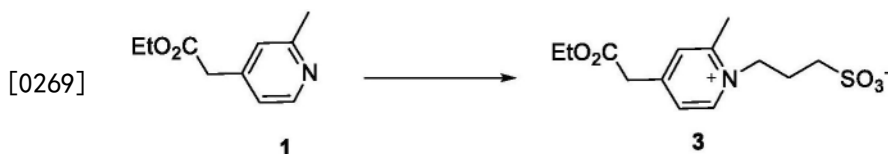
[0264] 实施例1. 烷基吡啶正离子香豆素染料的合成

[0265] 烷基吡啶正离子中间体的合成

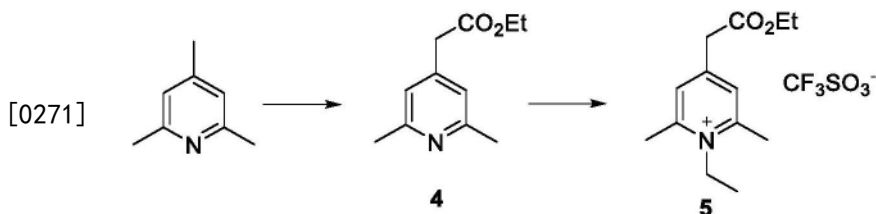


[0267] 将2,4-二甲基吡啶(5.8mL, 50mmol)在氮气下溶解于干燥烧瓶中的无水THF(20mL)中。在0℃下缓慢添加1M二异丙基氨基锂于THF/己烷(50mL, 50mmol)中的溶液。溶液变为深红色,并将其在室温下搅拌4小时。然后,缓慢添加溶解于10mL无水THF中的无水碳酸二乙酯(14.7mL, 2.5mmol),并将反应物在室温下搅拌过夜。然后将混合物用饱和NH₄Cl水溶液(10mL)淬灭,用200mL乙酸乙酯稀释并用200mL水洗涤。将有机相经MgSO₄干燥,过滤并在减压下蒸发。将残余物通过快速色谱法在硅胶上进行纯化。然后将中间体溶解于10mL浓HCl中并在100℃下加热1小时。在减压下去除挥发物,然后将残余物溶解于乙醇(50mL)中,并且添加几滴浓硫酸。在室温下搅拌溶液直到完成。将反应用5mL饱和NaHCO₃水溶液淬灭,在减压下去除挥发物。将残余物在饱和NaHCO₃(50mL)和乙酸乙酯(100mL)之间进行分配。将有机相经MgSO₄干燥,过滤并在减压下蒸发。以34%的产率获得呈无色粘稠油状物的化合物1。

[0268] 将6-溴己酸(228mg, 1.17mmol)和化合物1(210mg, 1.17mmol)混合,然后在100℃下加热过夜。然后将混合物在水与二氯甲烷之间进行分配。蒸发水相,得到化合物2,产率为70%(312mg, 0.838mmol)。LC-MS(ESI):(正离子)m/z 294(M+H⁺)。



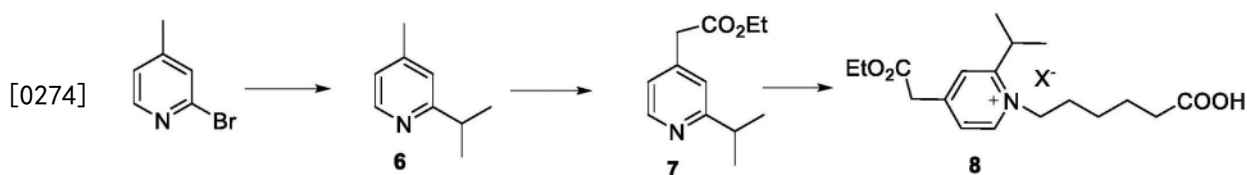
[0270] 将化合物1(199mg, 1.1mmol)和1,3-丙烷磺内酯(101μL, 1.15mmol)溶解于1mL丁腈中并在100℃下加热2小时。在减压下去除挥发物,并将残余物溶解于50mL水中并用2×20mL DCM洗涤。将水相蒸发至干燥。以88%产率(297mg, 0.98mmol)获得呈灰白色固体的化合物3。LC-MS(ESI):(正离子)m/z 302(M+H⁺)。



[0272] 将2,4,6-三甲吡啶(2.5mL, 20mmol)在氮气下溶解于干燥烧瓶中的无水THF(20mL)中。缓慢添加1M二异丙基氨基锂于THF/己烷(22mL, 22mmol)中的溶液。溶液变为深红色和浑浊,并将其在室温下搅拌2小时。然后在-70℃下将其缓慢添加到溶解于20mL无水THF中的无

水碳酸二乙酯(5mL, 40mmol)中,并将反应物缓慢升温至室温过夜。然后将混合物用饱和 NH_4Cl 水溶液(10mL)淬灭,用200mL乙酸乙酯稀释。将有机相分离,经 MgSO_4 干燥,过滤并在减压下蒸发。将粗产物通过快速色谱法在硅胶上进行纯化。然后将中间体溶解于20mL浓 HCl 中并在室温下静置过夜。在减压下去除挥发物,然后将残余物溶解于乙醇(50mL)中,并且添加0.5mL浓硫酸。将溶液回流4小时,然后用5mL饱和 NaHCO_3 水溶液淬灭,在减压下去除挥发物。将残余物在饱和 NaHCO_3 (50mL)和乙酸乙酯(100mL)之间进行分配。将有机相经 MgSO_4 干燥,过滤并在减压下蒸发。以12%的产率(474mg, 2.44mmol)获得呈无色粘稠油状物的化合物4。

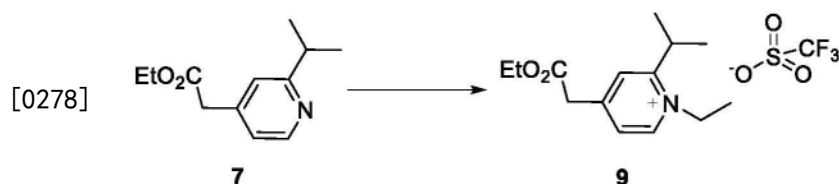
[0273] 向化合物4(100mg, 0.518mmol)中添加乙腈(300 μL)和三氟甲磺酸乙酯(67 μL , 0.518mmol)。将溶液在室温下搅拌3天,然后在减压下蒸发挥发物,得到呈色固体的化合物5,该化合物在不进一步纯化的情况下用于下一步骤。LC-MS(ESI):(正离子) m/z 222($\text{M}+\text{H}^+$)。



[0275] 将 CuCN (2g, 23.2mmol)和无水THF(75mL)在氮气下添加到烧瓶中,然后在搅拌的同时冷却到 -78°C 。在剧烈搅拌的同时逐滴添加异丙基溴化镁(于2-甲基四氢呋喃中的3M溶液, 15.5mL, 46.5mmol)。将悬浮液在 -78°C 下静置20分钟,然后添加2-溴-4-甲基吡啶(1g, 5.81mmol)。将反应在 -78°C 下搅拌3小时,然后升温至室温过夜。然后,将悬浮液在冰浴中冷却,并且用浓氢氧化铵水溶液缓慢淬灭。用 $2 \times 100\text{mL}$ 二氯甲烷萃取所得悬浮液。将有机相经 MgSO_4 干燥,过滤并在减压下蒸发。将粗产物通过快速柱色谱法在硅胶上进行纯化。以30%的产率(240mg, 1.77mmol)获得化合物6。LC-MS(ESI):(正离子) m/z 135($\text{M}+\text{H}^+$)。

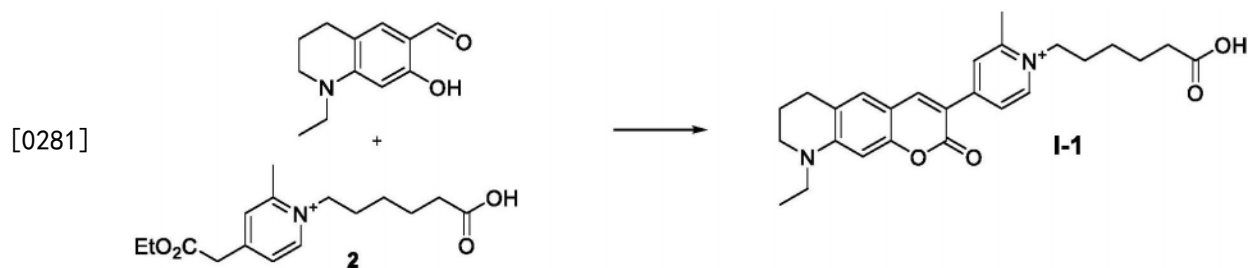
[0276] 将化合物6(240mg, 1.77mmol)在氮气下溶解于干燥烧瓶中的无水THF(5mL)中,并且冷却至 -78°C 。逐滴添加1M二异丙基氨基锂于THF/己烷(3.6mL, 3.54mmol)中的溶液。溶液变为淡红色,并将其在 -78°C 下搅拌1小时。然后缓慢添加溶解于5mL无水THF中的无水碳酸二乙酯(0.417mL, 3.54mmol),并将反应物升温至室温持续3.5小时。然后将混合物用饱和 NH_4Cl 水溶液(5mL)淬灭,用100mL乙酸乙酯稀释。将有机相经 MgSO_4 干燥,过滤并在减压下蒸发。将残余物通过快速色谱法在硅胶上进行纯化。然后将中间体溶解于1mL浓 HCl 中并在 100°C 下加热1小时。在减压下去除挥发物,然后将残余物溶解于乙醇(5mL)中,并且添加几滴浓硫酸。在室温下搅拌溶液直到完成。将反应用5mL饱和 NaHCO_3 水溶液淬灭,在减压下去除挥发物。将残余物在饱和 NaHCO_3 (50mL)和乙酸乙酯(100mL)之间进行分配。将有机相经 MgSO_4 干燥,过滤并在减压下蒸发。以29%的产率获得呈淡黄色粘稠油状物的化合物7。

[0277] 将6-溴己酸(103mg, 0.53mmol)、化合物7(100mg, 0.48mmol)和四丁基碘化铵(几mg, 催化量)溶解于1mL乙腈中,在 100°C 下在密封管中加热2天。然后将混合物在水与乙酸乙酯之间进行分配。蒸发水相,得到化合物8,产率为大约60%(86mg, 0.318mmol)。LC-MS(ESI):(正离子) m/z 322($\text{M}+\text{H}^+$)。

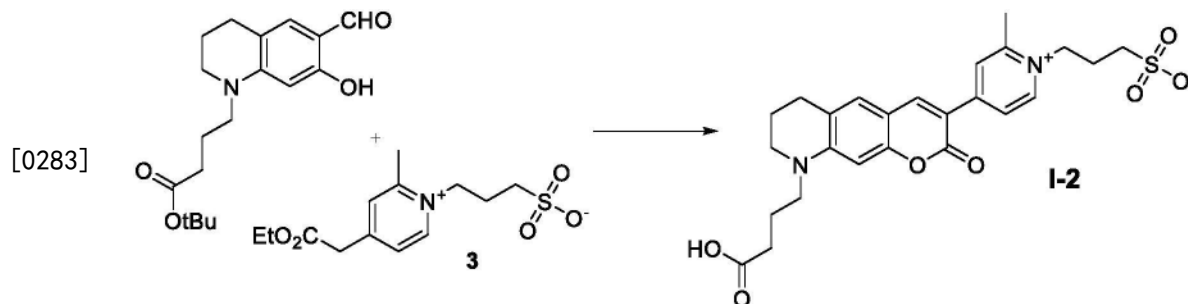


[0279] 使用针对5所述的相同程序制备化合物9。粗产物在不经纯化的情况下用于下一步骤。LC-MS (ESI) : (正离子) m/z 236 ($M+H^+$)。

