



(12) 发明专利申请

(10) 申请公布号 CN 118345072 A

(43) 申请公布日 2024. 07. 16

(21) 申请号 202410342599.0

(22) 申请日 2017.12.29

(30) 优先权数据

62/441,046 2016.12.30 US

62/492,001 2017.04.28 US

(62) 分案原申请数据

201780085248.4 2017.12.29

(71) 申请人 爱迪塔斯医药公司

地址 美国马萨诸塞州

(72) 发明人 J·海尔 S·金 S·萨科曼诺

S·凯普哈特 C·费尔南德斯

H·嘉亚拉穆 B·伊顿

K·Z·贝里

(74) 专利代理机构 北京市君合律师事务所

11517

专利代理师 何箴 张梦倩

(51) Int.Cl.

C12N 15/113 (2010.01)

C12N 9/22 (2006.01)

权利要求书3页 说明书208页

序列表(电子公布) 附图55页

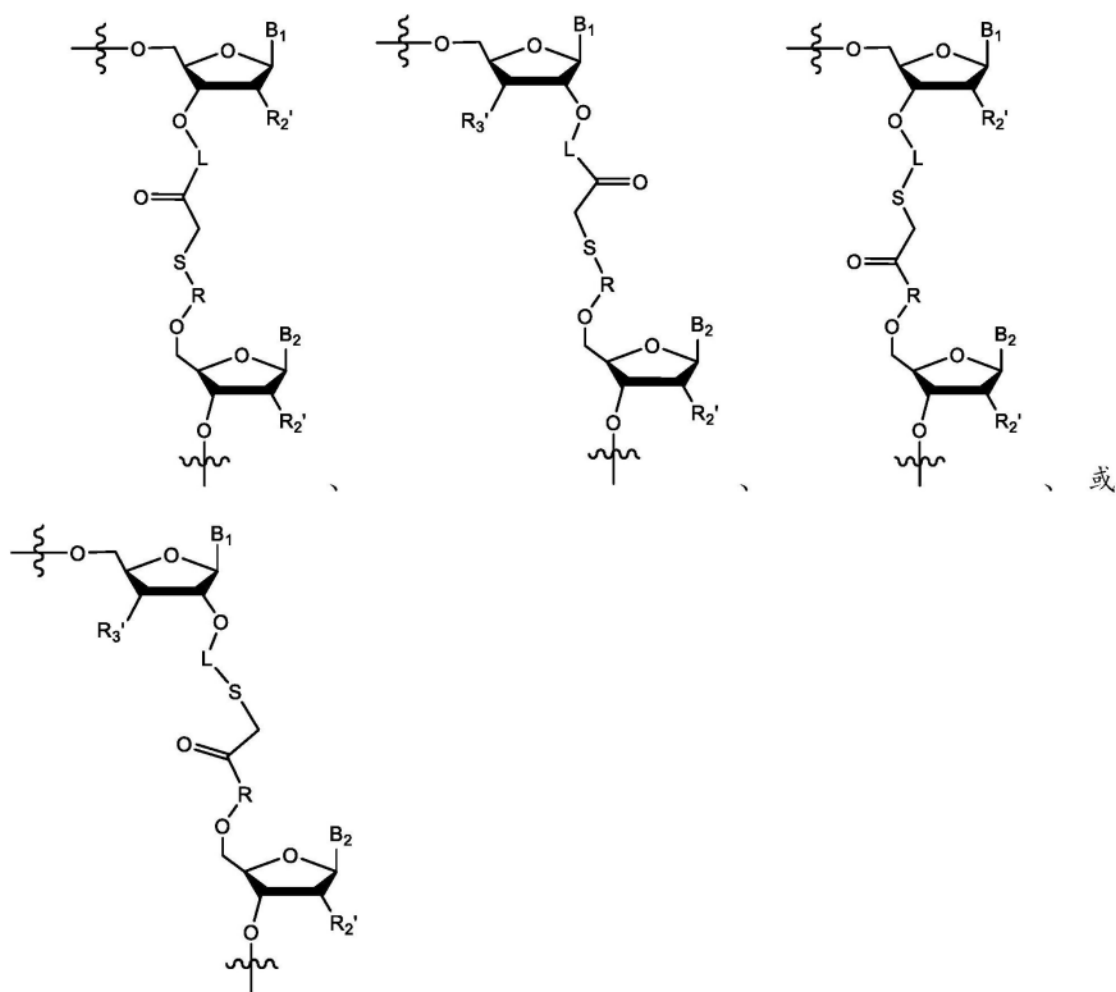
(54) 发明名称

合成的指导分子、组合物和与其相关的方法

(57) 摘要

披露了指导分子的化学合成,以及与其相关的组合物和方法。

1. 一种用于CRISPR系统的合成的单分子指导分子,其中所述指导分子包含具有下式的化学键合:



其中:

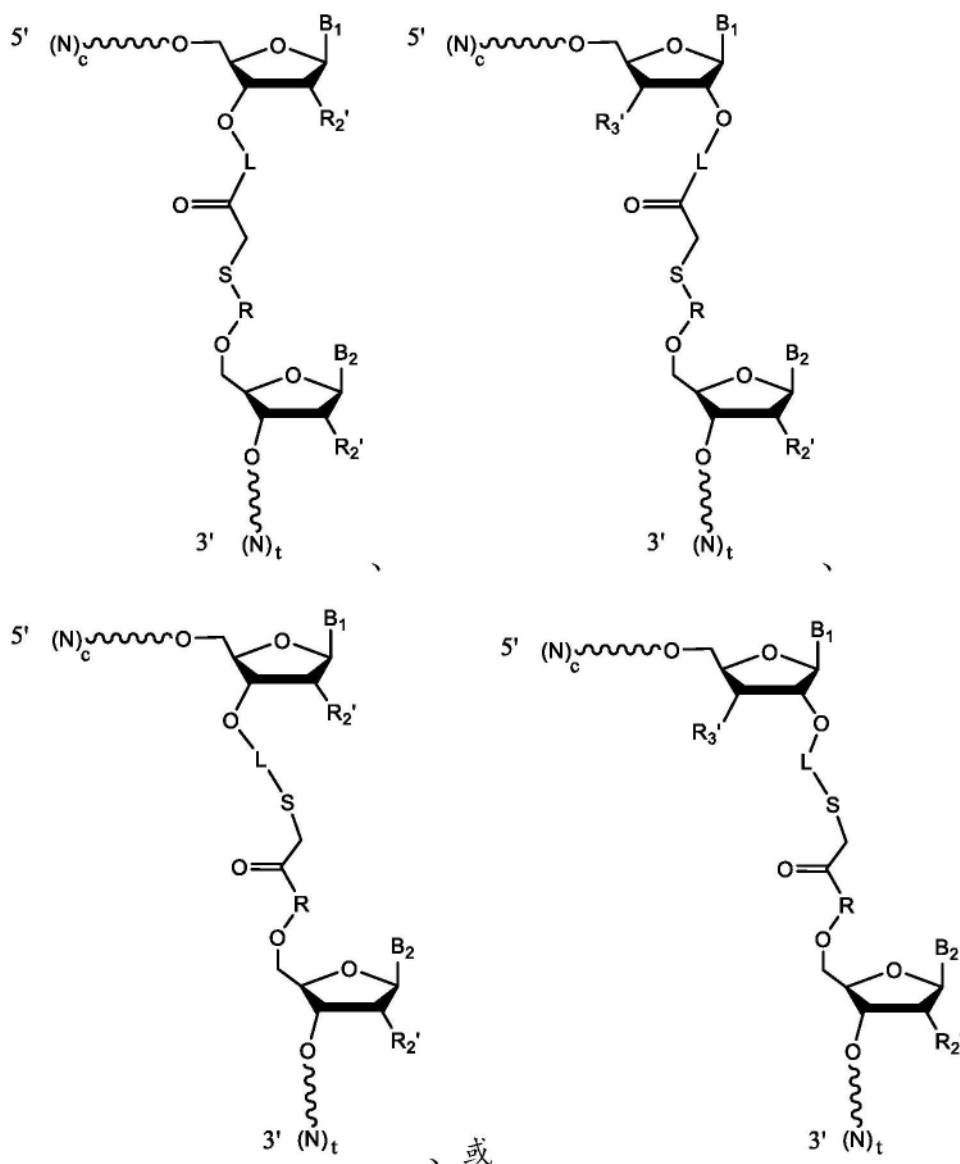
$R_2'$ 和 $R_3'$ 各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$ 或O- $R'$ ,其中每个 $R'$ 独立地是保护基团或烷基,其中所述烷基可任选地被取代;

L和R各自独立地是非核苷酸接头;并且

$B_1$ 和 $B_2$ 各自独立地是核碱基。

2. 根据权利要求1所述的指导分子,其中 $R_2'$ 选自由H、氟和O- $R'$ 组成的组,其中 $R'$ 是保护基团或任选地被取代的烷基。

3. 根据权利要求1或2所述的指导分子,其中所述指导分子具有下式:



其中：

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接；

(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区；

c是20或更大的整数；

t是20或更大的整数；并且

每个 独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

4. 根据权利要求3所述的指导分子,其中(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核糖核苷酸残基或糖修饰的核糖核苷酸残基。

5. 根据权利要求3或4所述的指导分子,其中(N)<sub>c</sub>或(N)<sub>t</sub>包含一个或多个脱氧核糖核苷酸残基。

6. 根据权利要求3-5中任一项所述的指导分子,其中(N)<sub>c</sub>或(N)<sub>t</sub>包含一个或多个2'-O-

甲基修饰的核糖核苷酸残基。

7. 根据权利要求3-6中任一项所述的指导分子, 其中  $(N)_c$  的5'末端的三个核苷酸中的每一个和/或  $(N)_t$  的3'末端的三个核苷酸中的每一个包含2'-O-甲基修饰的核糖核苷酸残基, 所述2'-O-甲基修饰的核糖核苷酸残基通过硫代磷酸酯键合与其相邻的一个或多个核苷酸连接。

8. 根据权利要求1-7中任一项所述的指导分子, 其中L和R各自独立地是非核苷酸接头, 所述非核苷酸接头包含选自下组的部分, 该组由以下组成: 聚乙烯、聚丙烯、聚乙二醇和聚丙二醇。

9. 根据权利要求3-8中任一项所述的指导分子, 其中所述指导分子用于II型CRISPR系统, 并且  $(N)_c$  包括5'区, 所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

10. 根据权利要求3-9中任一项所述的指导分子, 其中  $(N)_t$  包含3'区, 所述3'区包含一个或多个茎-环结构。

## 合成的指导分子、组合物和与其相关的方法

[0001] 本申请是申请号为201780085248.4、申请日为2017年12月29日、发明名称为“合成的指导分子、组合物和与其相关的方法”的中国发明专利申请的分案申请,原申请为国际申请号为PCT/US2017/069019的国际申请的中国国家阶段申请,该国际申请要求于2016年12月30日提交的美国申请号62/441,046和2017年4月28日提交的美国申请号62/492,001的优先权,其中每个的公开通过引用以其全部内容并入本文。

### 技术领域

[0002] 本披露涉及用于编辑靶核酸序列或调节靶核酸序列的表达的CRISPR/Cas相关方法和组分。更具体地,本披露涉及合成的指导分子和相关系统、方法和组合物。

### 背景技术

[0003] CRISPR (成簇规律间隔短回文重复序列) 在细菌和古细菌中作为适应性免疫系统进化,以防御病毒攻击。在暴露于病毒后,病毒DNA的短区段被整合进CRISPR基因座中。RNA是从包括病毒序列的CRISPR基因座的一部分转录。该RNA含有与病毒基因组互补的序列,介导将RNA指导的核酸酶蛋白如Cas9或Cpf1靶向病毒基因组中的靶序列。RNA指导的核酸酶反过来切割并因此使病毒靶标沉默。

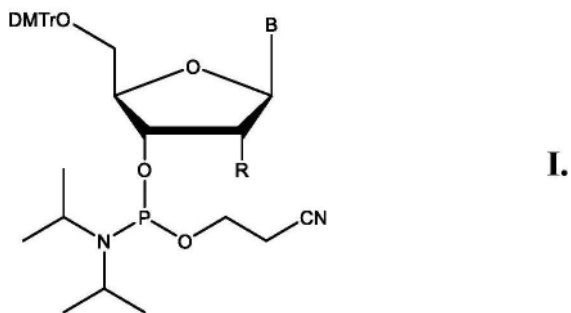
[0004] 最近,CRISPR系统已经适用于真核细胞中的基因组编辑。这些系统通常包括蛋白质组分(RNA指导的核酸酶)和核酸组分(通常称为指导分子,指导RNA或“gRNA”)。这两种组分形成复合物,其与由系统的两种组分识别或互补的特定靶DNA序列相互作用,并任选地编辑或改变靶序列,例如通过位点特异性DNA切割的方式。靶序列的编辑或改变还可涉及细胞DNA修复机制的募集,例如非同源末端连接(NHEJ)或同源定向修复(HDR)。

[0005] 作为治疗遗传疾病的手段的CRISPR系统的价值已被广泛认可,但是基于这些系统的治疗必须解决某些技术挑战以实现广泛的临床应用。除此之外,还需要高质量CRISPR系统组分的具有成本效益且直接的商业规模的合成。

[0006] 例如,大多数指导分子目前通过以下两种方法之一合成:体外转录(IVT)和化学合成。IVT通常涉及通过细菌RNA聚合酶如T7聚合酶的方式从DNA模板转录RNA。目前,根据美国 and 国外监管机构要求的良好生产规范(GMP)标准对指导分子的IVT制造可能成本高且规模有限。此外,IVT合成可能不适合所有指导RNA序列:T7聚合酶倾向于比用另外的5' 碱基引发的序列更有效地转录用5' 鸟嘌呤引发的序列,并且可以识别茎-环结构,然后是聚尿嘧啶束,这些结构存在于某些指导分子中,作为信号终止转录,导致截短的指导分子转录物。

[0007] 另一方面,化学合成是便宜的,并且可以容易地获得较短寡核苷酸(例如,长度小于100个核苷酸)的GMP产生。化学合成方法在整个文献中例如由Beaucage和Carruthers, Curr Protoc Nucleic Acid Chem[核酸化学实验方案].2001年5月;第3章:单元3.3 (Beaucage&Carruthers)描述,其全部内容通过引用并入本文并用于此处的所有目的。这些方法通常包括逐步添加反应性核苷酸单体,直至达到所需长度的寡核苷酸序列。在最常用的合成方案(例如亚磷酰胺方法)中,将单体添加到寡核苷酸的5' 末端。这些单体通常是3'

官能化的(例如用亚磷酰胺)并且包括5' 保护基团(例如4,4'-二甲氧基三苯甲基),例如根据式I,如下:



在式I中,DMTr是4,4'-二甲氧基三苯甲基,R是与寡核苷酸合成条件相容的基团,其非限制性实例包括H、F、O-烷基或受保护的羟基,并且B是任何合适的核碱基。(Beaucage& Carruthers)。使用5' 受保护的单体需要在每轮添加后进行脱保护步骤,其中除去5' 保护基以留下羟基。

[0008] 无论使用何种化学物质,5' 残基的逐步添加都不会定量地发生;一些寡核苷酸将“错过”一些残基的添加。这产生合成产物,所述合成产物包括所需的寡核苷酸,但被缺少各种残基的较短寡核苷酸(称为“n-1种类”,尽管它们可能包括n-2、n-3等,以及其他截短或缺失种类)污染。为了使n-1种类的污染最小化,许多化学合成方案在逐步添加步骤和脱保护步骤之间包括“加帽”反应。在加帽反应中,将非反应性部分添加到那些未被5' 保护基团终止的寡核苷酸的5' 末端;该非反应性部分阻止了单体进一步添加至寡核苷酸,并且在合成长度为约60或70个碱基的寡核苷酸期间有效地将n-1污染降低至可接受的低水平。然而,加帽反应也不是定量的,并且可能无法有效防止较长寡核苷酸如单分子指导RNA中的n-1污染。另一方面,在偶联反应期间存在DMT保护丧失的情况,这导致更长的寡核苷酸(称为“n+1种类”,尽管它们可包括n+2、n+3等)。被n-1种类和/或n+1种类污染的单分子指导RNA可能与通过其他方法制备的全长指导RNA的行为方式不同,可能使合成的指导RNA在治疗中的使用复杂化。

## 发明内容

[0009] 本披露通过提供除其他外的用于合成单分子指导分子的方法(所述方法涉及交联两个或更多个预退火的指导片段),解决了对具有最小n-1和/或n+1种类、截短种类和其他污染物的高纯度单分子指导分子的成本有效且直接的化学合成的需求。在一些实施例中,本文提供的单分子指导分子在5' 末端具有改善的序列保真度,减少了不希望的脱靶编辑。本文还提供了包含全长单分子指导分子(其基本上不含n-1和/或n+1污染物)或基本上由其组成的组合物。

[0010] 本公开的某些方面包括以下认识:当指导片段是同型多官能的(例如,同型双官能的)时,指导片段的预退火可能特别有用,例如本文所述的基于脲的交联方法中使用的胺官能化片段。实际上,将同型多官能指导片段预退火成异二聚体可以减少不期望的同型二聚体的形成。因此,本发明还提供了包含或基本上由全长单分子指导分子组成的组合物,所述全长单分子指导分子基本上不含副产物(例如,同型二聚体)。

[0011] 在一个方面,本披露涉及合成用于CRISPR系统的单分子指导分子的方法,所述方

法包括以下步骤:

使第一寡核苷酸和第二寡核苷酸退火以在所述第一寡核苷酸的3'区和所述第二寡核苷酸的5'区之间形成双链体,其中所述第一寡核苷酸包含第一反应性基团,所述第一反应性基团是2'反应性基团和3'反应性基团中的至少一个,并且其中所述第二寡核苷酸包含第二反应性基团,所述第二反应性基团是5'反应性基团;和

通过所述第一和第二反应性基团缀合所述退火的第一和第二寡核苷酸以形成单分子指导RNA分子,其包括连接所述第一和第二寡核苷酸的共价键。

[0012] 在一个方面,本披露涉及用于CRISPR系统的单分子指导分子。在一些实施例中,本文提供的单分子指导分子用于II型CRISPR系统。

[0013] 在一些实施例中,第一寡核苷酸的5'区包含与靶序列(例如,真核基因内的靶序列)内的靶结构域完全或部分互补的靶向结构域。

[0014] 在一些实施例中,第二寡核苷酸的3'区包含一个或多个茎-环结构。

[0015] 在一些实施例中,本文提供的单分子指导分子能够与Cas9分子相互作用并介导Cas9/指导分子复合物的形成。

[0016] 在一些实施例中,本文提供的单分子指导分子是与Cas9或RNA指导的核酸酶形成复合物。

[0017] 在一些实施例中,本文提供的单分子指导分子从5'至3'包含:

第一指导分子片段,其包含:

靶向结构域序列;

第一下茎序列;

第一凸起序列;

第一上茎序列;

非核苷酸化学键合;和

第二指导分子片段,其包含:

第二上茎序列;

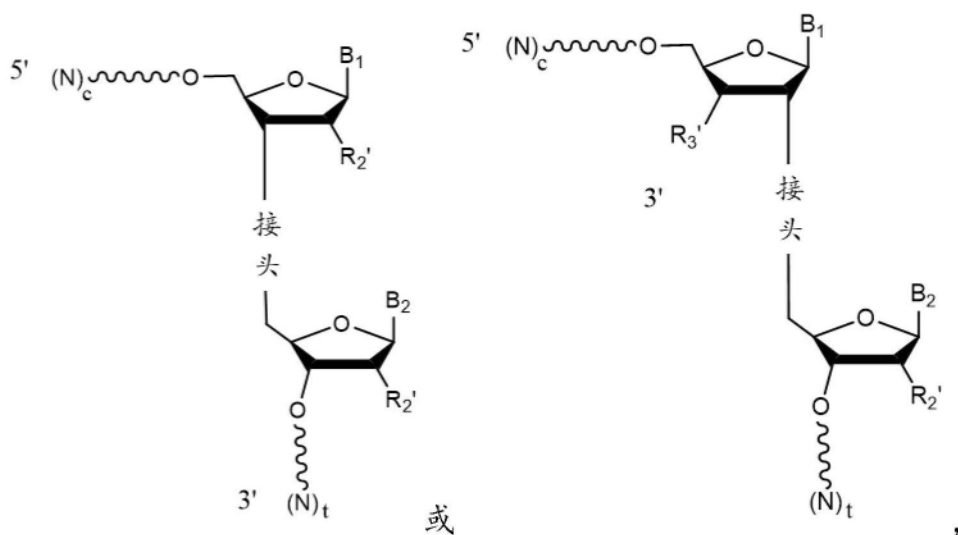
第二凸起序列;和

第二下茎序列,

其中(a)所述第一下茎序列中的至少一个核苷酸与所述第二下茎序列中的核苷酸碱基配对,和(b)所述第一上茎序列中的至少一个核苷酸与所述第二上茎序列中的核苷酸碱基配对。

[0018] 在一些实施例中,单分子指导分子不包括第一和第二上茎序列之间的四环序列。在一些实施例中,第一和/或第二上茎序列包含数目从4至22(包括端值)的核苷酸。

[0019] 在一些实施例中,单分子指导分子具有下式:



其中  $(N)_c$  和  $(N)_t$  中的每个  $N$  独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包括与  $(N)_t$  的 5' 区互补或部分互补并形成双链体的 3' 区;


$c$  是 20 或更大的整数;

$t$  是 20 或更大的整数;

接头是非核苷酸化学键合;

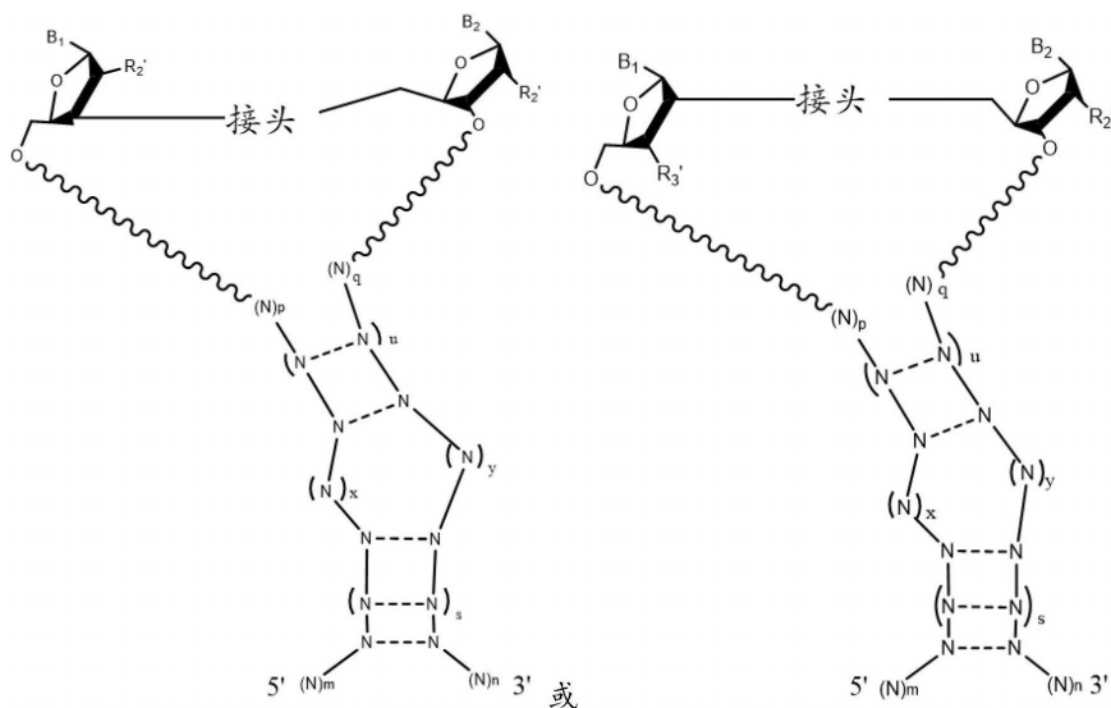
$B_1$  和  $B_2$  各自独立地是核碱基;

$R_2'$  和  $R_3'$  各自独立地是 H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$  或 O- $R'$ , 其中每个  $R'$  独立地是保护基团或烷基, 其中所述烷基可任选地被取代; 和

每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合。

[0020] 在一些实施例中,  $(N)_c$  包含 3' 区, 所述 3' 区包含来自 II 型 CRISPR 系统的重复的至少一部分。在一些实施例中,  $(N)_c$  包含 3' 区, 所述 3' 区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。在一些实施例中,  $(N)_t$  包含 3' 区, 所述 3' 区包含一个或多个茎-环结构。

[0021] 在一些实施例中, 单分子指导分子具有下式:



其中:

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;

p和q各为0;

u是2到22之间的整数(包括端值);

s是1到10之间的整数(包括端值);

x是1到3之间的整数(包括端值);

y是>x且3到5之间的整数(包括端值);

$m$ 是15或更大的整数;和

$n$ 是30或更大的整数。

[0022] 在一些实施例中,u是2到22之间的整数(包括端值);

s是1到8之间的整数(包括端值):

x是1到3之间的整数(包括端值);

y是>x且3到5之间的整数(包括端值);

m是15到50之间的整数(包括端值);和

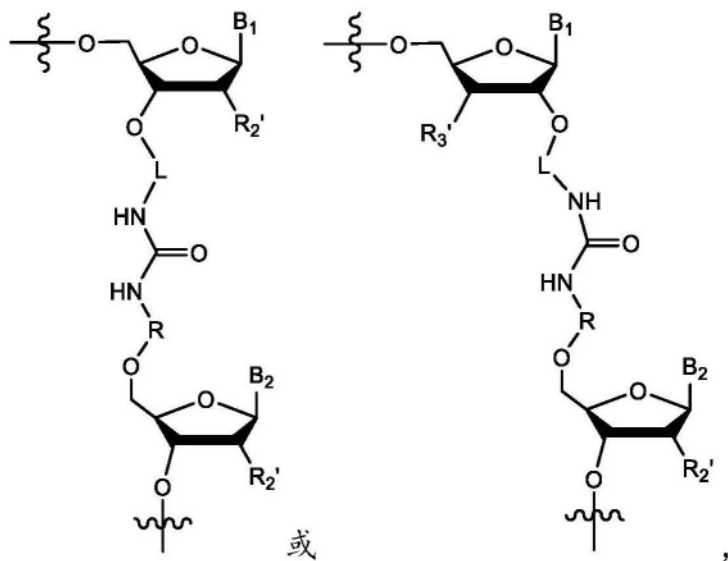
n是30到70之间的整数(包括端值)。

[0023] 在一些实施例中,指导分子不包含四环(p和q各自为0)。在一些实施例中,下茎序列和上茎序列不包含具有超过3个核苷酸的相同序列。在一些实施例中,u是3到22之间的整数(包括端值)。

[0024] 在一些实施例中,第一反应性基团和第二反应性基团各自为氨基,并且缀合步骤包括使第一和第二反应性基团的胺部分与含碳酸酯的双官能交联剂交联以形成脲键合。在

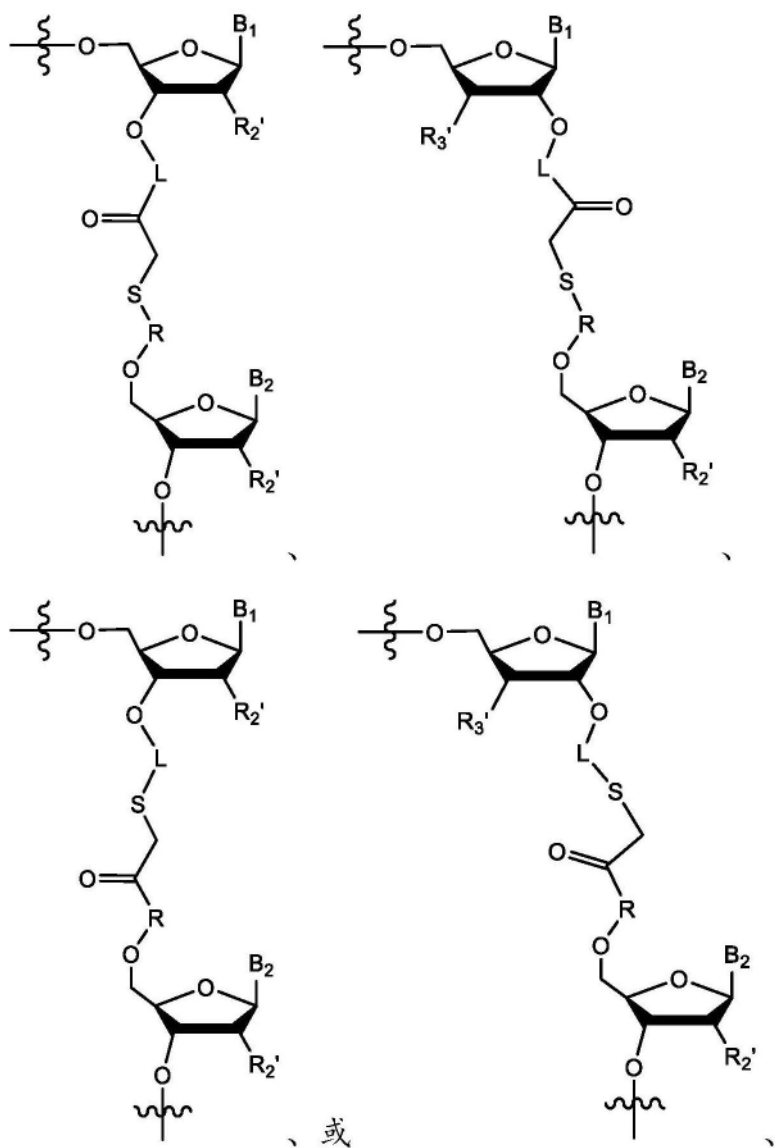
一些实施例中,第一反应性基团和第二反应性基团是溴乙酰基和巯基。在一些实施例中,第一反应性基团和第二反应性基团是磷酸基和羟基。

[0025] 在一些实施例中,单分子指导分子包含下式的化学键合:



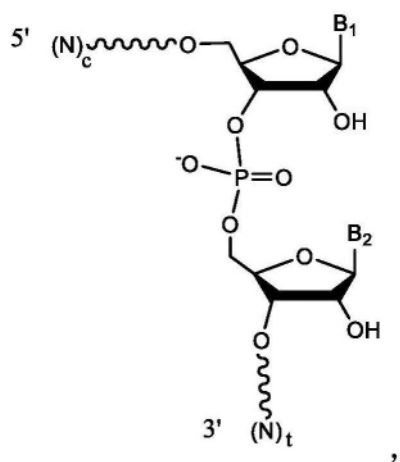
或其药学上可接受的盐,其中L和R各自独立地是非核苷酸化学接头。

[0026] 在一些实施例中,单分子指导分子包含下式的化学键合:

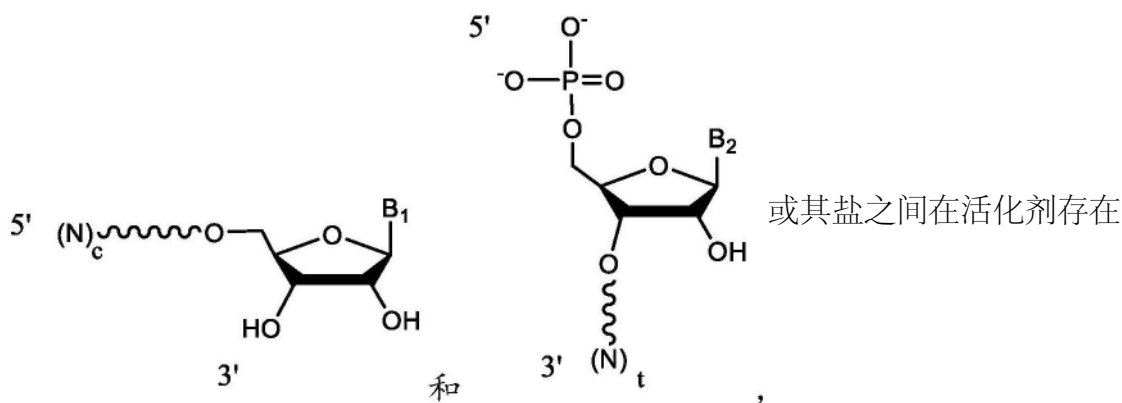


或其药学上可接受的盐,其中L和R各自独立地是非核苷酸化学接头。

[0027] 在一些实施例中,单分子指导分子具有下式:

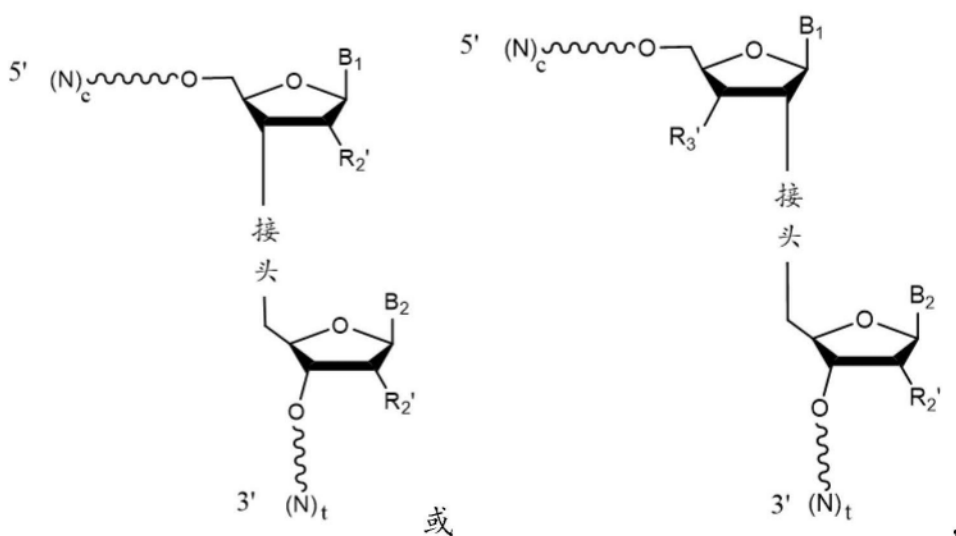


或其药学上可接受的盐,  
并且通过包括以下过程制备:



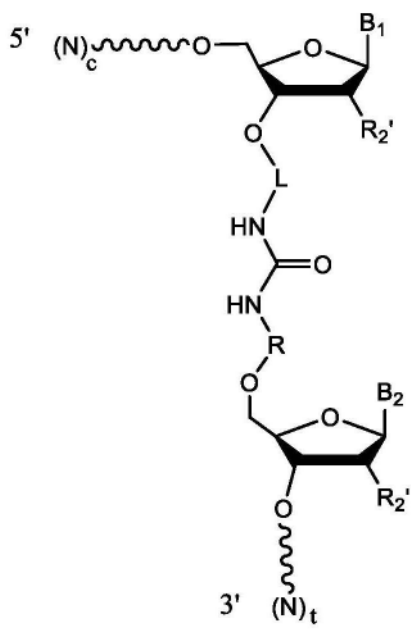
下反应,形成磷酸二酯键合。

[0028] 在一个方面,本披露涉及用于CRISPR系统的指导分子的组合物,其包含具有下式的单分子指导分子或基本上由其组成:



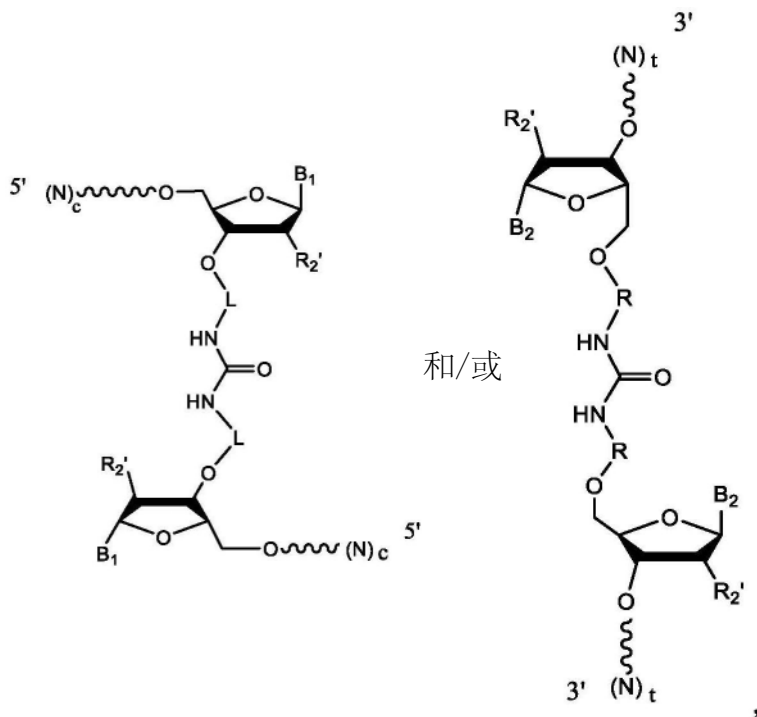
或其药学上可接受的盐。在一些实施例中,相对于参照指导分子序列,少于约10%的指导分子在5'末端包含截短。在一些实施例中,至少约99%的指导分子包含以下5'序列,所述5'序列包含指导分子的与参照指导分子序列的相应5'序列100%相同的核苷酸1-20。

[0029] 在一些实施例中,指导分子的组合物包含具有下式的指导分子或基本上由其组成:



或其药学上可接受的盐，

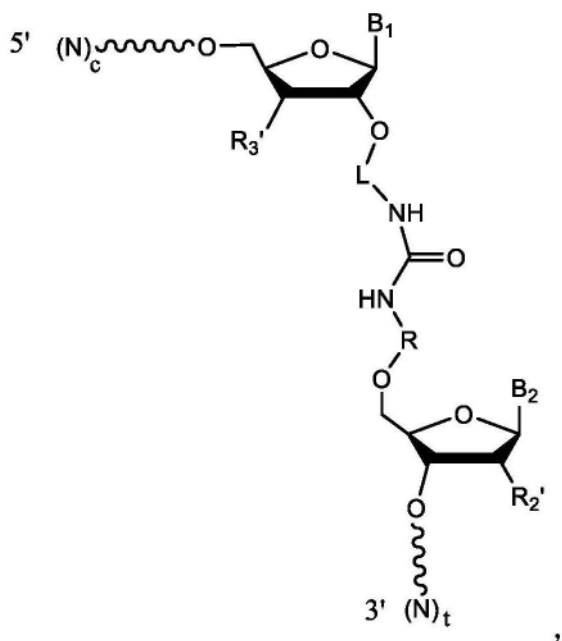
其中所述组合物基本上不含具有下式的分子：



和/或

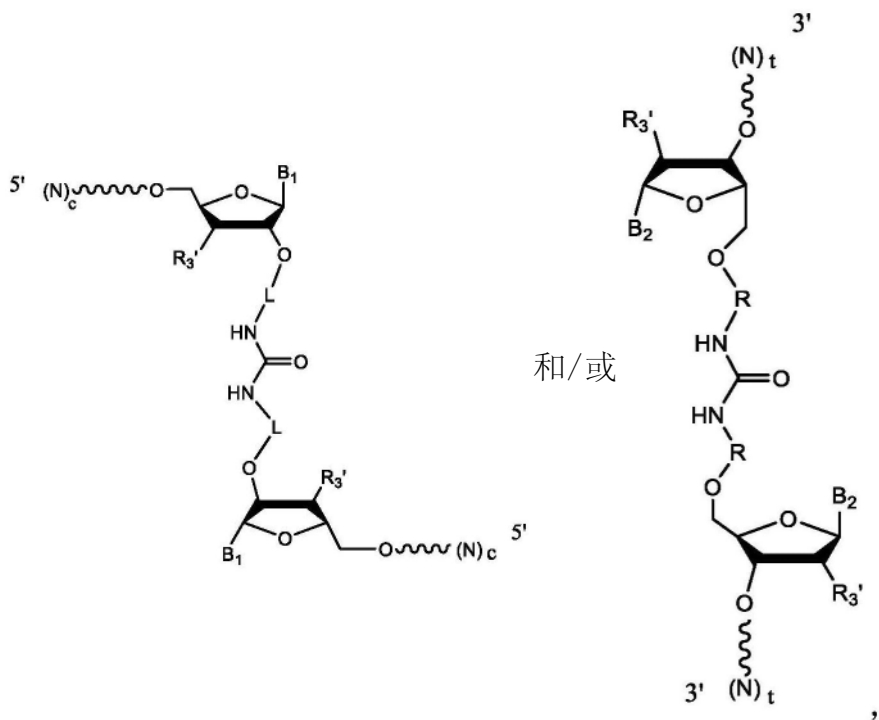
或其药学上可接受的盐。

[0030] 在一些实施例中，组合物包含具有下式的指导分子或基本上由其组成：



或其药学上可接受的盐,

其中所述组合物基本上不含具有下式的分子:

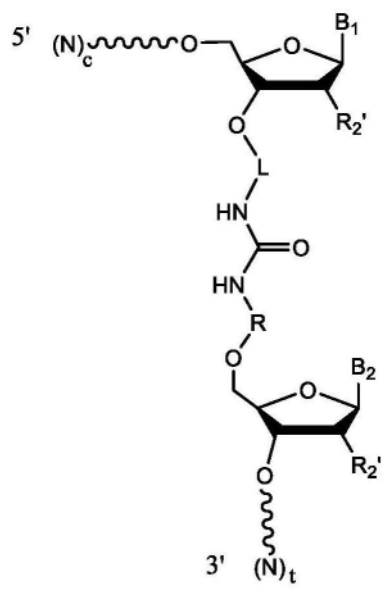


和/或

或其药学上可接

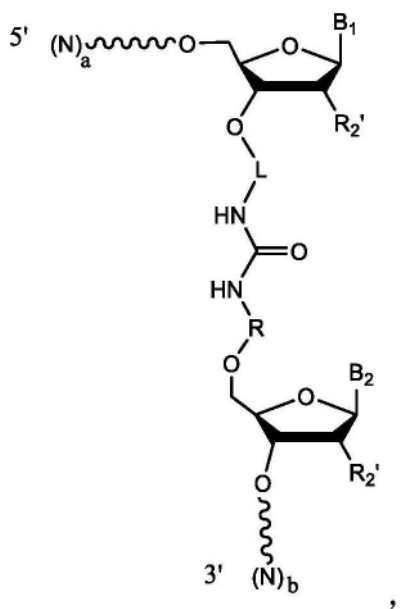
受的盐。

[0031] 在一些实施例中,组合物包含具有下式的指导分子或基本上由其组成:



或其药学上可接受的盐,

其中所述组合物基本上不含具有下式的分子:



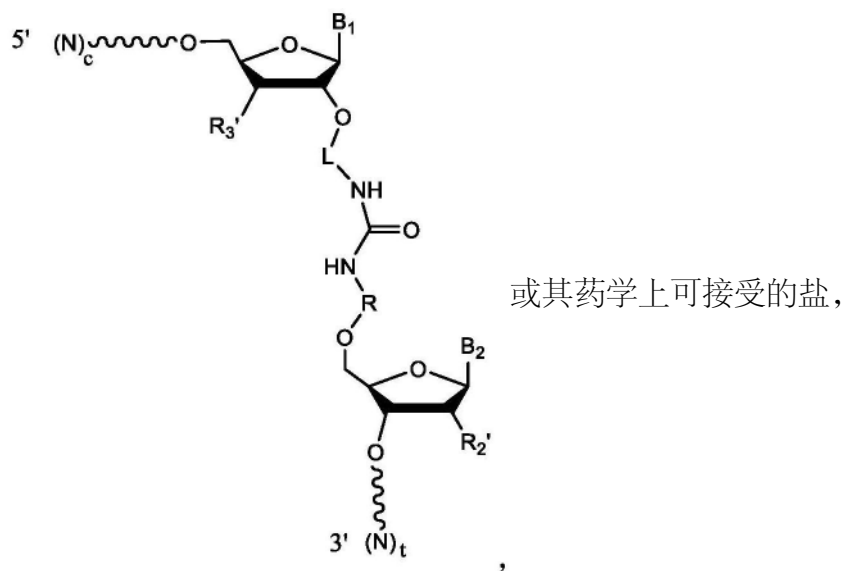
或其药学上可接受的盐,

其中:

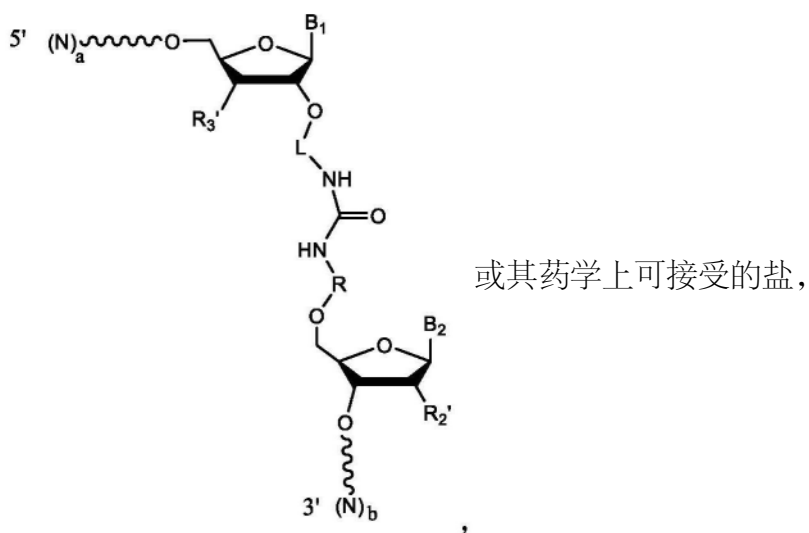
a不等于c;和/或

b不等于t。

[0032] 在一些实施例中,组合物包含具有下式的指导分子或基本上由其组成:



其中所述组合物基本上不含具有下式的分子：

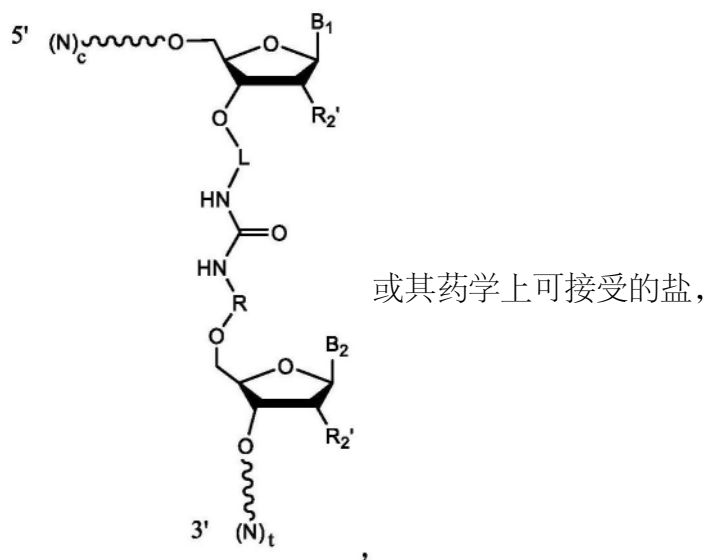


其中：

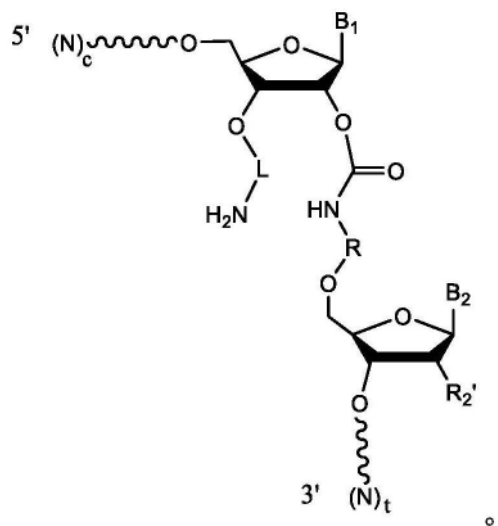
a不等于c；和/或

b不等于t。

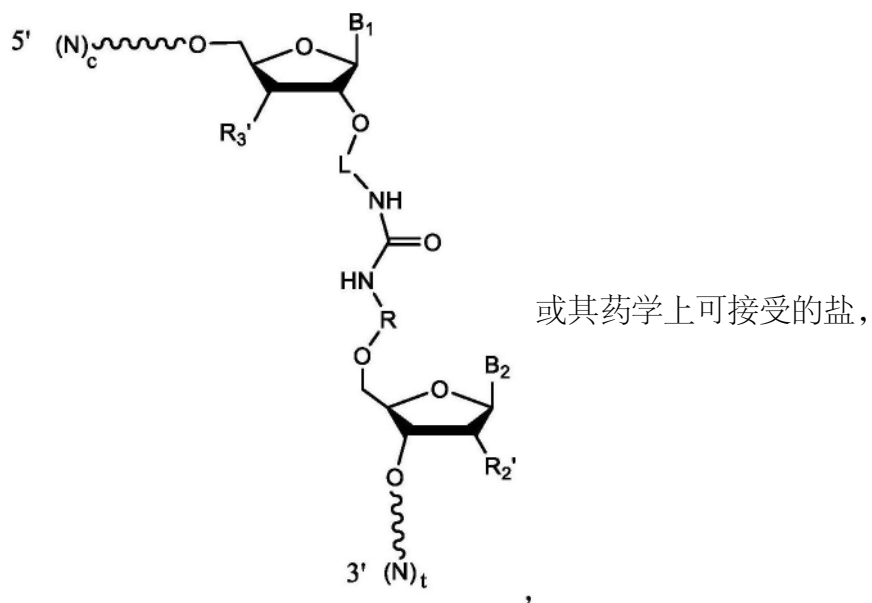
[0033] 在一些实施例中，组合物包含具有下式的指导分子或基本上由其组成：



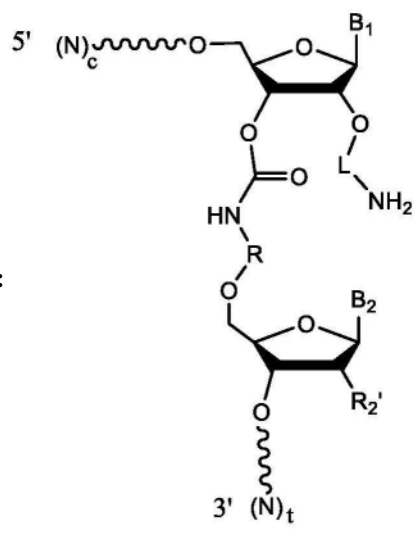
其中所述组合物基本上不含具有下式的分子：



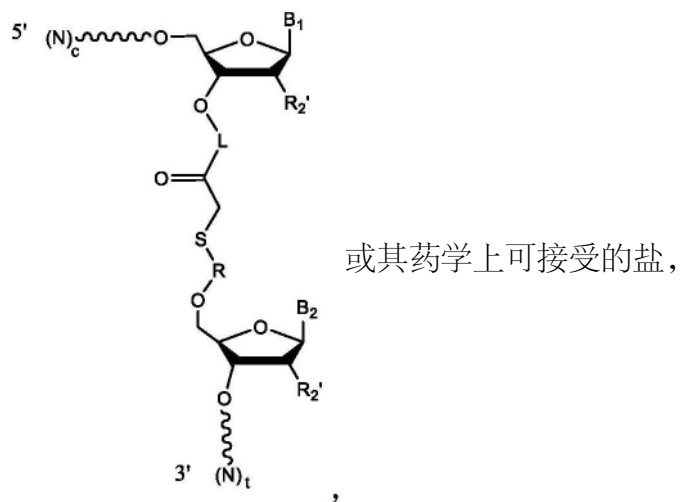
[0034] 在一些实施例中，组合物包含具有下式的指导分子或基本上由其组成：



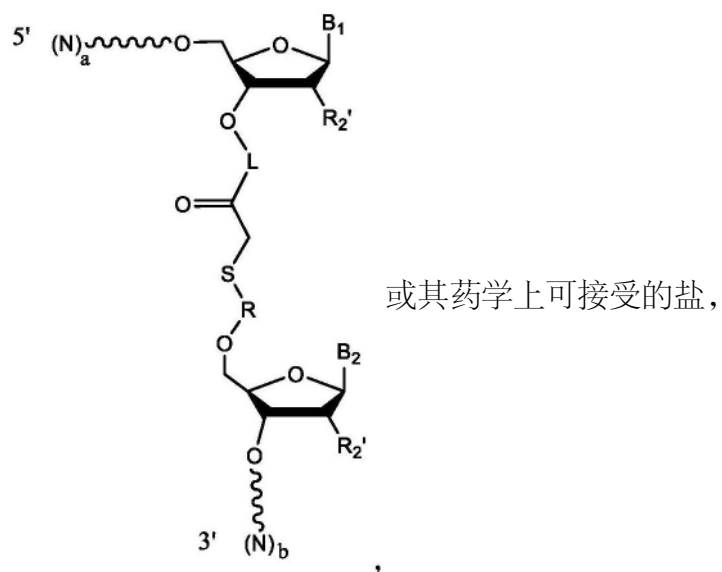
其中所述组合物基本上不含具有下式的分子：



[0035] 在一些实施例中，组合物包含具有下式的指导分子或基本上由其组成：



其中所述组合物基本上不含具有下式的分子：

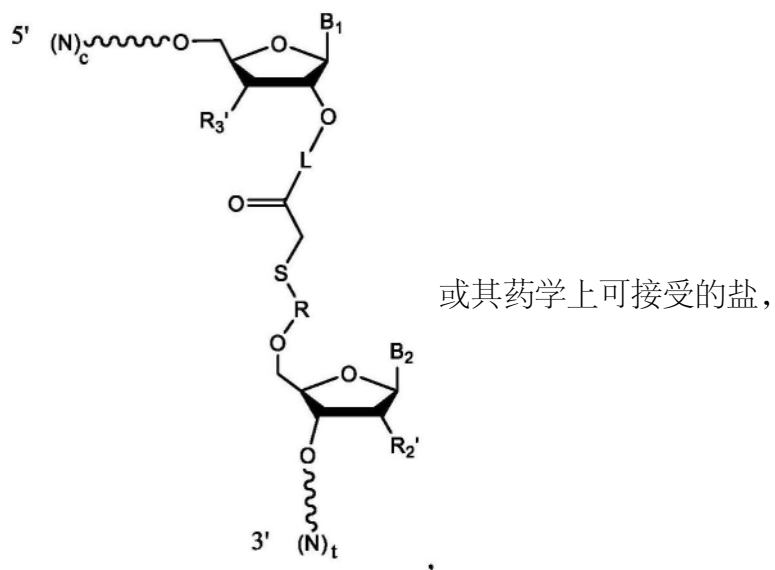


其中：

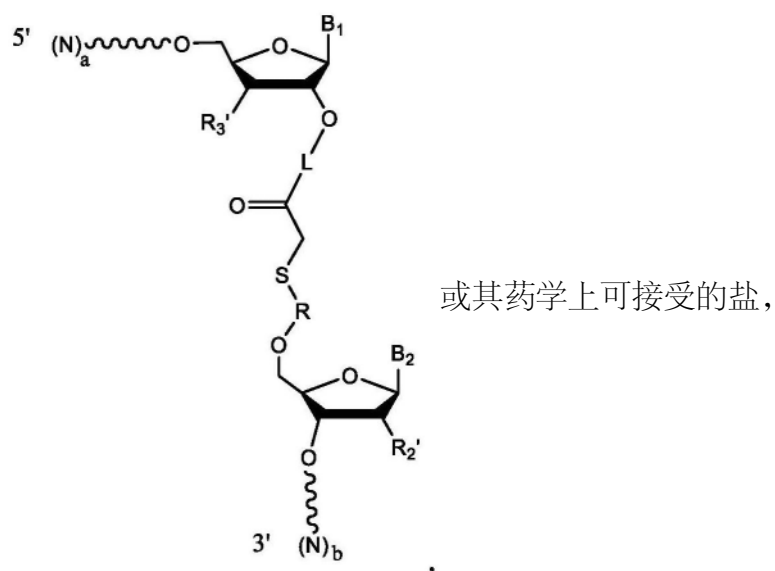
a不等于c；和/或

b不等于t。

[0036] 在一些实施例中,组合物包含具有下式的指导分子或基本上由其组成:



其中所述组合物基本上不含具有下式的分子:

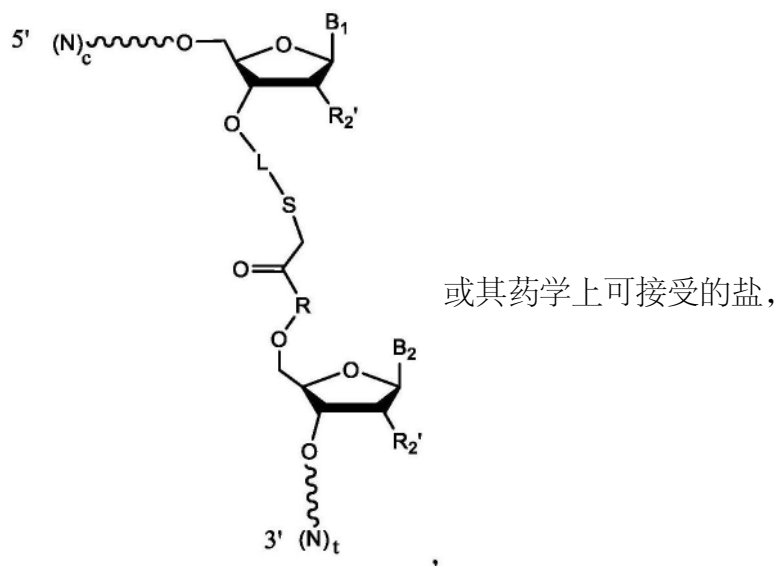


其中:

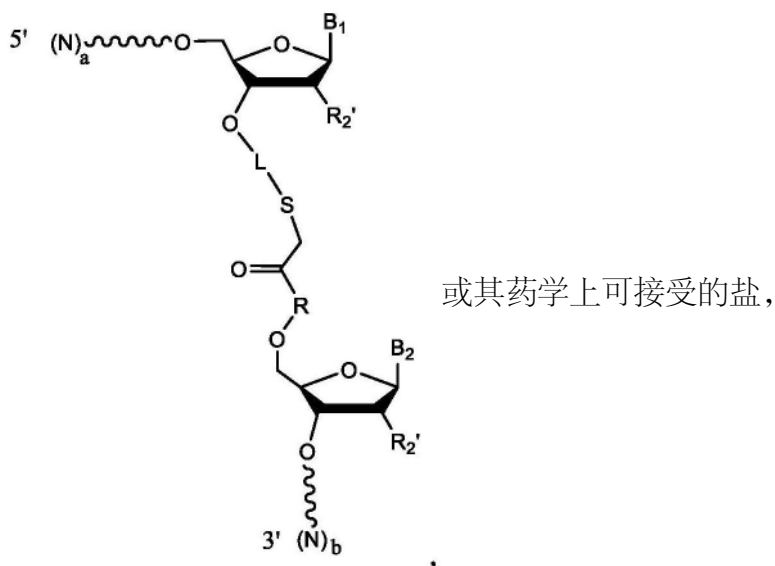
a不等于c;和/或

b不等于t。

[0037] 在一些实施例中,组合物包含具有下式的指导分子或基本上由其组成:



其中所述组合物基本上不含具有下式的分子：

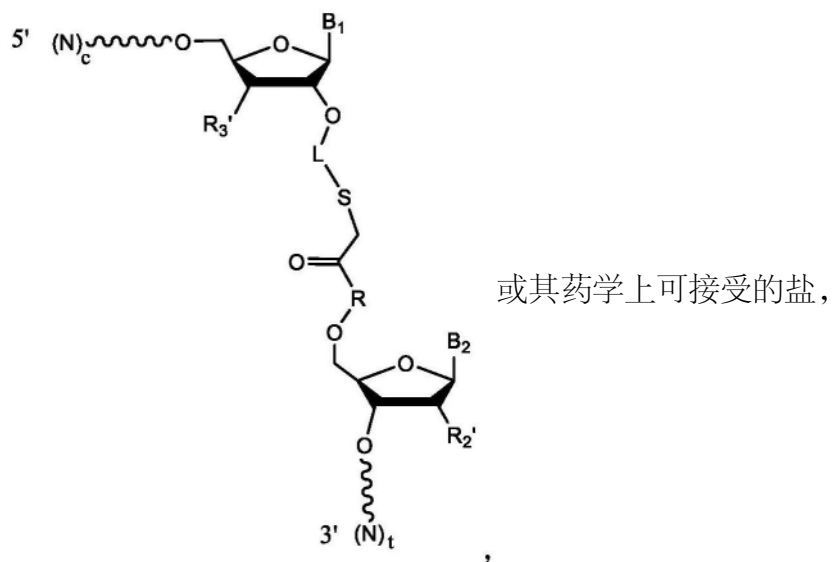


其中：

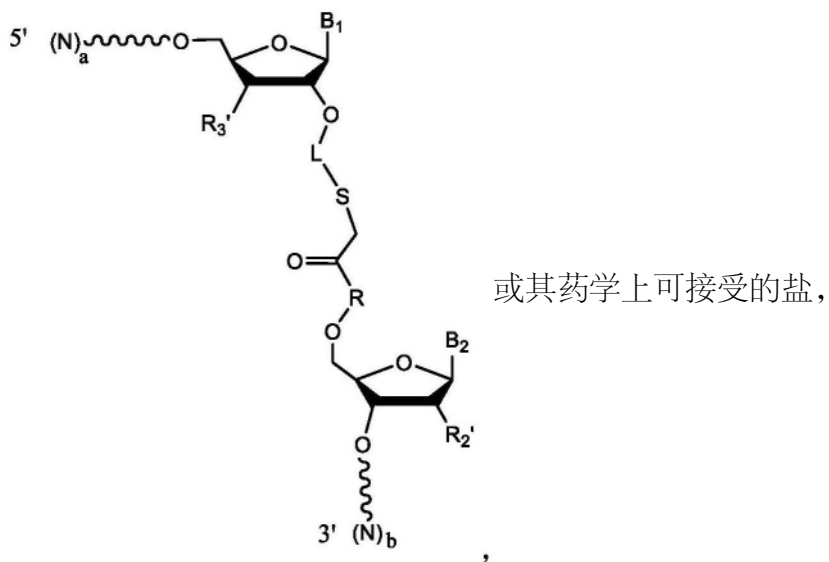
a不等于c；和/或

b不等于t。

[0038] 在一些实施例中，组合物包含具有下式的指导分子或基本上由其组成：



其中所述组合物基本上不含具有下式的分子：



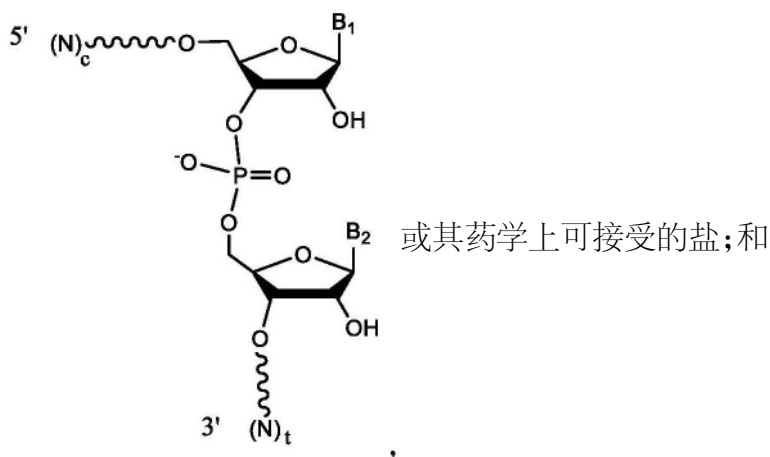
其中：

a不等于c；和/或

b不等于t。

[0039] 在一些实施例中，组合物包含

(a) 用于CRISPR系统的合成的单分子指导分子，其中所述指导分子具有下式：

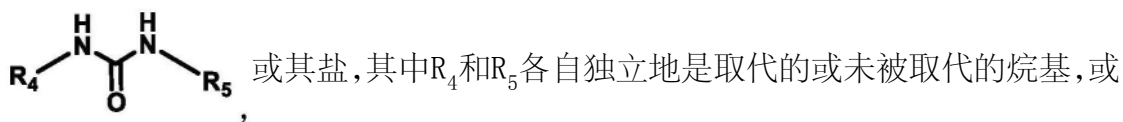


(b) 以下一项或多项:

(i) 碳二亚胺,或其盐;

(ii) 咪唑、氰基咪唑、吡啶和二甲基氨基吡啶,或其盐;和

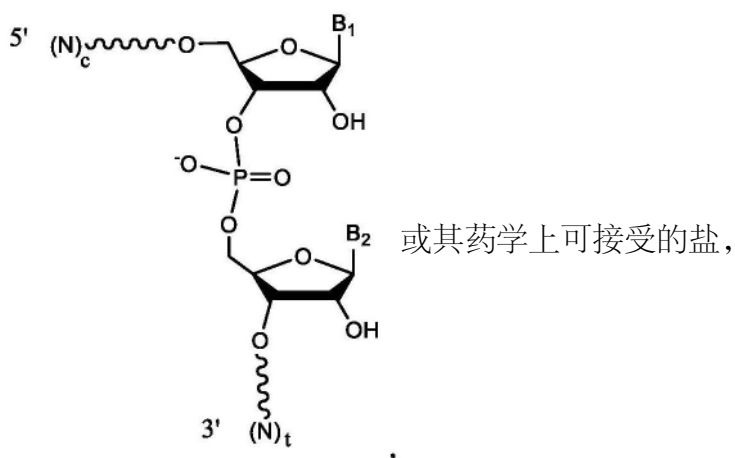
(iii) 具有下式的化合物:



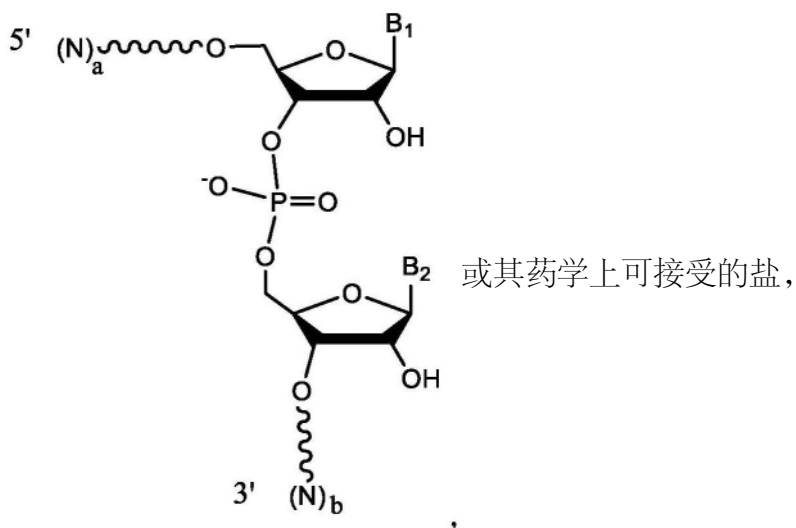
取代的或未被取代的碳环。

[0040] 在一些实施例中,组合物包含

用于CRISPR系统的合成的单分子指导分子,其中所述指导分子具有下式:



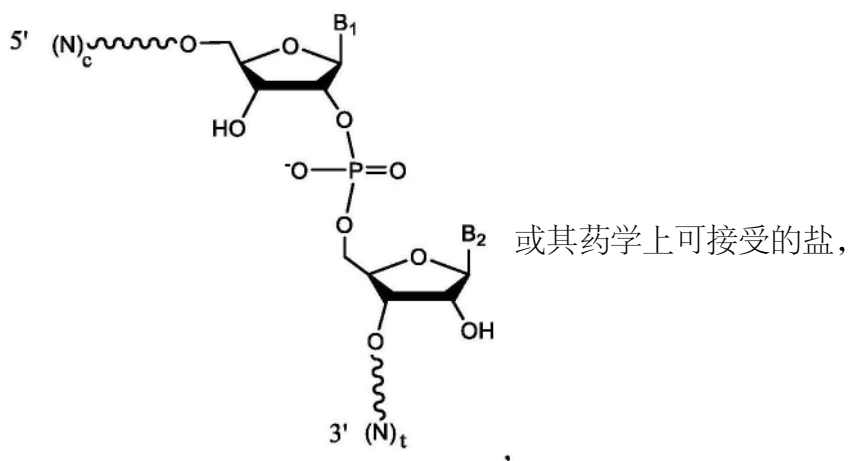
其中所述组合物基本上不含具有下式的分子:



其中:

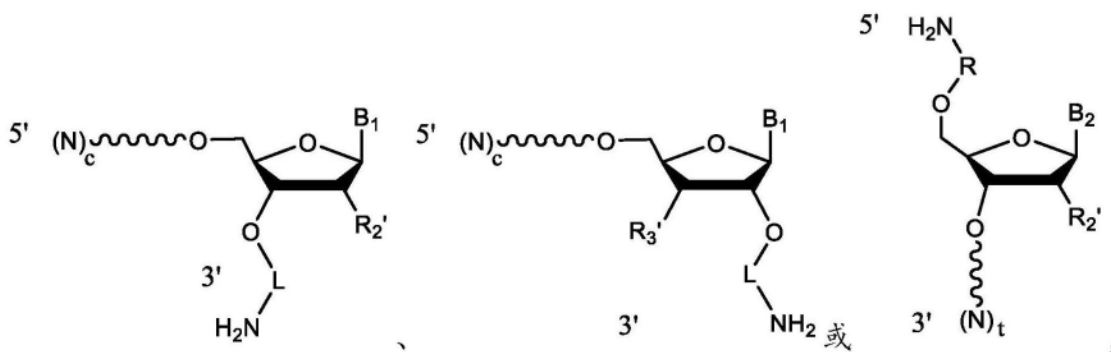
$a+b$ 是 $c+t-k$ ,其中 $k$ 是1-10。

[0041] 在一些实施例中,组合物包含用于CRISPR系统的合成的单分子指导分子或基本上由其组成,其中所述指导分子具有下式:

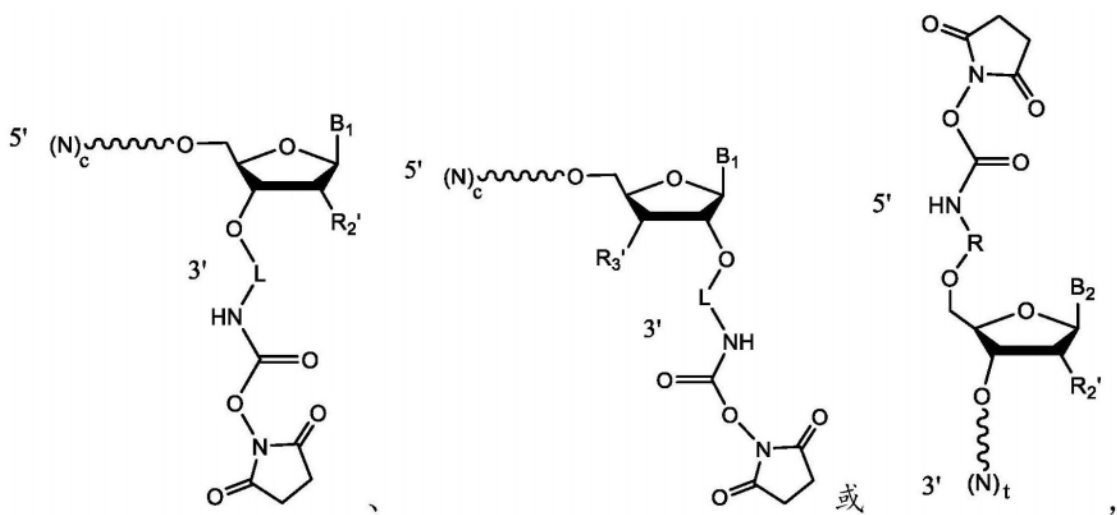


其中式中所示的2'-5'磷酸二酯键合是双链体中在 $(N)_c$ 的3'区和 $(N)_t$ 的5'区之间形成的两个核苷酸之间。

[0042] 在一个方面,本披露涉及用于合成本文提供的单分子指导分子和/或通过本文提供的方法合成单分子指导分子的寡核苷酸。在一些实施例中,寡核苷酸具有下式:

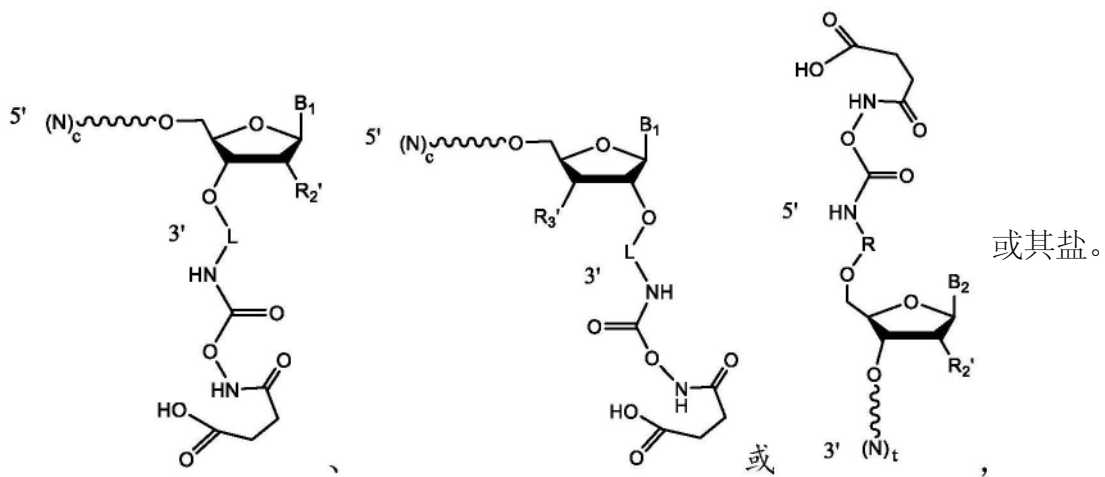


[0043] 在一些实施例中,寡核苷酸具有下式:



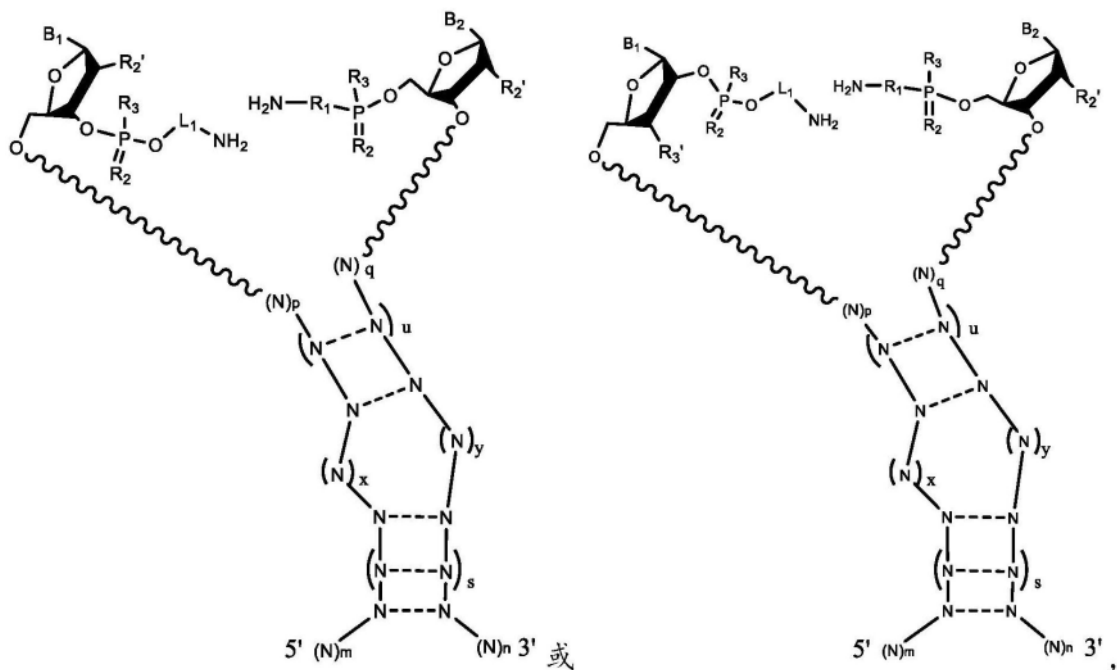
或其盐。

[0044] 在一些实施例中,寡核苷酸具有下式:



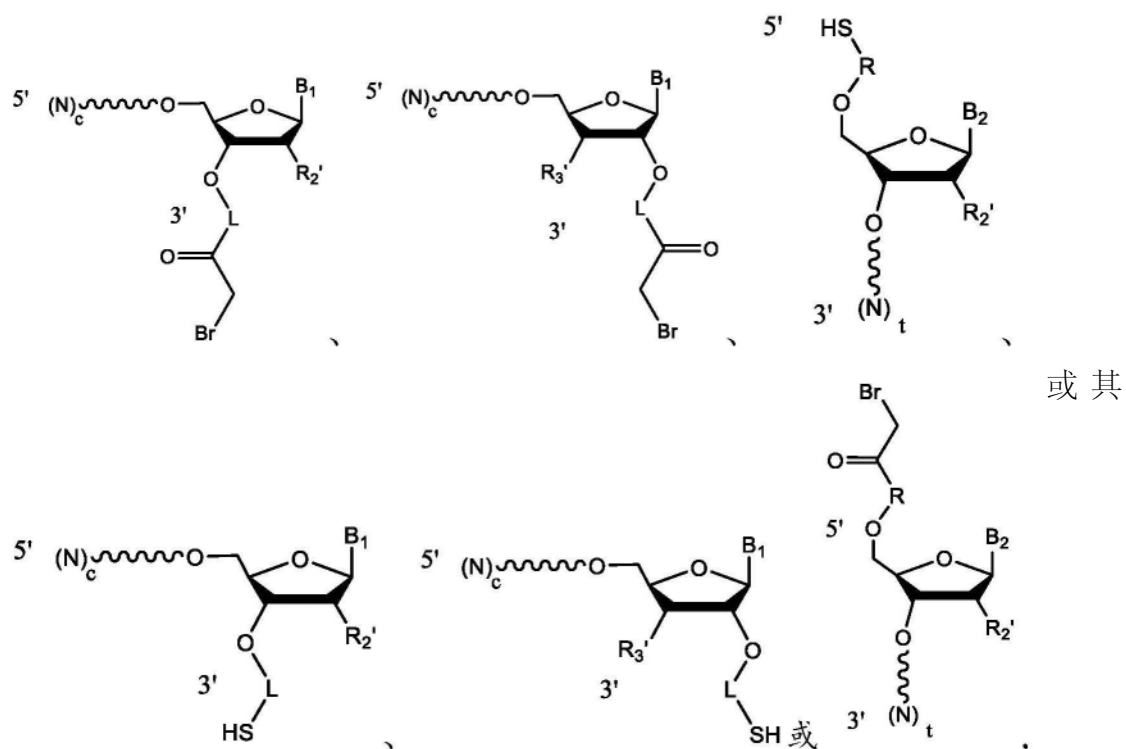
或其盐。

[0045] 在一些实施例中,组合物包含具有下式的退火双链体的寡核苷酸:



