



(12)发明专利

(10)授权公告号 CN 106604926 B

(45)授权公告日 2020.05.15

(21)申请号 201580048113.1

(22)申请日 2015.08.06

(65)同一申请的已公布的文献号
申请公布号 CN 106604926 A

(43)申请公布日 2017.04.26

(30)优先权数据
1414098.2 2014.08.08 GB

(85)PCT国际申请进入国家阶段日
2017.03.08

(86)PCT国际申请的申请数据
PCT/GB2015/052282 2015.08.06

(87)PCT国际申请的公布数据
W02016/020691 EN 2016.02.11

(73)专利权人 伊卢米纳剑桥有限公司
地址 英国剑桥郡

(72)发明人 吴晓琳 刘小海

(74)专利代理机构 北京市柳沈律师事务所
11105

代理人 曹立莉 牟科

(51)Int.Cl.
C07H 19/14(2006.01)
C07H 1/00(2006.01)
C12Q 1/6869(2018.01)

(56)对比文件

W0 2002088381 A2,2002.11.07,
W0 2013044018 A1,2013.03.28,
W0 2005044836 A2,2005.05.19,
W0 2009051807 A1,2009.04.23,

审查员 刘冬

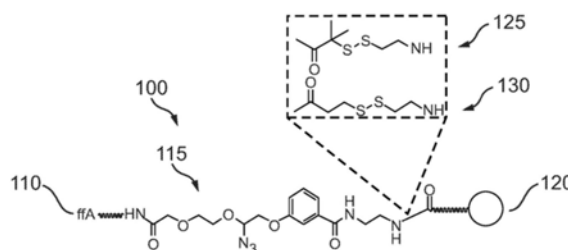
权利要求书1页 说明书21页 附图17页

(54)发明名称

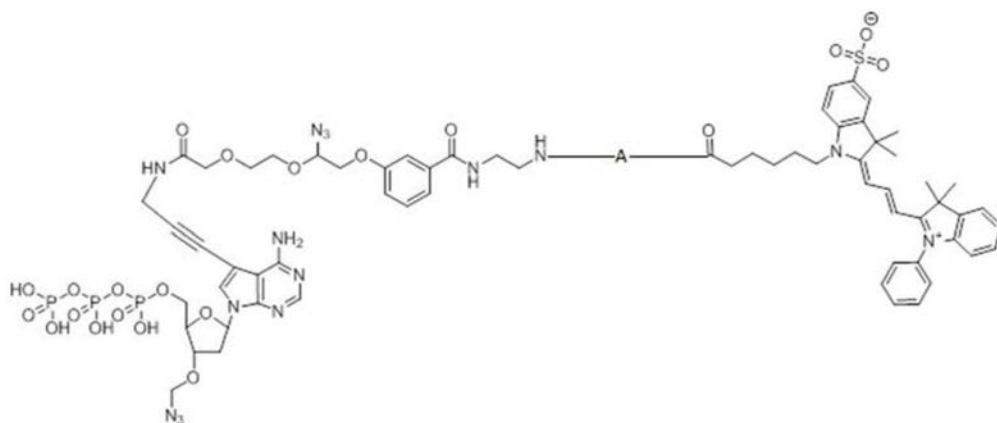
修饰的核苷酸连接基

(57)摘要

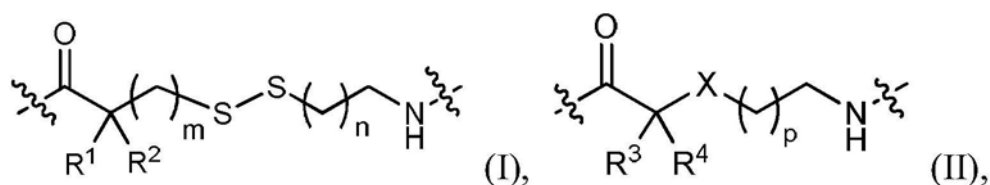
本申请的一些实施方案涉及在合成测序应用中用于增加核苷酸掺入的效率的新的修饰的核苷酸连接基。本申请还提供了制备这些修饰的核苷酸连接基的方法。



1. 通过连接基A共价连接到荧光团的核苷酸,所述核苷酸由下式表示:



其中连接基A具有式(I)或(II)的结构:



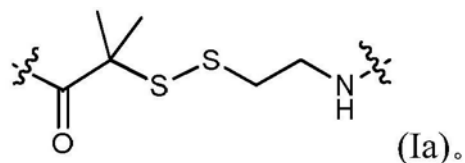
其中

R^1 和 R^2 为甲基, m 为0, n 为1;

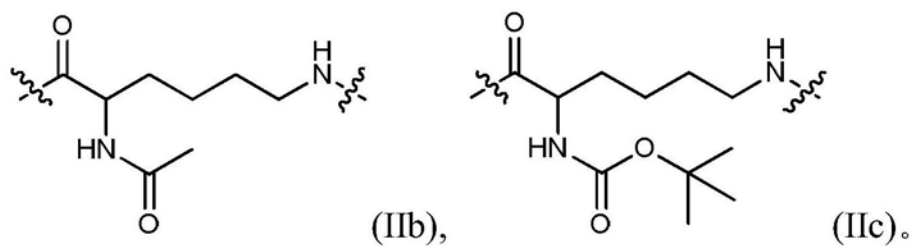
R^3 为 $-\text{NH}-\text{C}(=\text{O})-\text{O}-$ 叔丁基, R^4 为氢, X 为 CH_2 且 p 为2;或

R^3 为 $-\text{NH}-\text{C}(=\text{O})-\text{CH}_3$, R^4 为氢, X 为 CH_2 且 p 为2。

2. 根据权利要求1的核苷酸,其中连接基A具有式(Ia)的结构:



3. 根据权利要求1的核苷酸,其中连接基A具有式(IIb)或(IIc)的结构:



4. 试剂盒,其包含权利要求1-3中任一项所述的核苷酸。

修饰的核苷酸连接基

技术领域

[0001] 本申请的一些实施方案涉及用于在DNA测序和其它诊断应用(例如合成测序)中增加核苷酸掺入的新核苷或核苷酸连接基。

背景技术

[0002] 分子研究的进展部分地是通过改进用于表征分子或其生物反应的技术产生的。具体地,核酸DNA和RNA的研究已经从研发用于序列分析和杂交事件的技术中受益。

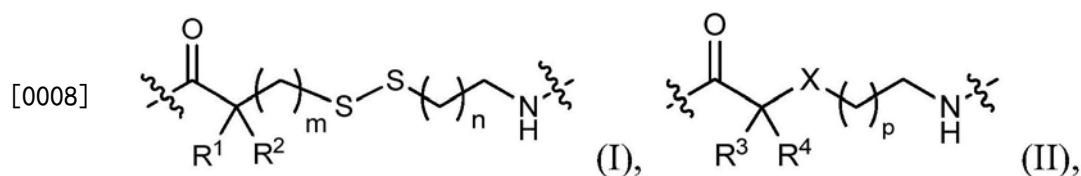
[0003] 已改进核酸研究的技术的一个实例是开发了固定化核酸的制备阵列。这些阵列通常具有固定在固体支持材料上的多核苷酸的高密度基质。参见,例如Fodor等,Trends Biotech.12:19-26,1994,其描述了使用化学敏化的玻璃表面组装不同核酸的方法,所述化学敏化玻璃表面由掩模保护,但在限定的区域暴露以允许连接适当修饰的核苷酸亚磷酸酰胺。制备的阵列也可以通过将已知的多核苷酸“点样”到固体支持物上的预定位置的技术来制造(例如,Stimpson等,Proc.Natl.Acad.Sci.92:6379-6383,1995)。

[0004] 确定与阵列结合的核酸的核苷酸序列的一种方法称为“合成测序”或“SBS”。用于测定DNA的核苷酸序列的这种技术理想地需要控制性(即一次一个)掺入与被测序的核酸相对的正确互补核苷酸。这允许通过在多个循环中添加核苷酸来进行精确测序,因为一次测序一个核苷酸残基,从而防止掺入不受控制的核苷酸序列。在去除标记部分和随后的下一轮测序之前,使用与其连接的合适标记来读取每个掺入的核苷酸。

[0005] 因此,在核酸测序反应的情形下,希望能够在合成测序过程中增加核苷酸掺入的速率,从而可以提高测序方法的效率。

[0006] 发明概述

[0007] 本文公开的一些实施方案涉及通过连接基共价连接到荧光团的核苷或核苷酸,其中所述连接基包含式(I)或(II)的结构,或两者的组合:



[0009] 每个 R^1 和 R^2 独立地选自氢或任选取代的 C_{1-6} 烷基;

[0010] R^3 选自氢、任选取代的 C_{1-6} 烷基、 $-NR^5-C(=O)R^6$ 或 $-NR^7-C(=O)-OR^8$;

[0011] R^4 选自氢或任选取代的 C_{1-6} 烷基;

[0012] 每个 R^5 和 R^7 独立地选自氢、任选取代的 C_{1-6} 烷基、任选取代的苯基、或任选取代的 C_{7-12} 芳烷基;

[0013] 每个 R^6 和 R^8 独立地选自任选取代的 C_{1-6} 烷基、任选取代的苯基、任选取代的 C_{7-12} 芳烷基、任选取代的 C_{3-7} 环烷基、或任选取代的5至10元杂芳基;

[0014]  中的每个亚甲基重复单元任选被取代;

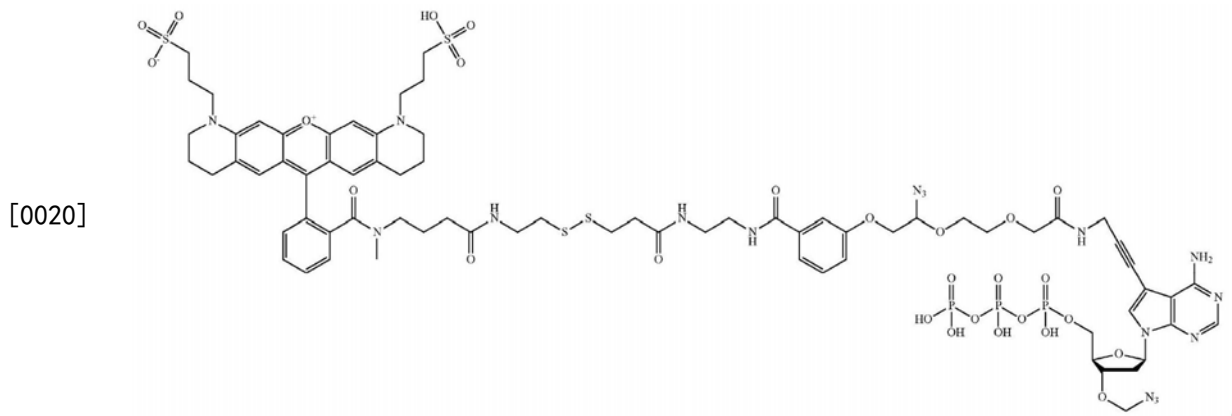
[0015] X选自亚甲基(CH₂)、氧(O)或硫(S)；

[0016] m为0至20的整数；

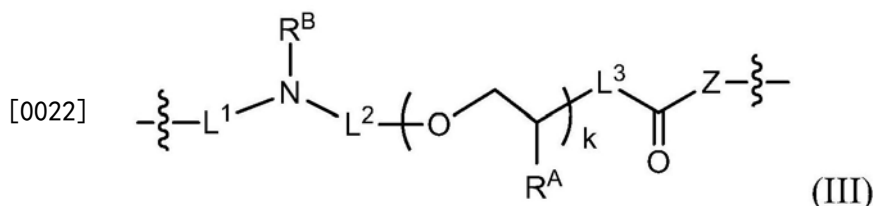
[0017] n为1至20的整数；和

[0018] p为1至20的整数。

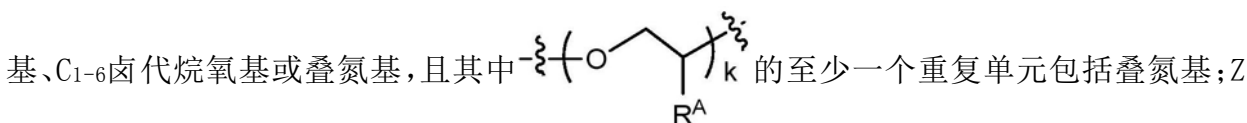
[0019] 在一些实施方案中,荧光团标记的核苷或核苷酸包含式(I)的结构,其不具有以下结构:



[0021] 本文公开的一些实施方案涉及通过连接基共价连接到荧光团的核苷或核苷酸,其中所述连接基包含式(III)的结构:



[0023] 其中L¹不存在或包括式(I)或(II)中所述的任一连接基、或保护基,或其组合;L²选自任选取代的C₁₋₂₀亚烷基、任选取代的C₁₋₂₀杂亚烷基、被取代的芳基中断的任选取代的C₁₋₂₀亚烷基、或被取代的芳基中断的任选取代的C₁₋₂₀杂亚烷基;L³选自任选取代的C₁₋₂₀亚烷基、或任选取代的C₁₋₂₀杂亚烷基;R^A选自氢、氰基、羟基、卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基、C₁₋₆卤代烷氧基或叠氮基,且其中



选自氧(O)或NR^B;每个R^B和R^C独立地选自氢或任选取代的C₁₋₆烷基;且k为1至50的整数。

[0024] 本文公开的一些实施方案涉及包含标记的核苷或核苷酸的试剂盒,所述核苷或核苷酸含有在荧光团与核苷或核苷酸之间的连接基,其中所述连接基包含式(I)、(II)或(III)中任一结构,或其组合。

[0025] 本文公开的一些实施方案涉及用于修饰核苷或核苷酸的试剂,其包含荧光团和连接基,其中所述连接基包含式(I)、(II)或(III)中任一结构,或其组合。

[0026] 本文公开的一些实施方案涉及用于检测已经掺入到多核苷酸中的核苷的方法,所述方法包括:(a)将包含连接基的标记的核苷或核苷酸掺入到多核苷酸中;和(b)检测步骤(a)中掺入的所述标记的核苷或核苷酸的荧光信号,其中所述连接基包含式(I)、(II)或(III)中任一结构,或其组合。在一些实施方案中,所述方法还包括:提供模板核酸链和部分

杂交的核酸链,其中步骤(a)将至少一个核苷或核苷酸掺入杂交链,所述核苷或核苷酸与模板链的相应位置的核苷或核苷酸互补,且其中步骤(b)鉴定了掺入的核苷或核苷酸的碱基,从而指示模板链的互补核苷或核苷酸的身份。

[0027] 本文公开的一些实施方案涉及对模板核酸分子测序的方法,所述方法包括:将一个或多个标记的核苷酸掺入到与模板核酸互补的核酸链中;确定存在于一个或多个掺入的标记核苷酸中的碱基的身份,以确定模板核酸分子的序列;其中通过检测由所述标记的核苷酸产生的荧光信号来确定存在于所述一个或多个标记的核苷酸中的碱基的身份;并且其中至少一个掺入的标记核苷酸包含如上所述的连接基,其中所述连接基包含式(I)、(II)或(III)中任一结构,或其组合。在一些实施方案中,在每个核苷酸掺入步骤后确定存在于一个或多个核苷酸中的碱基的身份。

附图说明

[0028] 图1A示出了标准的标记核苷酸的部分连接基团结构。

[0029] 图1B示出了图1A的标记的核苷酸,其具有可插入到图1A的标准连接基团中的两种可能的连接基125和130。

[0030] 图2显示使用图1A的标记核苷酸和图1B的经修饰的标记核苷酸的核苷酸掺入速率的图。

[0031] 图3A-3E示出了可插入到图1A的标准连接基团中的另外的连接基的结构式。

[0032] 图4显示用于评价图1B的125插入物和图3B的315插入物对测序质量影响的双染料测序运行的数据表。

[0033] 图5A和5B示出了使用连接基插入物125和315时,图4的测序运行的读数1的错误率图和读数2的错误率图。

[0034] 图6显示用于评估图1B的125插入物和图3A的310插入物对测序质量影响的测序运行的数据表。

[0035] 图7A和7B示出了使用连接基插入物125时,图6的测序运行的读数1的错误率图和读数2的错误率图。

[0036] 图8A显示了标准的LN₃连接基结构的一个实例。

[0037] 图8B、8C和8D显示了图8A的LN₃连接基的修饰结构的三个实例。

[0038] 图8E示出了将保护基插入到图8D的连接基中。

[0039] 图9A是显示在具有SS-连接基的ffa中出现杂质的色谱图。图9B是对比具有SS-连接基和AEDI-连接基的ffa的稳定性的表格。图9C是显示在IMX 60°中22小时具有SS-连接基和AEDI-连接基的ffa的对比色谱图,其再次显示SS连接基出现杂质。