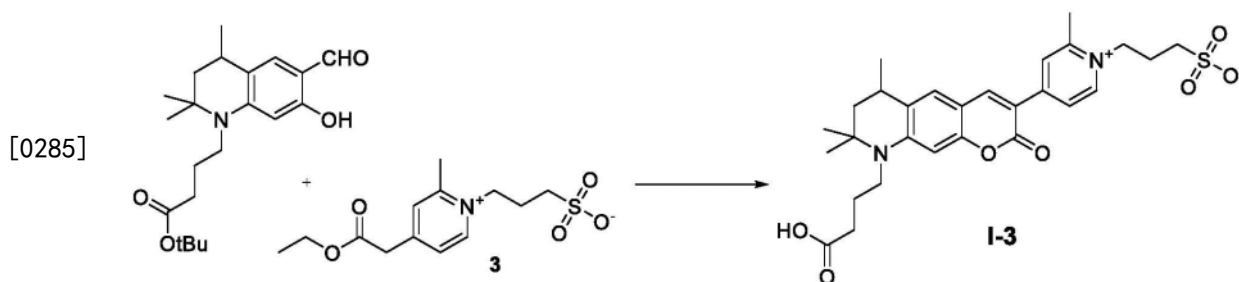
[0280] 香豆素染料的合成



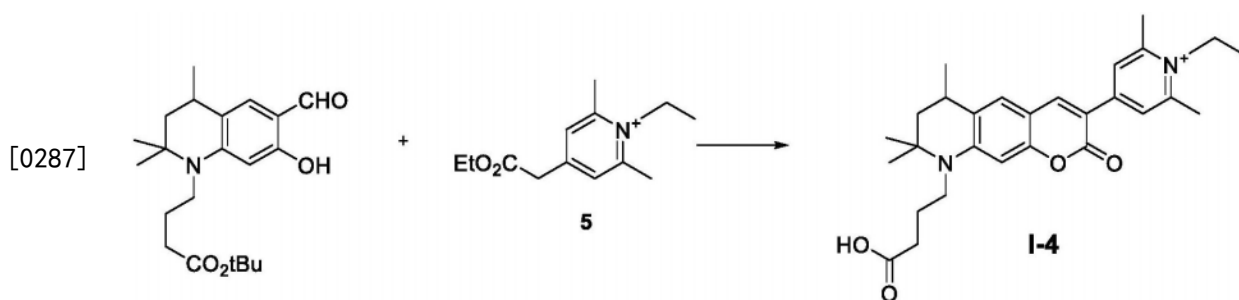
[0282] 将起始材料和催化量的吡啶乙酸盐溶解于乙醇中并回流直到反应完成。将粗产物蒸发至干燥并且通过快速柱色谱法在反相C18上进行纯化。以43%的产率(56mg, 0.129mmol)获得呈棕色固体的化合物I-1。 1H NMR (400MHz, MeOD) : δ (ppm) 8.71 (d, $J=6.9$ Hz, 1H, Ar-H pyr), 8.57 (s, 1H, Ar-H), 8.48 (d, $J=2.2$ Hz, 1H, Ar-H pyr), 8.42 (dd, $J=6.8, 2.3$ Hz, 1H, Ar-H pyr), 7.30 (s, 1H, Ar-H), 6.57 (s, 1H, Ar-H), 4.51 (d, $J=7.8$ Hz, 2H, CH_2-N^+), 3.58-3.48 (m, 4H, CH_2-N), 2.88 (s, 3H, CH_3 pyr), 2.83 (t, $J=6.2$ Hz, 2H, CH_2-Ar), 2.22 (t, $J=7.1$ Hz, 2H, CH_2-COO), 2.04-1.95 (m, 4H, CH_2-CH_2-N , $CH_2-CH_2-N^+$), 1.71 (p, $J=7.2$ Hz, 2H, CH_2-CH_2-COO), 1.58-1.44 (m, 2H, $CH_2-CH_2-CH_2-COO$), 1.27 (d, $J=7.1$ Hz, 3H, CH_3 Et)。LC-MS (ESI) : (正离子) m/z 435 ($M+H^+$) ; (负离子) m/z 433 (M^-)。



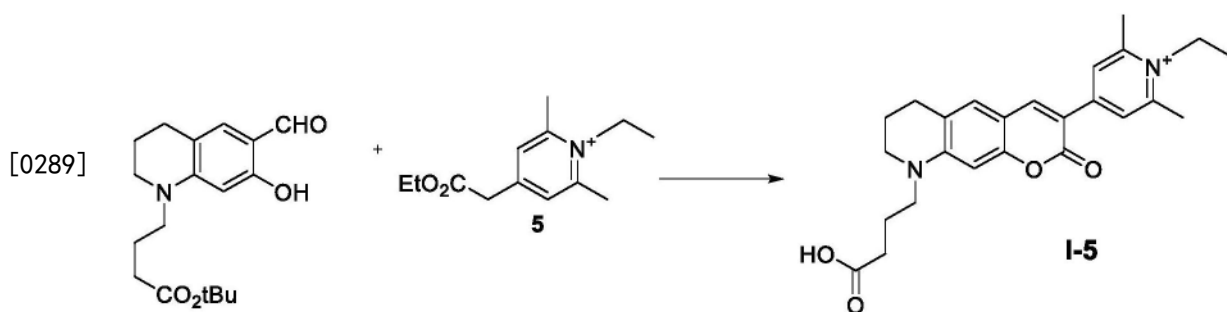
[0284] 将起始材料和催化量的吡啶乙酸盐溶解于乙醇中并回流直到反应完成。将粗产物蒸发至干燥,然后将残余物溶解于含1mL三氟乙酸的1mL水中,并且加热至90℃持续1小时。将反应在减压下蒸发,然后将粗产物通过快速柱色谱法在反相C18上进行纯化。以75%的产率(74mg, 0.123mmol)获得呈棕色固体的化合物I-2。 1H NMR (400MHz, MeOD) : δ (ppm) 8.73 (d, $J=6.1$ Hz, 1H, Ar-H pyr), 8.56 (s, 1H, Ar-H), 8.49 (br s, 1H, Ar-H pyr), 8.43 (br s, 1H, Ar-H pyr), 7.28 (s, 1H, Ar-H), 6.67 (s, 1H, Ar-H), 4.73 (t, $J=8.0$ Hz, 2H, CH_2-N^+), 3.58-3.45 (m, 4H, CH_2-N), 2.98 (t, $J=6.6$ Hz, 2H, $CH_2-SO_3^-$), 2.91 (s, 3H, CH_3 pyr), 2.82 (t, $J=6.2$ Hz, 2H, CH_2-Ar), 2.39 (m, 2H, $CH_2-CH_2-N^+$), 2.32 (t, $J=7.1$ Hz, 2H, CH_2-COO), 2.02-1.91 (m, 4H, CH_2-CH_2-N , $CH_2-CH_2-N^+$)。LC-MS (ESI) : (正离子) m/z 501 ($M+H^+$) ; (负离子) m/z 499 (M^-)。



[0286] 按照针对I-2所述的相同程序制备化合物I-3。以59%的产率(32mg, 0.064mmol)获得呈棕色固体的化合物I-3。¹H NMR (400MHz, MeOD): δ (ppm) 8.50 (d, $J=6.9$ Hz, 1H, Ar-H pyr), 8.41 (s, 1H, Ar-H), 8.27 (d, $J=2.3$ Hz, 1H, Ar-H pyr), 8.20 (dd, $J=6.8, 2.3$ Hz, 1H, Ar-H pyr), 7.25 (d, $J=1.5$ Hz, 1H, Ar-H), 6.45 (s, 1H, H-Ar), 4.50 (d, $J=8.2$ Hz, 2H, CH₂-N⁺), 3.38 (m, 1H, H-A CH₂-N), 3.17 (m, 1H, H-B CH₂-N), 2.74 (m, 3H, CH-Ar, CH₂-SO₃⁻), 2.68 (s, 3H, CH₃ pyr), 2.21-2.09 (m, 4H, CH₂-CH₂-N⁺, CH₂-COO), 1.71 (m, 4H, CH₂-CH₂-CH₂-&CH₂-CH), 1.24 (s, 3H, CH₃), 1.20 (d, $J=6.5$ Hz, 3H, CH₃), 1.12 (s, 3H, CH₃)。LC-MS (ESI): (正离子)m/z 543 (M+H⁺); (负离子)m/z 541 (M⁻)。



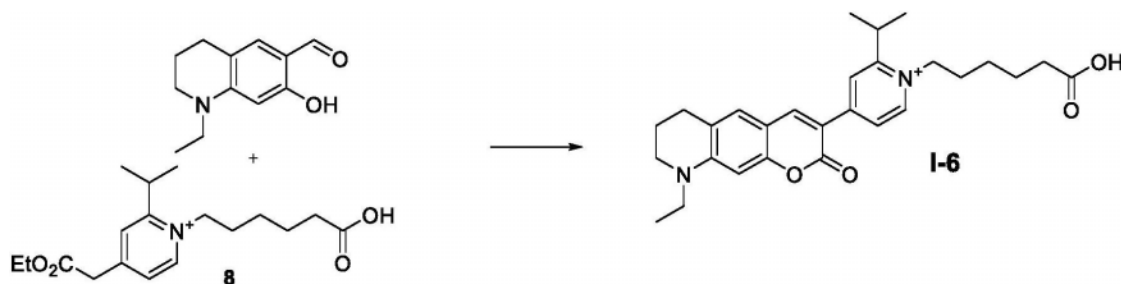
[0288] 按照针对I-2所述的相同程序制备化合物I-4。以84%的产率(77mg, 0.135mmol)获得呈棕色固体的化合物I-4。¹H NMR (400MHz, MeOD): δ (ppm) 8.57 (s, 1H, Ar-H), 8.35 (s, 2H, Ar-H pyr), 7.43 (d, $J=1.5$ Hz, 1H, Ar-H), 6.62 (s, 1H, H-Ar), 4.58 (q, $J=7.3$ Hz, 2H, CH₂-N⁺), 3.62 (m, 1H, H-A CH₂-N), 3.34 (m, 1H, H-B CH₂-N), 2.91 (s, 7H, CH₃ pyr, CH₃-CH-Ar), 2.29 (t, $J=7.1$ Hz, 2H, CH₂-COO), 1.92 (m, 4H, CH₂-CH₂-N, CH₂-CH), 1.56 (t, $J=7.3$ Hz, 3H, CH₃Et), 1.46 (s, 3H, CH₃), 1.41 (d, $J=6.5$ Hz, 3H, CH₃), 1.34 (s, 3H, CH₃)。LC-MS (ESI): (正离子)m/z 463 (M+H⁺); (负离子)m/z 461 (M⁻)。



[0290] 按照针对I-2所述的相同程序制备化合物I-5。以66%的产率(79mg, 0.139mmol)获得呈棕色固体的化合物I-4。¹H NMR (400MHz, MeOD): δ (ppm) 8.48 (s, 1H, Ar-H), 8.33 (s, 2H, Ar-H pyr), 7.24 (s, 1H, Ar-H pyr), 6.60 (s, 1H, Ar-H), 4.57 (q, $J=7.3$ Hz, 2H, CH₂-N⁺), 3.55-3.43 (m, 4H, CH₂-N), 2.90 (s, 6H, CH₃ pyr), 2.80 (t, $J=6.3$ Hz, 2H, CH₂-Ar), 2.27 (t, $J=7.2$ Hz, 2H, CH₂-COO), 2.02-1.89 (m, 4H, CH₂-CH₂-N, CH₂-CH₂-COO), 1.55 (t, $J=7.3$ Hz, 3H,