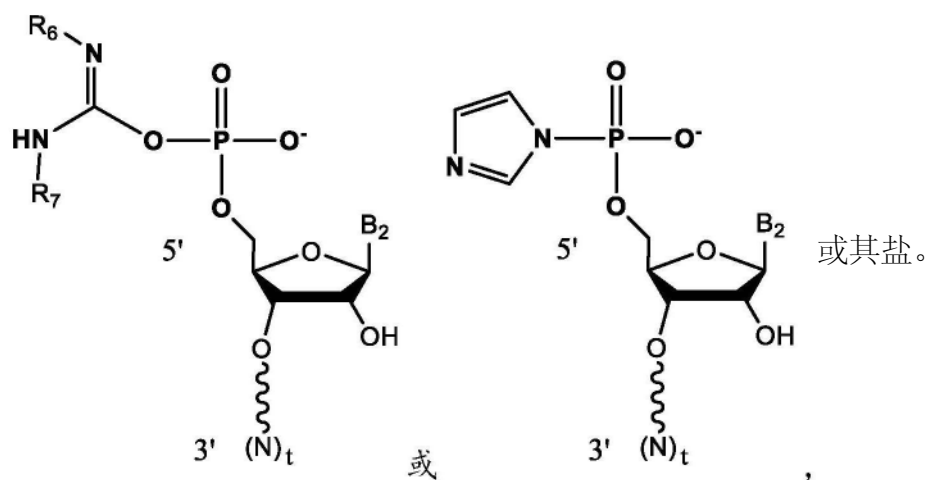
或其盐。

[0046] 在一些实施例中,寡核苷酸具有下式:

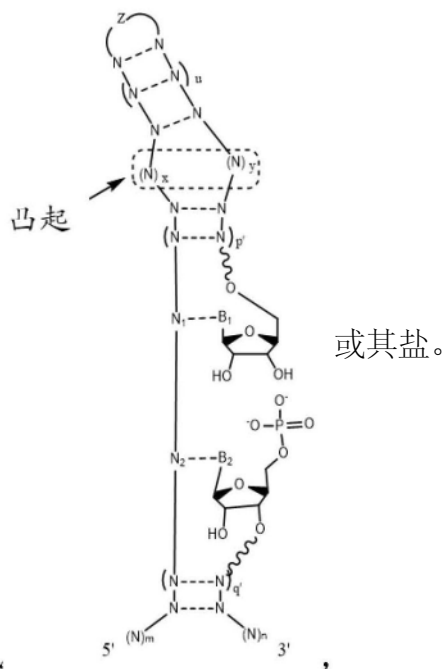
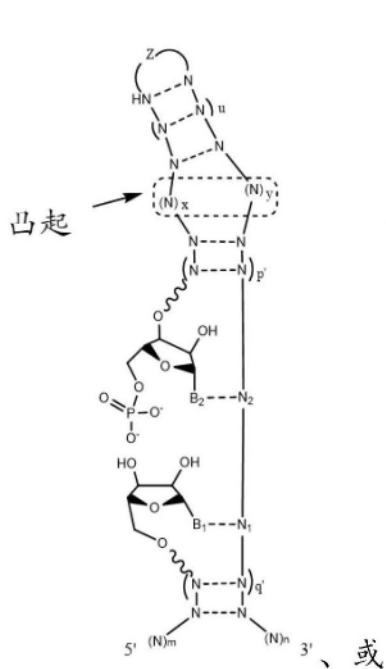
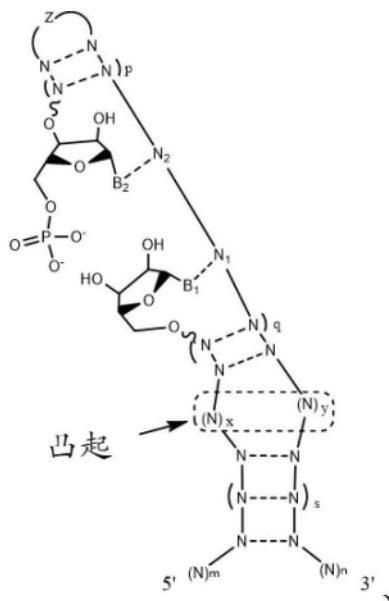
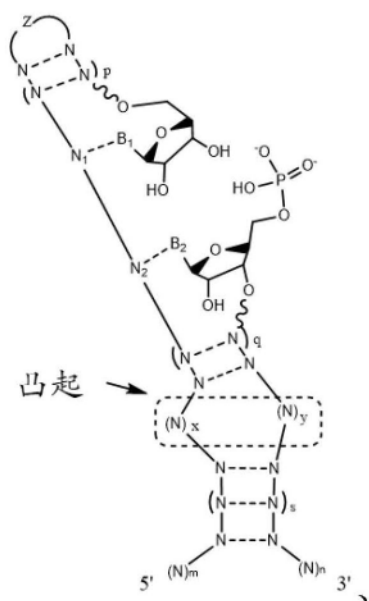


盐。

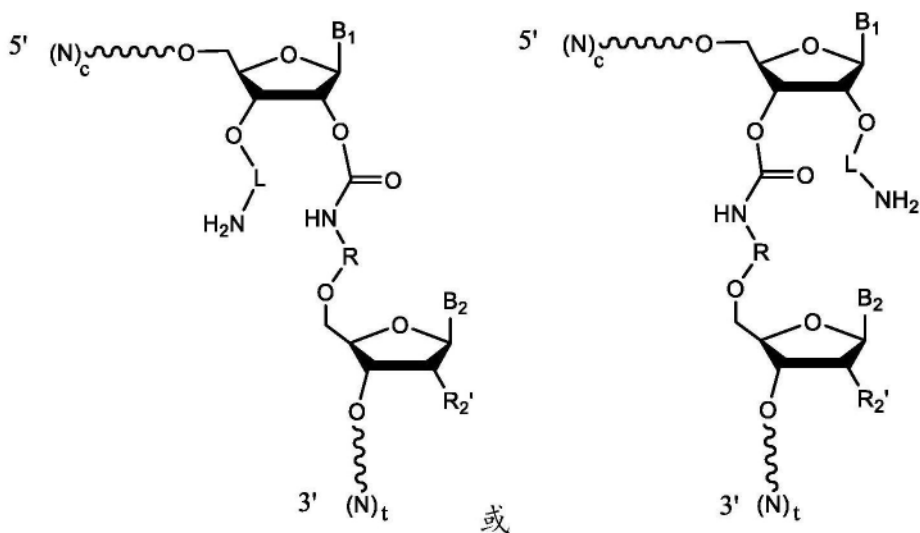
[0047] 在一些实施例中,寡核苷酸具有下式:



[0048] 在一些实施例中,组合物包含具有下式的退火双链体的寡核苷酸:



[0049] 在一个方面,本披露涉及具有下式的化合物:



[0050] 在一个方面,本披露涉及改变细胞或受试者中的核酸的方法,所述方法包括向受试者施用本文提供的指导分子或组合物。

[0051] 在一些实施例中,本文提供的组合物未经历任何纯化步骤。

[0052] 在一些实施例中,本文提供的组合物包含悬浮在溶液中或药学上可接受的载体中的单分子指导RNA分子。

[0053] 在一个方面,本披露涉及包含本文提供的指导分子的基因组编辑系统。在一些实施例中,基因组编辑系统和/或指导分子用于治疗。在一些实施例中,基因组编辑系统和/或指导分子用于生产药物。

#### 附图说明

[0054] 附图旨在提供本披露的某些方面和实施例的说明性和示意性而非综合性的实例。图式并不旨在限制或结合于任何具体理论或模型,并且不一定成比例。不限制前文,可将核酸和多肽描绘为线性序列,或描绘为示意性二维或三维结构;这些描绘旨在具有说明性,而不是限制或束缚于关于其结构的任何具体模型或理论。

[0055] 图1A描绘了根据本披露的某些实施例的示例性交联反应方法。

[0056] 图1B以二维示意图形式描绘了示例性化脓链球菌指导分子突出显示位置(具有星形),在此处根据本披露的各种实施例,第一和第二指导分子片段交联在一起。

[0057] 图1C以二维示意图形式描绘了示例性金黄色葡萄球菌指导分子突出显示位置(具有星形),在此处根据本披露的各种实施例,第一和第二指导分子片段交联在一起。

[0058] 图2A描绘了根据本披露的某些实施例的示例性交联反应方法中的步骤。

[0059] 图2B描绘了根据本披露的某些实施例的示例性交联反应方法中的步骤。

[0060] 图2C描绘了使用来自图2A和2B的反应产物的示例性交联反应方法中的另外步骤。

[0061] 图3A描绘了根据本披露的某些实施例的示例性交联反应方法。

[0062] 图3B描绘了根据本披露的某些实施例的示例性交联反应方法中的步骤。

[0063] 图3C以二维示意图形式描绘了示例性化脓链球菌指导分子突出显示位置,在此处根据本披露的各种实施例,第一和第二指导分子片段交联在一起。

[0064] 图3D以二维示意图形式描绘了示例性金黄色葡萄球菌指导分子突出显示位置,在

此处根据本披露的各种实施例,第一和第二指导分子片段交联在一起。

[0065] 图4显示了与通过商业供应商获得的IVT制备的未连接的退火的指导分子片段和指导分子相比,根据本披露的某些实施例的合成的单分子指导分子的DNA切割剂量-应答曲线。如本文所述通过T7E1测定法测定DNA切割。如图所示,缀合的指导分子以剂量依赖性方式支持HEK293细胞中的切割,这与用IVT产生的单分子指导分子或合成的单分子指导分子观察到的一致。应该注意的是,未缀合的退火的指导分子片段支持较低水平的切割,但是以类似的剂量依赖性方式。

[0066] 图5A显示代表性离子色谱,图5B显示根据实例1的方法与脲接头缀合的离子交换纯化的指导分子的解卷积质谱。图5C显示代表性离子色谱,图5D显示商业制备的合成的单分子指导分子的解卷积质谱。评估离子色谱中突出显示的峰的质谱。图5E显示了质谱的扩展版本。商业上制备的合成的单分子指导分子的质谱在左侧(按总质量计,34%纯度),而根据实例1的方法与脲接头缀合的指导分子的质谱在右侧(按总质量计,72%纯度)。

[0067] 图6A显示的图描绘从包含脲键合的合成的单分子指导分子产生的互补DNA(cDNA)的5'末端的每个位置处发生个体碱基和长度变化的频率,并且图6B显示的图描绘从商业制备的合成的单分子指导分子(即没有缀合地制备)产生的cDNA的5'末端的每个位置处发生个体碱基和长度变化的频率。盒围绕指导分子的20bp靶向结构域。图6C显示的图描绘从包含硫醚键合的合成的单分子指导分子产生的cDNA的5'末端的每个位置处发生个体碱基和长度变化的频率。

[0068] 图7A和图7B描绘从包含脲键合的各种合成的单分子指导分子(图7A)产生的cDNA和从商业制备(即,没有缀合地制备)的合成的单分子指导分子(图7B)的5'末端的前41个位置处的内部序列长度变化(+5至-5)。

[0069] 图8A-8H以二维示意图形式描绘了根据本披露的各种实施例的某些示例性指导分子的结构。能够碱基配对的互补碱基由碱基之间的一个(A-U或A-T配对)或两个(G-C)水平线表示。能够进行非沃森-克里克配对的碱基由带有圆的单水平线表示。

[0070] 图9A-9D以二维示意图形式描绘了根据本披露的各种实施例的某些示例性指导分子的结构。能够碱基配对的互补碱基由碱基之间的一个(A-U或A-T配对)或两个(G-C)水平线表示。能够进行非沃森-克里克配对的碱基由带有圆的单水平线表示。

[0071] 图10A-10D以二维示意图形式描绘了根据本披露的各种实施例的某些示例性指导分子的结构。能够碱基配对的互补碱基由碱基之间的一个(A-U或A-T配对)或两个(G-C)水平线表示。能够进行非沃森-克里克配对的碱基由带有圆的单水平线表示。

[0072] 图11显示了CD34+细胞中用包含来自表10的缀合的指导分子的一系列核糖核蛋白复合物进行的DNA切割的图。使用下一代测序技术评估切割以相对于野生型人参照序列定量%插入和缺失(indel)。根据实例1产生的连接的指导分子支持CD34+细胞中的DNA切割。发现%indel随着茎环长度的增加而增加,但是在茎环序列附近掺入U-A交换(参见gRNA 1E、1F和2D)减轻了所述作用。

[0073] 图12A显示了在T1核酸内切酶消化gRNA 1A后的液相色谱-质谱(LC-MS)迹线,并且图12B显示峰的质谱,保留时间为4.50min(A34:G39)。特别是,在保留时间为4.50分钟时检测到含有脲键合A-[UR]-AAUAG(A34:G39)的片段, $m/z=1190.7$ 。

[0074] 图13A显示了脲连接的指导分子的未纯化的组合物的LC-MS数据,其中存在主要产

物(A-2,保留时间为3.25min)和次要产物(A-1,保留时间为3.14min)。我们注意到,图13A中的次要产物(A-1)为了说明的目的而富集,并且通常在根据实例1的方法合成指导分子中以高达10%的产率检测。图13B显示峰A-2的解卷积质谱(保留时间为3.25min),并且图13C显示峰A-1的解卷积质谱(保留时间为3.14min)。通过质谱分析每个峰表明两种产物具有相同的分子量。

[0075] 图14A显示如实例10中所述的化学修饰后的指导分子组合物的LC-MS数据。主要产物(B-1,脲)具有与原始分析相同的保留时间(3.26min),而次要产物(B-2,氨基甲酸酯)的保留时间已经转移至3.86min,与游离胺部分的化学官能化相一致。图14B显示峰B-2的质谱(保留时间为3.86min)。对3.86min( $M+134$ )处的峰的分析表明已经发生了预测的官能化。

[0076] 图15A显示用含有主要产物(脲)和化学修饰的次要产物(氨基甲酸酯)的反应混合物的T1内切核酸酶消化后的片段混合物的LC-MS迹线。在保留时间分别为4.31min和5.77min时检测到脲键合( $G35-[UR]-C36$ )和化学修饰的氨基甲酸酯键合( $G35-[CA+PAA]-C36$ )。图15B显示了在4.31min处峰的质谱,其中 $m/z=532.13$ 被分配给 $[M-2H]^{2-}$ ,并且图15C显示了在5.77min处峰的质谱,其中 $m/z=599.15$ 被分配给 $[M-2H]^{2-}$ 。图15D和图15E显示图15B的 $m/z=532.1$ 的和图15C的 $m/z=599.1$ 的LC/MS-MS碰撞诱导解离(CID)实验。在图15D中,观察到典型的a-d和x-z离子,并且观察到UR键合的从5'末端( $m/z=487.1$ 和 $461.1$ )和3'末端( $m/z=603.1$ 和 $577.1$ )的任一侧的MS/MS片段离子。在图15E中,仅观察到两种产物离子,包括来自氨基甲酸酯键合的5'末端( $m/z=595.2$ )的和来自CA键合的3'末端( $m/z=603.1$ )的MS/MS片段离子。

[0077] 图16A显示与2'-H修饰的5'指导分子片段(上部光谱)反应的粗反应混合物的LC-MS数据,相比于与未修饰形式的相同5'指导分子反应的粗反应混合物(下部光谱)。用2'-H修饰的5'指导分子片段(上部光谱)观察不到氨基甲酸酯副产物形成。相反,用于与未修饰形式的相同5'指导分子片段反应的粗反应混合物(下部光谱)包括主要脲连接产物(A-2)和次要氨基甲酸酯副产物(A-1)的混合物。我们注意到,与实例10不同,氨基甲酸酯副产物没有富集,因此检测到的水平比实例10的图13A中的低得多。图16B显示峰B的解卷积质谱(保留时间3.14min,图16A的上部光谱),并且图16C显示峰A-2的解卷积质谱(保留时间3.45min,图16A的下部光谱)。分析与2'-H修饰的5'指导分子片段(B)的反应的产物得到M-16(与A-2(主要的未经修饰的脲连接产物)相比),如对于其中2'-OH已经被2'-H代替的分子所预期(参见图16B和图16C)。

[0078] 图17A显示了在T1核酸内切酶消化gRNA 1L后的LC-MS迹线,并且图17B显示峰的质谱,保留时间为4.65min(A34:G39)。特别是,在保留时间为4.65min时检测到含有脲键合A-[UR]-AAUAG(A34:G39)的片段, $m/z=1182.7$ 。

## 具体实施方式

### 定义和缩写

[0079] 除非另外指定,否则以下术语中的每一个具有此章节中与其相关的含义。

[0080] 不定冠词“一个”(“a”和“an”)是指至少一个相关名词,并且可与术语“至少一个”和“一个或多个”互换使用。例如,“一个模块”意指至少一个模块,或一个或多个模块。

[0081] 连词“或”和“和/或”可作为非排他析取词互换使用。

[0082] 短语“基本上由……组成”是指所叙述的种类是主要种类,但是其他种类可以以不影响主题组合物的结构、功能或行为的痕量或量存在。例如,基本上由特定种类组成的组合物通常包含90%、95%、96%或更多(按质量或摩尔浓度)的该种类。

[0083] 短语“基本上不含分子”是指所述分子不是所叙述的组合物中的主要组分。例如,基本上不含分子的组合物是指所述组合物中所述分子少于5%、4%、3%、2%、1%、0.5%或0.1%(按质量或摩尔浓度)。分子的量可通过各种分析技术测定,例如如实例中所述。在一些实施例中,本文提供的组合物基本上不含某些分子,其中如通过凝胶电泳所确定分子少于5%、4%、3%、2%、1%、0.5%或0.1%(按质量或摩尔浓度)。在一些实施例中,本文提供的组合物基本上不含某些分子,其中如通过质谱所确定分子少于5%、4%、3%、2%、1%、0.5%或0.1%(按质量或摩尔浓度)。

[0084] “结构域”用于描述蛋白质或核酸的区段。除非另外指明,否则不需要结构域具有任何特定功能特性。

[0085] 术语“互补”是指能够通过氢键形成稳定碱基对的核苷酸对。例如,U与A互补,G与C互补。本领域技术人员将理解,特定的一对互补核苷酸是否通过氢键碱基配对(例如,在指导分子双链体内)相关联可取决于背景(例如,周围的核苷酸和化学键合)和外部的条件(例如,温度和pH)。因此应理解,互补核苷酸不一定通过氢键碱基配对相关联。

[0086] “协变”序列与参考序列的不同之处在于,用互补核苷酸置换参照序列中的一个或多个核苷酸(例如,用As替换一个或多个Us,用Cs替换一个或多个Gs,等等)。当参考包含形成双链体的两个互补序列的区(例如,指导分子的上茎)使用时,术语“协变”包括在参照双链体的两个互补序列之间具有一个或多个核苷酸交换(即,一个或多个A-U交换和/或一个或多个G-C交换)的双链体,如下表1所示:

表1.具有三个核苷酸的序列的协变序列。

A---U	U---A
G---C	G---C
C---G	C---G
A---U	A---U
C---G	G---C
C---G	G---C
U---A	U---A
C---G	G---C
C---G	G---C
A---U	U---A
C---G	C---G
G---C	G---C

[0087] 在一些实施例中,协变序列可以表现出与参考序列基本相同的特定退火反应的能量有利性(例如,在本披露的指导分子的背景下形成双链体)。如本披露中其他地方所述,特定退火反应的能量有利性可以凭经验测量或使用计算模型预测。

[0088] “indel”是核酸序列中的插入和/或缺失。indel可为DNA双链断裂的修复产物,该DNA双链断裂例如通过本披露的基因组编辑系统形成的双链断裂。indel最常在通过“错误倾向”修复路径(例如下文所述的NHEJ路径)修复断裂时形成。

[0089] “基因转变”是指通过并入内源同源序列(例如基因阵列内的同源序列)改变DNA序列。“基因修正”是指通过并入外源同源序列(例如外源单链或双链供体模板DNA)改变DNA序列。基因转变和基因修正是通过HDR路径(例如下文所述的那些)修复DNA双链断裂的产物。

[0090] Indel、基因转变、基因修正和其他基因组编辑结果典型地通过测序(最常通过“新一代(next-gen)”或“边合成边测序(sequencing-by-synthesis)”方法进行,但仍可使用Sanger测序)来评价,并且通过所有测序读段之间所关注位点处的数值变化(例如,±1、±2或更多个碱基)的相对频率来定量。测序用DNA样品可通过本领域中已知的多种方法来制备,并且可包括通过聚合酶链式反应(PCR)扩增所关注位点、捕捉通过双链断裂产生的DNA末端,如在Tsai等人(Nat.Biotechnol.[自然生物技术]34(5):483(2016),通过引用并入本文)中所述的GUIDEseq方法中,或通过本领域熟知的其他手段。基因组编辑结果也可通过原位杂交法(例如FiberComb™系统,由基因组视觉公司(Genomic Vision)(法国巴涅)商品化)和通过本领域中已知的任何其他适宜方法来评价。

[0091] “alt-HDR”、“替代性同源定向修复”或“替代性HDR”可互换使用,是指使用同源核酸(例如,内源同源序列(例如姐妹染色单体)或外源核酸(例如模板核酸))修复DNA损伤的过程。alt-HDR与经典HDR的不同之处在于,该过程利用与经典HDR不同的路径,并且可以被经典HDR介体RAD51和BRCA2抑制。Alt-HDR的不同之处还在于涉及单链或带切口同源核酸模板,而经典HDR通常涉及双链同源模板。

[0092] “经典HDR”、“经典同源定向修复”或“cHDR”是指使用同源核酸(例如,内源同源序列(例如,姐妹染色单体)或外源核酸(例如,模板核酸))修复DNA损伤的过程。当在双链断裂处已有显著切除时,典型HDR通常起作用,形成DNA的至少一个单链部分。在正常细胞中,cHDR通常涉及一系列步骤,诸如断裂的识别、断裂的稳定、切除、单链DNA的稳定、DNA交叉中间体的形成、交叉中间体的拆分以及连接。该过程需要RAD51和BRCA2,并且同源核酸典型地为双链。

[0093] 除非另外指明,否则如本文所用的术语“HDR”涵盖经典HDR和alt-HDR两者。

[0094] “非同源末端连接”或“NHEJ”是指连接介导的修复和/或非模板介导的修复,包括经典NHEJ(cNHEJ)和替代性NHEJ(altNHEJ),替代性NHEJ又包括微同源介导的末端连接(MMEJ)、单链退火(SSA)和合成依赖性微同源介导的末端连接(SD-MMEJ)。

[0095] 在关于分子的修饰(例如核酸或蛋白质)使用时,“替代”或“替代的”不需要方法限制,但仅指示替代实体是存在的。

[0096] “受试者”意指人类或非人类动物。人类受试者可为任何年龄(例如,婴儿、儿童、青年人或成年人),并且可患有疾病,并且可能实际上具有基因改变。可替代地,受试者可为动物,该术语包括但不限于哺乳动物、鸟、鱼、爬行动物、两栖动物,并且更具体地非人灵长类动物、啮齿类动物(例如小鼠、大鼠、仓鼠等)、兔子、豚鼠、犬、猫等。在本披露的某些实施例中,受试者为家畜,例如牛、马、绵羊或山羊。在某些实施例中,受试者是家禽。

[0097] “治疗(Treat)”、“治疗(treating)”和“治疗(treatment)”意指治疗受试者(例如,人类受试者)的疾病,包括以下各项中的一种或多种:抑制疾病,即,阻止或预防其发展或进展;缓解疾病,即,引起疾病状态消退;减轻疾病的一种或多种症状;和治愈疾病。

[0098] “预防(prevent)”、“预防(preventing)”和“预防(prevention)”是指预防哺乳动物(例如人类)的疾病,包括:(a)避免或预先排除疾病;(b)影响朝向疾病的倾向;或(c)预防

或延迟疾病的至少一种症状的发作。

[0099] “试剂盒”是指两种或更多种组分的任何集合,该两种或更多种组分一起构成可用于特殊目的的功能单元。通过说明(而不是限制),根据本披露的一个试剂盒可以包括与RNA 指导的核酸酶复合或能够与该核酸酶复合的指导RNA,并且伴有(例如悬浮于,或可悬浮于) 药学上可接受的载体。该试剂盒可用于将复合物引入例如细胞或受试者中,用于在这种细胞或受试者中引起所需基因组改变的目的。试剂盒的组分可以包装在一起,或者这些组分可分开包装。根据本披露的试剂盒还任选地包括使用说明书(DFU),其描述例如根据本披露的方法使用该试剂盒。DFU可以物理方式与试剂盒包装在一起,或者可以使试剂盒的使用者能获得该DFU,例如通过电子方式获得。

[0100] 术语“多核苷酸”、“核苷酸序列”、“核酸”、“核酸分子”、“核酸序列”和“寡核苷酸”是指DNA和RNA中的一系列核苷酸碱基(也称为“核苷酸”),并且意指两个或更多个核苷酸的任何链。多核苷酸、核苷酸序列、核酸等可以是嵌合混合物或其衍生物或经修饰的形式,单链或双链。它们可以在碱基部分、糖部分或磷酸骨架处经修饰,例如以改进分子的稳定性、其杂交参数等。核苷酸序列典型地携带遗传信息,包括但不限于细胞器用于制造蛋白质和酶的信息。这些术语包括双链或单链基因组DNA、RNA、任何合成的和遗传操作的多核苷酸,以及有义和反义多核苷酸二者。这些术语还包括含有经修饰碱基的核酸。

[0101] 常规IUPAC表示法用于本文所呈现的核苷酸序列中,如下表2中所示(还参见 Cornish-Bowden A,Nucleic Acids Res.[核酸研究]1985年5月10日;13(9):3021-30,通过引用并入本文)。然而应注意,在序列可能由DNA或者RNA编码的那些情况下,例如在指导分子靶向结构域中,“T”表示“胸腺嘧啶或尿嘧啶”。

表2:IUPAC核酸表示法

符号	碱基
A	腺嘌呤
T	胸腺嘧啶或尿嘧啶
G	鸟嘌呤
C	胞嘧啶
U	尿嘧啶
K	G或T/U
M	A或C
R	A或G
Y	C或T/U
S	C或G
W	A或T/U
B	C、G或T/U
V	A、C或G
H	A、C或T/U
D	A、G或T/U
N	A、C、G或T/U

[0102] 术语“蛋白质”、“肽”和“多肽”可互换使用,是指通过肽键连接在一起的氨基酸的

连续链。这些术语包括个别蛋白质、缔合在一起的蛋白质的组或复合物,以及此类蛋白质的片段或部分、变体、衍生物和类似物。肽序列使用常规表示法呈现于本文中,在左侧以氨基或N末端开始,并且前进至右侧的羧基或C末端。可以使用标准单字母或三字母缩写。

[0103] 术语“变体”是指诸如多肽、多核苷酸或小分子的实体,其与参照实体显示出显著的结构同一性,但与参照实体相比,在一个或多个化学部分的存在或水平方面与参照实体在结构上不同。在许多实施例中,变体在功能上也与其参照实体不同。通常,特定实体是否被恰当地视为参照实体的“变体”是基于其与参照实体的结构同一性程度。

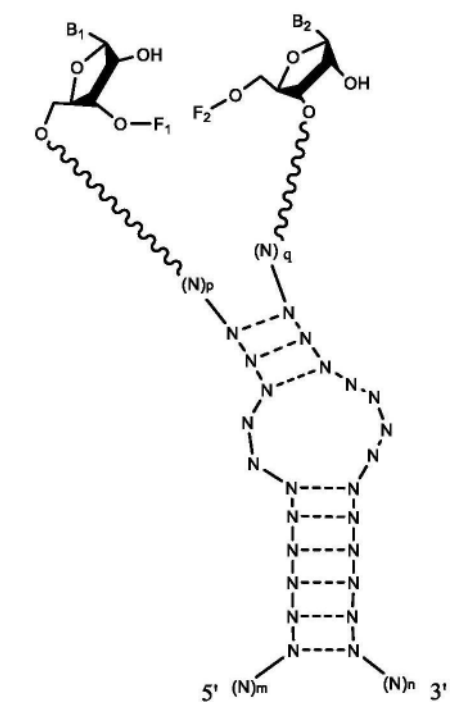
#### 概述

[0104] 本披露的某些实施例通常涉及用于合成指导分子的方法,其中两个或更多个指导片段(a)彼此退火,然后(b)使用适当的交联化学进行交联。发明人已经发现,包括在交联之前使指导片段预退火的步骤的方法提高了交联效率并且倾向于有利于形成所需的异二聚体产物,即使使用同型多官能交联接头也是如此。虽然不希望受任何理论的束缚,但认为交联效率的提高以及因此与未退火的同型二聚体相比,所需反应产物的产率的提高是由于退火的异型二聚体作为交联基质的稳定性增加,和/或通过预退火实现可用于形成同型二聚体等的游离RNA片段的比例降低。

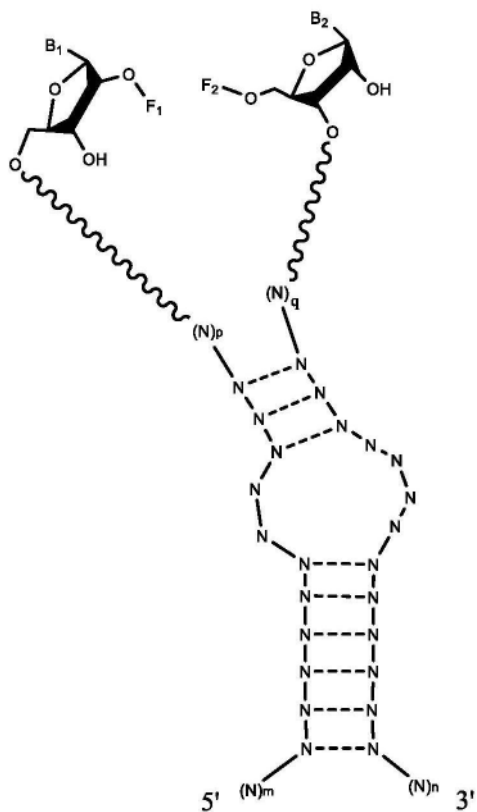
[0105] 包括指导片段的预退火的本披露的方法具有许多优点,包括但不限于:即使当片段是同型多官能的(例如,同型双官能的)时,例如本文所述的基于脲的交联方法中使用的胺官能化片段,它们也可以实现高产率;减少或不存在不希望的同型二聚体和其他反应产物可以转而简化下游纯化;并且由于用于交联的碎片往往比全长指导分子短,因此它们可能表现出较低的被以下污染的水平: $n-1$ 种类、截短种类、 $n+1$ 种类以及除了全长合成指导分子中观察到的其他污染物。

[0106] 关于预退火,本领域技术人员将理解,较长的退火碱基区可以比较短的区更稳定,并且在两个相似长度的区之间,更大程度的退火通常与更高的稳定性相关联。因此,在本披露的某些实施例中,设计片段以使片段之间的退火程度最大化,和/或将官能化的3'或5'末端定位在退火的碱基附近和/或彼此附近。

[0107] 如下面更详细讨论的,某些单分子指导分子,特别是单分子Cas9指导分子,其特征在于相对大的茎-环结构。例如,图1B和1C描绘了单分子化脓链球菌和金黄色葡萄球菌gRNA的二维结构,从图中可以明显看出,两种gRNA通常都包含相对较长的带有“凸起”的茎-环结构。在某些实施例中,合成的指导分子包括该茎环结构内的片段之间的交联。在某些情况下,这是通过交联分别在其3'和5'末端或其附近具有互补区的第一和第二片段来实现的;这些片段的3'和5'末端被官能化以促进交联反应,如例如下面的式II和III所示:



II.



III.


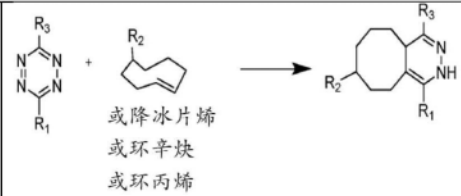
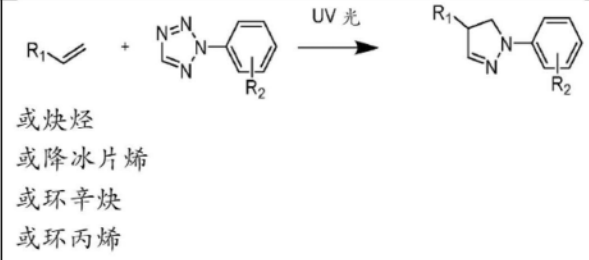
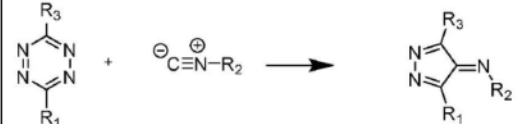
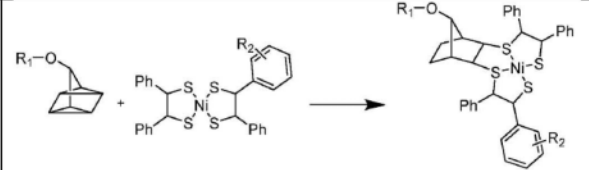
在这些式中,  $p$  和  $q$  各自独立地为  $0-6$ , 并且  $p+q$  为  $0-6$ ;  $m$  是  $20-40$ ;  $n$  是  $30-70$ ; 每个 --- 独立地代表相应核苷酸之间的氢键; 每个  代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合;  $N$  ---  $N$  独立地代表两个互补核苷酸, 任选地是氢键碱基配对的两个互补核苷酸;  $F_1$  和  $F_2$  各自包含官能团, 使得它们可以进行交联反应以交联两个指导片段。示例性交联化学品列于下表3中。

表3: 示例性交联化学品

反应类型	反应概要
巯基炔	$R_1-C\equiv C + HS-R_2 \longrightarrow R_1-CH=CH-S-R_2$
NHS 酯	$R_1-C(=O)-O-N \begin{array}{c} \diagup \\ \diagdown \end{array} \begin{array}{c} O \\ O \end{array} + H_2N-R_2 \longrightarrow R_1-C(=O)-NH-R_2$
巯基-烯	$R_1-CH=CH_2 + HS-R_2 \xrightarrow{h\nu, cat.} R_1-CH_2-CH_2-S-R_2$
异氰酸酯	$R_1-NCO + HX-R_2 \longrightarrow R_1-NH-C(=O)-X-R_2 \quad X = S \text{ 或 } NH$
环氧化物或氮丙啶	$R_1-\begin{array}{c} \diagup \\ \diagdown \end{array} O \text{ 或 } R_1-\begin{array}{c} \diagup \\ \diagdown \end{array} N + HS-R_2 \longrightarrow \begin{array}{c} HO \\   \\ R_1-CH_2-CH_2-S-R_2 \end{array} \text{ 或 } \begin{array}{c} H_2N \\   \\ R_1-CH_2-CH_2-S-R_2 \end{array}$
醛氨氧	$R_1-CHO + H_2N-O-R_2 \longrightarrow R_1-CH=N-O-R_2$
Cu 催化的叠氮化物-炔烃环加成	$R_1-N_3 + \begin{array}{c} \diagup \\ \diagdown \end{array} R_2 \xrightarrow{Cu^+} \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array}$
菌株促进的环加成	<p>环辛炔环加成 (与叠氮化物或氧化脒或硝酮)</p> $R_4-\begin{array}{c} \diagup \\ \diagdown \end{array} N^+-O^- \text{ 或 } R_1-\begin{array}{c} \diagup \\ \diagdown \end{array} N^+-O^- \text{ 或 } R_1-N_3 + \begin{array}{c} \diagup \\ \diagdown \end{array} R_2 \longrightarrow \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array}$ <p>降冰片烯环加成 (与叠氮化物或氧化脒或硝酮)</p> $R_1-\begin{array}{c} \diagup \\ \diagdown \end{array} + \begin{array}{c} R_2 \\   \\ N_3 \end{array} \text{ 或 } \begin{array}{c} R_2 \\   \\ N^+-O^- \end{array} \text{ 或 } \begin{array}{c} R_4 \\   \\ N^+-O^- \end{array} \longrightarrow \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} R_2 \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} R_4 \\   \\ R_1-N=N-N \end{array}$ <p>氧杂降冰片二烯环加成</p> $R_1-\begin{array}{c} \diagup \\ \diagdown \end{array} + \begin{array}{c} R_2 \\   \\ N_3 \end{array} \text{ 或 } \begin{array}{c} R_2 \\   \\ N^+-O^- \end{array} \text{ 或 } \begin{array}{c} R_4 \\   \\ N^+-O^- \end{array} \longrightarrow \begin{array}{c} F_3C \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} F_3C \\   \\ R_1-N=N-N \end{array} \text{ 或 } \begin{array}{c} F_3C \\   \\ R_1-N=N-N \end{array}$
施陶丁格连接 (Staudinger ligation)	$R_1-N_3 + \begin{array}{c} O \\    \\ R_2-C-PPh_2 \end{array} \longrightarrow \begin{array}{c} O \\    \\ R_2-C-NH-R_1 \end{array}$

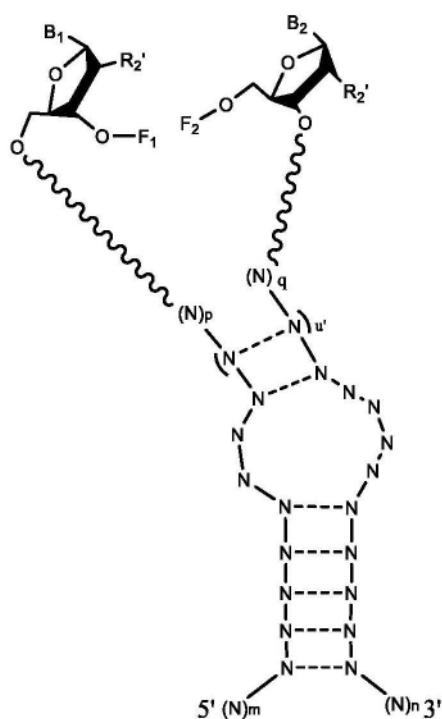
四嗪连接	 或降冰片烯 或环辛炔 或环丙烯
光诱导的四唑-烯烃环加成	 或炔烃 或降冰片烯 或环辛炔 或环丙烯
[4+1]环加成	
四环庚烷连接	

[0108] 虽然式II和III描述了位于指导分子重复-抗重复双链体中的“四环”结构(或交联接头替代“四环”结构)内的交联接头,但应理解可以在分子中的任何位置定位交联接头,例如,在指导分子内发生的任何茎环结构中,包括天然存在的茎环和工程化的茎环。特别地,本披露的某些实施例涉及缺乏四环结构并且包含位于第一和第二互补区末端(例如,在第一上茎区的3'末端和第二上茎区的5'末端)的交联接头的指导分子。

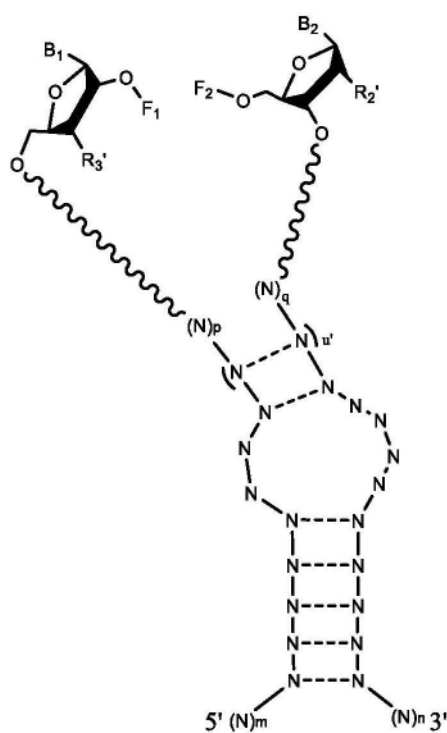
[0109] 式II和III描述了在重复-抗重复双链体中可以( $p>0$ 和 $q>0$ )或可以不( $p=0$ 和 $q=0$ )含有四环结构的指导分子。本发明的一个方面是认识到缺乏“四环”的指导分子可以表现出增强的连接效率,这是由于使官能化的3'和5'末端紧密接近并且处于合适的取向。

[0110] 可替代地/另外地,根据本披露的交联反应可包括“夹板”或单链寡核苷酸,其与官能化3'和5'末端处或附近的序列杂交,以稳定地使那些官能化末端彼此接近。

[0111] 本发明的另一个方面是认识到,与具有较短双链体的指导分子相比,具有较长双链体(例如,具有延伸的上部茎)的指导分子可以表现出增强的连接效率。这些较长的双链体结构在本披露中可互换地称为“延伸的双链体”,并且通常(但不是必须)位于指导片段中的官能化核苷酸附近。因此,在一些实施例中,本披露提供了具有以下式VIII和IX的指导分子:



VIII.



IX.

[0112] 在式VIII和IX中,  $p'$  和  $q'$  各自独立地是0和4之间的整数(包括端值),  $p' + q'$  是0和4之间的整数(包括端值),  $u'$  是2和22之间的整数(包括端值), 其他变量如式II和III中所定义。式VIII和IX描述了具有任选延伸的上茎以及任选的四环(即, 当  $p$  和  $q$  为0时)的双链体。由于较长的上茎导致的连接效率提高, 因此式VIII和IX的指导分子可以是有利的。此外, 较长的上茎和不存在四环的组合可有利于实现反应性基团  $F_1$  和  $F_2$  对于连接反应的适当取向。

[0113] 本发明的另一方面涉及识别指导片段可包括在单个指导片段内和/或在不同指导片段之间的多个互补区。例如, 在本披露的某些实施例中, 第一和第二指导片段设计有互补的上茎区和下茎区, 当完全退火时, 所述上茎区和下茎区产生异型二聚体, 其中 (a) 第一和

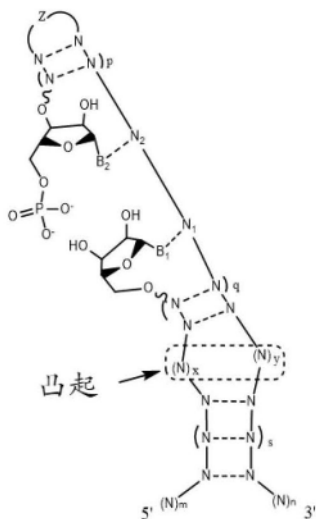
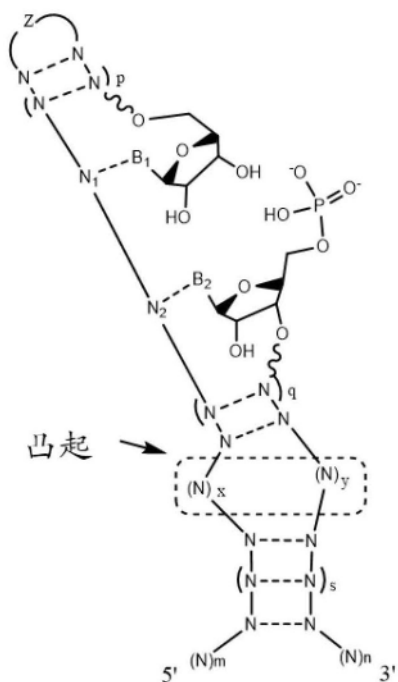
第二官能团位于双链上茎区的末端,对于交联反应适合地接近和/或(b)在第一和第二指导片段之间形成双链结构,其能够支持在指导分子和RNA指导的核酸酶之间形成复合物。然而,第一和第二指导片段可能彼此不完全退火,或形成内部双链体或同型二聚体,由此(a)和/或(b)不会发生。作为一个例子,在基于野生型crRNA和tracrRNA序列的化脓链球菌指导分子中,在下茎和上茎中可能存在多个高度互补的序列,例如聚-U或聚-A束,这可能导致涉及上茎和下茎区之间退火的不正确的“交错”异型二聚体,而不是期望的上茎区域彼此退火。类似地,在指导片段的靶向结构域序列与相同指导片段的另一个区或与不同片段之间可能形成不希望的双链体,并且在第一和第二指导片段的其他互补区之间可以发生错配,可能导致不完全的双链化、凸起和/或不成对的区段。

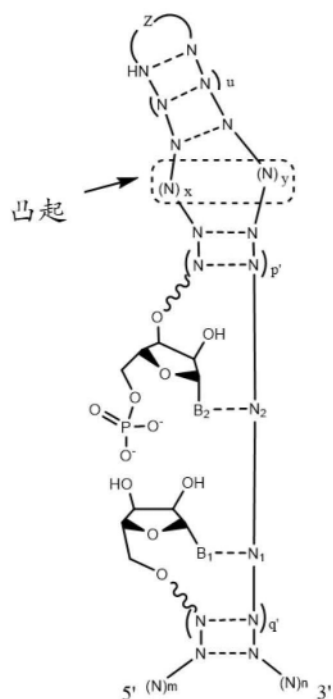
[0114] 虽然预测可能在指导片段之间形成的所有可能的不期望的内部或分子间双链结构是不实际的,但发明人已经发现,在一些情况下,进行修饰以减少或防止形成特定的错配和不希望的双链体可能对交联反应中所需的指导分子产物的产率具有显著影响,和/或导致从同一反应一种或多种污染物种类的减少。因此,在一些实施例,本披露提供了指导分子和方法,其中指导片段的一级序列已被设计为避免两个或特定的错配或不期望的双链体(例如,通过在第一和第二指导片段之间交换两个互补核苷酸)。例如,在上述野生型化脓链球菌片段的上茎中的A-U交换将产生包含不相同的UUUU和UAUU序列的第一指导片段和包含与第一片段的经修饰的序列互补的序列(即AAAA和AUAA序列)的第二指导片段。更广泛地说,指导片段可以包含序列变化,例如上茎或下茎的两个双链部分之间的核苷酸交换,上茎或下茎中序列的插入、缺失或替换,或结构变化,例如在选择用于减少或消除二级结构形成的位置的锁核酸(LNA)。

[0115] 尽管不希望受任何理论的束缚,但据信本文所述的双链体延伸、序列修饰和结构修饰通过相对于错配和不希望的双链体的形成增加双链体形成的能量有利性来促进所需双链体的形成并减少错配和不希望的双链体的形成。特定退火反应的能量有利性可以用吉布斯自由能( $\Delta G$ )表示;负 $\Delta G$ 值与自发反应相关,并且如果第一反应的 $\Delta G$ 小于(即,更负于)第二反应的 $\Delta G$ ,则第一退火反应在能量上比第二反应更有利。 $\Delta G$ 可以根据经验评估,基于特定双链体的热稳定性(熔化行为),例如使用NMR、荧光猝灭、UV吸光度、量热法等,如You, Tatourov和Owczarzy, “Measuring Thermodynamic Details of DNA Hybridization Using Fluorescence[用荧光测量DNA杂交的热力学细节]” Biopolymers[生物聚合物]第95卷,第7期,第472-486页(2011)所描述,将其通过引用并入本文用于所有目的。(参见,例如,第472-73页“引言”,和第473-475页“材料与方法”。)然而,在设计指导片段和退火反应时可能更实际使用计算模型来评估正确双链化的自由能以及选择的错配和不希望的双链化反应,并且可以使用许多工具来执行此类建模,包括由整合DNA技术公司(Integrated DNA Technologies) (科勒尔维尔(Coralville),爱荷华州)所有的biophysics.idtdna.com工具。可替代地或另外地,在文献中描述了许多利用热力学最近邻模型(TNN)的算法。参见例如,Tulpan, Andronescu和Leger, “Free energy estimation of short DNA duplex hybridizations[短DNA双链杂交的自由能估计],” BMC Bioinformatics[BMC生物信息学], 第11卷,第105期(2010)。(参见第1-2页的“背景”,描述TNN模型和MultiRNAFold包、Vienna包和UNAFold包)。其他算法也已在文献中描述,例如Kim等人, “An evolutionary Monte Carlo algorithm for predicting DNA hybridization[用于预测DNA杂交的进化蒙特卡

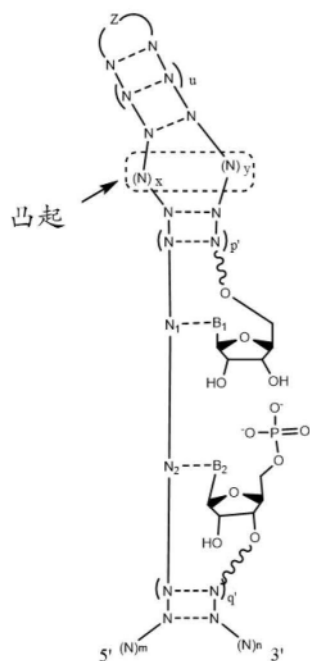
罗算法],”J.Biosystems[生物系统学杂志]第7卷,第5期(2007)。(参见描述所述模型的第71-2页第2节。)每个前述参考文献以其全部内容通过引用并入并用于此处的所有目的。

[0116] 当官能团位于包含多个碳的连接基团上时,式II和III中描述的排列可能是特别有利的。对于体积较小的交联接头,可能需要在官能化的3'和5'末端之间实现紧密并置。图3C和3D鉴定适于使用较短接头(包括但不限于磷酸二酯键)的化脓链球菌和金黄色葡萄球菌gRNA的双链体部分。通常选择这些位置以允许片段之间的退火,并定位官能化的3'和5'末端,使得它们在交联之前彼此紧邻。位于退火残基段内(而非邻近)的示例性3'和5'位置显示在下面的式IV、V、VI和VII中:






VI.



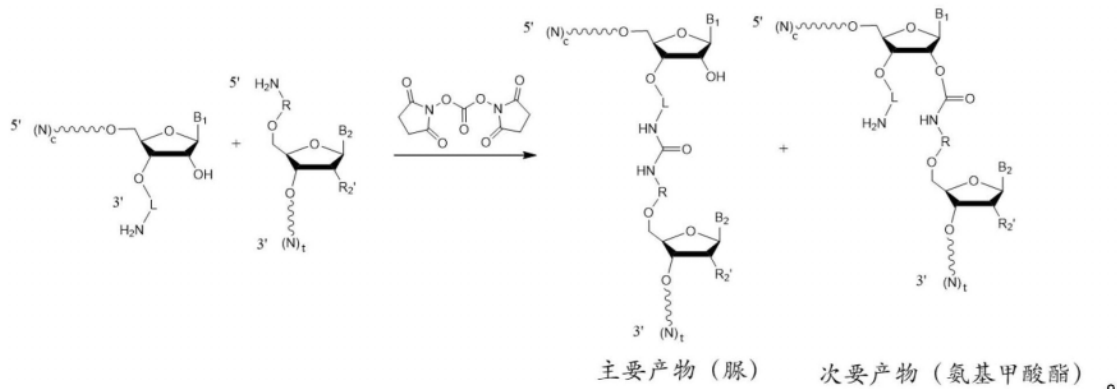
VII.

Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环；p和q各自独立地是0-2之间的整数(包括端值)，任选地是0；p'是0-4之间的整数(包括端值)，任选地是0；q'是2-4之间的整数(包括端值)，任选地是2；x是0-6之间的整数(包括端值)，任选地是2；y是0-6之间的整数(包括端值)，任选地是4；u是0-4之间的整数(包括端值)，任选地是2；s是2-6之间的整数(包括端值)，任选地是4；m是20-40之间的整数(包括端值)；n是30-70之间的整数(包括端值)；B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基；(N)<sub>m</sub>和(N)<sub>n</sub>中的每个N独立地是核苷酸残基；N<sub>1</sub>和N<sub>2</sub>各自独立地是核苷酸残基；N---N独立地代表两个互补核苷酸，任选地是氢键碱基配对的两个互补核苷酸；并且每个  代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

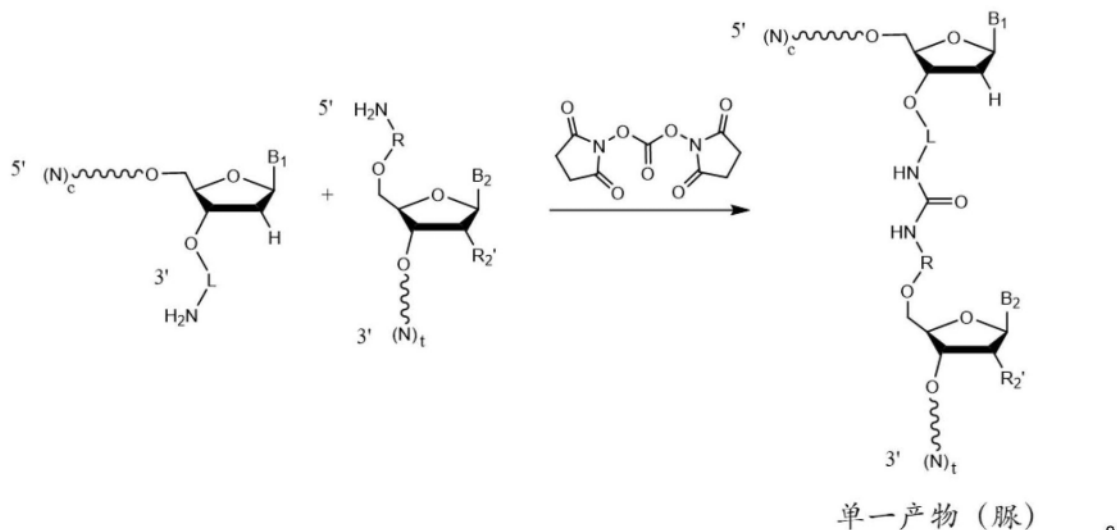
[0117] 本发明的另一方面是认识到式II、III、IV、V、VI或VII中任一个描述的排列可有利

于避免交联反应中的副产物,以及在没有同型二聚化的情况下允许发生同型双官能反应。两个异型二聚体链的预退火使反应性基团朝向期望的偶联取向,并且不利于与指导分子中的其他潜在反应性基团的反应。

[0118] 本发明的另一方面是认识到,优选地修饰靠近反应性基团的羟基(例如,第一片段的3'末端的2'-OH),以避免形成某些副产物。特别地,如下所示,发明人发现当胺官能化片段用于本文所述的基于脲的交联方法时,可形成氨基甲酸酯副产物:



[0119] 因此,在某些实施例中,第一片段3'末端的2'-OH被修饰(例如,被修饰为H、卤素、O-Me等),以防止形成氨基甲酸酯副产物。例如,2'-OH被修饰为2'-H:



[0120] 接下来转向交联,几个考虑因素与选择交联接头连接部分、官能团和反应性基团有关。其中包括接头尺寸,在水溶液中的溶解度和生物相容性,以及官能团反应性,交联的最佳反应条件,以及交联所需的任何必需的试剂,催化剂等。

[0121] 通常,选择接头大小和溶解度以保持或实现所需的RNA二级结构,并避免指导分子和RNA指导的核酸酶之间复合物的破坏或去稳定化。这两个因素在某种程度上是相关的,因为超过一定长度的有机接头可能在水溶液中难溶,并且可能与指导分子内的周围核苷酸和/或与指导分子复合的RNA指导的核酸酶中的氨基酸发生空间干扰。

[0122] 多种接头适用于本披露的各种实施例。某些实施例使用共同的连接部分,包括但不限于聚乙烯醚、聚乙烯、聚丙烯、聚乙二醇(PEG)、聚丙二醇(PEG)、聚乙烯醇(PVA)、聚乙交酯(PGA)、聚丙交酯(PLA)、聚己内酯(PCL)、及其共聚物。在一些实施例中,不使用接头。

[0123] 关于官能团,在其中双官能交联接头用于连接指导片段的5'和3'末端的实施例

中,待连接的指导片段的3'或5'末端用与交联接头的反应性基团反应的官能团修饰。通常,这些修饰包括胺、巯基、羧基、羟基、烯烃(例如末端烯烃)、叠氮化物和/或其他合适的官能团中的一种或多种。多官能(例如,双官能)交联剂也是本领域公知的,并且可以是异型官能的或同型官能的,并且可以包括任何合适的官能团,包括但不限于异硫氰酸酯、异氰酸酯、酰基叠氮化物、NHS酯、磺酰氯、甲苯磺酸酯、三苯甲基酯、醛、胺、环氧化物、碳酸酯(如、双(对硝基苯基)碳酸酯)、芳基卤、烷基卤、亚胺酯、羧酸酯、烷基磷酸酯、酸酐、氟苯酯、HOBt酯、羟甲基磷、0-甲基异脲、DSC、NHS氨基甲酸酯、戊二醛、活化双键、环状半缩醛、NHS碳酸酯、咪唑氨基甲酸酯、酰基咪唑、甲基吡啶鎓醚、吡内酯、氰酸酯、环状亚胺碳酸酯、氯三嗪、脱氢氮杂环庚三烯、6-磺基-胞嘧啶衍生物、马来酰亚胺、氮杂环丙烷、TNB硫醇、Ellman试剂、过氧化物、乙烯基砜、苯硫酯、重氮烷、重氮乙酰基、环氧化物、重氮、二苯甲酮、蒽醌、重氮衍生物、二氮杂环丙烯衍生物、补骨脂素衍生物、烯烃、苯基硼酸等。

[0124] 这些和其它交联化学物质是本领域已知的,并且在文献中进行了总结,包括Greg T.Hermanson, *Bioconjugate Techniques* [生物缀合物技术], 第3版. 2013, 学术出版社 (Academic Press) 出版, 将其通过引用以其全部内容并入本文并用于所有目的。

[0125] 在某些实施例中,包含通过披露开提供的方法合成的指导分子的组合物的特征在于高纯度的所需指导分子反应产物,低水平的不期望种类(包括n-1种类、截短、n+1种类、指导片段同型二聚体、未反应的官能化指导片段等)的污染。在本披露的某些实施例中,包含合成指导分子的纯化组合物可在组合物内包含多个种类(即,按质量或摩尔浓度,指导分子是组合物中最常见的种类)。可替代地或另外地,根据本披露的实施例的组合物可包含 $\geq 70\%$ 、 $\geq 75\%$ 、 $\geq 80\%$ 、 $\geq 85\%$ 、 $\geq 90\%$ 、 $\geq 95\%$ 、

$\geq 96\%$ 、 $\geq 97\%$ 、 $\geq 98\%$ 和/或 $\geq 99\%$ 的具有所需长度(例如,相对于参照指导分子序列在5'末端缺少截短)和所需序列(例如,包含参照指导分子序列的5'序列)的指导分子。

[0126] 例如,在一些实施例中,包含根据本披露的指导分子的组合物(例如,包含使用本文所述的适当交联化学交联的片段的指导分子)相对于参照指导分子序列包含少于20%、15%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或更少的在5'末端包含截短的指导分子。另外或可替代地,包含根据本披露的指导分子的组合物(例如,包含使用本文所述的适当交联化学物质交联的片段的指导分子)包含至少约90%、95%、96%、97%、98%、99%或100%的具有与参照指导分子序列的相应5'序列100%相同的5'序列(例如,包含指导分子的核苷酸1-30、1-25或1-20或由其组成的5'序列)的指导分子。在一些实施例中,如果组合物包含具有与参照指导分子序列的相应5'序列小于100%相同的5'序列的指导分子,并且此类指导分子以大于或等于0.1%的水平存在,则此类指导分子不包含潜在脱靶位点的靶向结构域。在一些实施例中,包含根据本披露的指导分子的组合物包含的至少约80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多的指导分子在5'末端不包含截短(相对于参照指导分子序列),并且至少约90%、95%、96%、97%、98%、99%或100%的此类指导分子(即,在5'末端不包含截短的此类指导分子)具有与参照指导分子序列的相应5'序列100%相同的5'序列(例如,包含指导分子的核苷酸1-30、1-25或1-20或由其组成的5'序列),并且如果组合物包含具有与参照指导分子序列的相应5'序列小于100%相同的5'序列的指导分子,并且此类指导分子以大于或等于0.1%的水平存在,则此类指导分子不包含潜在脱靶位点的靶向结构域。在一些实施例中,包含根据本披露的指导

分子的组合物包含少于约10%的相对于参照指导分子序列在5'末端包含截短的指导分子,并且表现出可接受水平的活性/功效。在一些实施例中,包含根据本披露的指导分子的组合物包含(i)至少约99%的指导分子具有与参照指导分子序列的相应5'序列100%相同的5'序列(例如,包含指导分子的核苷酸1-30、1-25或1-20或由其组成的5'序列),以及(ii)如果组合物包含具有与参照指导分子序列的相应5'序列小于100%相同的5'序列的指导分子,并且此类指导分子以大于或等于0.1%的水平存在,则此类指导分子不包含潜在脱靶位点的靶向结构域,并且组合物表现出可接受的特异性/安全性水平。组合物的纯度可以表示为组合物中总指导分子的分数的(按质量或摩尔浓度),表示为组合物中所有RNA或所有核酸的分数的(按质量或摩尔浓度),表示为组合物中所有溶质的分数的(按质量),和/或表示为组合物总质量分数。

[0127] 通过本领域已知的任何合适的方法评估包含根据本披露的指导分子的组合物的纯度。例如,可以通过凝胶电泳定性或半定量地评估所需指导分子种类的相对丰度。可替代地或另外地,通过色谱法(例如,液相色谱法、HPLC、FPLC、气相色谱法)、光谱法(例如,质谱法、无论是基于飞行时间、扇形场、四极杆质量、离子阱、轨道阱、傅里叶变换离子回旋共振或其他技术)、核磁共振(NMR)光谱(如可见光、红外或紫外)、热稳定性方法(如差示扫描量热法等)、测序方法(例如,使用模板转换寡核苷酸)及其组合(例如,色谱-光谱法等)评估所需指导分子种类的纯度。

[0128] 本文提供的合成的指导分子以与任何其他指导分子(例如,gRNA)基本相同的方式操作,并且通常通过以下来操作:(a)与RNA指导的核酸酶如Cas9形成复合物,(b)与靶序列相互作用,所述靶序列包括与指导分子的靶向序列互补的区域和由RNA指导的核酸酶识别的原型间隔子邻近基序(PAM),和任选地(c)修饰靶序列内或靶序列邻近的DNA,例如通过形成DNA双链断裂,单链断裂等,这可以通过在含有指导分子和RNA指导的核酸酶的细胞内的DNA修复途径操作来修复。

[0129] 在一些实施例中,本文所述的指导分子,例如使用本文所述方法产生的指导分子,可以充当作用于RNA的酶(例如逆转录酶)的底物。不希望受理论束缚,由于通过根据本披露的方法的预退火促进的反应性末端的紧密并置,存在于本文所述的指导分子内的交联接头可与这种进行性酶相容。

[0130] 上述示例性实施例集中于将本文所述的合成和交联方法应用于来自两个指导片段的指导分子的组装。然而,本文描述的方法具有多种应用,其中许多应用对于技术人员来说是显而易见的。这些应用在本披露的范围内。作为一个实例,本披露的方法可用于将异源序列连接至指导分子。异源序列可包括但不限于如Cotta-Ramusino等人的WO 2017/180711(将其通过引用并入本文用于所有目的)中所述的DNA供体模板。(参见,例如,第I部分,第23页的“gRNA融合分子”,描述共价连接的模板核酸,以及使用夹板寡核苷酸促进模板与指导分子3'末端的连接。)异源序列还可以包含被肽DNA或RNA结合结构域识别的核酸序列,例如MS2环,这也在上面的WO 2017/180711的第I部分中描述。

[0131] 该概述集中于少数示例性实施例,所述示例性实施例说明了与指导分子的合成有关的某些原理,以及包含此类指导分子的组合物。然而,为了清楚起见,本披露包含尚未描述但对于本领域技术人员显而易见的修改和变化。考虑到这一点,以下披露旨在更概括地说明基因组编辑系统的操作原理。下文不应被理解为限制性,而是说明基因组编辑系统的

某些原理,所述原理与本披露组合,将告知本领域技术人员关于在本披露范围内的其他实施方式和修改的信息。

### 基因组编辑系统

[0132] 术语“基因组编辑系统”是指具有RNA指导的DNA编辑活性的任何系统。本披露的基因组编辑系统包括至少两种从天然存在的CRISPR系统改适的组分:指导分子(例如,指导RNA或gRNA)和RNA指导的核酸酶。这两种组分形成复合物,所述复合物能够与特定核酸序列结合并在所述核酸序列中或其周围编辑DNA,例如通过制备一条或多条单链断裂(SSB或切口)、双链断裂(DSB)和/或点突变。

[0133] 天然存在的CRISPR系统进化性地组织化为两个类别和五种类型(Makarova等人, Nat Rev Microbiol.[自然综述:微生物学]2011年6月;9(6):467-477(Makarova),通过引用并入本文),并且虽然本披露的基因组编辑系统可改适任一类型或类别的天然存在的CRISPR系统的组分,但本文所呈现的实施例通常是从2类和II型或V型CRISPR系统改适。2类系统涵盖II型和V型,其特征为相对较大的多结构域RNA指导的核酸酶蛋白(例如,Cas9或Cpf1)以及一个或多个指导RNA(例如,crRNA和任选地tracrRNA),它们形成核糖核蛋白(RNP)复合物,该复合物缔合(即,靶向)并裂解与crRNA的靶向(或间隔序列)序列互补的特定基因座。根据本披露的基因组编辑系统类似地靶向并编辑细胞DNA序列,但与自然界中存在的CRISPR系统显著不同。例如,本文所述的单分子指导分子在自然界中不存在,并且根据本披露的指导分子和RNA指导的核酸酶二者可并入任一数目的非天然存在的修饰。

[0134] 基因组编辑系统可以用多种方式来实施(例如给予或递送至细胞或受试者),并且不同的实施可适合于不同应用。例如,在某些实施例中,基因组编辑系统是作为蛋白质/RNA复合物(核糖核蛋白,或RNP)来实施,其可以包括于药物组合物中,所述药物组合物任选地包括药学上可接受的载体和/或囊封剂,例如脂质或聚合物微粒或纳米颗粒、胶束、脂质体等。在某些实施例中,基因组编辑系统是作为一种或多种编码上述RNA指导的核酸酶和指导分子组分的核酸(任选地具有一种或多种其他组分)来实施;在某些实施例中,基因组编辑系统是作为一种或多种包含此类核酸的载体来实施,例如病毒载体,例如腺相关病毒;并且在某些实施例中,基因组编辑系统是作为前述任一种的组合来实施。根据本文所述原理操作的其他或经修改的实施将为技术人员所了解并且在本披露的范围内。

[0135] 应注意,本披露的基因组编辑系统可靶向单一特定核苷酸序列,或者可通过使用两个或更多个指导分子靶向(并且能平行编辑)两个或更多个特定核苷酸序列。在本披露通篇中,多个指导分子的使用称为“多重化”,并且可用于靶向多个无关的所关注靶序列,或用于在单一靶结构域内形成多个SSB或DSB,并且在一些情形中,用于在这种靶结构域内产生特定编辑。例如,Maeder等人的国际专利公开号WO 2015/138510(Maeder)(其是通过引用并入本文)描述用于修正人类CEP290基因中的点突变(C.2991+1655A至G)的基因组编辑系统,所述点突变导致产生隐蔽剪接位点,这又降低或消除该基因的功能。Maeder的基因组编辑系统利用两个指导RNA,这些指导RNA靶向该点突变任一侧上的(即,侧接)序列,并且形成侧接该突变的DSB。这又促进了包括突变在内的间插序列的缺失,由此消除隐蔽剪接位点并恢复正常的基因功能。

[0136] 作为另一实例,Cotta-Ramusino等人的WO 2016/073990(“Cotta-Ramusino”) (通过引用并入本文)描述利用两个gRNA与Cas9切口酶(制造单链切口的Cas9,例如化脓链球菌

(*S. pyogenes*) D10A) 的基因组编辑系统, 该布置称为“双重切口酶系统”。Cotta-Ramusino 的双重切口酶系统经配置以在所关注序列的相对链上制造两个偏移一个或多个核苷酸的切口, 所述切口组合产生具有悬突(在 Cotta-Ramusino 情形中是 5' 悬突, 但 3' 悬突也有可能) 的双链断裂。在一些情况下, 该悬突又可以促进同源定向修复事件。并且作为另一实例, Palestrant 等人的 WO 2015/070083 (“Palestrant”, 通过引用并入本文) 描述靶向编码 Cas9 的核苷酸序列的 gRNA (称为“管理 RNA”), 其可以包括于基因组编辑系统中, 所述基因组编辑系统包含一个或多个其他 gRNA 以容许 Cas9 的瞬时表达, 所述 Cas9 可能原本例如在一些经病毒转导的细胞中是组成型表达的。这些多重化应用旨在具有示例性而不是限制性, 并且技术人员将了解, 其他多重化应用通常与本文所述的基因组编辑系统相容。

[0137] 在一些情况下, 基因组编辑系统可以形成双链断裂, 这些双链断裂是通过细胞 DNA 双链断裂机制例如 NHEJ 或 HDR 来修复。这些机制描述于多处文献中, 例如 Davis 和 Maizels, PNAS, 111 (10): E924-932, 2014 年 3 月 11 日 (Davis) (描述 Alt-HDR); Frit 等人, DNA Repair [DNA 修复] 17 (2014) 81-97 (Frit) (描述 Alt-NHEJ); 以及 Iyama 和 Wilson III, DNA Repair [DNA 修复] (Amst.) 2013 年 8 月; 12 (8): 620-636 (Iyama) (概括描述经典 HDR 和 NHEJ 路径)。

[0138] 如果基因组编辑系统通过形成 DSB 来操作, 那么此类系统任选地包括促进或有助于特定双链断裂修复模式或特定修复结果的一个或多个组分。例如, Cotta-Ramusino 还描述其中添加单链寡核苷酸“供体模板”的基因组编辑系统; 将供体模板并入细胞 DNA 的靶区域中, 该靶区域由基因组编辑系统裂解, 并且可以导致靶序列中的变化。

[0139] 在某些实施例, 基因组编辑系统在不引起单链或双链断裂的情况下修饰靶序列, 或修饰靶序列中或附近的基因的表达。例如, 基因组编辑系统可包括融合至作用于 DNA 的功能结构域的 RNA 指导的核酸酶, 由此修饰靶序列或其表达。作为一个实例, RNA 指导的核酸酶可连接至(例如融合至) 胞苷脱氨酶功能结构域, 并且可通过产生所靶向的 C 至 A 取代来操作。示例性核酸酶/脱氨酶功能描述于 Komor 等人 Nature [自然] 533, 420-424 (2016 年 5 月 19 日) (“Komor”) 中, 所述文献是通过引用并入。可替代地, 基因组编辑系统可利用裂解失活的(即, “死”) 核酸酶, 例如死 Cas9 (dCas9), 并且可通过在细胞 DNA 的一个或多个所靶向区域上形成稳定复合物来操作, 由此干扰涉及一个或多个所靶向区域的功能, 包括但不限于 mRNA 转录、染色质重塑等。

### 指导分子

[0140] 本文使用的术语“指导分子”是指促进 RNA 指导的核酸酶(例如 Cas9 或 Cpf1) 与细胞中的靶序列(例如基因组或附加体序列) 的特异性缔合(或“靶向”) 的任何核酸。指导分子可以是 RNA 分子或杂合 RNA/DNA 分子。指导分子可以是单分子(包含单一分子, 并且可替代地称为嵌合) 或模块(包含多于一个、并且典型地两个单独的分子, 例如 crRNA 和 tracrRNA, 其通常例如通过双链化彼此缔合)。指导分子和其组成部分的描述遍及于文献中, 例如 Briner 等人 (Molecular Cell [分子细胞] 56 (2), 333-339, 2014 年 10 月 23 日 (Briner)), 将该文献是通过引用并入), 以及 Cotta-Ramusino。

[0141] 在细菌和古细菌中, II 型 CRISPR 系统通常包含 RNA 指导的核酸酶蛋白(例如 Cas9)、包括与外来序列互补的 5' 区的 CRISPR RNA (crRNA) 和包括与 crRNA 的 3' 区互补并形成双链体的 5' 区的反式激活 crRNA (tracrRNA)。虽然并非旨在受限于任何理论, 但认为此双链体有助于形成 Cas9/指导分子复合物, 并且是所述复合物的活性所需的。在 II 型 CRISPR 系统经改

适用于基因编辑中时,发现crRNA和tracrRNA可以接合成单一单分子或嵌合指导RNA,在一个非限制性实例中借助桥接crRNA(在其3'末端)和tracrRNA(在其5'末端)的互补区的四核苷酸(例如GAAA)“四环(tetraloop)”或“接头”序列来接合。(Mali等人Science.[科学]2013年2月15日;339(6121):823-826(“Mali”);Jiang等人Nat Biotechnol.[自然生物技术]2013年3月;31(3):233-239(“Jiang”);和Jinek等人,2012年,Science[科学]8月17日;337(6096):816-821(“Jinek”),将所有这些文献均通过引用并入本文。)

[0142] 指导分子不论是单分子或模块都包括“靶向结构域”,所述靶向结构域与靶序列内的靶结构域完全或部分互补,所述靶序列例如期望编辑的细胞基因组中的DNA序列。靶向结构域在文献中通过多种名称来提及,包括但不限于“指导序列”(Hsu等人,Nat Biotechnol.[自然生物技术]2013年9月;31(9):827-832(“Hsu”),通过引用并入本文)、“互补性区域”(Cotta-Ramusino)、“间隔序列”(Briner)和一般性地称为“crRNA”(Jiang)。不论给予其何种名称,靶向结构域典型地长度为10-30个核苷酸,并且在某些实施例中长度为16-24个核苷酸(例如,长度为16、17、18、19、20、21、22、23或24个核苷酸),并且在Cas9指导分子情形中位于5'末端处或5'末端附近,并且在Cpf1指导分子情形中位于3'末端处或3'末端附近。

[0143] 除了靶向结构域以外,指导分子典型地(但不一定,如下文所讨论)包括多个可影响指导分子/Cas9复合物的形成或活性的结构域。例如,如上文所提及,通过指导分子的第一和第二互补性结构域形成的双链化结构(也称为重复:抗重复双链体)与Cas9的识别(REC)叶相互作用,并且可以介导Cas9/指导分子复合物的形成。(Nishimasu等人,Cell[细胞]156,935-949,2014年2月27日(Nishimasu 2014)和Nishimasu等人,Cell[细胞]162,1113-1126,2015年8月27日(Nishimasu 2015),将这两篇文献均通过引用并入本文)。

[0144] 与第一和第二互补性结构域一起,Cas9指导分子典型地包括两个或更多个其他双链化区域,该两个或更多个其他双链化区域在体内但不一定在体外参与核酸酶活性。(Nishimasu 2015)。在第二互补性结构域的3'部分附近的第一茎环被不同地称为“近端结构域”(Cotta-Ramusino)、“茎环1”(Nishimasu 2014和2015)以及“连结(nexus)”(Briner)。一个或多个其他茎环结构通常存在于指导分子的3'末端附近,其数目依物种而变:化脓链球菌gRNA典型地包括2个3'茎环(总计4个茎环结构,包括重复:抗重复双链体),而金黄色葡萄球菌和其他物种仅具有一个(总计3个茎环结构)。对根据物种组织化的保守茎环结构(更通常地,和指导分子结构)的描述提供于Briner中。

[0145] 虽然前述说明集中于用于Cas9的指导分子,但应了解,已经(或可能在未来)发现或发明其他RNA指导的核酸酶,其利用在一些方面与针对这一点描述的那些指导分子不同的指导分子。例如,Cpf1(“来自普雷沃菌属(Prevotella)和弗朗西斯菌属(Franciscella)1的CRISPR”)是最近发现的RNA指导的核酸酶,其发挥功能不需要tracrRNA。(Zetsche等人,2015,Cell[细胞]163,759-771 2015年10月22日(Zetsche I),通过引用并入本文)。用于Cpf1基因组编辑系统的指导分子通常包括靶向结构域和互补性结构域(替代性地称为“把手”)。还应注意,在用于Cpf1的指导分子中,靶向结构域通常存在于3'末端处或附近,而不是如上文结合Cas9指导分子所述的5'末端(把手位于Cpf1指导分子的5'末端处或5'末端附近)。

[0146] 但本领域技术人员将了解,虽然在来自不同原核物种的指导分子之间或在Cpf1与Cas9指导分子之间可能存在结构差异,但指导分子的操作原理通常是一致的。因为这种操

作一致性,指导分子可在广义上通过其靶向结构域序列来定义,并且技术人员将了解,可将给定靶向结构域序列并入任何适宜指导分子中,包括单分子或嵌合指导分子,或包括一种或多种化学修饰和/或序列修饰(取代、额外核苷酸、截短等)的指导分子。因此,为了便于呈现本披露,指导分子可仅在其靶向结构域序列方面加以描述。

[0147] 更通常地,技术人员将了解,本披露的一些方面涉及可以使用多个RNA指导的核酸酶实施的系统、方法和组合物。为此,除非另外指定,否则术语指导分子应理解为不仅涵盖那些与Cas9或Cpf1的特定种类相容的那些指导分子,还涵盖可以用于任何RNA指导的核酸酶的任何适宜的指导分子(例如gRNA)。通过说明,在某些实施例中,术语指导分子可以包括用于存在于2类CRISPR系统中的任何RNA指导的核酸酶的指导分子,所述RNA指导的核酸酶例如II型或V型或CRISPR系统,或从其衍生或改适的RNA指导的核酸酶。

#### 交联的指导分子

[0148] 本公开的某些实施例涉及通过例如非核苷酸化学键合交联的指导分子。如上所述,键合的位置可以在指导分子的茎环结构中。在一些实施例中,指导分子包含

[0149] 在一些实施例中,单分子指导分子从5'至3'包含:

第一指导分子片段,其包含:

靶向结构域序列;

第一下茎序列;

第一凸起序列;

第一上茎序列;

非核苷酸化学键合;和

第二指导分子片段,其包含:

第二上茎序列;

第二凸起序列;和

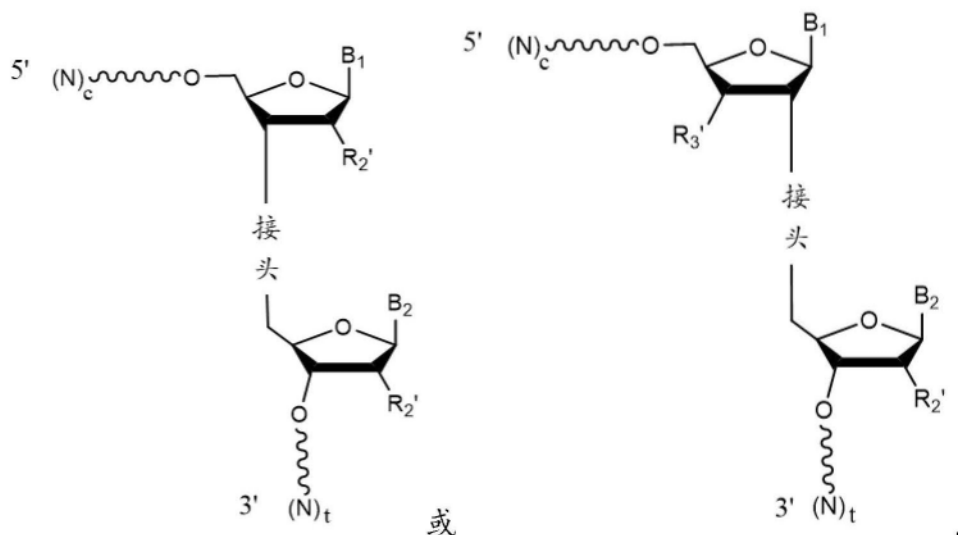
第二下茎序列,

其中(a)所述第一下茎序列中的至少一个核苷酸与所述第二下茎序列中的核苷酸碱基配对,和(b)所述第一上茎序列中的至少一个核苷酸与所述第二上茎序列中的核苷酸碱基配对。