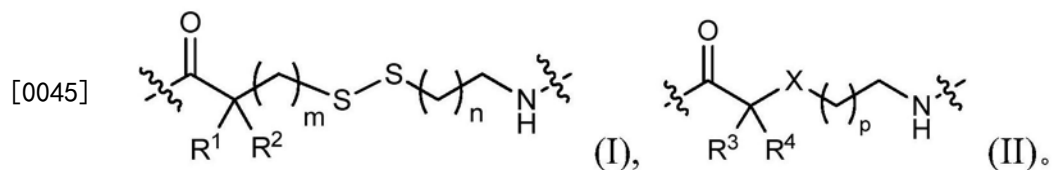
[0040] 图10A、10B和10C显示了在连接基改变的溶液中核苷酸掺入速度的意外增加。图10A是显示在1μM的掺入速率图。图10B显示列表的结果。图10C图示了具有NR550S0的AEDI和SS连接基。

[0041] 图11A显示了具有不同A-550S0(相同浓度)的V10组合的散点图。图11B显示了溶液中的Kcat FFA连接基。

[0042] 图12A和12B显示对于M111、Human550共2x151个循环的测序指标。

[0043] 优选实施方案的详细描述

[0044] 本文公开的一些实施方案涉及通过连接基共价连接到荧光团的核苷或核苷酸,其中所述连接基包含下式 (I) 或 (II), 或其组合的结构, 其中变量的定义如上定义。



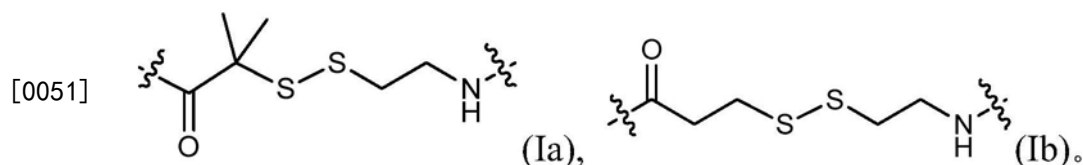
[0046] 在一些式 (I) 结构的实施方案中, R^1 为氢。在一些其他实施方案中, R^1 为任选取代的 C_{1-6} 烷基。在一些这样的实施方案中, R^1 为甲基。

[0047] 在本文所述的式 (I) 的 R^1 的任何实施方案中, R^2 是氢。在一些其他实施方案中, R^2 为任选取代的 C_{1-6} 烷基。在一些这样的实施方案中, R^2 是甲基。在一个实施方案中, R^1 和 R^2 都是甲基。在另一个实施方案中, R^1 和 R^2 都是氢。

[0048] 在一些式 (I) 结构的实施方案中, m 为 0。在一些其他实施方案中, m 为 1。

[0049] 在一些式 (I) 结构的实施方案中, n 为 1。

[0050] 在一些式 (I) 结构的实施方案中, 式 (I) 的结构也可以由式 (Ia) 或 (Ib) 表示:



[0052] 在本文所述的一些实施方案中, 式 (Ia) 被称为“AEDI”, 式 (Ib) 被称为“SS”。

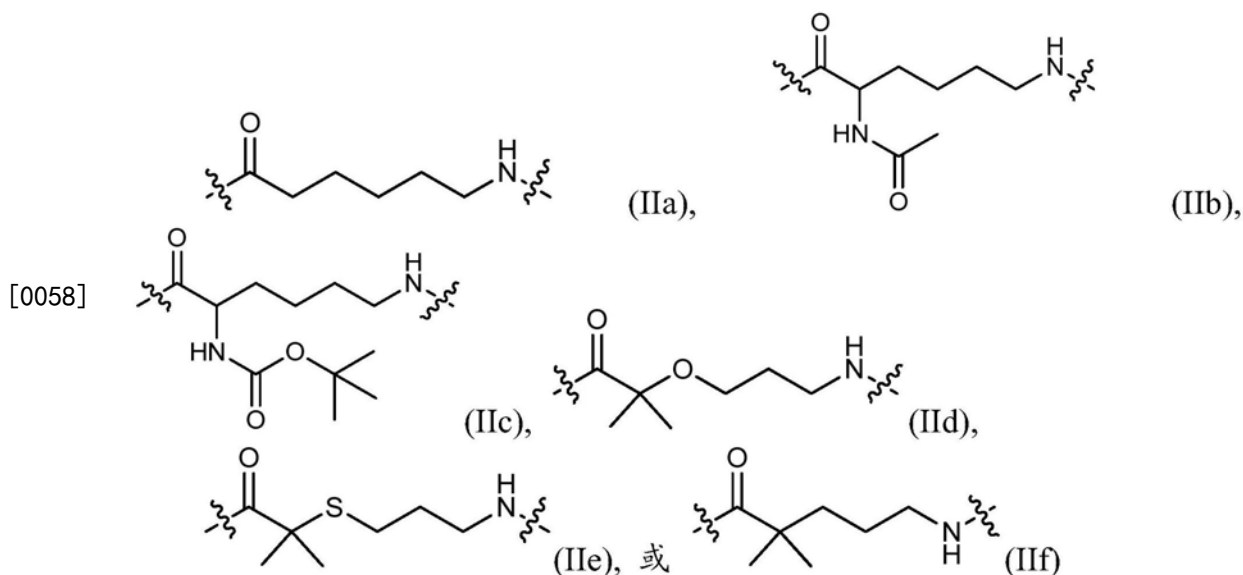
[0053] 在式 (II) 结构的一些实施方案中, R^3 是氢。在一些其他实施方案中, R^3 是任选取代的 C_{1-6} 烷基。在一些这样的实施方案中, R^3 是甲基。在一些实施方案中, R^3 是 $-NR^5-C(=O)R^6$ 。在一些这样的实施方案中, R^5 是氢。在一些这样的实施方案中, R^6 是任选取代的 C_{1-6} 烷基, 例如甲基。在一些实施方案中, R^3 是 $-NR^7-C(=O)OR^8$ 。在一些这样的实施方案中, R^7 是氢。在一些这样的实施方案中, R^8 是任选取代的 C_{1-6} 烷基, 例如叔丁基。

[0054] 在本文所述的式 (II) 的 R^3 的任何实施方案中, R^4 是氢。在一些其他实施方案中, R^4 是任选取代的 C_{1-6} 烷基。在一些这样的实施方案中, R^4 是甲基。在一个实施方案中, R^3 和 R^4 都是甲基。在另一个实施方案中, R^3 和 R^4 都是氢。在一个实施方案中, R^3 是 $-NH(C=O)CH_3$, 且 R^4 是氢。在另一个实施方案中, R^3 是 $-NH(C=O)O^tBu$ (Boc), R^4 是氢。

[0055] 在式 (II) 结构的一些实施方案中, X 是亚甲基, 其可以是任选取代的。在另一个实施方案中, X 是氧 (O)。在另一个实施方案中, X 是硫 (S)。

[0056] 在式 (II) 结构的一些实施方案中, p 为 1。在一些其他实施方案中, p 为 2。

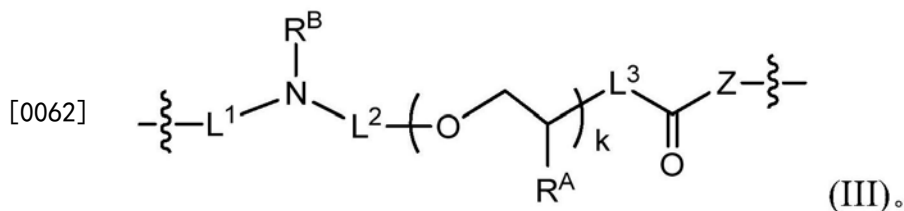
[0057] 在式 (II) 结构的一些实施方案中, 式 (II) 结构也可由式 (IIa)、(IIb)、(IIc)、(IId)、(IIe) 或 (IIf) 所代表:



[0059] 在本文所述的一些实施方案中,式(IIa)被称为“ACA”,式(IIb)被称为“AcLys”,式(IIc)被称为“BocLys”,式(IIId)被称为“dMeO”,式(IIe)被称为“dMeS”,式(IIIf)被称为“DMP”。

[0060] 在通过包含本文所述的式(I)或(II)结构的连接基的荧光团标记的核苷或核苷酸的任何实施方案中,所述核苷或核苷酸可以直接或通过另外的连接部分连接到连接基的左侧。

[0061] 本文公开的一些实施方案涉及通过连接基共价连接到荧光团的核苷或核苷酸,所述连接基包含式(III)的结构,并且其中变量的定义如上所定义,



[0063] 在式(III)结构的一些实施方案中, L^1 不存在。在一些其他实施方案中, L^1 是包含式(I)或(II),特别是式(Ia)、(Ib)、(II)、(IIa)、(IIb)、(IIc)、(IIId)、(IIe)或(IIIf)结构的上述连接基。在一些其他实施方案中, L^1 可以是包含防止DNA损伤的分子的保护基。在一些这样的实施方案中,保护基包括Trolox、没食子酸、对-硝基苄基(pNB)或抗坏血酸盐,或其组合。

[0064] 在式(III)结构的一些实施方案中, L^2 是任选取代的 C_{1-20} 亚烷基。在一些另外的实施方案中, L^2 是任选取代的 C_{4-10} 亚烷基。在一些这样的实施方案中, L^2 是亚庚基。在一些其他实施方案中, L^2 是任选取代的 C_{1-20} 杂亚烷基。在一些这样的实施方案中,所述任选取代的 C_{1-20} 杂亚烷基包含一个或多个氮原子。在一些这样的实施方案中, C_{1-20} 杂亚烷基中的至少一个碳原子被氧代基(=O)取代。在一些其他实施方案中, L^2 是任选取代的 C_{3-6} 杂亚烷基。在一些实施方案中, L^2 被取代的芳基(例如取代的 C_{6-10} 芳基)或包含一至三个杂原子的5至10元取代的杂芳基中断。在一些这样的实施方案中, L^2 被取代的苯基中断。在一些这样的实施方案中,苯基被一个或多个(至多4个)选自硝基、氰基、卤素、羟基、 C_{1-6} 烷基、 C_{1-6} 烷氧基、 C_{1-6} 卤代烷基、 C_{1-6} 卤代烷氧基或磺酸基的取代基所取代。在一些另外的实施方案中,苯基被一至四

个选自硝基、氰基、卤素或磺酸基(即, $-S(=O)_2OH$)的取代基所取代。

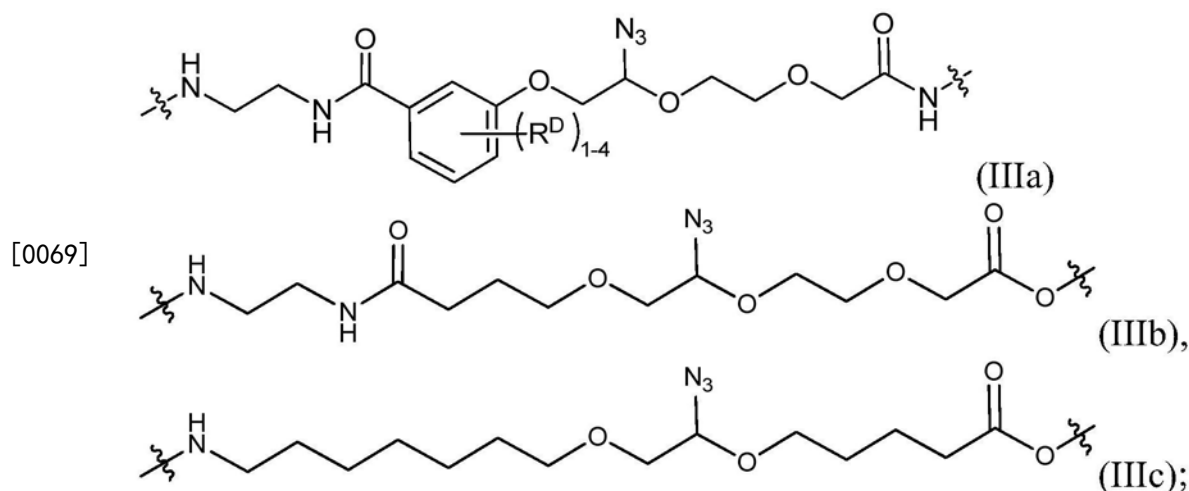
[0065] 在式(III)结构的一些实施方案中, $-\left(\begin{array}{c} \text{O} \\ | \\ \text{---} \end{array} \right)_k$ 中的 R^A 选自氢或叠氮基。在

一些这样的实施方案中, k 为 2, 其中一个 R^A 是叠氮基, 且另一个 R^A 是氢。

[0066] 在式(III)结构的一些实施方案中, L^3 是任选取代的 C_{1-20} 亚烷基。在一些另外的实施方案中, L^3 是任选取代的 C_{1-6} 亚烷基。在一些这样的实施方案中, L^3 是亚乙基。在一些其他实施方案中, L^3 是任选取代的 C_{1-20} 杂亚烷基。在一些这样的实施方案中, 任选取代的 C_{1-20} 杂亚烷基包含一个或多个氧原子。在一些这样的实施方案中, L^1 是任选取代的 C_{1-6} 氧杂亚烷基(alkylene oxide), 例如 C_{1-3} 氧杂亚烷基。

[0067] 在式(III)结构的一些实施方案中, R^B 是氢。在一些实施方案中, R^C 是氢。在一些另外的实施方案中, R^B 和 R^C 都是氢。

[0068] 在式(III)结构的一些实施方案中, 式(III)结构也可由式(IIIa)、(IIIb)或(IIIc)所代表:



[0070] 其中 R^D 选自硝基、氰基、卤素、羟基、 C_{1-6} 烷基、 C_{1-6} 烷氧基、 C_{1-6} 卤代烷基、 C_{1-6} 卤代烷氧基或磺酸基。在一些其它实施方案中, R^D 选自硝基、氰基、卤素或磺酸基。

[0071] 在通过包含本文所述的式(III)结构的连接基的荧光团标记的核苷或核苷酸的任何实施方案中, 所述荧光团可以直接或通过另外的连接部分连接到连接基的左侧。

[0072] 在文中所述的关于包含式(I)、(II)或(III)结构的连接基的任何实施方案中, 当术语“任选取代的”用于定义变量时, 该变量可以是未取代的。

[0073] 定义

[0074] 除非另有定义, 否则本文使用的所有技术和科学术语具有与本领域技术人员通常理解的相同的含义。使用术语“包括(including)”以及其他形式, 例如“包括(include)”、“包括(includes)”和“包括(included)”不是限制性的。使用术语“具有(having)”以及其他形式, 例如“具有(have)”、“具有(has)”和“具有(had)”不是限制性的。如本说明书中所使用的, 无论在权利要求的连接部分还是在主体部分中, 术语“包括(comprise(s))”和“包括(comprising)”都应被解释为具有开放式含义。也就是说, 上述术语应被解释为与短语“至少具有”或“至少包括”同义。例如, 当在方法的情形中使用时, 术语“包括”是指该方法包括

至少所述的步骤,但还可以包括另外的步骤。当在化合物、组合物或装置的情形中使用,术语“包含”是指化合物、组合物或装置至少包括所述的特征或组分,但也可包括另外的特征或组分。

[0075] 本文使用的章节标题仅用于组织编排的目的,并且不应被解释为限制所描述的主题。

[0076] 如本文所用的,常见的有机化学缩写定义如下:

[0077]	Ac	乙酰基
[0078]	Ac ₂ O	乙酸酐
[0079]	aq.	水溶液
[0080]	Bn	苄基
[0081]	Bz	苯甲酰基
[0082]	BOC或Boc	叔丁氧基羰基
[0083]	Bu	正丁基
[0084]	cat.	催化的
[0085]	°C	摄氏度
[0086]	CHAPS	3-[(3-胆酰氨基丙基)二甲基氨基]-1-丙磺酸内盐
[0087]	dATP	脱氧腺苷三磷酸
[0088]	dCTP	脱氧胞苷三磷酸
[0089]	dGTP	脱氧鸟苷三磷酸
[0090]	dTTP	脱氧胸苷三磷酸
[0091]	ddNTP(s)	双脱氧核苷酸
[0092]	DCM	二氯甲烷
[0093]	DMA	二甲基乙酰胺
[0094]	DMAP	4-二甲基氨基吡啶
[0095]	DMF	N,N'-二甲基甲酰胺
[0096]	DMSO	二甲基亚砷
[0097]	DSC	N,N'-二琥珀酰亚胺基碳酸酯
[0098]	EDTA	乙二胺四乙酸
[0099]	Et	乙基
[0100]	EtOAc	乙酸乙酯
[0101]	ffN	全功能化核苷酸
[0102]	ffA	全功能化腺苷核苷酸
[0103]	g	克
[0104]	GPC	凝胶渗透色谱法
[0105]	h或hr	小时
[0106]	Hunig's碱	N,N-二异丙基乙胺
[0107]	iPr	异丙基
[0108]	KPi	在pH 7.0下的10mM磷酸钾缓冲液
[0109]	IPA	异丙醇

[0110]	LCMS	液相色谱-质谱
[0111]	LDA	二异丙基氨基锂
[0112]	m或min	分钟
[0113]	MeCN	乙腈
[0114]	mL	毫升
[0115]	PEG	聚乙二醇
[0116]	PG	保护基
[0117]	Ph	苯基
[0118]	pNB	对-硝基-苄基
[0119]	ppt	沉淀
[0120]	rt	室温
[0121]	SBS	合成测序
[0122]	-S(O) ₂ OH	磺酸基
[0123]	TEA	三乙胺
[0124]	TEAB	四乙基溴化铵
[0125]	TFA	三氟乙酸
[0126]	Tert, t	叔
[0127]	THF	四氢呋喃
[0128]	TLC	薄层色谱法
[0129]	TSTU	O-(N-琥珀酰亚胺基)-N,N,N',N'-四甲基脒鎓四氟硼酸盐
[0130]	μL	微升