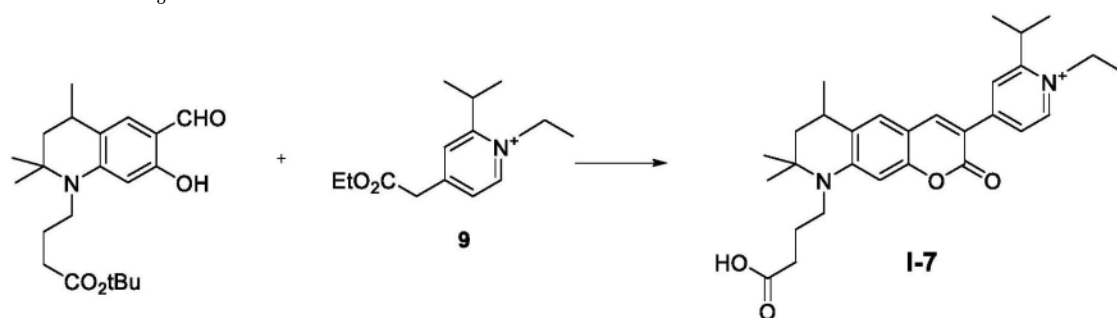
CH₃Et)。LC-MS (ESI) : (正离子)m/z 421 (M+H⁺)。

[0291]



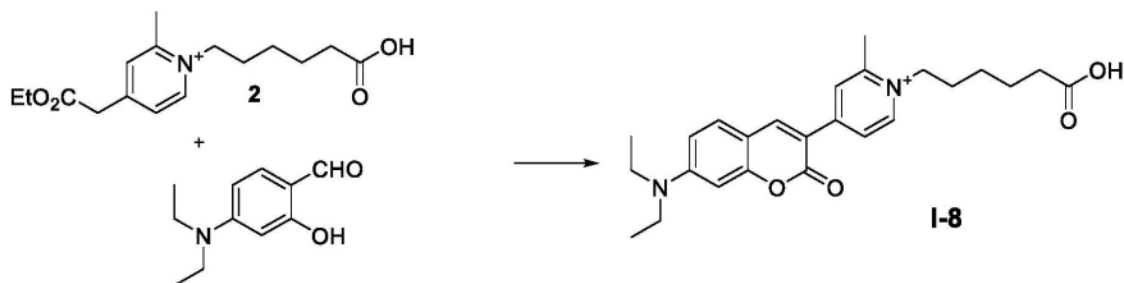
[0292] 按照针对I-1所述的相同程序制备化合物I-6。以33%的产率(47mg, 0.101mmol)获得呈棕色固体的化合物I-6。¹H NMR (400MHz, MeOD) : δ (ppm) 8.69 (d, J=6.9Hz, 1H, Ar-H pyr), 8.62 (m, 2H, Ar-H), 8.35 (dd, J=6.9, 2.3Hz, 1H, Ar-H pyr), 7.33 (d, J=1.2Hz, 1H, Ar-H), 6.58 (s, 1H, Ar-H), 4.62-4.54 (m, 2H, CH₂-N⁺), 3.60-3.48 (m, 5H, CH₂-N, CH iPr), 2.84 (t, J=6.2Hz, 2H, CH₂-Ar), 2.22 (t, J=7.1Hz, 2H, CH₂-COO), 2.05-1.92 (m, 4H, CH₂-CH₂-N, CH₂-CH₂-N⁺), 1.71 (p, J=7.2Hz, 2H, CH₂-CH₂-COO), 1.51 (m, 8H, CH₃ iPr, CH₂-CH₂-CH₂-COO), 1.27 (t, J=7.1Hz, 3H, CH₃ Et)。LC-MS (ESI) : (正离子)m/z 463 (M+H⁺)。

[0293]



[0294] 按照针对I-2所述的相同程序制备化合物I-7。以33%的产率(48mg, 0.1mmol)获得呈棕色固体的化合物I-7。¹H NMR (400MHz, MeOD) : δ (ppm) 8.68 (d, J=6.9Hz, 1H, Ar-H), 8.66 (s, 1H, Ar-H), 8.62 (d, J=2.3Hz, 1H, Ar-H), 8.38 (dd, J=6.9, 2.3Hz, 1H, Ar-H), 7.47 (d, J=1.5Hz, 1H, Ar-H), 6.64 (s, 1H, Ar-H), 4.65 (q, J=7.3Hz, 2H, CH₂-N⁺), 3.62 (m, 1H, H-A CH₂-N), 3.44-3.35 (m, 1H, H-B CH₂-N), 2.91 (m, 1H, CH iPr), 2.32 (t, J=7.2Hz, 2H, CH₂-COO), 2.00-1.87 (m, 4H, CH₂-CH₂-N, CH₂-C-N), 1.62 (t, J=7.3Hz, 3H, CH₃ Et), 1.52 (d, J=6.8Hz, 6H, CH₃ iPr), 1.47 (s, 3H, C-CH₃), 1.42 (d, J=6.5Hz, 3H, CH-CH₃), 1.36 (s, 3H, C-CH₃)。LC-MS (ESI) : (正离子)m/z 477 (M+H⁺)。

[0295]



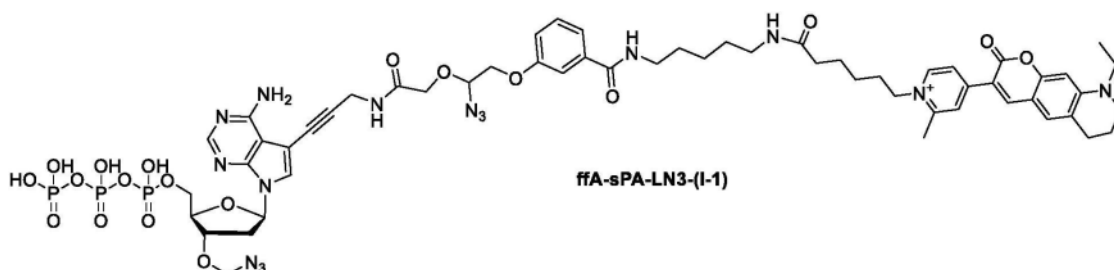
[0296] 按照针对I-1所述的相同程序制备化合物I-8。以61%的产率(69mg, 0.16mmol)获得呈棕色固体的化合物I-8。¹H NMR (400MHz, MeOD) : δ (ppm) 8.74 (d, J=6.9Hz, 1H, Ar-H pyr), 8.65 (s, 1H, Ar-H), 8.50 (d, J=2.3Hz, 1H, Ar-H pyr), 8.43 (dd, J=6.9, 2.3Hz, 1H, Ar-H pyr), 7.61 (d, J=9.0Hz, 1H, Ar-H), 6.88 (dd, J=9.1, 2.5Hz, 1H, Ar-H), 6.63 (d, J=

2.4Hz, 1H, Ar-H), 4.52 (d, $J=7.5\text{Hz}$, 2H, $\text{CH}_2\text{-N}^+$), 3.67-3.53 (m, 4H, $\text{CH}_2\text{-N}$), 2.89 (s, 3H, CH_3 pyr), 2.22 (t, $J=7.1\text{Hz}$, 2H, $\text{CH}_2\text{-COO}$), 1.98 (p, $J=7.7\text{Hz}$, 2H, $\text{CH}_2\text{-CH}_2\text{-N}^+$), 1.71 (p, $J=7.3\text{Hz}$, 2H, $\text{CH}_2\text{-CH}_2\text{-COO}$), 1.59-1.46 (m, 2H, $\text{CH}_2\text{-CH}_2\text{-CH}_2\text{-}$), 1.28 (t, $J=7.1\text{Hz}$, 6H, CH_3 Et)。
LC-MS (ESI): (正离子) m/z 423 ($\text{M}+\text{H}^+$); (负离子) m/z 421 (M^-)。

[0297] 实施例2. 烷基吡啶正离子香豆素染料标记的核苷酸的一般合成

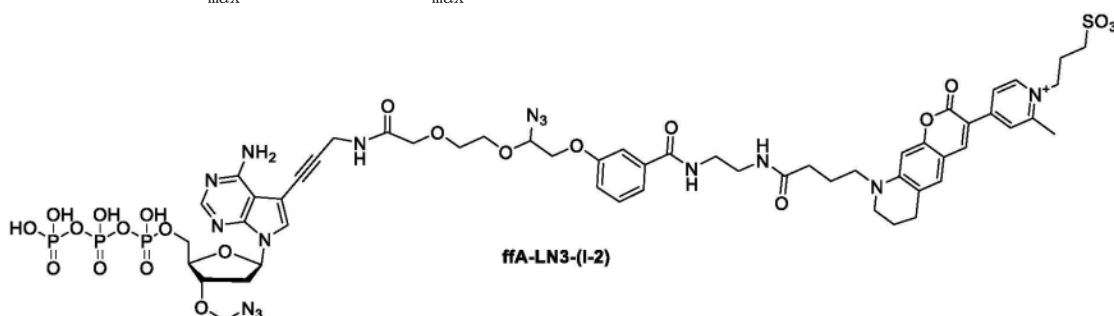
[0298] 将烷基吡啶正离子香豆素染料 (0.015mmol) 与 $2 \times 2\text{mL}$ 无水 $\text{N,N}'$ -二甲基甲酰胺 (DMF) 共蒸发, 然后溶解于 1mL 的无水 $\text{N,N}'$ -二甲基乙酰胺 (DMA) 中。添加 N,N -二异丙基乙胺 ($17\mu\text{L}$, 0.1mmol), 随后添加 $\text{N,N,N}',\text{N}'$ -四甲基- 0 -(N -琥珀酰亚胺基) 脲四氟硼酸盐 (TSTU, 4.8mg, 0.016mmol)。在氮气下在室温下搅拌反应30分钟。同时, 将核苷三磷酸 (0.01mmol) 的水溶液在减压下蒸发至干燥, 并重悬于 $100\mu\text{L}$ 的 0.1M 三乙基碳酸氢铵 (TEAB) 水溶液中。将活化的染料溶液添加到三磷酸盐中, 并将反应在室温下搅拌至多18小时并且通过 RP-HPLC 监测。粗产物通过离子交换色谱法在 DEAE-Sephadex A25 (25g) 上, 用三乙基碳酸氢铵水溶液 (TEAB, 从 0.1M 到 1M) 的线性梯度洗脱来进行纯化。汇集含有三磷酸盐的级分, 并将溶剂在减压下蒸发至干燥。使用 YMC-Pack-Pro C18 柱通过制备型 HPLC 进一步纯化粗材料。

[0299]



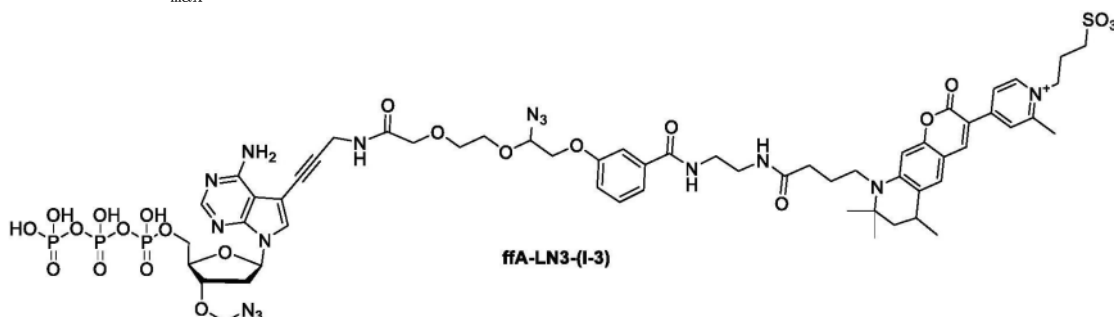
[0300] ffA-sPA-LN3- (I-1): 产率: $8.5\mu\text{mol}$, (53%)。LC-MS (ES): (负离子) m/z 1360 (M-H^+), 680 ($\text{M}-2\text{H}^+$)。UV-Vis λ_{max} : 507nm。荧光 λ_{max} : 566nm。