[0150] 在一些实施例中,指导分子不包含第一和第二上茎序列之间的四环序列。在一些实施例中,第一和/或第二上茎序列包含数目从4至22(包括端值)的核苷酸。在一些实施例中,第一和/或第二上茎序列包含数目从1至22(包括端值)的核苷酸。在一些实施例中,第一和/或第二上茎序列包含数目从4至22(包括端值)的核苷酸。在一些实施例中,第一和第二上茎序列包含数目从8至22(包括端值)的核苷酸。在一些实施例中,第一和第二上茎序列包含数目从12至22(包括端值)的核苷酸。

[0151] 在一些实施例中,指导分子的特征在于用于在第一和第二指导分子片段之间形成双链体的吉布斯自由能( $\Delta G$ )小于用于在两个第一指导分子片段之间形成双链体的 $\Delta G$ 。在一些实施例中,用于在第一和第二指导分子片段之间形成双链体的 $\Delta G$ 的特征在于(i)第一和第二上茎序列之间和(ii)第一和第二下茎序列之间各自的大于50%、60%、70%、80%、90%或95%碱基配对小于形成以下双链体的 $\Delta G$ ,所述双链体的特征在于(i)和(ii)之间少于50%、60%、70%、80%、90%或95%碱基配对。

[0152] 在一些实施例中,合成指导分子具有下式:



其中  $(N)_c$  和  $(N)_t$  中的每个  $N$  独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包括与  $(N)_t$  的 5' 区互补或部分互补并形成双链体的 3' 区;

$c$  是 20 或更大的整数;

$t$  是 20 或更大的整数;

接头是非核苷酸化学键合;

$B_1$  和  $B_2$  各自独立地是核碱基;

$R_2'$  和  $R_3'$  各自独立地是 H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$  或 O- $R'$ , 其中每个  $R'$  独立地是保护基团或烷基, 其中所述烷基可任选地被取代; 和

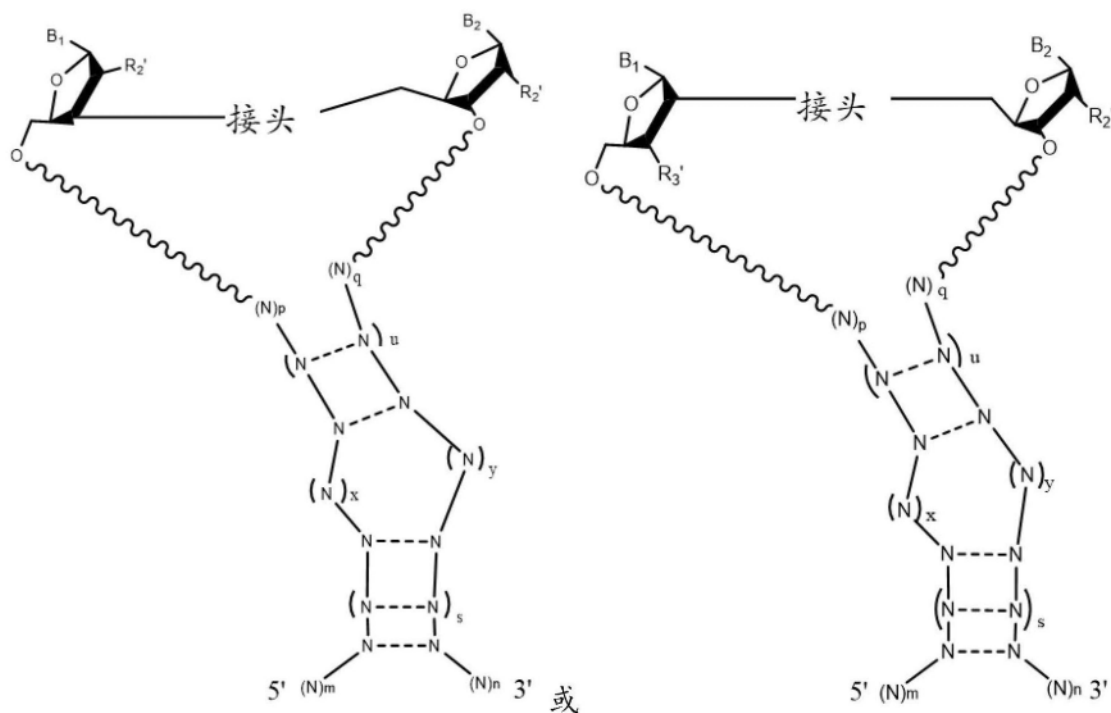
每个  $\sim$  独立地代表磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合。

[0153] 在一些实施例中,  $(N)_t$  和  $(N)_c$  的双链体区包含表 4 中列出的序列。

表 4.  $(N)_t$  和  $(N)_c$  的示例性序列

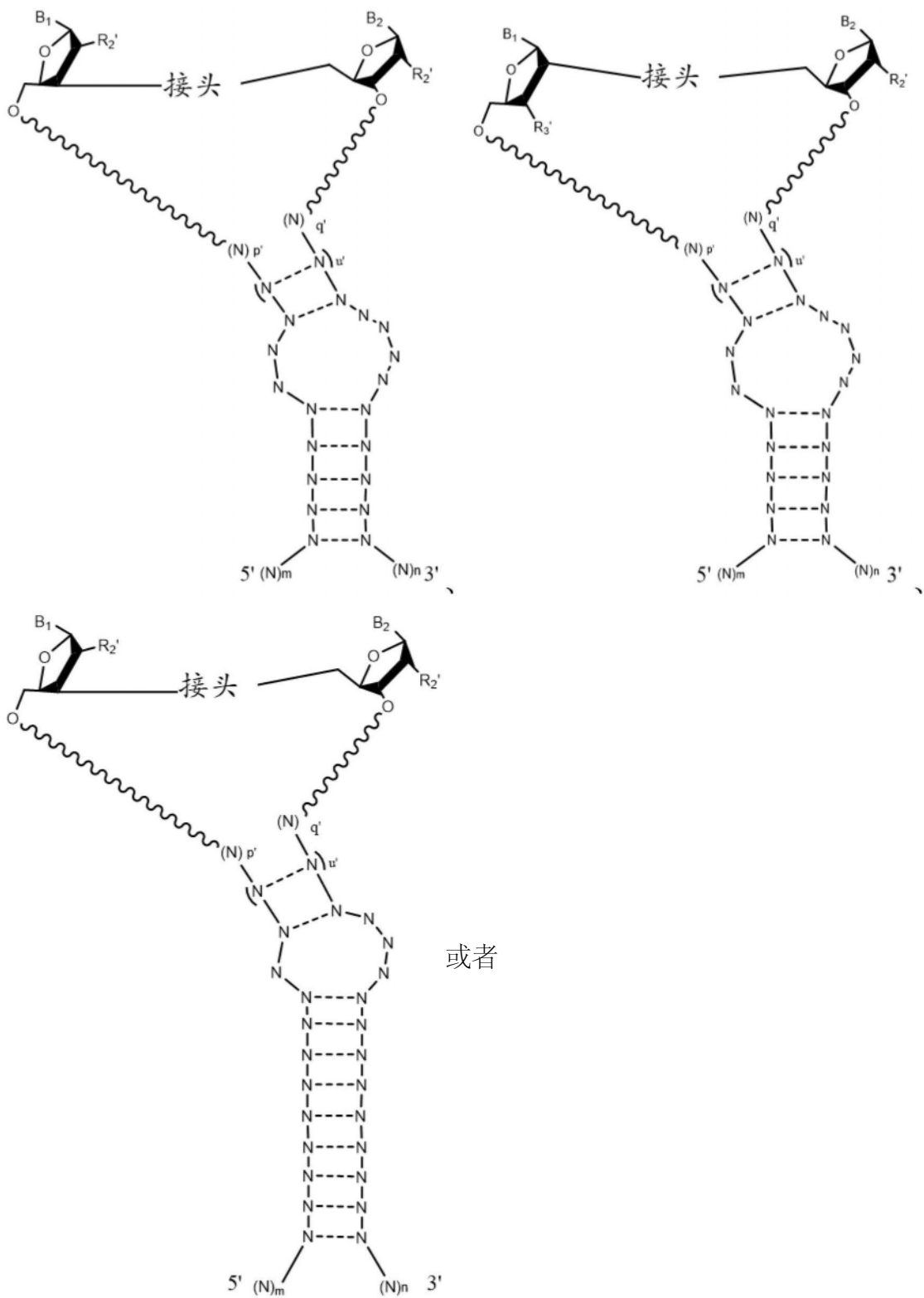
SEQ ID NO.	序列
1	GUUUUAGAGCUAG
2	AUAGCAAGUUAAAAU
3	GUUUUAGAGCU
4	AGCAAGUUAAAAU
5	GUUUUAGAGCUAG
6	CUAGCAAGUUAAAAU
7	GUUUUAGAGCUAUG
8	CAUAGCAAGUUAAAAU
9	GUAUUAGAGCUAUGCUGUUUU
10	AAAACAGCAUAGCAAGUUAUUAU
11	GUAUUAGAGCUAUGCU
12	AGCAUAGCAAGUUAUUAU
13	GUUUUAGAGCUAUGCUGUUUU
14	AAAACAGCAUAGCAAGUUAUUAU
15	GUUUUAGAGCUAUGCU
16	AGCAUAGCAAGUUAUUAU
17	GUUUUAGAGCUAAAG
18	AUUUAGCAAGUUAUUAU
19	GUUUUAGAGCUAA
20	UUAGCAAGUUAUUAU
21	GUUUUAGAGCUAAAGG
22	ACCUUUAGCAAGUUAUUAU
23	GUUUUAGAGCUAG
24	GUUUUAGUACUCU
25	AGAAUCUACUAAAAC
26	GUUUUAGUACUCUGUA
27	UACAGAAUCUACUAAAAC
28	GUUUUAGUACUCUGUAAUUUUAGG
29	CCUAAAAUUACAGAAUCUACUAAAAC
30	GUUUUAGUACUCUGUAAUUUUAGGUAUGA
31	UCAUACCUAAAAUUACAGAAUCUACUAAAAC

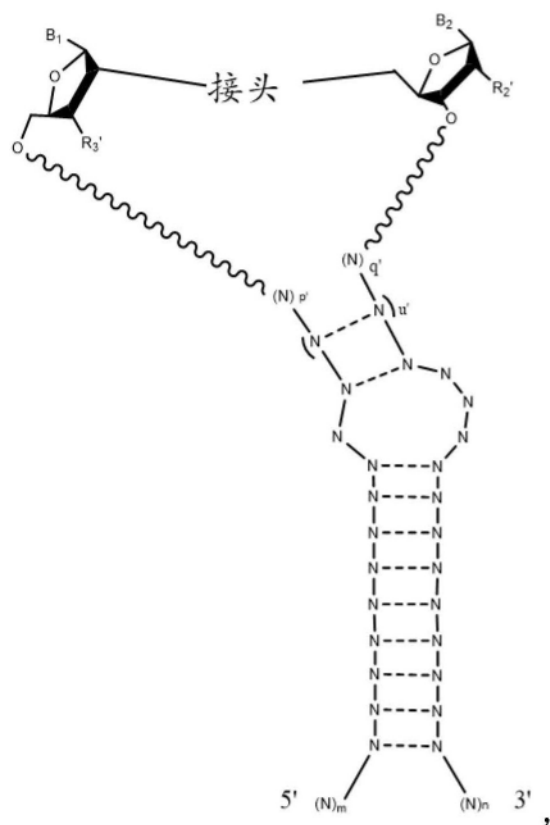
[0154] 在一些实施例中,指导分子具有下式:



[0155] 在一些实施例中，(N---N)<sub>u</sub>和(N---N)<sub>s</sub>不包含具有3个或更多个核苷酸的相同序列。在一些实施例中，(N---N)<sub>u</sub>和(N---N)<sub>s</sub>不包含具有4个或更多个核苷酸的相同序列。在一些实施例中，(N---N)<sub>s</sub>包含N'UUU、UN'UU、UUN'U或UUUN'序列并且(N---N)<sub>u</sub>包含UUUU序列，其中N'是A、G或C。在一些实施例中，(N---N)<sub>s</sub>包含UUUU序列并且(N---N)<sub>u</sub>包含N'UUU、UN'UU、UUN'U或UUUN'序列，其中N'是A、G或C。在一些实施例中，N'是A。在一些实施例中，N'是G。在一些实施例中，N'是C。

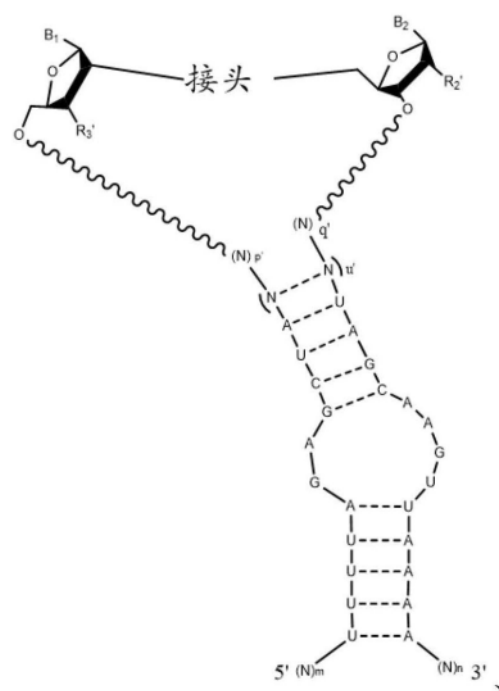
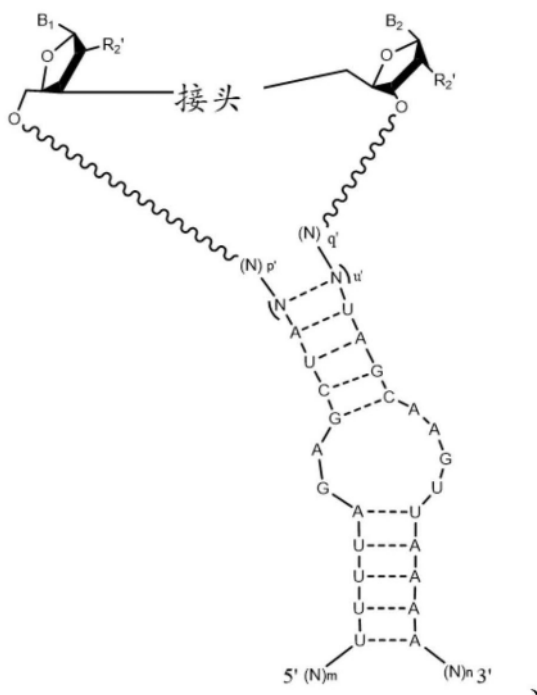
[0156] 在一些实施例中，指导分子基于化脓链球菌或金黄色葡萄球菌Cas9系统中使用的gRNA。在一些实施例中，指导分子具有下式：

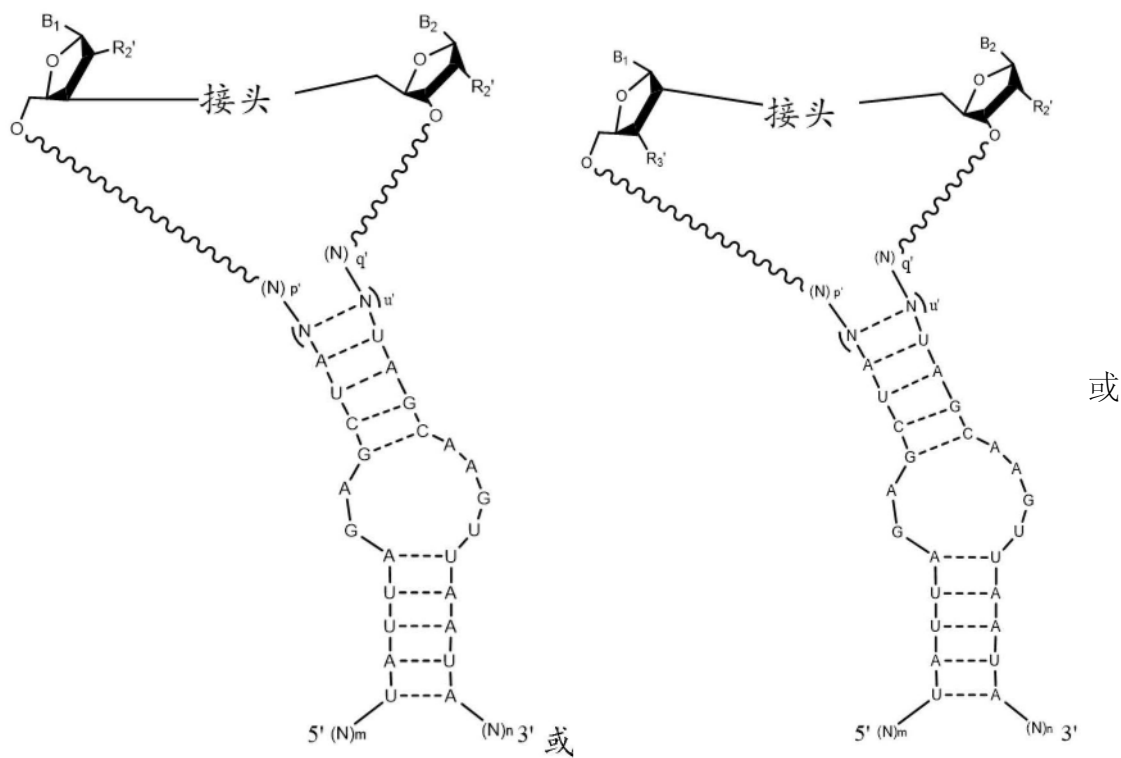




$q'$  各自独立地是0到6之间的整数(包括端值), 并且  $p' + q'$  是0到6之间的整数(包括端值)。

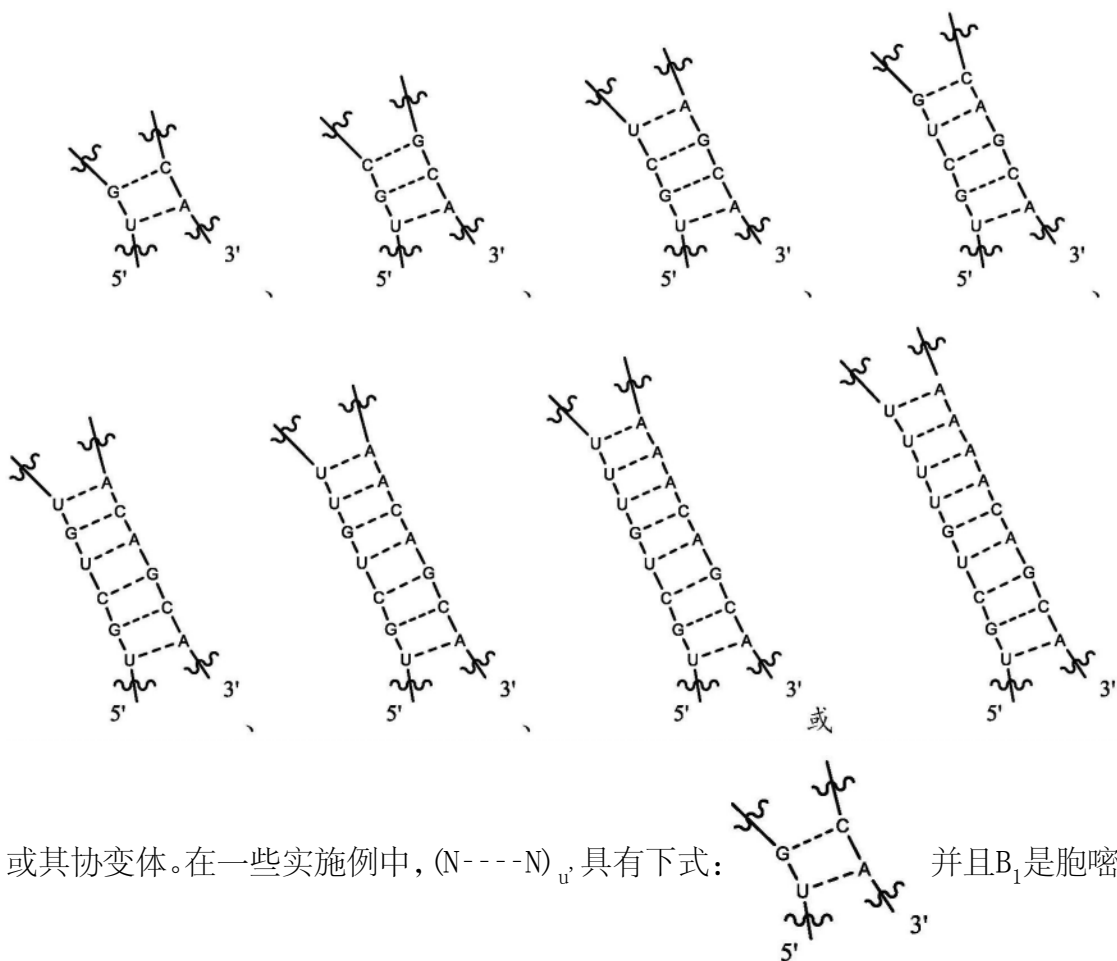
[0157] 在一些实施例中, 指导分子具有下式:



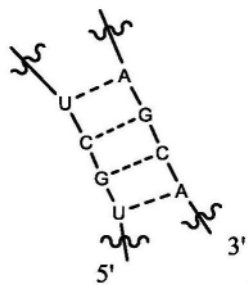


其协变体。

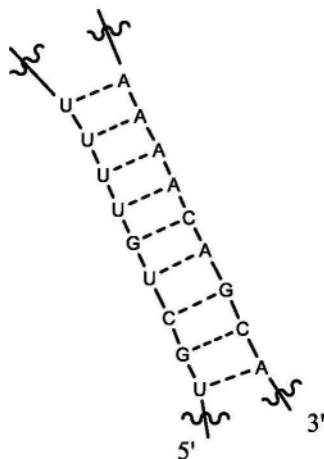
在一些实施例中,  $(N \cdots N)_u$  具有下式:



啶残基, B<sub>2</sub>是鸟嘌呤残基, 或其协变体。在一些实施例中, (N---N)<sub>u</sub>, 具有下式:



并且B<sub>1</sub>是鸟嘌呤残基, B<sub>2</sub>是胞嘧啶残基, 或其协变体。在一些实施例中,

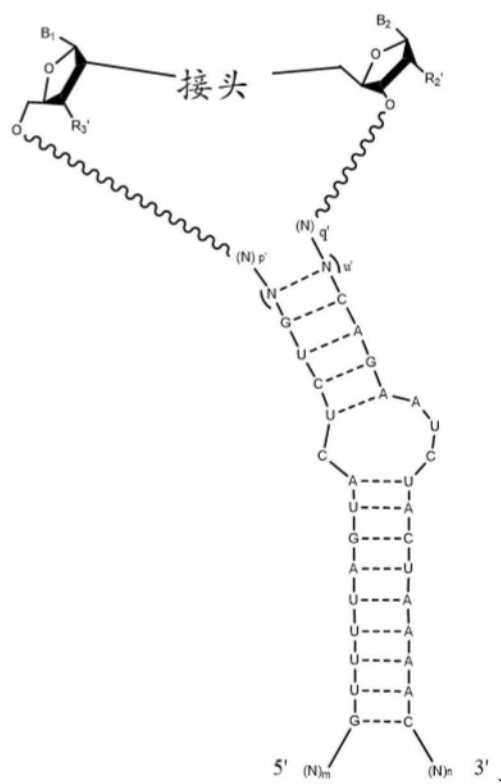
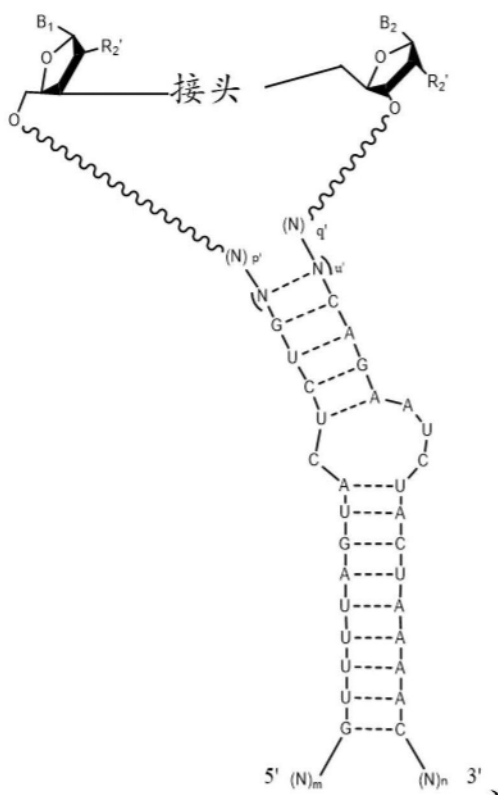


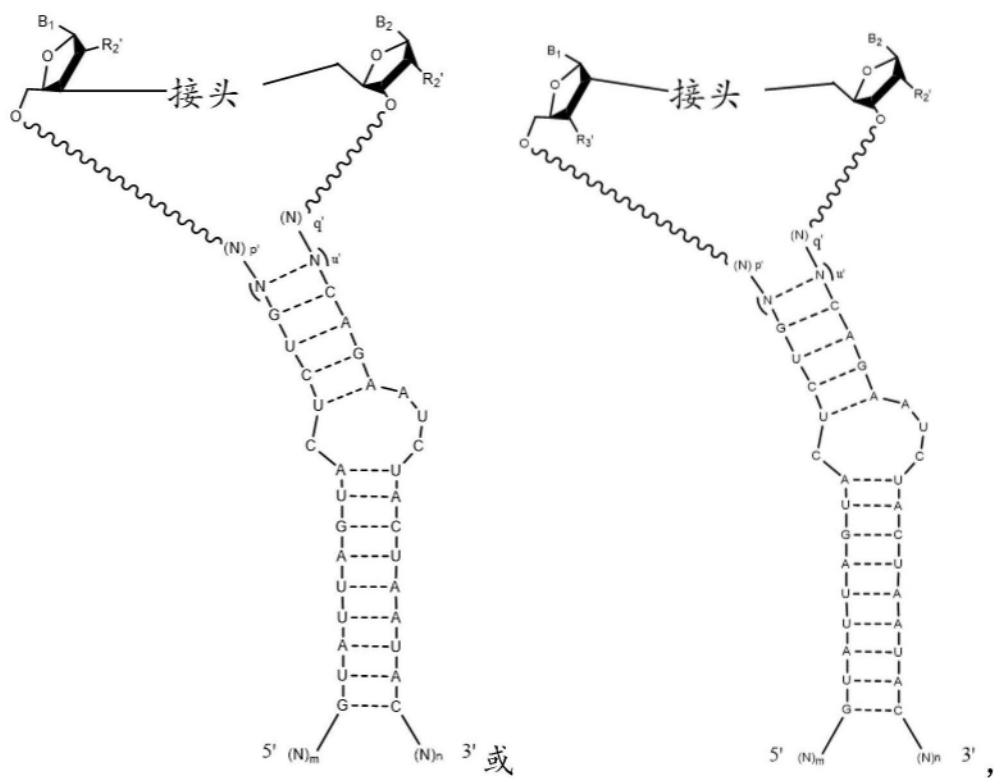
$(N\cdots N)_u$ , 具有下式:

并且B<sub>1</sub>是鸟嘌呤残基,B<sub>2</sub>是胞嘧啶残基,或其

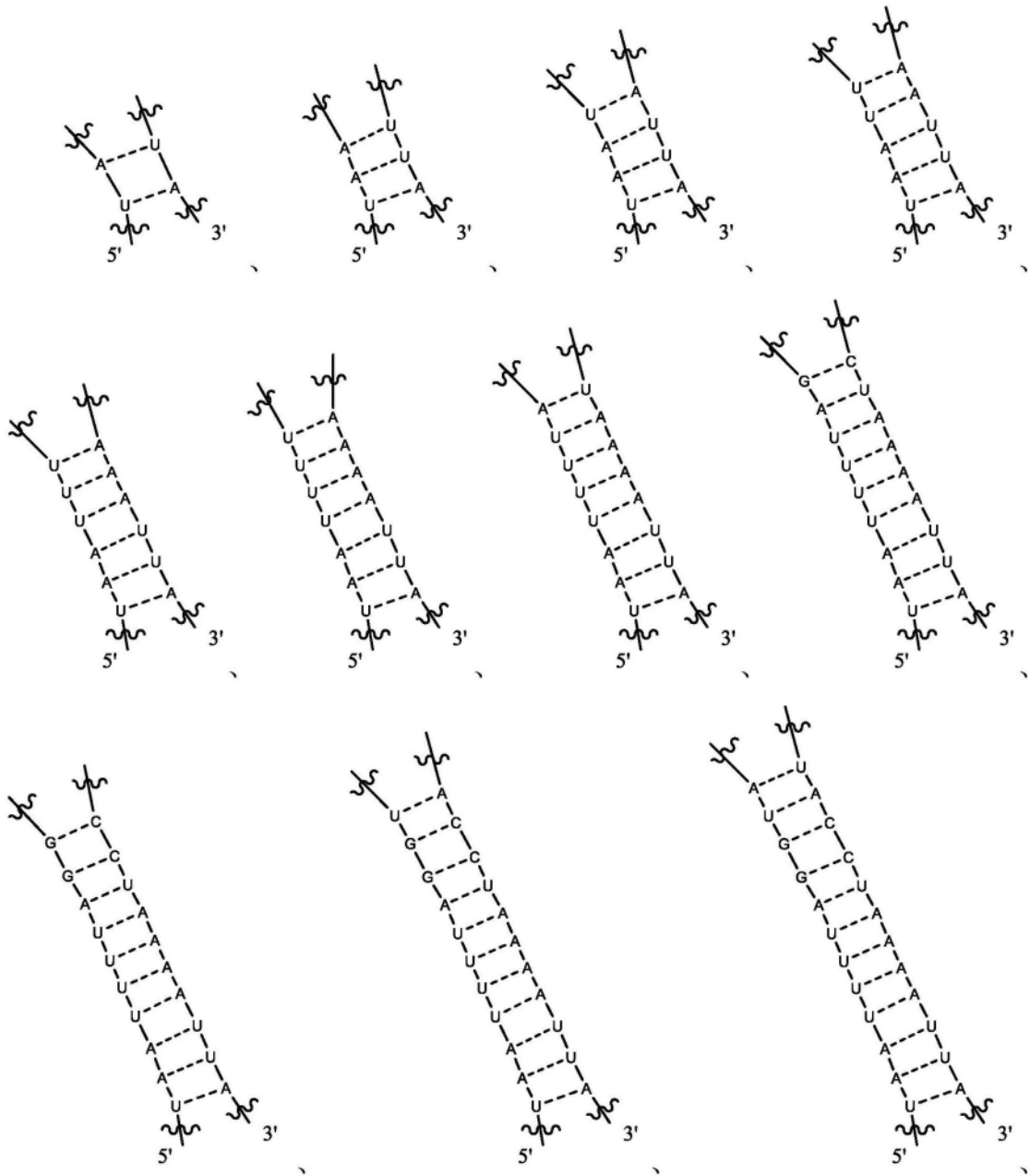
协变体。

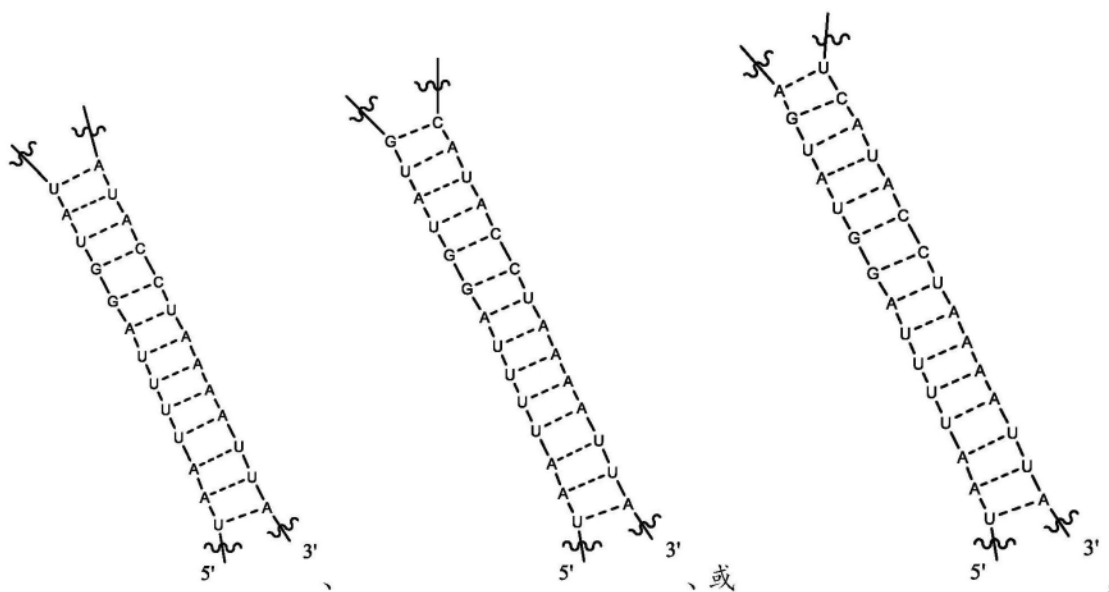
[0158] 在一些实施例中,指导分子具有下式:



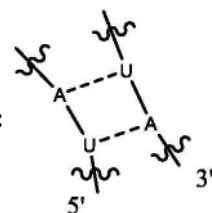


变体。在一些实施例中,  $(N\text{---}N)_u$  具有下式:



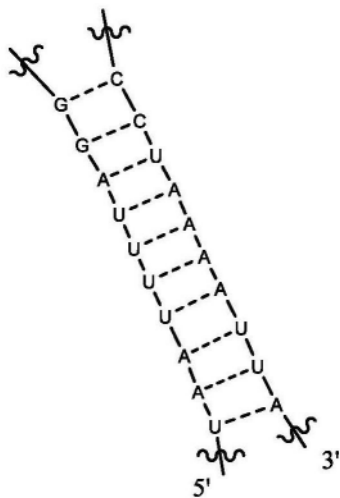


或其协变体。在一些实施例中,  $(N\cdots N)_u$  具有下式:

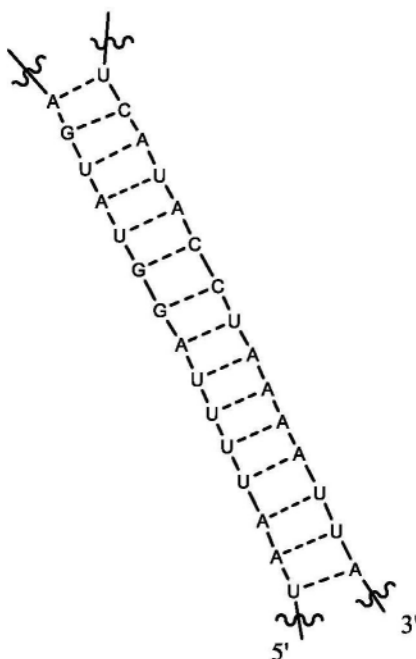


并且 $B_1$ 是腺嘌呤残基,

$B_2$ 是尿嘧啶残基,或其协变体。在一些实施例中,  $(N\cdots N)_u$  具有下式:



并且 $B_1$ 是尿嘧啶残基, $B_2$ 是腺嘌呤残基,或其协变体。在

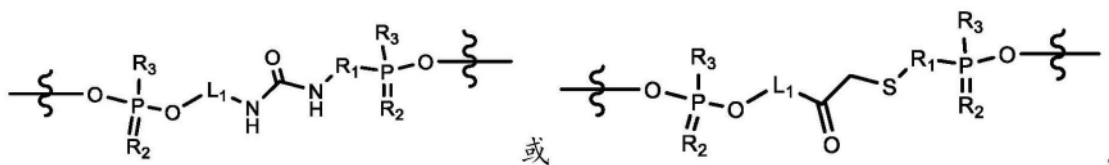


一些实施例中,  $(N\cdots N)_n$  具有下式:

并且B<sub>1</sub>是鸟嘌呤残基,

B<sub>2</sub>是胞嘧啶残基,或其协变体。

[0159] 在一些实施例中,接头具有下式:



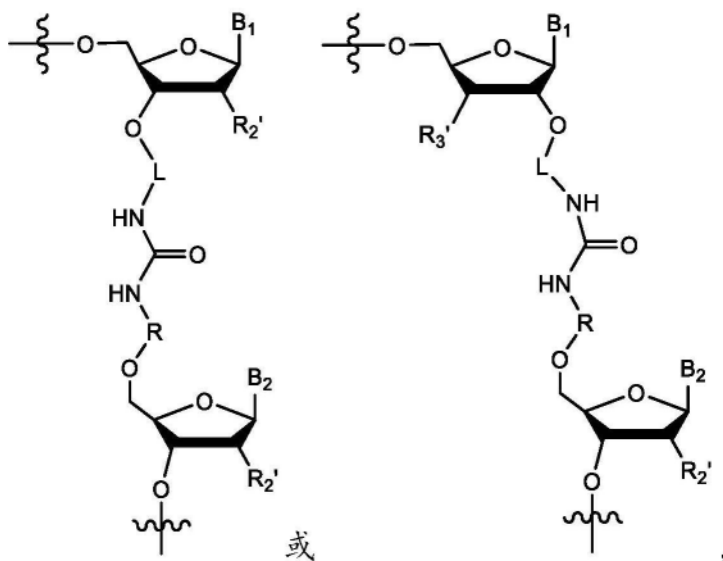
其中：

每个 $R_i$ 独立地是0或S;

每个 $R_3$ 独立地是 $O^-$ 或 $COO^-$ ;和

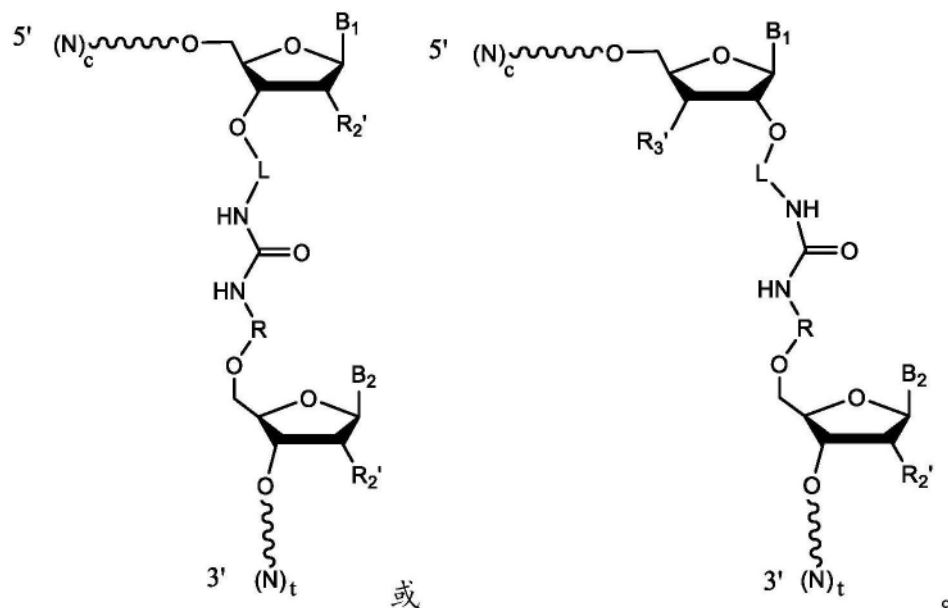
$L_1$ 和 $R_1$ 各自是非核苷酸化学接头。

[0160] 在一些实施例中,交联的指导分子的化学键合包含脲。在一些实施例中,包含脲的指导分子具有下式:

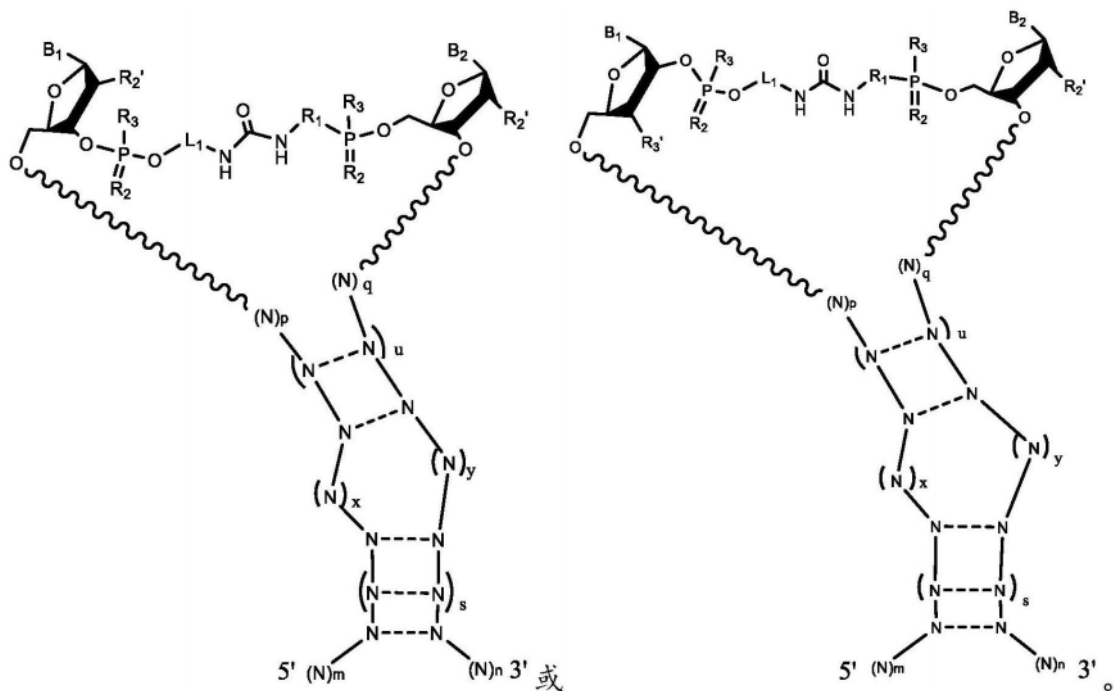


其中L和R各自独立地是非核苷酸接头。

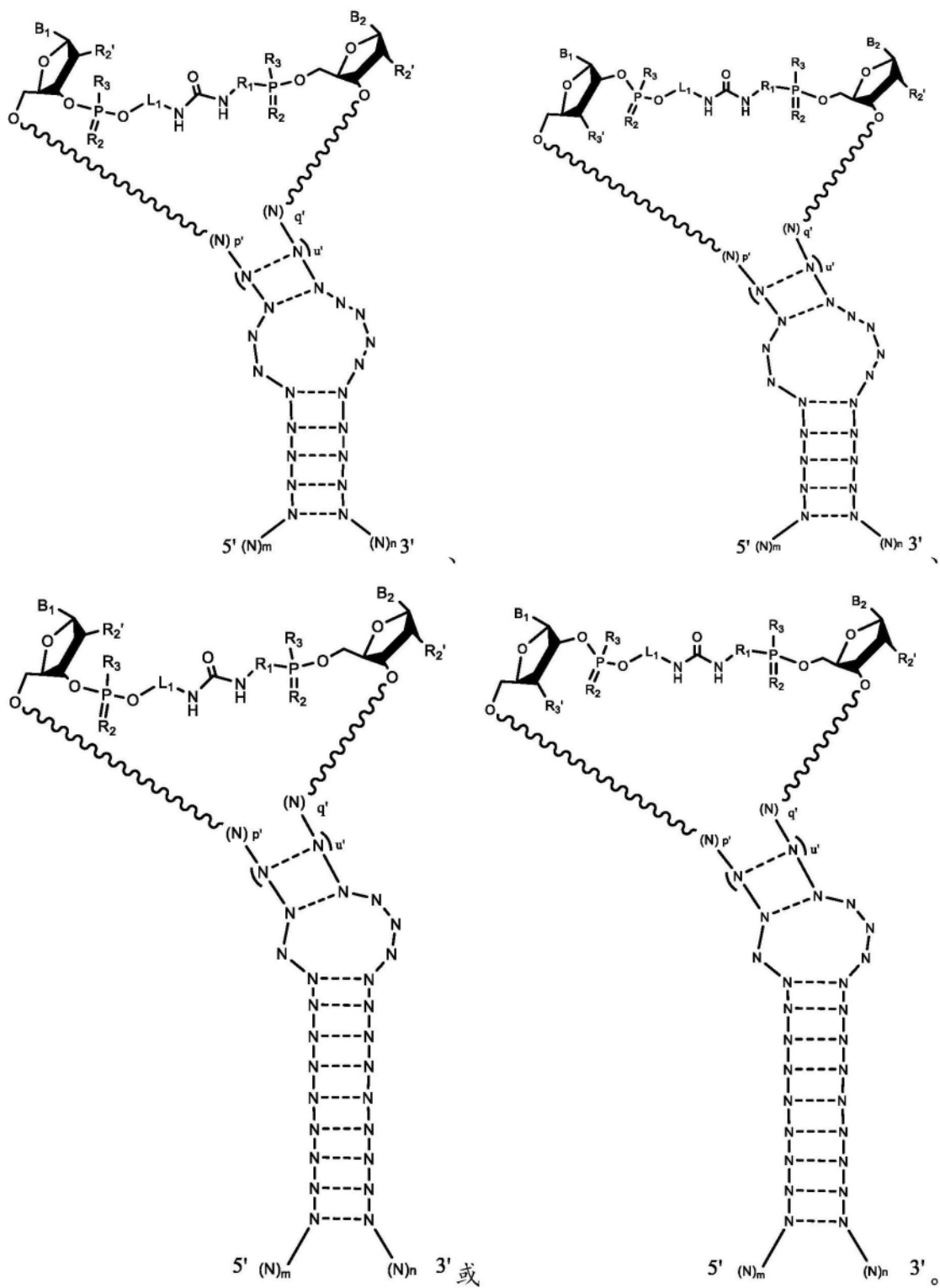
[0161] 在一些实施例中,包含脲的指导分子具有下式:



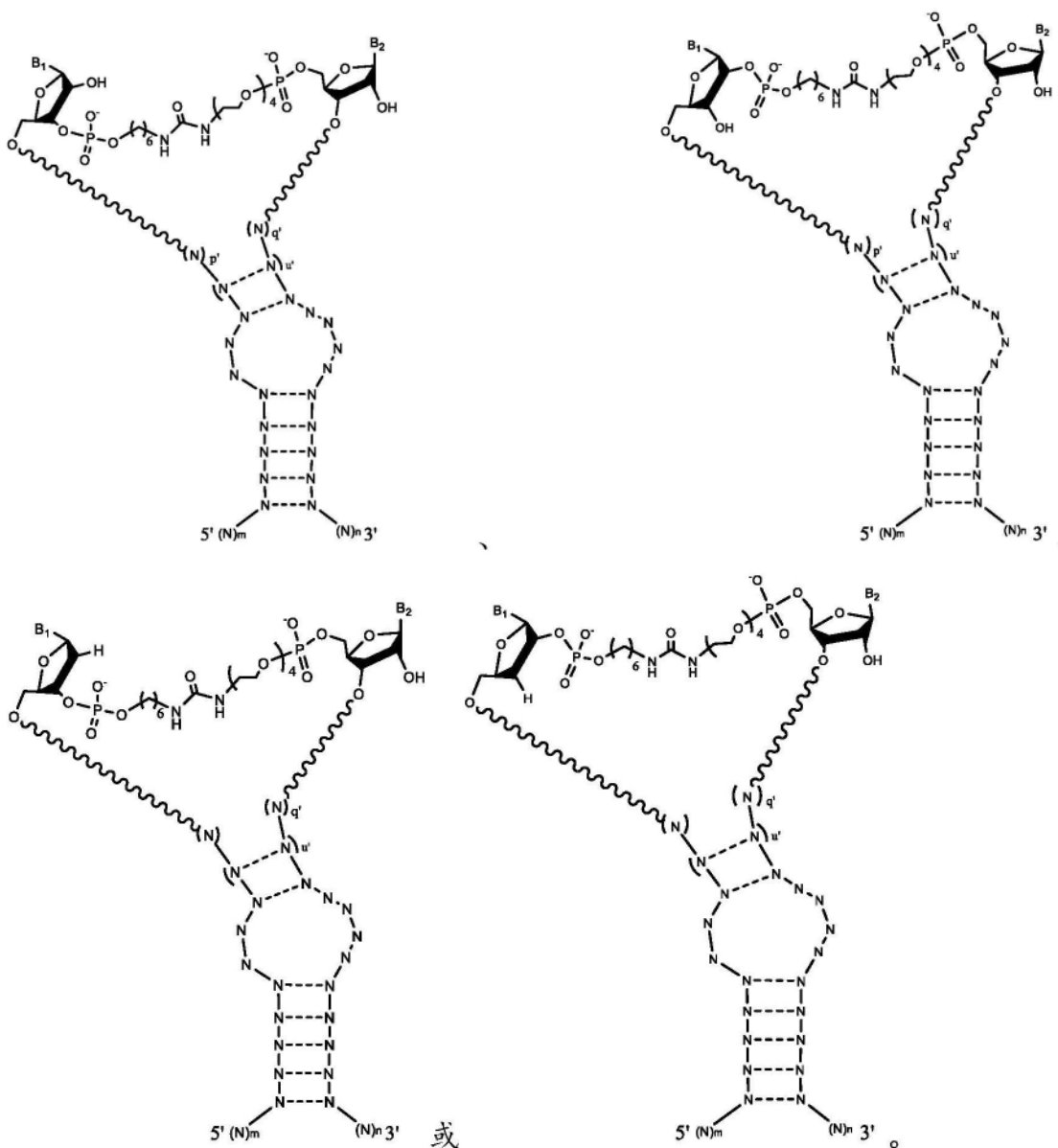
[0162] 在一些实施例中,包含脲的指导分子具有下式:



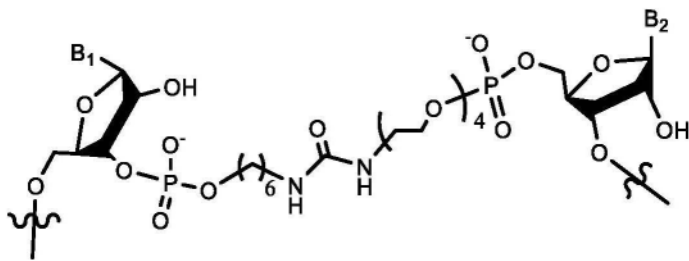
[0163] 在一些实施例中,包含脲的指导分子具有下式:



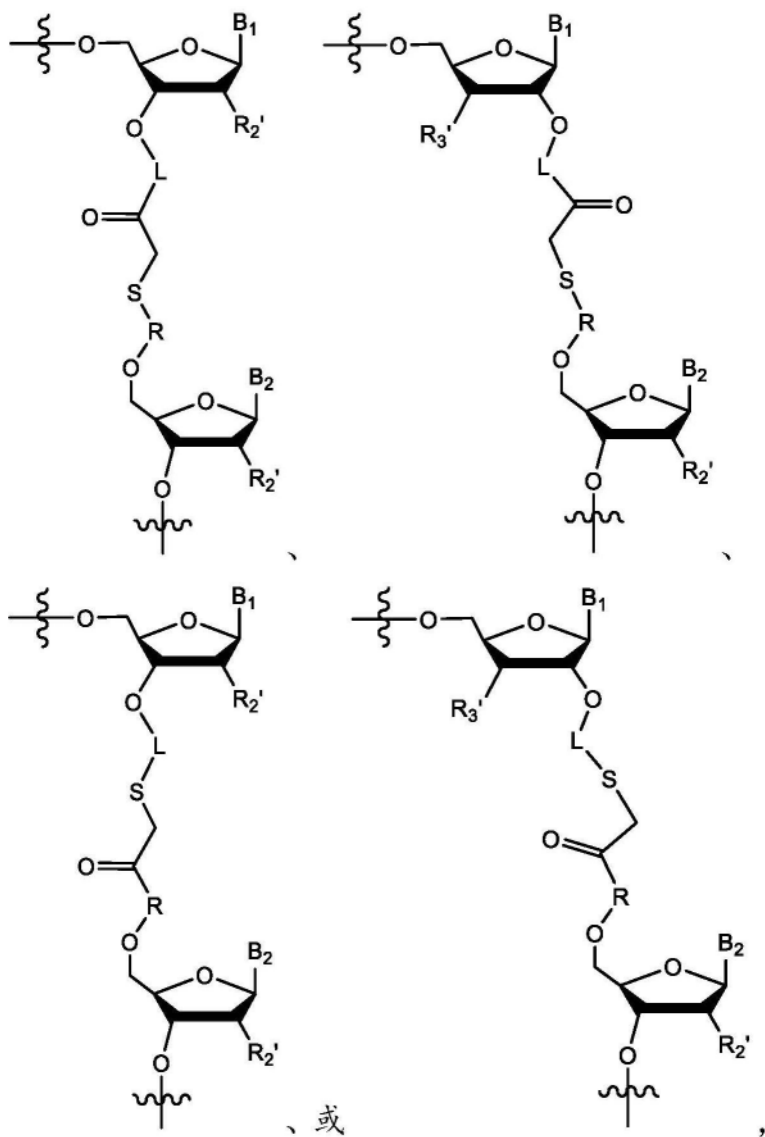
[0164] 在一些实施例中,包含脲的指导分子具有下式:



[0165] 在一些实施例中,包含脲的指导分子具有来自实例部分的表10中列出的序列,其中[UR]是包含脲的非核苷酸键合。在一些实施例中,[UR]表示具有核碱基 $B_1$ 和 $B_2$ 的两个核苷酸之间的以下键合:

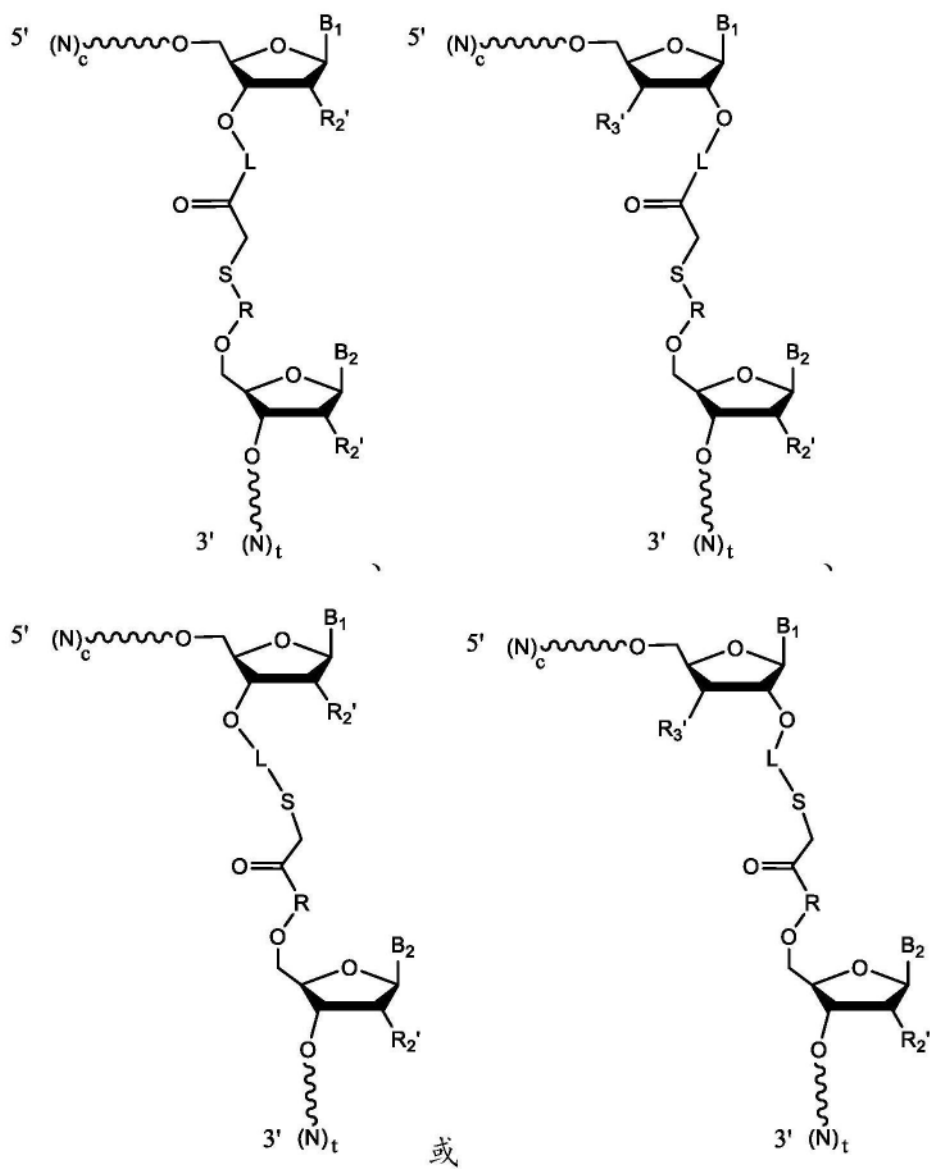


[0166] 在一些实施例中,交联的指导分子的化学键合包硫醚。在一些实施例中,包含硫醚的指导分子具有下式:

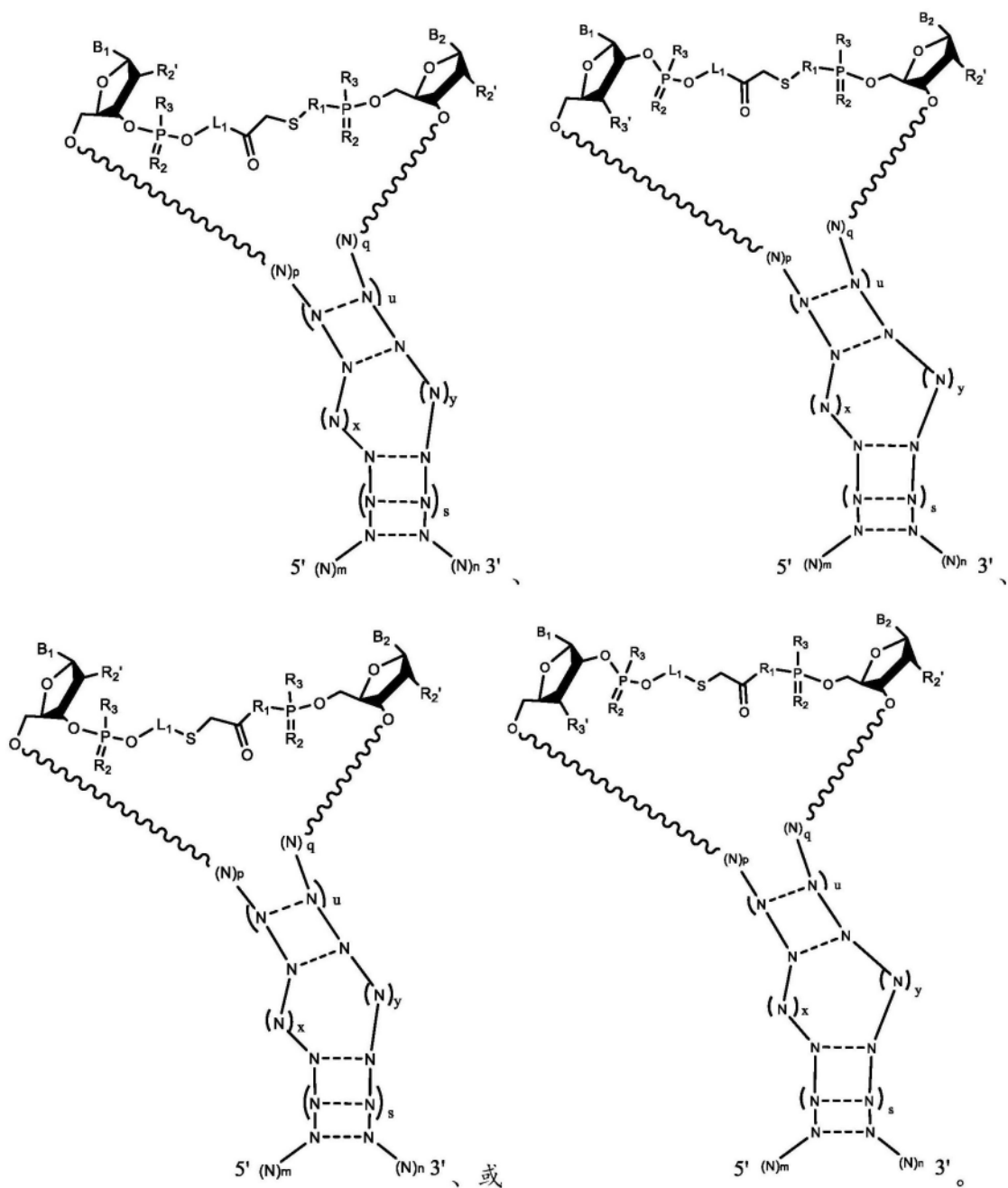


非核苷酸接头。

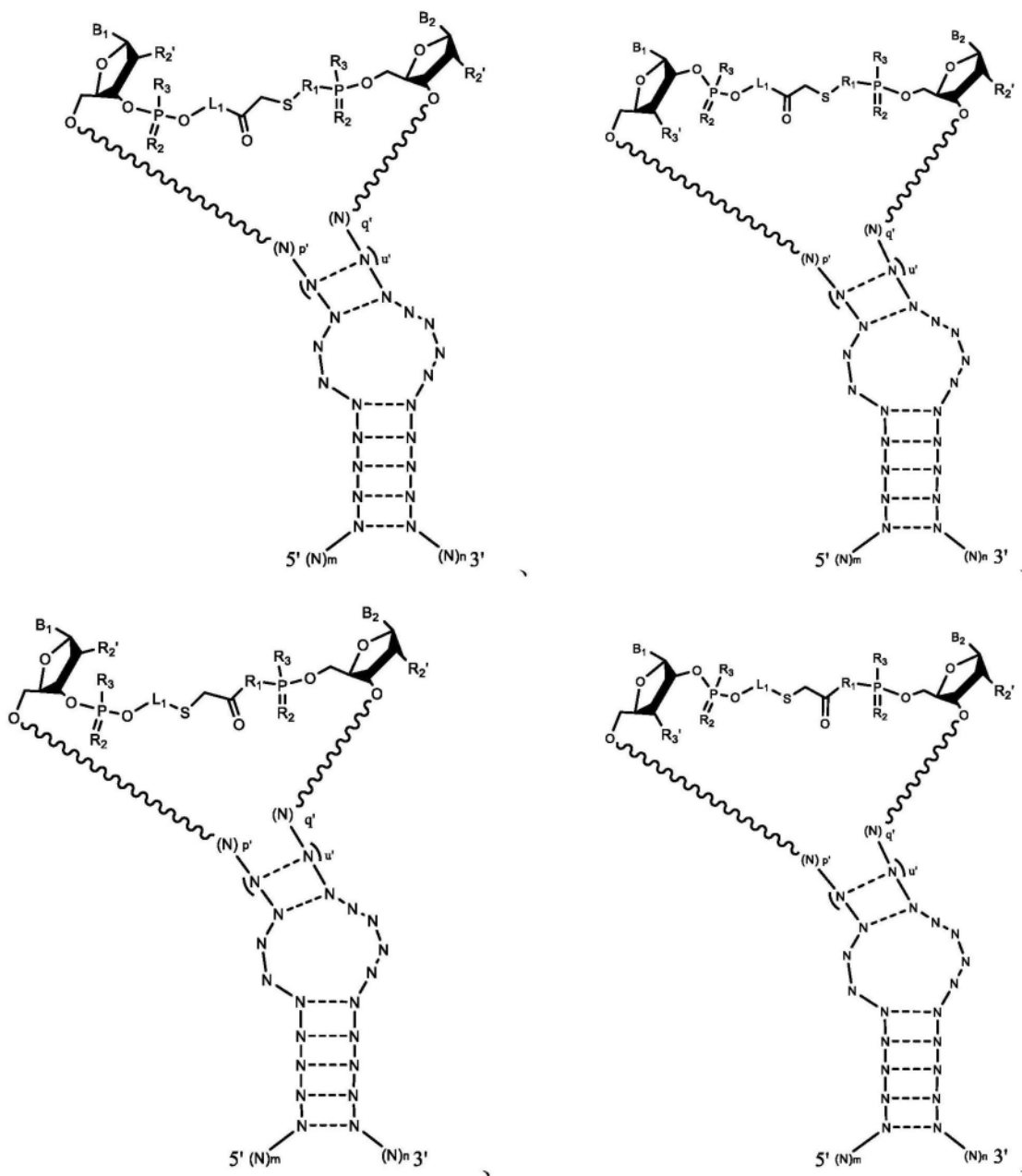
[0167] 在一些实施例中,包含硫醚的指导分子具有下式:

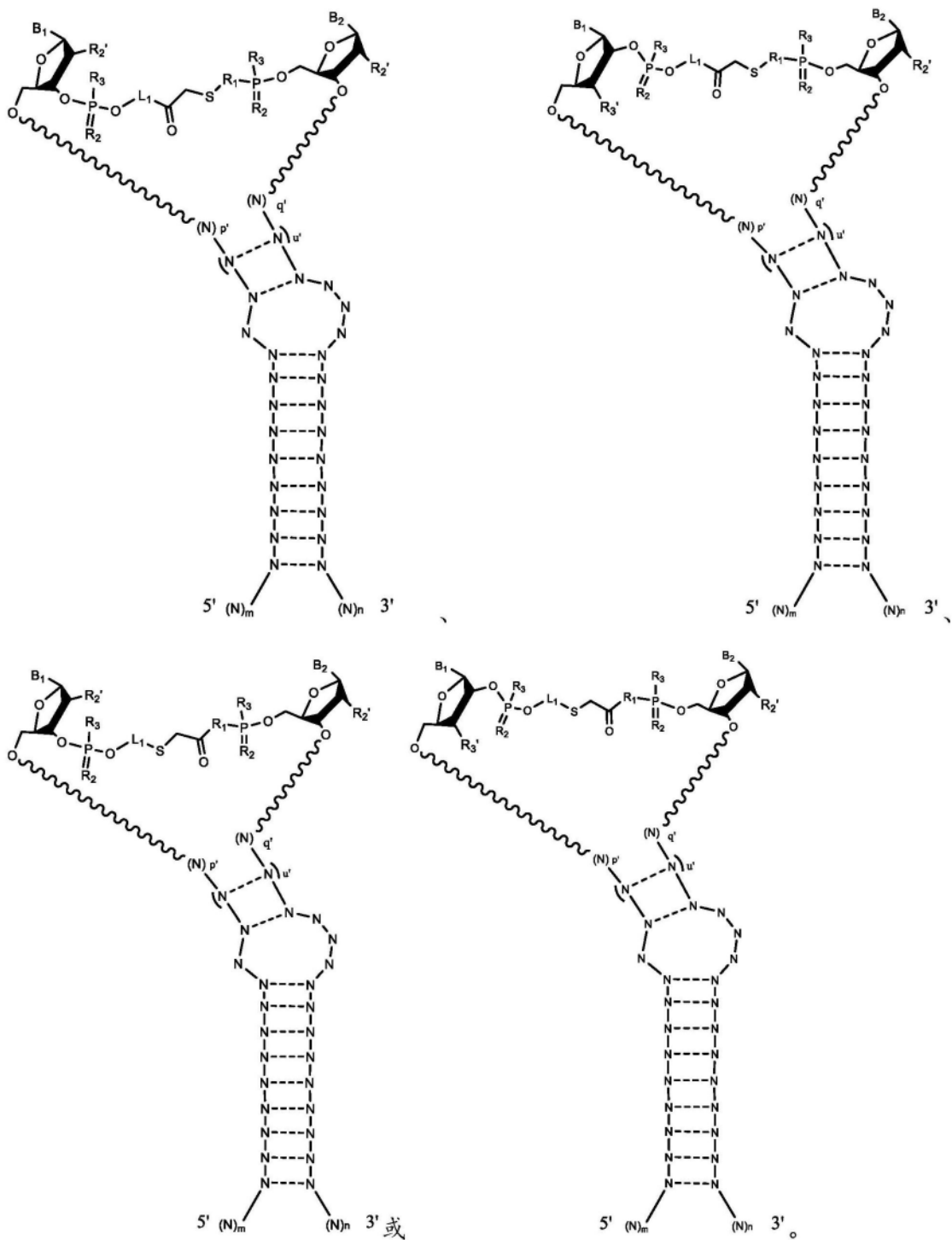


[0168] 在一些实施例中,包含硫醚的指导分子具有下式:

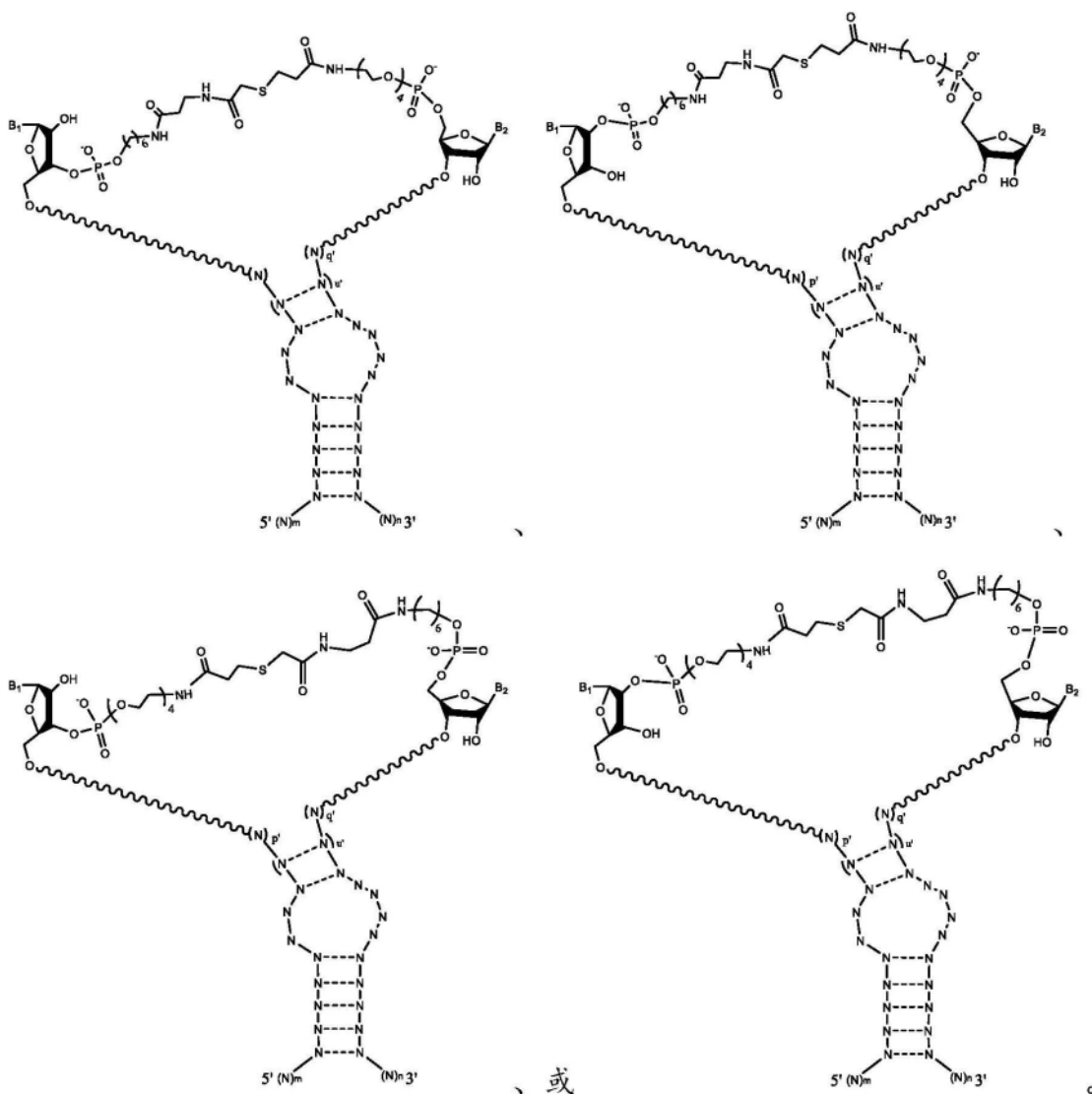


[0169] 在一些实施例中,包含硫醚的指导分子具有下式:

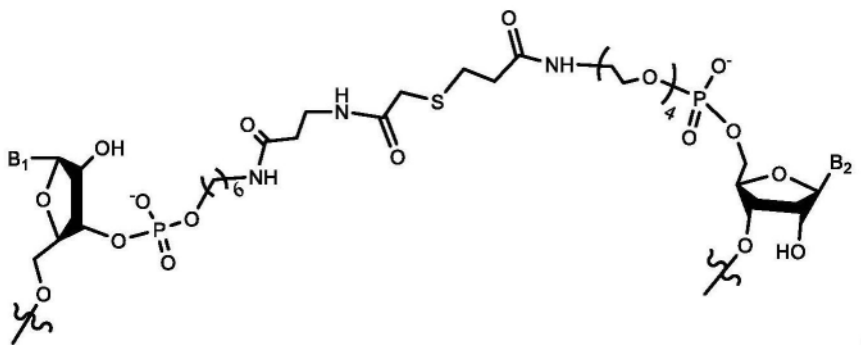




[0170] 在一些实施例中,包含硫醚的指导分子具有下式:



[0171] 在一些实施例中,包含脲的指导分子具有来自实例部分的表9中列出的序列,其中 [L] 是硫醚键合。在一些实施例中, [L] 表示具有核碱基  $B_1$  和  $B_2$  的两个核苷酸之间的以下键合:



[0172] 在一些实施例中,在本申请的任何式中,  $R_2$  和  $R_3$  各自独立地是 H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R' 或 O-R', 其中每个 R' 独立地是保护基团或任选地被取代的烷基。在一些实施例中,  $R_2$  和  $R_3$  各自独立地是 H、OH、卤素、 $NH_2$ 、或 O-R', 其中每个 R' 独立地是保护基团或任选地被取代的烷基。在一些实施例中,  $R_2$  和  $R_3$  各自独立地是 H、氟和 O-R', 其中 R 是保护基团或任

选地被取代的烷基。在一些实施例中, $R_2$  是H。在一些实施例中, $R_3$  是H。在一些实施例中, $R_2$  是卤素。在一些实施例中, $R_3$  是卤素。在一些实施例中, $R_2$  是氟。在一些实施例中, $R_3$  是氟。在一些实施例中, $R_2$  是O-R'。在一些实施例中, $R_3$  是O-R'。在一些实施例中, $R_2$  是O-Me。在一些实施例中, $R_3$  是O-Me。

[0173] 在一些实施例中,在本申请的任何式中,p和q各自独立地是0、1、2、3、4、5或6。在一些实施例中,p和q各自独立地是2。在一些实施例中,p和q各自独立地是0。在一些实施例中,p'和q'各自独立地是0、1、2、3或4。在一些实施例中,p'和q'各自独立地是2。在一些实施例中,p'和q'各自独立地是0。

[0174] 在一些实施例中,在本申请的任何式中,u是2到22之间的整数(包括端值)。在一些实施例中,u是3到22之间的整数(包括端值)。在一些实施例中,u是4到22之间的整数(包括端值)。在一些实施例中,u是8到22之间的整数(包括端值)。在一些实施例中,u是12到22之间的整数(包括端值)。在一些实施例中,u是0到22之间的整数(包括端值)。在一些实施例中,u是2到14之间的整数(包括端值)。在一些实施例中,u是4到14之间的整数(包括端值)。在一些实施例中,u是8到14之间的整数(包括端值)。在一些实施例中,u是0到14之间的整数(包括端值)。在一些实施例中,u是0到4之间的整数(包括端值)。在一些实施例中,在本申请的任何式中,u' 是2到22之间的整数(包括端值)。在一些实施例中,u' 是3到22之间的整数(包括端值)。在一些实施例中,u' 是4到22之间的整数(包括端值)。在一些实施例中,u' 是8到22之间的整数(包括端值)。在一些实施例中,u' 是12到22之间的整数(包括端值)。在一些实施例中,u' 是0到22之间的整数(包括端值)。在一些实施例中,u' 是2到14之间的整数(包括端值)。在一些实施例中,u' 是4到14之间的整数(包括端值)。在一些实施例中,u' 是8到14之间的整数(包括端值)。在一些实施例中,u' 是0到14之间的整数(包括端值)。在一些实施例中,u' 是0到4之间的整数(包括端值)。

[0175] 在一些实施例中,在本申请的任何式中,N独立地是核糖核苷酸,脱氧核糖核苷酸,经修饰的核糖核苷酸或经修饰的脱氧核糖核苷酸。核苷酸修饰在下面讨论。

[0176] 在一些实施例中,在本申请的任何式中,c是20或更大的整数。在一些实施例中,c是20到60之间的整数(包括端值)。在一些实施例中,c是20到40之间的整数(包括端值)。在一些实施例中,c是40到60之间的整数(包括端值)。在一些实施例中,c是30到60之间的整数(包括端值)。在一些实施例中,c是20到50之间的整数(包括端值)。

[0177] 在一些实施例中,在本申请的任何式中,t是20或更大的整数。在一些实施例中,t是20到80之间的整数(包括端值)。在一些实施例中,t是20到50之间的整数(包括端值)。在一些实施例中,t是50到80之间的整数(包括端值)。在一些实施例中,t是20到70之间的整数(包括端值)。在一些实施例中,t是30到80之间的整数(包括端值)。

[0178] 在一些实施例中,在本申请的任何式中,s是1到10之间的整数(包括端值)。在一些实施例中,s是3到9之间的整数(包括端值)。在一些实施例中,s是1到8之间的整数(包括端值)。在一些实施例中,s是0到10之间的整数(包括端值)。在一些实施例中,s是2到6之间的整数(包括端值)。

[0179] 在一些实施例中,在本申请的任何式中,x是1到3之间的整数(包括端值)。在一些实施例中,x是1。在一些实施例中,x是2。在一些实施例中,x是3。在一些实施例中,在本申请的任何式中,y大于x。在一些实施例中,y是3到5之间的整数(包括端值)。在一些实施例中,y

是3。在一些实施例中,y是4。在一些实施例中,y是5。在一些实施例中,x是1且y是3。在一些实施例中,x是2且y是4。

[0180] 在一些实施例中,在本申请的任何式中,m是15或更大的整数。在一些实施例中,m是15到50之间的整数(包括端值)。在一些实施例中,m是16或更大的整数。在一些实施例中,m是17或更大的整数。在一些实施例中,m是18或更大的整数。在一些实施例中,m是19或更大的整数。在一些实施例中,m是20或更大的整数。在一些实施例中,m是20到40之间的整数(包括端值)。在一些实施例中,m是30到50之间的整数(包括端值)。在一些实施例中,m是15到30之间的整数(包括端值)。

[0181] 在一些实施例中,在本申请的任何式中,n是30或更大的整数。在一些实施例中,n是30到70之间的整数(包括端值)。在一些实施例中,n是30到60之间的整数(包括端值)。在一些实施例中,n是40到70之间的整数(包括端值)。