[0131] 如本文所用,术语“阵列”是指连接到一个或多个底物的不同探针分子的群体,使得不同的探针分子可根据相对位置彼此区分。阵列可以包括不同的探针分子,其各自位于底物上不同的可寻址位置上。或者或另外,阵列可包括各自承载不同探针分子的分开的底物,其中不同的探针分子可根据在底物所附着的表面上的底物位置或根据液体中的底物的位置来识别。其中分离的底物位于表面上的示例性阵列包括但不限于那些包括例如在美国专利号6,355,431B1、US 2002/0102578和PCT公开号W0 00/63437中描述的孔中的珠。例如使用微流体装置(例如荧光活化细胞分选仪(FACS))来区分液体阵列中的珠的本发明中可使用的示例性形式描述于例如美国专利号6,524,793中。可用于本发明的阵列的其它实例包括但不限于在下文中描述的那些:美国专利号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W0 93/17126;W0 95/11995;W0 95/35505;EP 742 287;和EP 799 897。

[0132] 如本文所用,术语“共价连接”或“共价键合”是指形成化学键,其特征在于原子之间共享电子对。例如,共价连接的聚合物涂层是指与通过其他方式(例如粘附或静电相互作用)附着到表面相比,与底物的官能化表面形成化学键的聚合物涂层。应当理解,共价连接到表面的聚合物也可以通过除共价连接之外的方式结合。

[0133] 如本文所用,其中“a”和“b”为整数的“C_a至C_b”或“C_{a-b}”是指特定基团中的碳原子数。也就是说,该基团可以包含“a”至“b”(包括端值)个碳原子。因此,例如“C₁至C₄烷基”或

“C₁₋₄烷基”是指具有1至4个碳的所有烷基,即CH₃-、CH₃CH₂-、CH₃CH₂CH₂-、(CH₃)₂CH-、CH₃CH₂CH₂CH₂-、CH₃CH₂CH(CH₃)-和(CH₃)₃C-。

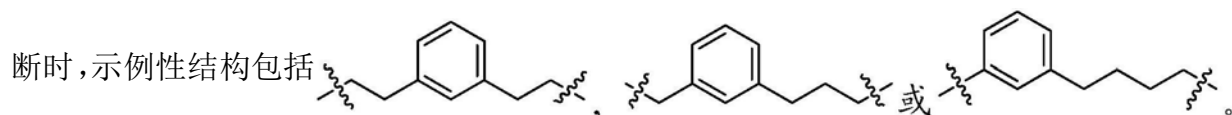
[0134] 本文所用的术语“卤素”或“卤代”是指元素周期表第7列的放射性-稳定的原子中的任何一个,例如氟、氯、溴或碘,其中优选氟和氯。

[0135] 本文所用的“烷基”是指完全饱和(即不含双键或叁键)的直链或支链烃链。烷基可以具有1至20个碳原子(每当在文中出现时,数值范围如“1至20”是指给定范围内的每个整数;例如“1至20个碳原子”是指烷基可以由1个碳原子、2个碳原子、3个碳原子等,直到20个碳原子组成,但是本定义还涵盖其中没有指定数值范围的术语“烷基”)。烷基还可以是具有1至9个碳原子的中等大小的烷基。烷基还可以是具有1至4个碳原子的低级烷基。烷基可以指定为“C₁₋₄烷基”或类似的名称。仅作为实例,“C₁₋₄烷基”表示在烷基链中具有一至四个碳原子,即所述烷基链选自甲基、乙基、丙基、异丙基、正丁基、异丁基、仲丁基和叔丁基。典型的烷基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。烷基可以是取代的或未取代的。

[0136] 本文所用的“烷氧基”是指式-OR,其中R是如上定义的烷基,例如“C₁₋₉烷氧基”,其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。

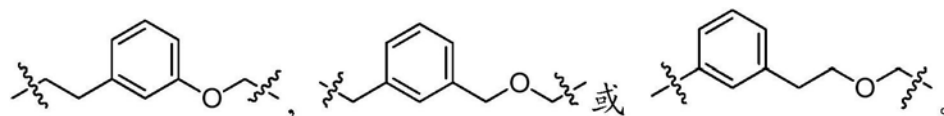
[0137] 本文所用的“杂烷基”是指在主链中含有一个或多个杂原子的直链或支链烃链,所述杂原子即除碳以外的元素,包括但不限于氮、氧和硫。杂烷基可以具有1至20个碳原子,但是本定义还涵盖没有指定数值范围的术语“杂烷基”的情形。杂烷基还可以是具有1至9个碳原子的中等大小的杂烷基。杂烷基还可以是具有1至4个碳原子的低级杂烷基。杂烷基可以称为“C₁₋₄杂烷基”或类似的名称。杂烷基可以含有一个或多个杂原子。仅作为实例,“C₁₋₄杂烷基”表示在杂烷基链中有一至四个碳原子,并且在主链中另外还含有一个或多个杂原子。

[0138] 本文所用的“亚烷基”是指仅含有碳和氢的支链或直链完全饱和的二价化学基团,其通过两个连接点(即烷二基)连接到分子的其余部分。亚烷基可以具有1至20个碳原子,但是本定义还涵盖其中没有指定数值范围的术语亚烷基的情形。亚烷基还可以是具有1至9个碳原子的中等大小的亚烷基。亚烷基还可以是具有1至4个碳原子的低级亚烷基。亚烷基可以被称为“C₁₋₄亚烷基”或类似的名称。仅作为实例,“C₁₋₄亚烷基”表示在亚烷基链中有一至四个碳原子,即亚烷基链选自亚甲基、亚乙基、乙-1,1-二基、亚丙基、丙-1,1-二基、丙-2,2-二基、1-甲基-亚乙基、亚丁基、丁-1,1-二基、丁-2,2-二基、2-甲基-丙-1,1-二基、1-甲基-亚丙基、2-甲基-亚丙基、1,1-二甲基-亚乙基、1,2-二甲基-亚乙基和1-乙基-亚乙基。如本文所用,当亚烷基被芳基中断时,其是指芳基通过两个连接点插入到亚烷基链的一个碳-碳键之间,或者芳基通过一个连接点连接到亚烷基链的一个末端。例如,当正亚丁基被苯基中断时,示例性结构包括



[0139] 如本文所用,术语“杂亚烷基”是指亚烷基链,其中亚烷基的一个或多个骨架原子选自非碳原子,例如氧、氮、硫、磷或其组合。杂亚烷基链可以具有2至20,000的长度。示例性的杂亚烷基包括但不限于-OCH₂-、-OCH(CH₃)-、-OC(CH₃)₂-、-OCH₂CH₂-、-CH(CH₃)O-、-CH₂OCH₂-、-CH₂OCH₂CH₂-、-SCH₂-、-SCH(CH₃)-、-SC(CH₃)₂-、-SCH₂CH₂-、-CH₂SCH₂CH₂-、-

NHCH₂-、-NHCH(CH₃)-、-NHC(CH₃)₂-、-NHCH₂CH₂-、-CH₂NHCH₂-、-CH₂NHCH₂CH₂-等。如本文所使用的,当杂亚烷基被芳基中断时,其是指芳基通过两个连接点插入到杂亚烷基链的一个碳-碳键或碳-杂原子键之间,或者芳基通过一个连接点连接到杂亚烷基链的一个末端。例如,当氧杂正亚丙基被苯基中断时,示例性结构包括



[0140] 本文所用的“烯基”是指在直链或支链烃链中含有一个或多个双键的烷基。烯基可以是未取代的或取代的。

[0141] 本文所用的“炔基”是指在直链或支链烃链中含有一个或多个叁键的烷基。炔基可以是未取代的或取代的。

[0142] 本文所用的“环烷基”是指完全饱和的(无双键或叁键)单环或多环烃环体系。当由两个或更多个环组成时,这些环可以以稠合的方式连接在一起。环烷基在环中可以含有3至10个原子。在一些实施方案中,环烷基可以在环中含有3至8个原子。环烷基可以是未取代的或取代的。典型的环烷基包括但绝不限于环丙基、环丁基、环戊基、环己基、环庚基和环辛基。

[0143] 术语“芳香的”是指具有共轭 π 电子系统的环或环体系,并且包括碳环芳香基团(例如苯基)和杂环芳香基团(例如吡啶)。该术语包括单环或稠环多环(即,这些环共享相邻的原子对),条件是整个环系是芳香的。

[0144] 本文所用的“芳基”是指在环骨架中仅含有碳的芳环或环系(即,共享两个相邻碳原子的两个或多个稠环)。当芳基是环体系时,体系中的每个环是芳香的。芳基可以具有6至18个碳原子,但是本定义还涵盖没有指定数值范围的术语“芳基”的情形。在一些实施方案中,芳基具有6至10个碳原子。芳基可以称为“C₆₋₁₀芳基”,“C₆或C₁₀芳基”或类似的名称。芳基的实例包括但不限于苯基、萘基、蒽基和蒽基。

[0145] “芳烷基”或“芳基烷基”是通过亚烷基连接作为取代基的芳基,例如“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,亚烷基是低级亚烷基(即,C₁₋₄亚烷基)。

[0146] 本文所用的“杂芳基”是指在环骨架中含有一个或多个杂原子(即除碳以外的元素,包括但不限于氮、氧和硫)的芳环或环系(即,共享两个相邻原子的两个或多个稠环)。当杂芳基是环系时,体系中的每个环是芳香的。杂芳基可以具有5-18个环成员(即,构成环骨架的原子数,包括碳原子和杂原子),但是本定义还涵盖其中没有指定数值范围的术语“杂芳基”的情形。在一些实施方案中,杂芳基具有5至10个环成员或5至7个环成员。杂芳基可以指定为“5-7元杂芳基”,“5-10元杂芳基”或类似的名称。杂芳基环的实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、吡啶基、异吡啶基和苯并噻吩基。

[0147] “杂芳烷基”或“杂芳基烷基”是通过亚烷基连接作为取代基的杂芳基。实例包括但不限于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,亚烷基是低级亚烷基(即, C₁₋₄亚烷基)。

[0148] 本文所用的“环烷基”是指完全饱和的碳环或环体系。实例包括环丙基、环丁基、环戊基和环己基。

[0149] “O-羧基”基团是指“ $-OC(=O)R$ ”基团，其中R如文中所定义选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。

[0150] “C-羧基”是指“ $-C(=O)OR$ ”基团，其中R如本文所定义选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。非限制性实例包括羧基(即， $-C(=O)OH$)。

[0151] “氰基”是指“ $-CN$ ”基团。

[0152] “叠氮基”是指“ $-N_3$ ”基团。

[0153] “O-氨基甲酰基”是指“ $-OC(=O)NR_AR_B$ ”基团，其中R_A和R_B如本文所定义各自独立地选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。

[0154] “N-氨基甲酰基”是指“ $-N(R_A)OC(=O)R_B$ ”基团，其中R_A和R_B如本文所定义各自独立地选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。

[0155] “C-酰氨基”是指“ $-C(=O)NR_AR_B$ ”基团，其中R_A和R_B如本文所定义各自独立地选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。

[0156] “N-酰氨基”是指“ $-N(R_A)C(=O)R_B$ ”基团，其中R_A和R_B如本文所定义各自独立地选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。

[0157] “氨基”是指“ $-NR_AR_B$ ”基团，其中R_A和R_B如本文所定义各自独立地选自氢、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基、C₃₋₇碳环基、C₆₋₁₀芳基、5-10元杂芳基和5-10元杂环基。非限制性实例包括游离氨基(即 $-NH_2$)。

[0158] 如本文所用，术语“Trolox”是指6-羟基-2,5,7,8-四甲基苯并二氢吡喃-2-甲酸。

[0159] 如本文所用，术语“抗坏血酸盐”是指抗坏血酸的盐。

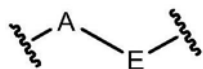
[0160] 如本文所用，术语“没食子酸”是指3,4,5-三羟基苯甲酸。

[0161] 如本文所用，取代的基团衍生自未取代的母体基团，其中已经有一个或多个氢原子与另一个原子或基团交换。除非另有说明，当基团被认为是“取代的”时，是指该基团被一个或多个取代基取代，所述取代基独立地选自：C₁₋₆烷基、C₁₋₆烯基、C₁₋₆炔基、C₁₋₆杂烷基、C₃₋₇碳环基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、C₃₋₇碳环基-C₁₋₆-烷基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、5-10元杂环基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、5-10元杂环基-C₁₋₆-烷基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、芳基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、芳基(C₁₋₆)烷基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、5-10元杂芳基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、5-10元杂芳基(C₁₋₆)烷基(其任选被卤素、C₁₋₆烷基、C₁₋₆烷氧基、C₁₋₆卤代烷基和C₁₋₆卤代烷氧基取代)、卤素、氰基、羟基、C₁₋₆烷氧基、C₁₋₆烷氧基(C₁₋₆)烷基(即醚)、芳氧基、巯基、卤代(C₁₋₆)烷基(例如 $-CF_3$)、卤代(C₁₋₆)烷氧基(例如 $-OCF_3$)、C₁₋₆烷硫基、芳硫基、氨基、氨基(C₁₋₆)烷基、硝基、O-氨基甲酰基、N-氨基甲酰基、O-硫代氨基甲酰基、N-硫代氨基甲酰基、C-酰氨基、N-酰氨基。

基、S-磺酰氨基、N-磺酰氨基、C-羧基、O-羧基、酰基、氰酰基(cyanato)、异氰酰基、硫氰酸酯基(thiocyanato)、异硫氰酸酯基(isothiocyanato)、亚磺酰基、磺酰基和氧代基(=O)。无论在何处将基团描述为“任选取代的”，该基团可以被上述取代基取代。

[0162] 应当理解，取决于上下文，某些基团命名规则可以包括一价基(mono-radical)或二价基(di-radical)。例如，当取代基需要与分子其余部分的两个点连接时，应理解该取代基是二价基。例如，需要两个连接点的确定为烷基的取代基包括二价基，例如-CH₂-、-CH₂CH₂-、-CH₂CH(CH₃)CH₂-等。类似地，需要两个连接点的确定为氨基的取代基包括二价基，例如-NH-、-N(CH₃)-等。其它基团命名规则清楚地表明该基团是二价基，例如“亚烷基”或“亚烯基”。

[0163] 当取代基被描述为二价基(即，具有两个与分子其余部分的连接点)时，应当理解，除非另有说明，该取代基可以以任何方向的构型连接。因此，例如描述为-AE-或



的取代基包括其取向使得A连接于分子的最左侧连接点的取代基以及A连接于分子的最右侧连接点的情形。

[0164] 当本文公开的化合物具有至少一个立体中心时，它们可以作为单独的对映异构体和非对映异构体或作为这些异构体的混合物(包括外消旋体)存在。单个异构体的分离或单个异构体的选择性合成通过应用本领域技术人员公知的各种方法来实现。除非另有说明，否则所有这些异构体及其混合物都包括在本文公开的化合物的范围内。

[0165] 如本文所用，“核苷酸”包括含氮杂环碱基、糖和一个或多个磷酸基。它们是核酸序列的单体单元。在RNA中，糖是核糖，在DNA中，糖是脱氧核糖，即缺少存在于核糖中的羟基的糖。含氮杂环碱基可以是嘌呤碱基或嘧啶碱基。嘌呤碱基包括腺嘌呤(A)和鸟嘌呤(G)，及其修饰的衍生物或类似物。嘧啶碱基包括胞嘧啶(C)、胸腺嘧啶(T)和尿嘧啶(U)，及其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。

[0166] 如本文所用，“核苷”在结构上与核苷酸相似，但缺少磷酸部分。核苷类似物的实例是其中标记连接到碱基上并且没有磷酸基团连接到糖分子上的类似物。术语“核苷”在本文中以本领域技术人员所理解的普通含义使用。实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。改性的戊糖部分是其中氧原子已经被碳替代和/或碳已经被硫或氧原子替代的戊糖部分。“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。

[0167] 如本文所用，术语“多核苷酸”是指通常包括DNA(例如基因组DNA cDNA)、RNA(例如mRNA)、合成的寡核苷酸和合成的核酸类似物的核酸。多聚核苷酸可以包括天然或非天然碱基，或其组合，以及天然或非天然主链键例如，硫代磷酸酯、PNA或2'-O-甲基-RNA，或其组合。

[0168] 如本文所用，术语“定相(phasing)”是指在SBS中在给定的测序循环，由3'终止子和荧光团的不完全去除，和聚合酶不能完成将一部分DNA链掺入到聚簇内而引起的现象。预定相(Pre-phasing)由掺入核苷酸而没有有效的3'终止子引起，并且掺入事件提前1个循环。定相和预定相使得特定循环的提取强度包括当前循环的信号以及来自前面和后面循环的噪声。随着循环数目的增加，每个簇受到定相影响的序列片段增加，阻碍了正确碱基的鉴定。预定相可以由合成测序(SBS)期间存在痕量的未保护或未封闭的3'-OH核苷酸引起。未