[0301]



[0302] ffA-LN3- (I-2): 产率: $1.9\mu\text{mol}$, (38%)。LC-MS (ES): (负离子) m/z 1428 (M-H^+)。UV-Vis λ_{max} : 502nm。荧光 λ_{max} : 563nm。

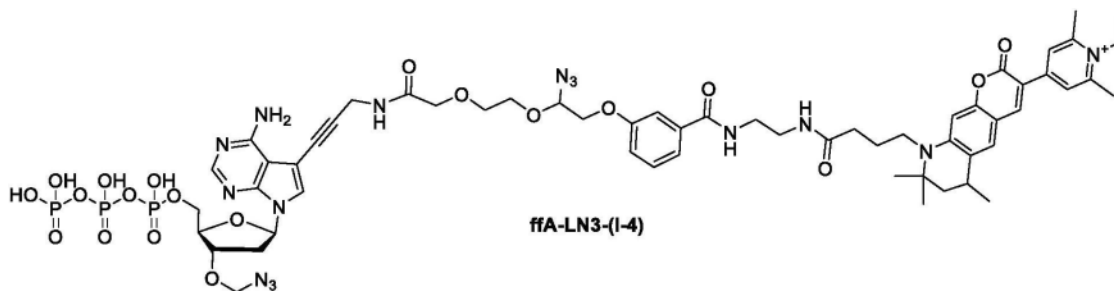
[0303]



[0304] ffA-LN3- (I-3): 产率: $30.7\mu\text{mol}$, (61%)。LC-MS (ES): (负离子) m/z 1470 (M-H^+)。

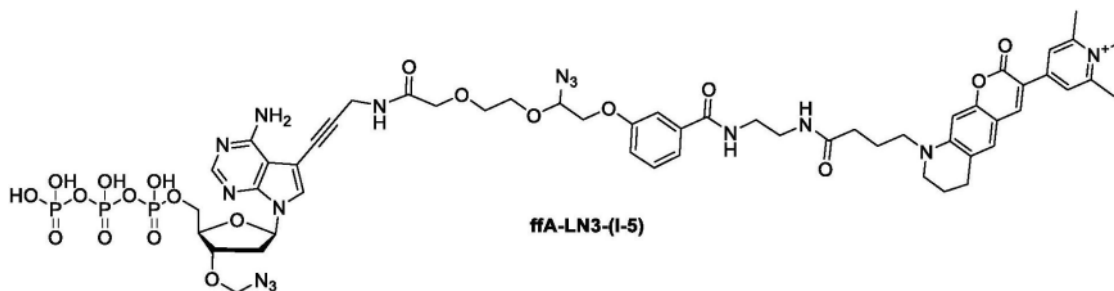
UV-Vis $\lambda_{\max}$:501nm。荧光 $\lambda_{\max}$:565nm。

[0305]



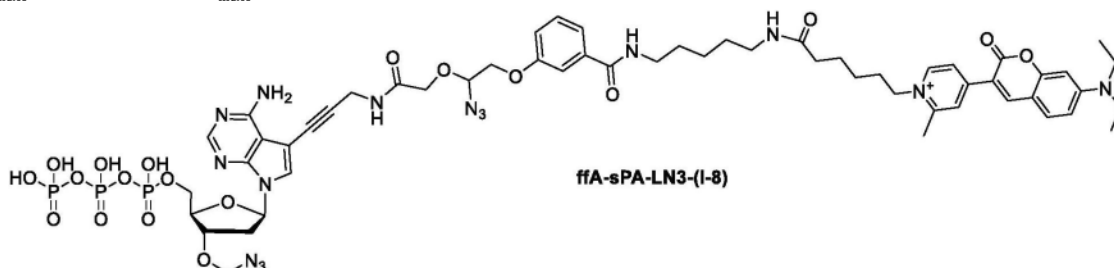
[0306] ffA-LN3- (I-4) :产率:1.4 μ mol, (29%)。LC-MS (ES) : (负离子) m/z 1390 (M-H⁺)。UV-Vis $\lambda_{\max}$:493nm。荧光 $\lambda_{\max}$:551nm。

[0307]



[0308] ffA-LN3- (I-5) :产率:2.7 μ mol, (54%)。LC-MS (ES) : (负离子) m/z 1348 (M-H⁺)。UV-Vis $\lambda_{\max}$:491nm。荧光 $\lambda_{\max}$:555nm。

[0309]



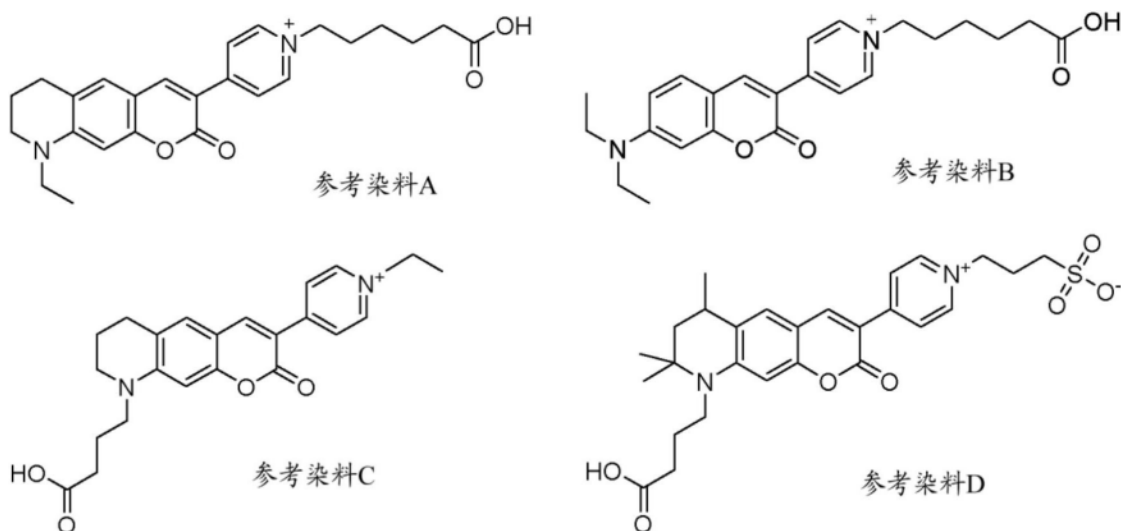
[0310] ffA-sPA-LN3- (I-8) :产率:7.8 μ mol, (78%)。LC-MS (ES) : (负离子) m/z 1348 (M-H⁺)。UV-Vis $\lambda_{\max}$:495nm。荧光 $\lambda_{\max}$:556nm。

[0311] 实施例3.烷基吡啶正离子香豆素染料的光谱特性

[0312] 在此实施例中,将本文所述的几种烷基吡啶正离子香豆素染料的光谱特性与没有甲基化的对应参考染料进行比较。在图1A和图1B中,分别在450nm(“蓝光”)和520nm(“绿光”)激发波长下将甲基吡啶正离子香豆素染料I-1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)的溶液中的荧光发射与参考染料A的荧光发射进行比较。使用石英比色皿在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。观察到与参考染料A相比,染料I-1显示出,在绿光激发时,荧光发射增加了大约2倍,并且在蓝光激发时,荧光发射增加了大约3倍。

[0313] 类似地,图2A和图2B分别展示了在450nm和520nm激发波长下,与参考染料C的荧光发射相比,甲基吡啶正离子香豆素染料I-5在USM溶液中的荧光发射。图3A和图3B分别展示了在450nm和520nm激发波长下,与参考染料B的荧光发射相比,甲基吡啶正离子香豆素染料I-8在USM溶液中的荧光发射。与参考染料C相比,香豆素染料I-5显示出,在绿光激发时,荧光发射增加了大约2.5倍,并且在蓝光激发时,荧光发射增加了大约8倍。与参考染料B相比,

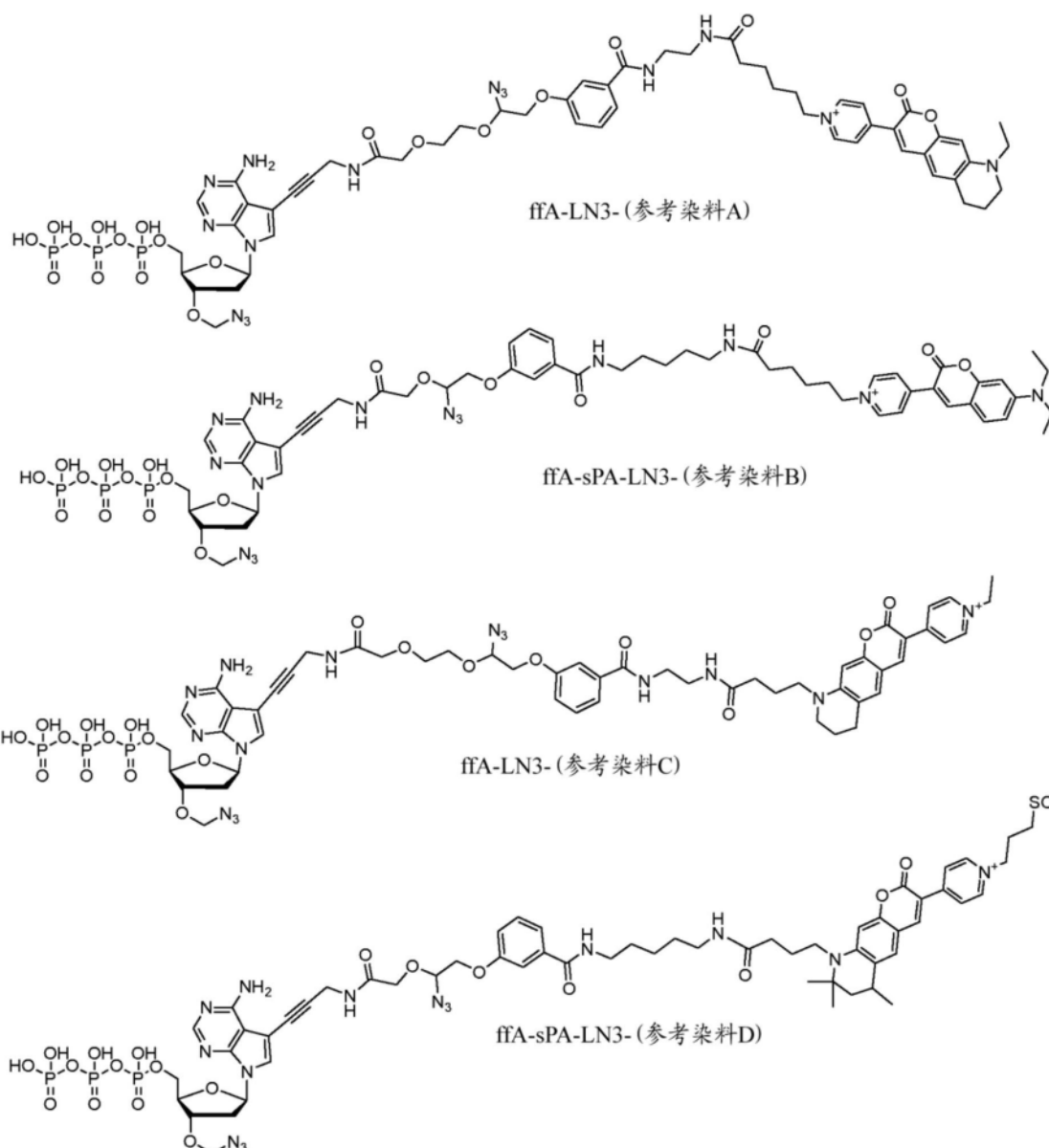
香豆素染料I-8显示出,在绿光激发时,荧光发射类似,并且在蓝光(450nm)激发时,荧光发射增加了大约2.5倍。