[0182] 在一些实施例中,在本申请的任何式中,L、R、 $L_1$ 和 $R_1$ 各自独立地是非核苷酸接头。在一些实施例中,L、R、 $L_1$ 和 $R_1$ 各自独立地包含选自以下组成的组的部分:聚乙烯、聚丙烯、聚乙二醇和聚丙二醇。在一些实施例中, $L_1$ 和 $R_1$ 各自独立地是 $-(CH_2)_w-$ 、 $-(CH_2)_w-NH-C(O)-(CH_2)_w-NH-$ 、 $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ 、或 $-(CH_2CH_2O)_v-$ ,并且每个w是1-20之间的整数(包括端值),每个v是1-10之间的整数(包括端值)。在一些实施例中, $L_1$ 是 $-(CH_2)_w-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_w-NH-C(O)-(CH_2)_w-NH-$ 。在一些实施例中, $L_1$ 是 $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_6-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_6-NH-C(O)-(CH_2)_1-NH-$ 。在一些实施例中, $L_1$ 是 $-(OCH_2CH_2)_4-NH-C(O)-(CH_2)_2-$ 。在一些实施例中, $R_1$ 是 $-(CH_2CH_2O)_v-$ 。在一些实施例中, $R_1$ 是 $-(CH_2)_w-NH-C(O)-(CH_2)_w-NH-$ 。在一些实施例中, $R_1$ 是 $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ 。在一些实施例中, $R_1$ 是 $-(CH_2CH_2O)_4-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_6-NH-C(O)-(CH_2)_1-NH-$ 。在一些实施例中, $R_1$ 是 $-(OCH_2CH_2)_4-NH-C(O)-(CH_2)_2-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_6-$ 并且 $R_1$ 是 $-(CH_2CH_2O)_4-$ 。在一些实施例中, $L_1$ 是 $-(CH_2)_6-NH-C(O)-(CH_2)_1-NH-$ 并且 $R_1$ 是 $-(OCH_2CH_2)_4-NH-C(O)-(CH_2)_2-$ 。

[0183] 在一些实施例中,在本申请的任何式中, $R_2$ 是O,并且在一些实施例中, $R_2$ 是S。在一些实施例中, $R_3$ 是 $O^-$ ,并且在一些实施例中, $R_3$ 是 $COO^-$ 。在一些实施例中, $R_2$ 是O并且 $R_3$ 是 $O^-$ 。在一些实施例中, $R_2$ 是O并且 $R_3$ 是 $COO^-$ 。在一些实施例中, $R_2$ 是S并且 $R_3$ 是 $O^-$ 。在一些实施例中, $R_2$ 是S并且 $R_3$ 是 $COO^-$ 。本领域技术人员将认识到 $R_3$ 也可以以质子化形式(OH和COOH)存在。在整个申请中,我们打算包括 $R_3$ 的去质子化和质子化形式。

[0184] 在一些实施例中,在本申请的任何式中,每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。在一些实施例中,所有N---N代表氢键碱基配对的两个互补核苷酸。在一些实施例中,一些N---N代表两个互补核苷酸,一些N---N代表氢键碱基配对的两个互补核苷酸。

[0185] 在一些实施例中,在本申请的任何式中, $B_1$ 和 $B_2$ 各自独立地是核碱基。在一些实施例中, $B_1$ 是鸟嘌呤并且 $B_2$ 是胞嘧啶。在一些实施例中, $B_1$ 是胞嘧啶并且 $B_2$ 是鸟嘌呤。在一些实施例中, $B_1$ 是腺嘌呤并且 $B_2$ 是尿嘧啶。在一些实施例中, $B_1$ 是尿嘧啶并且 $B_2$ 是腺嘌呤。在一些实施例中, $B_1$ 和 $B_2$ 是互补的。在一些实施例中, $B_1$ 和 $B_2$ 是互补的并通过氢键碱基配对。在一些实施例中, $B_1$ 和 $B_2$ 是互补的,而不是通过氢键碱基配对。在一些实施例中, $B_1$ 和 $B_2$ 不是互补的。

### 指导分子的合成

[0186] 本发明的另一方面是合成单分子指导分子的方法,所述方法包括以下步骤:

使第一寡核苷酸和第二寡核苷酸退火以在所述第一寡核苷酸的3'区和所述第二寡核苷酸的5'区之间形成双链体,其中所述第一寡核苷酸包含第一反应性基团,所述第一反应性基团是2'反应性基团和3'反应性基团中的至少一个,并且其中所述第二寡核苷酸包含第二反应性基团,所述第二反应性基团是5'反应性基团;和

通过所述第一和第二反应性基团缀合所述退火的第一和第二寡核苷酸以形成单分子指导分子,所述单分子指导分子包含连接所述第一和第二寡核苷酸的共价键。

[0187] 在一些实施例中,第一反应性基团和第二反应性基团选自上文“概述”下列出的官能团。在一些实施例中,第一反应性基团和第二反应性基团各自独立地是胺部分、巯基部分、溴乙酰基部分、羟基部分或磷酸酯部分。在一些实施例中,第一反应性基团和第二反应性基团都是胺部分。在一些实施例中,第一反应性基团是巯基部分并且第二反应性基团是溴乙酰基部分。在一些实施例中,第一反应性基团是溴乙酰基部分并且第二反应性基团是巯基部分。在一些实施例中,第一反应性基团是羟基部分并且第二反应性基团是磷酸酯部分。在一些实施例中,第一反应性基团是磷酸酯部分并且第二反应性基团是羟基部分。

[0188] 在一些实施例中,缀合步骤包含 $10\mu\text{M}$ 至 $1\text{mM}$ 范围内的第一核苷酸浓度。在一些实施例中,缀合步骤包含 $10\mu\text{M}$ 至 $1\text{mM}$ 范围内的第二核苷酸浓度。在一些实施例中,第一或第二核苷酸的浓度是 $10\mu\text{M}$ 、 $50\mu\text{M}$ 、 $100\mu\text{M}$ 、 $200\mu\text{M}$ 、 $400\mu\text{M}$ 、 $600\mu\text{M}$ 、 $800\mu\text{M}$ 、或 $1\text{mM}$ 。

[0189] 在一些实施例中,缀合步骤包括 $5.0$ 至 $9.0$ 范围内的pH。在一些实施例中,pH是 $5.0$ 、 $5.5$ 、 $6.0$ 、 $6.5$ 、 $7.0$ 、 $7.5$ 、 $8.0$ 、 $8.5$ 、或 $9.0$ 。在一些实施例中,pH为 $6.0$ 。在一些实施例中,pH为 $8.0$ 。在一些实施例中,pH为 $8.5$ 。

[0190] 在一些实施例中,缀合步骤在氩气下进行。在一些实施例中,缀合步骤在环境大气下进行。

[0191] 在一些实施例中,缀合步骤在水中进行。在一些实施例中,缀合步骤在具有共溶剂的水中进行。在一些实施例中,共溶剂是DMSO、DMF、NMP、DMA、吗啉、吡啶或MeCN。在一些实施例中,共溶剂是DMSO。在一些实施例中,共溶剂是DMF。

[0192] 在一些实施例中,缀合步骤在 $0^{\circ}\text{C}$ 至 $40^{\circ}\text{C}$ 范围内的温度下进行。在一些实施例中,温度是 $0^{\circ}\text{C}$ 、 $4^{\circ}\text{C}$ 、 $10^{\circ}\text{C}$ 、 $20^{\circ}\text{C}$ 、 $25^{\circ}\text{C}$ 、 $30^{\circ}\text{C}$ 、 $37^{\circ}\text{C}$ 、或 $40^{\circ}\text{C}$ 。在一些实施例中,温度是 $25^{\circ}\text{C}$ 。在一些实施例中,温度是 $4^{\circ}\text{C}$ 。

[0193] 在一些实施例中,缀合步骤在二价金属阳离子的存在下进行。在一些实施例中,二价金属阳离子是 $\text{Mg}^{2+}$ 、 $\text{Ca}^{2+}$ 、 $\text{Sr}^{2+}$ 、 $\text{Ba}^{2+}$ 、 $\text{Cr}^{2+}$ 、 $\text{Mn}^{2+}$ 、 $\text{Fe}^{2+}$ 、 $\text{Co}^{2+}$ 、 $\text{Ni}^{2+}$ 、 $\text{Cu}^{2+}$ 或 $\text{Zn}^{2+}$ 。在一些实施例中,二价金属阳离子是 $\text{Mg}^{2+}$ 。

[0194] 在一些实施例中,缀合步骤包括交联剂或交联接头(参见上文“概述”)。在一些实施例中,交联接头是多官能的,并且在一些实施例中,交联接头是双官能的。在一些实施例中,多官能交联接头是异型官能或同型官能的。在一些实施例中,交联接头含有碳酸酯。在一些实施例中,含碳酸酯的交联接头是二琥珀酰亚胺基碳酸酯、二咪唑碳酸酯或双-(对硝基苯基)碳酸酯。在一些实施例中,含碳酸酯的交联接头是二琥珀酰亚胺基碳酸酯。

[0195] 在一些实施例中,缀合步骤包括浓度为 $1\text{mM}$ 至 $100\text{mM}$ 的双官能交联剂。在一些实施例中,双官能交联剂的浓度是 $1\text{mM}$ 、 $10\text{mM}$ 、 $20\text{mM}$ 、 $40\text{mM}$ 、 $60\text{mM}$ 、 $80\text{mM}$ 、或 $100\text{mM}$ 。在一些实施例中,

双官能交联剂的浓度比第一和第二寡核苷酸中的每一种的浓度大100至1000倍。在一些实施例中,双官能交联剂的浓度比第一寡核苷酸的浓度大100、200、400、600、800或1000倍。在一些实施例中,双官能交联剂的浓度比第二寡核苷酸的浓度大100、200、400、600、800或1000倍。

[0196] 在一些实施例中,缀合步骤在螯合剂的存在下进行。在一些实施例中,螯合剂是乙二胺四乙酸(EDTA)或其盐。

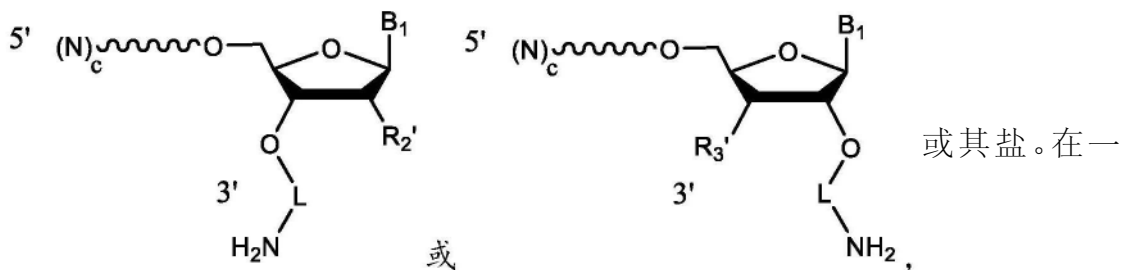
[0197] 在一些实施例中,缀合步骤在活化剂的存在下进行。在一些实施例中,活化剂是碳二亚胺或其盐。在一些实施例中,碳二亚胺是1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)、N,N'-二环己基碳二亚胺(DCC)或N,N'-二异丙基碳二亚胺(DIC)或其盐。在一些实施例中,碳二亚胺是1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)或其盐。

[0198] 在一些实施例中,缀合步骤包含1mM至100mM范围内的活化剂浓度。在一些实施例中,活化剂的浓度是1mM、10mM、20mM、40mM、60mM、80mM、或100mM。在一些实施例中,活化剂的浓度比第一和第二寡核苷酸中的每一种的浓度大100至1000倍。在一些实施例中,活化剂的浓度比第一寡核苷酸的浓度大100、200、400、600、800或1000倍。在一些实施例中,活化剂的浓度比第二寡核苷酸浓度大100、200、400、600、800或1000倍。

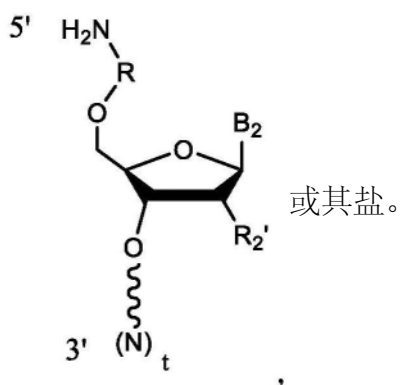
[0199] 在一些实施例中,缀合步骤在稳定剂的存在下进行。在一些实施例中,稳定剂是咪唑、氰基咪唑、吡啶或二甲基氨基吡啶、或其盐。在一些实施例中,稳定剂是咪唑。在一些实施例中,缀合步骤在活化剂和稳定剂的存在下进行。在一些实施例中,缀合步骤在1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)和咪唑或其盐的存在下进行。

[0200] 在一些实施例中,合成单分子指导分子的方法产生具有上文披露的任何式的指导分子。

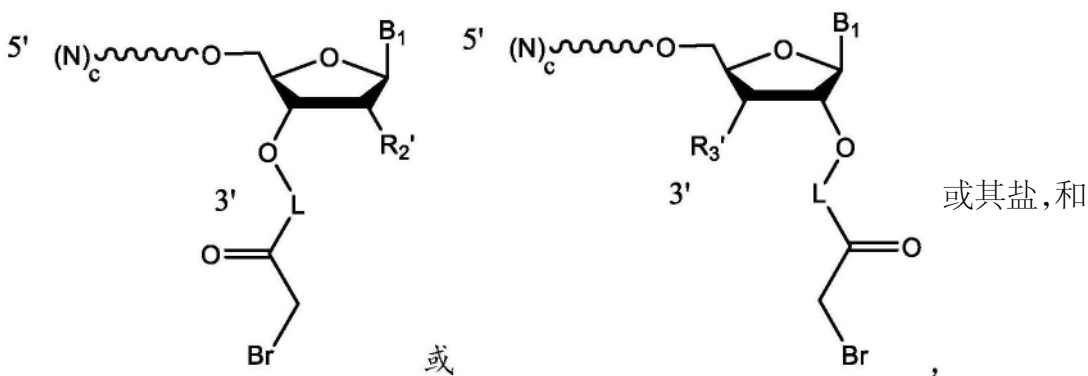
[0201] 在一些实施例中,合成单分子指导分子的方法产生具有脲接头的指导分子。在一些实施例中,第一反应性基团和第二反应性基团均为胺,并且第一和第二反应性基团与含碳酸酯的双官能交联剂交联以形成脲接头。在一些实施例中,含碳酸酯的双官能交联剂是二琥珀酰亚胺基碳酸酯。在一些实施例中,所述方法包括具有下式的第一寡核苷酸:



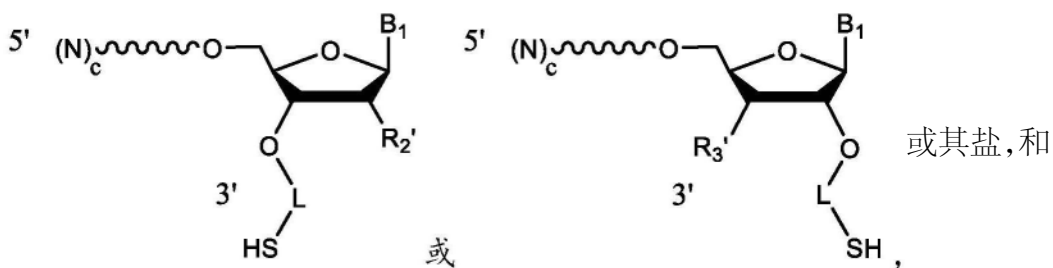
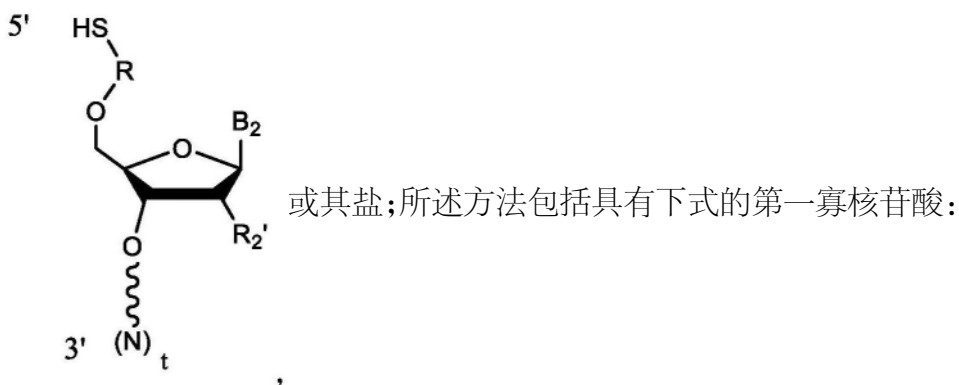
些实施例中,所述方法包括具有下式的第二寡核苷酸:



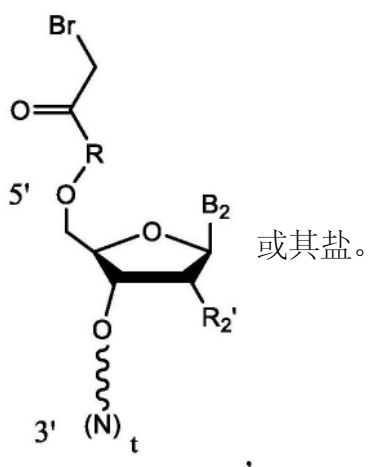
[0202] 在一些实施例中,合成单分子指导分子的方法产生具有硫醚接头的指导分子。在一些实施例中,第一反应性基团是巯基并且第二反应性基团是溴乙酰基,或第一反应性基团是溴乙酰基并且第二反应性基团是巯基。在一些实施例中,第一反应性基团和第二反应性基团在螯合剂存在下反应以形成硫醚键合。在一些实施例中,第一反应性基团和第二反应性基团经历取代反应以形成硫醚键合。在一些实施例中,所述方法包括具有下式的第一寡核苷酸:



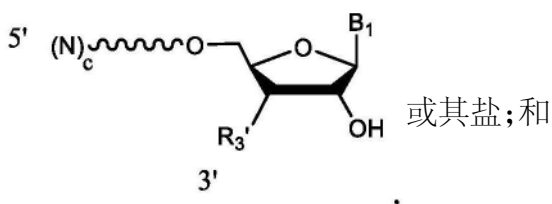
第二寡核苷酸具有下式:



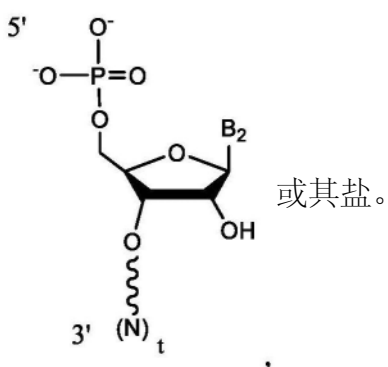
第二寡核苷酸具有下式:



[0203] 在一些实施例中,合成单分子指导分子的方法产生具有磷酸二酯接头的指导分子。在一些实施例中,第一反应性基团包含2'或3'羟基并且第二反应性基团包含5'磷酸部分。在一些实施例中,第一和第二反应性基团在活化剂存在下缀合以形成磷酸二酯接头。在一些实施例中,活化剂是EDC。在一些实施例中,所述方法包括具有下式的第一寡核苷酸:



第二寡核苷酸具有下式:

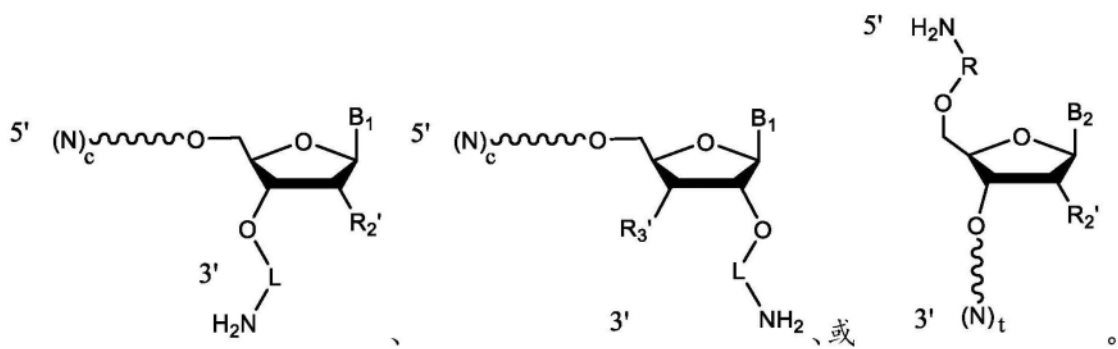


[0204] 在一些实施例中,合成单分子指导分子的方法产生单分子指导分子,其在双链体区中具有至少一个2'-5'磷酸二酯键合。

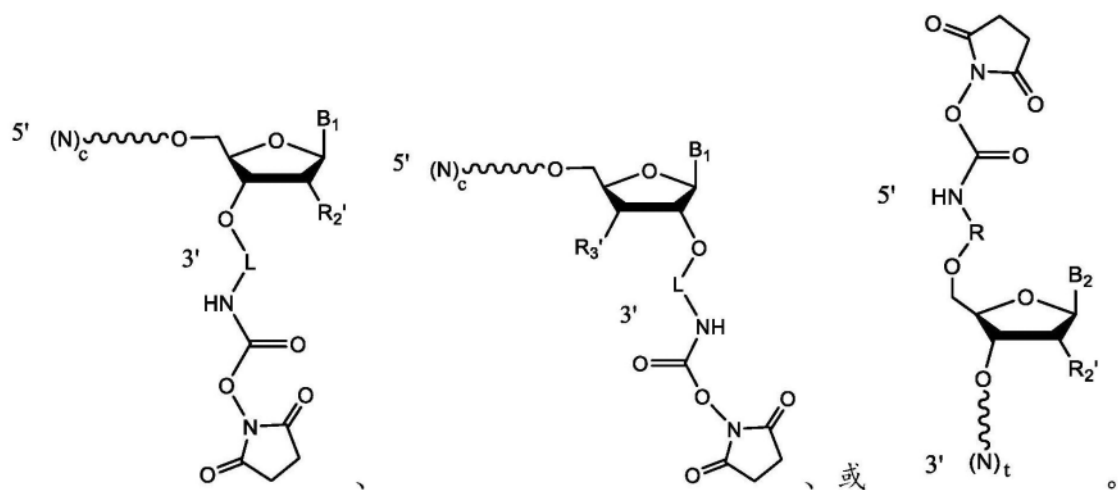
#### 寡核苷酸中间体

[0205] 本披露的某些实施例涉及可用于合成交联的合成指导分子的寡核苷酸中间体。在一些实施例中,寡核苷酸中间体可用于合成包含脲键合,硫醚键合或磷酸二酯键合的指导分子。在一些实施例中,寡核苷酸中间体包含退火的双链体。

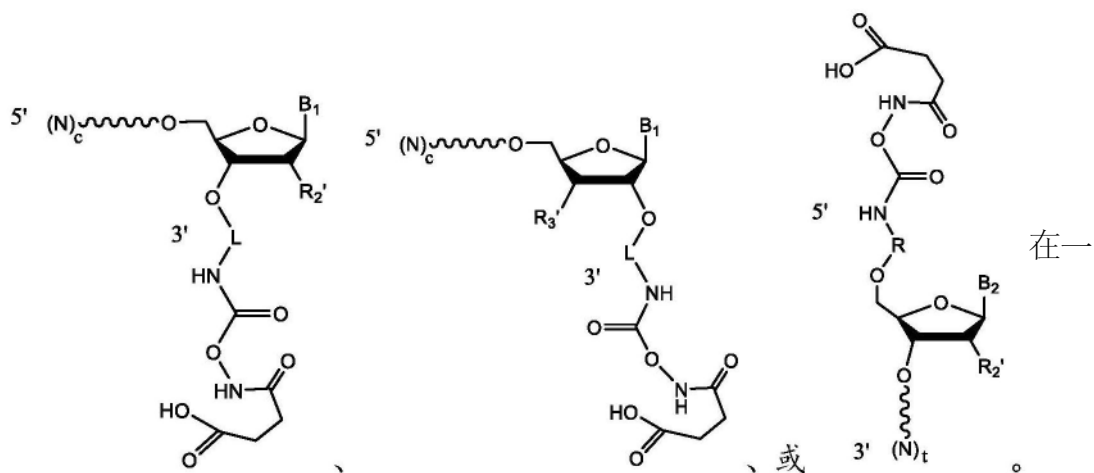
[0206] 在某些实施例中,寡核苷酸中间体可用于合成包含脲键合的指导分子。在一些实施例中,寡核苷酸中间体具有下式:



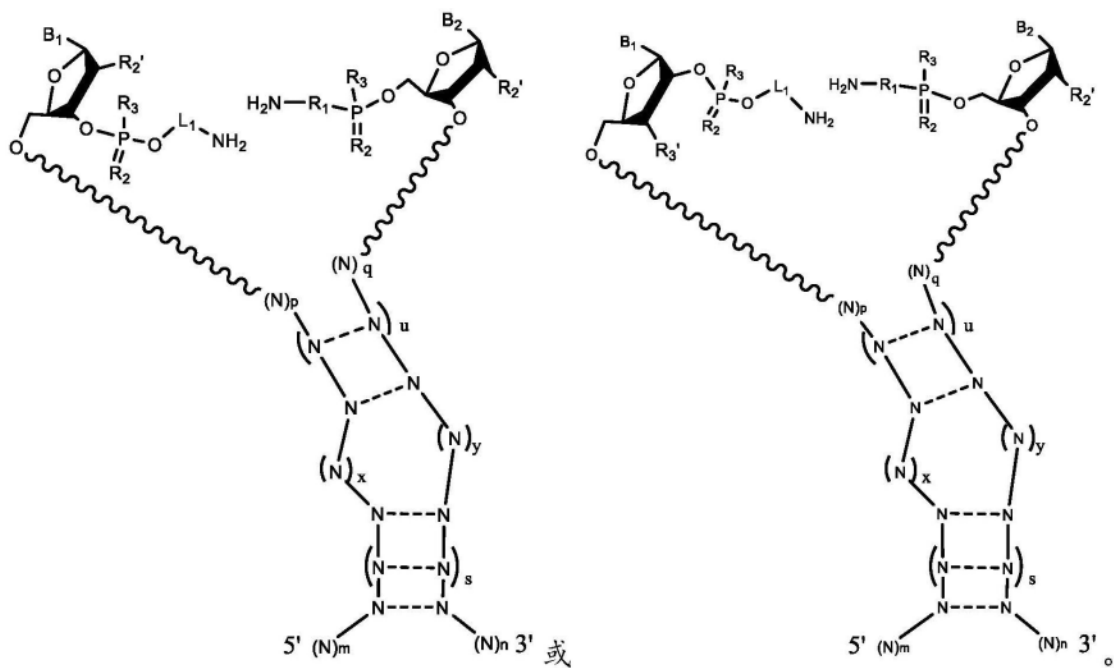
在一些实施例中,寡核苷酸中间体具有下式:



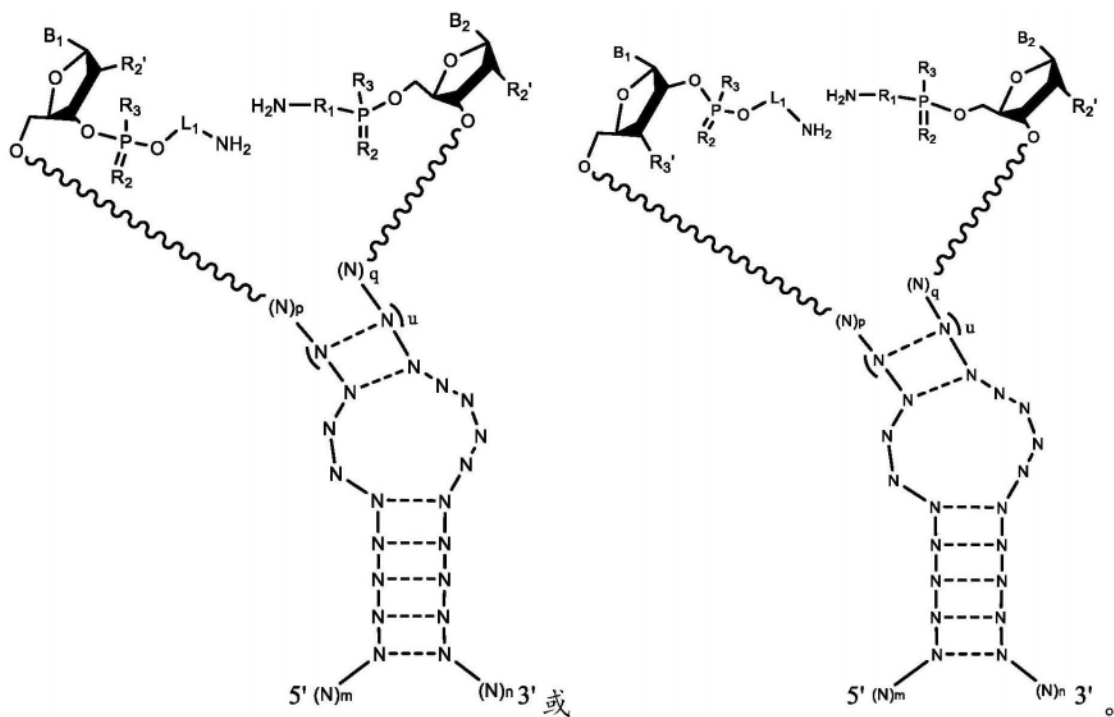
在一些实施例中,寡核苷酸中间体具有下式:



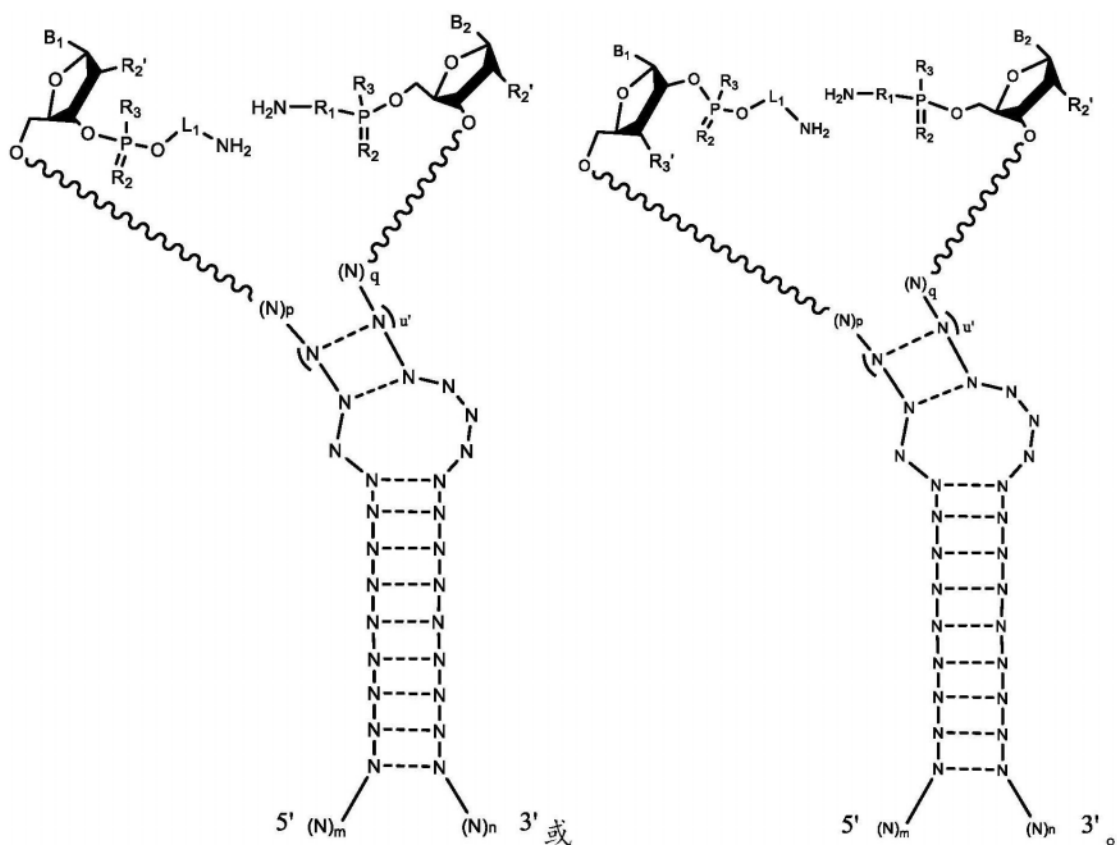
些实施例中,寡核苷酸中间体具有下式:



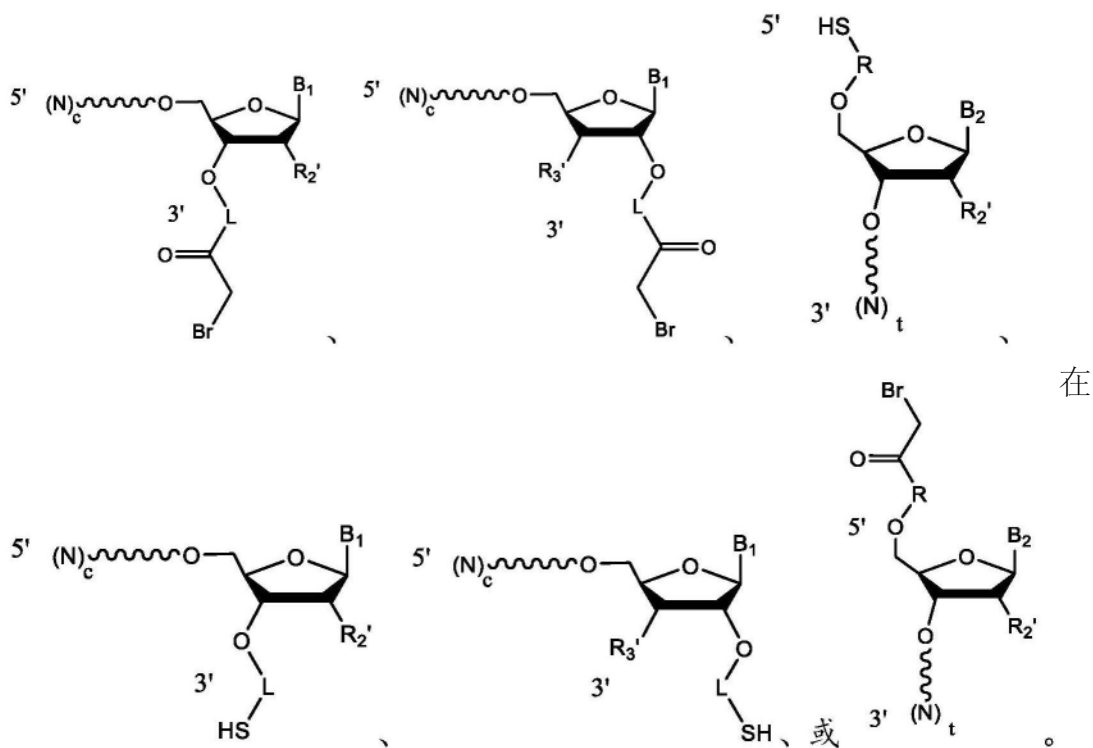
在一些实施例中,寡核苷酸中间体具有下式:



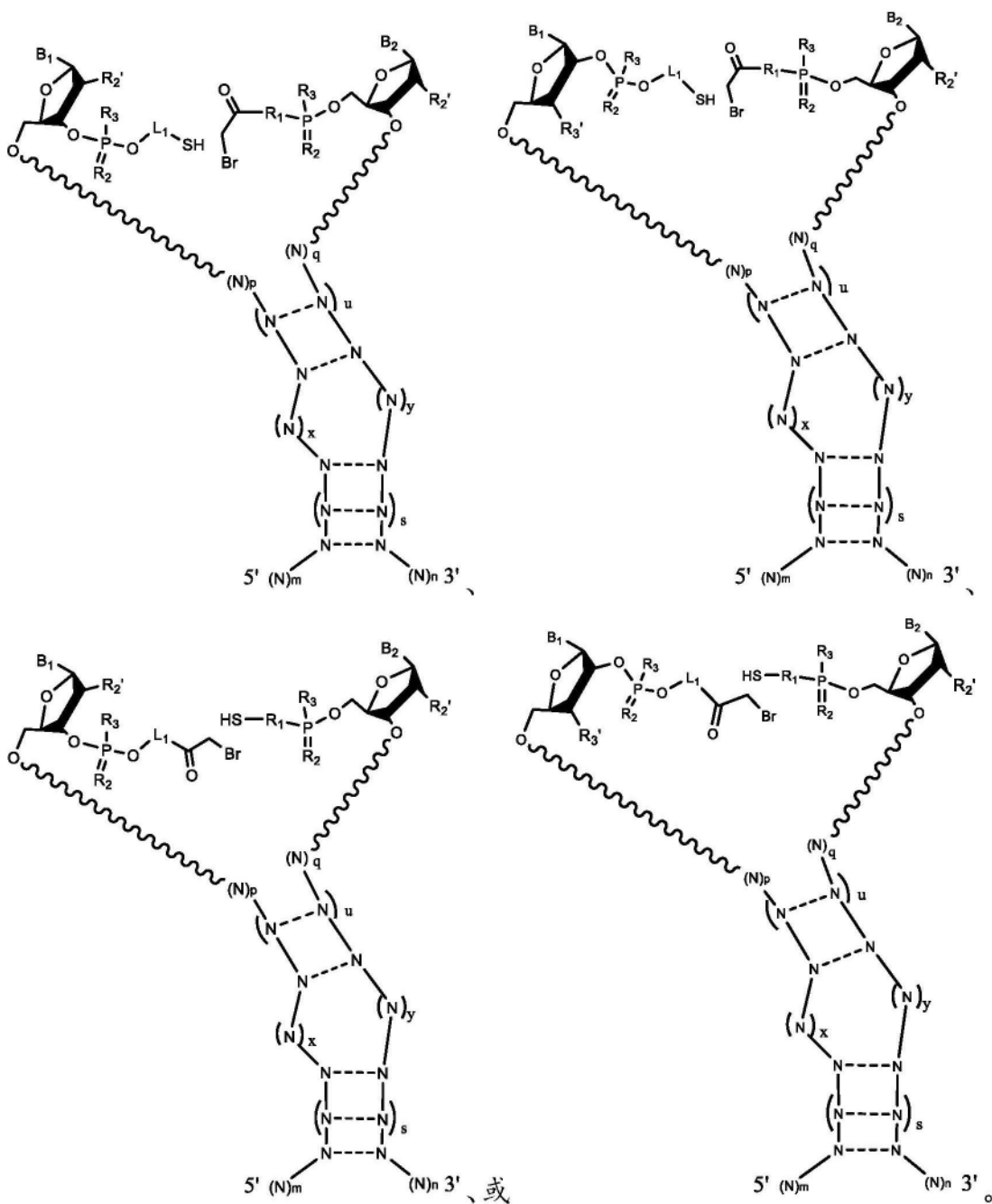
在一些实施例中,寡核苷酸中间体具有下式:



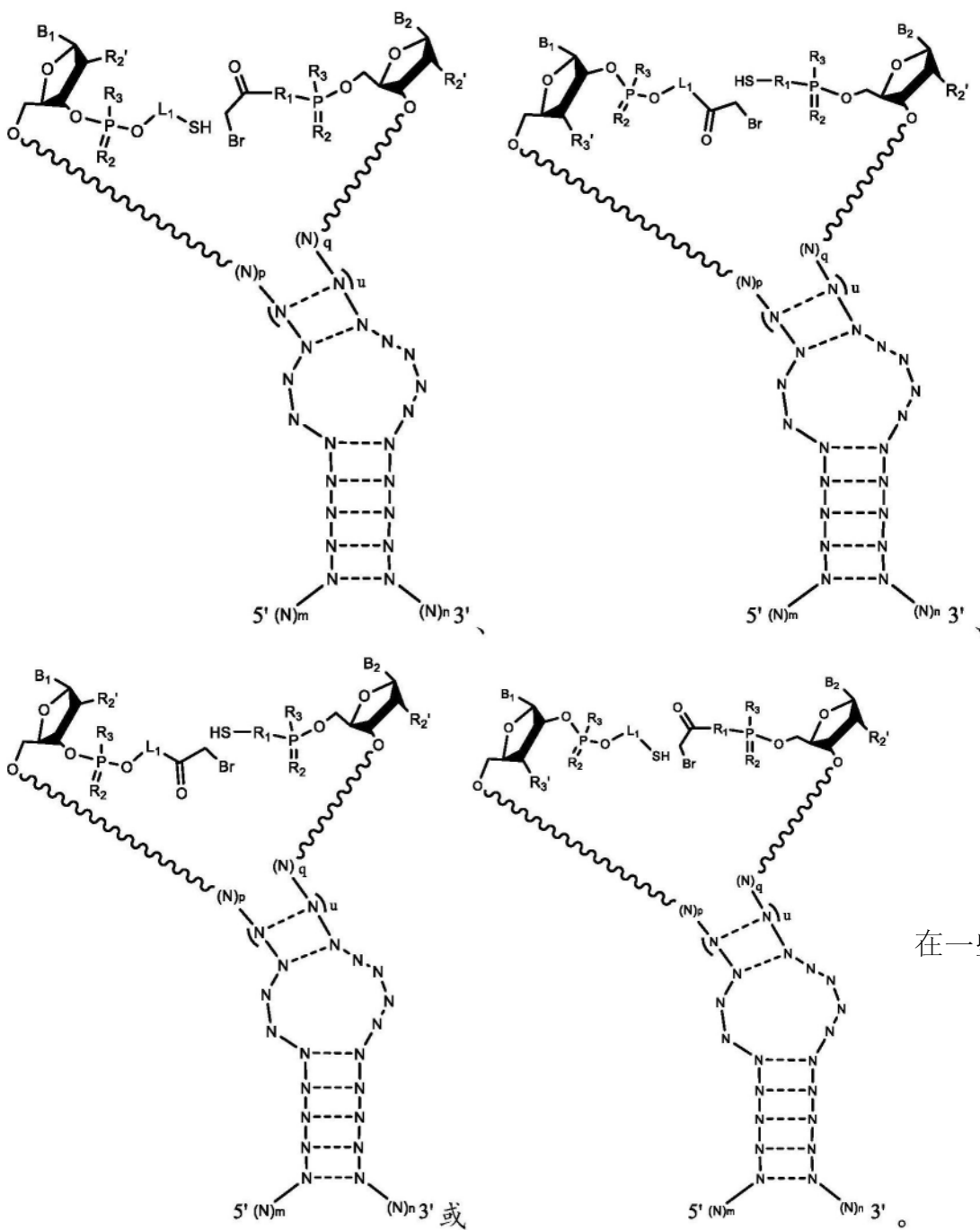
[0207] 在某些实施例中,寡核苷酸中间体可用于合成包含硫醚键合的指导分子。在一些实施例中,寡核苷酸中间体具有下式:



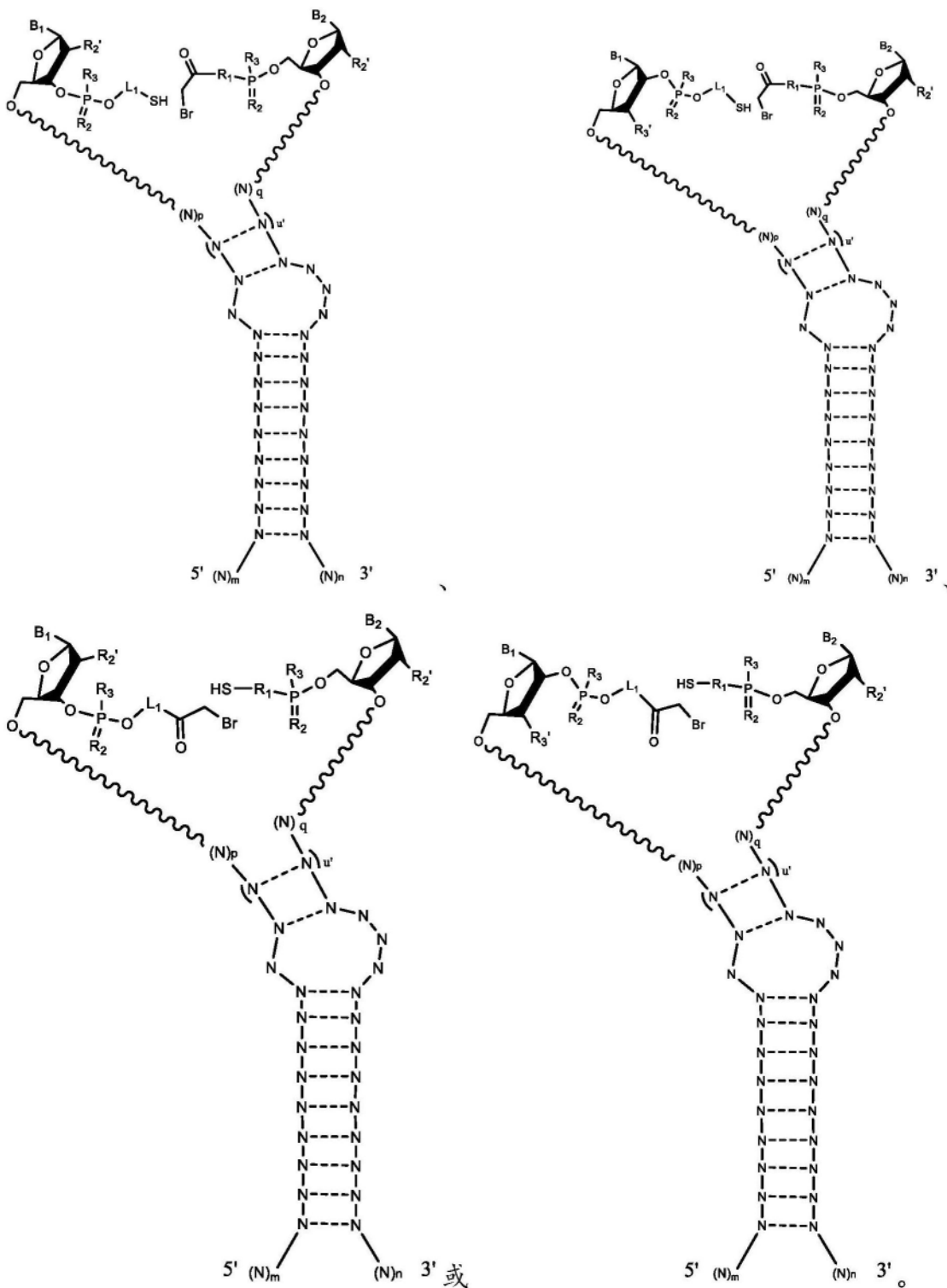
一些实施例中,寡核苷酸中间体具有下式:



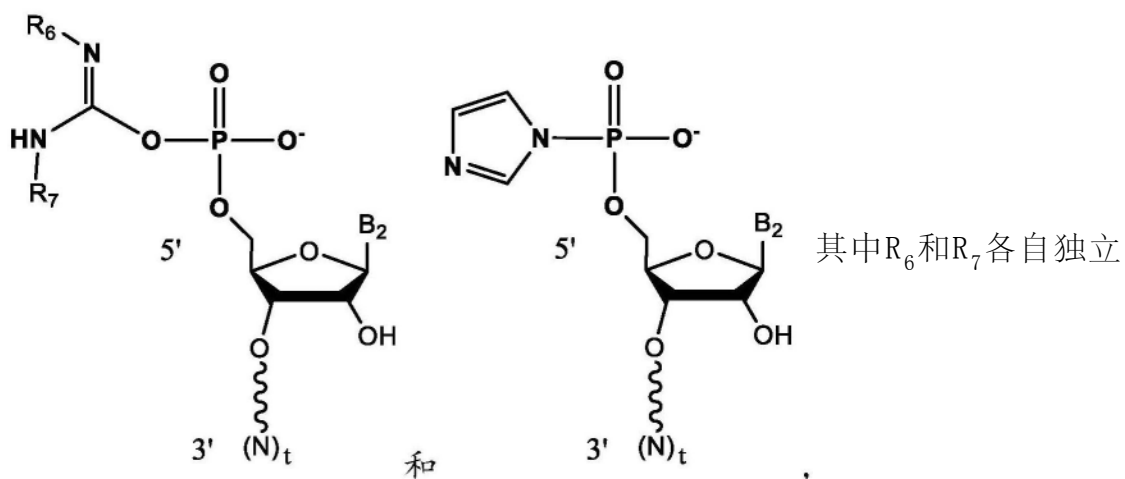
在一些实施例中,寡核苷酸中间体具有下式:



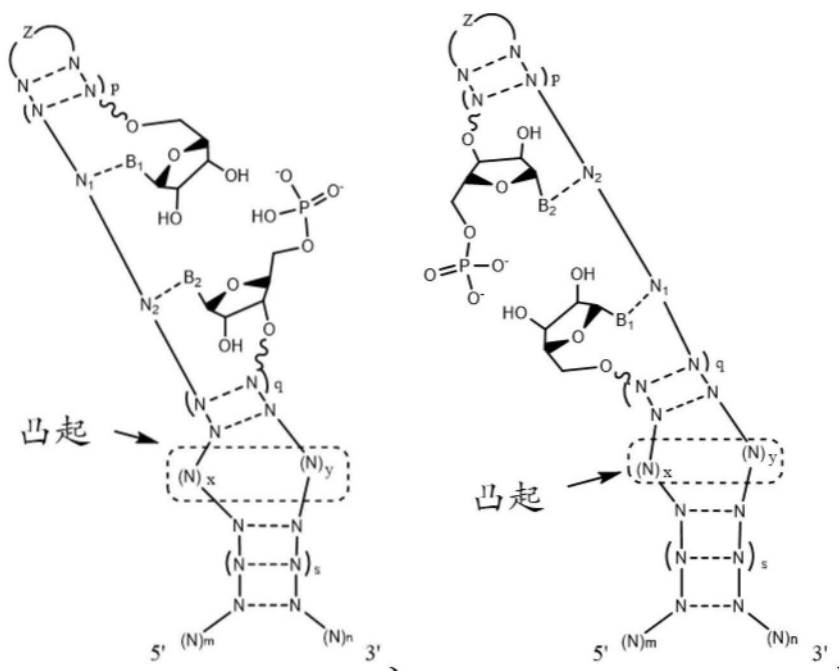
实施例中,寡核苷酸中间体具有下式:

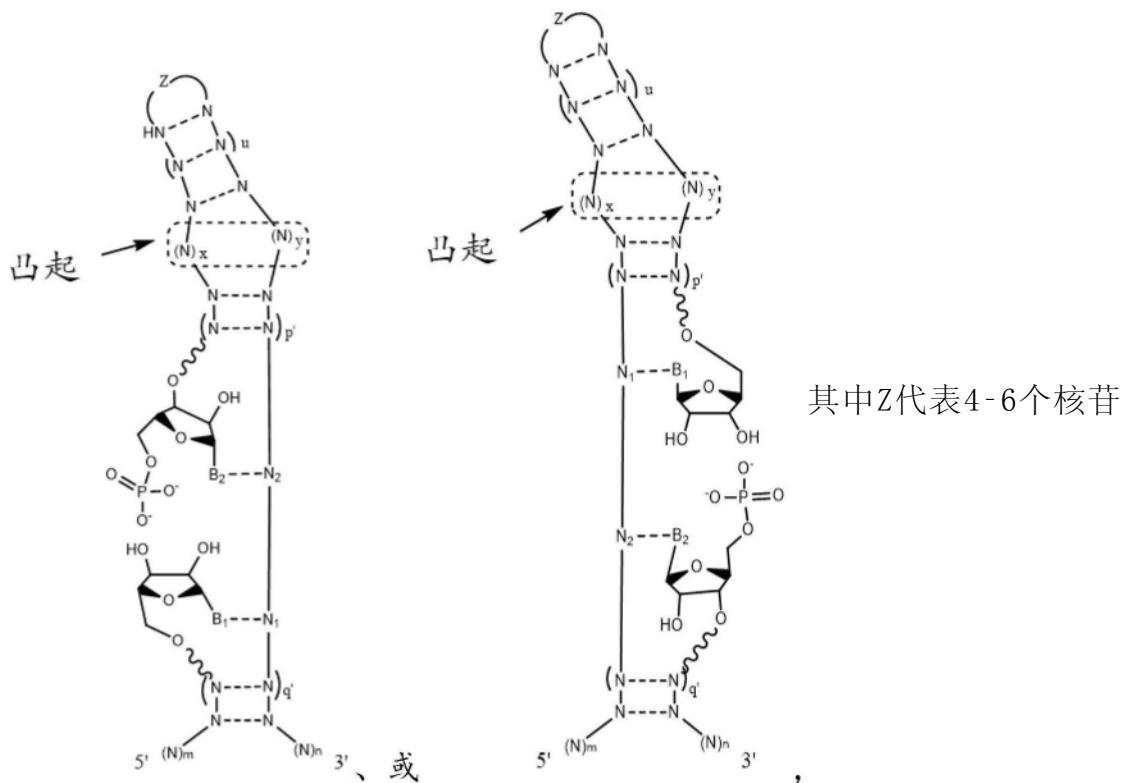


[0208] 在某些实施例中,寡核苷酸中间体可用于合成包含磷酸二酯键合的指导分子。在一些实施例中,寡核苷酸中间体具有下式:



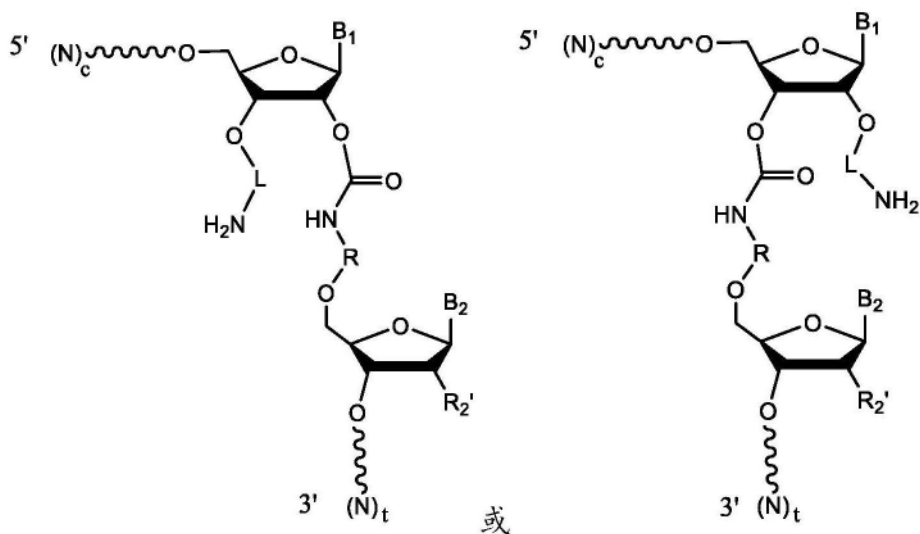
地是取代的或未被取代的烷基,或取代的或未被取代的碳环。在一些实施例中,寡核苷酸中间体具有下式:





酸长、任选地4或6个核苷酸长的核苷酸环。

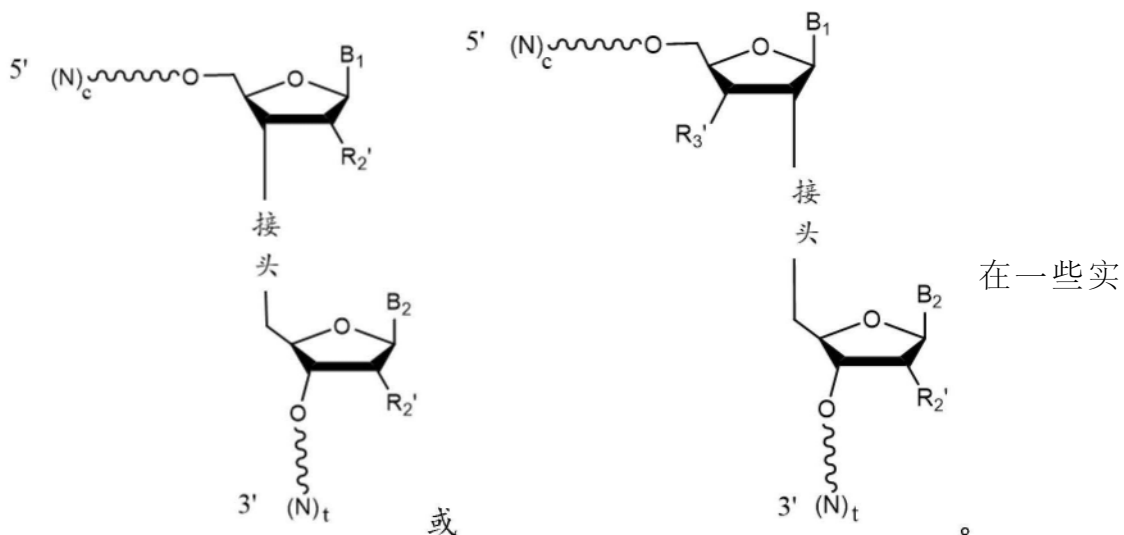
[0209] 本披露的某些实施例涉及在交联反应中作为副产物形成的寡核苷酸化合物。这些寡核苷酸化合物可以用作或不用作指导分子。在一些实施例中,寡核苷酸化合物具有下式:



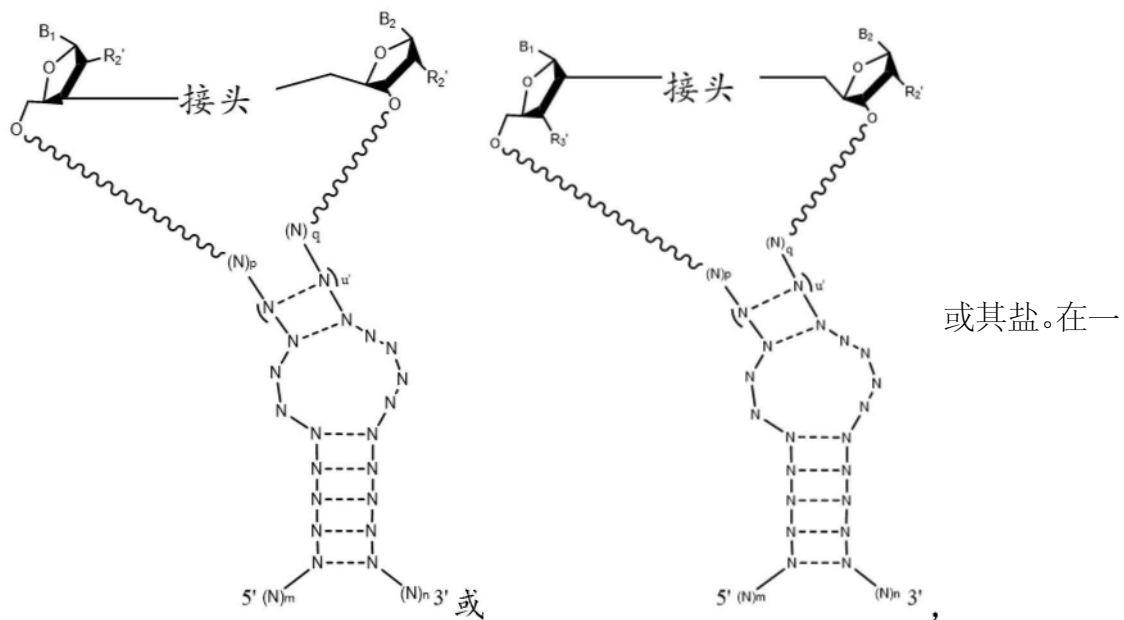
## 化学缀合的指导分子的组合物

[0210] 本披露的某些实施例涉及包含上述合成指导分子的组合物和通过上述方法产生的组合物。在一些实施例中,组合物的特征在于组合物中多于90%的指导分子是全长指导分子。在一些实施例中,组合物的特征在于组合物中多于85%的指导分子包含相同的靶向结构域序列。

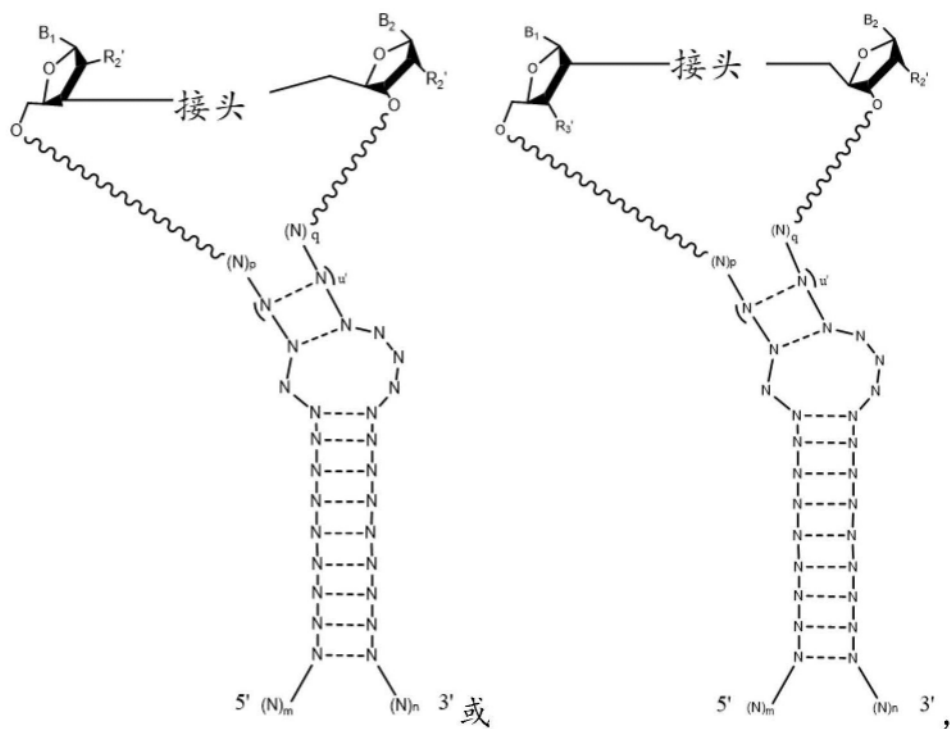
[0211] 在一些实施例中,组合物未经过纯化步骤。在一些实施例中,CRISPR系统的指导分子的组合物基本上由具有下式的指导分子组成:



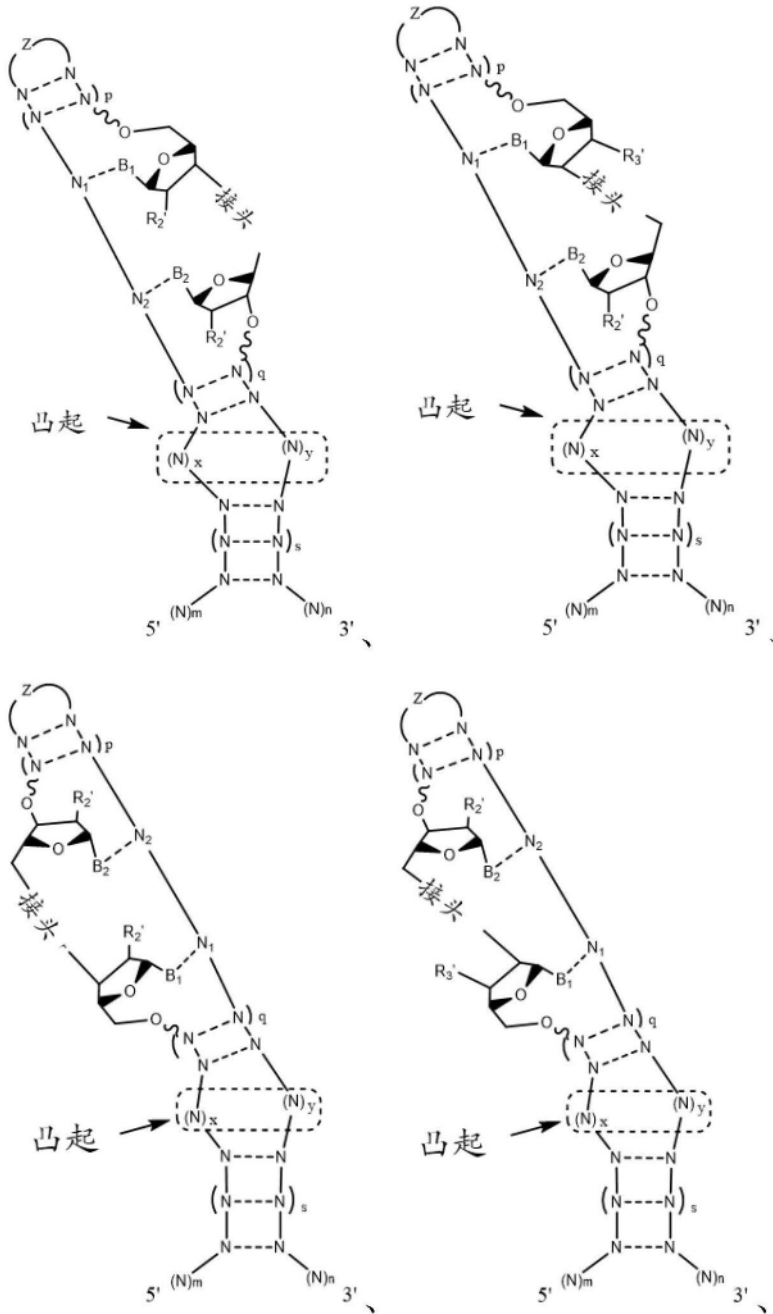
实施例中,组合物基本上由具有下式的指导分子组成:

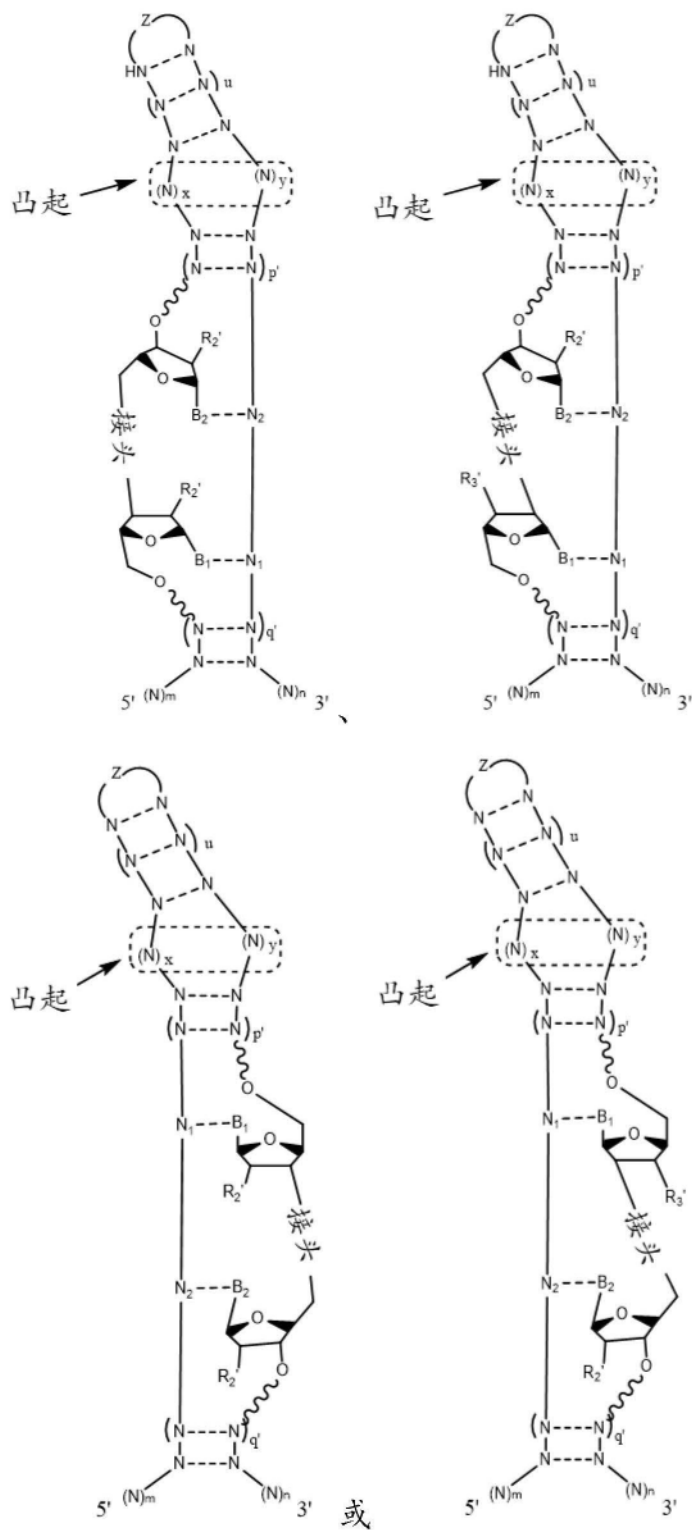


些实施例中,组合物基本上由具有下式的指导分子组成:

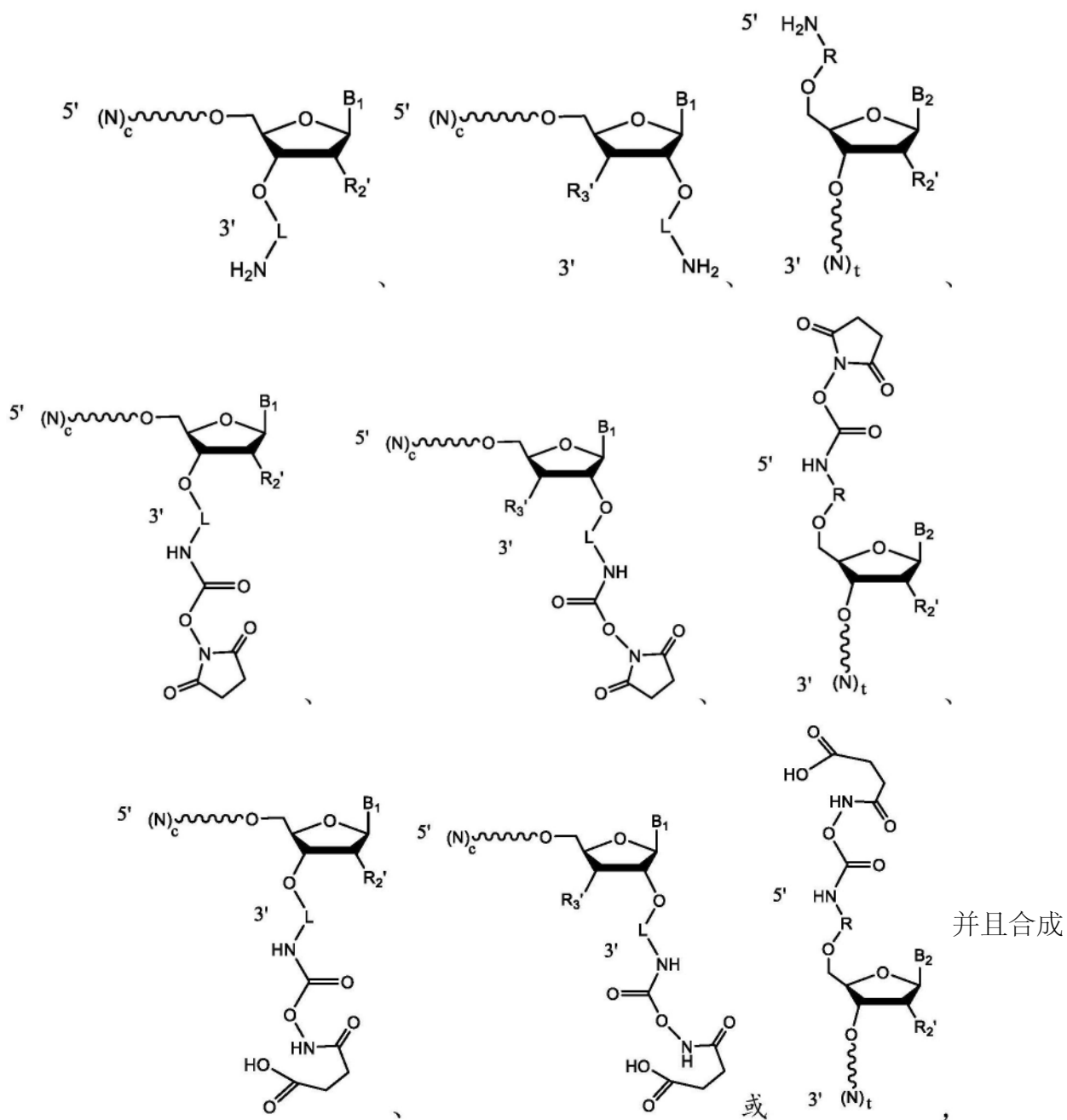


一些实施例中,组合物基本上由具有下式的指导分子组成:

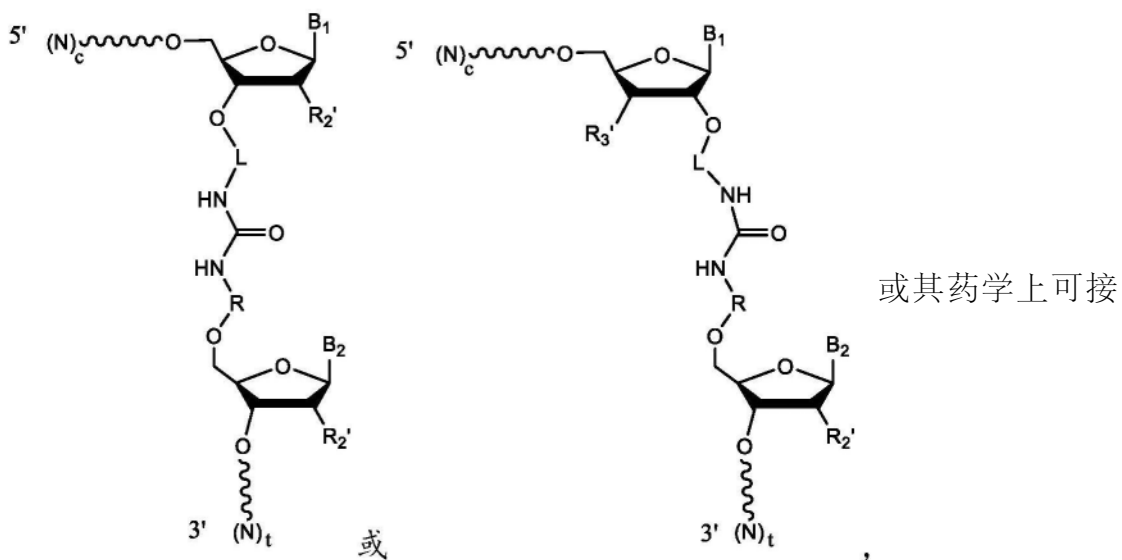




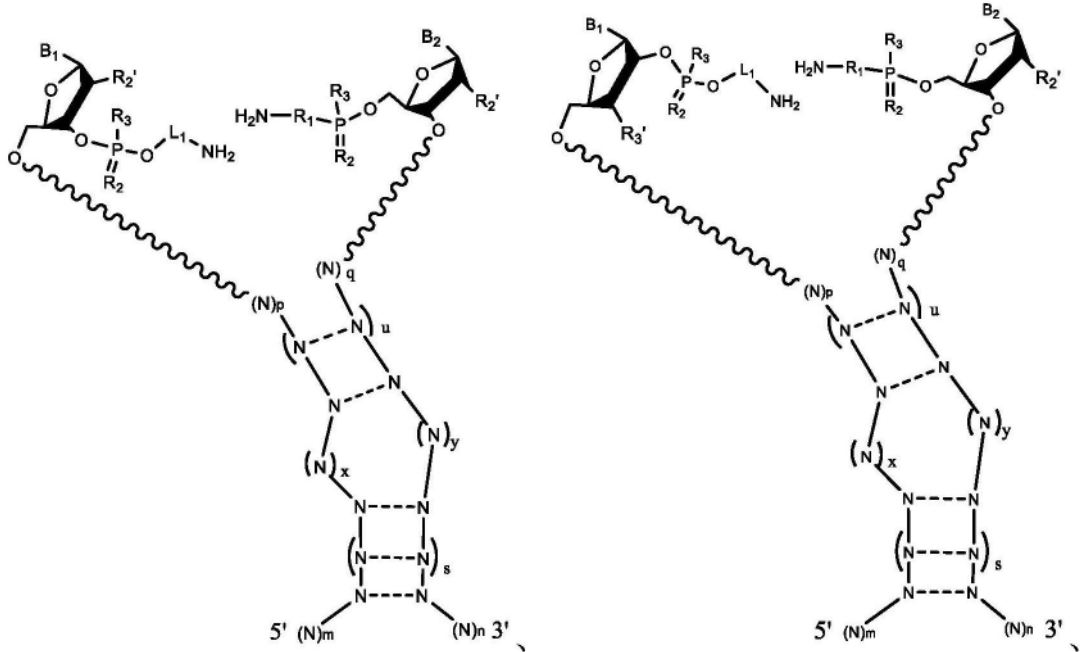
[0212] 在一些实施例中,组合物包含在合成指导分子存在或不存在下的寡核苷酸中间体(如上所述)。在一些实施例中,组合物的寡核苷酸中间体具有下式:

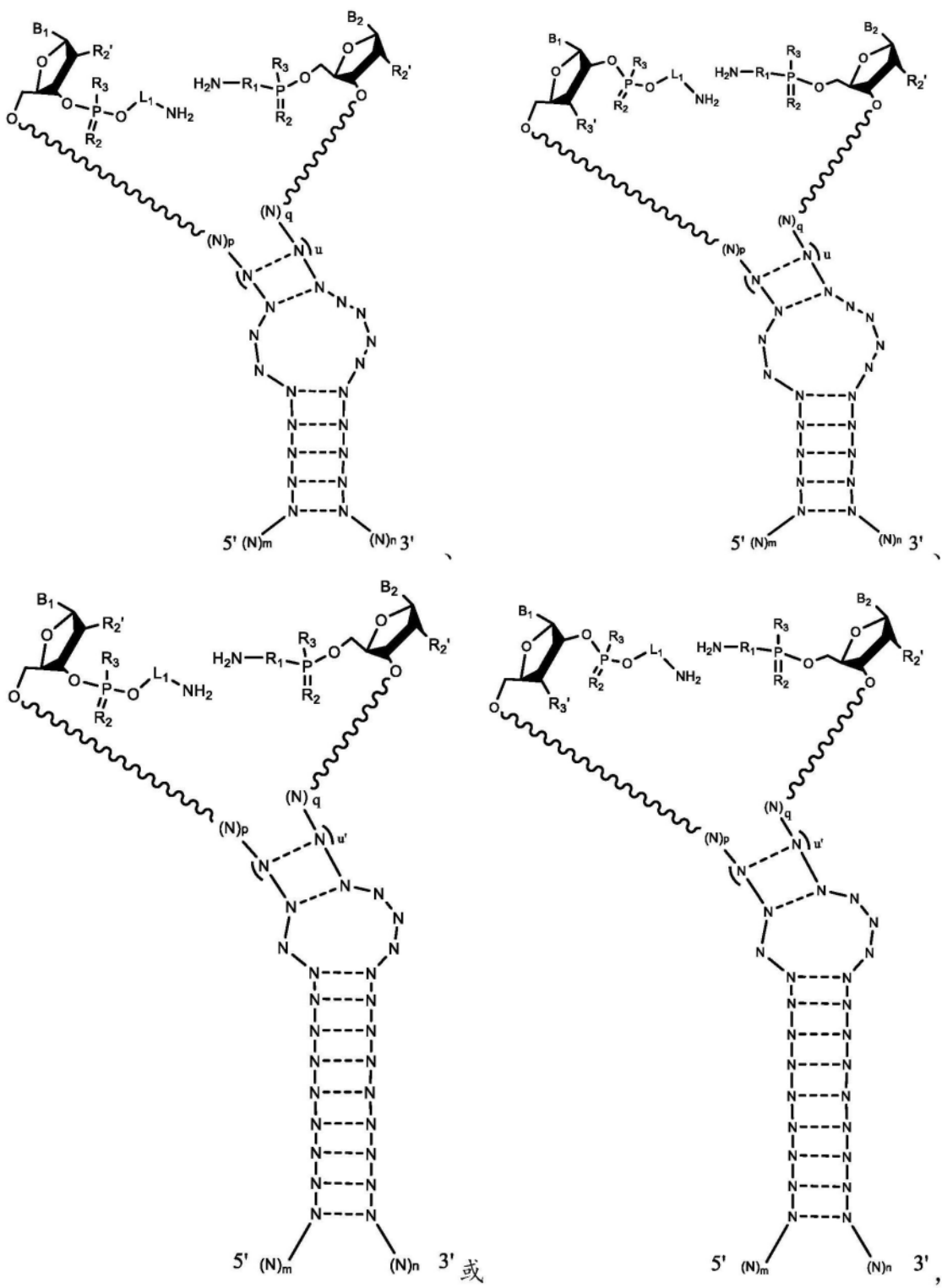


的指导分子具有下式：

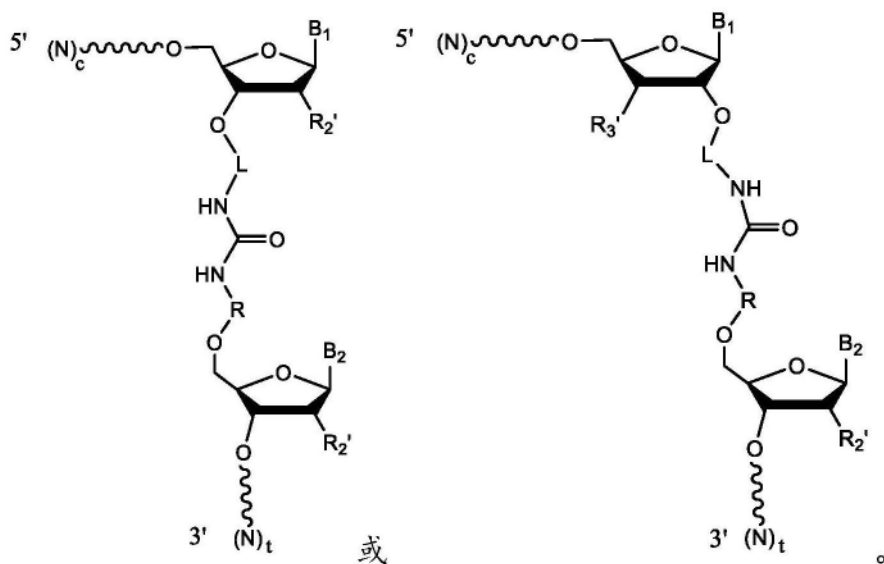


受的盐。在一些实施例中,组合物包含具有下式的退火双链体的寡核苷酸中间体:

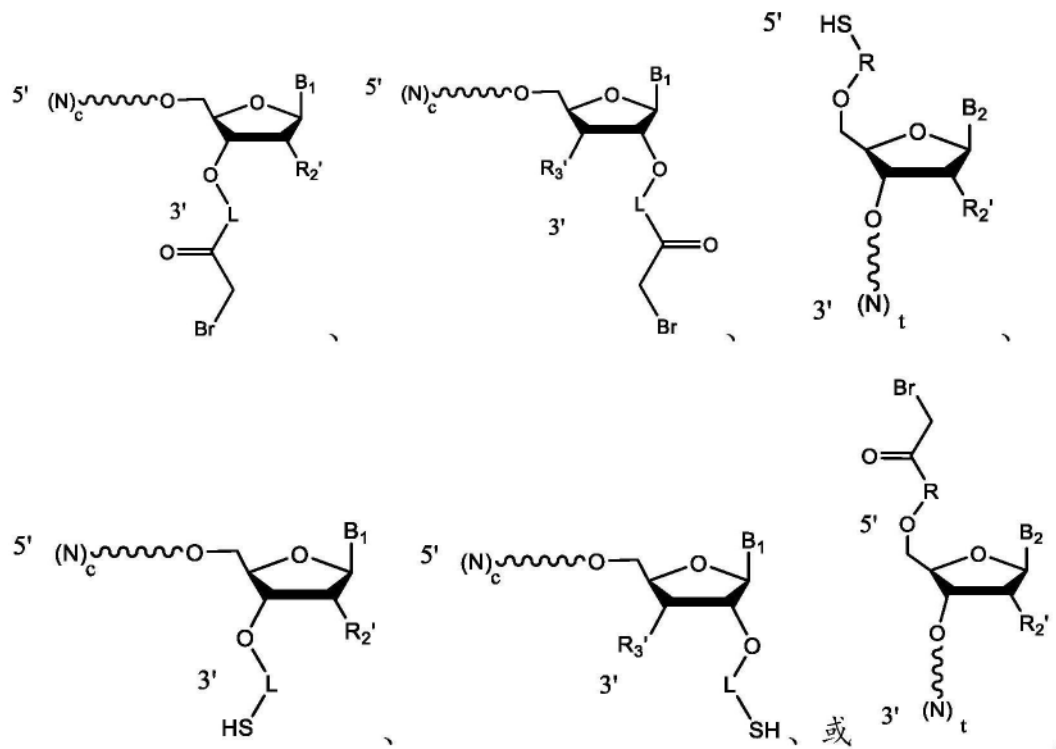




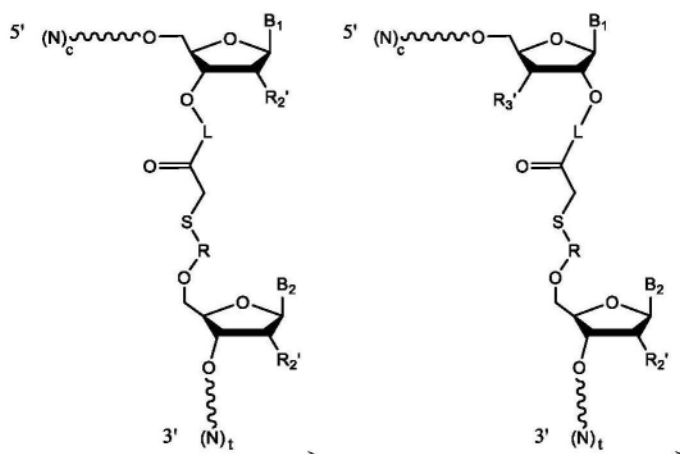
或其盐,这是在存在或不存在具有下式的合成指导分子的情况下:



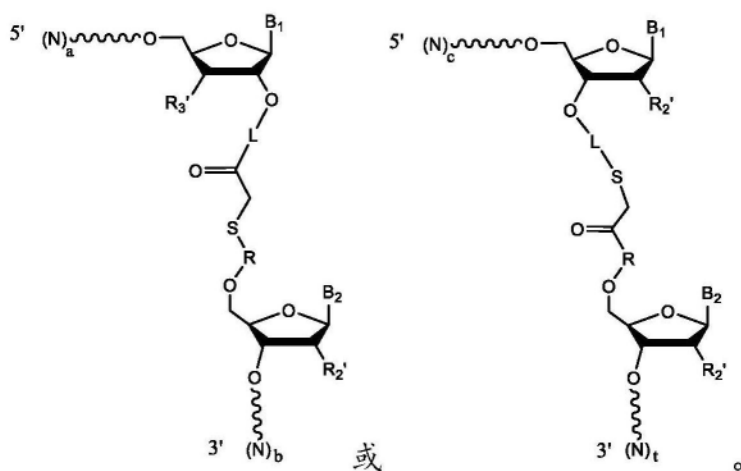
[0213] 在一些实施例中,组合物中的寡核苷酸中间体具有下式:



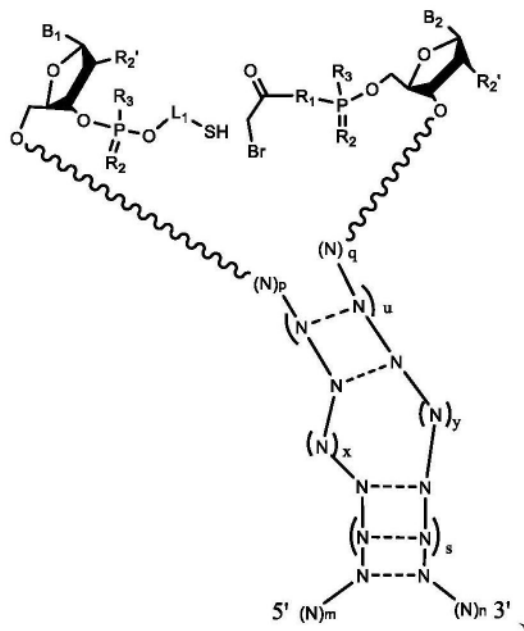
或其盐,并且合成的指导分子具有下式:



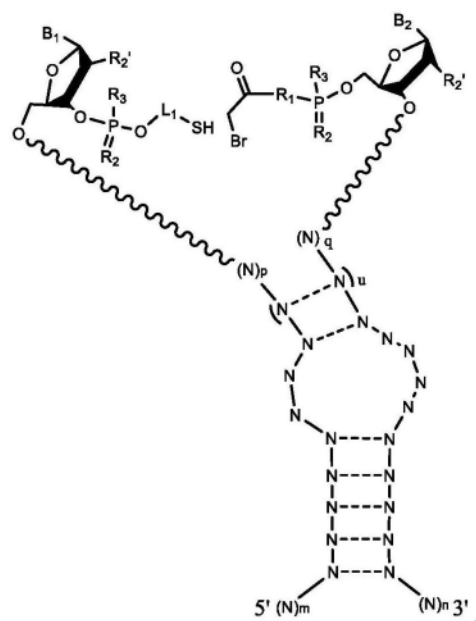
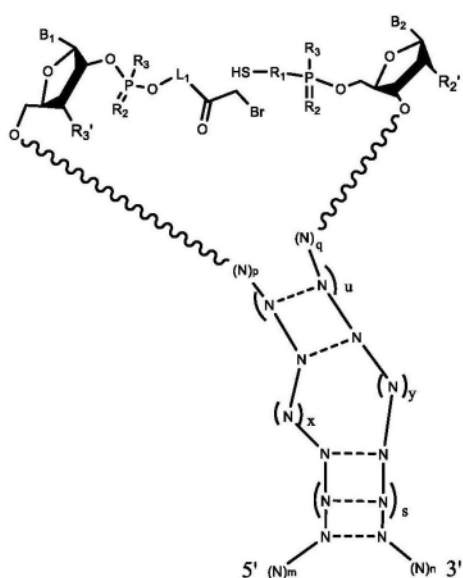
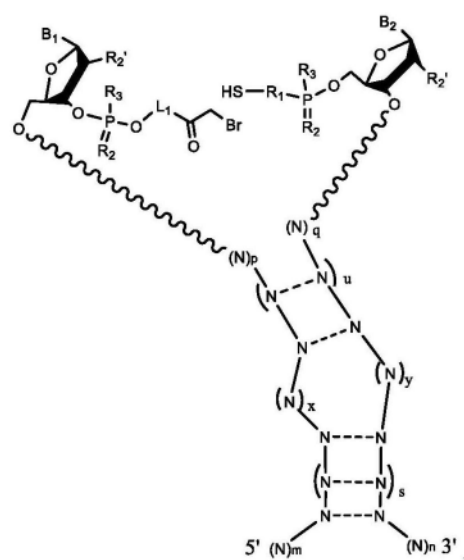
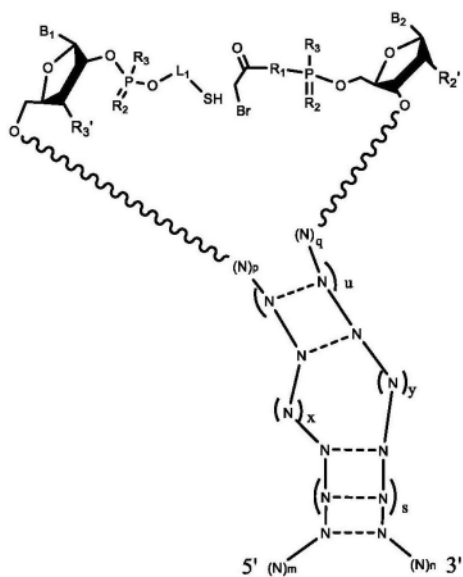
在一些实施例中,组合物包

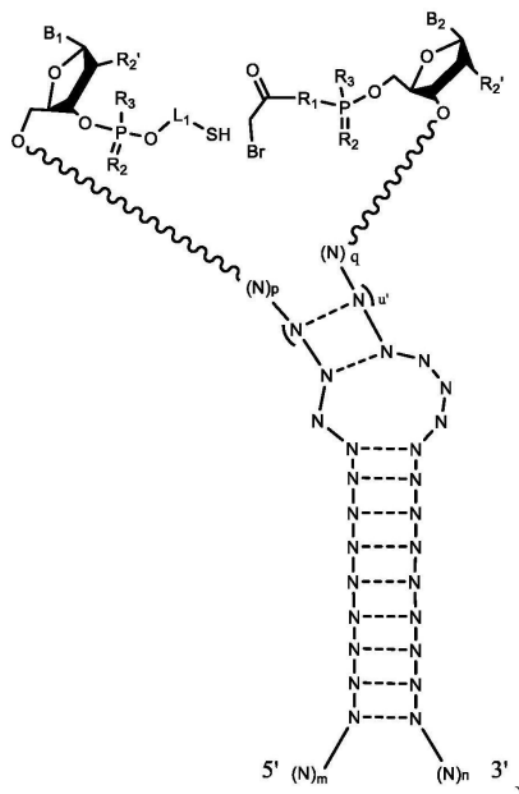
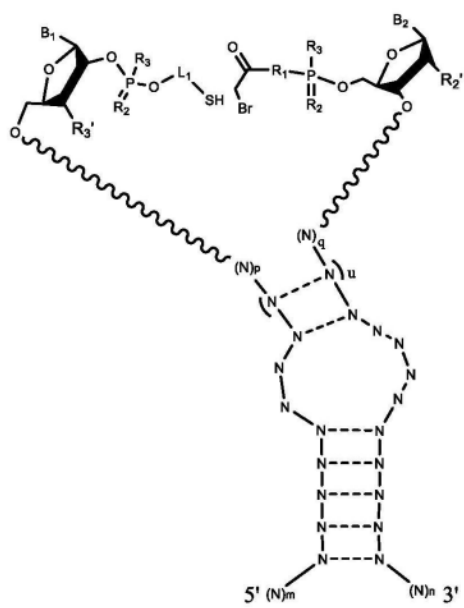
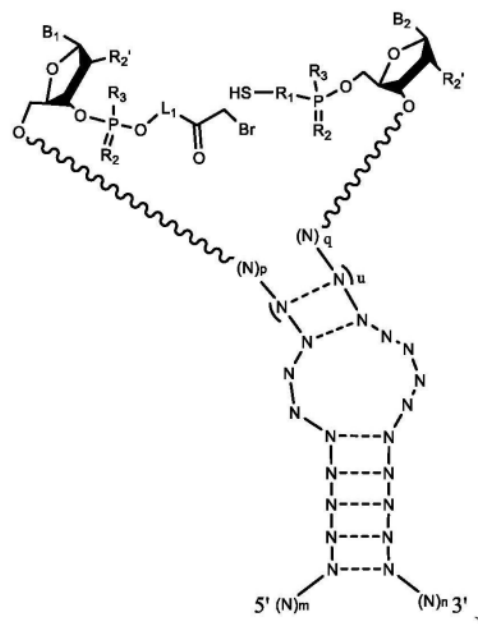
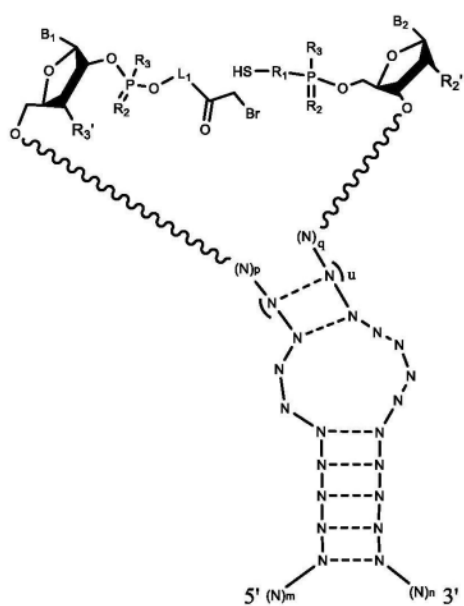


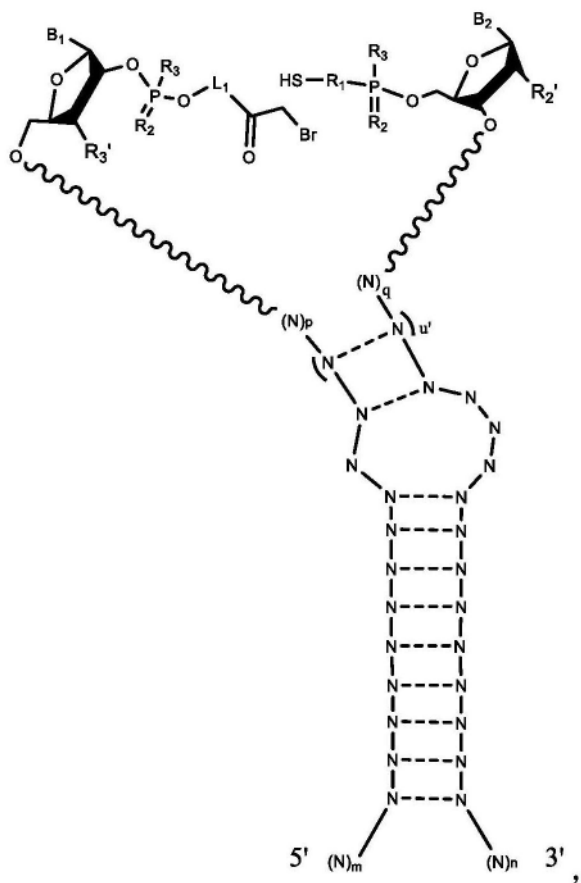
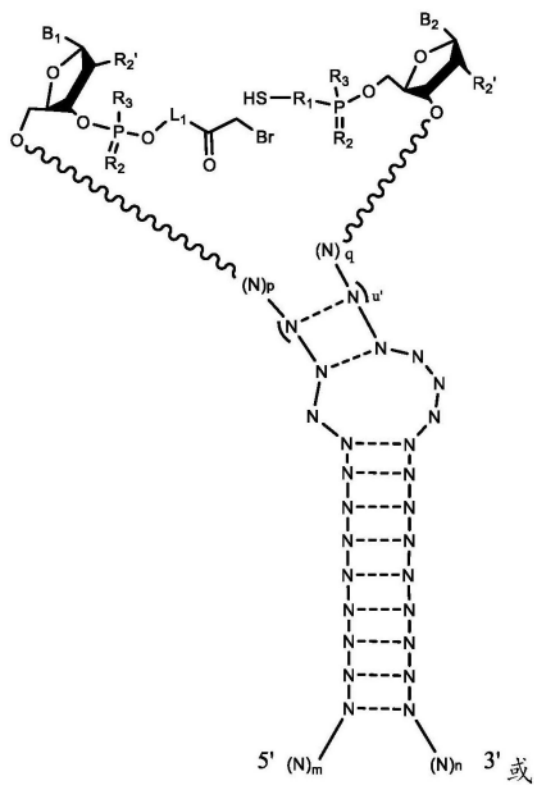
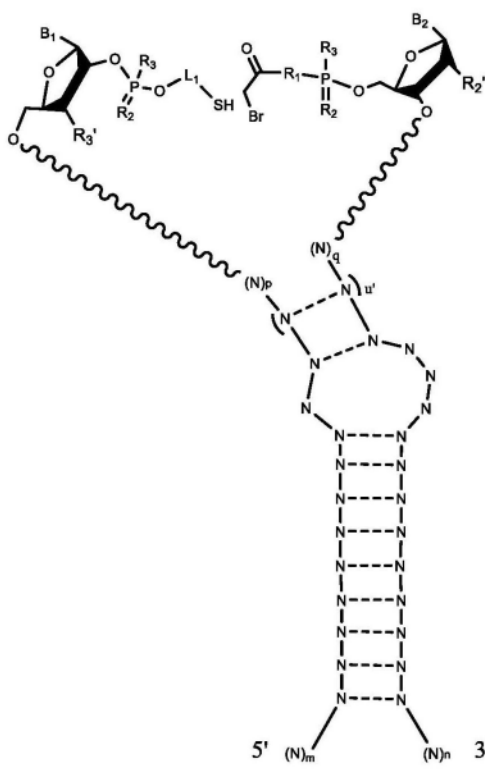
或



含具有下式的退火双链体的寡核苷酸中间体:

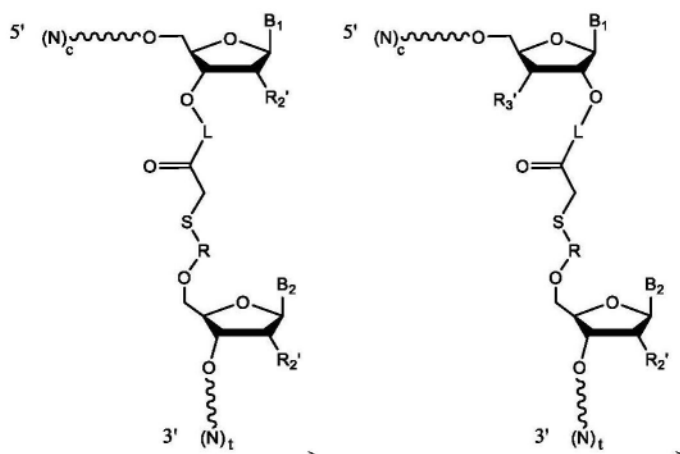




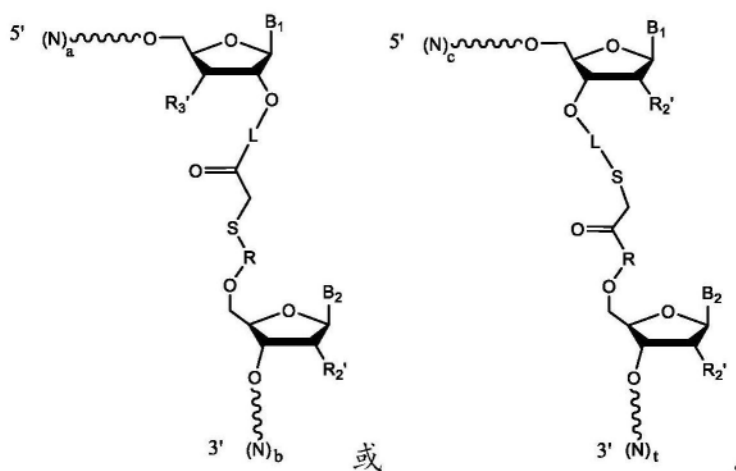


这是在存在或不存在具有下式的合成指导分子

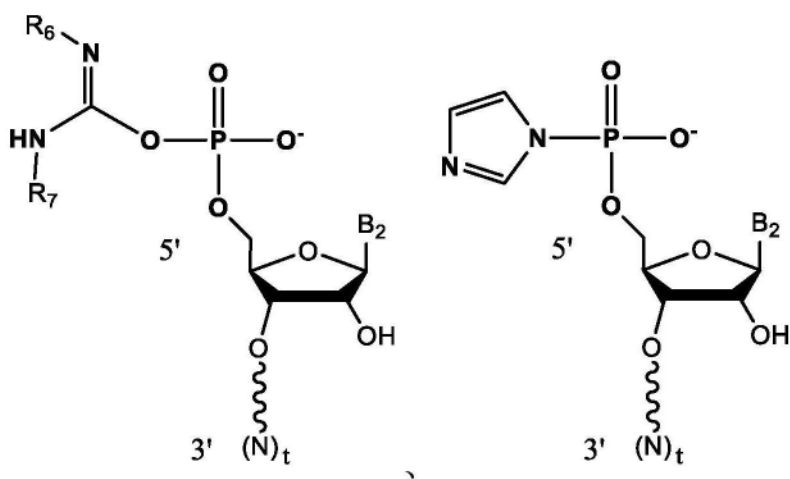
的情况下:



或其盐。

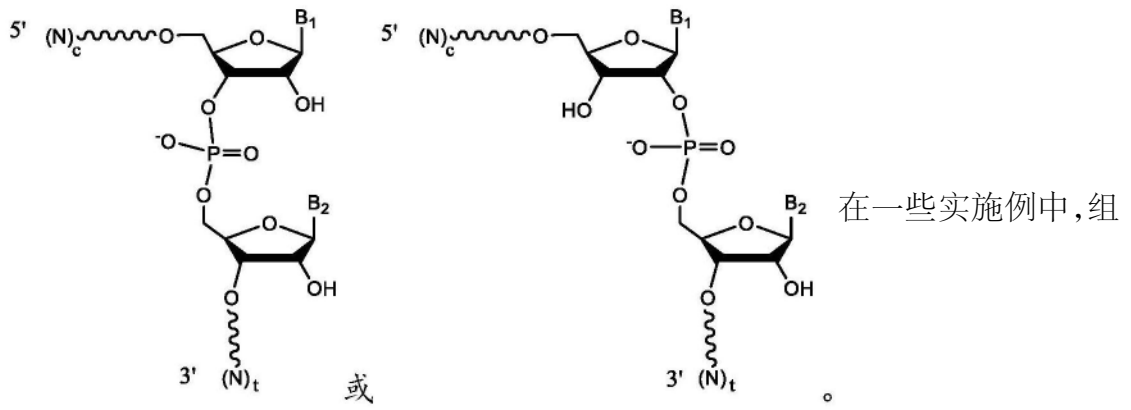


[0214] 在一些实施例中,组合物的寡核苷酸中间体具有下式:

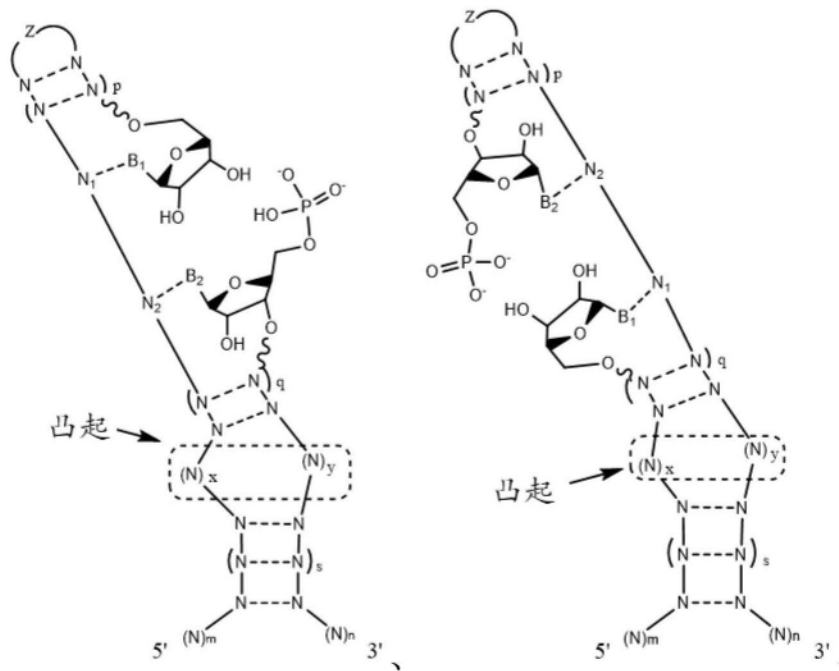


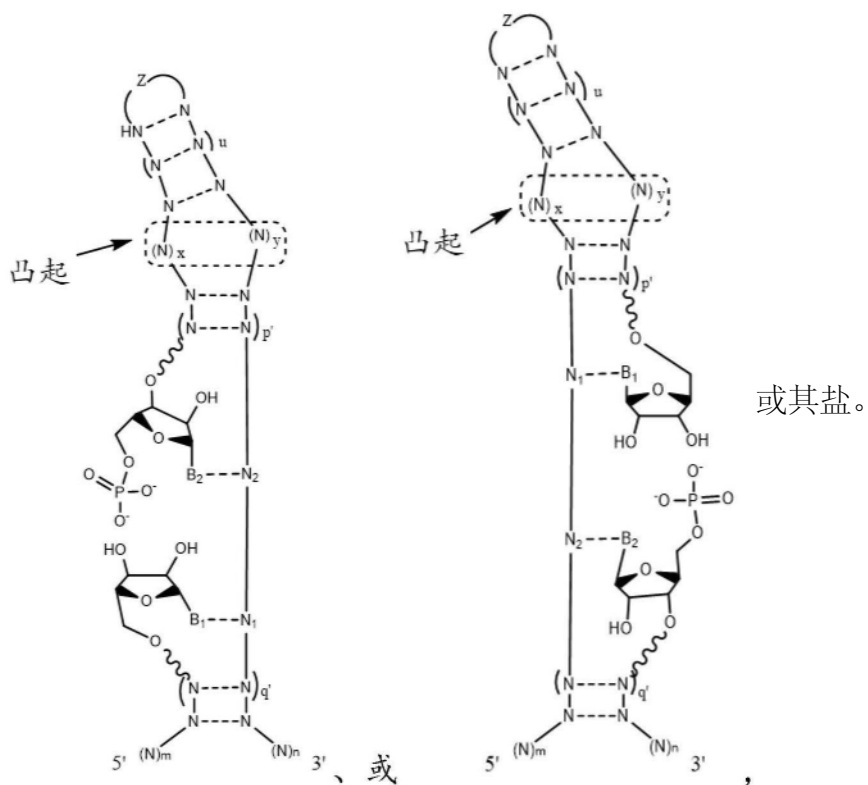
或,并且合成的指导分

子具有下式:

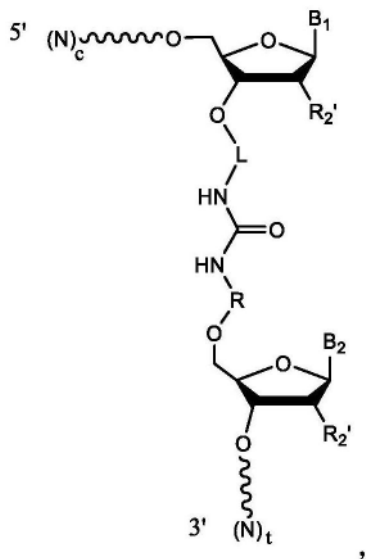


合物包含具有下式的退火双链体的寡核苷酸中间体：



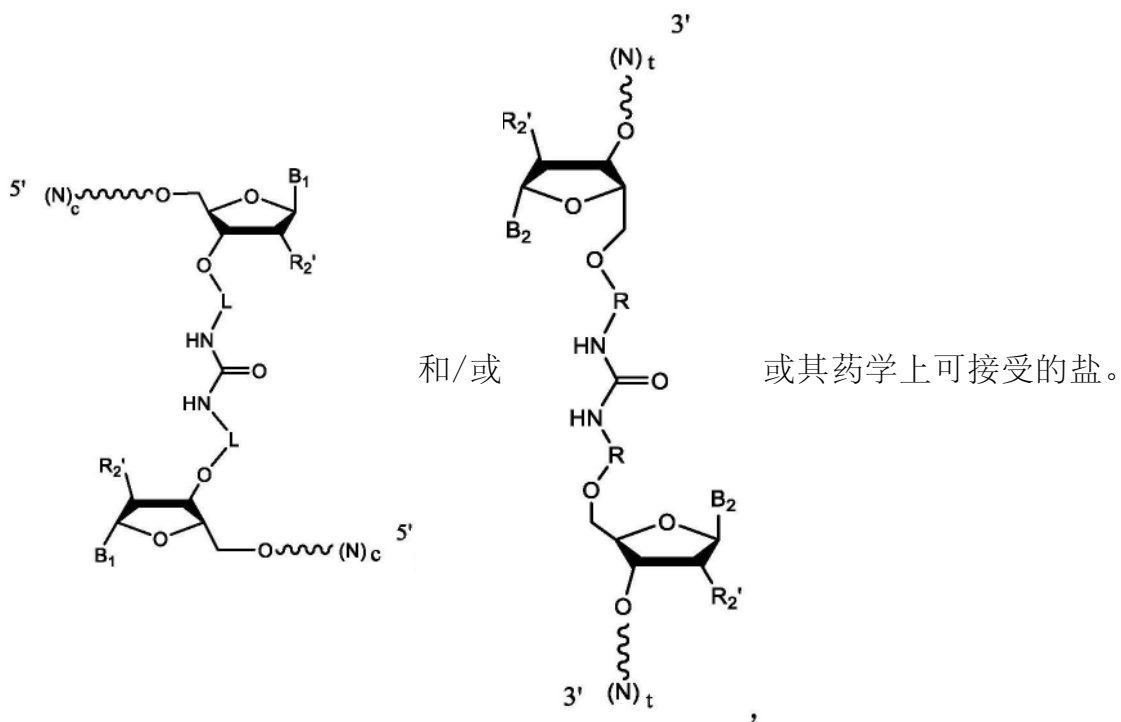


[0215] 在一些实施例中,组合物基本上不含同型二聚体。在一些实施例中,基本上不含同型二聚体和/或副产物的组合物包含使用包含同型双官能交联剂的方法合成的指导分子。在一些实施例中,基本上不含同型二聚体和/或副产物的组合物包含具有脲键合的指导分子。在一些实施例中,指导分子具有下式:

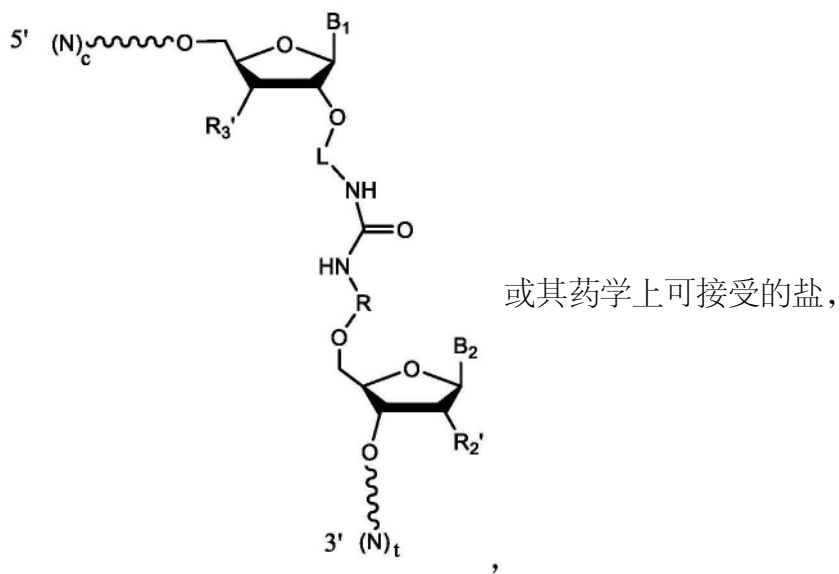


或其药学上可接受的盐,其中组合物基本上不含具有

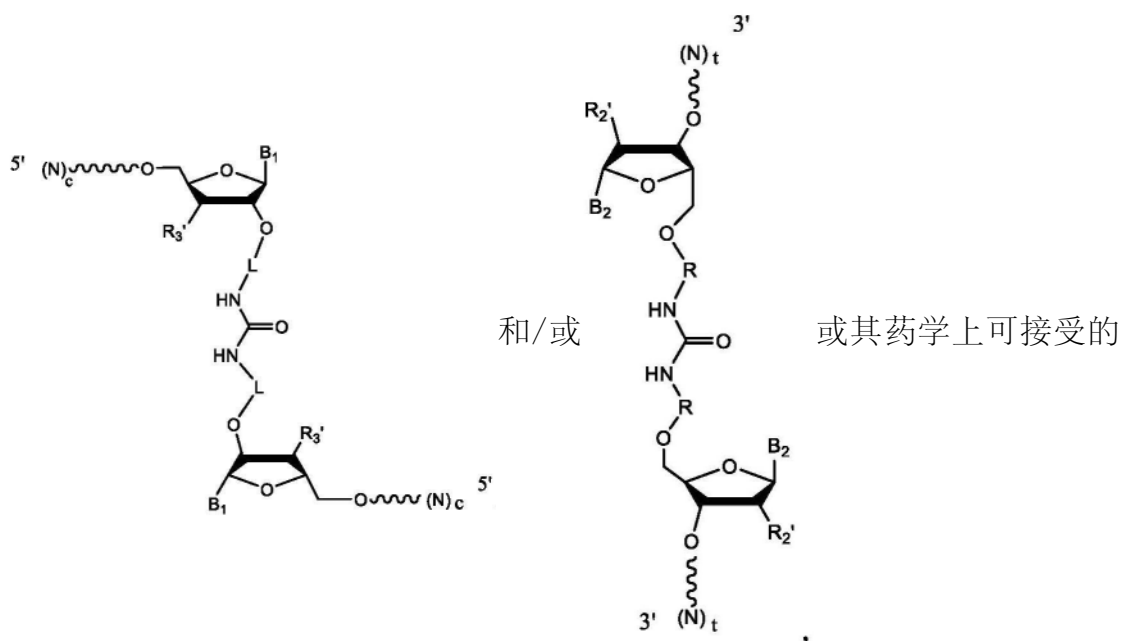
下式的分子:



在一些实施例中,指导分子具有下式:

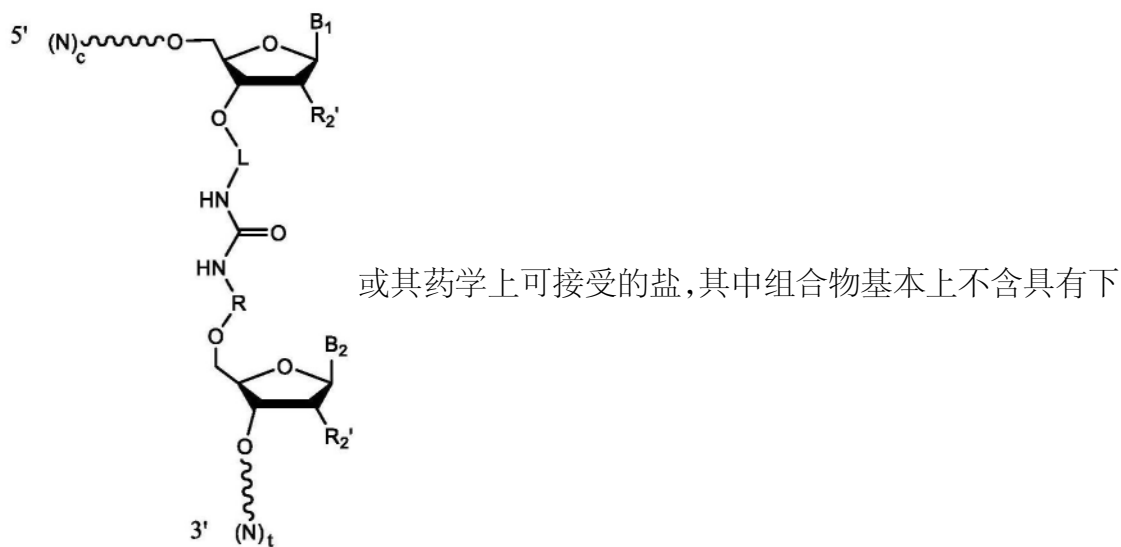


其中所述组合物基本上不含具有下式的分子:

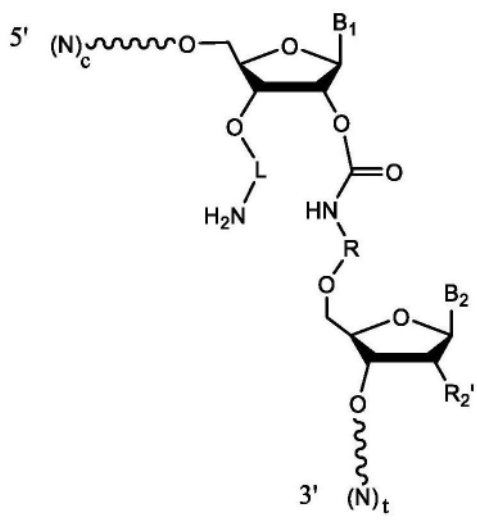


盐。

[0216] 在一些实施例中,组合物基本上不含副产物。在一些实施例中,组合物包含含有脲键合的指导分子。在一些实施例中,组合物包含具有下式的指导分子:

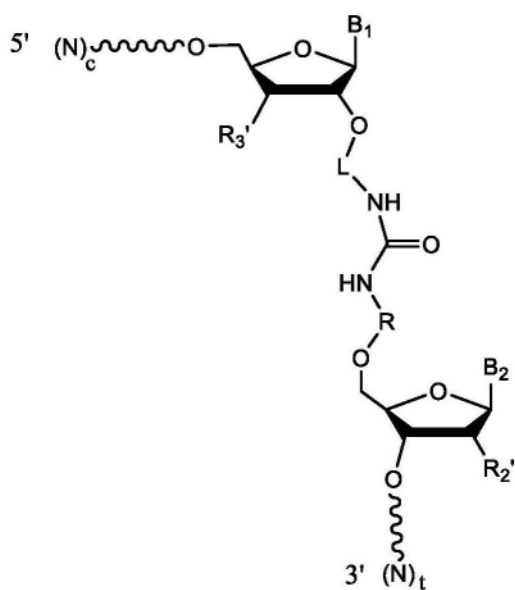


式的分子:



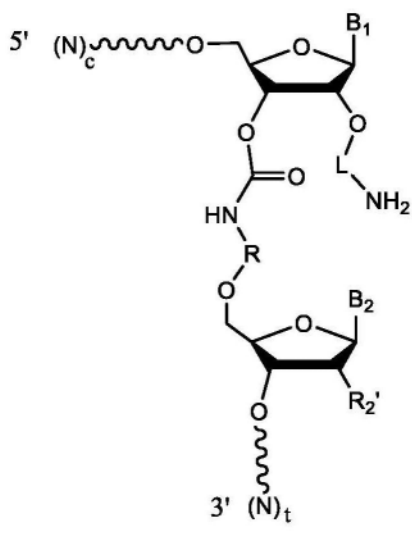
在一些实施例中,组合物包含具有下式的指

导分子:



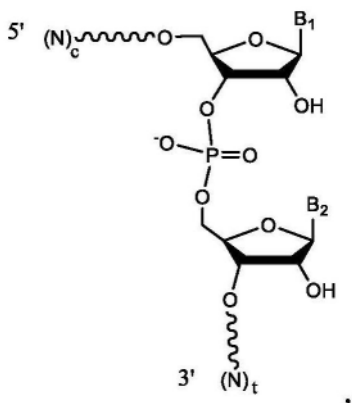
或其药学上可接受的盐,其中组合物基本上

不含具有下式的分子:



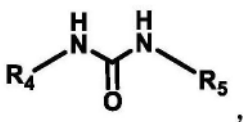
[0217] 在一些实施例中,组合物不是基本上不含副产物。在一些实施例中,组合物包含

(a) 用于CRISPR系统的合成的单分子指导分子,其中所述指导分子具有下式:



或其药学上可接受的盐;以及(b)以下一项或多项:(i)

碳二亚胺,或其盐;(ii)咪唑、氰基咪唑、吡啶和二甲基氨基吡啶,或其盐;以及(iii)具有下式的化合物:

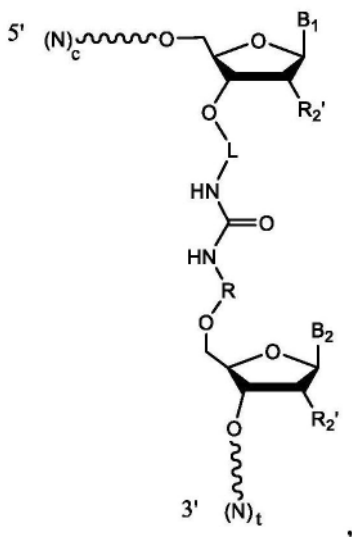


或其盐,其中R<sub>4</sub>和R<sub>5</sub>各自独立地是取代的或未被取代的烷基,或

取代的或未被取代的碳环。在一些实施例中,碳二亚胺是EDC、DCC或DIC。在一些实施例中,组合物包含EDC。在一些实施例中,组合物包含咪唑。

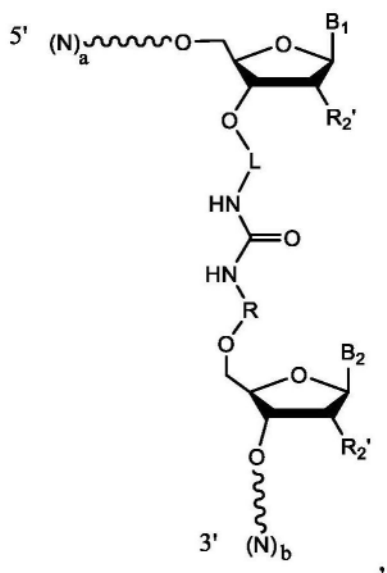
[0218] 在一些实施例中,组合物基本上不含n+1和/或n-1种类。在一些实施例中,组合物包含少于约10%、5%、2%、1%或0.1%的指导分子,所述指导分子包含相对于参照指导分子序列的截短。在一些实施例中,至少约85%、90%、95%、98%或99%的指导分子包含以下5'序列,所述5'序列包含指导分子的与参照指导分子序列的相应5'序列100%相同核苷酸1-20。

[0219] 在一些实施例中,组合物基本上包含具有下式的指导分子:



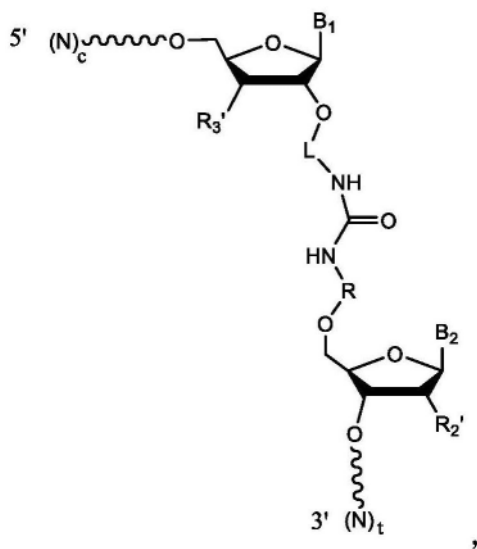
或其药学上可接受的盐,

其中所述组合物基本上不含具有下式的分子:



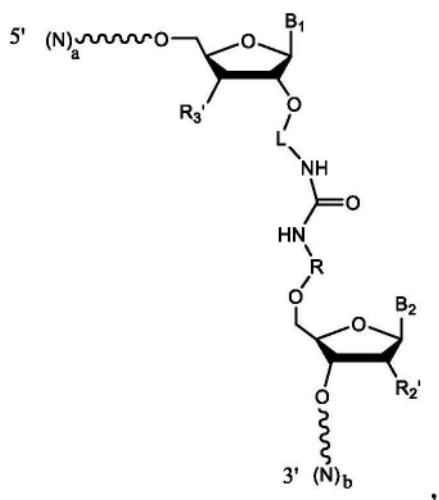
或其药学上可接受的盐,

其中a不等于c;和/或b不等于t。在一些实施例中,组合物基本上包含具有下式的指导分子:



或其药学上可接受的盐,其中所述组合物基本

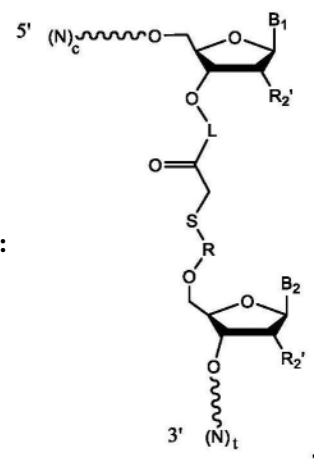
上不含具有下式的分子:



或其药学上可接受的盐,其中a不等于c;和/或b不

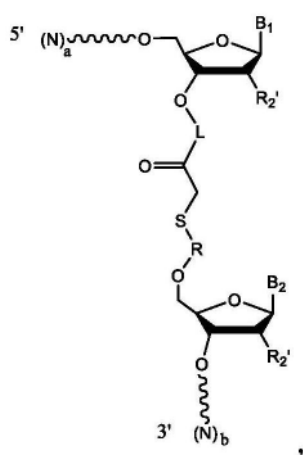
等于t。

[0220] 在一些实施例中,组合物基本上包含具有下式的指导分子:



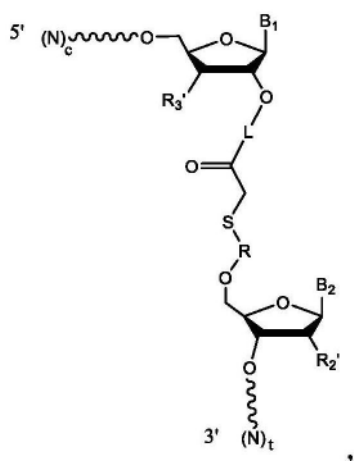
或其药学上可接受的盐,

其中所述组合物基本上不含具有下式的分子:



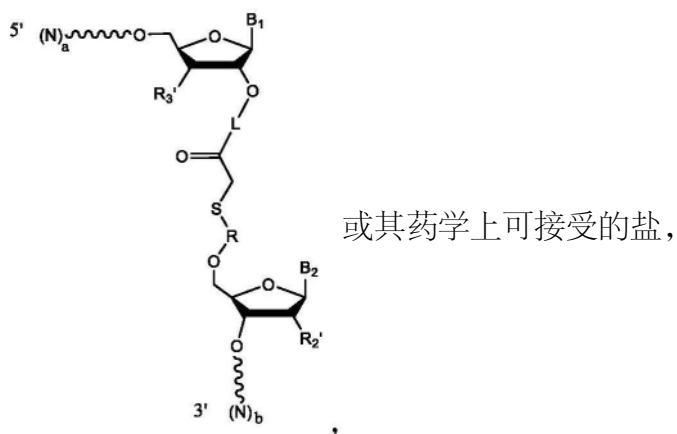
或其药学上可接受的盐,

其中a不等于c;和/或b不等于t。在一些实施例中,组合物基本上包含具有下式的指导分子:

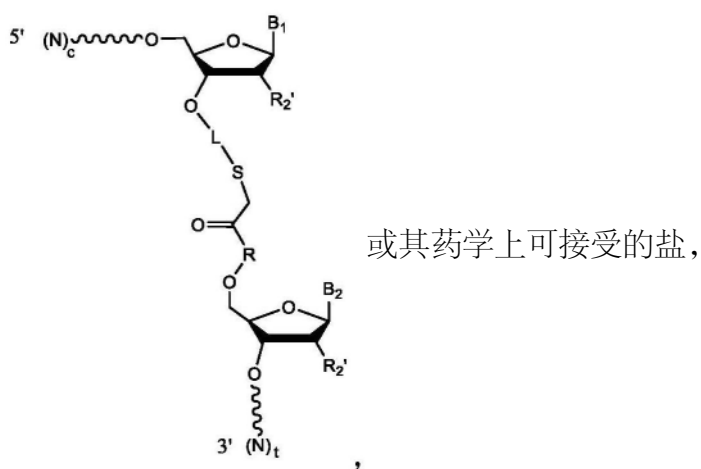


或其药学上可接受的盐,

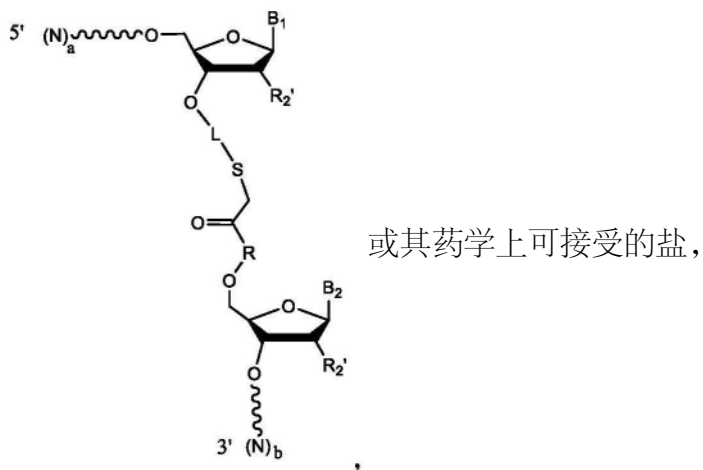
其中所述组合物基本上不含具有下式的分子:



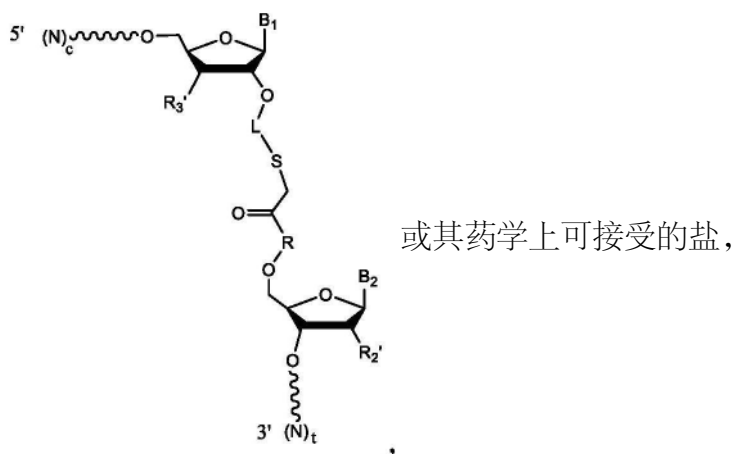
其中a不等于c；和/或b不等于t。在一些实施例中，组合物基本上包含具有下式的指导分子：



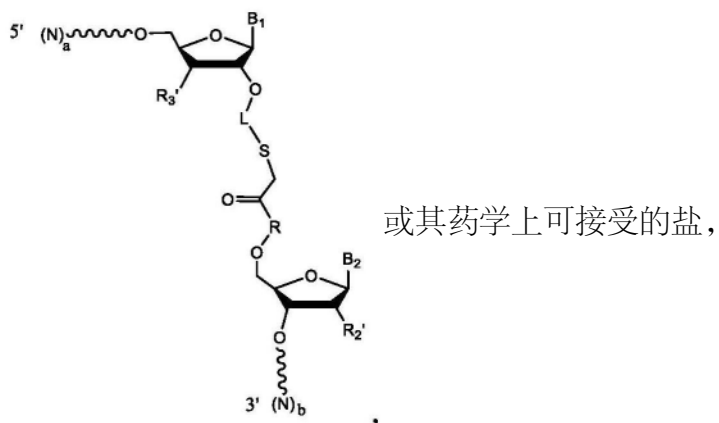
其中所述组合物基本上不含具有下式的分子：



其中a不等于c；和/或b不等于t。在一些实施例中，组合物基本上包含具有下式的指导分子：

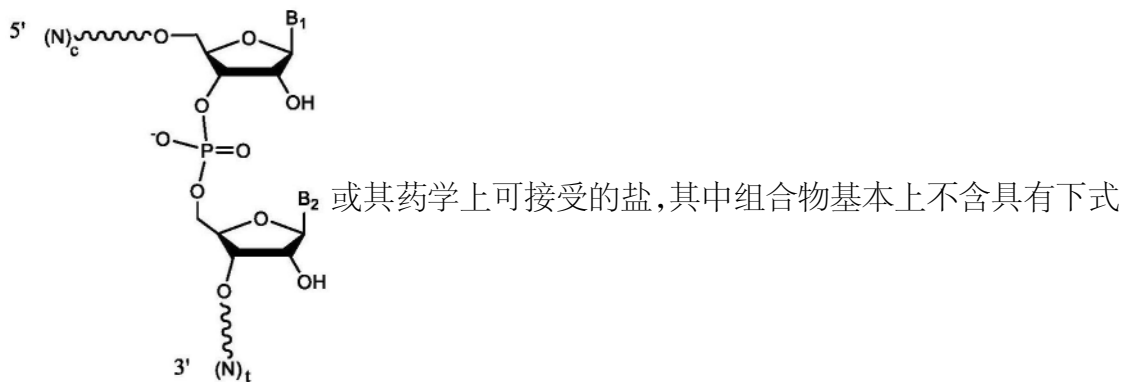


其中所述组合物基本上不含具有下式的分子:

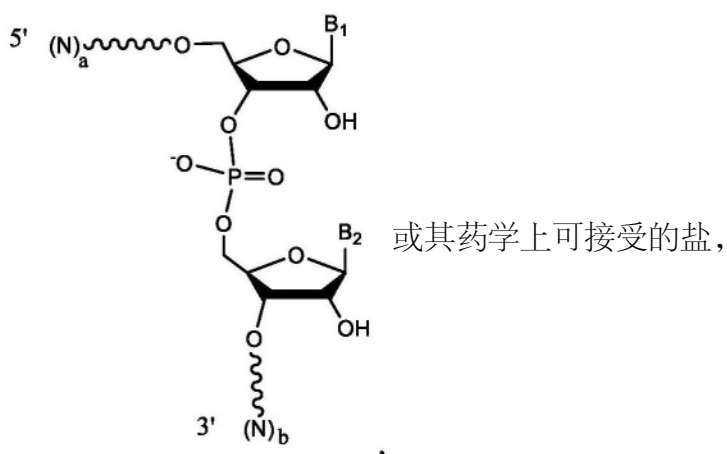


其中a不等于c;和/或b不等于t。在一些实施例中,a小于c,和/或b小于t。

[0221] 在一些实施例中,组合物包含具有下式的指导分子:

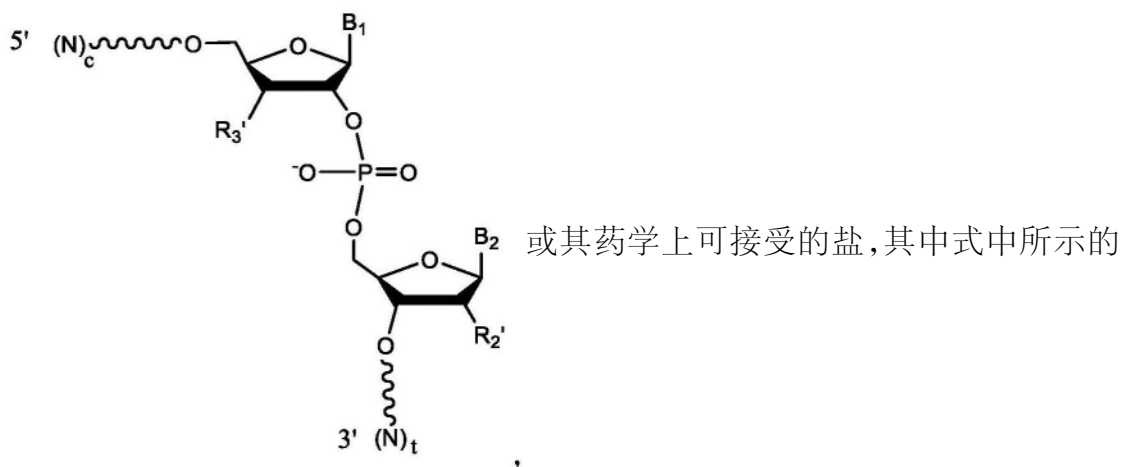


的分子:

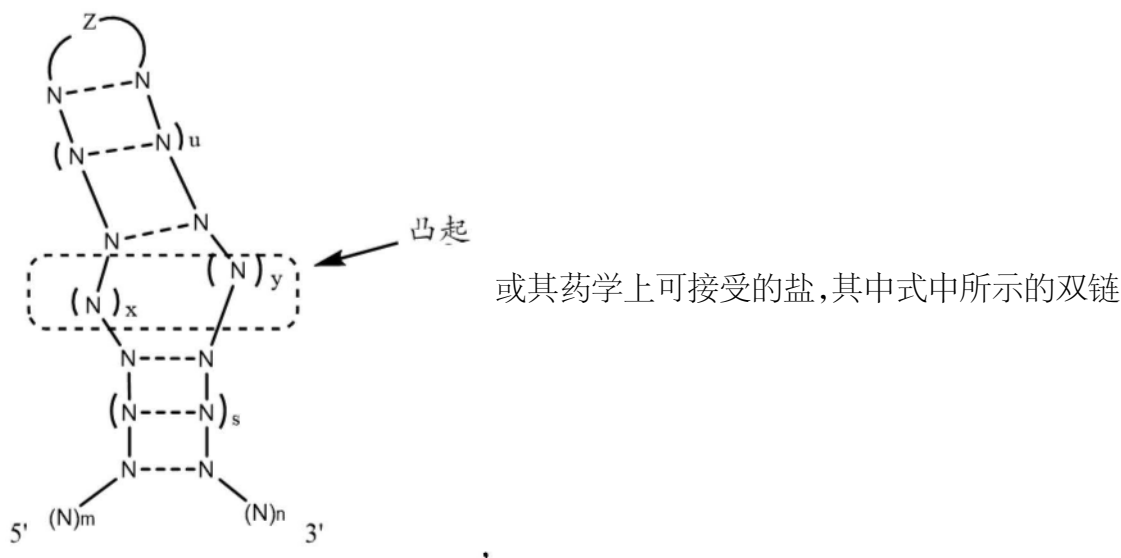


其中 $a+b$ 是 $c+t-k$ ,其中 $k$ 是1到10之间的整数(包括端值)。

[0222] 在一个实施例中,组合物包含用于CRISPR系统的合成的单分子指导分子,其中所述指导分子具有下式:



2'-5' 磷酸二酯键合在双链体中的两个核苷酸之间。在一些实施例中,指导分子具有下式:



体区中两个核苷酸之间的至少一个磷酸二酯键合是2'-5' 磷酸二酯键合。在一些实施例中,2'-5' 磷酸二酯键合在位于凸起5'的两个核苷酸之间。在一些实施例中,2'-5' 磷酸二酯键合在位于核苷酸环Z的5'和凸起的3'的两个核苷酸之间。在一些实施例中,2'-5' 磷酸二酯

键合在位于核苷酸环Z的3'和凸起的5'的两个核苷酸之间。在一些实施例中,2'-5'磷酸二酯键合在位于凸起3'的两个核苷酸之间。

#### 指导分子设计

[0223] 用于选择和验证靶序列的方法以及脱靶分析先前已描述于例如以下文献中: Mali、Hsu、Fu等人,2014Nat biotechnol[自然生物技术]32(3):279-84;Heigwer等人,2014Nat methods[自然方法]11(2):122-3;Bae等人,Bioinformatics[生物信息学]30(10):1473-5;和Xiao A等人(2014)Bioinformatics[生物信息学]30(8):1180-1182。这些参考文献中的每一篇都是通过引用并入本文。作为非限制性实例,指导分子设计可包括使用软件工具来优化对应于使用者的靶序列的潜在靶序列的选择,例如以使全基因组的总脱靶活性降至最低。虽然脱靶活性不限于裂解,但每个脱靶序列处的裂解效率可以例如使用实验衍生的加权方案来预测。这些和其他指导选择方法详细描述于Maeder和Cotta-Ramusino中。

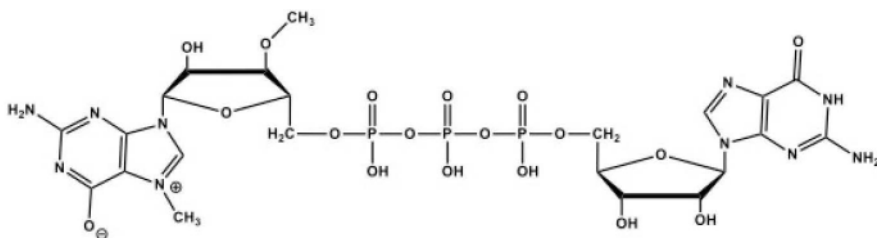
[0224] 还可以设计合成的单分子指导分子中的化学键合的茎环结构和位置。发明人认识到使用吉布斯自由能差( $\Delta G$ )来预测化学缀合反应的连接效率的价值。使用OligoAnalyzer(可从[www.idtdna.com/calc/analyzer](http://www.idtdna.com/calc/analyzer)获得)或类似工具进行 $\Delta G$ 的计算。比较异型二聚化的 $\Delta G$ 以形成所需的退火双链体和两个相同寡核苷酸的同型二聚化的 $\Delta G$ 可以预测化学缀合的实验结果。当异型二聚化的 $\Delta G$ 小于同型二聚化的 $\Delta G$ 时,预测连接效率是高的。该预测方法在实例XX中进一步解释。

#### 指导分子修饰

[0225] 指导分子的活性、稳定性或其他特征可以通过并入某些修饰来改变。作为一个实例,瞬时表达或递送的核酸可能易于被例如细胞核酸酶降解。因此,本文所述的指导分子可以含有引入针对核酸酶的稳定性的一个或多个经修饰核苷或核苷酸。虽然不希望受理论束缚,但还应相信,本文所述的某些经修饰的指导分子在引入细胞中时可展现降低的先天性免疫反应。本领域技术人员将了解一般在细胞(例如哺乳动物细胞)中响应外源核酸(尤其那些病毒或细菌来源的核酸)观察到的某些细胞反应。此类反应可以包括诱导细胞因子表达和释放以及细胞死亡,可通过本文所呈现的修饰来减少或完全消除。

[0226] 这个章节中讨论的某些示例性修饰可以包括于指导分子序列内的任一位置,包括但不限于在5'末端处或5'末端附近(例如,在5'末端的1-10、1-5、1-3或1-2个核苷酸内)和/或在3'末端处或3'末端附近(例如,在3'末端的1-10、1-5、1-3或1-2个核苷酸内)。在一些情形中,修饰定位于功能基序内,例如Cas9指导分子的重复-抗重复双链体、Cas9或Cpf1指导分子的茎环结构和/或指导分子的靶向结构域。

[0227] 作为一个实例,指导分子的5'末端可以包括真核mRNA帽结构或帽类似物(例如,G(5')ppp(5')G帽类似物、m7G(5')ppp(5')G帽类似物或3'-O-Me-m7G(5')ppp(5')G抗反向帽类似物(ARCA)),如下文所示:

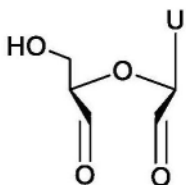


在指导分子的化学或酶促合成过程中可以包括帽或帽类似物。

[0228] 按类似方式,指导分子的5'末端可能缺少5'三磷酸基团。例如,体外转录的指导分子可能经磷酸酶处理(例如,使用小牛肠碱性磷酸酶)以移除5'三磷酸基团。

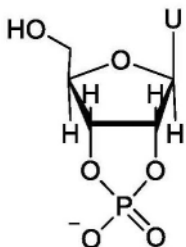
[0229] 另一常见修饰包括在指导分子的3'末端添加多个(例如,1-10、10-20或25-200个)腺嘌呤(A)残基,称为聚A段。使用聚腺苷聚合酶(例如,大肠杆菌聚(A)聚合酶),可以在化学或酶促合成期间将聚A段添加到指导分子中。

[0230] 指导RNA可以在3'末端U核糖处经修饰。例如,U核糖的两个末端羟基可以被氧化为醛基,并且伴随核糖环的打开,以提供如下所示的经修饰核苷:



其中“U”可以是未经修饰或经修饰的尿苷。

[0231] 3'末端U核糖可以经如下所示的2'3'环状磷酸酯修饰:



其中“U”可以是未经修饰或经修饰的尿苷。

[0232] 指导RNA可以例如通过并入本文所述的一个或多个经修饰核苷酸而含有可以针对降解稳定的3'核苷酸。在某些实施例中,尿苷可以由经修饰尿苷(例如,5-(2-氨基)丙基尿苷和5-溴代尿苷)或由本文所述的任何经修饰尿苷替代;腺苷和鸟苷可以由经修饰腺苷和鸟苷(例如,在8位处具有修饰,例如8-溴代鸟苷)或由本文所述的任何经修饰腺苷和鸟苷替代。

[0233] 在某些实施例中,可以将糖修饰的核糖核苷酸并入指导分子中,例如,其中2'OH-基团由选自以下的基团替代:H、-OR、-R(其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖)、卤基、-SH、-SR(其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖)、氨基(其中氨基可以是例如NH<sub>2</sub>;烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基、二杂芳基氨基或氨基酸);或氰基(-CN)。在某些实施例中,磷酸骨架可以如本文所述经修饰,例如经硫代磷酸酯(PhTx)基团修饰。在某些实施例中,指导分子的一个或多个核苷酸可以各自独立地是经修饰或未经修饰的核苷酸,包括但不限于2'-糖修饰的,例如2'-O-甲基、2'-O-甲氧基乙基,或2'-氟修饰的,包括例如,2'-F或2'-O-甲基腺苷(A)、2'-F或2'-O-甲基胞苷(C)、2'-F或2'-O-甲基尿苷(U)、2'-F或2'-O-甲基胸苷(T)、2'-F或2'-O-甲基鸟苷(G)、2'-O-甲氧基乙基-5-甲基尿苷(Teo)、2'-O-甲氧基乙基腺苷(Aeo)、2'-O-甲氧基乙基-5-甲基胞苷(m5Ceo)及其任何组合。

[0234] 指导RNA还可以包括“锁定”核酸(LNA),其中2'OH-基团可以例如通过C1-6亚烷基C1-6亚杂烷基桥连接至同一核糖的4'碳。可使用任何适宜部分来提供此类桥,包括但不限

于亚甲基、亚丙基、醚或氨基桥；0-氨基（其中氨基可以是例如NH<sub>2</sub>；烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基或二杂芳基氨基、乙二胺或聚氨基）和氨基烷氧基或O(CH<sub>2</sub>)<sub>n</sub>-氨基（其中氨基可以是例如NH<sub>2</sub>；烷基氨基、二烷基氨基、杂环基、芳基氨基、二芳基氨基、杂芳基氨基或二杂芳基氨基、乙二胺或聚氨基）。

[0235] 在某些实施例中，指导分子可以包括经修饰核苷酸，其为多环状（例如，三环；和“解锁”形式，例如二醇核酸（GNA）（例如，R-GNA或S-GNA，其中核糖由附接至磷酸二酯键的二醇单元替代），或苏糖核酸（TNA，其中核糖由 $\alpha$ -L-苏式呋喃糖基-（3'→2'）替代）。

[0236] 通常，指导分子包括糖基核糖，它是具有氧的5元环。示例性的经修饰指导分子可以包括但不限于核糖中氧的替代（例如，经硫（S）、硒（Se）或亚烷基，例如亚甲基或亚乙基）；双键的添加（例如，以用环戊烯基或环己烯基替代核糖）；核糖的缩环（例如，以形成环丁烷或氧杂环丁烷的4元环）；核糖的扩环（例如，以形成具有额外碳或杂原子的6元环或7元环，例如脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代，其也具有氨基磷酸酯骨架）。尽管大多数的糖类物改变位于2'位，但其他位点也适于修饰，包括4'位。在某些实施例中，指导分子包含4'-S、4'-Se或4'-C-氨基甲基-2'-O-Me修饰。

[0237] 在某些实施例中，可以将脱氮核苷酸（例如，7-脱氮-腺苷）并入指导分子中。在某些实施例中，可以将0-烷基化和N-烷基化核苷酸（例如，N6-甲基腺苷）并入指导分子中。在某些实施例中，指导分子中的一个或多个或所有核苷酸是脱氧核苷酸。

[0238] 指导分子的核苷酸也可以在磷酸二酯键合处被修饰。这种修饰可包括磷酰乙酸酯，硫代磷酸酯，硫代磷酰乙酸酯或磷酰胺酯键合。在一些实施例中，核苷酸可通过硫代磷酸酯键合与其相邻核苷酸连接。此外，对磷酸二酯键合的修饰可以是对核苷酸的唯一修饰，或者可以与上述其他核苷酸修饰组合。例如，经修饰的磷酸二酯键合可以与对核苷酸的糖基的修饰组合。在一些实施例中，5'或3'核苷酸包含2'-OMe修饰的核糖核苷酸残基，其通过硫代磷酸酯键合与其相邻的一个或多个核苷酸连接。

#### RNA指导的核酸酶

[0239] 根据本披露的RNA指导的核酸酶包括但不限于天然存在的2类的CRISPR核酸酶，例如Cas9和Cpf1，以及由其衍生或获得的其他核酸酶。在功能方面，RNA指导的核酸酶定义为如下的那些核酸酶：(a) 与指导分子（例如，gRNA）相互作用（例如复合）；和(b) 与指导分子（例如，gRNA）一起缔合或任选地裂解或修饰DNA的靶区域，所述靶区域包括(i) 与指导分子（例如，gRNA）的靶向结构域互补的序列，以及任选地，(ii) 称为“原型间隔子邻近基序”或“PAM”的另一序列，其更详细地描述于下文中。在说明以下实例时，RNA指导的核酸酶可以在广义上依据其PAM特异性和裂解活性来定义，即使在共享相同PAM特异性或裂解活性的个别RNA指导的核酸酶之间可能存在变异。技术人员将了解，本披露的一些方面涉及可以使用具有某种PAM特异性和/或裂解活性的任何适宜RNA指导的核酸酶实施的系统、方法和组合物。因此，除非另有说明，术语RNA指导的核酸酶应理解为通用术语，并不限于任何特定类型（例如，Cas9与Cpf1），物种（例如，化脓链球菌与金黄色葡萄球菌）或RNA指导的核酸酶的变异（例如，全长与截短的或分裂；天然存在的PAM特异性与工程化PAM特异性等）。

[0240] PAM序列的名称来源于其与“原型间隔子”序列的顺序关系，所述原型间隔子与指导分子靶向结构域（或“间隔序列”）互补。与原型间隔子一起，PAM序列定义特定RNA指导的核酸酶/指导分子组合的靶区域或序列。

[0241] 多种RNA指导的核酸酶可能需要PAM与原型间隔子之间的不同顺序关系。通常,如相对于指导分子可视化的,Cas9s识别原型间隔子3'的PAM序列。

[0242] 另一方面,如相对于指导分子可视化的,Cpf1通常识别原型间隔子5'的PAM序列。

[0243] 除了识别PAM和原型间隔子的特定顺序定向以外,RNA指导的核酸酶还可以识别特定PAM序列。例如,金黄色葡萄球菌Cas9识别NNGRRT或NNGRRV的PAM序列,其中N个残基紧靠由指导分子靶向结构域所识别区域的3'。化脓链球菌Cas9识别NGG PAM序列。并且新凶手弗朗西斯菌(*F. novicida*)Cpf1识别TTN PAM序列。已鉴别多种RNA指导的核酸酶的PAM序列,并且鉴别新颖PAM序列的策略已描述于Shmakov等人,2015,Molecular Cell[分子细胞]60,385-397,2015年11月5日中。还应注意,工程化RNA指导的核酸酶可具有与参考分子的PAM特异性不同的PAM特异性(例如,在工程化RNA指导的核酸酶的情形中,参考分子可以是衍生出RNA指导的核酸酶的天然存在的变体,或与工程化RNA指导的核酸酶具有最大氨基酸序列同源的天然存在的变体)。

[0244] 除了其PAM特异性以外,RNA指导的核酸酶可通过其DNA裂解活性来表征:天然存在的RNA指导的核酸酶典型地在靶核酸中形成DSB,但是已产生仅生成SSB的工程化变体(上文所讨论)(Ran和Hsu等人,Cell[细胞]154(6),1380-1389,2013年9月12日(Ran),通过引用并入本文),或完全不切割的工程化变体。

#### Cas9

[0245] 已确定化脓链球菌Cas9的晶体结构(Jinek 2014)以及与单分子指导RNA和靶DNA复合的金黄色葡萄球菌Cas9的晶体结构(Nishimasu 2014;Anders 2014;和Nishimasu 2015)。

[0246] 天然存在的Cas9蛋白包含两个叶:识别(REC)叶和核酸酶(NUC)叶;每一叶包含特定的结构和/或功能结构域。REC叶包含富精氨酸桥螺旋(BH)结构域,以及至少一个REC结构域(例如REC1结构域和任选地REC2结构域)。REC叶与其他已知蛋白不共享结构相似性,指示其是独特的功能结构域。不希望受限于任何理论,突变分析提出BH和REC结构域的特殊功能作用:BH结构域似乎在指导分子:DNA识别中起作用,而REC结构域被认为与指导分子的重复:抗重复双链体相互作用并且介导Cas9/指导分子复合物的形成。

[0247] NUC叶包含RuvC结构域、HNH结构域和PAM相互作用(PI)结构域。RuvC结构域与逆转录病毒整合酶超家族成员共享结构相似性,并且裂解靶核酸的非互补(即底部)链。其可从两个或更多个分裂RuvC基序形成(例如化脓链球菌和金黄色葡萄球菌中的RuvC I、RuvC II和RuvC III)。同时,HNH结构域在结构上与HNN内切核酸酶基序类似,并且裂解靶核酸的互补(即顶部)链。顾名思义,PI结构域有助于PAM特异性。

[0248] 虽然Cas9的某些功能与上文所述的特定结构域相关(但不一定完全取决于所述结构域),但这些和其他功能可以由其他Cas9结构域或任一叶上的多个结构域来介导或影响。例如,在化脓链球菌Cas9中,如在Nishimasu 2014中所述,指导分子的重复:抗重复双链体落在REC叶与NUC叶之间的沟中,并且双链体中的核苷酸与BH、PI和REC结构域中的氨基酸相互作用。第一茎环结构中的一些核苷酸也与多个结构域(PI、BH和REC1)中的氨基酸相互作用,第二和第三茎环(RuvC和PI结构域)中的一些核苷酸也是如此。

#### Cpf1

[0249] 与crRNA复合的氨基酸球菌属(*Acidaminococcus* sp.)Cpf1的晶体结构和包括

TTN PAM序列的双链(ds)DNA靶标已经由Yamano等人解析(Cell[细胞].2016年5月5日;165(4):949-962(Yamano),通过引用并入本文)。Cpf1像Cas9一样,具有两个叶:REC(识别)叶和NUC(核酸酶)叶。REC叶包括REC1和REC2结构域,其缺少与任何已知蛋白质结构的相似性。同时,NUC叶包括三个RuvC结构域(RuvC-I、-II和-III)和BH结构域。然而,与Cas9相反,Cpf1 REC叶缺少HNH结构域,并且包括同样缺少与已知蛋白质结构的相似性的其他结构域:结构上独特的PI结构域、三个楔形(WED)结构域(WED-I、-II和-III)和核酸酶(NUC)结构域。

[0250] 虽然Cas9和Cpf1共享结构和功能的相似性,但应了解,某些Cpf1活性是由与任何Cas9结构域不同的结构域介导的。例如,靶DNA的互补链的裂解似乎是由Nuc结构域介导,所述Nuc结构域在顺序上和空间上与Cas9的HNH结构域不同。另外,Cpf1指导分子的非靶向部分(把手)采用假结结构,而不是Cas9指导分子中由重复:抗重复双链体形成的茎环结构。

#### RNA指导的核酸酶的修饰

[0251] 上述RNA指导的核酸酶具有可用于多种应用的活性和特性,但技术人员将了解,RNA指导的核酸酶在某些情况下也可以经修饰,以改变裂解活性、PAM特异性或其他结构或功能特征。

[0252] 首先参看改变裂解活性的修饰,上文已描述降低或消除NUC叶内结构域活性的突变。可在RuvC结构域中、在Cas9 HNH结构域中或在Cpf1 Nuc结构域中进行的示例性突变描述于Ran和Yamano中,以及Cotta-Ramusino中。通常,降低或消除两个核酸酶结构域之一中的活性的突变导致具有切口酶活性的RNA指导的核酸酶,但应注意,切口酶活性的类型根据哪个结构域失活而变化。作为一个实例,Cas9的RuvC结构域的失活将导致裂解互补或顶部链的切口酶。