保护的3'-OH核苷酸可能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,修饰核苷酸类似物或连接基,导致更快的SBS循环时间、更低的定相和预定相值以及更长的测序读序长度,在SBS应用中提供了更大的优势。

[0169] 如本文所用,术语“保护基团”包括但不限于可防止DNA损伤(例如光损伤或其它化学损伤)的分子。一些具体实例包括抗氧化剂,例如维生素C、维生素E衍生物、酚酸、多酚,及其衍生物和类似物。应当理解,在定义术语“保护基”的某些上下文中,其是指由保护基的一个或多个官能团与本文所述的连接基的相应官能团之间反应所生成的基团。例如,当保护基是“没食子酸”时,其可以指的是没食子酸的酰胺和酯,而不是具有游离羧基的没食子酸本身。

[0170] 可检测的标记

[0171] 本文所述的一些实施方案涉及常规可检测标记的使用。检测可以通过任何合适的方法包括荧光光谱法或通过其它光学方法进行。优选的标记是荧光团,其在能量吸收之后发射确定波长的辐射。许多合适的荧光标记是已知的。例如,Welch等人(Chem.Eur.J.5(3):951-960,1999)公开了可以用于本发明的丹磺酰基官能化的荧光基团。Zhu等人(Cytometry 28:206-211,1997)描述了也可以在本发明中使用的荧光标记Cy3和Cy5的使用。适合使用的标记也公开于Prober等人(Science 238:336-341,1987);Connell等人(BioTechniques 5(4):342-384,1987),Ansorge等人(Nucl.Acids Res.15(11):4593-4602,1987)和Smith等人(Nature 321:674,1986)中。其它市售的荧光标记包括但不限于荧光素、罗丹明(包括TMR、德克萨斯红和Rox)、alexa、氟硼荧(bodipy)、吖啶、香豆素、苾、苯并蒽和花青素。

[0172] 多标记也可以用于本申请中,例如双荧光团FRET盒(Tet.Let.46:8867-8871,2000)。也可使用多氟树枝状体系(J.Am.Chem.Soc.123:8101-8108,2001)。尽管优选荧光标记,但是显然其他形式的可检测标记对于本领域技术人员也是有用的。例如,微粒,包括量子点(Empodocles等人,Nature 399:126-130,1999)、金纳米颗粒(Reichert等人,Anal.Chem.72:6025-6029,2000)和微球(Lacoste等人,Proc.Natl.Acad.Sci USA 97(17):9461-9466,2000)都可被使用。

[0173] 多组分标记也可以用于本申请中。多组分标记是取决于与用于检测的另一化合物的相互作用的标记。生物学中使用的最常见的多组分标记是生物素-链亲和素系统。生物素用作连接于核苷酸的碱基的标记。然后分别加入链亲和素以使检测发生。可采用其他多组分系统。例如,二硝基酚具有可用于检测的市售荧光抗体。

[0174] 除非另有说明,否则对核苷酸的参考也意图适用于核苷。本申请还将参考DNA进一步描述,除非另有说明,否则本描述也适用于RNA、PNA和其他核酸。

[0175] 测序方法

[0176] 本文所述的核苷或核苷酸可以与多种测序技术联合使用。在一些实施方案中,用于确定靶核酸的核苷酸序列的方法可以是自动化方法。

[0177] 本文提供的核苷酸类似物可用于测序过程,例如合成测序(SBS)技术。简言之,可以通过使靶核酸与一种或多种标记的核苷酸、DNA聚合酶等接触来启动SBS。将使用靶核酸作为模板延伸引物的那些特征掺入可被检测的标记的核苷酸。任选地,标记的核苷酸还可以包括一旦将核苷酸添加到引物中就会终止进一步引物延伸的可逆终止的特性。例如,可以将具有可逆终止子部分的核苷酸类似物加入到引物中,使得随后的延伸不会发生,直到

递送去封闭剂以除去该部分。因此,对于使用可逆终止的实施方案,可以将去封闭试剂递送到流动池(flow cell)(在检测发生之前或之后)。可以在各个递送步骤之间进行洗涤。然后将该循环重复n次以将引物延伸n个核苷酸,从而检测长度n的序列。可以容易地适合与本发明公开的方法所产生的阵列一起使用的示例性SBS方法、流体系统和检测平台描述于例如,Bentley等人,Nature 456:53-59(2008),WO 04/018497;WO 91/06678;WO 07/123744;美国专利号7,057,026;7,329,492;7,211,414;7,315,019或7,405,281,以及美国专利申请公开号2008/0108082A1中,其每个引入本文作为参考。

[0178] 可以使用利用循环反应的其他测序过程,例如焦磷酸测序。焦磷酸测序是当特定核苷酸被掺入到新生核酸链中时检测无机焦磷酸(PPi)的释放(Ronaghi,等人,Analytical Biochemistry 242(1),84-9(1996);Ronaghi,Genome Res.11(1),3-11(2001);Ronaghi等人Science 281(5375),363(1998);美国专利号6,210,891;6,258,568和6,274,320,其每个并入本文作为参考)。在焦磷酸测序中,释放的PPi可以通过由ATP硫酸化酶转化为三磷酸腺苷(ATP)来检测,并且所得的ATP可以通过荧光素酶产生的光子检测。因此,可以通过发光检测系统监测测序反应。基于荧光的检测系统所用的激发辐射源对于焦磷酸测序过程不是必需的。可用于将焦磷酸测序应用于本发明公开的阵列的有用的流体系统、检测器和方法例如描述于WIPO专利申请序列号PCT/US11/57111、美国专利申请公开号2005/0191698A1、美国专利号7,595,883和美国专利号7,244,559中,其每个引入本文作为参考。

[0179] 连接反应测序也是有用的,包括例如Shendure等人Science 309:1728-1732(2005);美国专利号 5,599,675;和美国专利号5,750,341中所述的那些方法,其每个引入本文作为参考。一些实施方案可以包括例如在下列文献中所述的杂交测序方法:Bains等人,Journal of Theoretical Biology 135(3),303-7(1988);Drmanac等人,Nature Biotechnology 16,54-58(1998);Fodor等人,Science 251(4995),767-773(1995);和WO 1989/10977,其每个引入本文作为参考。在连接测序和杂交测序方法中,存在于含有凝胶的孔(或其它凹面特征)中的核酸经历寡核苷酸递送和检测的重复循环。用于本文或本文引用的参考文献中的SBS方法的流体系统可以容易地适用于递送连接测序或杂交测序方法中的试剂。通常,寡核苷酸是荧光标记的,并且可以使用类似于本文中所述的SBS方法或本文引用的参考文献中所述的那些荧光检测器来检测。

[0180] 一些实施方案可以利用涉及DNA聚合酶活性的实时监测的方法。例如,可以通过带有荧光团的聚合酶和 γ -磷酸标记的核苷酸之间的荧光共振能量转移(FRET)相互作用,或者用zeromode波导来检测核苷酸掺入。基于FRET的测序的技术和试剂描述于例如,Levene等人Science 299,682-686(2003);Lundquist等人Opt.Lett.33,1026-1028(2008);Korlach等人Proc.Natl.Acad.Sci.USA 105,1176-1181(2008),其公开内容引用本文作为参考。

[0181] 一些SBS实施方案包括检测在将核苷酸掺入到延伸产物中时释放的质子。例如,基于检测所释放的质子的测序可以使用由Ion Torrent(Guilford,CT,Life Technologies的子公司)商购获得的电检测器和相关技术,或者可以使用美国专利申请公开号2009/0026082 A1;2009/0127589 A1;2010/0137143 A1;或2010/0282617 A1中所述的测序方法和系统,其每个并入本文作为参考。

[0182] 示例性修饰的连接基

[0183] 在以下实施例中进一步详细公开了其它实施方案,其不以任何方式限制权利要求的范围。

[0184] 图1A表示标记的核苷酸100的部分结构式。标记的核苷酸100包括完全功能化的腺苷核苷酸(ffA) 110、标准的连接基部分115和荧光染料120。标准的连接基部分115可以是在合成测序(SBS)中通常用于合成标记的核苷酸的连接基部分。在一个实例中,荧光染料120是NR550S0。在该实例中,标记的核苷酸100可以描述为“ffA-NR550S0”。在另一个实例中,荧光染料120是S07181,且标记的核苷酸100可以被描述为“ffA-S07181”。

[0185] 图1B示出了图1A的标记的核苷酸100,其具有对标准的连接基部分115的两种可能的结构修饰。在一个实例中,标准的连接基部分115包括在酰胺基的羰基(即-C(=O)-)和氨基(即-NH-)之间的AEDI插入物125。在该实例中,修饰的标记核苷酸可以描述为“ffA-AEDI-NR550S0”。在另一个实例中,标准的连接基部分115包括SS插入物130,并且修饰的标记核苷酸可以命名为“ffA-SS-NR550S0”。

[0186] 图2显示使用图1A的标记核苷酸100和图1B的经修饰的标记核苷酸100的核苷酸掺入速率的图。在55℃,在40mM乙醇胺(pH 9.8)、9mM MgCl₂、40mM NaCl、1mM EDTA、含有20nM引物:模板DNA和30μg/ml聚合酶812(MiSeq Kit V2)的0.2%CHAPS、1mM核苷酸中进行测定。将酶与DNA结合,然后在淬灭流动池中与核苷酸快速混合短时间(最多10秒),然后用500mM EDTA猝灭。对于每个核苷酸采取几个时间点。然后在变性凝胶上分析所产生的样品,测定转化为DNA+1的DNA的百分比,并相对于时间作图以确定每个核苷酸的一级速率常数。数据总结在下表1中。数据显示,与包含标准的连接基115的标记核苷酸(ffA-NR550S0)的掺入速率相比,包含SS插入物130的标记核苷酸(ffA-SS-NR550S0)的掺入速率为约2倍。包含AEDI插入物125的标记核苷酸(ffA-AEDI-NR550S0)的掺入速率为ffA-NR550S0的掺入速率的约4倍。数据还显示包含标准的连接基115和荧光染料S07181的标记核苷酸(ffA-S07181)的掺入速率为ffA-NR550S0的掺入速率的约4倍。

表 1.	
ffA	K (μM/min)
ffA-NR550S0	22 (1X)
ffA-SS-NR550S0	54 (2X)
ffA-AEDI-NR550S0	97 (4X)
ffA-S07181	99 (4X)

[0188] 图3A-3F示出了图1A的标准的连接基部分115的另外的插入物310、315、320、325、330和335的结构式。在ACA插入物310中,与插入物125相比,除去了二甲基取代基,并将硫-硫(S-S)键用碳-碳键代替。硫-硫键对于SBS不是必需的(例如2-染料或4-染料SBS)。

[0189] 在AcLys插入物315中,乙酰基保护的赖氨酸用于代替插入物125。

[0190] 在BocLys插入物320中,叔丁氧基羰基保护的赖氨酸用于代替插入物125。

[0191] 在dMe0插入物325中,与插入物125相比,将硫-硫(S-S)键用氧-碳(O-CH₂)键代替。

[0192] 在dMeS插入物330中,与插入物125相比,将硫-硫(S-S)键用硫-碳(S-CH₂)键代替。

[0193] 在DMP插入物335中,与插入物125相比,将硫-硫(S-S)键用碳(CH₂)键代替。

[0194] 在本文所述的各种实例中,发现包含二甲基取代模式(例如,AEDI插入物125,dMeO插入物325和dMeS插入物330)的插入物在SBS期间具有增加的核苷酸掺入速率。

[0195] 在本文所述的各种实例中,插入物中的碳链长度也是可以变化的。

[0196] 图4显示用于评价图1B的AEDI插入物125和图3B的AcLys插入物315对测序质量影响的双染料测序运行的数据表。测序在具有人550bp模板和2次150个循环的Miseq杂交平台上进行。将新的染料组、V10/cyan-peg4 A-AEDI550S0、V10/cyan-peg4 A-AcLys550S0和V10/cyan A-AcLys 550S0与标准商业染料组V4和Nova平台V5.75的改进染料组相比较。对于V10/cyan-peg4 A-AEDI550S0、V10/cyan-peg4 A-AcLys550S0和V10/cyan A-AcLys 550S0样品中的每一个,定相值(Ph R1)低于没有AEDI或AcLys插入物的样品的定相值。因此,包含另外的插入物125和315的标记的核苷酸证明了测序质量的改进。

[0197] 图5A和5B分别示出了图4的测序运行的读数1的错误率图和读数2的错误率图。对于读数1,V10/cyan-peg4 A-AEDI550S0和V10/cyan-peg4 A-AcLys550S0的错误率低于没有插入物的相同染料组V10/Cyan-peg4的错误率。对于读数2,当使用AcLys插入物315时更加显著,其中与无插入物染料组相比最终错误率降低30%。因此,已证明插入物125和315明显改善了测序质量。

[0198] 图6显示用于评估AEDI插入物125和ACA插入物310对测序质量影响的测序运行的数据表。测序在具有人550bp模板和2次150个循环的Miseq杂交平台上进行。将新的染料组、V10/cyan-peg4 A-AEDI550S0、V10/cyan-peg4 A-ACALys550S0与标准商业染料组V4和Nova平台V5.75的改进染料组相比较。此外,V10/cyan-peg4 A-AEDI550S0和V10/cyan-peg4 A-ACA550S0样品中的每一个具有比没有AEDI或ACA插入物的样品的定相值更低的定相值(Ph R1),因此显示测序质量的改进。

[0199] 图7A和7B分别示出了图6的测序运行的读数1的错误率图和读数2的错误率图。包含新插入物AEDI的组(V10/cyan-peg4 A-AEDI550S0)的读数1的错误率低于不含插入物V10/Cyan-peg4的相同染料组的错误率。插入物ACA(V10/Cyan-peg4 ACA550S0)在两个读数都获得了与标准V10/Cyan-peg4相似的错误率图。AEDI显示再次提高了测序的质量。那些数据显示插入物本身的结构也对测序质量的提高具有影响。

[0200] 图8A显示了标准的LN₃连接基800的结构式。LN₃连接基800包括第一取代的酰胺官能基团810,第二叠氮基取代的PEG官能基团815和第三酯官能基团820,所述第三酯官能基团在用于将染料分子825连接到核苷酸830的连接基结构中可能是需要的。第一官能基团810可以例如用于将染料分子825连接到LN₃连接基800。第二官能基团815可例如为可用于从LN₃连接基800上裂解染料分子825的可裂解的官能团。第三官能基团820可例如用于将核苷酸830连接到LN₃连接基800。图8B示出了对其中苯氧基基团850被一至四个选自-NO₂、-CN、卤素或-SO₃H的取代基取代的标准LN₃连接基的一些修饰。此外,酯基团820被酰氨基855替代。

[0201] 图9A是显示在具有SS-连接基的ffa中出现杂质的图。ffa-SS-NR550S0在HPLC纯化后:在弱碱性条件(pH8-9)下过夜出现杂质。在室温在0.1M TEAB/CH₃CN中过夜。图9B比较了具有SS-连接基和AEDI-连接基的ffa的稳定性,其中与SS连接基相比,AEDI-连接基显示出明显改善的稳定性。使用ffa-LN₃-NR550S0作为内标。二硫化物副产物:(NR550S0-S-) 2。图9C显示在IMX 60°中22小时具有SS-连接基和AEDI-连接基的ffa的对比,再次显示SS连接基

出现杂质。内标是ffA-LN3-NR550S0。Di-P：二磷酸。

[0202] 图10A、10B和10C显示了在连接基改变的溶液中核苷酸掺入速度的意外增加。图10A显示在1 μ M下的掺入速率。结果显示染料和连接基对掺入动力学具有显著影响，其中表10B清楚地显示了AEDI连接基在掺入动力学中的益处。图10C图示了具有NR550S0的AEDI和SS连接基。

[0203] 图11A显示了具有不同A-550S0 (相同浓度)的V10组合的散点图。散点图针对格(tile)1循环2。图11B显示了溶液中的Kcat FFA连接基。可以看出在表面上就掺入率而言，AEDI连接基比没有连接基的更快：‘A’-云(在散点图上)略微向中心移动。AcLys连接基比AEDI连接基和没有连接基的更慢：‘A’-云向x轴移动。BocLys连接基看起来类似于没有连接基的并且与AEDI连接基差距不大。在溶液中可以看出ACA连接基是最慢的，而AEDI连接基和ACLys连接基具有相似的Kcat，其次是BocLys连接基。

[0204] 图12A和12B显示了对于M111、Human550共2 \times 151个循环的测序指标。使用ffNs与不同的A-550S0 (相同浓度)组合。可以看出，AEDI连接基和BocLys连接基都给出了相似的良好测序结果。虽然AEDI和AcLys的解决方案Kcat相似，但AEDI的测序结果更好。