[0315] 实施例4. 烷基吡啶正离子香豆素缀合的ffA核苷酸的光谱特性

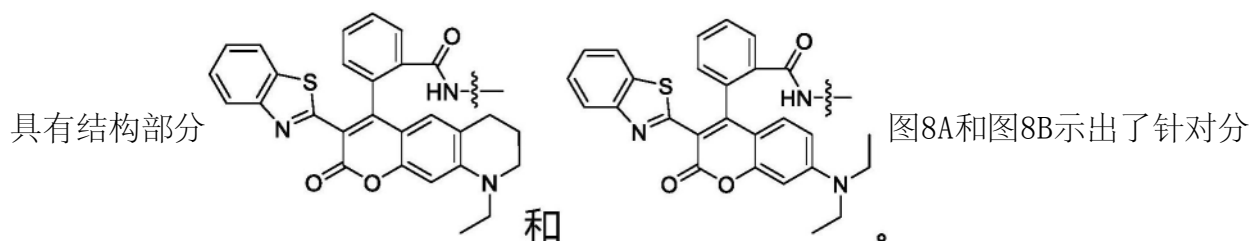
[0316] 在此实施例中,将与本文所述的甲基吡啶正离子香豆素染料缀合的几种完全官能化A核苷酸(ffA)的光谱特性与没有甲基化的对应参考染料进行比较。在图4A和图4B中,分别在450nm(“蓝光”)和520nm(“绿光”)激发波长下将与甲基吡啶正离子香豆素染料I-1缀合的ffA(呈于USM中的2 μ M溶液的形式)的荧光发射与参考染料A的荧光发射进行比较。使用石英比色皿在Agilent Cary 100UV-Vis分光光度计上和Cary Eclipse荧光分光光度计上获取光谱。观察到与ffA-sPA-LN3-(参考染料A)相比,ffA-sPA-LN3-(I-1)显示出,在绿光激发时,荧光发射类似,并且在蓝光激发时,荧光发射增加了大约3倍。

[0317] 类似地,图5A和图5B分别展示了在450nm和520nm激发波长下,与参考染料C的荧光发射相比,与甲基吡啶正离子香豆素染料I-5缀合的ffA(呈于USM中的2 μ M溶液的形式)的荧光发射。图6A和图6B分别展示了在450nm和520nm激发波长下,与参考染料B的荧光发射相比,与甲基吡啶正离子香豆素染料I-8缀合的ffA(呈于USM中的2 μ M溶液的形式)的荧光发射。图7A和图7B分别展示了在450nm和520nm激发波长下,与参考染料D的荧光发射相比,与甲基吡啶正离子香豆素染料I-3缀合的ffA(呈于USM中的2 μ M溶液的形式)的荧光发射。与ffA-sPA-LN3-(参考染料C)相比,ffA-LN3-(I-5)显示出,在绿光(520nm)激发时,荧光发射类似,并且在蓝光(450nm)激发时,荧光发射增加了大约2.5倍。与ffA-sPA-LN3-(参考染料B)相比,ffA-sPA-LN3-(I-8)显示出,在绿光(520nm)激发时,荧光发射类似,并且在蓝光(450nm)激发时,荧光发射增加了大约1.6倍。与ffA-sPA-LN3-(参考染料D)相比,ffA-sPA-LN3-(I-3)显示出,在绿光(520nm)激发时,荧光发射类似,并且在蓝光(450nm)激发时,荧光发射增加了大约1.2倍。



[0319] 实施例5. 因美纳iSeq™100仪器上的测序实验

[0320] 在此实施例中, 在因美纳iSeq™100仪器上测试了本文所述的ffA-连接基-染料化合物, 该仪器已经设置为在绿色激发光(约520nm)的情况下拍摄第一图像并且在蓝色激发光(约450nm)的情况下拍摄第二图像。修改测序配方以便执行标准SBS循环(掺入, 随后成像, 随后进行裂解)。这些实验中的每个实验中使用的掺入混合物含有以下四种ffN: 1) 与本文所述的烷基吡啶正离子染料缀合的ffA或与本文所述的参考染料缀合的ffA; 2) 在使用与本文所述的烷基吡啶正离子染料缀合的ffA的情况下, 可用450nm的蓝光激发的ffC(例如ffC-连接基-香豆素染料E), 以及在使用与本文所述的参考染料缀合的ffA的情况下的ffC-连接基-香豆素染料F; 3) 可用绿光激发的ffT(例如ffT-LN3-NR550s0); 以及4) 未标记的ffG(暗色ffG), 在50mM乙醇胺缓冲液, pH 9.6、50mM NaCl、1mM EDTA、0.2% CHAPS、4mM MgSO₄和DNA聚合酶中。香豆素染料E和F公开于美国公开第2018/0094140号中, 当与ffC缀合时分别



别含有ffA-sPA-LN3- (I-1) 和ffA-sPA-LN3- (参考染料A) 的掺入混合物获得的散点图。图8C和图8D示出了针对分别含有ffA-sPA-LN3- (I-3) 和ffA-sPA-LN3- (参考染料D) 的掺入混合物获得的散点图。在两种情况下都观察到, 与在吡啶正离子部分处没有甲基取代的对应参考染料相比, 含有与烷基吡啶正离子香豆素染料缀合的ffA的掺入混合物提供增加的强度和更好的A-云 (A-cloud) 分离。

[0321] 表1展示了与使用标准试剂和标准配方在因美纳iSeq™100仪器上运行的 2×151 个循环获得的度量相比, ffA-sPA-LN3- (I-1)、ffA-LN3- (I-3) 和ffA-sPA- (参考染料D) 的在iSeq™100上运行的 2×151 个循环的定相、预定相和PhiX错误率度量。这些实验中的每个实验中使用的掺入混合物含有以下四种ffN: 1) 与本文所述的烷基吡啶正离子染料缀合的ffA或与本文所述的参考染料缀合的ffA; 2) 在使用与本文所述的烷基吡啶正离子染料缀合的ffA的情况下, 可用450nm的蓝光激发的ffC (例如, ffC-连接基-香豆素染料E), 以及在使用与本文所述的参考染料缀合的ffA的情况下的ffC-连接基-香豆素染料F; 3) 可用绿光激发的ffT (例如ffT-LN3-NR550s0); 以及4) 未标记的ffG (暗色ffG), 在50mM乙醇胺缓冲液, pH 9.6、50mM NaCl、1mM EDTA、0.2% CHAPS、4mM MgSO₄和DNA聚合酶中。与ffA-sPA- (参考染料D) 和标准iSeq™100试剂相比, 用化合物ffA-sPA-LN3- (I-1) 和ffA-LN3- (I-3) 观察到了PhiX错误率度量的改善。

[0322] 表1: iSeq™100测序度量比较 (2×151 个循环)

	标准 iSeq™100 试剂	ffA-sPA-LN3- (参考染料 D)	ffA-sPA-LN3- (I-1)	ffA-sPA-LN3- (I-3)
定相	0.13	0.1	0.07	0.17
预定相	0.17	0.3	0.18	0.28
错误率 (读段 1, 读段 2)	0.29, 0.43	0.5, 0.75	0.23, 0.24	0.38, 0.43

[0324] 此外, 选择了化合物ffA-sPA-LN3- (I-1) 以在因美纳iSeq™100仪器上运行 2×300 个循环。仪器被设置为在绿色激发光的情况下拍摄第一图像并且在蓝色激发光的情况下拍摄第二图像, 并且修改配方以针对 2×300 个循环执行标准SBS循环 (掺入, 随后成像, 随后进行切割)。这些实验中使用的掺入混合物含有以下四种ffN: 1) ffA-sPA-LN3- (I-1); 2) 1可用450nm的蓝光激发的ffC (例如, ffC-连接基-香豆素染料E); 3) 可用绿光激发的ffT (例如ffT-连接基-NR550s0); 以及4) 暗色ffG, 在50mM乙醇胺缓冲液, pH 9.6、50mM NaCl、1mM EDTA、0.2% CHAPS、4mM MgSO₄和DNA聚合酶中。定相、预定相、PhiX错误率和%Q30度量示于下表2中。

[0325] 表2: iSeq™100测序度量 (2×300 个循环)

[0326]

	定相	预定相	错误率 (%)	%≥ Q30	%≥ Q30 (最后 10 个循环)
读段 1	0.099	0.108	0.79	88.89	72.64
读段 2	0.104	0.114	1.51	79.75	51.11