[0253] 另一方面,Cas9 HNH结构域的失活导致裂解底部或非互补链的切口酶。

[0254] 对于化脓链球菌(Kleinstiver等人,Nature[自然].2015年7月23日;523(7561):481-5(Kleinstiver I)和金黄色葡萄球菌(Kleinstiver等人,Nat Biotechnol.[自然生物技术]2015年12月;33(12):1293-1298(Kleinstiver II)),相对于天然存在的Cas9参考分子的PAM特异性的修饰已经由Kleinstiver等人描述。Kleinstiver等人还已描述改进Cas9的靶向保真性的修饰(Nature[自然],2016年1月28日;529,490-495(Kleinstiver III))。这些参考文献中的每一篇都是通过引用并入本文。

[0255] 已将RNA指导的核酸酶分裂成两个或更多个部分,如Zetsche等人(Nat Biotechnol.[自然生物技术]2015年2月;33(2):139-42(Zetsche II),通过引用并入)和Fine等人(Sci Rep.[科学报告]2015年7月1日;5:10777(Fine),通过引用并入)所描述。

[0256] 在某些实施例中,RNA指导的核酸酶可以是尺寸经优化的或截短的,例如通过一个或多个缺失来进行,所述缺失减小核酸酶的尺寸,同时仍保留指导分子关联、靶标和PAM识别以及裂解活性。在某些实施例中,RNA指导的核酸酶以共价或非共价方式,任选地通过接头结合至另一多肽、核苷酸或其他结构。示例性结合的核酸酶和接头描述于Guilinger等人,Nature Biotechnology[自然生物技术]32,577-582(2014)中,所述文献出于所有目的通过引用并入本文中。

[0257] RNA指导的核酸酶还任选地包括标签,例如但不限于核定位信号,以促进RNA指导的核酸酶蛋白移动至细胞核中。在某些实施例中,RNA指导的核酸酶可以并入C末端和/或N末端核定位信号。核定位序列是本领域中已知的并且描述于Maeder和其他文献中。

[0258] 前述修饰列表旨在具有示例性,并且技术人员根据本披露将了解,在某些应用中,其他修饰可以是可能的或期望的。因此,为简洁起见,参考特定RNA指导的核酸酶呈现本披露的示例性系统、方法和组合物,但应理解,所用RNA指导的核酸酶可以不改变其操作原理的方式经修饰。此类修饰在本披露的范围内。

#### 编码RNA指导的核酸酶的核酸

[0259] 本文提供编码RNA指导的核酸酶(例如Cas9、Cpf1或其功能片段)的核酸。先前已描述编码RNA指导的核酸酶的示例性核酸(参见,例如,Cong 2013;Wang 2013;Mali 2013;Jinek 2012)。

[0260] 在一些情形中,编码RNA指导的核酸酶的核酸可以是合成的核酸序列。例如,合成核酸分子可以经化学修饰。在某些实施例,编码RNA指导的核酸酶的mRNA将具有一种或多种(例如,所有)以下特性:其可以被加帽;多聚腺苷酸化;以及经5-甲基胞苷和/或假尿苷取代。

[0261] 合成的核酸序列也可以经密码子优化,例如,至少一个非常见密码子或较不常见的密码子已经见密码子替代。例如,合成核酸可以引导经优化信使mRNA的合成,例如,针对在哺乳动物表达系统中的表达进行优化,例如,本文所述。密码子优化的Cas9编码序列的实例呈现于Cotta-Ramusino中。

[0262] 另外,或可替代地,编码RNA指导的核酸酶的核酸可包含核定位序列(NLS)。核定位序列是本领域中已知的。

#### 候选分子的功能分析

[0263] 候选RNA指导的核酸酶、指导分子和其复合物可以通过本领域中已知的标准方法来评估。参见,例如Cotta-Ramusino。RNP复合物的稳定性可通过差示扫描荧光法来评估,如下文所述。

#### 差示扫描荧光法(DSF)

[0264] 包含指导分子和RNA指导的核酸酶的核糖核蛋白(RNP)复合物的热稳定性可以通过DSF来测量。DSF技术测量蛋白质的热稳定性,所述热稳定性可以在有利条件下(例如添加结合RNA分子,例如指导分子)增加。

[0265] DSF测定可以根据任何适宜方案来进行,并且可以用于任何适宜环境中,包括但不限于(a)测试不同条件(例如指导分子:RNA指导的核酸酶蛋白的不同化学计量比,不同缓冲溶液等)以鉴别RNP形成的最佳条件;和(b)测试RNA指导的核酸酶和/或指导分子的修饰(例如化学修饰、序列改变等)以鉴别改进RNP形成或稳定性的那些修饰。DSF测定的一个读出是RNP复合物的熔融温度的位移;相对高的位移表明,相对于特征为较低位移的参考RNP复合物,RNP复合物更稳定(并且可能因此具有更高活性或更有利的形成动力学、降解动力学或另一功能特征)。在作为筛选工具布置DSF测定时,可指定阈值熔融温度位移,使得输出是熔融温度位移等于或高于阈值的一种或多种RNP。例如,阈值可以是5°C-10°C(例如5°、6°、7°、8°、9°、10°)或更高,并且输出可以是一种或多种特征为熔融温度位移大于或等于该阈值的RNP。

[0266] DSF测定条件的两个非限制性实例陈述于下文中:

[0267] 为了确定形成RNP复合物的最佳溶液,将水+10xSYPRO Orange®(生命技术公司(Life Technologies)目录号S-6650)中固定浓度(例如2μM)的Cas9分配至384孔板中。然

后添加稀释于具有不同pH和盐的溶液中的等摩尔量的指导分子。在室温下孵育10分钟并短暂离心以移除任何气泡后,使用Bio-Rad CFX384<sup>TM</sup> Real-Time System C1000 Touch<sup>TM</sup>热循环仪和Bio-Rad CFX Manager软件运行从20°C至90°C的梯度,每10秒温度增加1°C。

[0268] 第二测定由以下步骤组成:在来自上文测定1的最适缓冲液中混合不同浓度的指导分子与固定浓度(例如2 $\mu$ M)的Cas9,并在384孔板中孵育(例如,在室温下10分钟)。添加等体积的最适缓冲液+10xSYPRO Orange<sup>®</sup> (生命技术公司目录号S-6650),并且将板用Microseal<sup>®</sup>B粘合剂(MSB-1001)密封。在短暂离心以移除任何气泡后,使用Bio-Rad CFX384<sup>TM</sup> Real-Time System C1000 Touch<sup>TM</sup>热循环仪和Bio-Rad CFX Manager软件运行从20°C至90°C的梯度,每10秒温度增加1°C。

#### 基因组编辑策略

[0269] 在本披露的各个实施例中,上述基因组编辑系统用于在细胞内或从细胞获得的DNA的靶向区中产生编辑(即改变)。本文描述产生特定编辑的各种策略,并且这些策略通常是在所需修复结果、个别编辑(例如SSB或DSB)的数目和定位以及此类编辑的靶位点方面加以描述。

[0270] 涉及SSB或DSB的形成的基因组编辑策略是通过修复结果来表征,这些修复结果包括:(a)所靶向区域的全部或部分的缺失;(b)所靶向区域的全部或部分中的插入或其替代;或(c)所靶向区域的全部或部分的中断。此分组并不旨在具有限制性,或结合于任何具体理论或模型,而仅是为了便于呈现而提供。技术人员将了解,所列结果不会相互排斥,并且一些修复可导致其他结果。除非另外指定,否则对特定编辑策略或方法的描述不应理解为需要特定的修复结果。

[0271] 所靶向区域的替代通常涉及用同源序列替代所靶向区域内现有序列的全部或部分,例如通过基因修正或基因转变来进行,两种修复结果是通过HDR路径介导。HDR是通过使用供体模板来促进,所述供体模板可以是单链或双链的,如下文更详细地描述。单链或双链模板可以是外源的,在这种情形中其将促进基因修正,或者所述模板可以是内源的(例如细胞基因组内的同源序列),以促进基因转变。外源模板可以具有不对称悬突(即,模板中与DSB位点互补的部分可在3'或5'方向上偏移,而不是位于供体模板内的中心),例如如Richardson等人所述(Nature Biotechnology[自然生物技术]34,339-344(2016), (Richardson),通过引用并入)。在模板为单链的情况下,其可以对应于所靶向区域的互补(顶部)或非互补(底部)链。

[0272] 在一些情形中,通过在所靶向区域中或周围形成一个或多个切口来促进基因转变和基因修正,如在Ran和Cotta-Ramusino中所述。在一些情形中,双重切口酶策略用于形成两个偏移SSB,其又形成具有悬突(例如5'悬突)的单一DSB。

[0273] 所靶向序列的全部或部分的中断和/或缺失可通过多种修复结果来实现。作为一个实例,可通过同时产生两个或更多个侧接所靶向区域的DSB使序列缺失,然后在修复DSB时切除所述所靶向区域,如在Maeder中针对LCA10突变所述。作为另一实例,可在修复前通过以下方式产生的缺失来中断序列:形成具有单链悬突的双链断裂,之后对悬突进行核酸外切加工。

[0274] 靶序列中断的一个特定子集是通过在所靶向序列内形成indel来介导,其中修复结果典型地是通过NHEJ路径(包括Alt-NHEJ)来介导。NHEJ由于其与indel突变的关联而称

为“错误倾向”修复路径。然而,在一些情形中,DSB是通过NHEJ修复,并且不改变其周围的序列(所谓的“完美”或“无瘢痕”修复);这通常需要DSB的两端完美连接。同时,Indel被认为是从游离DNA末端在连接前的酶加工产生,其在一个或两个游离末端的一条或两条链中添加和/或移除核苷酸。

[0275] 由于游离DSB末端的酶加工可具有随机性,indel突变往往是可变的,沿分布发生,并且可能受到多种因素影响,包括特定靶位点、所用细胞类型、所用基因组编辑策略等。即使如此,有可能引起关于indel形成的有限泛化:通过修复单一DSB形成的缺失最常在1-50bp范围内,但是可能达到大于100-200bp。通过修复单一DSB形成的插入往往较短,并且常包括紧密围绕断裂位点的序列的短重复。然而,有可能获得大的插入,并且在这些情形中,插入的序列通常已经被追溯至基因组的其他区域或追溯至存在于细胞中的质粒DNA。

[0276] Indel突变和经配置以产生indel的基因组编辑系统可用于例如在不产生特定的最终序列时和/或在可耐受移码突变的情况下中断靶序列。其还可以用于偏好特定序列的环境中,只要某些所需序列往往优先通过给定位点处的SSB或DSB的修复而发生即可。Indel突变还是可用于评估或筛选特定基因组编辑系统及其组分的活性的工具。在这些和其他环境中,indel可以通过以下各项来表征:(a) 其在与基因组编辑系统接触的细胞的基因组中的相对和绝对频率,和(b) 相对于未编辑序列的数值差异的分布,例如 $\pm 1$ 、 $\pm 2$ 、 $\pm 3$ 等。作为一个实例,在先导发现(lead-finding)环境中,可基于在受控条件下的indel读出筛选多个指导分子,以鉴别最有效驱动靶位点处的切割的那些指导分子。可以选择以阈值频率或以高于阈值的频率产生indel或产生indel的特定分布的指导以供进一步研究和开发。Indel频率和分布还可以用作读出,用于评估不同的基因组编辑系统实施或配置和递送方法,例如通过保持指导分子不变并改变某些其他反应条件或递送方法。

#### 多重策略

[0277] 虽然上文讨论的示例性策略集中于通过单一DSB介导的修复结果,但根据本披露的基因组编辑系统还可用于产生在相同基因座中或在不同基因座中的两个或更多个DSB。涉及形成多个DSB或SSB的编辑策略描述于例如Cotta-Ramusino中。

#### 供体模板设计

[0278] 供体模板设计详细描述于文献中,例如Cotta-Ramusino中。DNA寡聚体供体模板(寡脱氧核苷酸或ODN)可以是单链(ssODN)或双链(dsODN)的,可以用于促进基于HDR的DSB修复,并且尤其可用于将改变引入靶DNA序列中、将新序列插入靶序列中、或完全替代靶序列。

[0279] 不论是单链或双链,供体模板通常包括与待裂解的靶序列内或附近(例如侧接或邻近)的DNA区域同源的区域。这些同源区域在本文中称作“同源臂”,并且示意性地示于下文中:

[5'同源臂]-[替代序列]-[3'同源臂]。

[0280] 同源臂可具有任何适宜长度(如果仅使用一个同源臂,包括0个核苷酸),并且3'和5'同源臂可以具有相同长度或可以具有不同长度。适当同源臂长度的选择可能受到多种因素影响,例如对避免与某些序列(例如Alu重复序列或其他极为常见的元件)的同源性或微同源性的期望。例如,可以缩短5'同源臂以避免序列重复元件。在其他实施例中,可以缩短3'同源臂以避免序列重复元件。在一些实施例中,可以同时缩短5'和3'同源臂以避免包括

某些序列重复元件。另外,一些同源臂设计可以改进编辑效率或增加所需修复结果的频率。例如,Richardson等人(Nature Biotechnology[自然生物技术]34,339-344(2016)(Richardson),所述文献是通过引用并入)发现,单链供体模板的3'和5'同源臂的相对不对称性影响修复率和/或结果。

[0281] 供体模板中的替代序列已描述于其他文献中,包括Cotta-Ramusino等人中。替代序列可以是任何适宜长度(如果所需修复结果是缺失,那么包括0个核苷酸),并且相对于需要编辑的细胞内的天然存在的序列,典型地包括1、2、3或更多个序列修饰。一种常见序列修饰涉及改变天然存在的序列以修复突变,所述突变与需要治疗的疾病或病症相关。另一常见序列修饰涉及改变一个或多个序列,所述序列与RNA指导的核酸酶的PAM序列或用于产生SSB或DSB的一个或多个指导分子的靶向结构域互补或编码所述PAM序列或靶向结构域,以在将替代序列并入靶位点中之后减少或消除靶位点的重复裂解。

[0282] 如果使用线性ssODN,其可以经配置以(i)退火至靶核酸的带切口链,(ii)退火至靶核酸的完整链,(iii)退火至靶核酸的正链,和/或(iv)退火至靶核酸的负链。ssODN可具有任何适宜长度,例如约、至少或不大于150-200个核苷酸(例如,150、160、170、180、190或200个核苷酸)。

[0283] 应注意,模板核酸也可以是核酸载体,例如病毒基因组或环状双链DNA,例如质粒。包含供体模板的核酸载体可以包括其他编码或非编码元件。例如,模板核酸可以作为病毒基因组的部分来递送(例如在AAV或慢病毒基因组中),其包括某些基因组骨架元件(例如在AAV基因组情形中,末端反向重复序列)并且任选地包括编码指导分子和/或RNA指导的核酸酶的其他序列。在某些实施例中,供体模板可以邻近或侧接由一个或多个指导分子识别的靶位点,以促进在供体模板的一端或两端上形成游离DSB,所述供体模板可以参与使用相同指导分子修复在细胞DNA中形成的相应SSB或DSB。适合用作供体模板的示例性核酸载体描述于Cotta-Ramusino中。

[0284] 不论使用何种形式,模板核酸都可以设计为避免不期望的序列。在某些实施例中,可以缩短一个或两个同源臂以避免与某些序列重复元件(例如,Alu重复、LINE元件等)重叠。

### 靶细胞

[0285] 根据本披露的基因组编辑系统可以用于操作或改变细胞,例如以编辑或改变靶核酸。在各个实施例中,操作可在体内进行或离体进行。

[0286] 可以根据本披露的实施例操作或改变多种细胞类型,并且在一些情形中,例如在体内应用中,例如通过将根据本披露的基因组编辑系统递送至多个细胞类型来改变或操作多个细胞类型。然而,在其他情形中,可能需要将操作或改变限制于特定的一个或多个细胞类型。例如,在一些情况下可能需要编辑具有有限分化潜力的细胞或最终分化细胞,例如在Maeder情形中的感光细胞,其中预期基因型的修饰会导致细胞表型变化。然而,在其他情形中,可能需要编辑分化程度较低的、多潜能或多能性的干细胞或祖细胞。举例来说,细胞可以是胚胎干细胞、诱导型多能干细胞(iPSC)、造血干细胞/祖细胞(HSPC)或其他干细胞或祖细胞类型,其分化为与给定应用或适应证相关的细胞类型。

[0287] 作为推论,根据所靶向的一种或多种细胞类型和/或所需编辑结果,所改变或操作的细胞不同地为分裂细胞或非分裂细胞。

[0288] 在离体操作或改变细胞时,细胞可以立即使用(例如给予受试者),或可以维持或储存细胞以供将来使用。本领域技术人员将了解,可以使用本领域已知的任何适宜方法将细胞维持在培养中或储存(例如冷冻于液氮中)。

#### 基因组编辑系统的实施:递送、配制和给予途径

[0289] 如上文所讨论的,本披露的基因组编辑系统可以用任何适宜方式来实现,意味着此类系统的组分(包括但不限于RNA指导的核酸酶、指导分子和可选供体模板核酸)可以用任何适宜形式或形式的组合来递送、配制或给予,从而在细胞、组织或受试者中导致基因组编辑系统的转导、表达或引入和/或引起所需的修复结果。表5和6展示基因组编辑系统实施的若干非限制性实例。但本领域技术人员将了解,这些列表不是综合性的,并且其他实施是可能的。尤其参照表5,该表列示包含单一指导分子和可选供体模板的基因组编辑系统的若干示例性实施。然而,根据本披露的基因组编辑系统可以并入多个指导分子、多个RNA指导的核酸酶以及其他组分,例如蛋白质,并且基于该表中所示的原理,多种实施将为技术人员所了解。在该表中,[N/A]指示,基因组编辑系统不包括所指示的组分。

表5

基因组编辑系统组分			
RNA 指导的 核酸酶	指导分子	供体模板	注解
蛋白质	RNA	[N/A]	与 gRNA 分子复合的 RNA 指导的核酸酶蛋白（RNP 复合物）
蛋白质	RNA	DNA	如上所述的 RNP 复合物加单链或双链供体模板。
蛋白质	DNA	[N/A]	RNA 指导的核酸酶蛋白加从 DNA 转录的 gRNA。
蛋白质	DNA	DNA	RNA 指导的核酸酶蛋白加 gRNA 编码 DNA 和单独的 DNA 供体模板。
蛋白质	DNA		RNA 指导的核酸酶蛋白以及编码 gRNA 和供体模板二者的单一 DNA。
DNA			编码 RNA 指导的核酸酶、gRNA 和供体模板的 DNA 或 DNA 载体。

DNA	DNA	[N/A]	两个单独的 DNA 或两个单独的 DNA 载体，其分别编码 RNA 指导的核酸酶和 gRNA。
DNA	DNA	DNA	三个单独的 DNA 或三个单独的 DNA 载体，其分别编码 RNA 指导的核酸酶、gRNA 和供体模板。
DNA		[N/A]	编码 RNA 指导的核酸酶和 gRNA 的 DNA 或 DNA 载体
DNA		DNA	编码 RNA 指导的核酸酶和 gRNA 的第一 DNA 或 DNA 载体，以及编码供体模板的第二 DNA 或 DNA 载体。
DNA	DNA		编码 RNA 指导的核酸酶的第一 DNA 或 DNA 载体以及编码 gRNA 和供体模板的第二 DNA 或 DNA 载体。
DNA		DNA	编码 RNA 指导的核酸酶和供体模板的第一 DNA 或 DNA 载体，以及编码 gRNA 的第二 DNA 或 DNA 载体
DNA		RNA	编码 RNA 指导的核酸酶和供体模板的 DNA 或 DNA 载体，以及 gRNA
RNA		[N/A]	RNA 或 RNA 载体，其编码 RNA 指导的核酸酶并包含 gRNA
RNA		DNA	编码 RNA 指导的核酸酶并包含 gRNA 的 RNA 或 RNA 载体，以及编码供体模板的 DNA 或 DNA 载体。

[0290] 表6归纳用于如本文所述的基因组编辑系统的组分的各种递送方法。同样，该列表旨在是示例性而不是限制性。

表6

递送载体/方式	递送到非分裂细胞中	表达的持续时间	基因组整合	所递送的分子类型
物理的（例如，电穿孔、基因枪、磷酸钙转染、细	是	瞬时的	否	核酸和蛋白质

<b>胞压缩或挤压)</b>					
<b>病毒的</b>	<b>逆转录病毒</b>	否	稳定的	是	RNA
	<b>慢病毒</b>	是	稳定的	是/否 具有修饰	RNA
	<b>腺病毒</b>	是	瞬时的	否	DNA
	<b>腺相关病毒 (AAV)</b>	是	稳定的	否	DNA
	<b>痘苗病毒</b>	是	非常瞬时的	否	DNA
	<b>单纯疱疹病毒</b>	是	稳定的	否	DNA
<b>非病毒的</b>	<b>阳离子脂质体</b>	是	瞬时的	取决于递送什么	核酸和蛋白质
	<b>聚合物纳米颗粒</b>	是	瞬时的	取决于递送什么	核酸和蛋白质
<b>生物非病毒递送运载体</b>	<b>减毒细菌</b>	是	瞬时的	否	核酸
	<b>工程化的噬菌体</b>	是	瞬时的	否	核酸
	<b>哺乳动物病毒样颗粒</b>	是	瞬时的	否	核酸
	<b>生物脂质体：红细胞血影和外泌体</b>	是	瞬时的	否	核酸

#### 基因组编辑系统的基于核酸的递送

[0291] 可以通过本领域已知的方法或如本文所述将编码根据本披露的基因组编辑系统的各种元件的核酸给予受试者或递送至细胞。例如,可以通过例如载体(例如,病毒或非病毒载体)、非载体基方法(例如,使用裸DNA或DNA复合物)或其组合递送编码RNA指导的核酸酶的DNA和/或编码指导分子的DNA、以及供体模板核酸。

[0292] 编码基因组编辑系统或其组分的核酸可以作为裸DNA或RNA直接递送至细胞,例如借助转染或电穿孔来递送,或者可以缀合至促进靶细胞(例如,红血球、HSC)摄取的分子(例如,N-乙酰半乳糖胺)。也可以使用核酸载体,例如表6中归纳的载体。

[0293] 核酸载体可以包含编码基因组编辑系统组分(例如RNA指导的核酸酶、指导分子和/或供体模板)的一个或多个序列。载体还可以包含编码信号肽(例如,用于核定位、核仁定位或线粒体定位)的序列,其与编码蛋白质的序列缔合(例如插入其中或与其融合)。作为一个实例,核酸载体可以包括Cas9编码序列,其包括一个或多个核定位序列(例如,来自SV40的核定位序列)。

[0294] 核酸载体还可以包括任何适宜数目的调节/控制元件,例如,启动子、增强子、内含子、多聚腺苷酸化信号、Kozak共有序列或内部核糖体进入位点(IRES)。这些元件为本领域中所熟知,并且描述于Cotta-Ramusino中。

[0295] 根据本披露的核酸载体包括重组病毒载体。示例性病毒载体展示于表6中,并且其他适宜病毒载体以及其使用和产生描述于Cotta-Ramusino中。还可以使用本领域已知的其他病毒载体。另外,可以使用病毒颗粒来递送呈核酸和/或肽形式的基因组编辑系统组分。例如,可以组装“空”病毒颗粒以含有任何适宜负荷。病毒载体和病毒颗粒也可以经工程化以并入靶向配体,从而改变靶组织特异性。

[0296] 除了病毒载体以外,可以使用非病毒载体来递送编码根据本披露的基因组编辑系统的核酸。非病毒核酸载体的一个重要分类是纳米颗粒,其可以有机的或无机的。纳米颗粒为本领域所熟知,并且概述于Cotta-Ramusino中。可以使用任何适宜的纳米颗粒设计来递送基因组编辑系统组分或编码此类组分的核酸。例如,在本披露的某些实施例中,有机(例如脂质和/或聚合物)纳米颗粒可以适合用作递送运载体。用于纳米颗粒配制物和/或基因转移的示例性脂质显示于表7中,并且表8列示用于基因转移和/或纳米颗粒配制物的示例性聚合物。

表7:用于基因转移的脂质

脂质	缩写	特征
1,2-二油酰基-sn-甘油基-3-磷脂酰胆碱	DOPC	辅助物
1,2-二油酰基-sn-甘油基-3-磷脂酰乙醇胺	DOPE	辅助物
胆固醇		辅助物
N-[1-(2,3-二油烯基氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子的
1,2-二油烯基氧基-3-三甲基铵-丙烷	DOTAP	阳离子的
双十八烷基氨基甘氨酸精胺	DOGS	阳离子的
N-(3-氨基丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-溴化丙铵	GAP-DLRIE	阳离子的

溴化十六烷基三甲基铵	CTAB	阳离子的
6-月桂氧己基鸟氨酸	LHON	阳离子的
1-(2,3-二油酰氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子的
2,3-二油烯基氧基-N-[2(精胺甲酰胺基-乙基)-N,N-二甲基-1-三氟乙酸丙铵	DOSPA	阳离子的
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子的
N-(2-羟基乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻酰氧基丙基二甲基羟基乙基溴化铵	DMRI	阳离子的
3β-[N-(N',N'-二甲基氨基乙烷)-氮甲酰基]胆固醇	DC-Chol	阳离子的
双-胍鎓-三氮乙基胺-胆固醇	BGTC	阳离子的
1,3-双脱氧基-2-(6-羧基-精胺基)-丙酰胺	DOSPER	阳离子的
二甲基十八烷基溴化铵	DDAB	阳离子的
双十八烷基氨基甘氨酸酰基亚精胺	DSL	阳离子的
外消旋-[(2,3-双十八烷氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子的
外消旋-[2(2,3-双十六烷氧基丙基-氧基甲氧基)乙基]三甲基溴化铵	CLIP-6	阳离子的
乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子的
1,2-二硬脂氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子的
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子的
O,O'-二肉豆蔻基-N-赖氨酸酰基天冬氨酸酯	DMKE	阳离子的
1,2-二硬脂酰基-sn-甘油基-3-乙基磷酸胆碱	DSEPC	阳离子的
N-棕榈酰 D-赤-鞘氨醇基氮甲酰基-精胺	CCS	阳离子的
N-叔丁基-N'-十四烷基-3-十四烷基氨基丙脒	二 C14-脒	阳离子的
十八碳烯醇基氧基[乙基-2-十七烯基-3 羟基乙基]氯化咪唑啉鎓	DOTIM	阳离子的
N1-胆固醇氧基羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子的
2-(3-[双(3-氨基-丙基)-氨基]丙胺基)-N-双十四烷基氨基甲酰基甲-乙基-乙酰胺	RPR209120	阳离子的
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子的
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子的
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子的

表8:用于基因转移的聚合物

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP

二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚( <i>N</i> -乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚( $\beta$ -氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚( $\alpha$ -[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚( <i>N</i> -乙基-4-乙烯基溴化吡啶鎓)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚( <i>N</i> -2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲基氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨基乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
<i>N</i> -十二烷基化的壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0297] 非病毒载体任选地包括靶向修饰以改进摄取和/或选择性靶向某些细胞类型。这些靶向修饰可以包括例如细胞特异性抗原、单克隆抗体、单链抗体、适体、聚合物、糖(例如, *N*-乙酰半乳糖胺(GalNAc))和细胞穿透肽。此类载体还任选地使用致融性和核内体去稳定肽/聚合物,经历酸触发的构象变化(例如,加速负荷的核内体逃逸),和/或并入刺激可裂解的聚合物,例如用于在细胞区室中释放。例如,可以使用在还原性细胞环境中裂解的基于二硫化物的阳离子型聚合物。

[0298] 在某些实施例中,除了基因组编辑系统的组分(例如,本文所述的RNA指导的核酸酶组分和/或指导分子组分)以外,递送一种或多种核酸分子(例如,DNA分子)。在某些实施例中,该核酸分子是与基因组编辑系统的一个或多个组分同时递送。在某些实施例中,该核酸分子是在递送基因组编辑系统的一个或多个组分之前或之后(例如,小于约30分钟、1小时、2小时、3小时、6小时、9小时、12小时、1天、2天、3天、1周、2周或4周)递送。在某些实施例中,该核酸分子是通过与递送基因组编辑系统的一个或多个组分(例如,RNA指导的核酸酶组分和/或指导分子组分)不同的方式来递送。该核酸分子可以通过本文所述的任何递送方法来递送。例如,该核酸分子可以通过病毒载体(例如,整合缺陷型慢病毒)来递送,并且RNA指导的核酸酶分子组分和/或指导分子组分可以通过电穿孔来递送,例如,使得可以降低由核酸(例如,DNA)引起的毒性。在某些实施例中,该核酸分子编码治疗性蛋白质,例如,本文

所述的蛋白质。在某些实施例中,该核酸分子编码RNA分子,例如,本文所述的RNA分子。

#### RNP和/或编码基因组编辑系统组分的RNA的递送

[0299] 可以通过本领域已知的方法将RNP(指导分子和RNA指导的核酸酶的复合物)和/或编码RNA指导的核酸酶和/或指导分子的RNA递送至细胞中或给予受试者,其中一些方法描述于Cotta-Ramusino中。在体外,编码RNA指导的核酸酶的和/或编码指导分子的RNA可以通过例如显微注射、电穿孔、瞬时细胞压缩或挤压来递送(参见,例如, Lee 2012)。还可以使用脂质介导的转染、肽介导的递送、GalNAc或其他缀合物介导的递送和其组合进行体外和体内递送。

[0300] 在体外,通过电穿孔递送包含将细胞与编码RNA指导的核酸酶和/或指导分子的RNA(具有或不具有供体模板核酸分子)在盒、室或比色皿中混合,并且施加一个或多个限定持续时间和幅度的电脉冲。用于电穿孔的系统和方案为本领域中已知,并且任何适宜的电穿孔工具和/或方案可以结合本披露的各个实施例使用。

#### 给予途径

[0301] 可以通过任何适宜模式或途径(局部或全身)将基因组编辑系统或使用此类系统改变或操作的细胞给予受试者。全身给予模式包括口服和肠胃外途径。肠胃外途径包括例如静脉内、骨髓内、动脉内、肌内、真皮内、皮下、鼻内和腹膜内途径。全身给予的组分可以经修饰或配制以靶向例如HSC(造血干细胞/祖细胞)或红系祖细胞或前体细胞。

[0302] 局部给予模式包括例如骨髓内注射至骨小梁中或股骨内注射到髓腔中,以及输注至门静脉中。在某些实施例中,与全身给予(例如,静脉内)时相比,在局部给予(例如,直接给予至骨髓中)时,显著较少量的组分(与全身方法相比)可以发挥作用。局部给予模式可以降低或消除在全身性给予治疗有效量的组分时可能发生的潜在毒副作用的发生率。

[0303] 给予可以作为定期推注(例如静脉内)或作为持续输注从内部储库或从外部储库(例如从静脉内注射袋或可植入泵)提供。组分可以局部给予,例如,通过从持续释放药物递送装置中连续释放。

[0304] 另外,组分可以经配制以容许经延长时段释放。释放系统可以包括生物可降解材料或通过扩散释放所并入组分的材料的基质。组分可以均匀或非均匀地分布于释放系统内。多种释放系统可以是有用的,但适当系统的选择将取决于具体应用所需要的释放速率。不可降解和可降解的释放系统均可以被使用。适宜释放系统包括聚合物和聚合基质、非聚合基质或无机和有机赋形剂和稀释剂,例如但不限于碳酸钙和糖(例如,海藻糖)。释放系统可以是天然的或合成的。然而,合成释放系统是优选的,因为其通常更可靠、更具可重现性并且产生更明确的释放曲线。可以选择释放系统材料,使得具有不同分子量的组分是通过穿过材料扩散或通过材料降解而释放。

[0305] 代表性的合成生物可降解聚合物包括例如:聚酰胺,例如聚(氨基酸)和聚(肽);聚酯,例如聚(乳酸)、聚(乙醇酸)、聚(乳酸-共-乙醇酸)和聚(己内酯);聚(酸酐);聚原酸酯;聚碳酸酯;和其化学衍生物(取代、添加化学基团,例如,烷基、亚烷基、羟化、氧化和本领域技术人员常规进行的其他修饰)、共聚物和其混合物。代表性的合成非生物可降解聚合物包括例如:聚醚,例如聚(氧化乙烯)、聚(乙二醇)和聚(环氧丁烷);乙烯基聚合物-聚丙烯酸酯和聚甲基丙烯酸酯,例如甲基、乙基、其他烷基、甲基丙烯酸羟乙酯、丙烯酸和甲基丙烯酸和其他,例如聚(乙烯醇)、聚(乙烯基吡咯烷酮)和聚(乙酸乙烯酯);聚(氨酯);纤维素和其衍

生物,例如烷基、羟烷基、醚、酯、硝化纤维素和各种醋酸纤维素;聚硅氧烷;和其任何化学衍生物(取代、添加化学基团,例如,烷基、亚烷基、羟化、氧化和本领域技术人员常规进行的其他修饰)、共聚物和其混合物。

[0306] 也可以使用聚(丙交酯共乙交酯)微球。典型地,微球是由乳酸和乙醇酸的聚合物构成,其经结构化以形成空心球体。球体的直径可以是约15-30微米,并且可以加载本文所述的组分。

#### 组分的双模或差别递送

[0307] 技术人员根据本披露将了解,本文所披露的基因组编辑系统的不同组分可以一起或分开并且同时或不同时递送。可能尤其期望基因组编辑系统组分的分开和/或异步递送以提供对基因组编辑系统功能的时间或空间控制并限制由其活性引起的某些效应。

[0308] 如本文所用的不同或差别模式是指递送模式,这些递送模式赋予受试组分分子(例如, RNA 指导的核酸酶分子、指导分子、模板核酸或有效负载)不同的药效学或药物代谢动力学特性。例如,递送模式可以导致不同的组织分布,不同的半衰期或不同的时间分布,例如,在所选区室、组织或器官中。

[0309] 一些递送的模式(例如,通过例如通过自主复制或插入细胞核酸中而在细胞或细胞子代中持续存在的核酸载体来递送)导致组分更持久的表达和存在。实例包括病毒(例如, AAV 或慢病毒)递送。

[0310] 举例来说,基因组编辑系统的组分(例如, RNA 指导的核酸酶和指导分子)可以通过在所递送组分在体内或在特定区室、组织或器官中的所得半衰期或持久性方面不同的模式进行递送。在某些实施例中,指导分子可以通过此类模式进行递送。RNA 指导的核酸酶分子组分可以通过导致在身体或特定区室或组织或器官中更低持久性或更少暴露的模式进行递送。

[0311] 更通常地,在某些实施例中,使用第一递送模式来递送第一组分,并且使用第二递送模式来递送第二组分。第一递送模式赋予第一药效学或药物代谢动力学特性。第一药效学特性可以是例如组分或编码该组分的核酸在体内、区室、组织或器官中的分布、持久性或暴露。第二递送模式赋予第二药效学或药物代谢动力学特性。第二药效学特性可以是例如组分或编码该组分的核酸在体内、区室、组织或器官中的分布、持久性或暴露。

[0312] 在某些实施例中,第一药效学或药物代谢动力学特性(例如,分布、持久性或暴露)比第二药效学或药物代谢动力学特性更有限。

[0313] 在某些实施例中,第一递送模式经选择以优化(例如,最小化)药效学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0314] 在某些实施例中,第二递送模式经选择以优化(例如,最大化)药效学或药物代谢动力学特性(例如,分布、持久性或暴露)。

[0315] 在某些实施例中,第一递送模式包含使用相对持久的元件,例如,核酸,例如,质粒或病毒载体,例如, AAV 或慢病毒。由于此类载体相对持久,从其转录的产物将相对持久。

[0316] 在某些实施例中,第二递送模式包含相对短暂的元件,例如, RNA 或蛋白质。

[0317] 在某些实施例中,第一组分包含指导分子,并且递送模式相对持久,例如,指导分子是从质粒或病毒载体(例如, AAV 或慢病毒)转录。这些基因的转录将具有极小生理学意义,因为这些基因不编码蛋白质产物,并且这些指导分子不能够单独起作用。第二组分(RNA

指导的核酸酶分子)是以瞬时方式递送,例如,作为mRNA或作为蛋白质递送,从而确保完整的RNA指导的核酸酶分子/指导分子复合物仅在短时段内存在并有活性。

[0318] 此外,这些组分可以不同的分子形式或用不同的互为补充以增强安全性和组织特异性的递送载体进行递送。

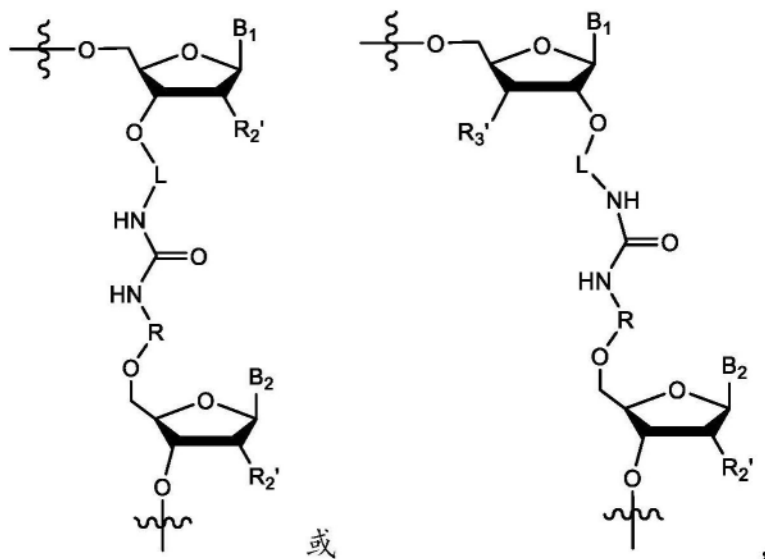
[0319] 差别递送模式的使用可以增强性能、安全性和/或功效,例如,可以减少最终脱靶修饰的可能性。通过较不持久的模式递送免疫原性组分(例如,Cas9分子)可以降低免疫原性,因为来自细菌源Cas酶的肽通过MHC分子展示于细胞表面上。两部分式递送系统可以改善这些缺点。

[0320] 差别递送模式可以用于将组分递送至不同但重叠的靶区域。在靶区域的重叠以外,活性复合物的形成被最小化。因此,在某些实施例中,第一组分(例如,指导分子)是通过第一递送模式进行递送,其导致第一空间(例如,组织)分布。第二组分(例如,RNA指导的核酸酶分子)是通过第二递送模式进行递送,其导致第二空间(例如,组织)分布。在某些实施例中,第一模式包含选自脂质体、纳米颗粒(例如,聚合纳米颗粒)和核酸的第一元件,例如,病毒载体。第二模式包含选自该组的第二元件。在某些实施例中,第一递送模式包含第一靶向元件,例如,细胞特异性受体或抗体,并且第二递送模式不包括该元件。在某些实施例中,第二递送模式包含第二靶向元件,例如,第二细胞特异性受体或第二抗体。

[0321] 当在病毒递送载体、脂质体或聚合纳米颗粒中递送RNA指导的核酸酶分子时,存在递送至多个组织并且在多个组织中具有治疗活性的可能性,但此时可能希望仅靶向单一组织。两部分式递送系统可以解决这一挑战并且增强组织特异性。如果将指导分子和RNA指导的核酸酶分子包装于具有不同但重叠的组织向性的分开的递送运载体中,那么完全功能复合物仅在两种载体所靶向的组织中形成。

非限制性实施方式

实施方式1.一种用于CRISPR系统的合成的单分子指导分子,其中所述指导分子包含具有下式的化学键合:



其中:

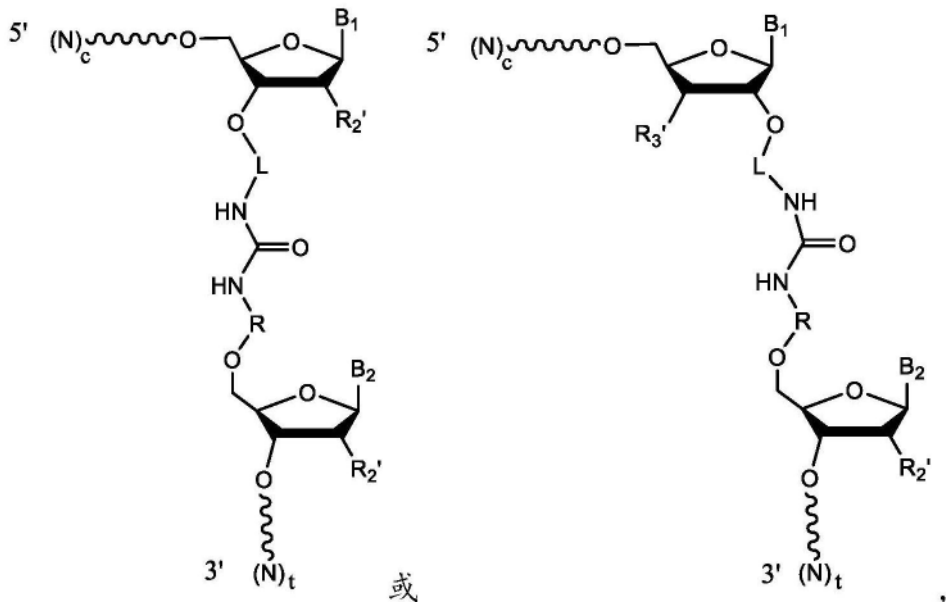
$R_2'$  和  $R_3'$  各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$  或O- $R'$ , 其中每个 $R'$ 独立地是保护基团或烷基,其中所述烷基可任选地被取代;

L和R各自独立地是非核苷酸接头;并且

$B_1$ 和 $B_2$ 各自独立地是核碱基。

实施方式2.根据实施方式1所述的指导分子,其中R<sub>2</sub>'选自由H、氟和O-R'组成的组,其中R'是保护基团或任选地被取代的烷基。

实施方式3.根据实施方式1或2所述的指导分子,其中所述指导分子具有下式:



其中：

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

(N)<sub>2</sub>包括与(N)<sub>1</sub>的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

$t$ 是20或更大的整数;并且

每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式4. 根据实施方式3所述的指导分子, 其中  $(N)_c$  和  $(N)_t$  中的每个N独立地是核糖核苷酸残基或糖修饰的核糖核苷酸残基。

实施方式5. 根据实施方式3或4所述的指导分子, 其中(N)<sub>c</sub>或(N)<sub>t</sub>包含一个或多个脱氧核糖核苷酸残基。

实施方式6.根据实施方式3-5中任一项所述的指导分子,其中(N)<sub>c</sub>或(N)<sub>t</sub>包含一个或多个2'-O-甲基修饰的核糖核苷酸残基。

实施方式7.根据实施方式3-6中任一项所述的指导分子,其中(N)<sub>c</sub>的5'末端的三个核苷酸中的每一个和/或(N)<sub>t</sub>的3'末端的三个核苷酸中的每一个包含2'-O-甲基修饰的核糖核苷酸残基,所述2'-O-甲基修饰的核糖核苷酸残基通过硫代磷酸酯键合与其相邻的一个或多个核苷酸连接。

实施方式8. 根据实施方式1-7中任一项所述的指导分子, 其中L和R各自独立地是非核苷酸接头, 所述非核苷酸接头包含选自下组的部分, 该组由以下组成: 聚乙烯、聚丙烯、

聚乙二醇和聚丙二醇。

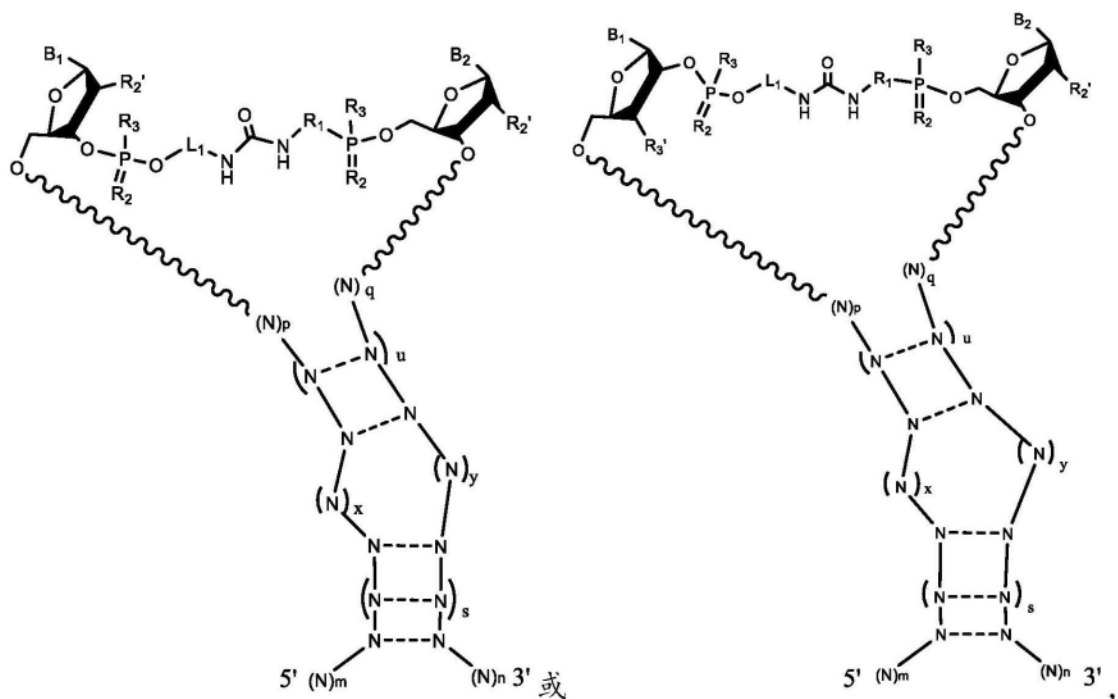
实施方式9.根据实施方式3-8中任一项所述的指导分子,其中所述指导分子用于II型CRISPR系统,并且(N)<sub>q</sub>包括5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

实施方式10.根据实施方式3-9中任一项所述的指导分子,其中(N)<sub>t</sub>包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式11.根据实施方式1-10中任一项所述的指导分子,其中所述指导分子能够与Cas9分子相互作用并介导Cas9/指导分子核糖核蛋白复合物的形成。

实施方式12.根据实施方式3-11中任一项所述的指导分子,其中(N)<sub>c</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式13.根据实施方式1-12中任一项所述的指导分子,其中所述指导分子具有下式:



其中:

L<sub>1</sub>和R<sub>1</sub>各自独立地是非核苷酸接头;

每个R<sub>2</sub>独立地是O或S;

每个R<sub>3</sub>独立地是O<sup>-</sup>或COO<sup>-</sup>;

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u是2到22之间的整数,包括端值;

s是1到10之间的整数,包括端值;

x是1到3之间的整数,包括端值;

y是>x且3到5之间的整数,包括端值;

m是15或更大的整数;

n是30或更大的整数;

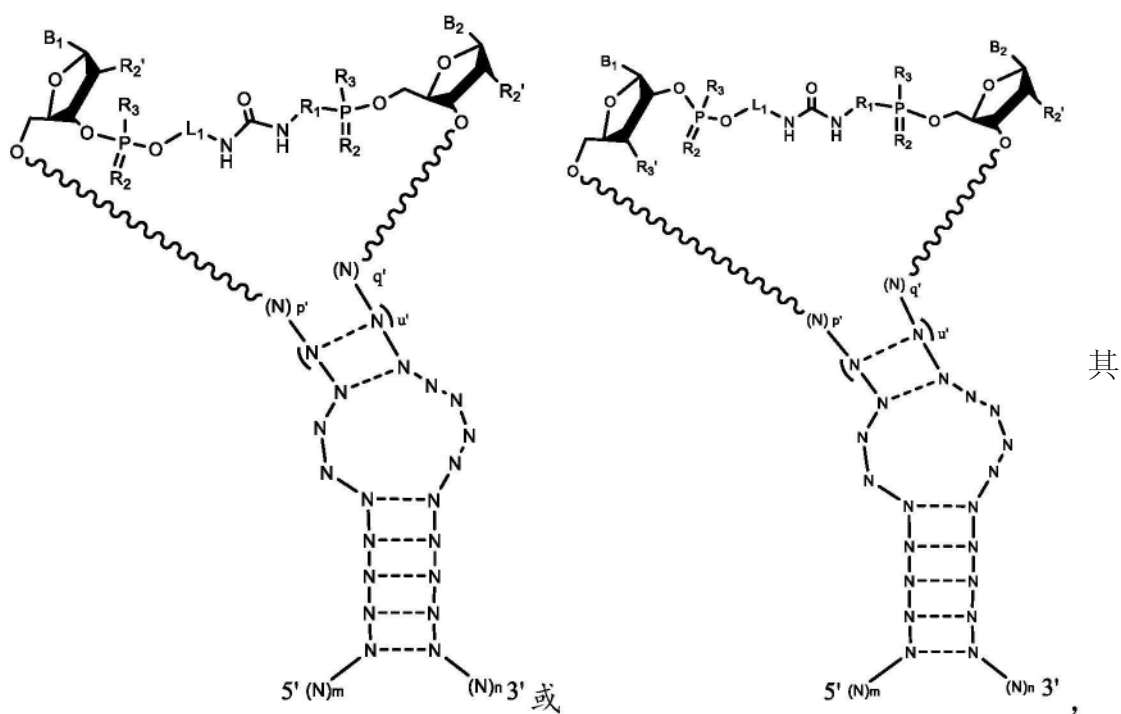
每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。

实施方式14.根据实施方式13所述的指导分子,其中p和q各为0。

实施方式15.根据实施方式13所述的指导分子,其中u是3到22之间的整数,包括端值。

实施方式16.根据实施方式1-15中任一项所述的指导分子,其中所述指导分子具有下式:

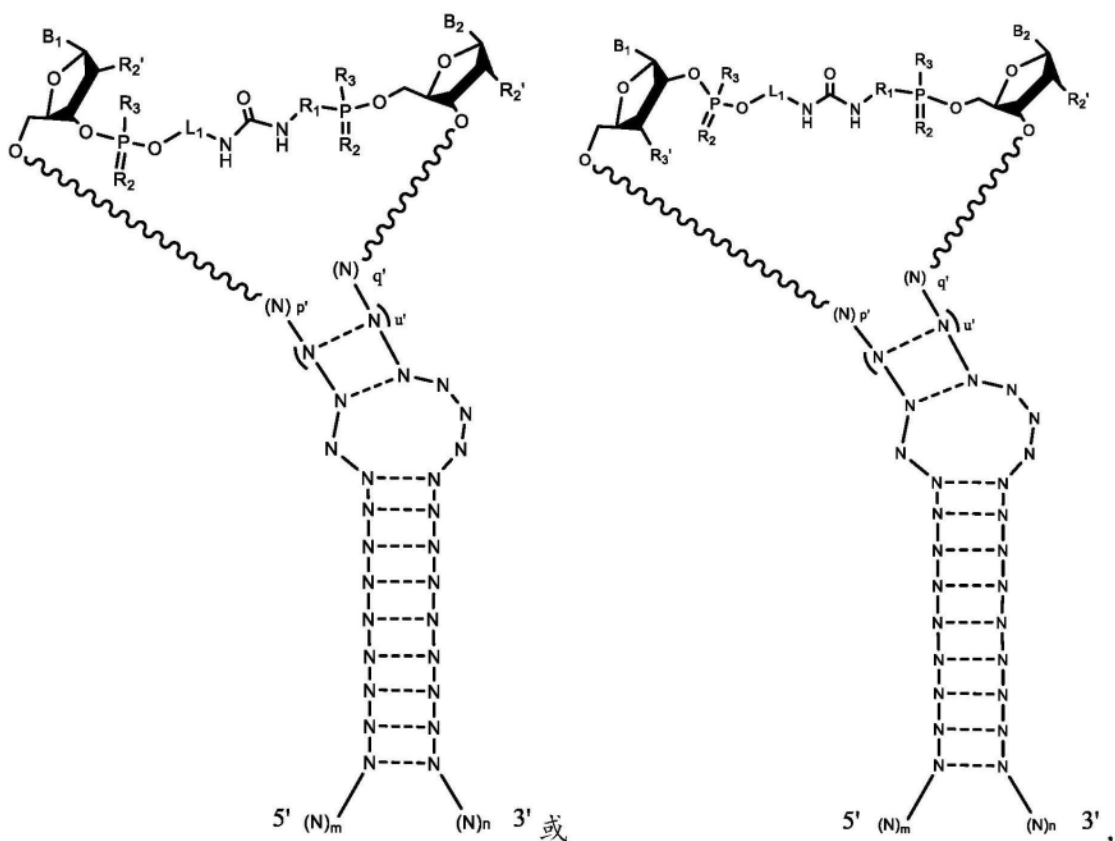


中:

u' 是2到22之间的整数,包括端值;并且

p' 和q' 各自独立地是0到4之间的整数,包括端值;并且p' +q' 是0到4之间的整数,包括端值。

实施方式17.根据实施方式1-15中任一项所述的指导分子,其中所述指导分子具有下式:



其中：

$u'$  是2到22之间的整数, 包括端值; 并且

$p'$  和  $q'$  各自独立地是0到4之间的整数, 包括端值; 并且  $p' + q'$  是0到4之间的整数, 包括端值。

实施方式18. 根据实施方式14或15所述的指导分子, 其中  $p'$  和  $q'$  各为0。

实施方式19. 根据实施方式14或15所述的指导分子, 其中  $u'$  是3到22之间的整数, 包括端值。

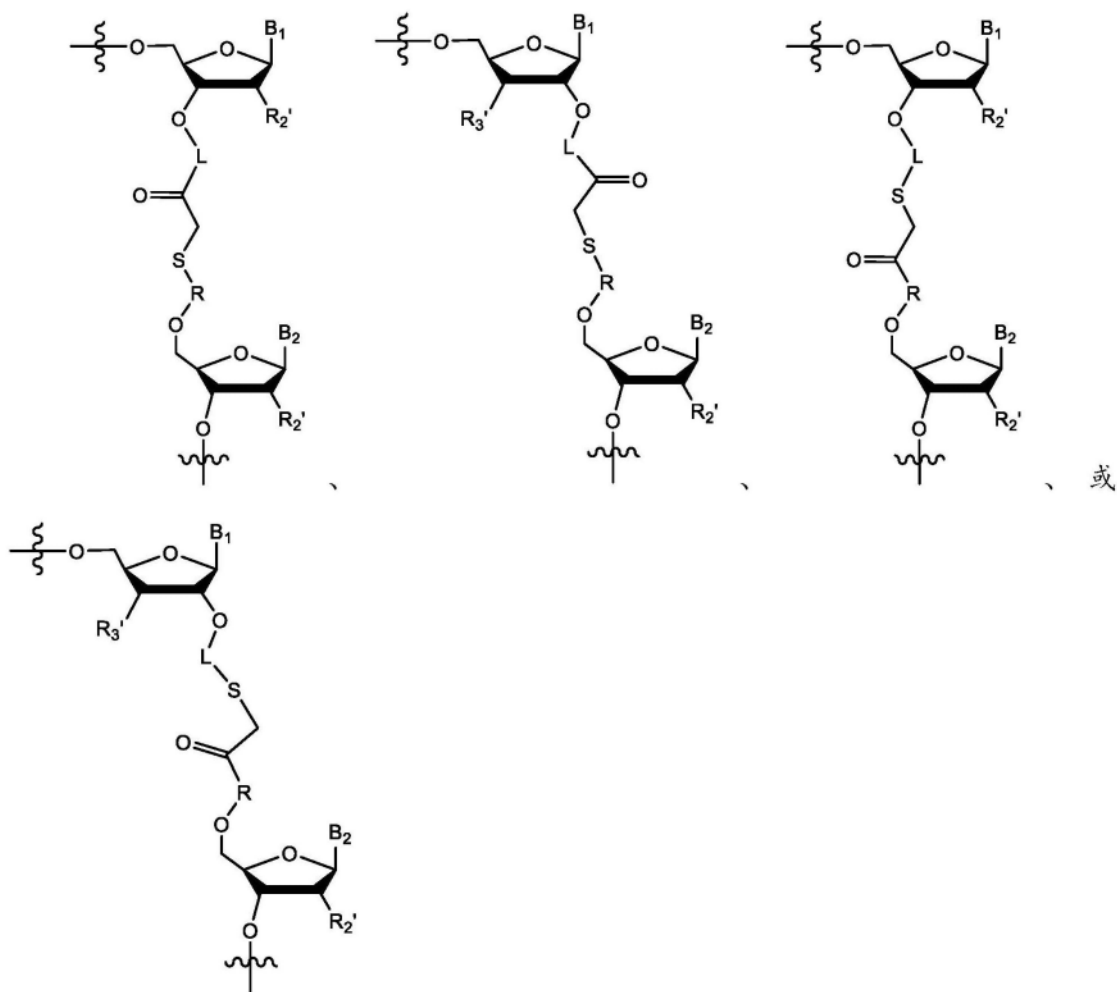
实施方式20. 根据实施方式13-19中任一项所述的指导分子, 其中  $L_1$  是  $-(CH_2)_w-$ , 并且  $w$  是1-20之间的整数, 包括端值。

实施方式21. 根据实施方式13-20中任一项所述的指导分子, 其中  $R_1$  是  $-(CH_2CH_2O)_v-$ , 并且  $v$  是1-10之间的整数, 包括端值。

实施方式22. 根据实施方式21所述的指导分子, 其中  $w$  是6并且  $v$  是4。

实施方式23. 根据前述实施方式中任一项所述的指导分子, 其中所述指导分子包含选自 SEQ ID NO.36-56 的序列。

实施方式24. 一种用于CRISPR系统的合成的单分子指导分子, 其中所述指导分子包含具有下式的化学键合：



其中：

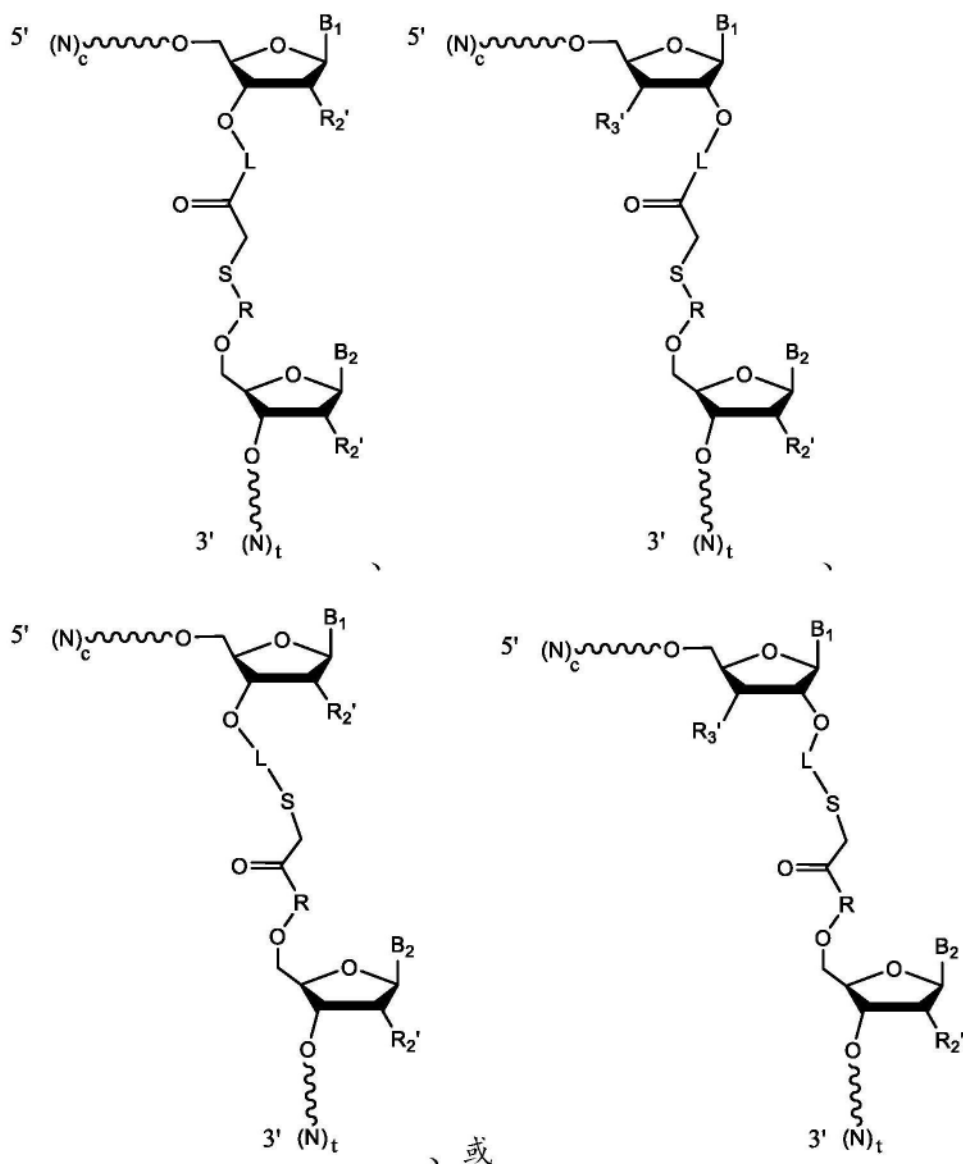
$R_2'$  和  $R_3'$  各自独立地是 H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R' 或 O-R'，其中每个 R' 独立地是保护基团或烷基，其中所述烷基可任选地被取代；

L 和 R 各自独立地是非核苷酸接头；并且

$B_1$  和  $B_2$  各自独立地是核碱基。

实施方式 25. 根据实施方式 24 所述的指导分子，其中  $R_2'$  选自由 H、氟和 O-R' 组成的组，其中 R' 是保护基团或任选地被取代的烷基。

实施方式 26. 根据实施方式 24 或 25 所述的指导分子，其中所述指导分子具有下式：



一个或多个2'-O-甲基修饰的核糖核苷酸残基。

实施方式30.根据实施方式26-29中任一项所述的指导分子,其中(N)<sub>c</sub>的5'末端的三个核苷酸中的每一个和/或(N)<sub>t</sub>的3'末端的三个核苷酸中的每一个包含2'-O-甲基修饰的核糖核苷酸残基,所述2'-O-甲基修饰的核糖核苷酸残基通过硫代磷酸酯键合与其相邻的一个或多个核苷酸连接。

实施方式31.根据实施方式24-30中任一项所述的指导分子,其中L和R各自独立地是非核苷酸接头,所述非核苷酸接头包含选自下组的部分,该组由以下组成:聚乙烯、聚丙烯、聚乙二醇和聚丙二醇。

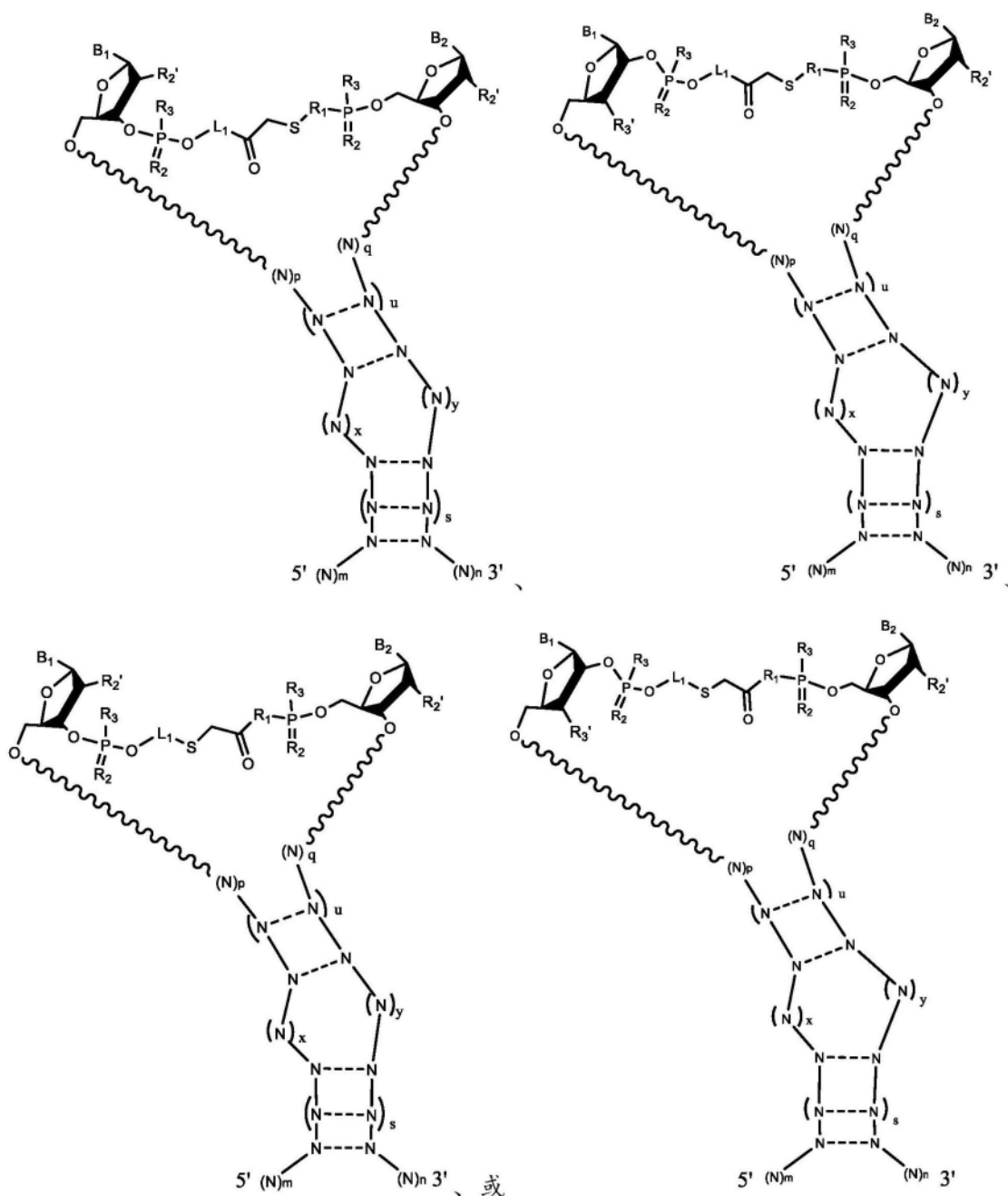
实施方式32.根据实施方式26-31中任一项所述的指导分子,其中所述指导分子用于II型CRISPR系统,并且(N)<sub>c</sub>包括5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

实施方式33.根据实施方式26-32中任一项所述的指导分子,其中(N)<sub>t</sub>包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式34.根据实施方式24-33中任一项所述的指导分子,其中所述指导分子能够与Cas9分子相互作用并介导Cas9/指导分子核糖核蛋白复合物的形成。

实施方式35.根据实施方式26-34中任一项所述的指导分子,其中(N)<sub>c</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式36.根据实施方式24-35中任一项所述的指导分子,其中所述指导分子具有下式:



其中:

L<sub>1</sub>和R<sub>1</sub>各自独立地是非核苷酸接头;

每个R<sub>2</sub>独立地是O或S;

每个R<sub>3</sub>独立地是O<sup>-</sup>或COO<sup>-</sup>;

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u是2到22之间的整数,包括端值;

s是1到10之间的整数,包括端值;

x是1到3之间的整数,包括端值;

y是>x且3到5之间的整数,包括端值;

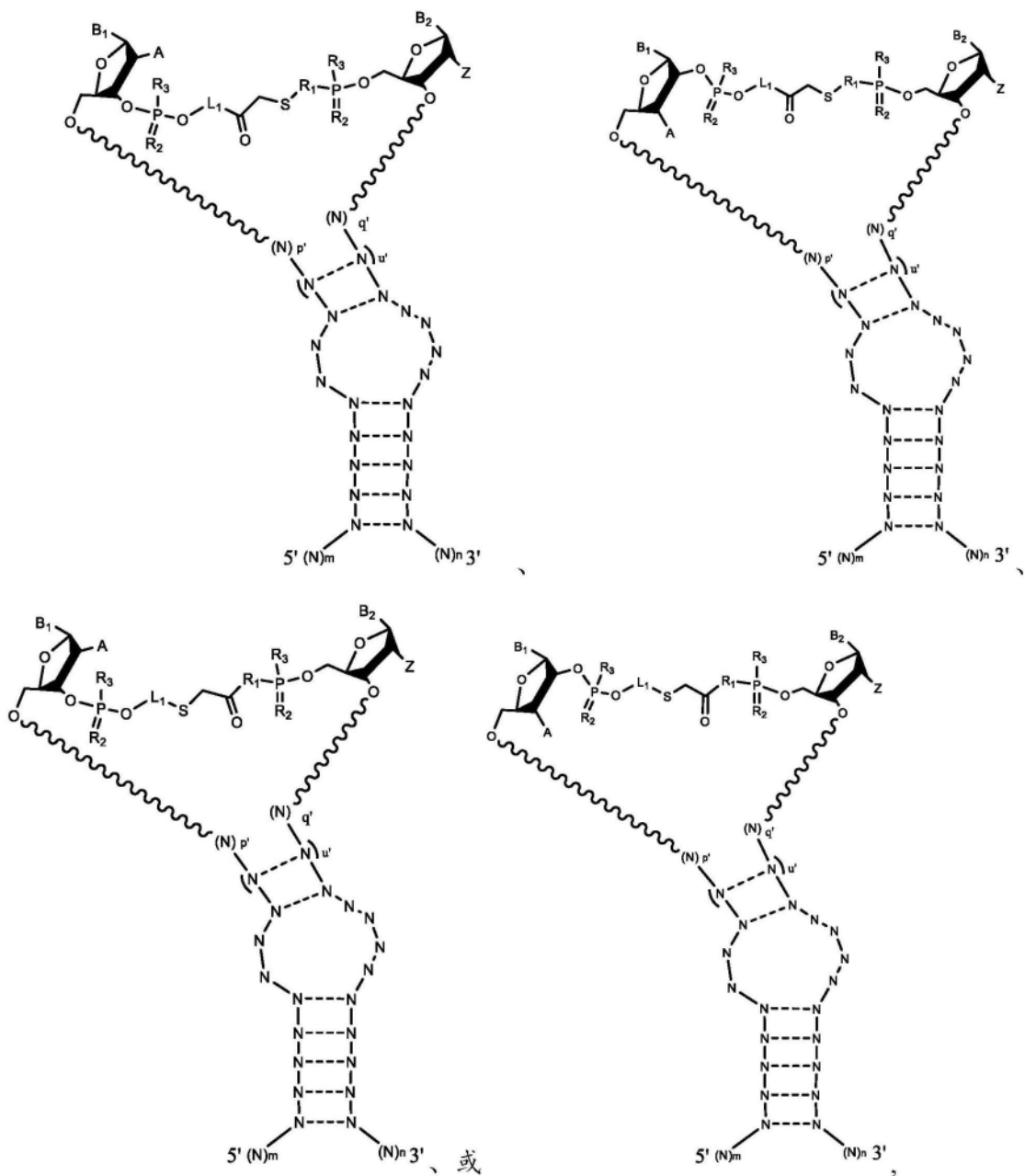
m是15或更大的整数;

$n$ 是30或更大的整数;

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。

实施方式37.根据实施方式24-36中任一项所述的指导分子,其中所述指导分子具有下式:

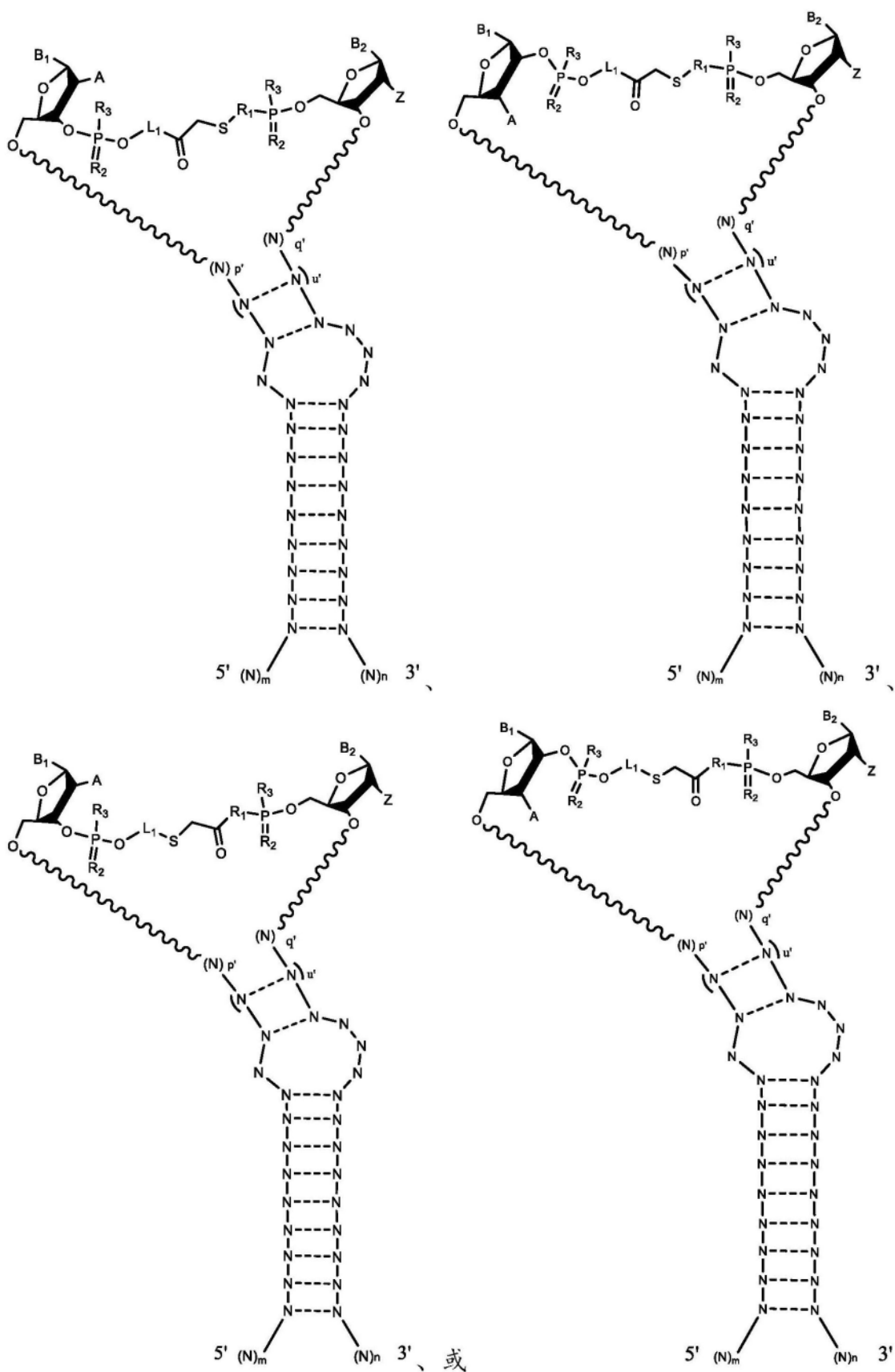


其中：

$u'$  是2到22之间的整数,包括端值;并且

$p'$  和  $q'$  各自独立地是0到4之间的整数,包括端值;并且 $p' + q'$  是0到4之间的整数,包括端值。

实施方式38.根据实施方式24-36中任一项所述的指导分子,其中所述指导分子具有下式:



其中: $u'$ 是2到22之间的整数,包括端值;并且

p' 和 q' 各自独立地是 0 到 4 之间的整数, 包括端值; 并且 p' + q' 是 0 到 4 之间的整数, 包括端值。

实施方式 39. 根据实施方式 36-38 中任一项所述的指导分子, 其中  $L_1$  是  $-(CH_2)_w-NH-C(O)-(CH_2)_w-NH-$  或  $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ ; 每个 w 独立地是 1-20; 并且每个 v 独立地是 1-10。

实施方式 40. 根据实施方式 36-39 中任一项所述的指导分子, 其中  $R_1$  是  $-(CH_2)_w-NH-C(O)-(CH_2)_w-$  或  $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ ; 每个 w 独立地是 1-20; 并且每个 v 独立地是 1-10。

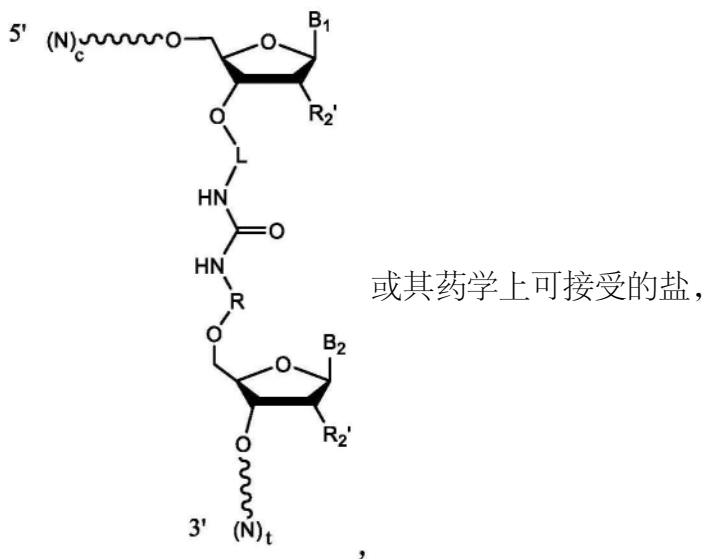
实施方式 41. 根据实施方式 40 所述的指导分子, 其中  $L_1$  是  $-(CH_2)_6-NH-C(O)-(CH_2)_1-NH-$  并且  $R_1$  是  $-(OCH_2CH_2)_4-NH-C(O)-(CH_2)_2-$ 。

实施方式 42. 根据实施方式 24-41 中任一项所述的指导分子, 其中所述指导分子包含选自 SEQ ID NO. 35 的序列。

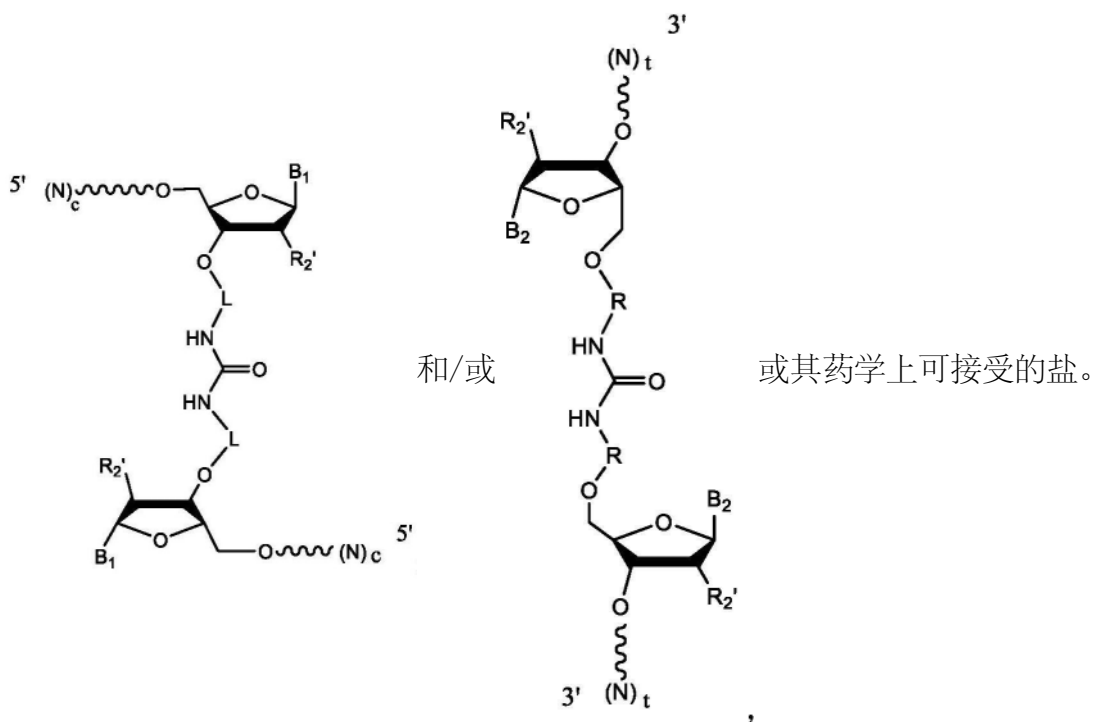
实施方式 43. 一种组合物, 其包含多个如前述实施方式中任一项所述的合成的指导分子, 其中相对于参照指导分子序列, 少于约 10% 的所述指导分子在 5' 末端包含截短。