[0205] LN₃连接基800包括可以从LN₃连接基800除去的任选的苯氧基835，如图8C所示。LN₃连接基800还包括任选的酰胺基840和任选的醚基845，如图8D所示，两者都可以从LN₃连接基800上去除。去除某些官能团如酰胺基810、苯氧基835的目的是测试它们在核苷酸掺入期间是否与酶具有负相互作用，这可能降低掺入效率。

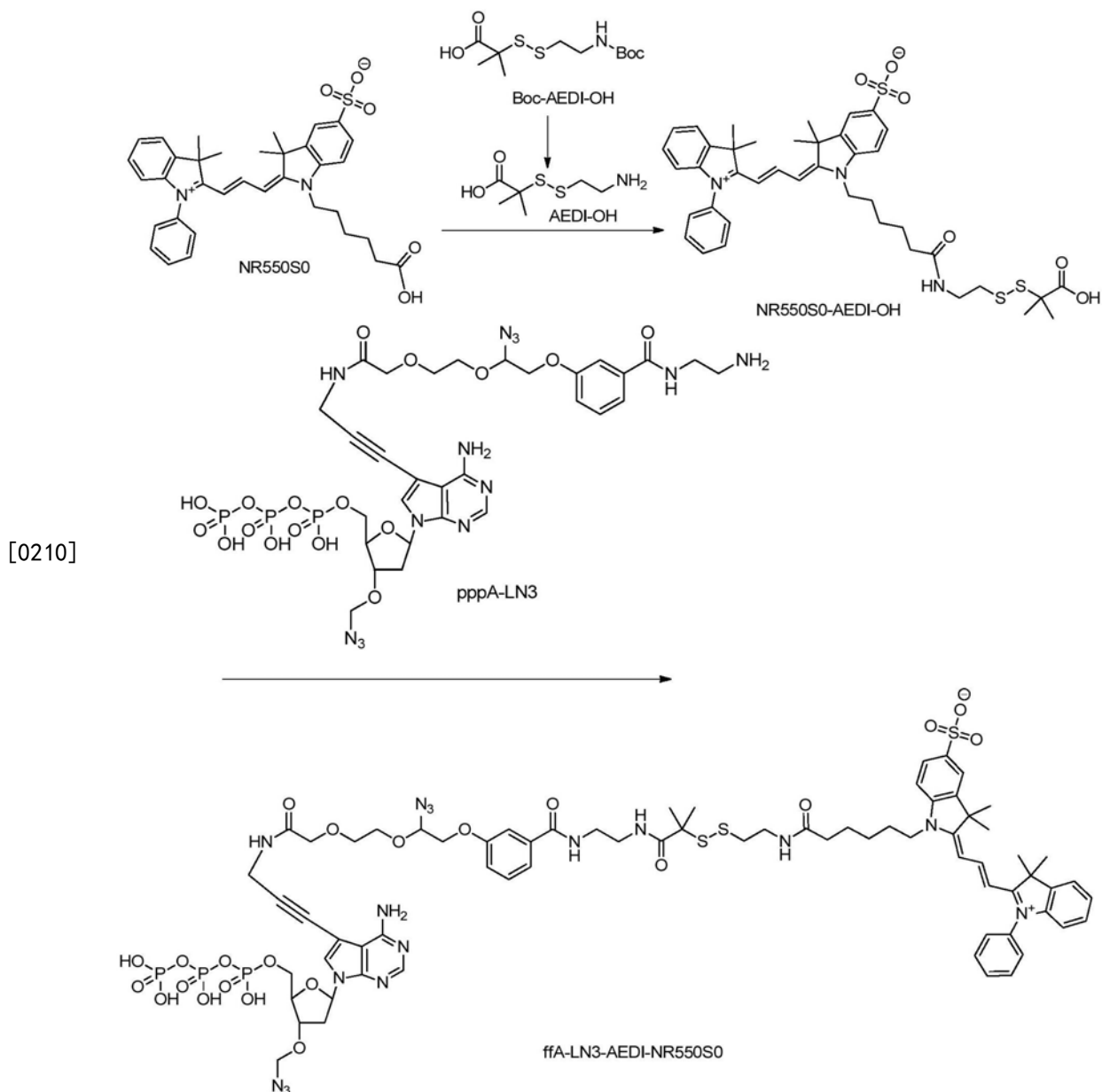
[0206] 图8E显示保护基860插入或加入到连接基中。保护基860插入到官能基团810(或可以连接到图8A-8C中的苯氧基835或850上)和染料分子825之间。保护基860可以是例如防止DNA损伤的分子。包括光损伤或其它化学损伤的DNA损伤是SBS的累积效应(即不断循环(cycle by cycle))之一。基本上减少或消除DNA损伤可以提供更有效的SBS和更长的测序读数。在一些实施方案中，保护基860可选自三线态猝灭剂例如Trolox、没食子酸、2-巯基乙醇(BME)等。在一些其他实施方案中，保护基860可选自淬灭剂或保护剂，例如4-硝基苄基醇或抗坏血酸的盐如抗坏血酸钠。在一些其他实施方案中，保护剂可物理混合到缓冲液中，而不是与标记的核苷或核苷酸形成共价键。然而，该方法可能需要更高浓度的保护剂，并且可能效率更低。或者，共价连接于核苷或核苷酸的保护基可以提供更好的对抗DNA损伤的保护。在图8C、8D和8E的一些另外的实施方案中，酯基基团820也可以被酰胺部分855和苯氧基基团835进一步取代。

[0207] 在图8A-8E所示的任何实例中，图1B的AEDI插入物125和SS插入物130以及图3A-3F的插入物300可以例如插入到连接基800的第一功能基团810和染料分子825之间。

实施例

[0208] 在以下实施例中进一步详细公开了其它实施方案，其不以任何方式限制权利要求的范围。

[0209] 制备ffA-LN₃-AEDI-NR550S0的一般反应过程：



[0211] 在50ml圆底烧瓶中,将Boc-AEDI-OH (1g, 3.4mmol) 溶于DCM (15ml) 中,并在室温将TFA (1.3ml, 17ml) 加入到溶液中。将反应混合物搅拌2小时。TLC (DCM:MeOH=9:1) 表明Boc-AEDI-OH完全消耗。将反应混合物蒸发至干。然后向残余物中加入TEAB (2M, ~15ml), 监测pH直到中性。然后将混合物溶于H₂O/CH₃CN (1:1, ~15ml) 中并蒸发至干。将该过程重复3次, 以除去过量的TEAB盐。将白色固体残余物用CH₃CN (20ml) 处理并搅拌0.5小时。滤出溶液, 并用CH₃CN洗涤固体, 得到纯的AEDI-OH-TFA盐 (530mg, 80%)。¹H NMR (400MHz, D₂O, δ (ppm)) : 3.32 (t, J=6.5Hz, 2H, NH₂-CH₂) ; 2.97 (t, J=6.5Hz, 2H, S-CH₂) ; 1.53 (s, 6H, 2x CH₃)。 ¹³C NMR (400MHz, D₂O, δ (ppm)) : 178.21 (s, CO) ; 127.91, 117.71 (2s, TFA) ; 51.87 (s, C-(CH₃)₂) ; 37.76 (t, CH₂-NH₂) ; 34.23 (t, S-CH₂) ; 23.84 (q, 2x CH₃)。 ¹⁹F NMR (400MHz, D₂O, δ (ppm)) : -75.64。

[0212] 在50ml圆底烧瓶中,将染料NR550S0 (114mg, 176μmol) 溶于DMF (无水, 20mL) 中, 并蒸发至干。将该过程重复3次。然后将无水DMA (10mL) 和Hunig's碱 (92μL, 528μmol, 3当量) 移液到圆底烧瓶中。一次性加入TSTU (69mg, 228μmol, 1.3当量)。将反应混合物保持在室温。30分钟后, TLC (CH₃CN:H₂O=85:15) 分析表明反应完成。向反应混合物中加入在0.1M TEAB中的

AEDI-OH (68mg, 352 μ mol, 2当量), 并在室温搅拌3小时。TLC (CH₃CN:H₂O=8:2) 显示活化酯完全消耗, 并且在活化酯下面出现红色斑点。同时, 分析型HPLC还表明活化酯完全消耗且产物形成。用TEAB缓冲液 (0.1M, 10ml) 淬灭反应, 通过减压蒸发 (HV) 除去挥发性溶剂, 并在Axia柱上纯化, 得到NR550S0-AEDI-OH。产率: 60%。

[0213] 在25ml圆底烧瓶中, 将NR550S0-AEDI-OH (10 μ mol) 溶于DMF (无水, 5mL) 中, 并蒸发至干。将该过程重复3次。然后将无水DMA (5mL) 和DMAP (1.8mL, 15 μ mol, 1.5当量) 加入到圆底烧瓶中。一次性加入DSC (5.2mg, 20 μ mol, 2当量)。将反应混合物保持在室温。30分钟后, TLC (CH₃CN:H₂O=8:2) 分析表明反应完成。将Hunig碱 (3.5 μ l, 20 μ mol) 移液到反应混合物中。然后将pppA-LN₃ (20 μ mol在0.5ml H₂O中, 2当量) 和Et₃N (5 μ l) 的溶液加入到反应混合物中, 并在室温搅拌过夜。TLC (CH₃CN:H₂O=8:2) 显示活化酯完全消耗, 并且在基线上出现红色斑点。同时, 分析型HPLC还表明活化酯完全消耗, 且最终产物形成。用TEAB缓冲液 (0.1M, 10ml) 淬灭反应, 并装载在DEAE Sephadex柱 (25g Biotage柱) 上。用如下表2所示的梯度洗脱柱。

[0214] A: 0.1M TEAB缓冲液 (10%CH₃CN)

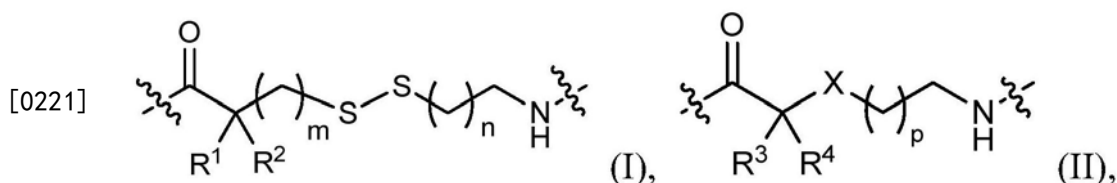
[0215] B: 1M TEAB缓冲液 (10%CH₃CN)。

[0216] 梯度:

表2		
步骤	溶剂混合物(B%)	体积(ml)
[0217]	1	0
	2	0-45
	3	45
[0218]	4	45-100
	5	100

[0219] 由45%至100%1M TEAB缓冲液中洗脱出所需产物。合并含有产物的级分, 蒸发并通过HPLC (YLC柱, 8ml/min) 纯化。产率: 53%。

[0220] 总之, 本发明涉及通过连接基共价连接到荧光团的核苷或核苷酸, 其中所述连接基包含式 (I) 或 (II) 的结构或两者的组合:



[0222] 其中

[0223] 每个R¹和R²独立地选自氢或任选取代的C₁₋₆烷基;

[0224] R³选自氢、任选取代的C₁₋₆烷基、-NR⁵-C(=O)R⁶或-NR⁷-C(=O)-OR⁸;

[0225] R⁴选自氢或任选取代的C₁₋₆烷基;

[0226] 每个R⁵和R⁷独立地选自氢、任选取代的C₁₋₆烷基、任选取代的苯基、或任选取代的

C₇₋₁₂芳烷基；

[0227] 每个R⁶和R⁸独立地选自任选取代的C₁₋₆烷基、任选取代的苯基、任选取代的C₇₋₁₂芳烷基、任选取代的C₃₋₇环烷基、或任选取代的5至10元杂芳基；

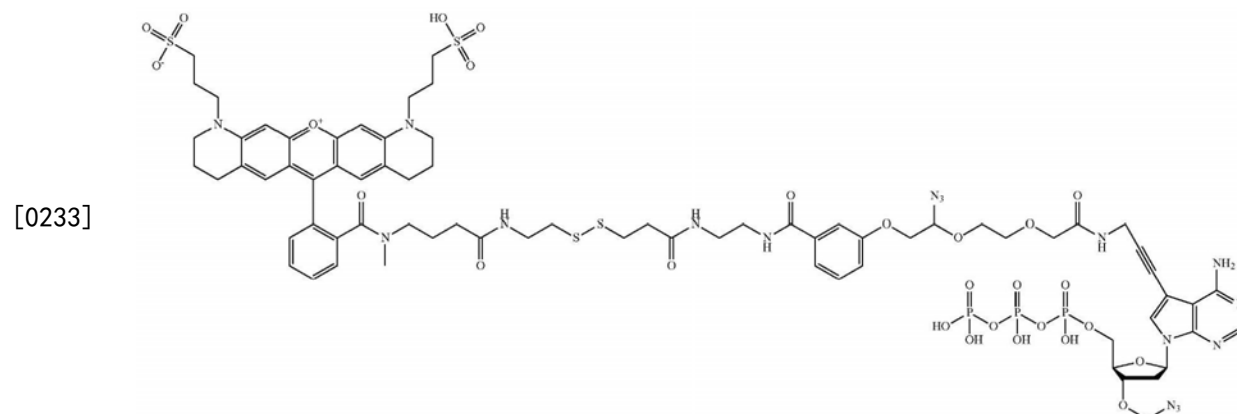
[0228]  中的每个亚甲基重复单元任选被取代；

[0229] X选自亚甲基(CH₂)、氧(O)或硫(S)；

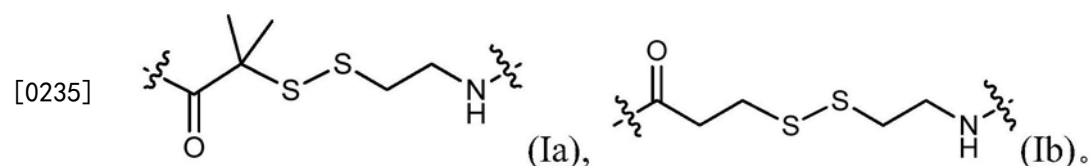
[0230] m为0至20的整数；

[0231] n为1至20的整数；和

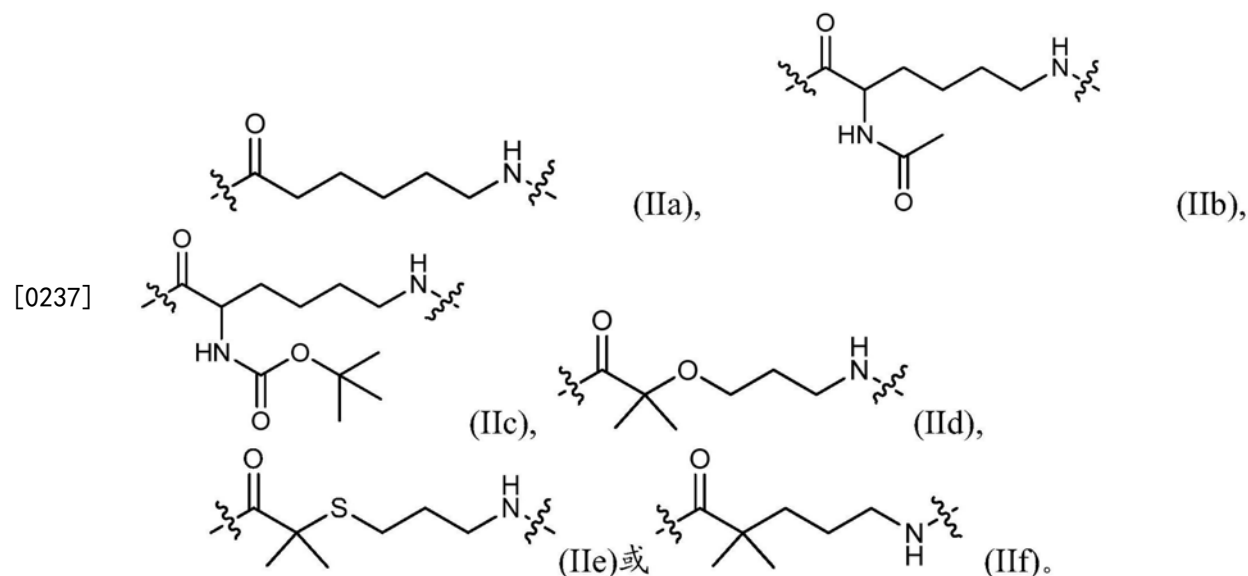
[0232] p为1至20的整数，条件是荧光团标记的核苷或核苷酸不具有以下结构：



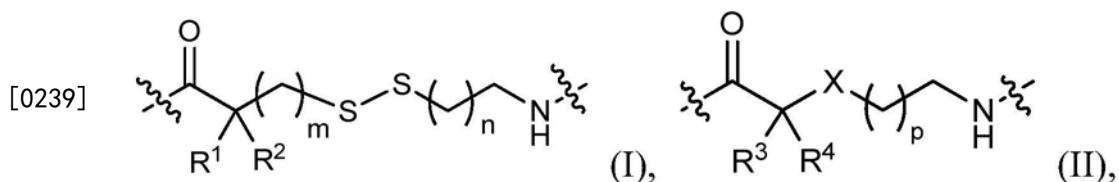
[0234] 在上述核苷或核苷酸的一些情况下，式(I)的结构也由式(Ia)或(Ib)表示：