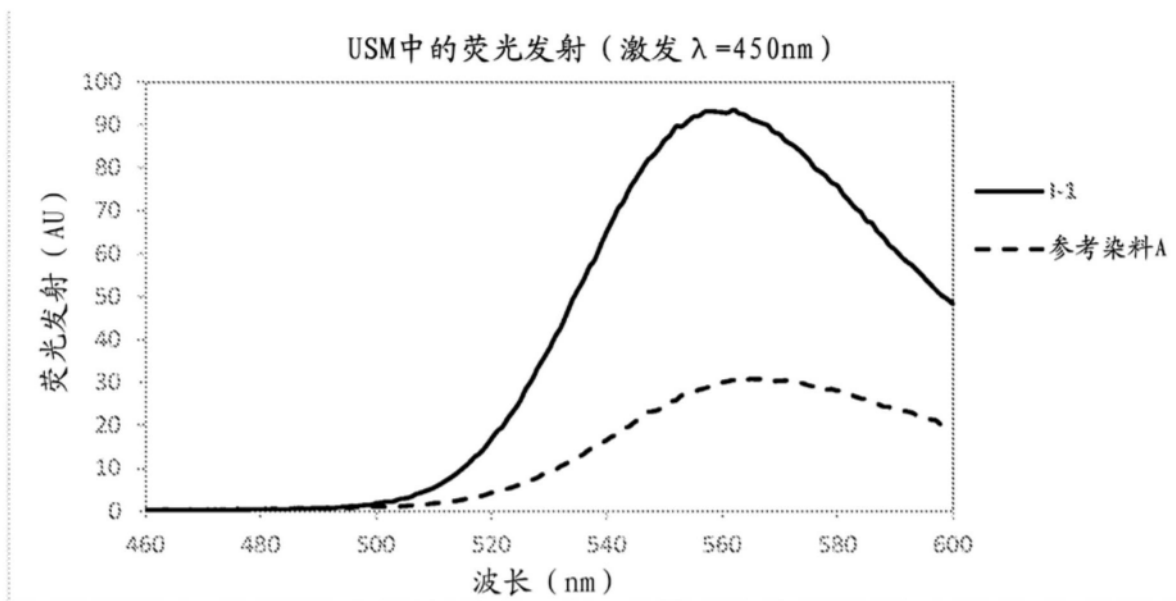


图1A

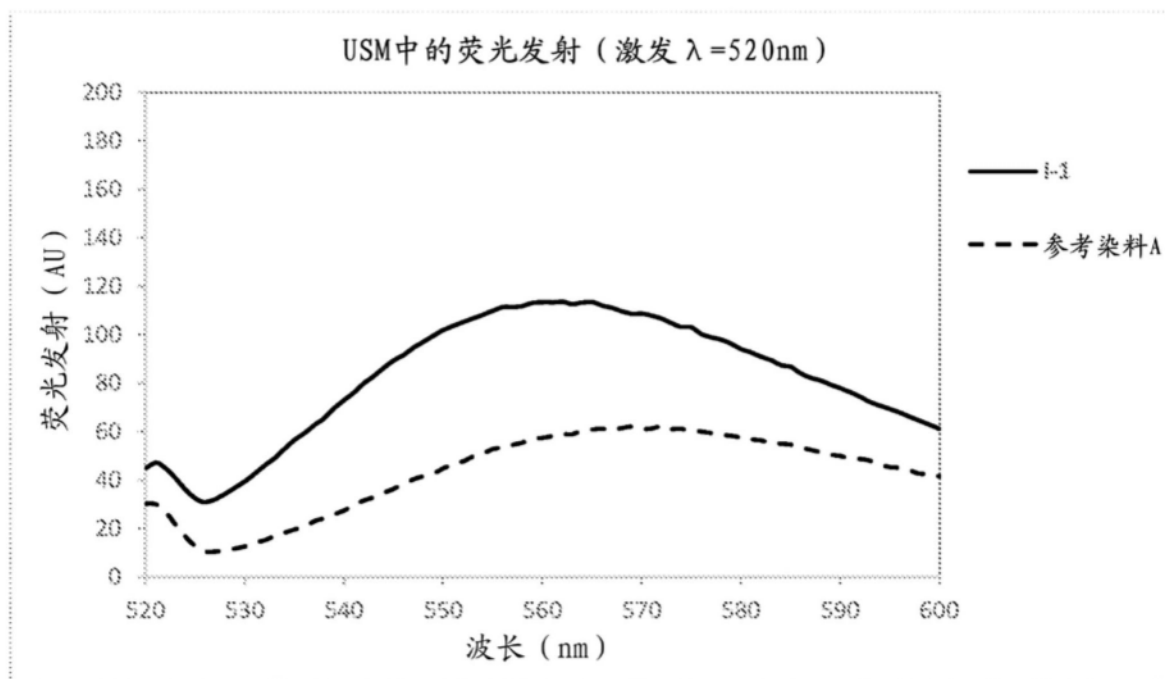


图1B

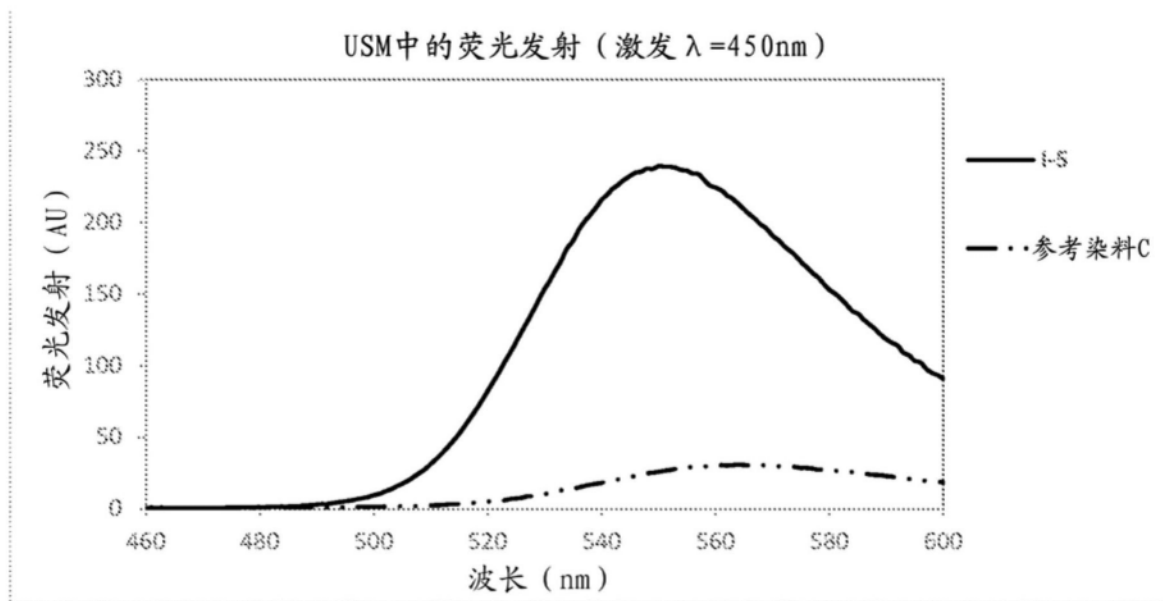


图2A

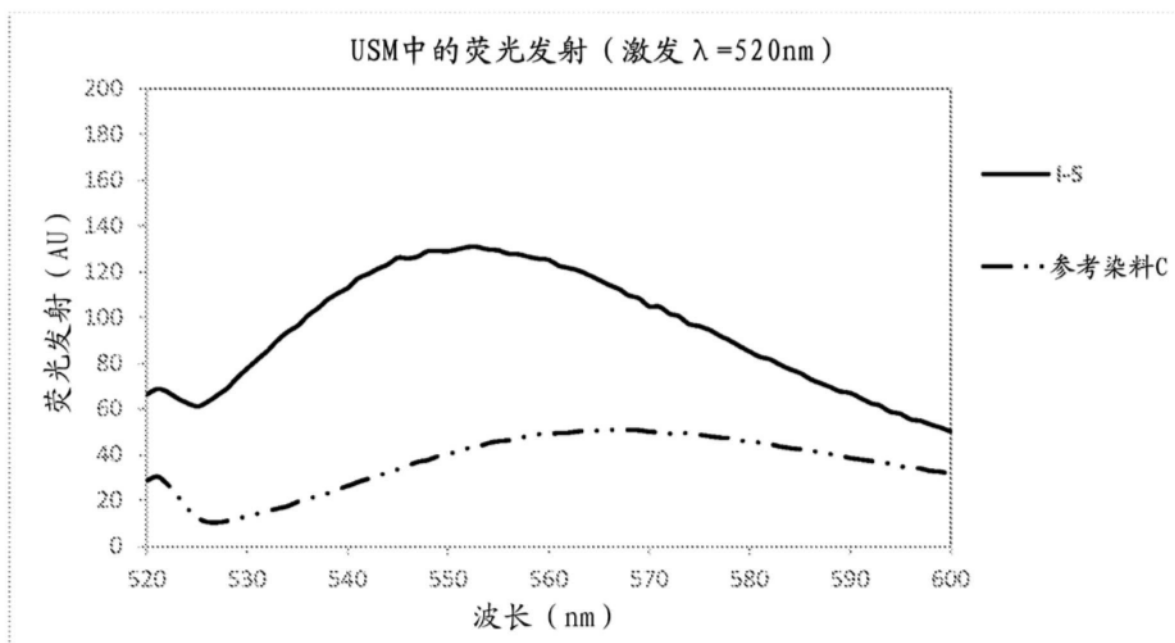


图2B

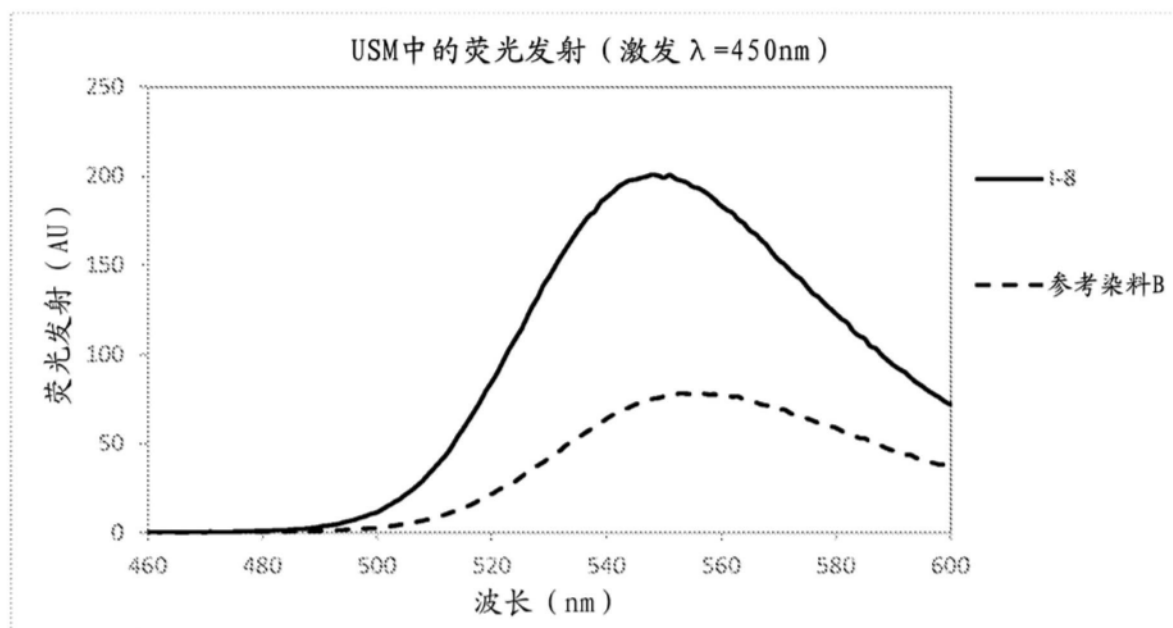


图3A

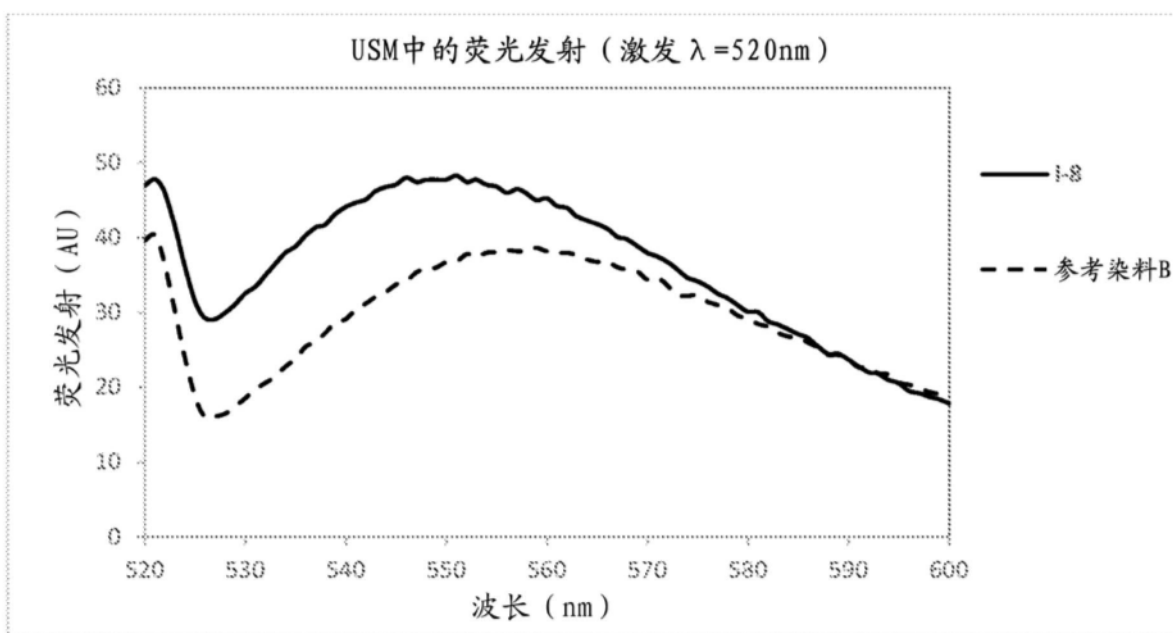


图3B