实施方式 44. 根据实施方式 39 所述的组合物, 其中至少约 99% 的所述指导分子包含以下 5' 序列, 所述 5' 序列包含所述指导分子的与参照指导分子序列的相应 5' 序列 100% 相同的核苷酸 1-20。

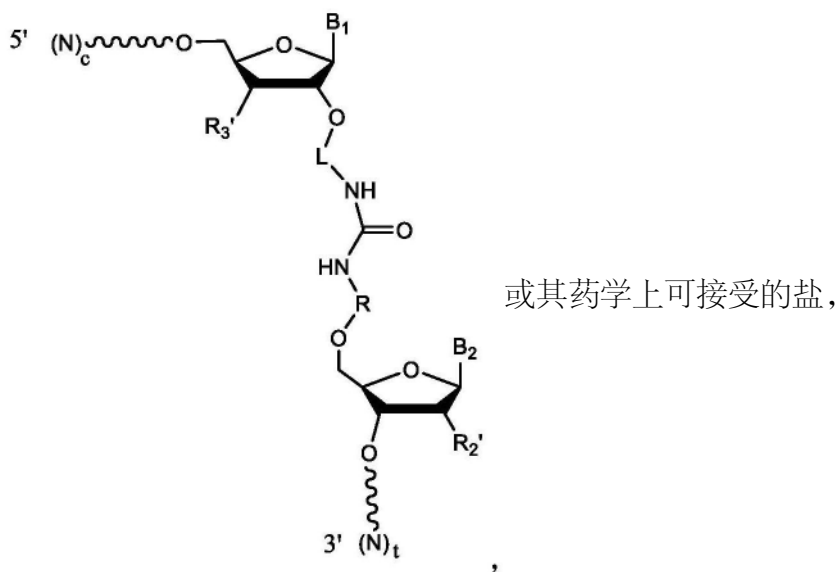
实施方式 45. 一种组合物, 所述组合物包含以下或基本上由以下组成: 根据实施方式 1-23 中任一项所述的具有下式的指导分子



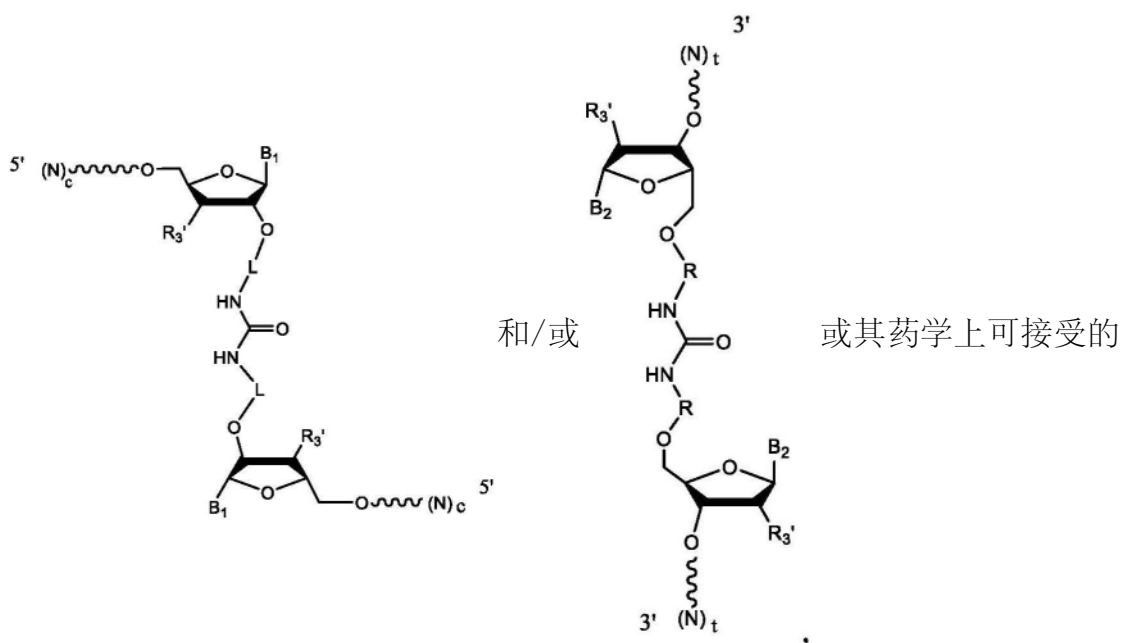
其中所述组合物基本上不含具有下式的分子:



实施方式46. 一种组合物, 所述组合物包含以下或基本上由以下组成: 根据实施方式1-23中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:



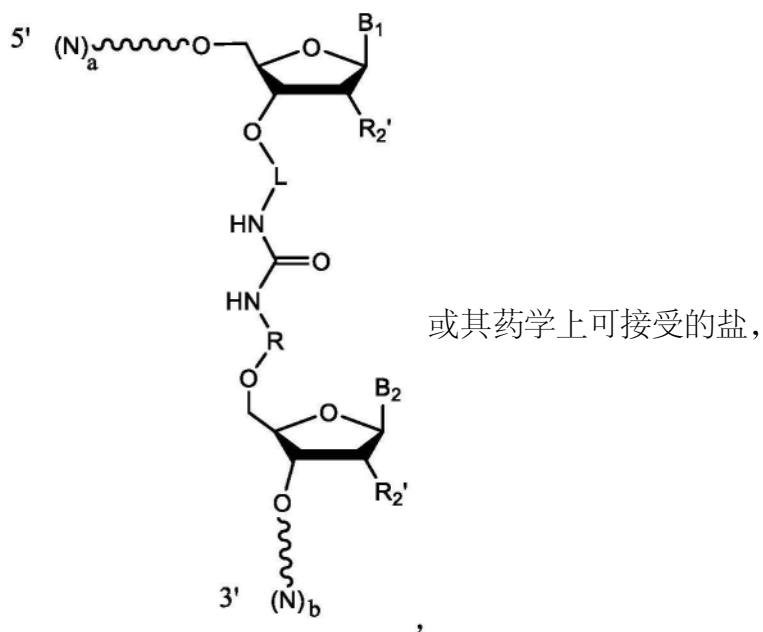
盐。

实施方式47.根据实施方式45或46所述的组合物,其中所述组合物未经过任何纯化步骤。

实施方式48.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式1-23中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:

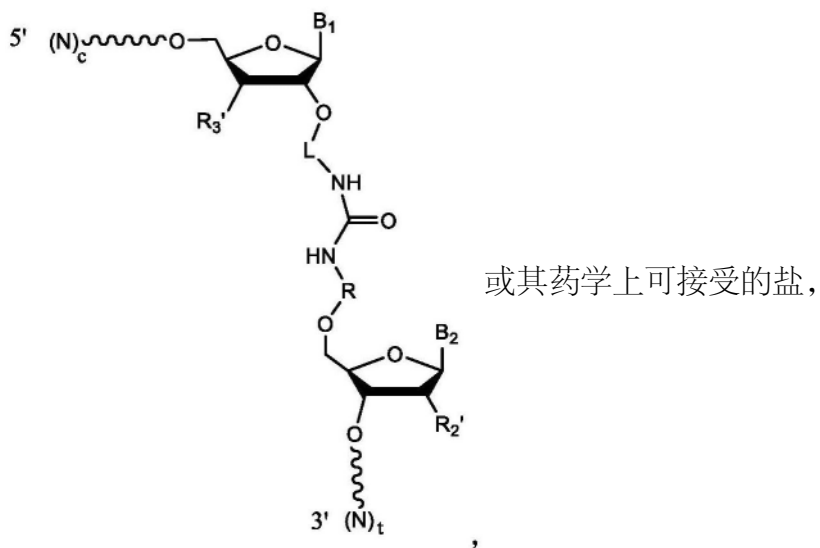


其中：

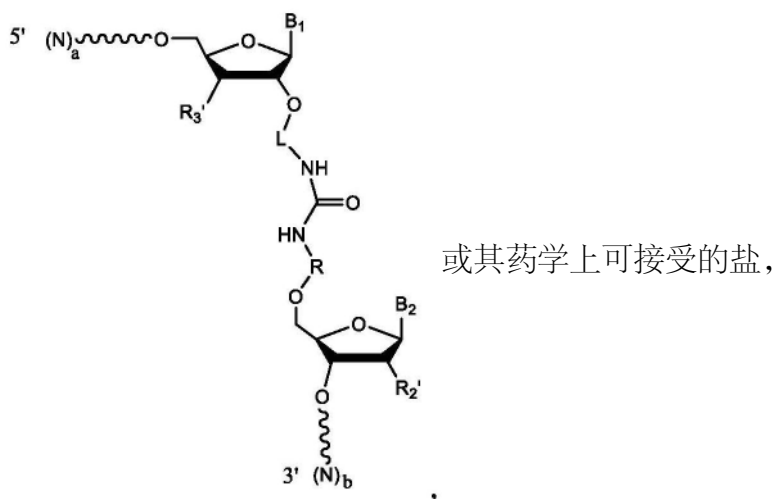
a不等于c；和/或

b不等于t。

实施方式49. 一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式1-23中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子：



其中:

a不等于c;和/或

b不等于t。

实施方式50. 根据实施方式48或49所述的组合物, 其中a小于c, 和/或b小于t。

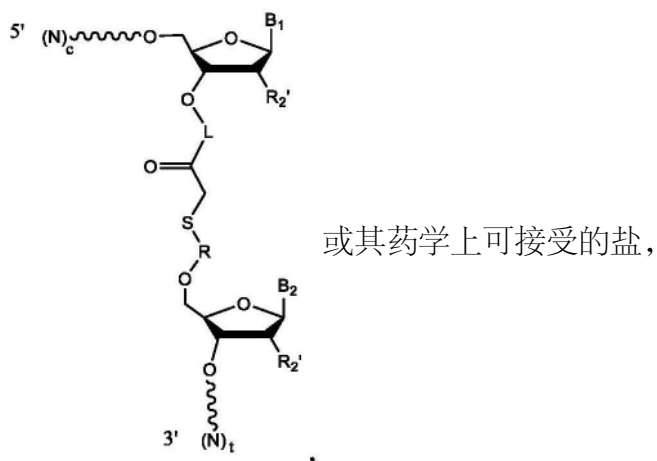
实施方式51.根据实施方式48-50中任一项所述的组合物,其中所述组合物未经过任何纯化步骤。

实施方式52.根据实施方式45-51中任一项所述的组合物,所述组合物包含所述指导分子与Cas9或RNA指导的核酸酶的复合物。

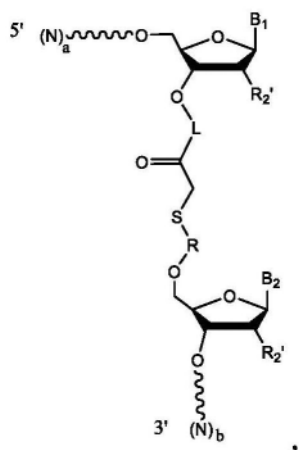
实施方式53.根据实施方式45-52中任一项所述的组合物,其中所述指导分子悬浮在溶液中或药学上可接受的载体中。

实施方式54.根据实施方式45-53中任一项所述的组合物,其中(N)<sub>2</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式55.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式24-42中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子：



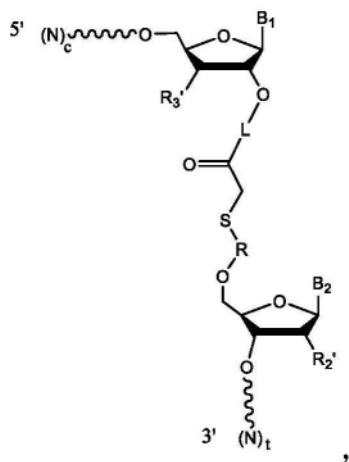
或其药学上可接受的盐，

其中：

a不等于c；和/或

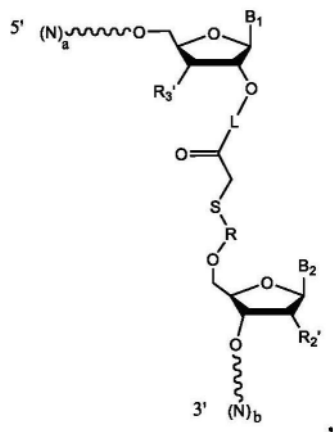
b不等于t。

实施方式56.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式24-42中任一项所述的具有下式的指导分子



或其药学上可接受的盐，

其中所述组合物基本上不含具有下式的分子：



或其药学上可接受的盐，

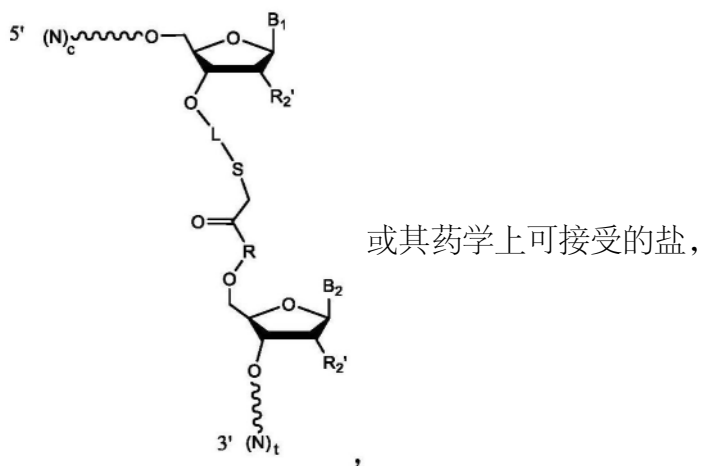
其中：

a不等于c；和/或

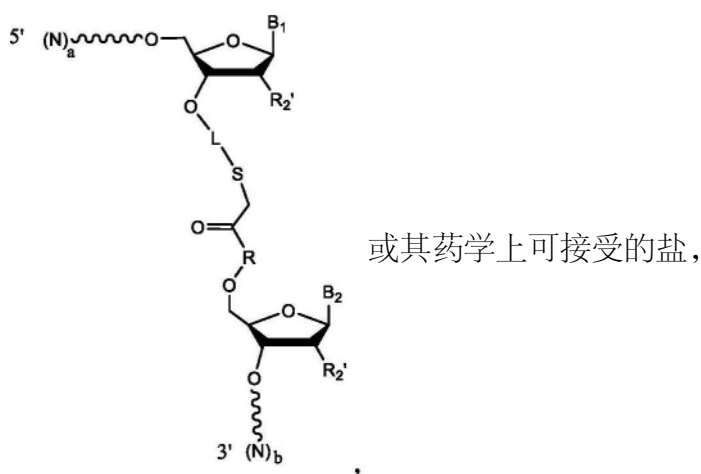
b不等于t。

实施方式57.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方

式24-42中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:

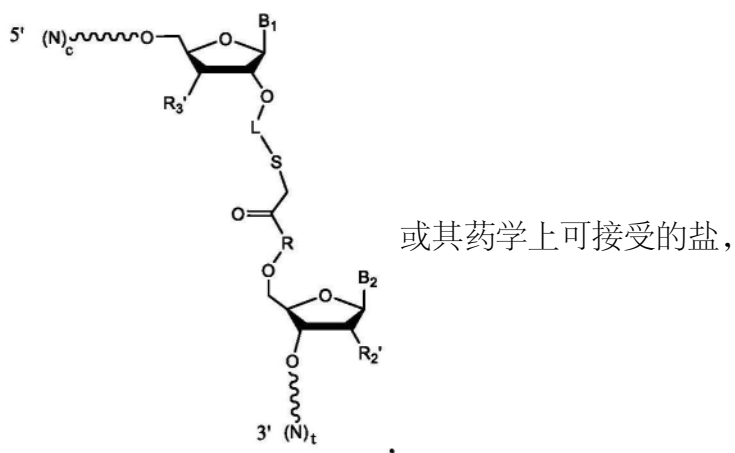


其中:

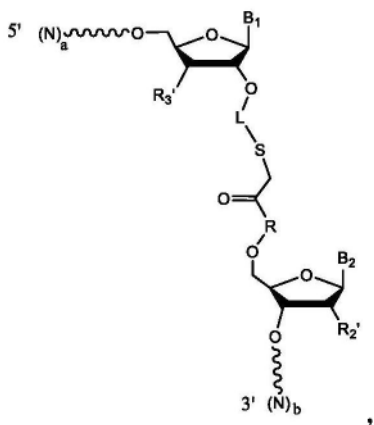
a不等于c;和/或

b不等于t。

实施方式58.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式24-42中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:



或其药学上可接受的盐,

其中:

a不等于c;和/或

b不等于t。

实施方式59.根据实施方式55-58中任一项所述的组合物,其中a小于c,和/或b小于t。

实施方式60.根据实施方式55-58中任一项所述的组合物,其中所述组合物未经过任何纯化步骤。

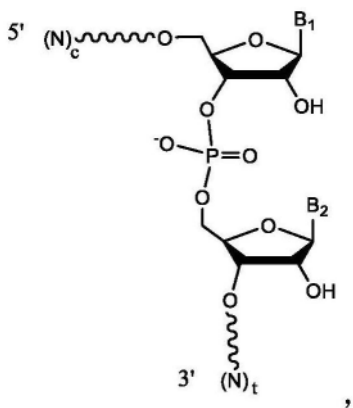
实施方式61.根据实施方式55-60中任一项所述的组合物,所述组合物包含所述指导分子与Cas9或RNA指导的核酸酶的复合物。

实施方式62.根据实施方式55-61中任一项所述的组合物,其中所述指导分子悬浮在溶液中或药学上可接受的载体中。

实施方式63.根据实施方式55-62中任一项所述的组合物,其中(N)<sub>c</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式64.一种组合物,其包含

(a) 用于CRISPR系统的合成的单分子指导分子,其中所述指导分子具有下式:



或其药学上可接受的盐,

其中:


(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数；

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基；并且

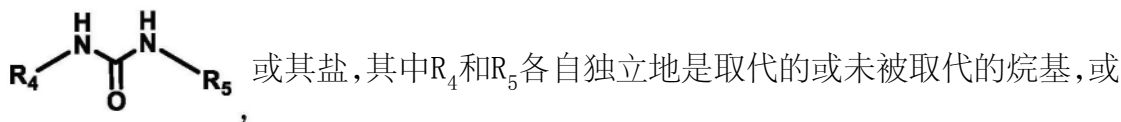
每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合；和

(b) 以下一项或多项：

(i) 碳二亚胺，或其盐；

(ii) 咪唑、氰基咪唑、吡啶和二甲基氨基吡啶，或其盐；和

(iii) 具有下式的化合物：

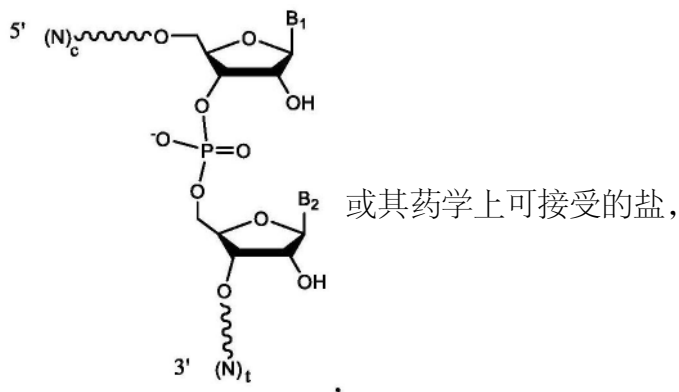


取代的或未被取代的碳环。

实施方式65. 根据实施方式64所述的组合物，其中碳二亚胺是EDC、DCC或DIC。

实施方式66. 一种组合物，其包含

用于CRISPR系统的合成的单分子指导分子，其中所述指导分子具有下式：



其中：


(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基，任选地是经修饰的核苷酸残基，各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接；

(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区；

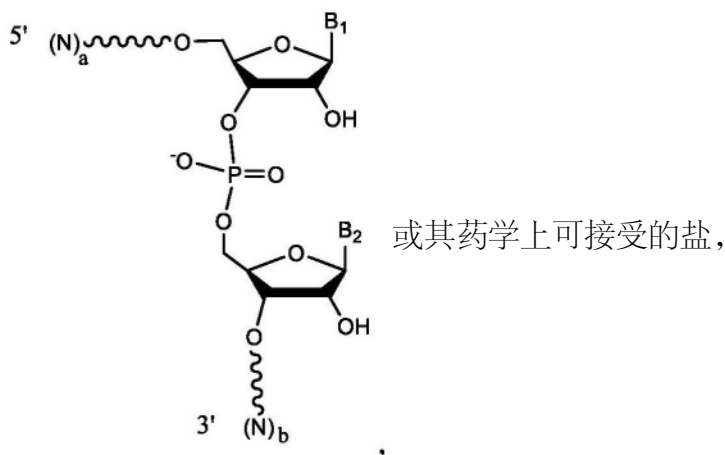
c是20或更大的整数；

t是20或更大的整数；

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基；并且

每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合；

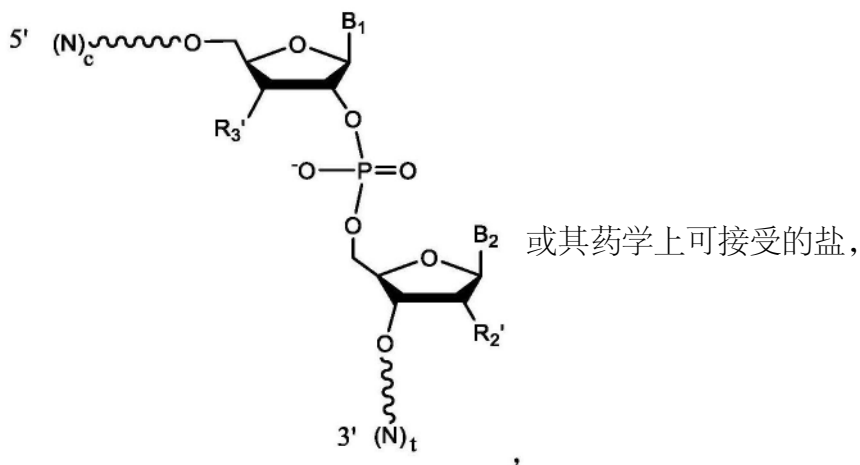
其中所述组合物基本上不含具有下式的分子：



其中:

$a+b$ 是 $c+t-k$ ,其中 $k$ 是1到10之间的整数,包括端值。

实施方式67.一种组合物,其包含用于CRISPR系统的合成的单分子指导分子或基本上由其组成,其中所述指导分子具有下式:



其中:

$(N)_c$ 和 $(N)_t$ 中的每个 $N$ 独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$ 包括与 $(N)_t$ 的5'区互补或部分互补并形成双链体的3'区;

所述式中所示的2'-5'磷酸二酯键合在所述双链体中的两个核苷酸之间;

$c$ 是20或更大的整数;

$t$ 是20或更大的整数;

$B_1$ 和 $B_2$ 各自独立地是核碱基;

$R_2$ 和 $R_3$ '各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代;并且

每个 $\sim$ 独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

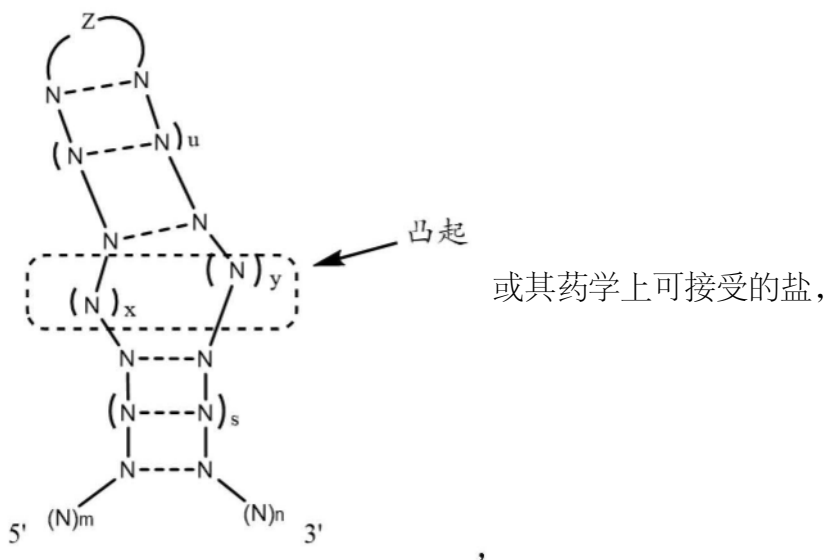
实施方式68.根据实施方式67所述的组合物,其中所述指导分子用于II型CRISPR系统,并且 $(N)_c$ 包括5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结

构域。

实施方式69.根据实施方式67或68所述的组合物,其中 $(N)_t$ 包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式70.根据实施方式67-69中任一项所述的组合物,其中所述指导分子能够与Cas9分子相互作用并介导Cas9/指导分子复合物的形成。

实施方式71.根据实施方式67-70中任一项所述的组合物,其中所述指导分子具有下式:



其中:

Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环;

u是0到22之间的整数,包括端值,任选地是2;

s是1到10之间的整数,包括端值,任选地是4;

x是1到3之间的整数,包括端值,任选地是2;

y是 $>x$ 且3到5之间的整数,包括端值,任选地是4;

m是15或更大的整数;

n是30或更大的整数;

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

所述式中所示的双链体区中两个核苷酸之间的至少一个磷酸二酯键合是2'-5'磷酸二酯键合;并且

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接。

实施方式72.根据实施方式67-71中任一项所述的组合物,其中所述2'-5'磷酸二酯键合在位于实施方式71的式中所示的凸起的5'的两个核苷酸之间。

实施方式73.根据实施方式67-71中任一项所述的组合物,其中所述2'-5'磷酸二酯键合在位于核苷酸环Z的5'和实施方式71的式中所示的凸起的3'的两个核苷酸之间。

实施方式74.根据实施方式67-71中任一项所述的组合物,其中所述2'-5'磷酸二

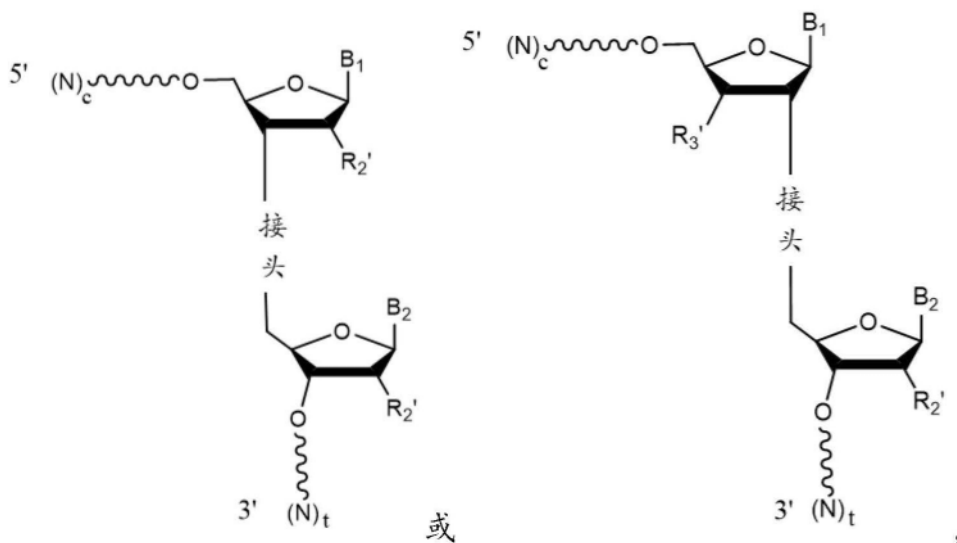
酯键合在位于核苷酸环Z的3'和实施方式71的式中所示的凸起的5'的两个核苷酸之间。

实施方式75. 根据实施方式67-71中任一项所述的组合物, 其中所述2'-5'磷酸二酯键合在位于实施方式71的式中所示的凸起的3'的两个核苷酸之间。

实施方式76. 根据实施方式45-75中任一项所述的组合物, 其中相对于参照指导分子序列, 少于约10%的所述指导分子在5'末端包含截短。

实施方式77. 根据实施方式76所述的组合物, 其中至少约99%的所述指导分子包含以下5'序列, 所述5'序列包含所述指导分子的与参照指导分子序列的相应5'序列100%相同的核苷酸1-20。

实施方式78. 一种用于CRISPR系统的指导分子的组合物, 其中所述组合物基本上由以下组成: 具有下式的指导分子



或其药学上可接受的盐, 其中:

$(N)_c$  和  $(N)_t$  中的每个N独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包括与  $(N)_t$  的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数; 并且

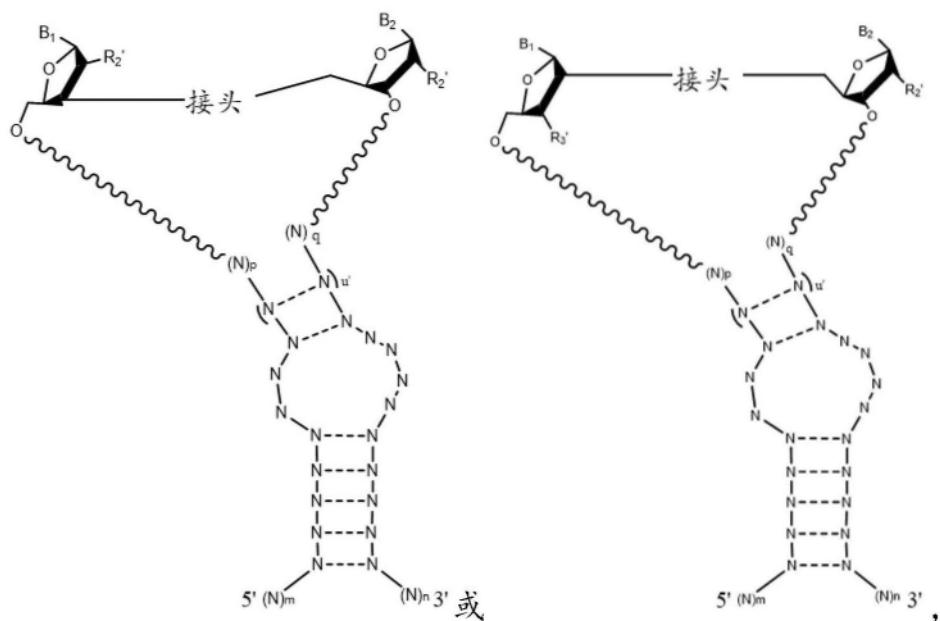
每个  $\sim$  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合;

接头是非核苷酸化学键合;

$B_1$  和  $B_2$  各自独立地是核碱基; 并且

$R_2'$  和  $R_3'$  各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R' 或 O-R', 其中每个R'独立地是保护基团或烷基, 其中所述烷基可任选地被取代。

实施方式79. 根据实施方式78所述的指导分子的组合物, 所述指导分子基本上由以下组成: 具有下式的指导分子



或其盐, 其中:

p和q各自独立地是0到6之间的整数, 包括端值; 并且p+q是0到6之间的整数, 包括端值;

u'是2到22之间的整数, 包括端值;

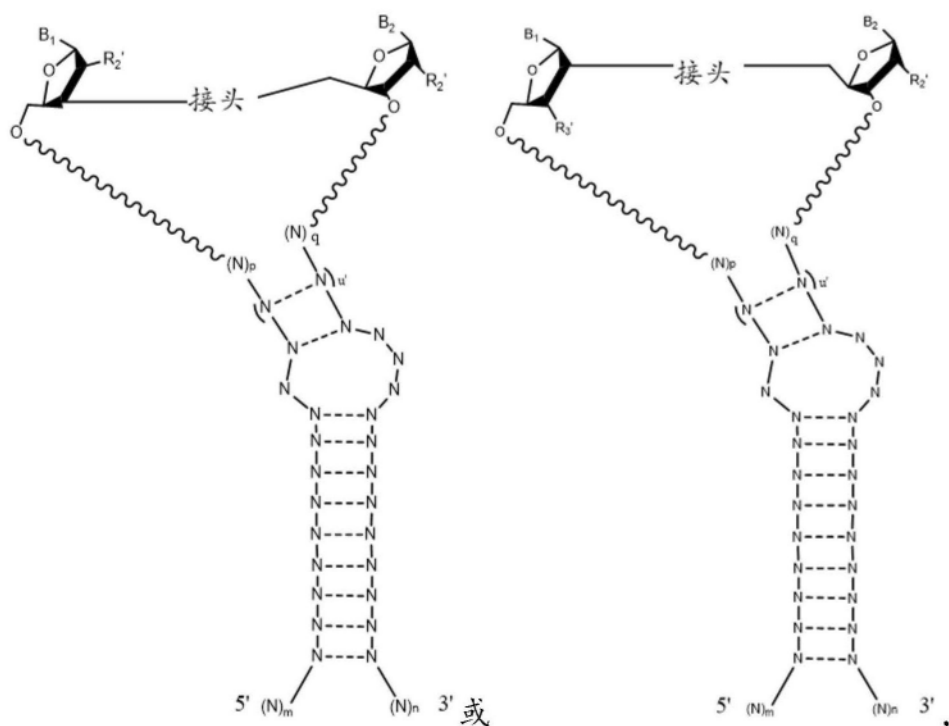
m是15或更大的整数;

n是30或更大的整数;

每个N---N独立地代表两个互补核苷酸, 任选地是氢键碱基配对的两个互补核苷酸; 并且

每个N独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接。

实施方式80. 根据实施方式78所述的指导分子的组合物, 所述指导分子基本上由以下组成: 具有下式的指导分子



或其盐,其中:

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u'是2到22之间的整数,包括端值;

m是15或更大的整数;

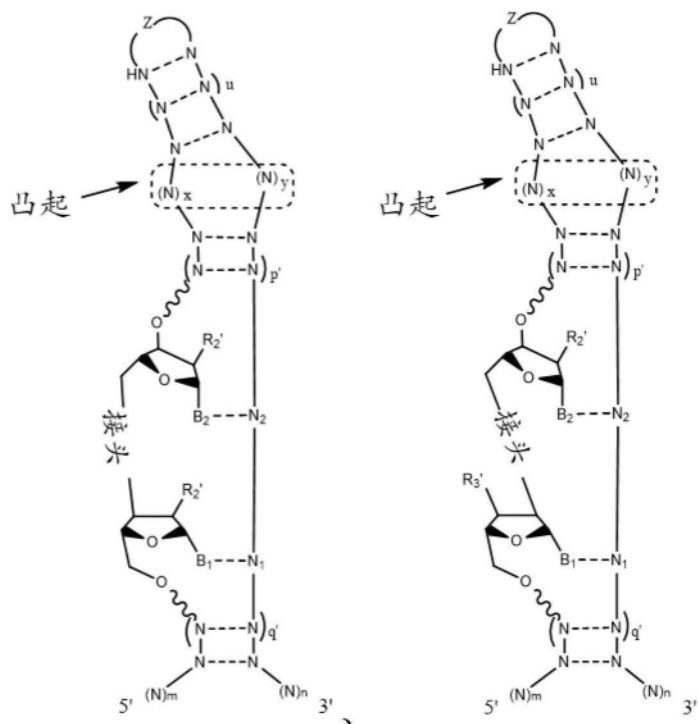
n是30或更大的整数;

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

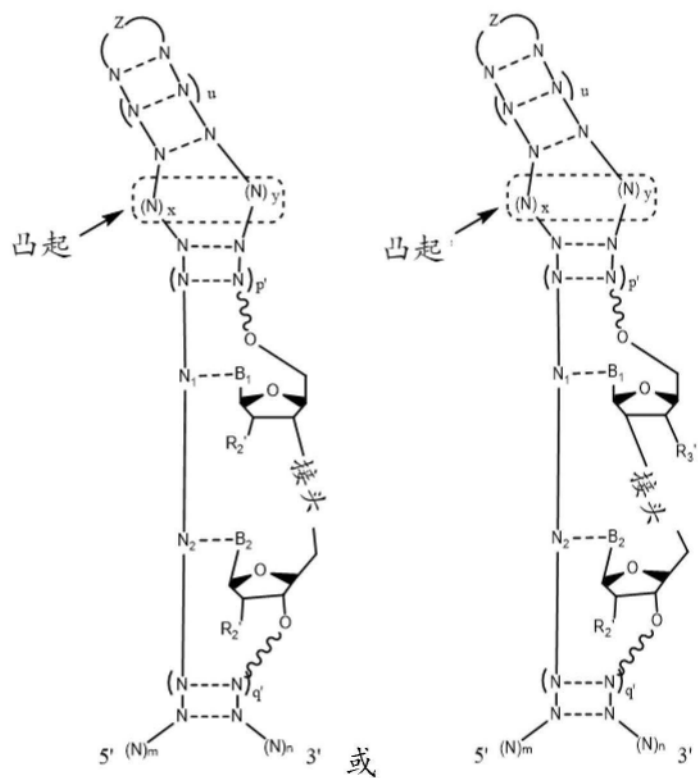
每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接。

实施方式81. 根据实施方式78所述的指导分子的组合物,所述指导分子基本上由以下组成:具有下式的指导分子





或其盐,其中:



Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环;p和q各自独立地是0到2之间的整数,包括端值,任选地是0;p'是0到4之间的整数,包括端值,任选地是0;q'是0到4之间的整数,包括端值,任选地是2;x是1到3之间的整数,包括端值,任选地是2;y是>x且3到5之间的整数,包括端值,任选地是4;u是2到22之间的整数,包括端值,任选地是2;

s是1到10之间的整数,包括端值,任选地是4;

m是15或更大的整数;

n是30或更大的整数;

$B_1$ 和 $B_2$ 各自独立地是核碱基;

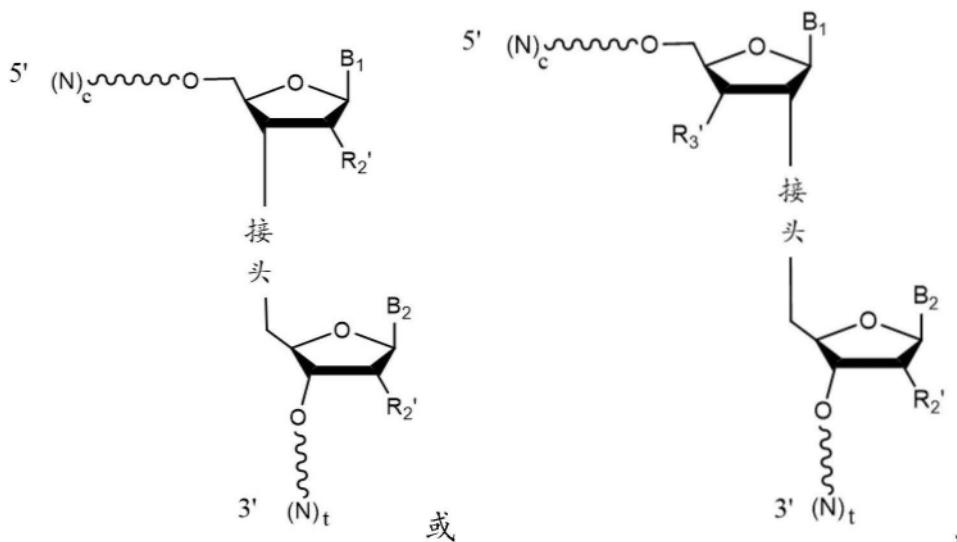
每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;

$N_1$ 和 $N_2$ 各自独立地是核苷酸残基;

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

每个 $\sim$ 独立地代表磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合。

实施方式82.一种用于CRISPR系统的指导分子的组合物,其中所述指导分子具有下式:



或其药学上可接受的盐,其中:

$(N)_c$ 和 $(N)_t$ 中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$ 包括与 $(N)_t$ 的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数;并且

每个 $\sim$ 独立地代表磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合;

接头是非核苷酸化学键合;

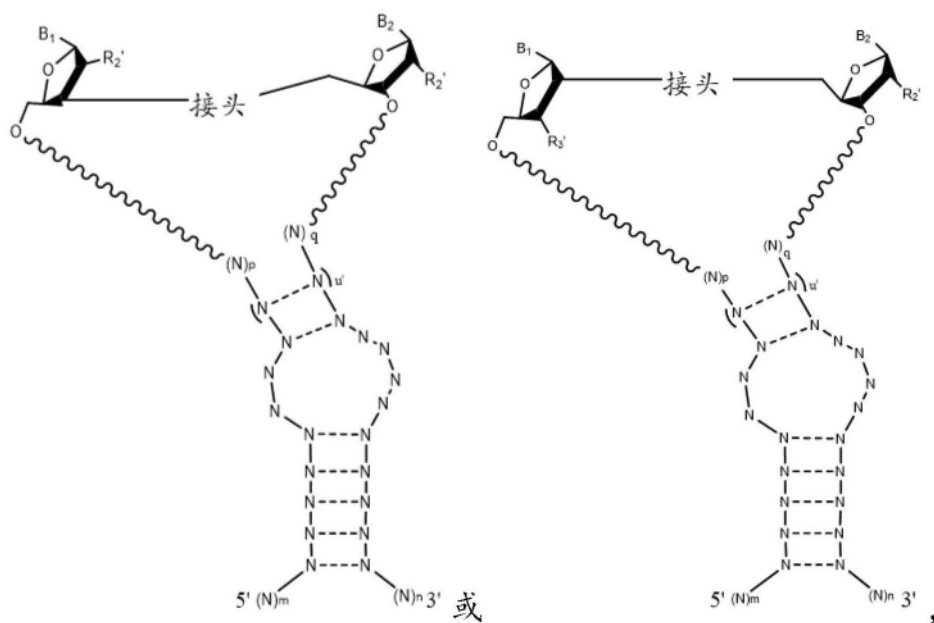
$B_1$ 和 $B_2$ 各自独立地是核碱基;并且

$R_2'$ 和 $R_3'$ 各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代,

其中相对于参照指导分子序列,少于约10%的所述指导分子在5'末端包含截短,并且

其中至少约99%的所述指导分子包含以下5'序列,所述5'序列包含所述指导分子的与参照指导分子序列的相应5'序列100%相同的核苷酸1-20。

实施方式83.根据实施方式82所述的指导分子的组合物,其中所述指导分子具有下式:



或其盐,其中:

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u'是2到22之间的整数,包括端值;

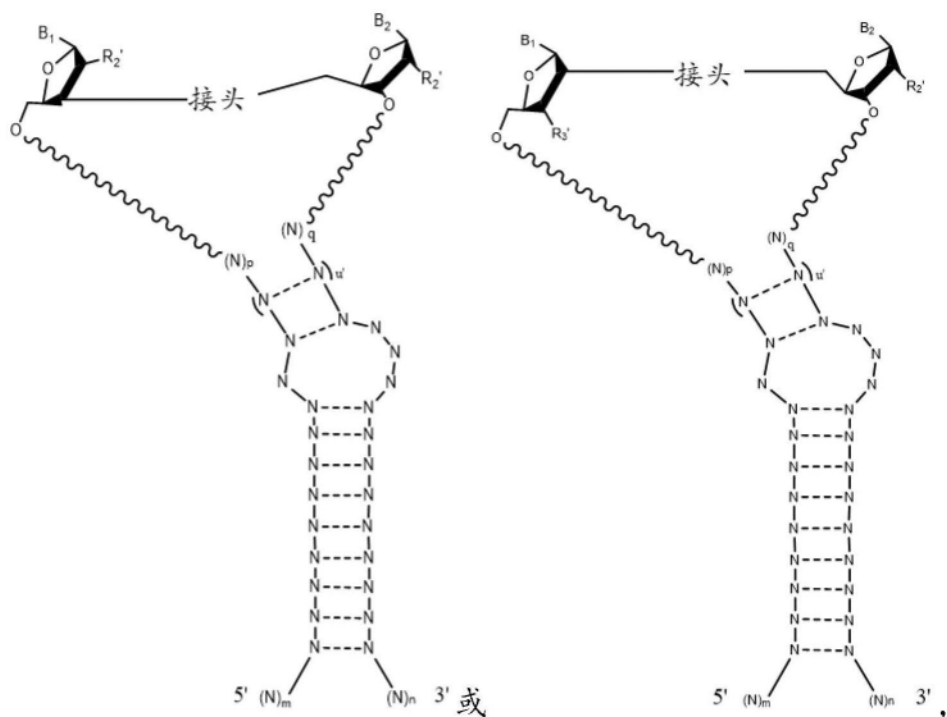
m是15或更大的整数;

n是30或更大的整数;

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接。

实施方式84.根据实施方式82所述的指导分子的组合物,其中所述指导分子具有下式:



或其盐,其中:

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u'是2到22之间的整数,包括端值;

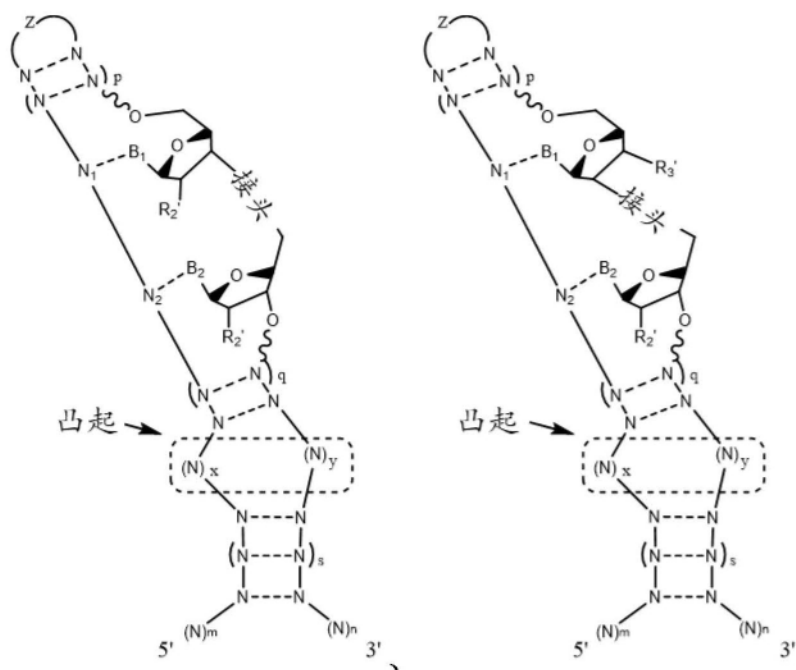
m是15或更大的整数;

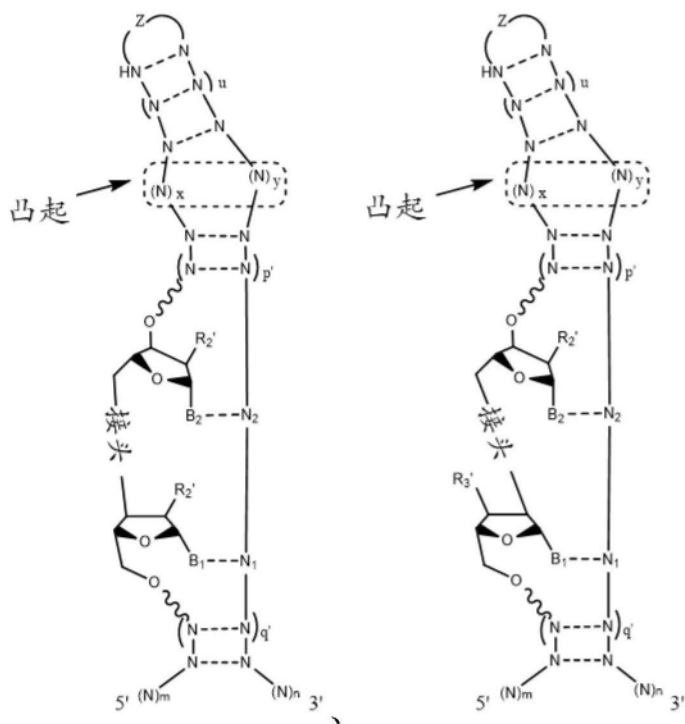
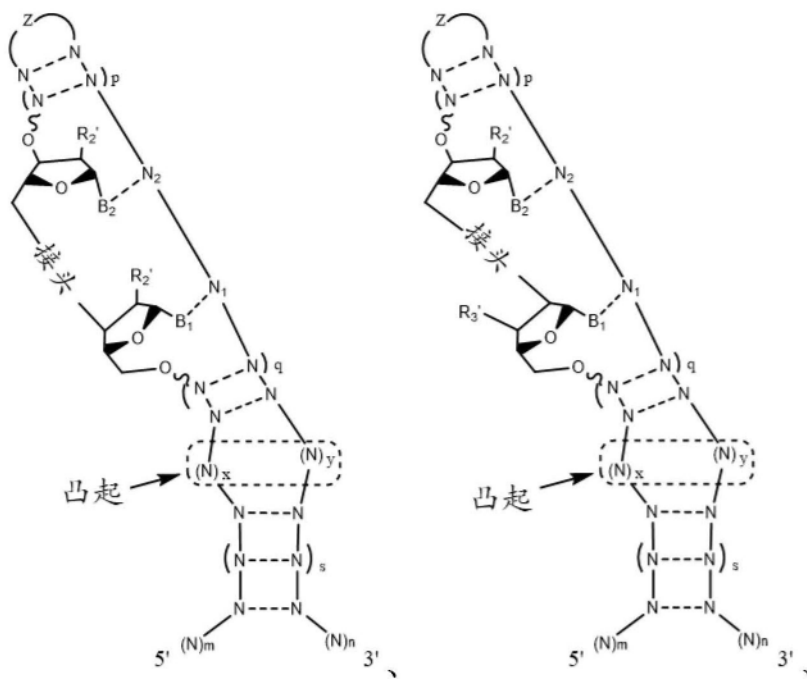
n是30或更大的整数;

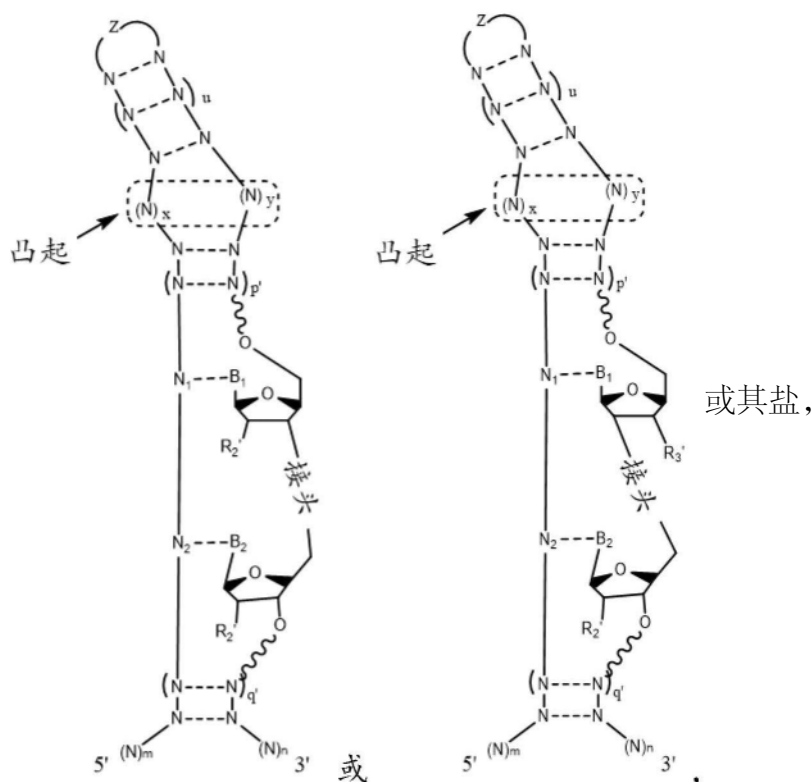
每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接。

实施方式85.根据实施方式82所述的指导分子的组合物,其中所述指导分子具有下式:







其中:

Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环;

p和q各自独立地是0到2之间的整数,包括端值,任选地是0;

p' 是0到4之间的整数,包括端值,任选地是0;

q' 是0到4之间的整数,包括端值,任选地是2;

x是1到3之间的整数,包括端值,任选地是2;

y是>x且3到5之间的整数,包括端值,任选地是4;

u是2到22之间的整数,包括端值,任选地是2;

s是1到10之间的整数,包括端值,任选地是4;

m是15或更大的整数;

n是30或更大的整数;

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基;

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

N<sub>1</sub>和N<sub>2</sub>各自独立地是核苷酸残基;

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸;并且

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式86.一种合成用于CRISPR系统的单分子指导分子的方法,所述方法包括以下步骤:

使第一寡核苷酸和第二寡核苷酸退火以在所述第一寡核苷酸的3'区和所述第二寡核苷酸的5'区之间形成双链体,其中所述第一寡核苷酸包含第一反应性基团,所述第一反应性基团是2'反应性基团和3'反应性基团中的至少一个,并且其中所述第二寡核苷酸包含第二反应性基团,所述第二反应性基团是5'反应性基团;和

通过所述第一和第二反应性基团缀合所述退火的第一和第二寡核苷酸以形成单分子指导分子,所述单分子指导分子包含连接所述第一和第二寡核苷酸的共价键。

实施方式87.根据实施方式86所述的方法,其中所述指导分子用于II型CRISPR系统,并且所述第一寡核苷酸的5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

实施方式88.根据实施方式86或87所述的方法,其中所述第二寡核苷酸的3'区包含一个或多个茎-环结构。

实施方式89.根据实施方式86-88中任一项所述的方法,其中所述指导分子能够与Cas9分子相互作用并介导Cas9/指导分子复合物的形成。

实施方式90.根据实施方式86-88中任一项所述的方法,其中所述第一和第二反应性基团均包含胺部分,并且缀合步骤包括使所述第一和第二反应性基团的胺部分与含碳酸酯的双官能交联剂交联以形成脲键合。

实施方式91.根据实施方式90所述的方法,其中所述含碳酸酯的双官能交联剂是二琥珀酰亚胺基碳酸酯、二咪唑碳酸酯或双-(对硝基苯基)碳酸酯。

实施方式92.根据实施方式86-91中任一项所述的方法,其中所述第一和第二寡核苷酸中的每一个的浓度在10 $\mu$ M至1mM的范围内。

实施方式93.根据实施方式90-92中任一项所述的方法,其中所述含碳酸酯的双官能交联剂的浓度在1mM至100mM的范围内。

实施方式94.根据实施方式90-93中任一项所述的方法,其中所述含碳酸酯的双官能交联剂的浓度比所述第一和第二寡核苷酸中每一个的浓度高100-1,000倍。

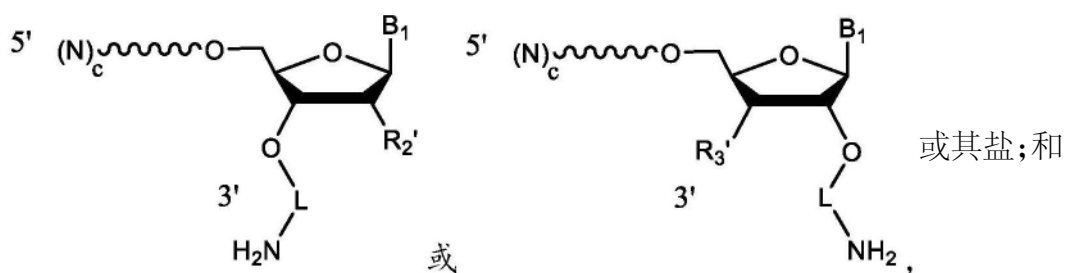
实施方式95.根据实施方式86-94中任一项所述的方法,其中所述缀合步骤在7-9的pH范围内进行。

实施方式96.根据实施方式86-95中任一项所述的方法,其中所述缀合步骤在具有DMSO、DMF、NMP、DMA、吗啉、吡啶或MeCN作为共溶剂的水中进行。

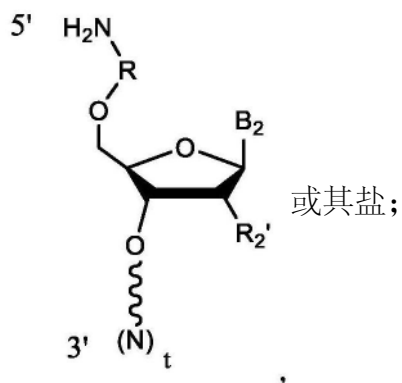
实施方式97.根据实施方式86-96中任一项所述的方法,其中所述缀合步骤在二价金属阳离子的存在下进行。

实施方式98.根据实施方式86-97中任一项所述的方法,其中所述缀合步骤在0°C至40°C的温度范围内进行。

实施方式99.根据实施方式86-98中任一项所述的方法,其中所述第一寡核苷酸具有下式:



所述第二寡核苷酸具有下式:



其中:

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;

(N)<sub>c</sub>包括与(N)<sub>+</sub>的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数;

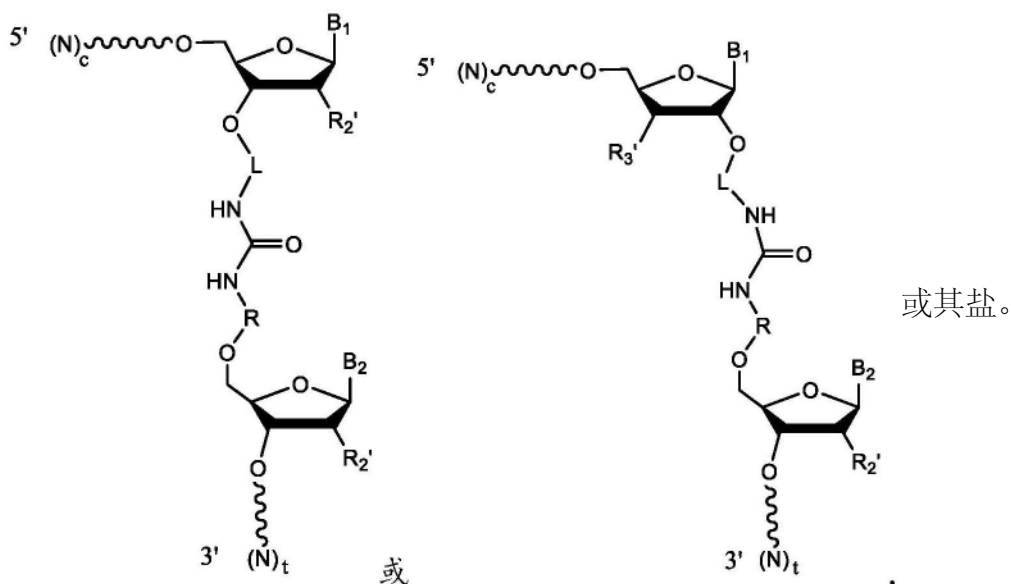
$R_2'$ 和 $R_3'$ 各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$ 或O- $R'$ ,其中每个 $R'$ 独立地是保护基团或烷基,其中所述烷基可任选地被取代;

L和R各自独立地是非核苷酸接头;并且

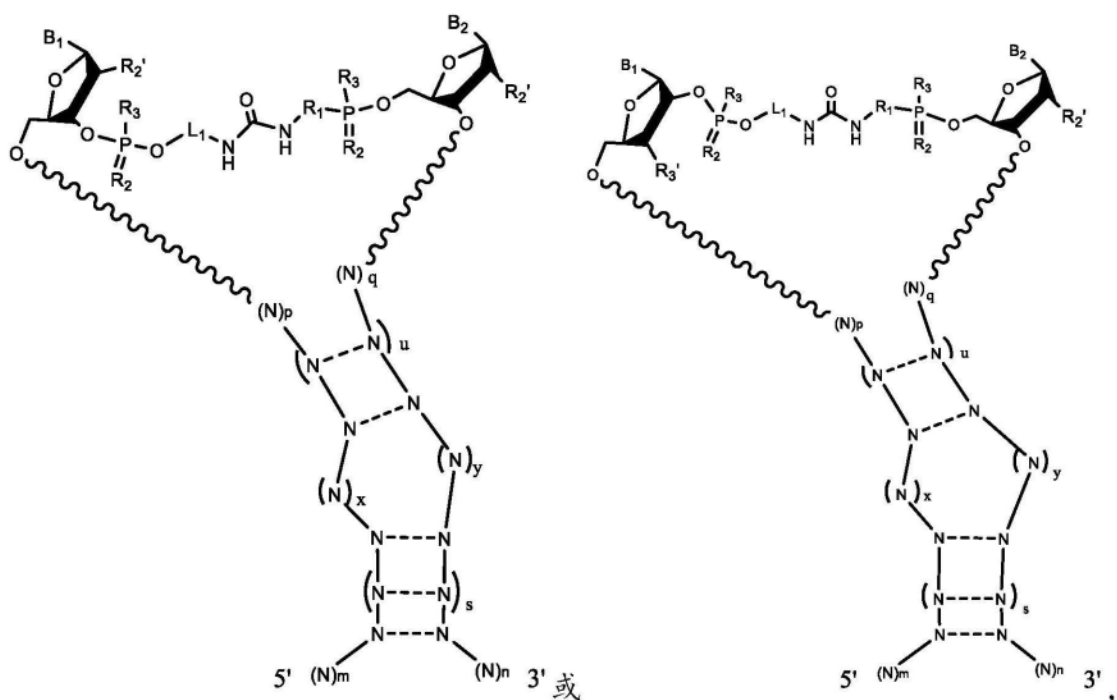
$B_1$ 和 $B_2$ 各自独立地是核碱基;并且

每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式100.根据实施方式86-99中任一项所述的方法,其中所述单分子指导分子具有下式:



实施方式101. 根据实施方式82-96中任一项所述的方法, 其中所述单分子指导分子具有下式:



其中:

$L_1$ 和 $R_1$ 各自独立地是非核苷酸接头;

每个 $R_2$ 独立地是O或S;

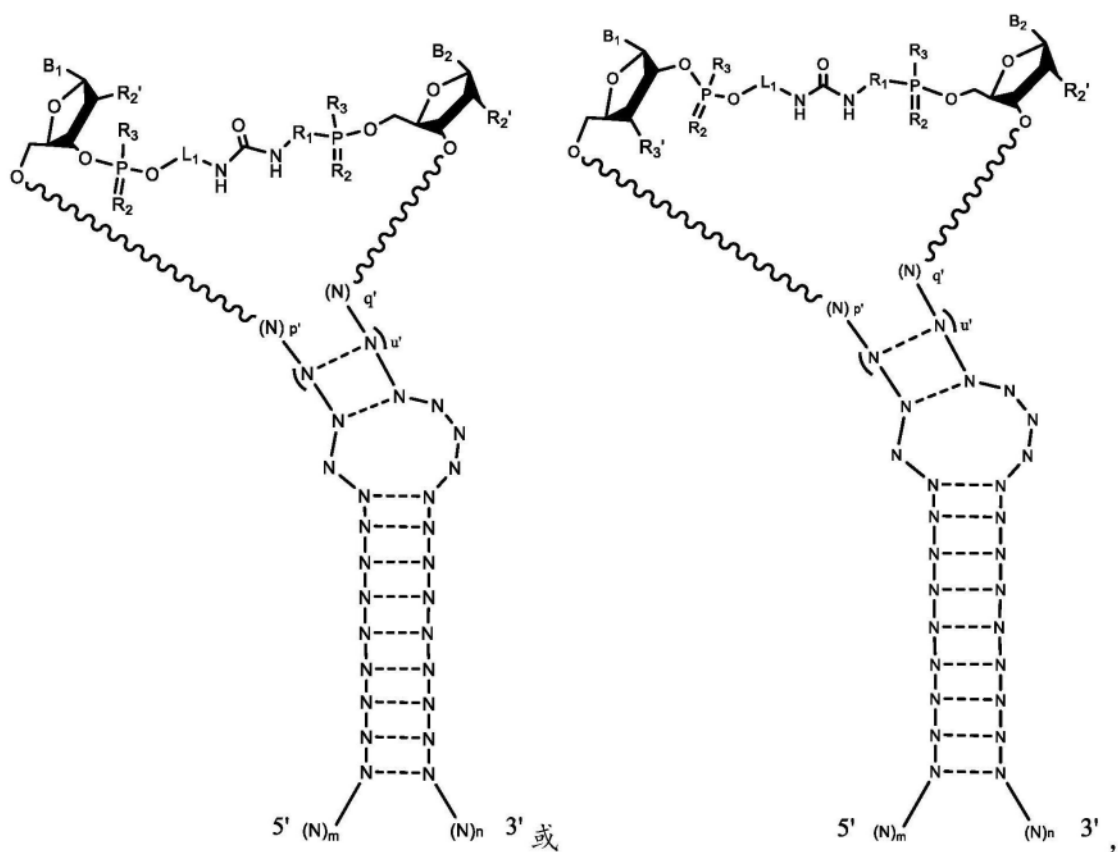
每个 $R_3$ 独立地是O<sup>-</sup>或COO<sup>-</sup>;

$p$ 和 $q$ 各自独立地是0到6之间的整数, 包括端值; 并且 $p+q$ 是0到6之间的整数, 包括端值;

$u$ 是2到22之间的整数, 包括端值;

$s$ 是1到10之间的整数, 包括端值;





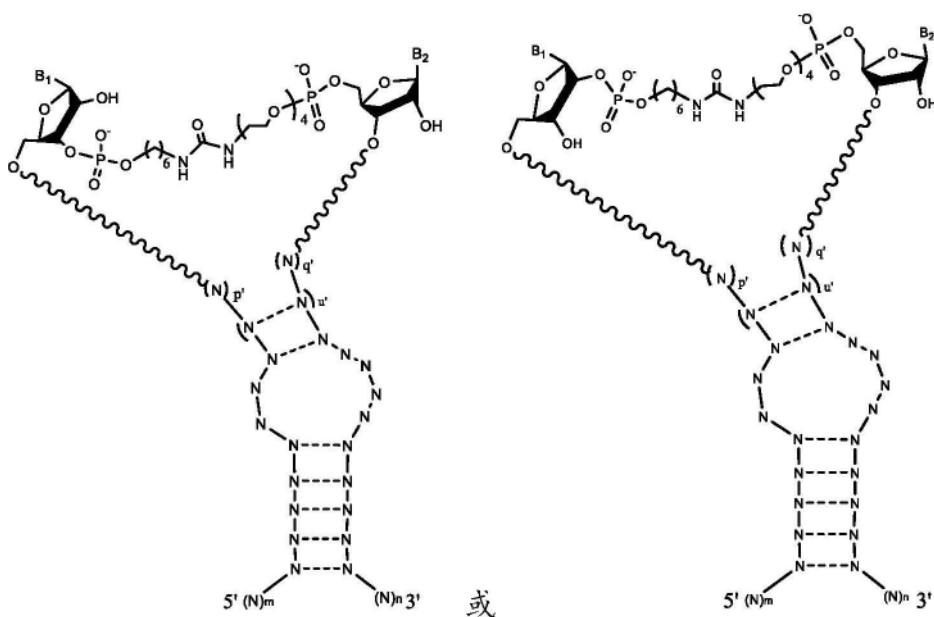
或其盐,其中 $p'$ 和 $q'$ 各自独立地是0到4之间的整数,包括端值;并且 $p'+q'$ 是0到4之间的整数,包括端值;并且 $u'$ 是2到22之间的整数,包括端值。

实施方式104.根据实施方式102或103所述的方法,其中 $p'=q'$ ,任选地其中 $p'=q'=0$ 、 $p'=q'=1$ 、或 $p'=q'=2$ 。

实施方式105.根据实施方式101-104中任一项所述的方法,其中 $L_1$ 是 $-(CH_2)_w-$ ,并且 $w$ 是1-20。

实施方式106. 根据实施方式101-104中任一项所述的方法, 其中R<sub>1</sub>是-(CH<sub>2</sub>CH<sub>2</sub>O)<sub>v</sub>-, 并且v是1-10。

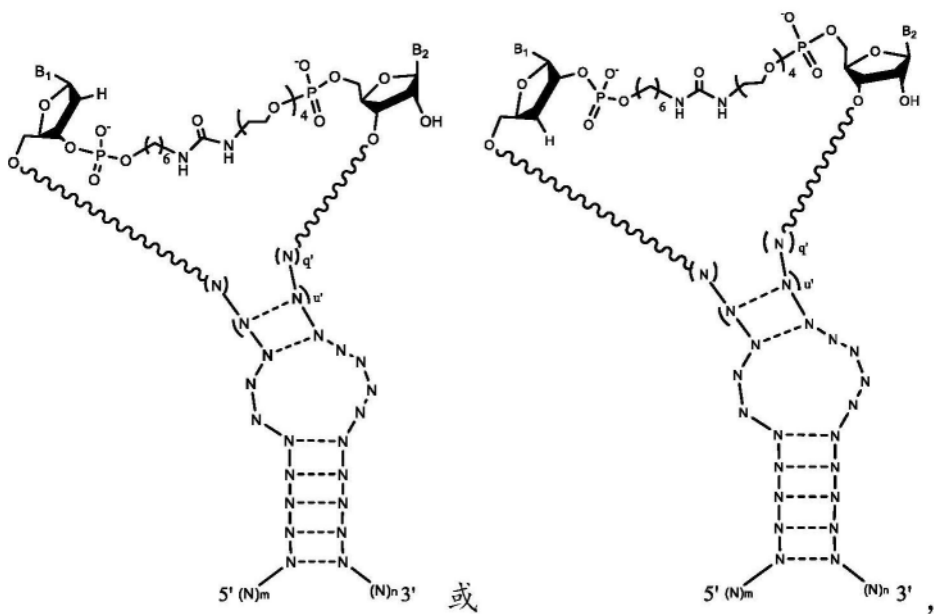
实施方式107.根据实施方式99-102中任一项所述的方法,其中所述单分子指导分子具有下式:



其盐, 或其

盐, 其中 $p'$ 和 $q'$ 各自独立地为0至4之间的整数, 并且 $p'+q'$ 为0至4之间的整数, 包括端值; 并且 $u'$ 是2到14之间的整数, 包括端值。

实施方式108. 根据实施方式99-102中任一项所述的方法, 其中所述单分子指导分子具有下式:



或其盐, 其中

$p'$ 和 $q'$ 各自独立地是0到4之间的整数, 包括端值; 并且 $p'+q'$ 是0到4之间的整数, 包括端值; 并且 $u'$ 是2到14之间的整数, 包括端值。

实施方式109. 根据实施方式86-89中任一项所述的方法, 其中 (a) 所述第一反应性基团包含溴乙酰基部分并且所述第二反应性基团包含巯基部分, 或 (b) 所述第一反应性基团包含巯基部分并且所述第二反应性基团包含溴乙酰基部分, 并且所述缀合步骤包括使所述溴乙酰基部分与所述巯基部分反应形成溴乙酰基-硫醇键合。

实施方式110. 根据实施方式109所述的方法, 其中所述第一和第二寡核苷酸中的每一个的浓度在 $10\mu\text{M}$ 至 $1\text{mM}$ 的范围内。

实施方式111.根据实施方式109或110所述的方法,其中所述缀合步骤在7-9的pH范围内进行。

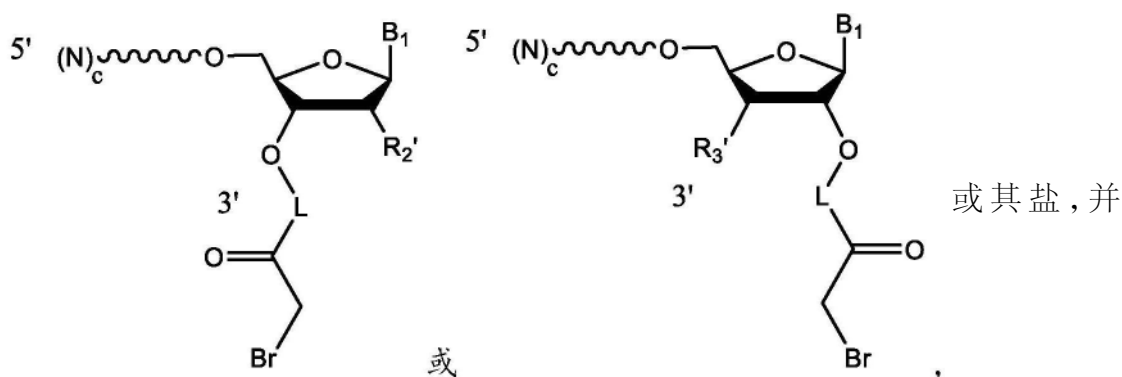
实施方式112.根据实施方式109-111中任一项所述的方法,其中所述缀合步骤在氩气下进行。

实施方式113.根据实施方式109-112中任一项所述的方法,其中所述缀合步骤在螯合剂、任选地在乙二胺四乙酸(EDTA)或其盐的存在下进行。

实施方式114.根据实施方式109-113中任一项所述的方法,其中所述缀合步骤在0℃至40℃的温度范围内进行。

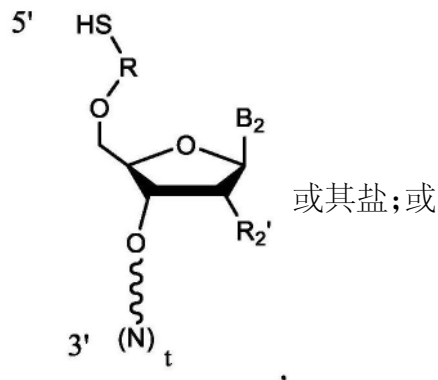
实施方式115.根据实施方式109-114中任一项所述的方法,其中:

(a) 所述第一寡核苷酸具有下式:

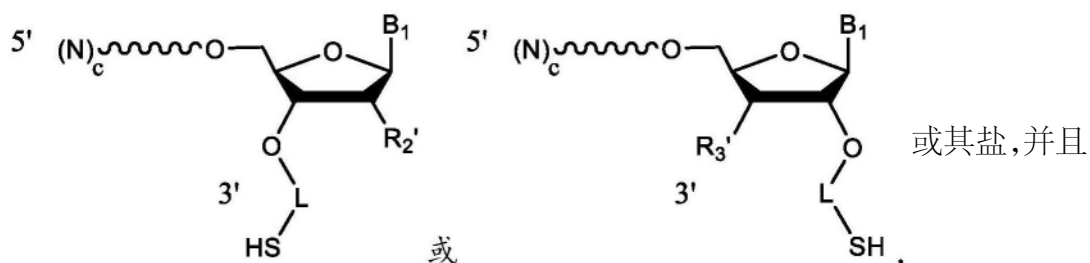


且

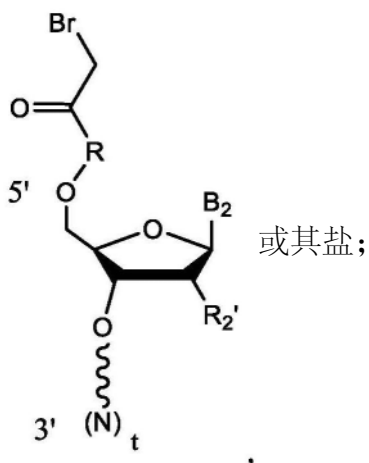
所述第二寡核苷酸具有下式:



(b) 所述第一寡核苷酸具有下式:



所述第二寡核苷酸具有下式:



其中:

$(N)_c$  和  $(N)_t$  中的每个 N 独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、酞酰乙酸酯键合、硫代酞酰乙酸酯键合或酞酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包括与  $(N)_t$  的 5' 区互补或部分互补并形成双链体的 3' 区;


c 是 20 或更大的整数;

t 是 20 或更大的整数;

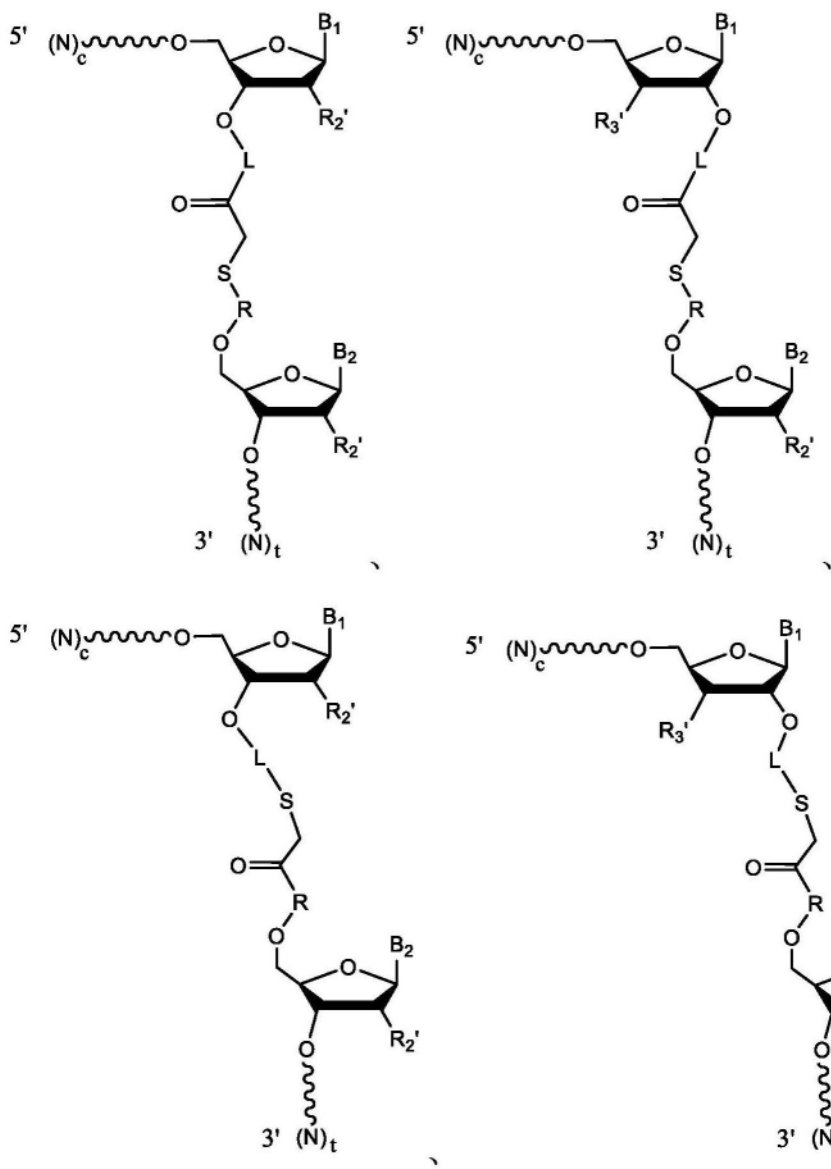
$R_2'$  和  $R_3'$  各自独立地是 H、OH、氟、氯、溴、 $NH_2$ 、SH、S- $R'$  或 O- $R'$ , 其中每个  $R'$  独立地是保护基团或烷基, 其中所述烷基可任选地被取代;