[0236] 此外，式(II)的结构也可由式(IIa)、(IIb)、(IIc)、(IId)、(IIe)或(IIf)所代表：



[0238] 更具体地，本发明可涉及通过连接基共价连接到荧光团的核苷或核苷酸，其中所述连接基包含式(I)或(II)的结构或两者的组合：



[0240] 其中：

[0241] R^1 选自任选取代的 C_{1-6} 烷基；

[0242] R^2 选自氢或任选取代的 C_{1-6} 烷基；

[0243] R^3 选自任选取代的 C_{1-6} 烷基、 $-NR^5-C(=O)R^6$ 或 $-NR^7-C(=O)-OR^8$ ；

[0244] R^4 选自氢或任选取代的 C_{1-6} 烷基；

[0245] 每个 R^5 和 R^7 独立地选自氢、任选取代的 C_{1-6} 烷基、任选取代的苯基、或任选取代的 C_{7-12} 芳烷基；

[0246] 每个 R^6 和 R^8 独立地选自任选取代的 C_{1-6} 烷基、任选取代的苯基、任选取代的 C_{7-12} 芳烷基、任选取代的 C_{3-7} 环烷基、或任选取代的5至10元杂芳基；

[0247] 中的每个亚甲基重复单元任选被取代；

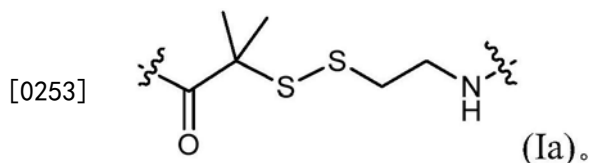
[0248] X选自亚甲基(CH_2)、氧(O)或硫(S)；

[0249] m为0至20的整数；

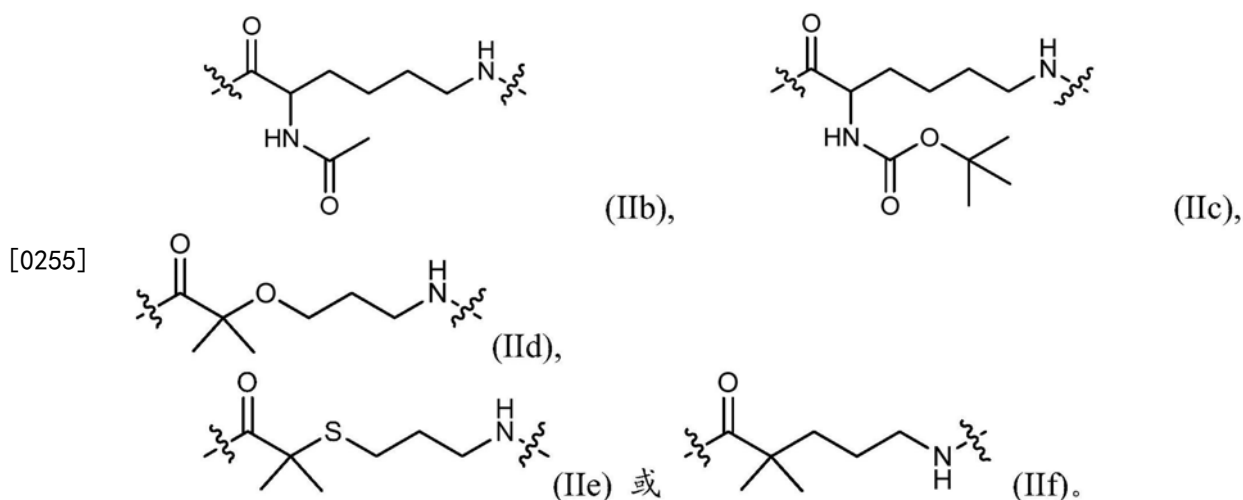
[0250] n为1至20的整数；和

[0251] p为1至20的整数。

[0252] 优选地，式(I)的结构也由式(Ia)表示：



[0254] 此外，式(II)的结构也由式(IIb)、(IIc)、(IId)、(IIe)或(IIf)所代表：



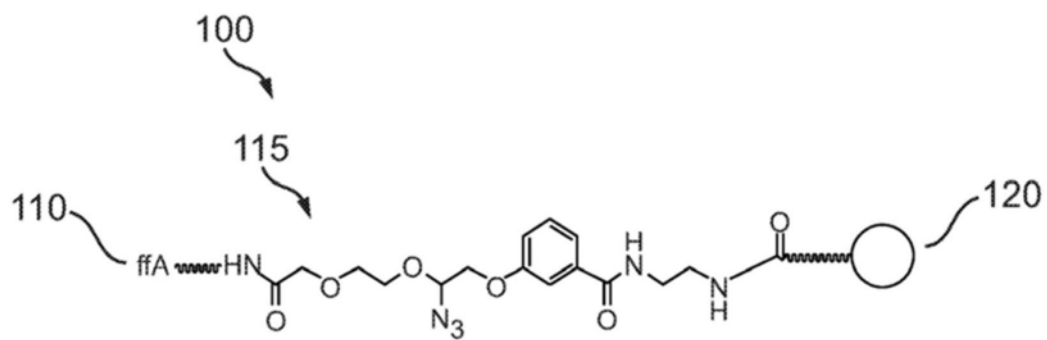


图1A

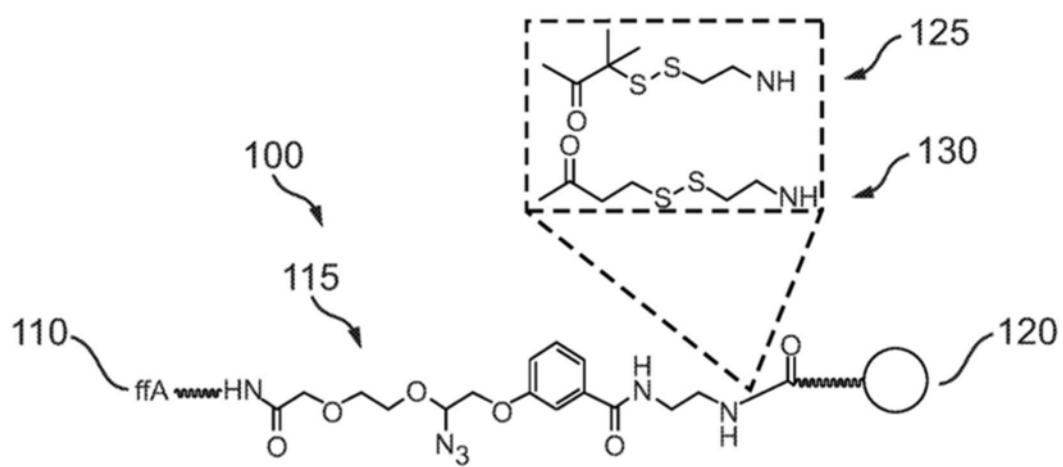


图1B

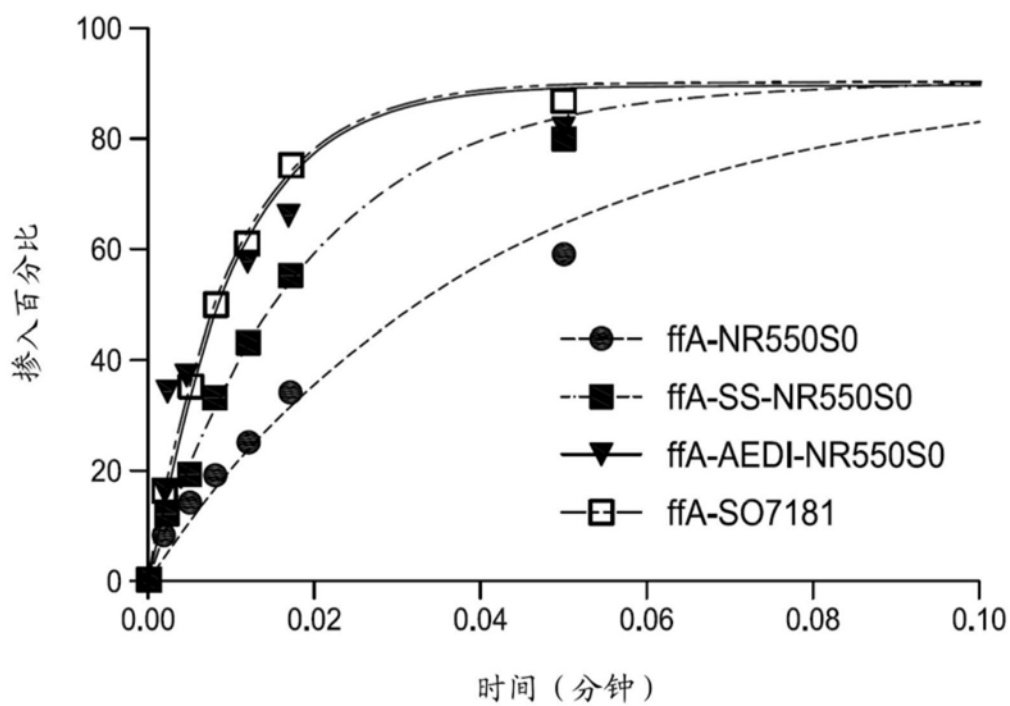
在 $1\ \mu\text{M}$ 下的掺入速率

图2

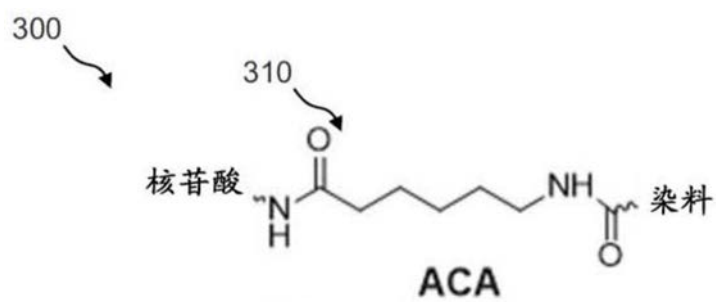


图3A

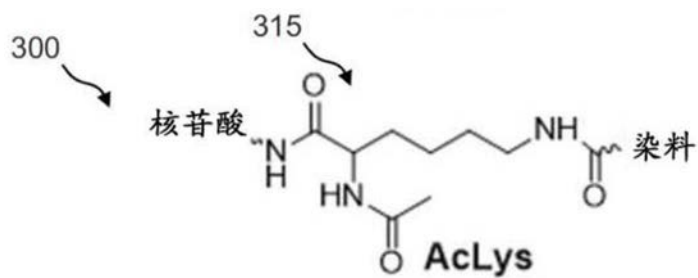


图3B

图 3C

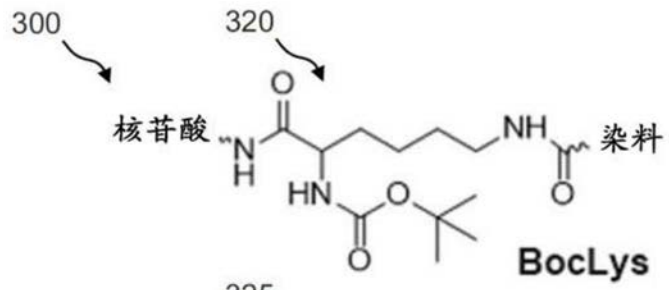


图 3D

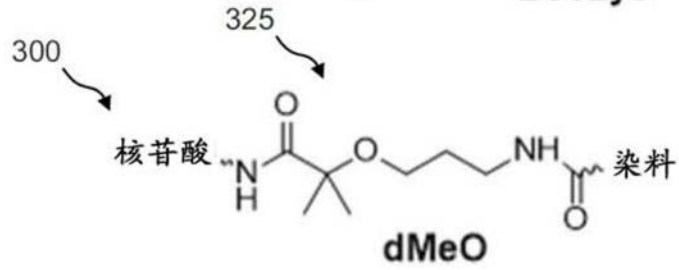


图 3E

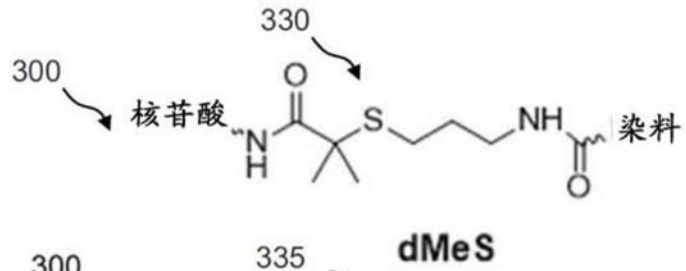
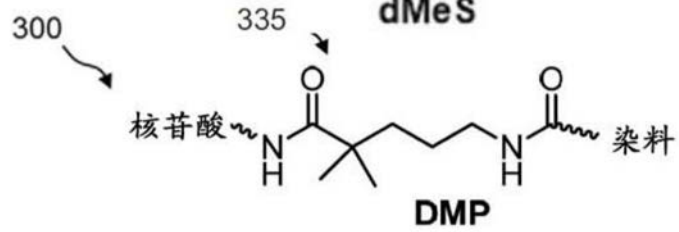


图 3F



染料组	密度	%PF	Ph R1	PPh R1	Ph R2	PPh R2	% 比对 R1	% 比对 R2	错误 R1	错误 R2
V4	230	79.6	0.205	0.111	0.35	0.149	82.2	79.3	1.41	3.19
V5.75	187	77.65	0.371	0.048	0.442	0.164	80.06	79.99	2.16	3.94
V10/cyan-peg4	222	79.88	0.305	0.084	0.53	0.179	81.89	79.45	1.42	3.48
V10/cyan-peg4, A-AEDI550S0	223	81.83	0.14	0.123	0.294	0.197	81.61	67.12	0.85	5.84
V10/cyan-peg4, A-AcLys550S0	227	78.52	0.2	0.128	0.319	0.272	82.23	78.52	1.44	3.48
V10/cyan, A-AcLys550S0	145	86.27	0.178	0.211	0.283	0.29	82.4	79.84	1.33	3.33

图4

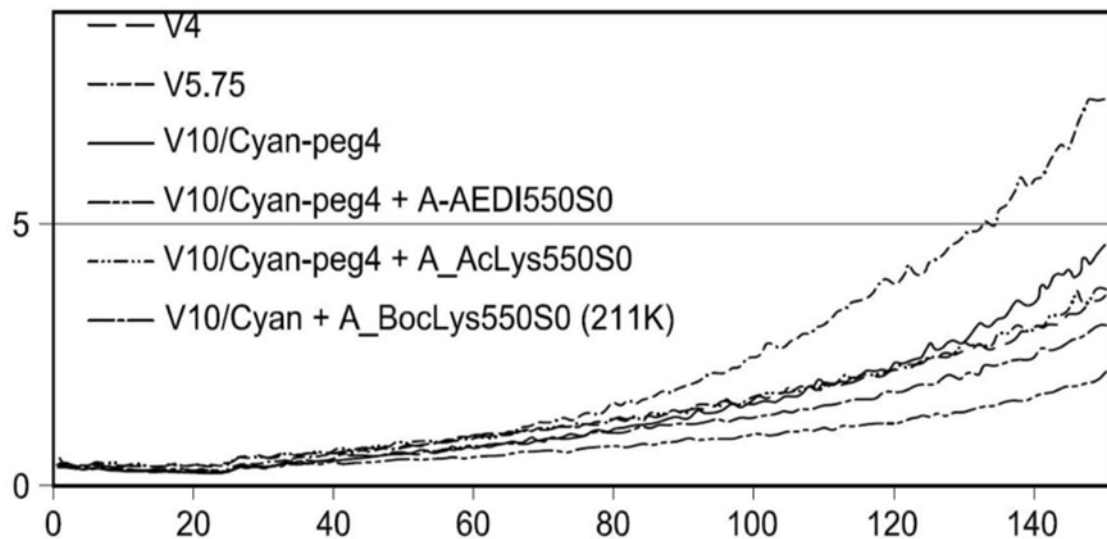


图5A

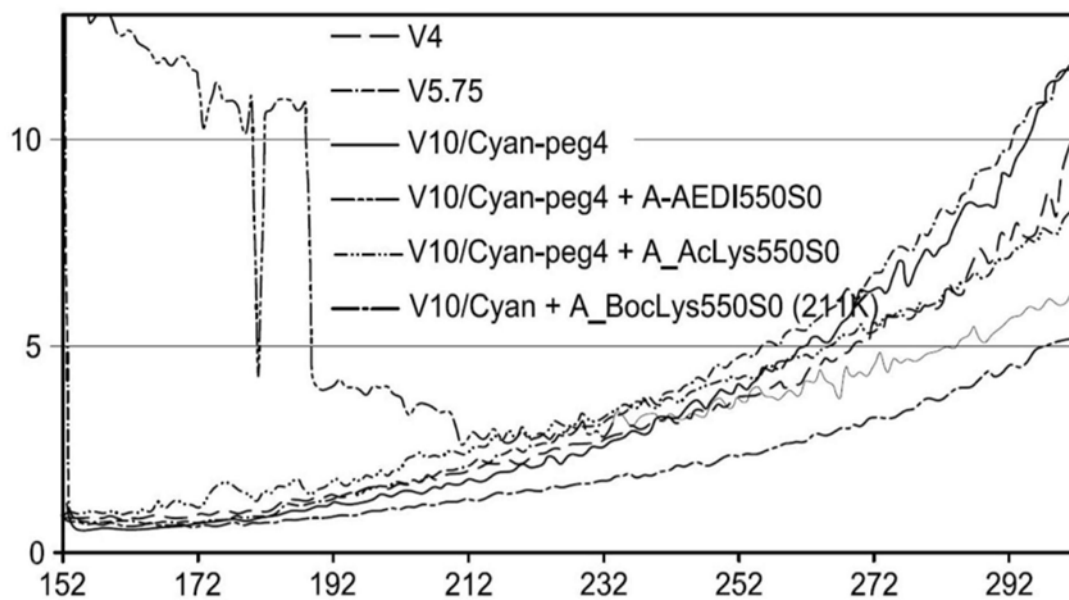


图5B

染料组	密度	%PF	Ph R1	PPh R1	Ph R2	PPh R2	% 比对 R1	% 比对 R2	错误 R1	错误 R2
V4	191	82.62	0.132	0.151	0.242	0.163	81.91	79.59	1.34	2.43
V5.75	235	79.02	0.212	0.05	0.304	0.187	82.54	78.74	1.4	4.2
V10/Cyan-peg4	213	84.7	0.2	0.063	0.348	0.179	83.3	81.19	0.96	2.02
V10/Cyan-peg4	220	83.03	0.188	0.115	0.317	0.207	82.92	80.9	1.11	2.16
V10/Cyan-peg4, A-AEDI-550S0	252	76.37	0.102	0.097	0.132	0.412	80.1	34.9	0.84	25.58
V10/Cyan-peg4, A-ACA550S0	204	85.88	0.145	0.076	0.271	0.222	83.21	80.58	0.96	2.39