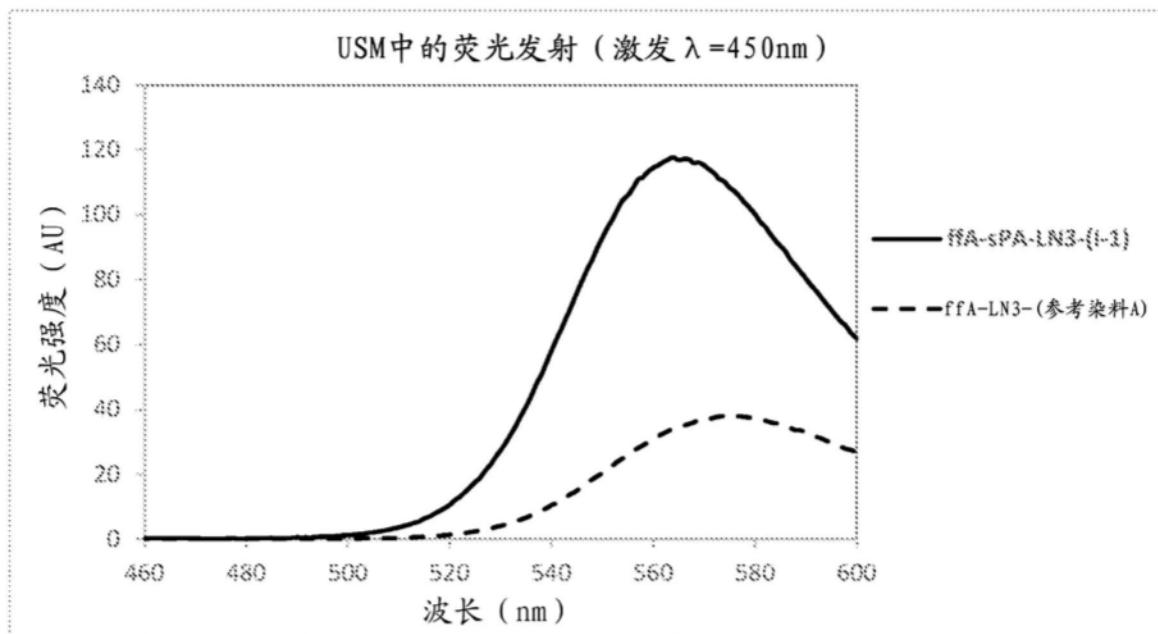


图4A

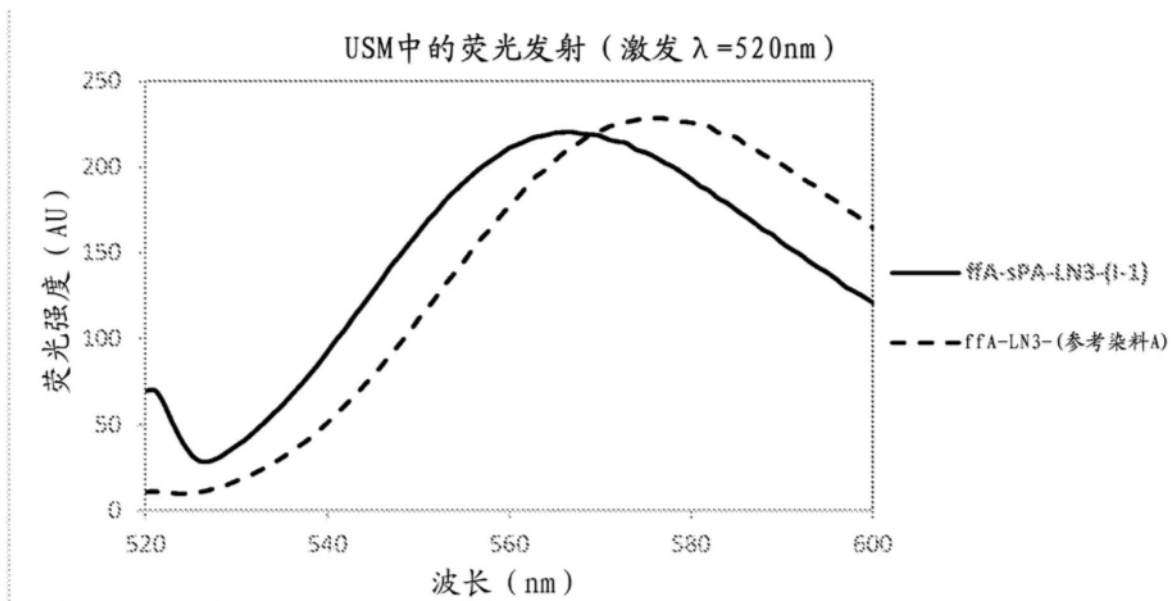


图4B

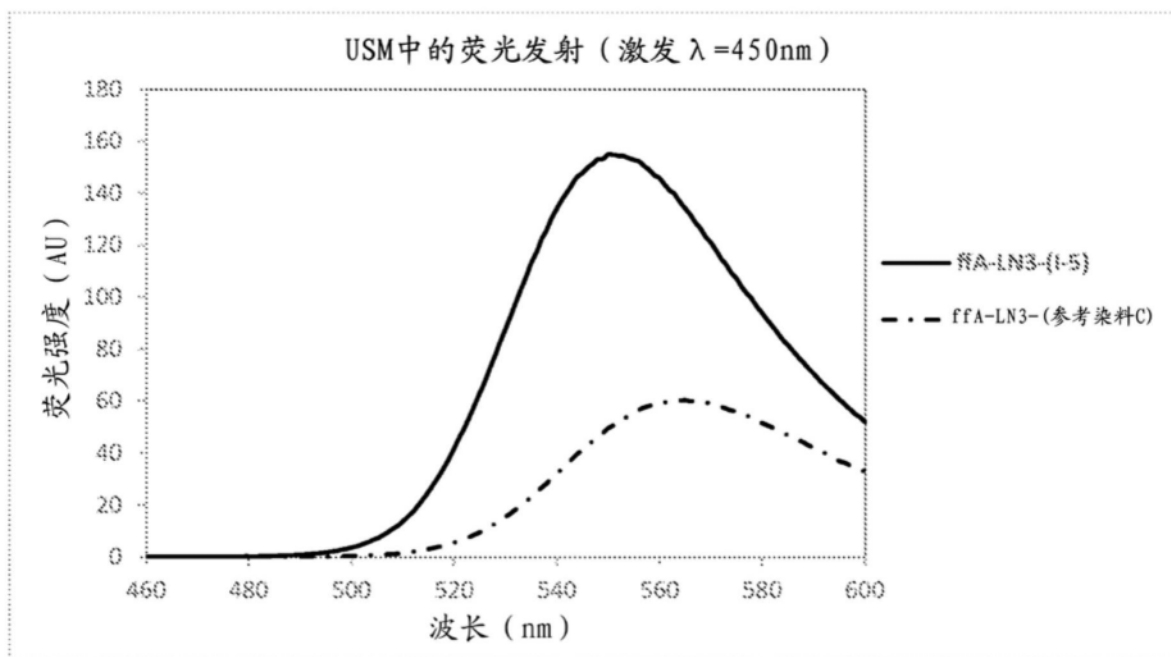


图5A

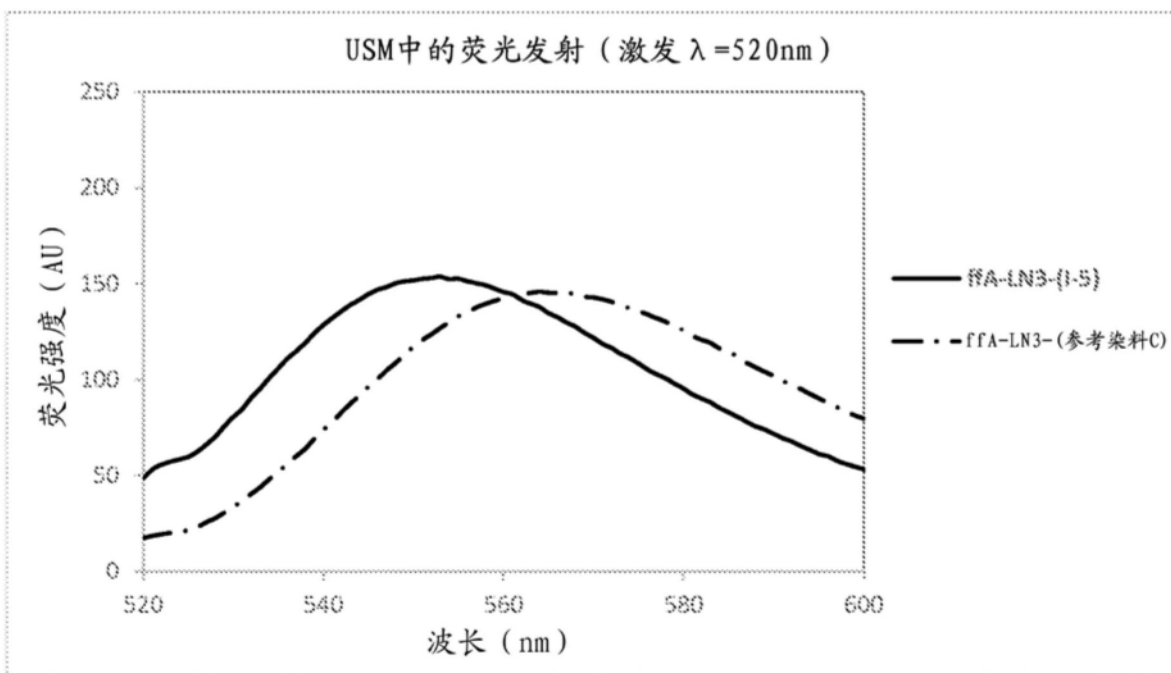


图5B

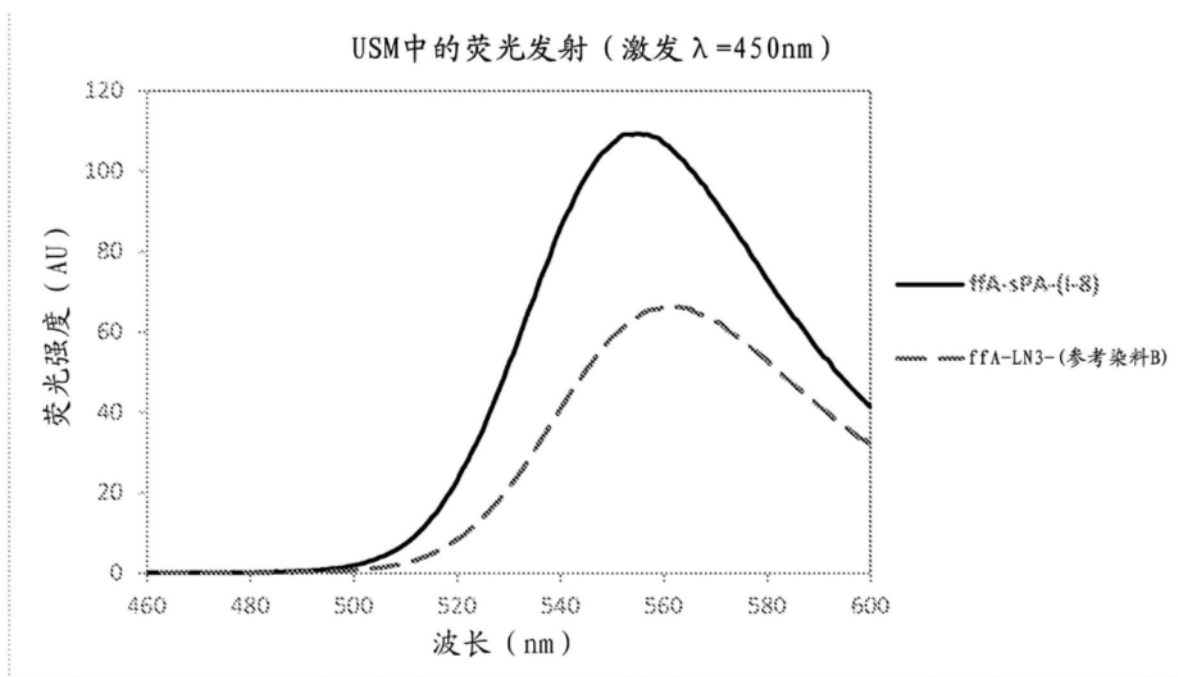


图6A

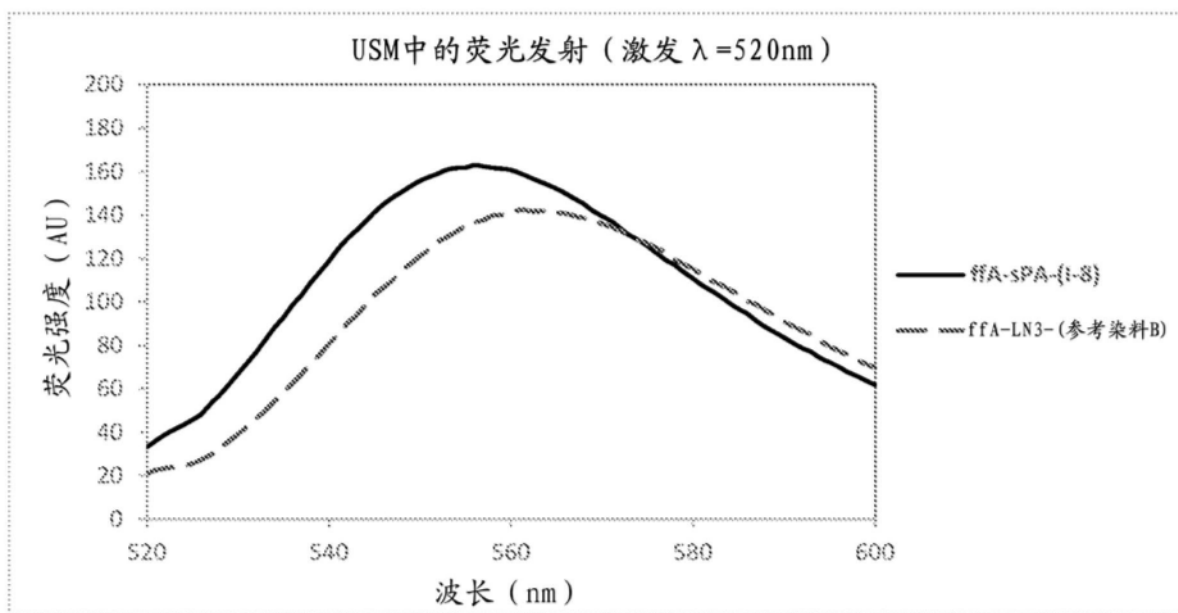


图6B