图6

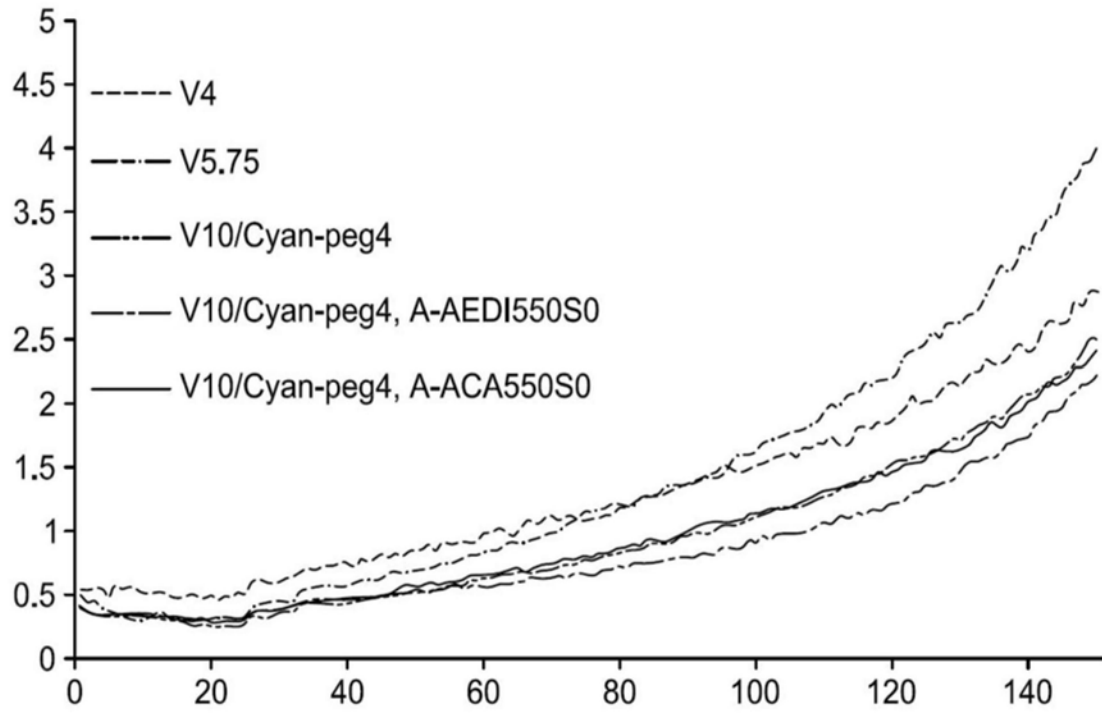


图7A

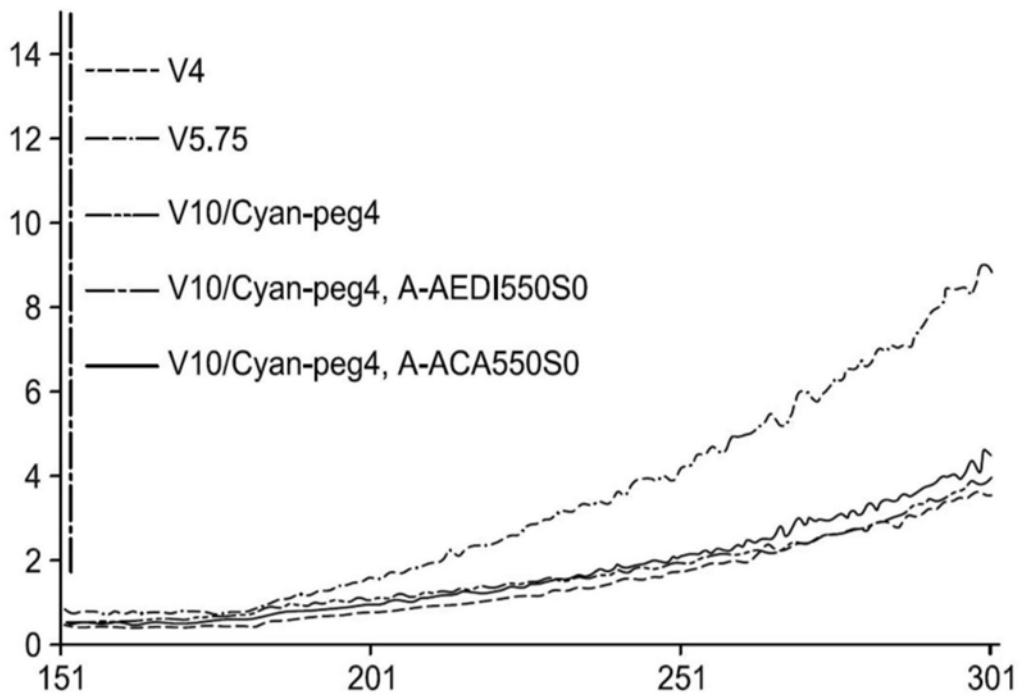


图7B

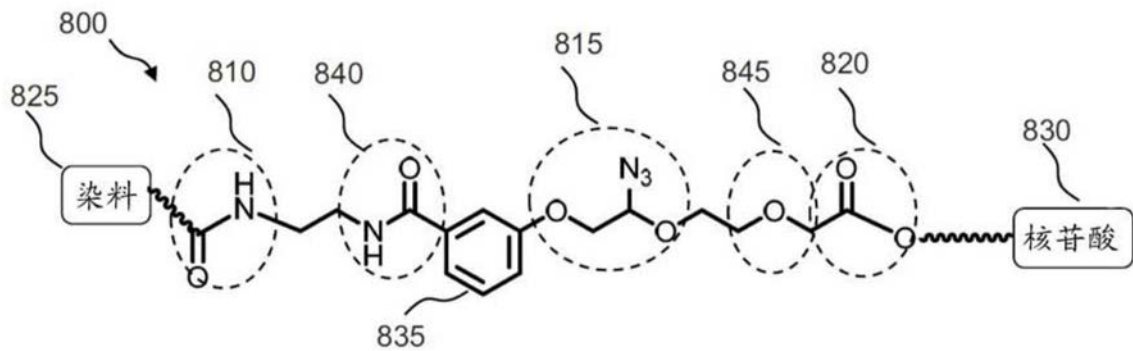


图8A

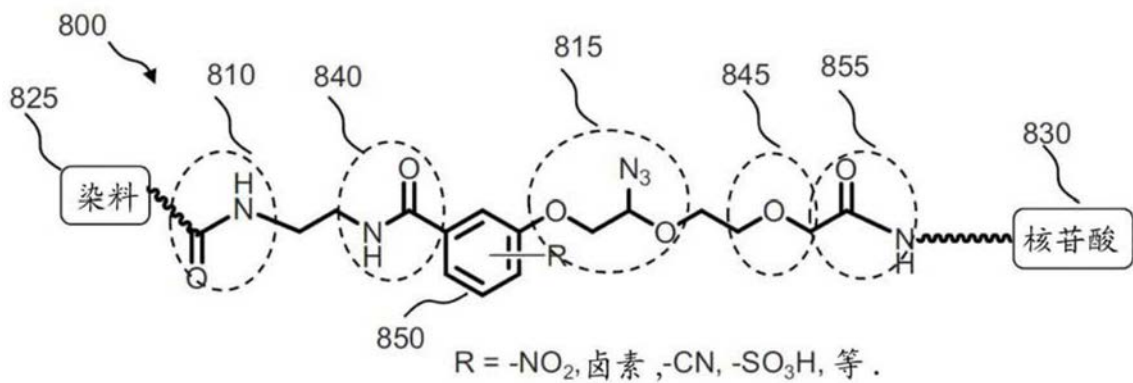


图8B

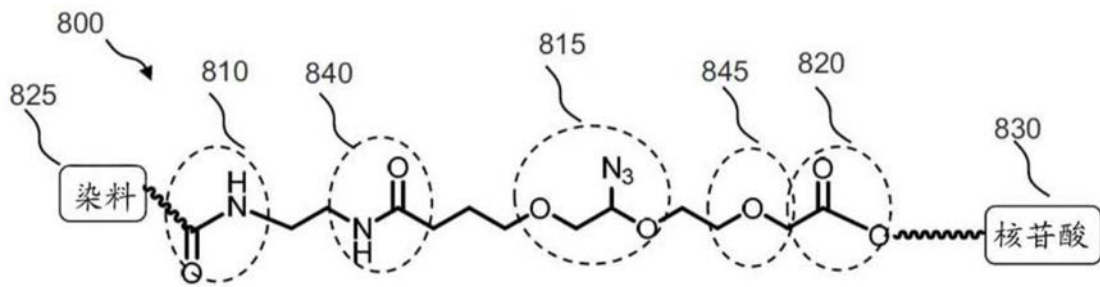


图8C

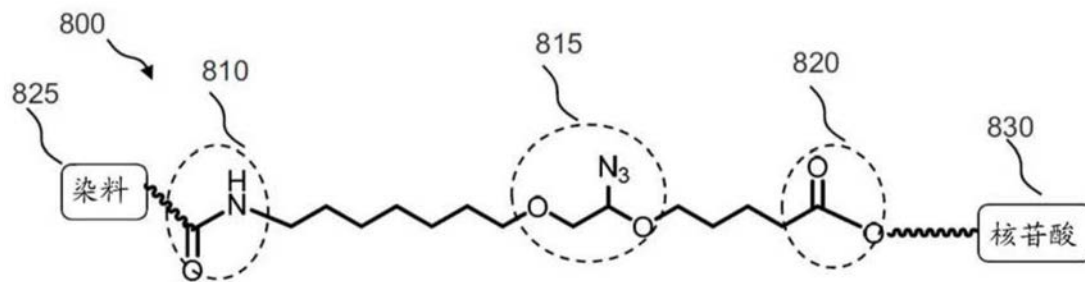


图8D

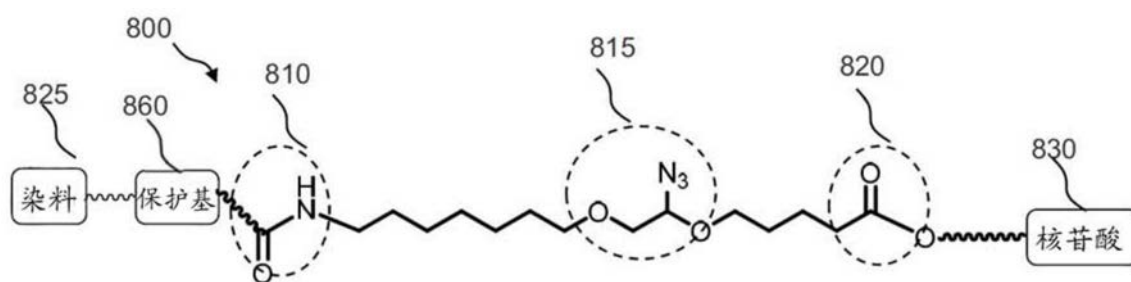


图8E

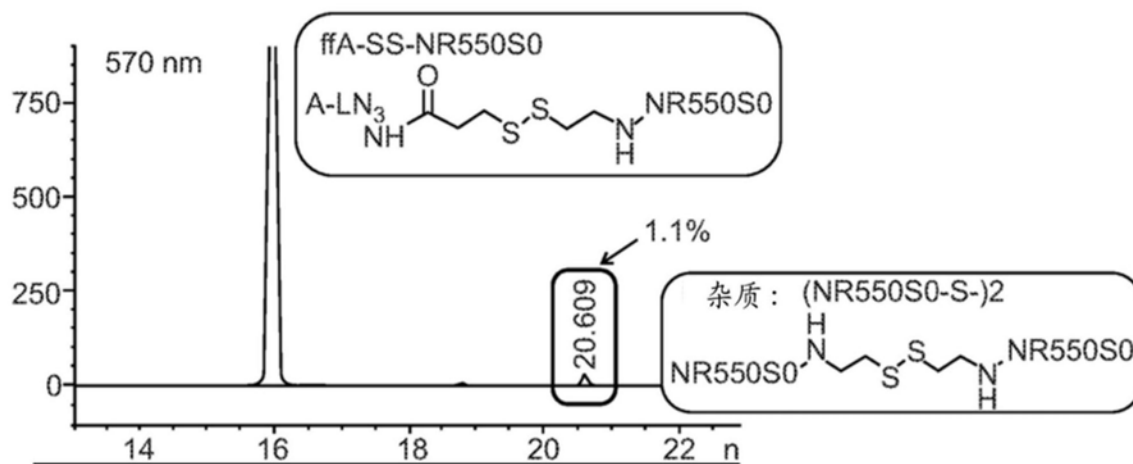


图9A

	缓冲液	结果*	
	21 小时 @ 60°C	ffA-SS-NR550S0	ffA-AEDI-NR550S0
1	PR2 缓冲液	~50% 二硫化物副产物	~6% 二硫化物副产物
2	掺入缓冲液 (pH 9.9) (无 Mg ²⁺ , 无 Chaps)	~10% 二硫化物副产物	没有二硫化物副产物
3	IMX	~38% 二硫化物副产物	没有二硫化物副产物
4	IMX + Pol 812	~34% 二硫化物副产物	没有二硫化物副产物
5	SRE	没有二硫化物副产物	没有二硫化物副产物
6	TEAB (0.1M): CH ₃ CN (1.1)	~20% 二硫化物副产物	没有二硫化物副产物

图9B

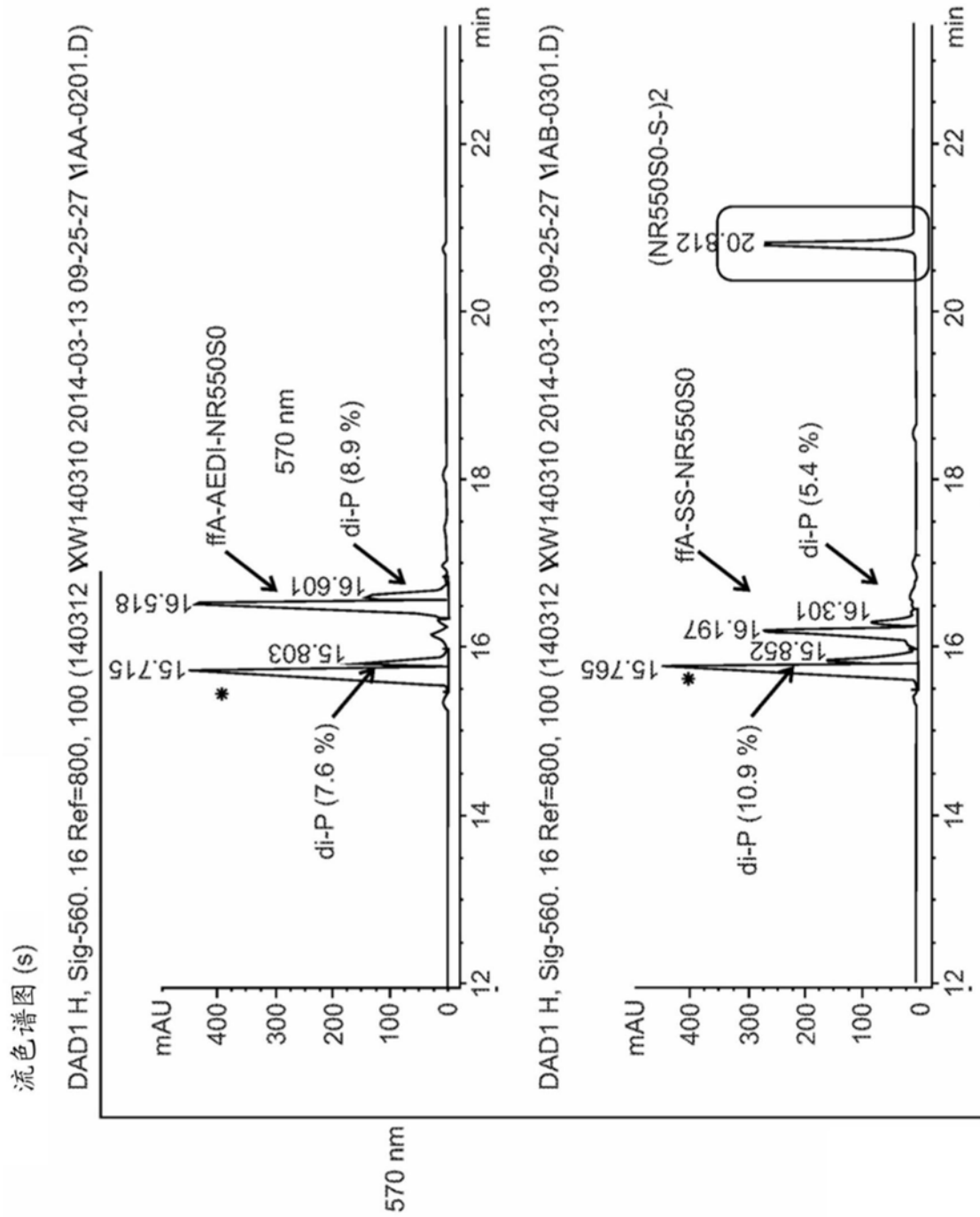


图9C

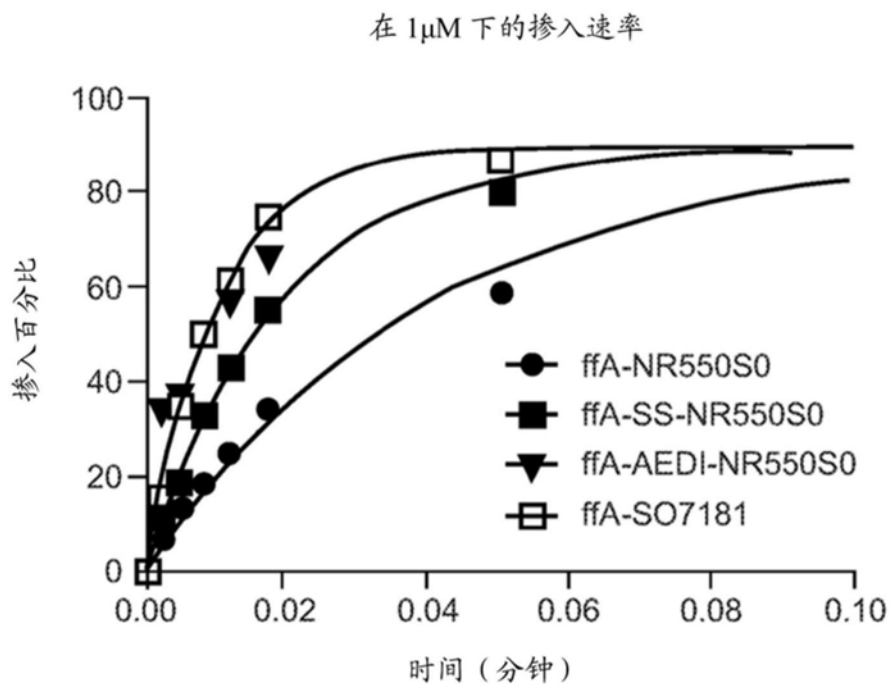


图10A

ffA	K (uM/分钟)
NR550S0	22 (1X)
-SS-NR550S0	54 (2X)
-AEDI-NR550S0	97 (4X)
Std -SO7181	99 (4X)

图10B

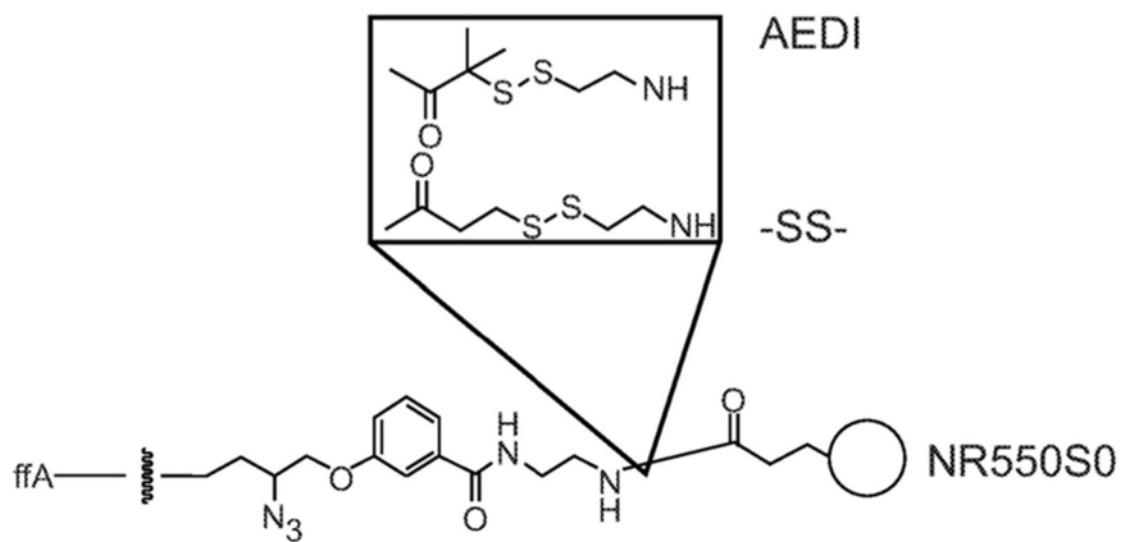


图10C

具有不同 A-550S0 (相同浓度) 的
V10 组合的散点图, 格 1 循环 2

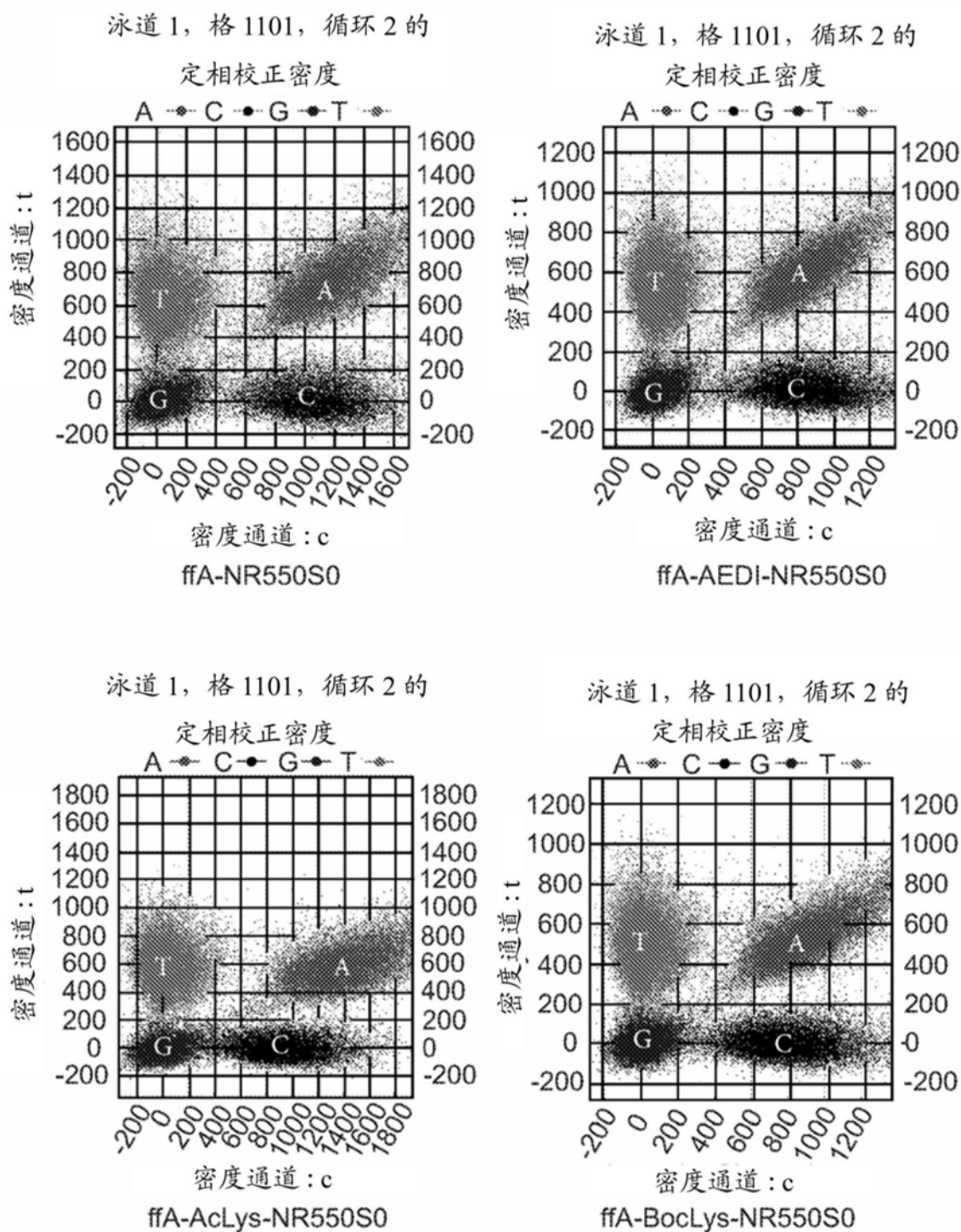


图11A

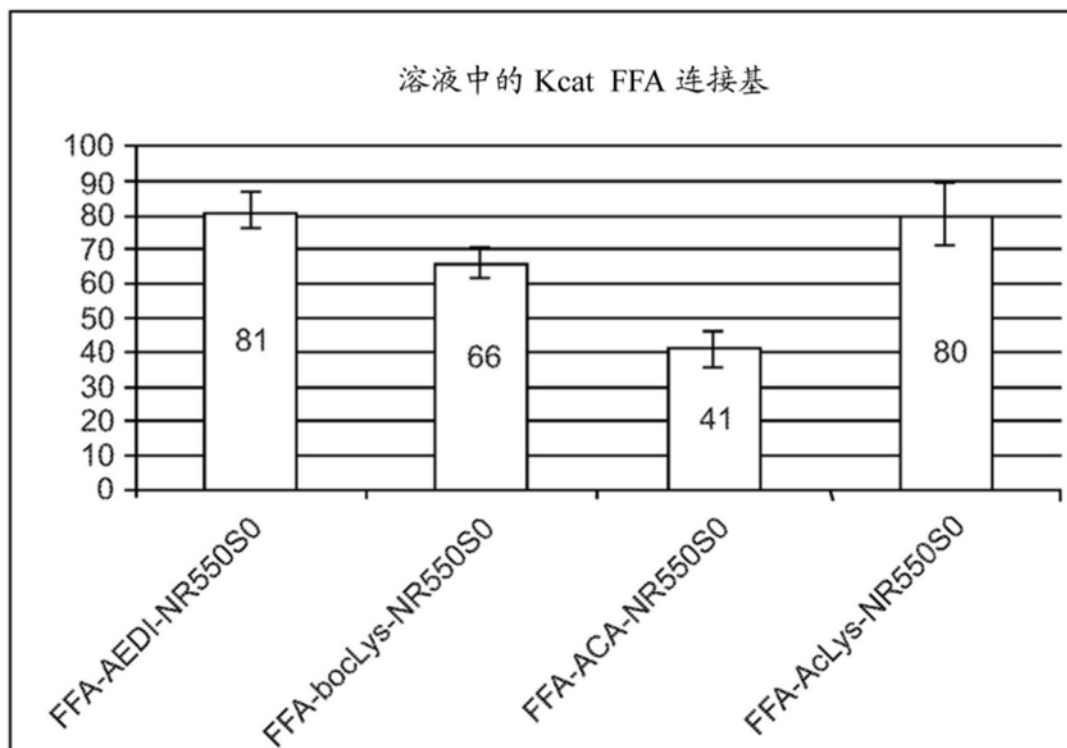


图11B

染料组	密度	%PF	Ph R1	PPh R1	Ph R2	PPh R2	% 比对 R1	% 比对 R2	错误 R1	错误 R2
ffNs + ffA-NR550S0	222	79.88	0.305	0.084	0.53	0.179	81.89	79.45	1.42	3.48
ffNs + ffA-AEDI-NR550S0	223	81.83	0.14	0.123	0.294	0.197	81.61	67.12*	0.85	5.84*
ffNs + ffA-AcLys-NR550S0	227	78.52	0.2	0.128	0.319	0.272	82.23	78.52	1.44	3.48
ffNs + ffA-BocLys-NR550S0	211	82.97	0.126	0.146	0.226	0.249	82.54	80.45	1.17	1.99

图12A

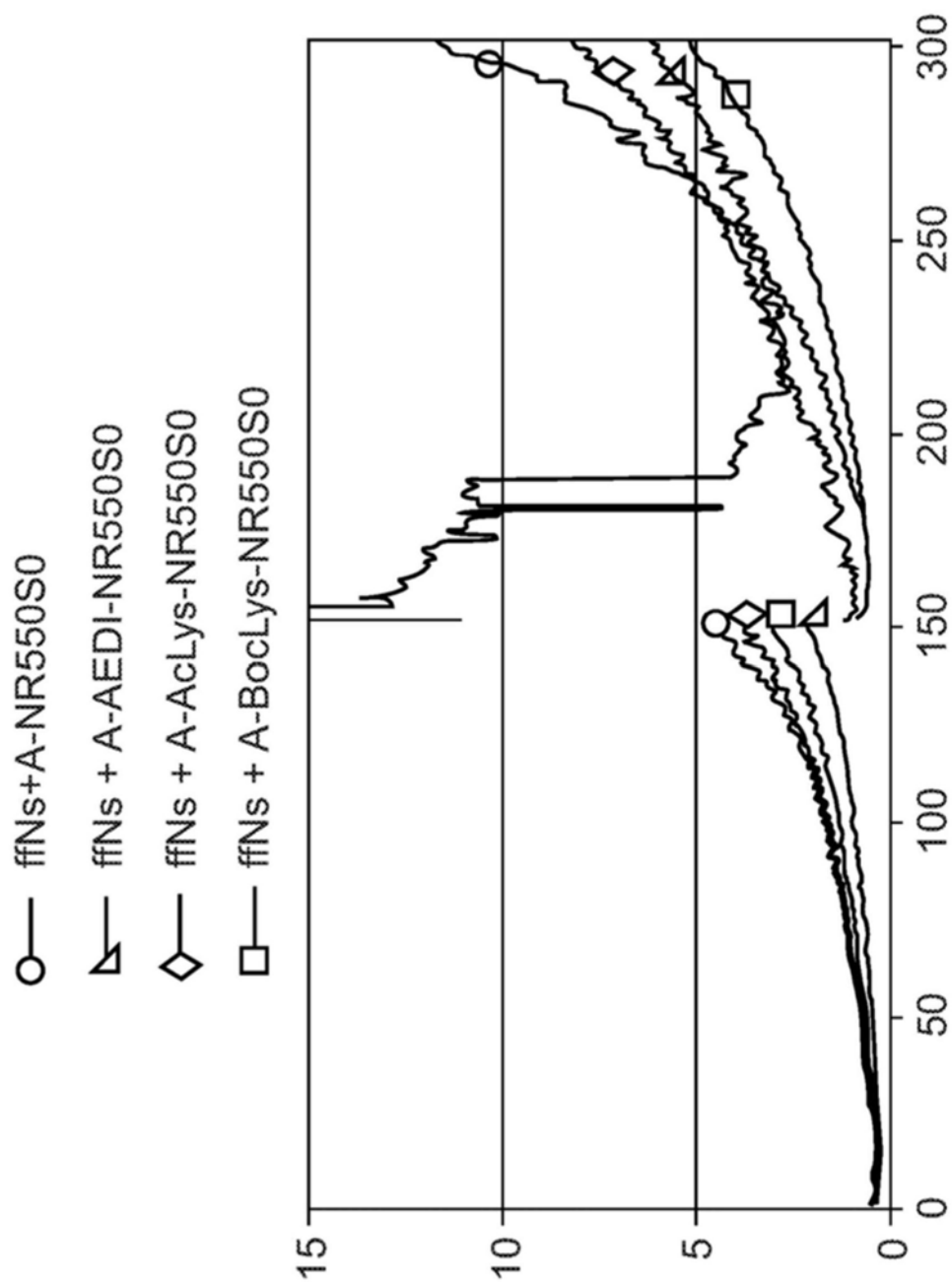


图12B