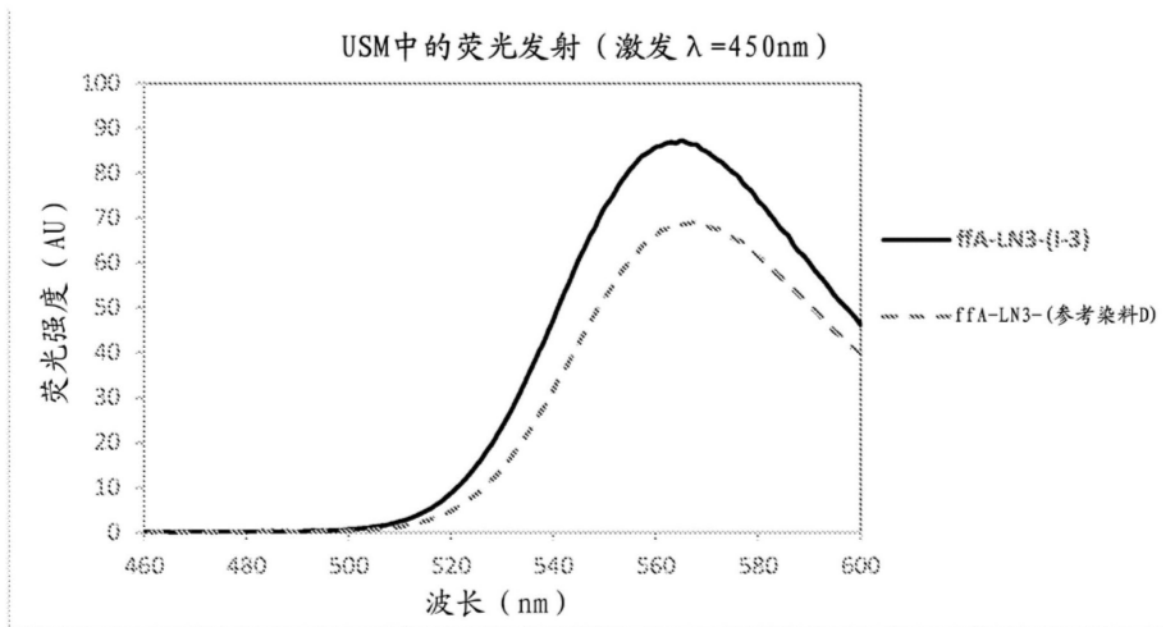


图7A

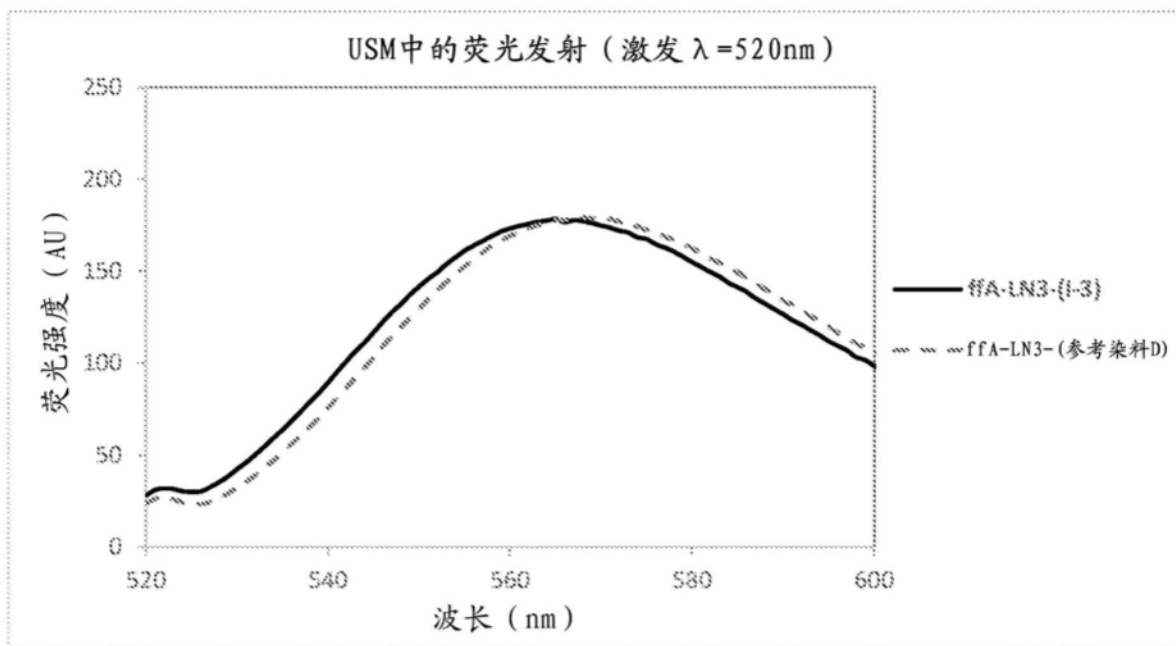


图7B

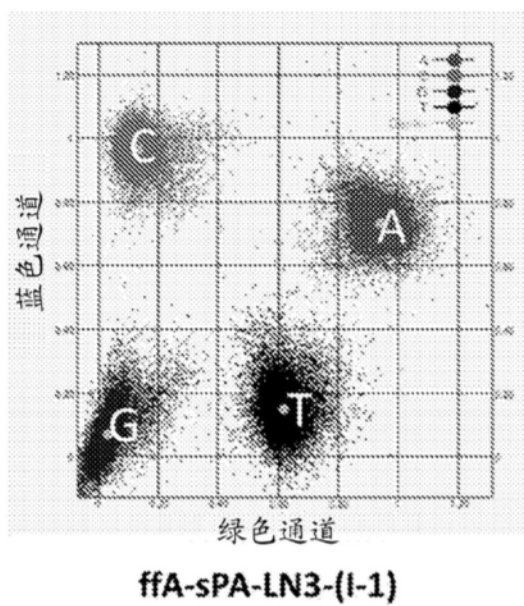


图8A

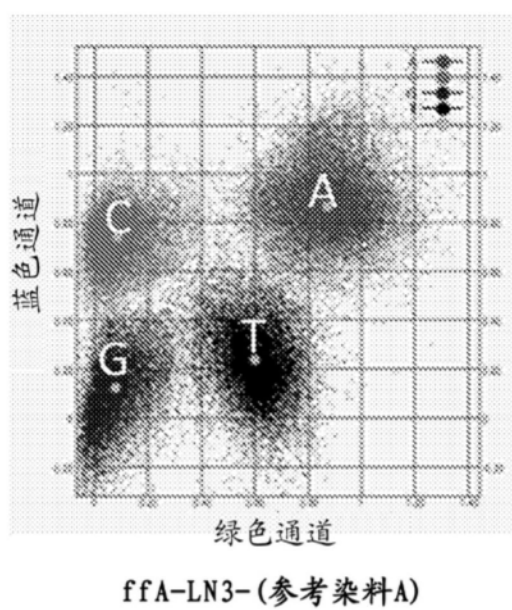


图8B

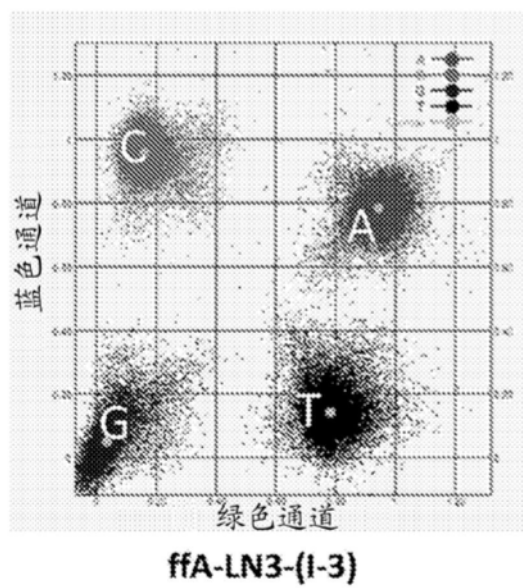


图8C

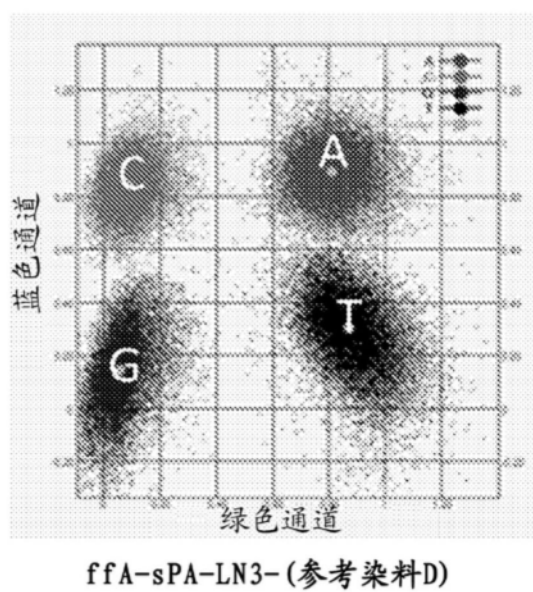


图8D