L 和 R 各自独立地是非核苷酸接头; 并且

$B_1$  和  $B_2$  各自独立地是核碱基; 并且

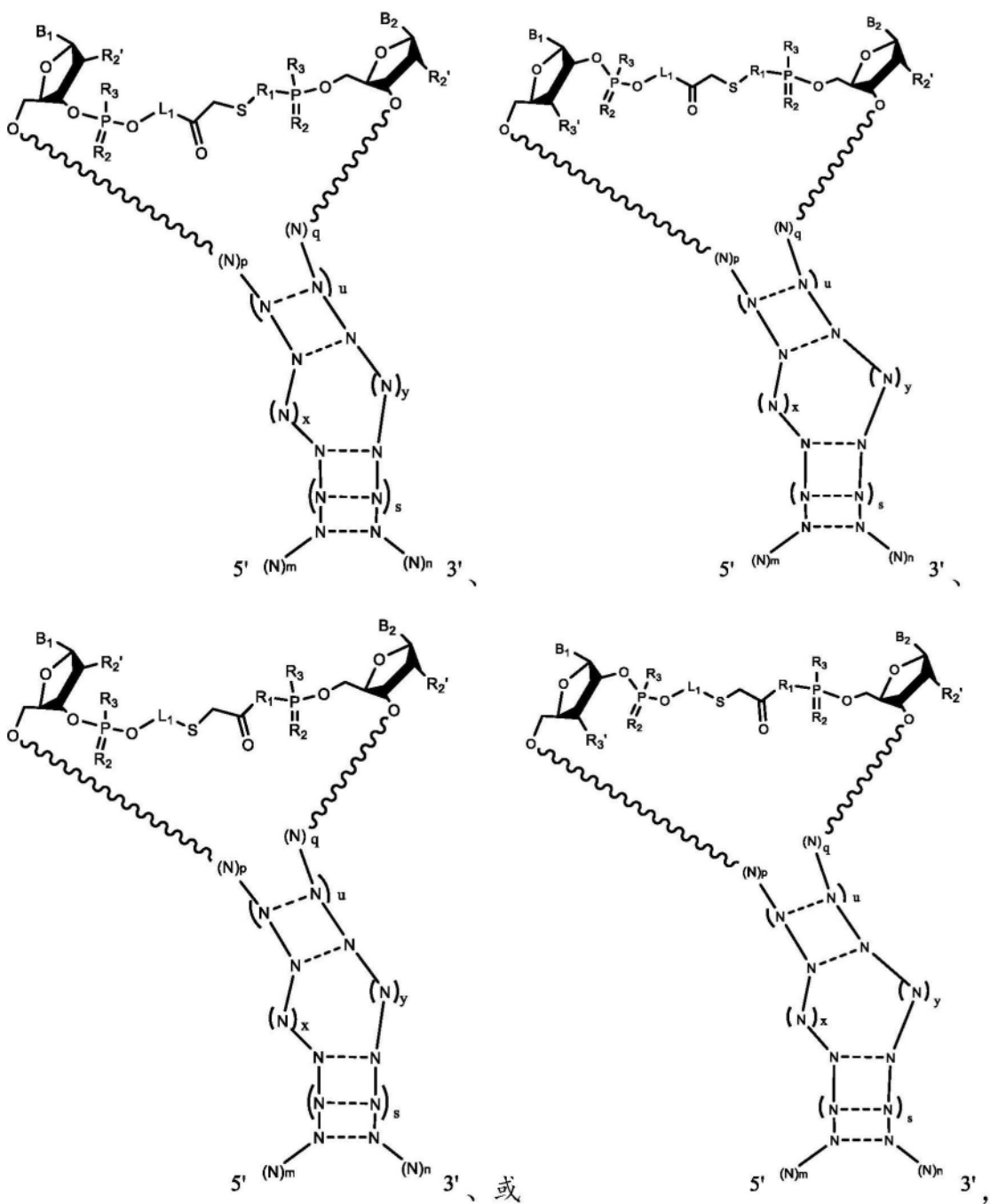
每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、酞酰乙酸酯键合、硫代酞酰乙酸酯键合或酞酰胺酯键合。

实施方式 116. 根据实施方式 109-115 中任一项所述的方法, 其中: 所述单分子指导分子具有下式:



或其盐。

实施方式117. 根据实施方式109-116中任一项所述的方法, 其中所述单分子指导分子具有下式:



y是 $\geq x$ 且3到5之间的整数,包括端值;

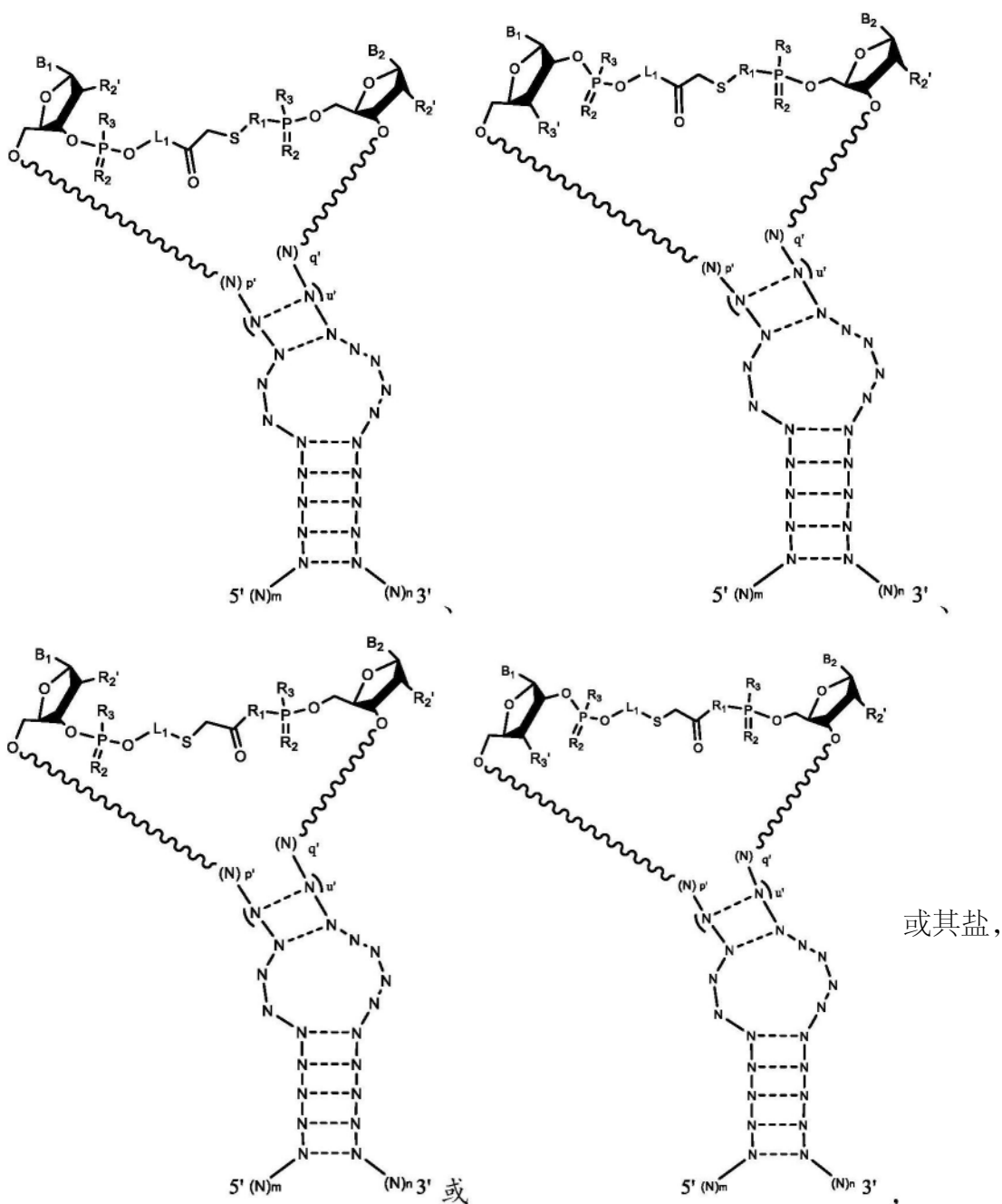
$m$ 是15或更大的整数;

$n$ 是30或更大的整数;

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或膦酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。

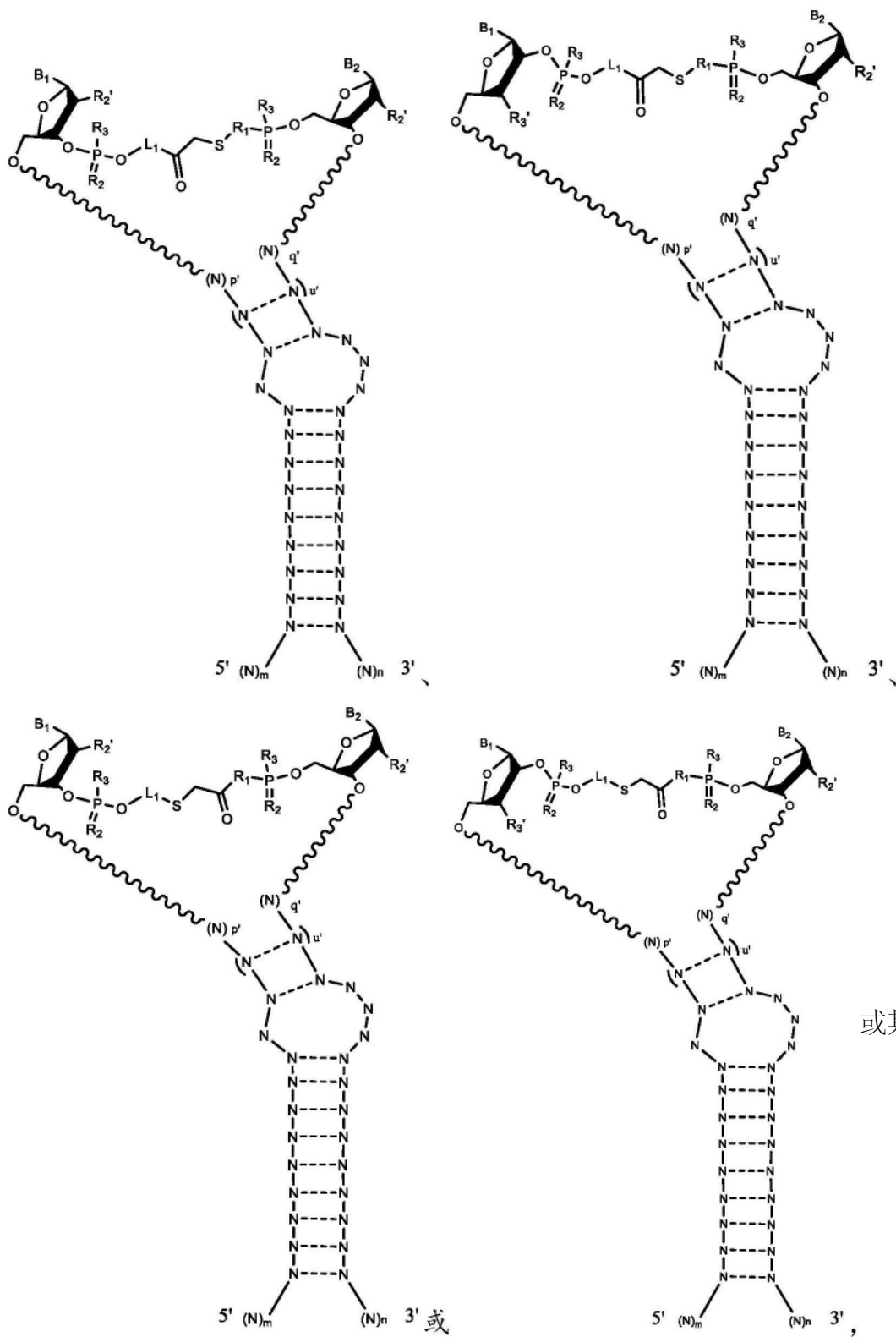
实施方式118.根据实施方式109-117中任一项所述的方法,其中所述单分子指导分子具有下式:



其中 $p'$  和 $q'$  各自独立地是0到4之间的整数,包括端值;并且 $p' + q'$  是0到4之间的整

数,包括端值;并且 $u'$ 是2到14之间的整数,包括端值。

实施方式119.根据实施方式109-117中任一项所述的方法,其中所述单分子指导分子具有下式:



盐,

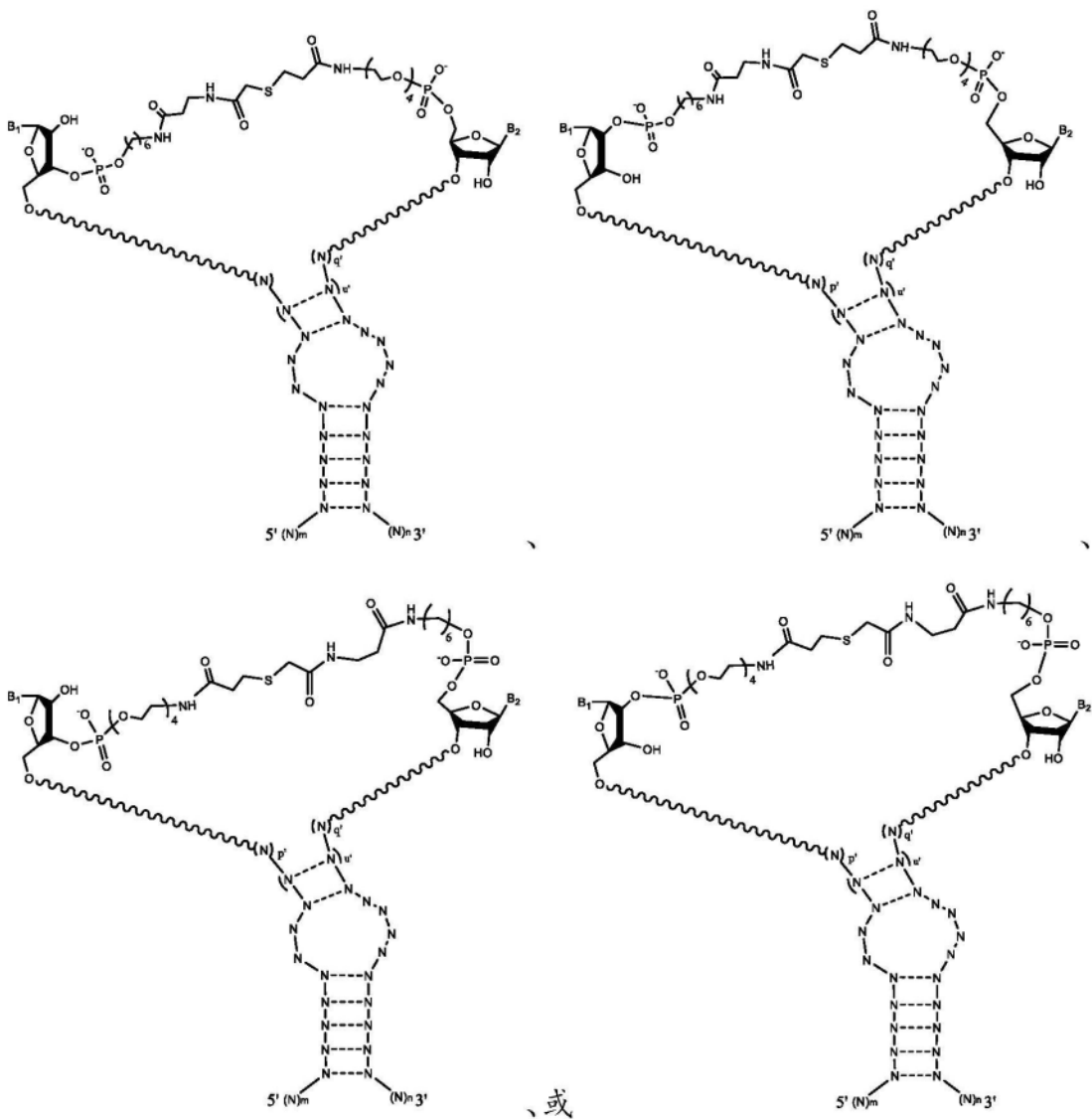
其中 $p'$ 和 $q'$ 各自独立地是0到4之间的整数,包括端值;并且 $p' + q'$ 是0到4之间的整

数,包括端值;并且 $u'$ 是2到22之间的整数,包括端值。

实施方式120.根据实施方式118或119所述的方法,其中 $p' = q'$ ,任选地其中 $p' = q' = 0$ 、 $p' = q' = 1$ 或 $p' = q' = 2$ 。

实施方式121.根据实施方式117-120中任一项所述的方法,其中 $L_1$ 和 $R_1$ 各自独立地是 $-(CH_2)_w-NH-C(O)-(CH_2)_w-NH-$ 或 $-(OCH_2CH_2)_v-NH-C(O)-(CH_2)_w-$ ;每个 $w$ 独立地是1-20;并且每个 $v$ 独立地是1-10。

实施方式122.根据实施方式115-121中任一项所述的方法,其中:所述单分子指导分子具有下式:



或其盐,其中 $p'$ 和 $q'$ 各自独立地是0到4之间的整数,包括端值;并且 $p' + q'$ 是0到4之间的整数,包括端值;并且 $u'$ 是2到14之间的整数,包括端值。

实施方式123.根据实施方式86-89中任一项所述的方法,其中(a)所述第一反应性基团包含2'或3'羟基部分并且所述第二反应性基团包含5'磷酸部分或(b)所述第一反应性基团包含5'磷酸部分并且所述第二反应性基团包含2'或3'羟基部分,并且所述缀合步骤包括在活化剂存在下使所述羟基部分和所述磷酸部分反应形成磷酸二酯键合。

实施方式124.根据实施方式123所述的方法,其中所述活化剂是碳二亚胺或其盐。

实施方式125.根据实施方式124所述的方法,其中所述碳二亚胺是1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)、N,N'-二环己基碳二亚胺(DCC)或N,N'-二异丙基碳二亚胺(DIC)或其盐。

实施方式126.根据实施方式123-125中任一项所述的方法,其中所述缀合步骤包括在活化剂和稳定剂存在下使羟基部分和磷酸部分反应。

实施方式127.根据实施方式126所述的方法,其中所述稳定剂是咪唑、氰基咪唑、吡啶或二甲基氨基吡啶、或其盐。

实施方式128.根据实施方式123-127中任一项所述的方法,其中所述第一和第二寡核苷酸中的每一个的浓度在10 $\mu$ M至1mM的范围内。

实施方式129.根据实施方式123-128中任一项所述的方法,其中所述活化剂的浓度在1mM至100mM的范围内。

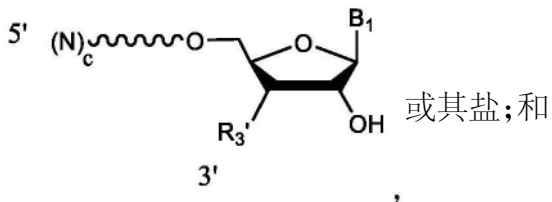
实施方式130.根据实施方式123-129中任一项所述的方法,其中所述活化剂的浓度比所述第一和第二寡核苷酸中每一个的浓度高100-1,000倍。

实施方式131.根据实施方式123-130中任一项所述的方法,其中所述缀合步骤在5-9的pH范围内进行。

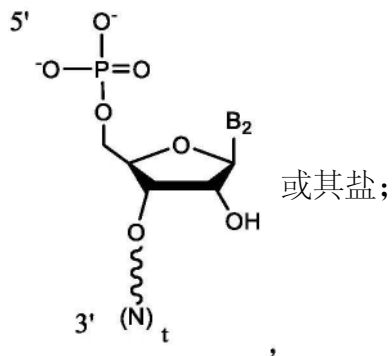
实施方式132.根据实施方式123-131中任一项所述的方法,其中所述缀合步骤在二价金属阳离子的存在下进行。

实施方式133.根据实施方式123-132中任一项所述的方法,其中所述缀合步骤在0 $^{\circ}$ C至40 $^{\circ}$ C的温度范围内进行。

实施方式134.根据实施方式123-133中任一项所述的方法,其中所述第一寡核苷酸具有下式:



所述第二寡核苷酸具有下式:



其中:

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;


(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区;

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基;

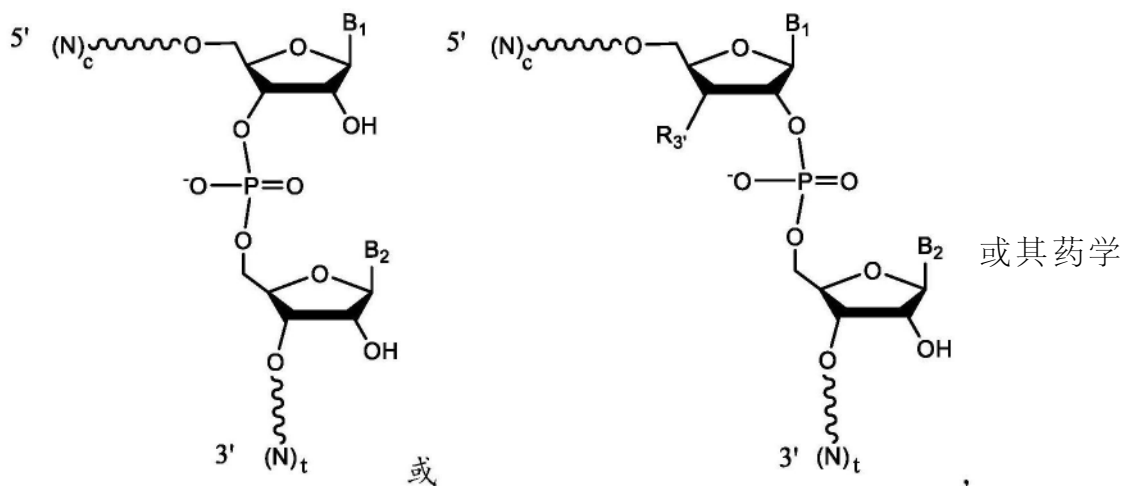
c是20或更大的整数;

t是20或更大的整数;

R<sub>3</sub>' 独立地是H、OH、氟、氯、溴、NH<sub>2</sub>、SH、S-R' 或O-R', 其中每个R' 独立地是保护基团或烷基, 其中所述烷基可任选地被取代; 并且

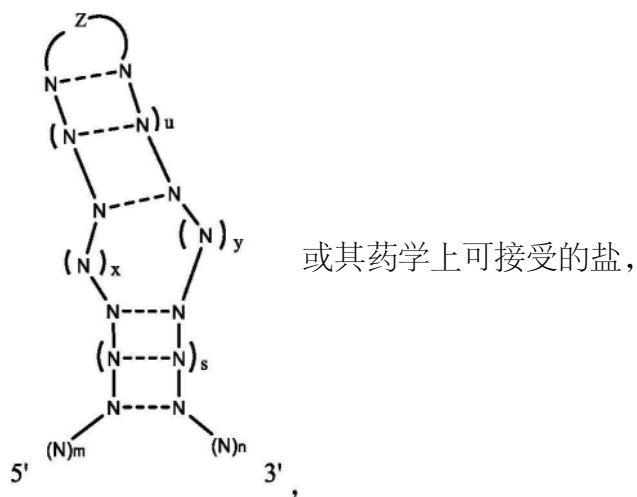
每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式135.根据实施方式123-134中任一项所述的方法,其中所述单分子指导分子具有下式:



上可接受的盐。

实施方式136.根据实施方式123-135中任一项所述的方法,其中所述单分子指导分子具有下式:



其中：

Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环;

u是2到22之间的整数,包括端值;

s是1到10之间的整数,包括端值;

x是1到3之间的整数,包括端值;

y是>x且3到5之间的整数,包括端值:

m是15或更大的整数；

n是30或更大的整数；

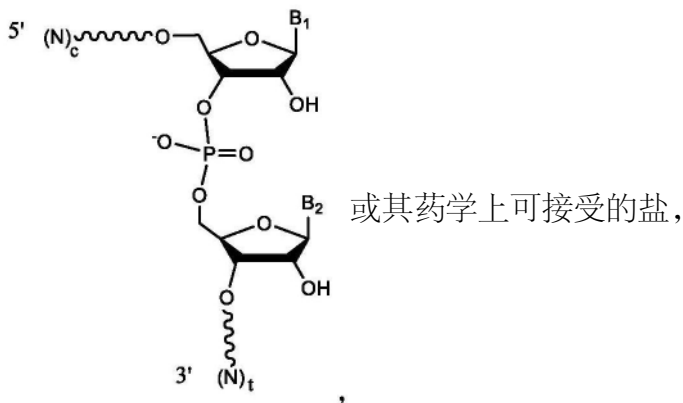
每个N独立地是核苷酸残基，任选地是经修饰的核苷酸残基，各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接；并且

每个N---N独立地代表两个互补核苷酸，任选地是氢键碱基配对的两个互补核苷酸，任选地在双链体区包含至少一个2' -5' 磷酸二酯键合。

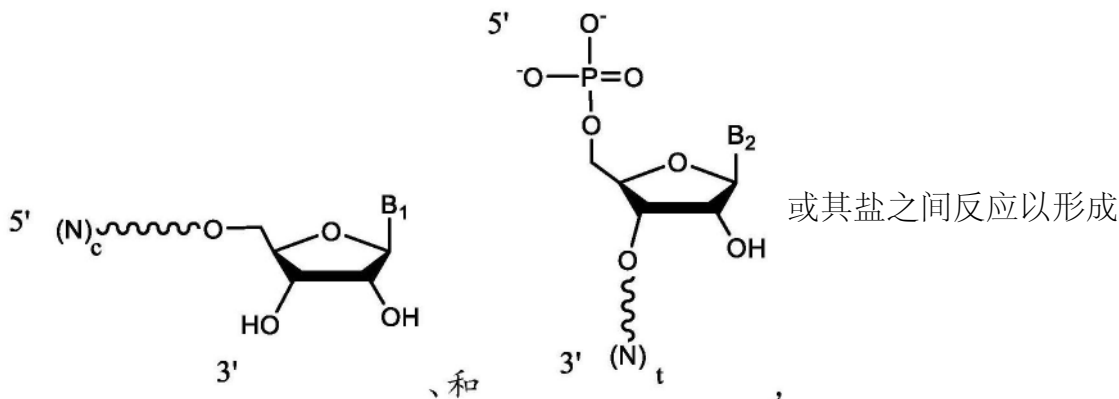
实施方式137.一种组合物，其包含多个通过根据实施方式86-136中任一项所述的方法产生的指导分子，其中相对于参照指导分子序列，少于约10%的所述指导分子在5'末端包含截短。

实施方式138.根据实施方式137所述的组合物，其中至少约99%的所述指导分子包含以下5'序列，所述5'序列包含所述指导分子的与参照指导分子序列的相应5'序列100%相同的核苷酸1-20。

实施方式139.一种组合物，其包含用于CRISPR系统的具有下式的合成的单分子指导分子：



所述单分子指导分子通过以下方法制备，所述方法包括在活化剂存在下



磷酸二酯键合，

其中：

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基，任选地是经修饰的核苷酸残基，各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接；

(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区；

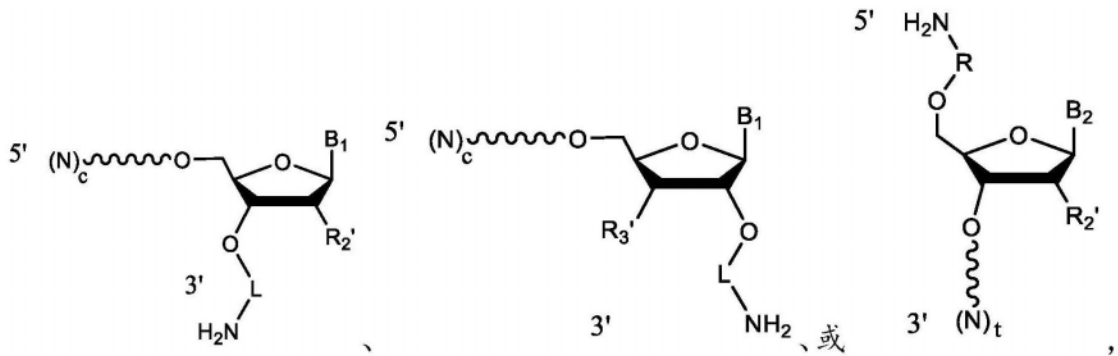
c是20或更大的整数；

t是20或更大的整数；

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基；并且

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式140.一种用于合成II型CRISPR系统的单分子指导分子的寡核苷酸,其中所述寡核苷酸具有下式:



或其盐,其中:

R<sub>2</sub>'和R<sub>3</sub>'各自独立地是H、OH、氟、氯、溴、NH<sub>2</sub>、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代;

(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

(N)<sub>c</sub>包含5'区和3'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域,所述3'区包含来自II型CRISPR系统的重复的至少一部分;

(N)<sub>t</sub>包含5'区,所述5'区包含来自II型CRISPR系统的抗重复的至少一部分;

c是20或更大的整数;

t是20或更大的整数;

L和R各自独立地是非核苷酸接头;

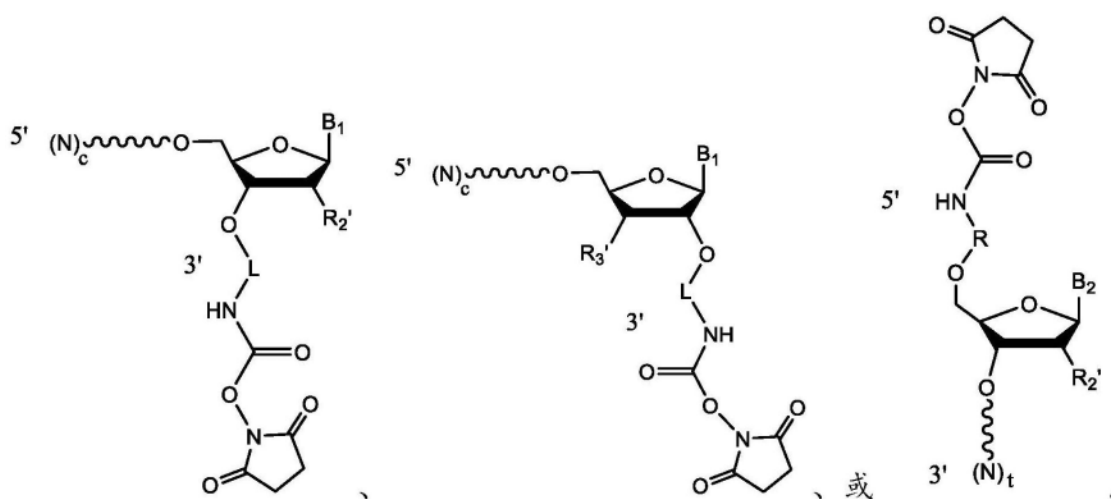
B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基;并且

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式141.根据实施方式140所述的寡核苷酸,其中(N)<sub>c</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式142.根据实施方式140或141所述的寡核苷酸,其中(N)<sub>t</sub>包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式143.一种用于合成II型CRISPR系统的单分子指导分子的寡核苷酸中间体,其中所述寡核苷酸中间体具有下式:



或其盐,其中:

$R_2'$  和  $R_3'$  各自独立地是 H、OH、氟、氯、溴、 $\text{NH}_2$ 、SH、S- $R'$  或 O- $R'$ , 其中每个  $R'$  独立地是保护基团或烷基, 其中所述烷基可任选地被取代;

$(N)_c$  和  $(N)_t$  中的每个 N 独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包含 5' 区和 3' 区, 所述 5' 区包含与靶序列内的靶结构域完全或部分互补的靶向结构域, 所述 3' 区包含来自 II 型 CRISPR 系统的重复的至少一部分;


$(N)_t$  包含 5' 区, 所述 5' 区包含来自 II 型 CRISPR 系统的抗重复的至少一部分;

c 是 20 或更大的整数;

t 是 20 或更大的整数;

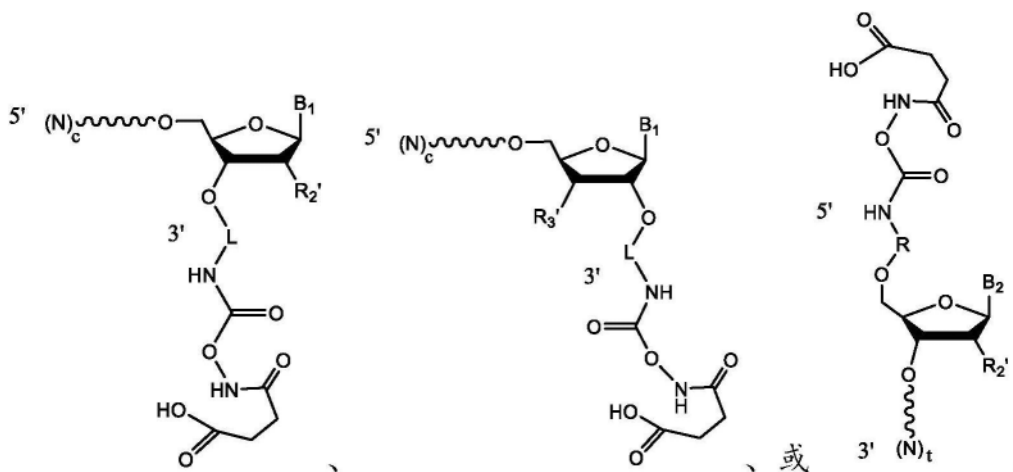
L 和 R 各自独立地是非核苷酸接头;

$B_1$  和  $B_2$  各自独立地是核碱基; 并且

每个  独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式 144. 根据实施方式 143 所述的寡核苷酸, 其中  $(N)_t$  包含 3' 区, 所述 3' 区包含一个或多个茎-环结构。

实施方式 145. 一种用于合成 II 型 CRISPR 系统的单分子指导分子的寡核苷酸中间体, 其中所述寡核苷酸中间体具有下式:



或其药学上可接受的盐,其中:

$R_2'$  和  $R_3'$  各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代;

$(N)_c$  和  $(N)_t$  中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$  包含5'区和3'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域,所述3'区包含来自II型CRISPR系统的重复的至少一部分;

$(N)_t$  包含5'区,所述5'区包含来自II型CRISPR系统的抗重复的至少一部分;

c是20或更大的整数;

t是20或更大的整数;

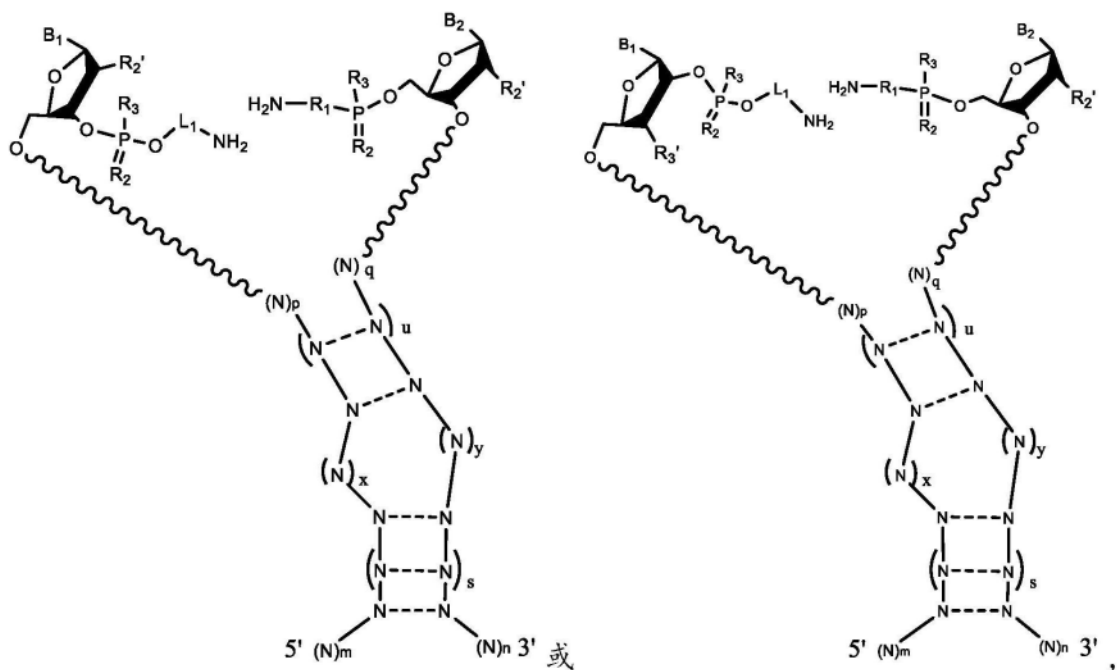
L和R各自独立地是非核苷酸接头;

$B_1$  和  $B_2$  各自独立地是核碱基;并且

每个 ~ 独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式146.根据实施方式145所述的寡核苷酸,其中  $(N)_t$  包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式147.一种组合物,其包含用于合成II型CRISPR系统的单分子指导分子的具有退火双链体的中间体,其中所述中间体具有下式:



或其盐,其中:

$L_1$ 和 $R_1$ 各自独立地是非核苷酸接头;

每个 $R_2$ 独立地是O或S;

每个 $R_3$ 独立地是 $O^-$ 或 $COO^-$ ;

$p$ 和 $q$ 各自独立地是0到6之间的整数,包括端值;并且 $p+q$ 是0到6之间的整数,包括端值;

$u$ 是2到22之间的整数,包括端值;

$s$ 是1到10之间的整数,包括端值;

$x$ 是1到3之间的整数,包括端值;

$y$ 是 $>x$ 且3到5之间的整数,包括端值;

$m$ 是15或更大的整数;

$n$ 是30或更大的整数;

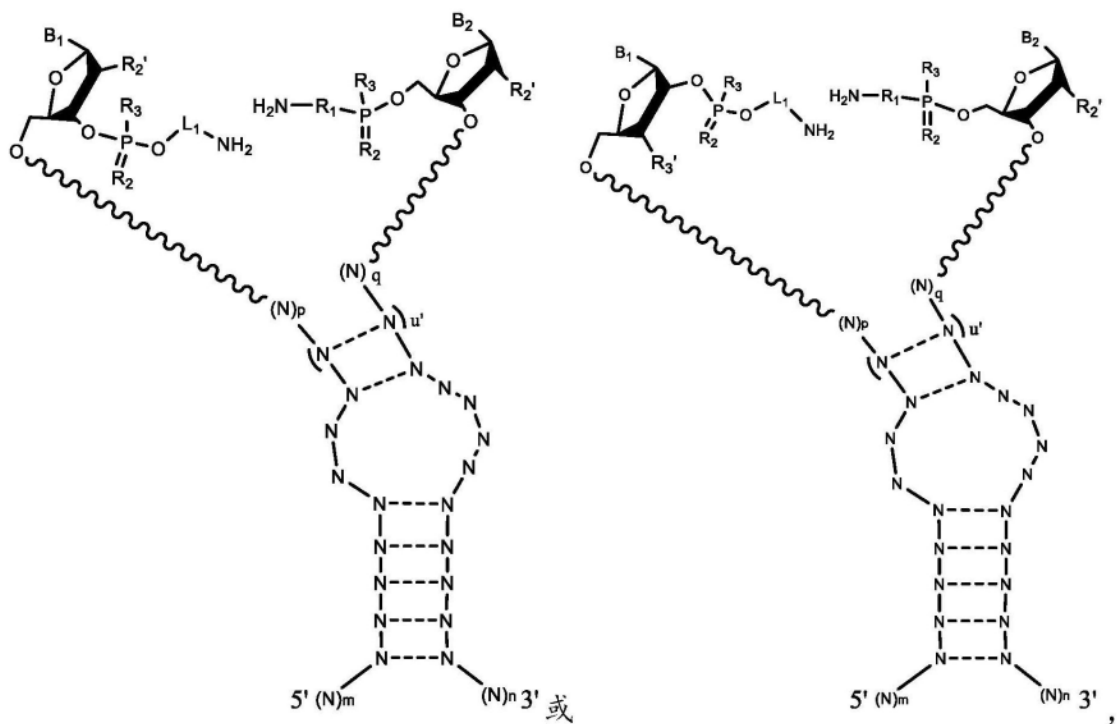
每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。

实施方式148. 根据实施方式147所述的组合物,其中 $p$ 和 $q$ 各为0。

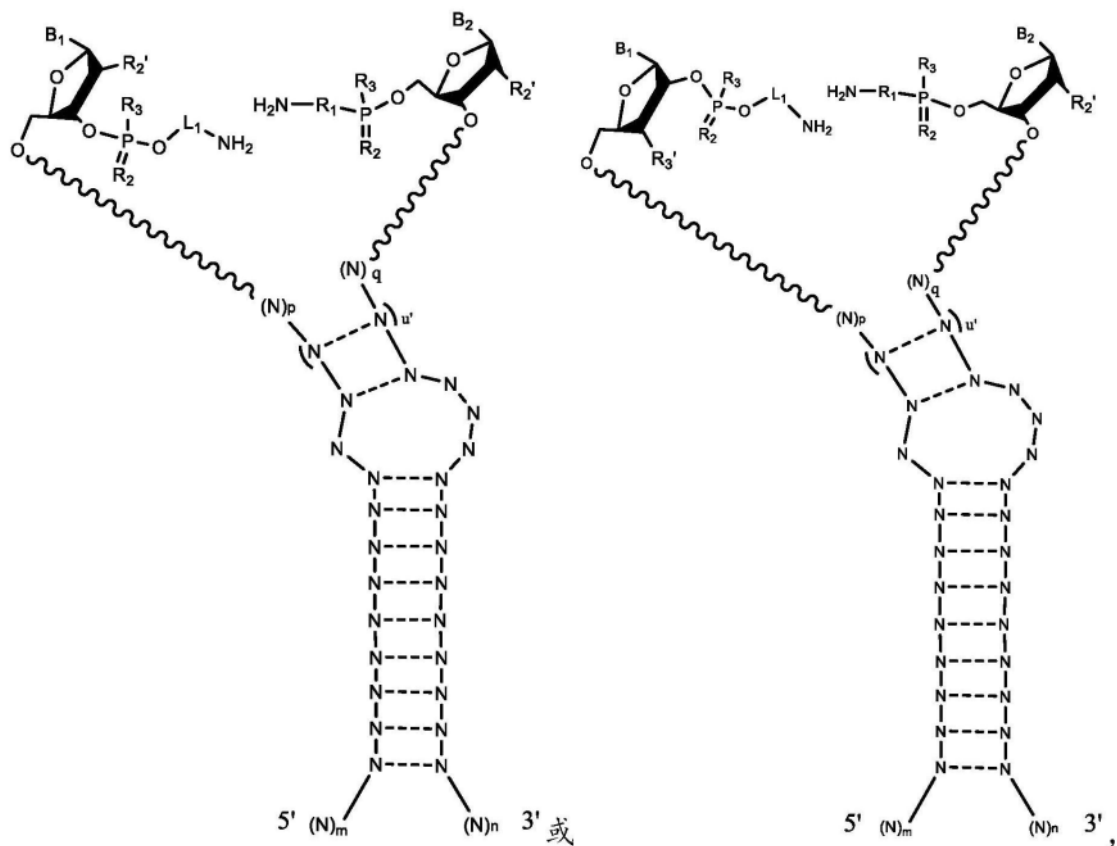
实施方式149. 根据实施方式147所述的组合物,其中 $u$ 是3到22之间的整数。

实施方式150. 根据实施方式147-149中任一项所述的组合物,其中所述中间体具有下式:



或其盐。

实施方式151.根据实施方式147-149中任一项所述的组合物,其中所述中间体具有下式:



或其盐。

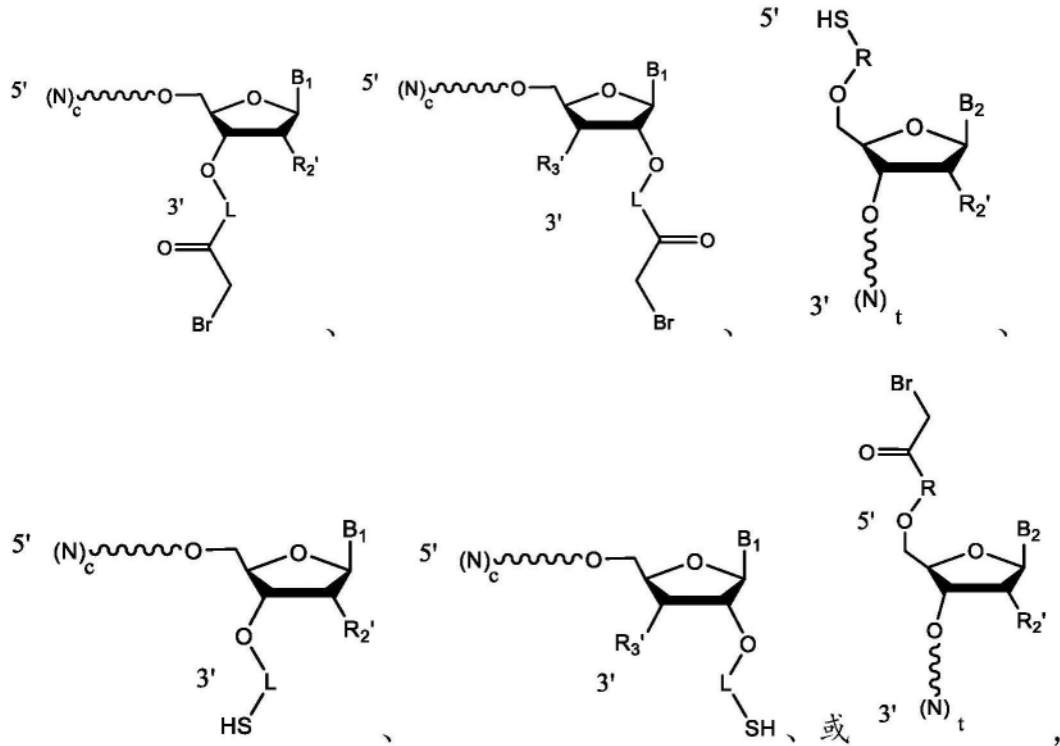
实施方式152. 根据实施方式150或151所述的组合物, 其中p'和q'各为0。

实施方式153.根据实施方式150或151所述的组合物,其中 $u'$ 是3到22之间的整数。

实施方式154.根据实施方式147-153中任一项所述的组合物,其中 $(N)_m$ 包含5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

实施方式155.根据实施方式147-154中任一项所述的组合物,其中 $(N)_n$ 包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式156.一种用于合成II型CRISPR系统的单分子指导分子的寡核苷酸,其中所述寡核苷酸具有下式:



或其盐,其中:

$R_2'$ 和 $R_3'$ 各自独立地是H、OH、氟、氯、溴、 $NH_2$ 、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代;

L和R各自独立地是非核苷酸接头;

$B_1$ 和 $B_2$ 各自独立地是核碱基;

$(N)_c$ 和 $(N)_t$ 中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$ 包括与 $(N)_t$ 的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数;并且

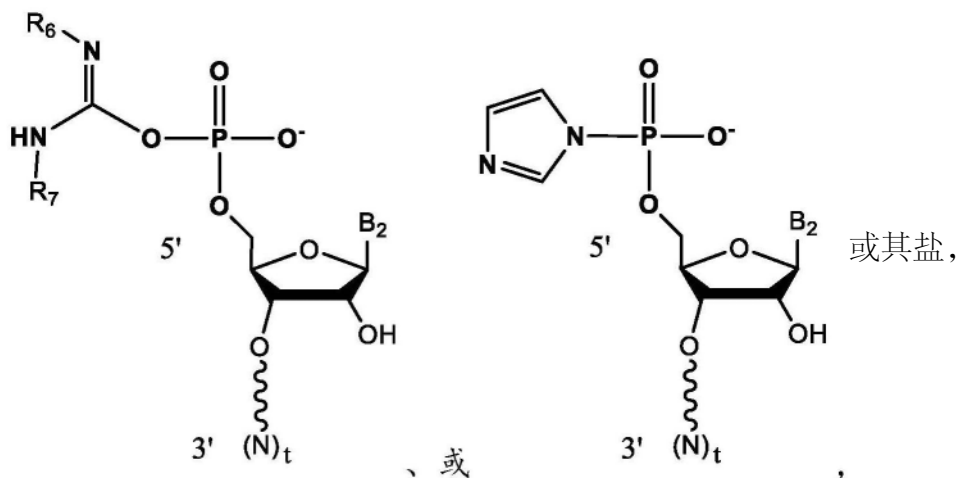
每个 $\sim$ 独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式157.根据实施方式156所述的寡核苷酸,其中 $(N)_c$ 包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式158.根据实施方式156或157所述的寡核苷酸,其中 $(N)_t$ 包含3'区,所述

3' 区包含一个或多个茎-环结构。

实施方式159. 一种用于合成II型CRISPR系统的单分子指导分子的寡核苷酸中间体, 其中所述寡核苷酸中间体具有下式:



其中:

$R_6$ 和 $R_7$ 各自独立地是取代的或未被取代的烷基, 或取代的或未被取代的碳环;

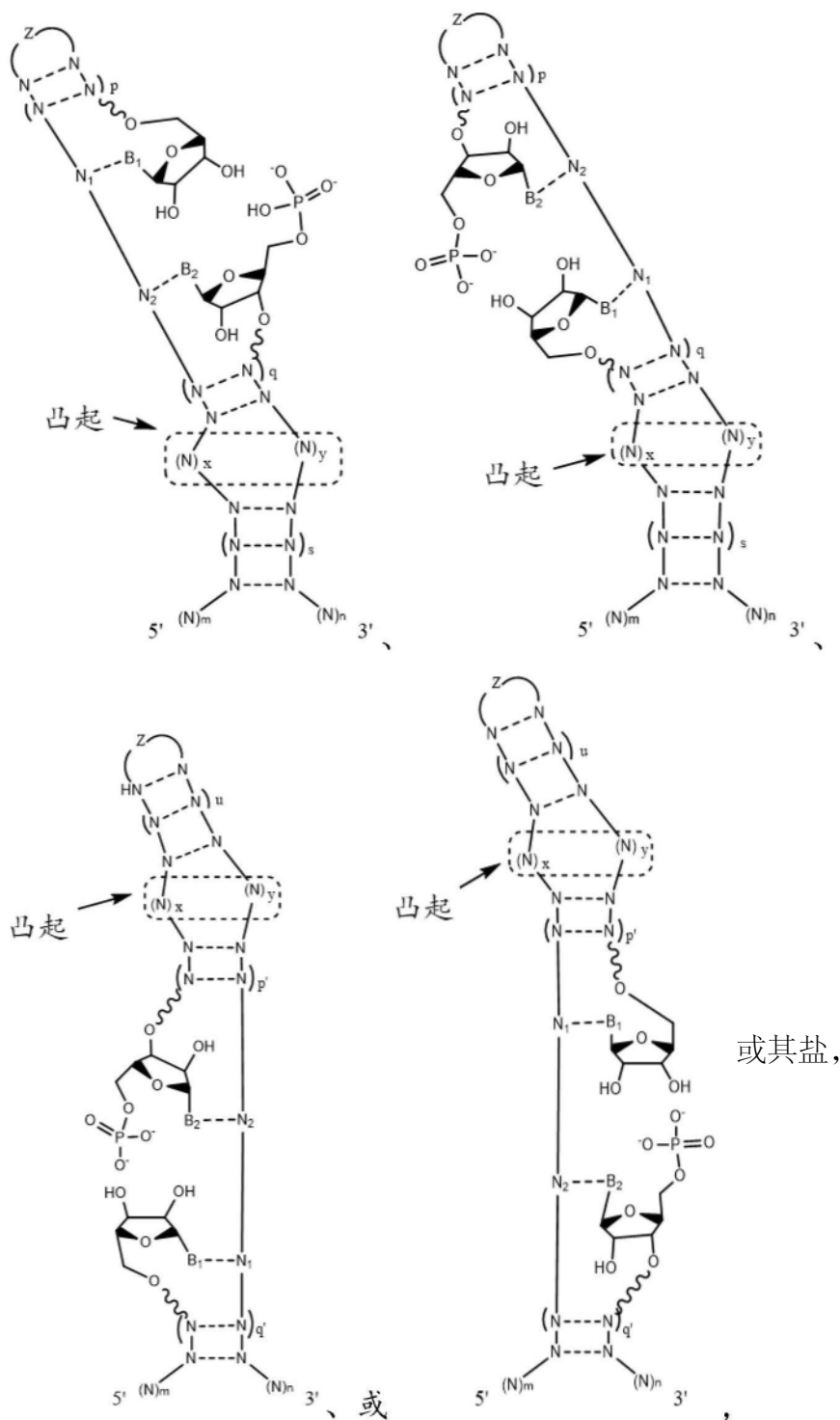
$(N)_t$ 包含5' 区, 所述5' 区包含来自II型CRISPR系统的抗重复的至少一部分, 其中 $(N)_t$ 中的每个N独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$t$ 是20或更大的整数;

$B_2$ 是核碱基; 并且

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合。

实施方式160. 一种组合物, 其包含用于合成II型CRISPR系统的单分子指导分子的具有退火双链体的中间体, 其中所述中间体具有下式:



其中:

Z代表4-6个核苷酸长、任选地4或6个核苷酸长的核苷酸环;

p和q各自独立地是0到6之间的整数,包括端值;并且p+q是0到6之间的整数,包括端值;

u是2到22之间的整数,包括端值;

s是1到10之间的整数,包括端值;

x是1到3之间的整数,包括端值;

y是>x且3到5之间的整数,包括端值;

m是15或更大的整数；

n是30或更大的整数；

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸。

实施方式161.根据实施方式160所述的组合物,其中 $(N)_m$ 包含5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

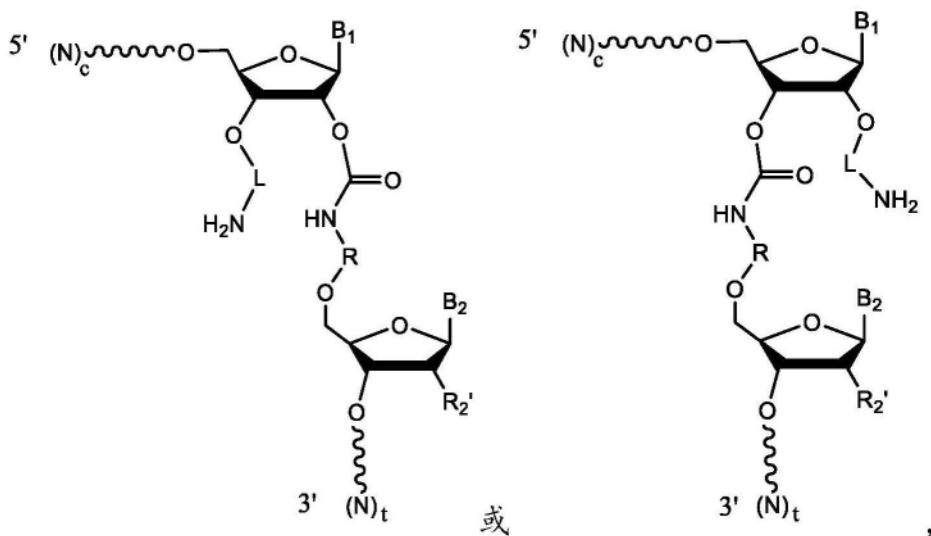
实施方式162.根据实施方式160或161所述的组合物,其中 $(N)_n$ 包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式163.根据实施方式160-162中任一项所述的组合物,其中所述组合物还包含碳二亚胺或其盐,和/或咪唑、氰基咪唑、吡啶和二甲基氨基吡啶,或其盐。

实施方式164.根据实施方式163所述的组合物,其中碳二亚胺是EDC、DCC或DIC。

实施方式165.根据实施方式160-164中任一项所述的组合物,其中所述组合物包含碳二亚胺或其盐以及咪唑或其盐。

实施方式166.一种具有下式的化合物:



其中:

$(N)_c$ 和 $(N)_t$ 中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$ 包括与 $(N)_t$ 的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

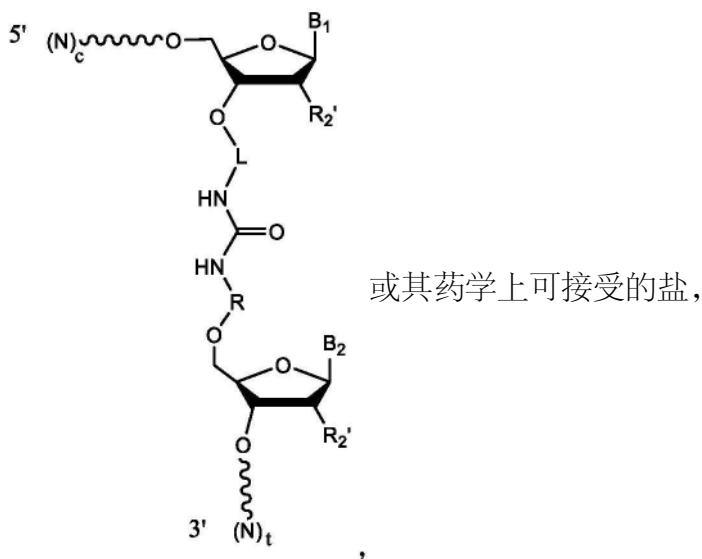
t是20或更大的整数;并且

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合;

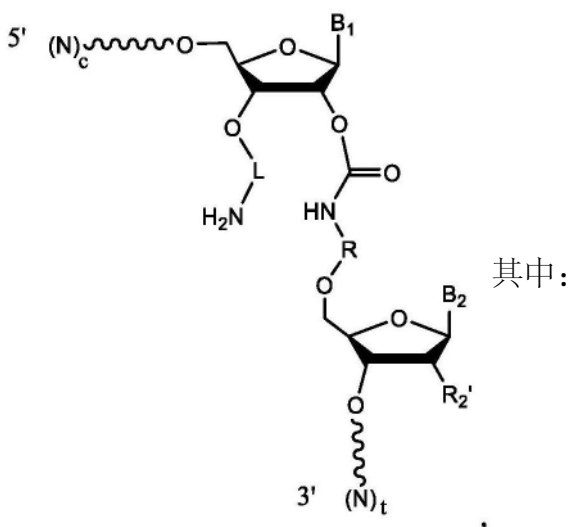
L和R各自独立地是非核苷酸接头;并且

$B_1$ 和 $B_2$ 各自独立地是核碱基。

实施方式167.一种组合物,所述组合物包含以下或基本上由以下组成:根据实施方式1-23中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:



$(N)_c$ 和 $(N)_t$ 中的每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

$(N)_c$ 包括与 $(N)_t$ 的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数;并且

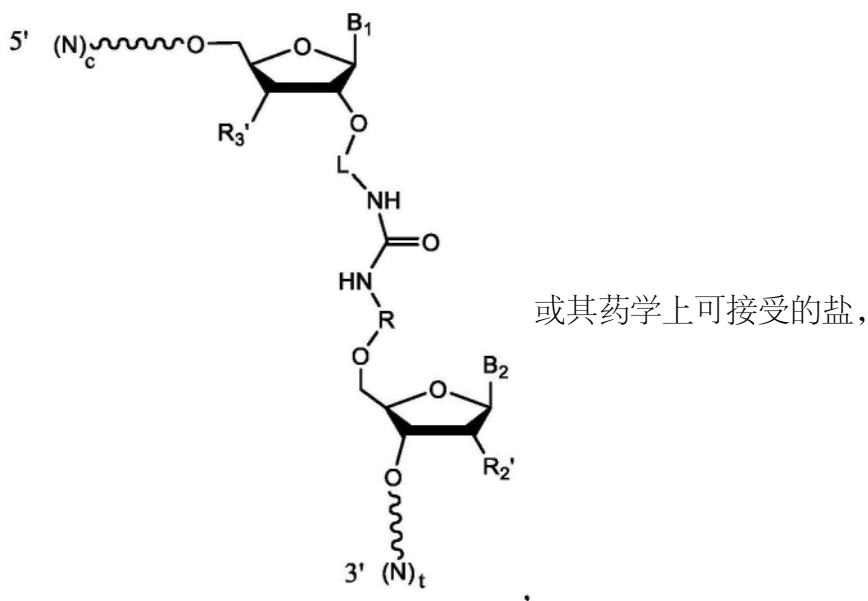
每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合;

L和R各自独立地是非核苷酸接头;并且

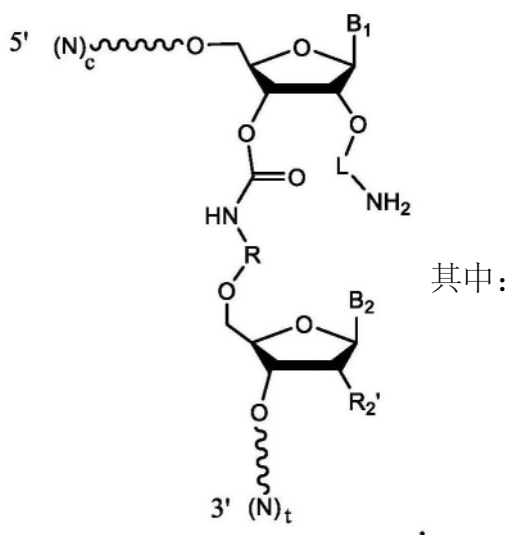
$B_1$ 和 $B_2$ 各自独立地是核碱基,

和/或,或其药学上可接受的盐。

实施方式168. 一种组合物, 所述组合物包含以下或基本上由以下组成: 根据实施方式1-23中任一项所述的具有下式的指导分子



其中所述组合物基本上不含具有下式的分子:



(N)<sub>c</sub>和(N)<sub>t</sub>中的每个N独立地是核苷酸残基, 任选地是经修饰的核苷酸残基, 各自独立地通过磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;

(N)<sub>c</sub>包括与(N)<sub>t</sub>的5'区互补或部分互补并形成双链体的3'区;

c是20或更大的整数;

t是20或更大的整数; 并且

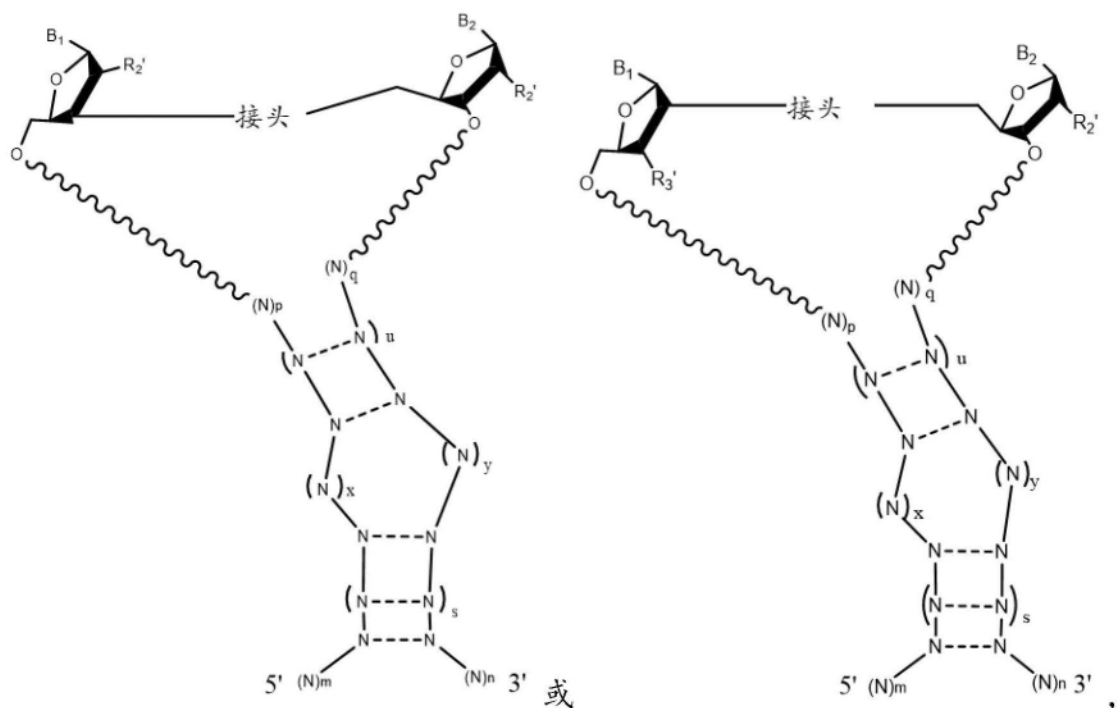
每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、磷酰乙酸酯键合、硫代磷酰乙酸酯键合或磷酰胺酯键合;

L和R各自独立地是非核苷酸接头; 并且

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基,

和/或, 或其药学上可接受的盐。

实施方式169.一种用于CRISPR系统的具有下式的合成的单分子指导分子:



其中:

每个N独立地是核苷酸残基,任选地是经修饰的核苷酸残基,各自独立地通过磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合与其相邻的一个或多个核苷酸连接;并且

每个N---N独立地代表两个互补核苷酸,任选地是氢键碱基配对的两个互补核苷酸

每个~独立地代表磷酸二酯键合、硫代磷酸酯键合、膦酰乙酸酯键合、硫代膦酰乙酸酯键合或磷酰胺酯键合;

p和q各为0;

u是2到22之间的整数,包括端值;

s是1到10之间的整数,包括端值;

x是1到3之间的整数,包括端值;

y是>x且3到5之间的整数,包括端值;

m是15或更大的整数;

n是30或更大的整数;

接头是非核苷酸化学键合;

B<sub>1</sub>和B<sub>2</sub>各自独立地是核碱基;并且

R<sub>2</sub>'和R<sub>3</sub>'各自独立地是H、OH、氟、氯、溴、NH<sub>2</sub>、SH、S-R'或O-R',其中每个R'独立地是保护基团或烷基,其中所述烷基可任选地被取代。

实施方式170.根据实施方式169所述的指导分子,其中R<sub>2</sub>'和R<sub>3</sub>'各自独立地选自由H、氟和O-R'组成的组,其中R'是保护基团或任选地被取代的烷基。

实施方式171.根据实施方式169或170所述的指导分子,其中每个N独立地是核糖核苷酸残基或糖修饰的核糖核苷酸残基。

实施方式172.根据实施方式169-171中任一项所述的指导分子,其中一个或多个N是脱氧核糖核苷酸残基。

实施方式173.根据实施方式169-172中任一项所述的指导分子,其中一个或多个N是2'-O-甲基修饰的核糖核苷酸残基。

实施方式174.根据实施方式169-173中任一项所述的指导分子,其中(N)<sub>m</sub>的5'末端的三个核苷酸中的每一个和/或(N)<sub>n</sub>的3'末端的三个核苷酸中的每一个包含2'-O-甲基修饰的核糖核苷酸残基,所述2'-O-甲基修饰的核糖核苷酸残基通过硫代磷酸酯键合与其相邻的一个或多个核苷酸连接。

实施方式175.根据实施方式169-174中任一项所述的指导分子,其中所述指导分子用于II型CRISPR系统,并且(N)<sub>m</sub>包括5'区,所述5'区包含与靶序列内的靶结构域完全或部分互补的靶向结构域。

实施方式176.根据实施方式169-175中任一项所述的指导分子,其中(N)<sub>n</sub>包含3'区,所述3'区包含一个或多个茎-环结构。

实施方式177.根据实施方式169-176中任一项所述的指导分子,其中所述指导分子能够与Cas9分子相互作用并介导Cas9/指导分子核糖核蛋白复合物的形成。

实施方式178.根据实施方式169-177中任一项所述的指导分子,其中(N)<sub>m</sub>包含3'区,所述3'区包含来自II型CRISPR系统的重复的至少一部分。

实施方式179.根据实施方式169-178中任一项所述的指导分子,其中(N---N)<sub>u</sub>和(N---N)<sub>s</sub>不包含具有3个或更多个核苷酸的相同序列。

实施方式180.根据实施方式179所述的指导分子,其中(N---N)<sub>u</sub>和(N---N)<sub>s</sub>不包含具有4个或更多个核苷酸的相同序列。

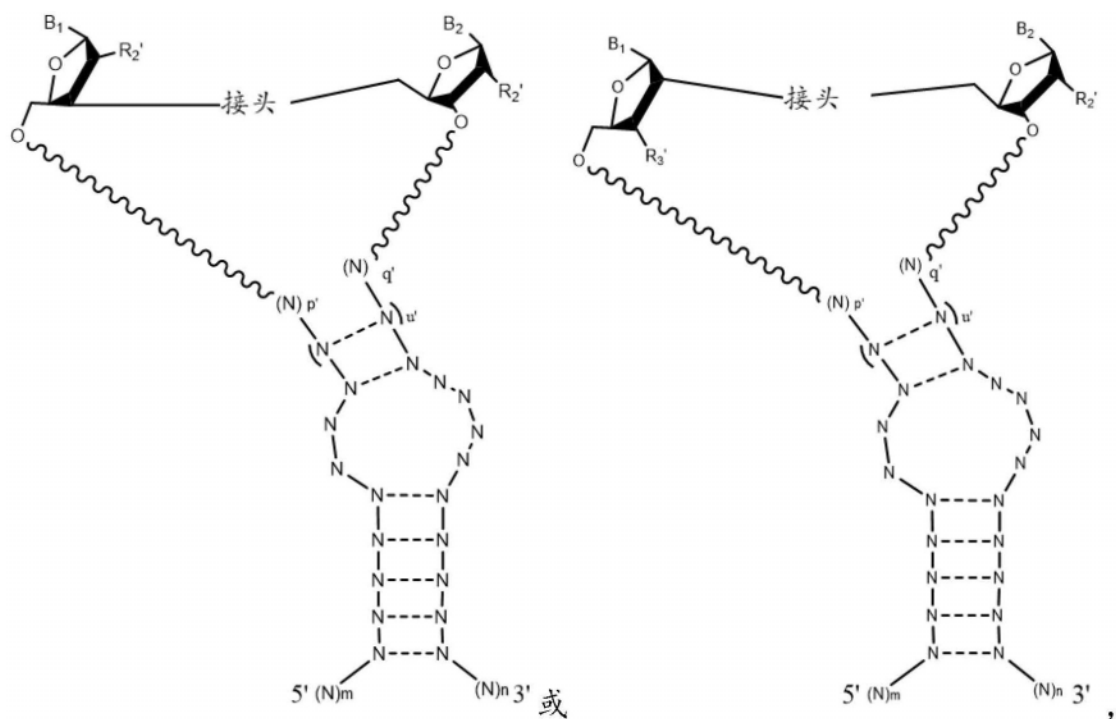
实施方式181.根据实施方式180所述的指导分子,其中(N---N)<sub>s</sub>包含N'UUU、UN'UU、UUN'U或UUUN'序列并且(N---N)<sub>u</sub>包含UUUU序列,其中N'是A、G或C。

实施方式182.根据实施方式181所述的指导分子,其中N'是A。

实施方式183.根据实施方式180所述的指导分子,其中(N---N)<sub>s</sub>的下茎序列包含UUUU序列并且(N---N)<sub>u</sub>包含N'UUU、UN'UU、UUN'U或UUUN'序列,其中N'是A、G或C。

实施方式184.根据实施方式183所述的指导分子,其中N'是A。

实施方式185.根据实施方式169-184中任一项所述的指导分子,其中所述指导分子具有下式:

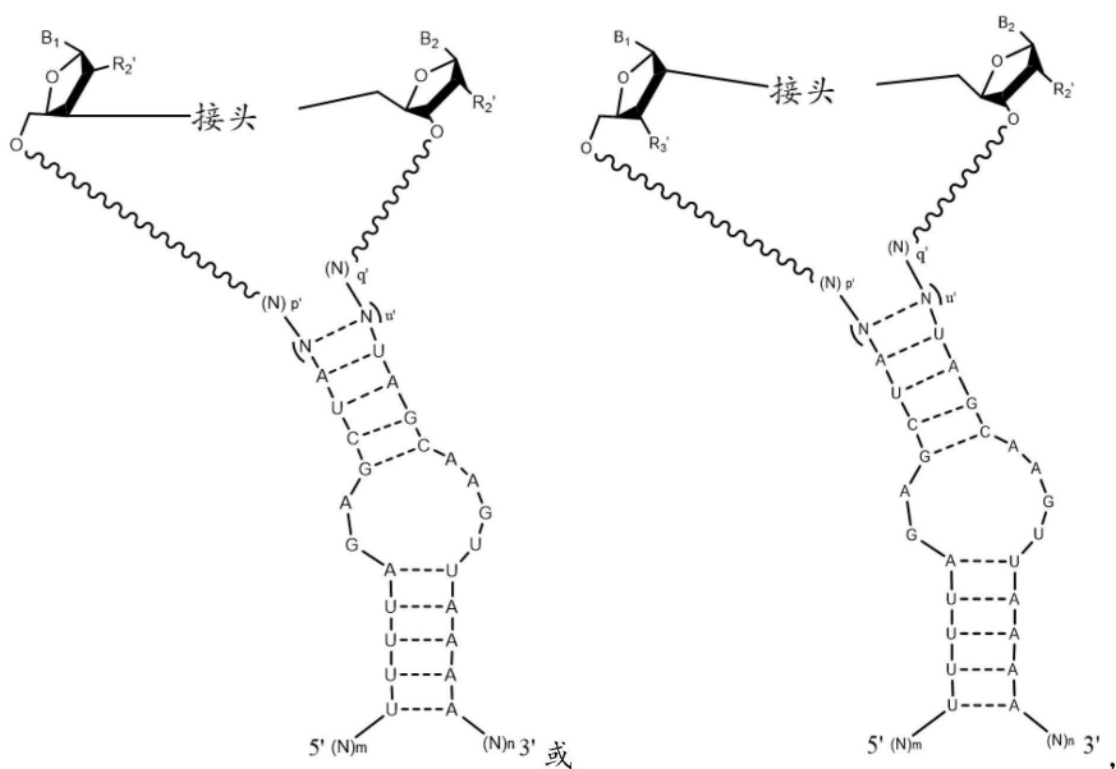


其中：

u' 是2到22之间的整数,包括端值;和

$p'$ 和 $q'$ 各为0。

实施方式186.根据实施方式185所述的指导分子,其中所述指导分子具有下式:

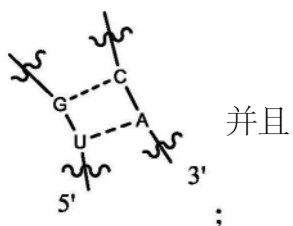


或其协变体,其中:

$p'$ 和 $q'$ 各为0;并且

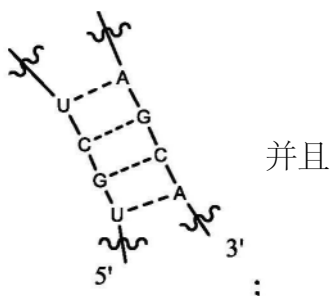
u' 是0到15之间的整数,包括端值。

实施方式187. 根据实施方式186所述的指导分子, 其中  $(N\cdots N)_w$  具有下式:



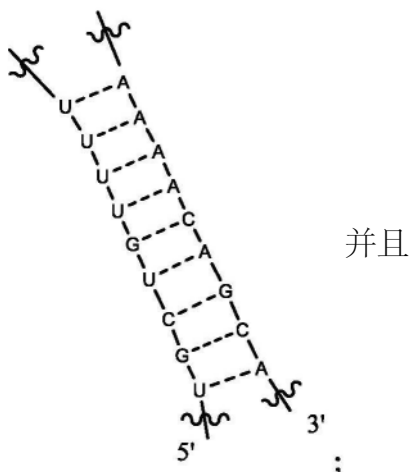
其中B<sub>1</sub>是胞嘧啶残基并且B<sub>2</sub>是鸟嘌呤残基;或其协变体。

实施方式188. 根据实施方式186所述的指导分子, 其中 (N---N)<sub>u</sub>, 具有下式:



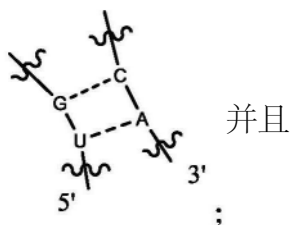
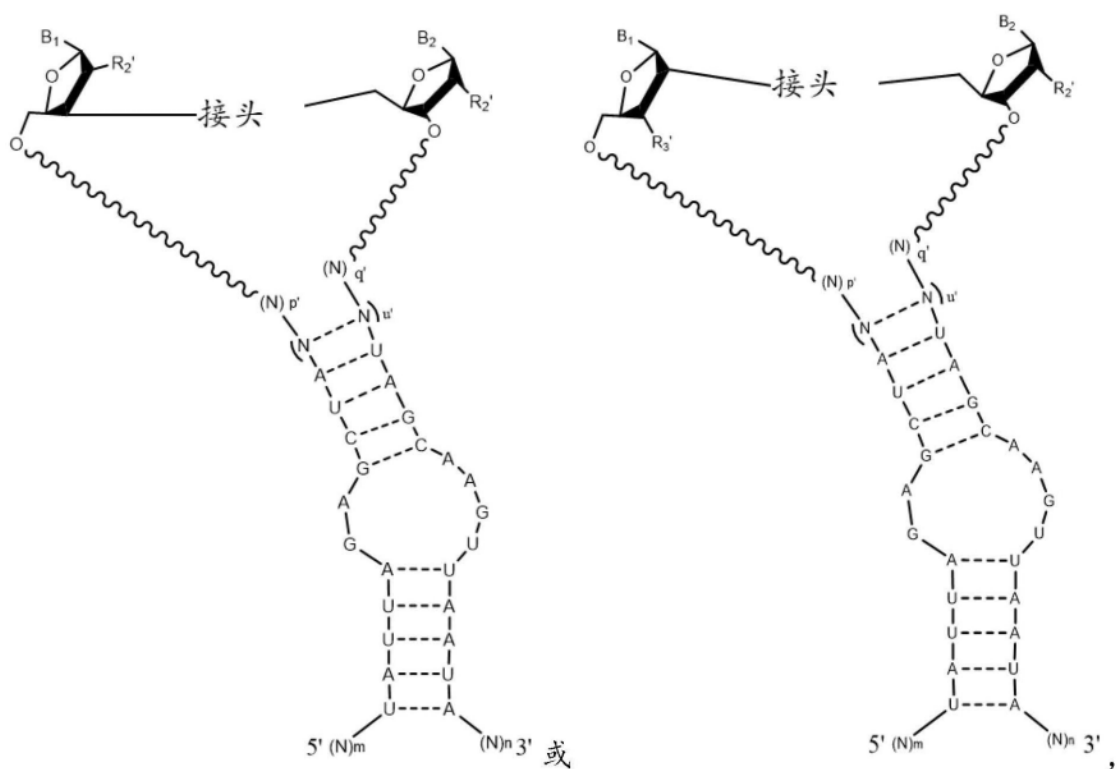
其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式189. 根据实施方式186所述的指导分子, 其中 (N---N)<sub>n</sub>, 具有下式:



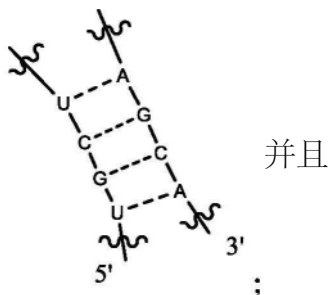
其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式190.根据实施方式185所述的指导分子,其中所述指导分子具有下式:



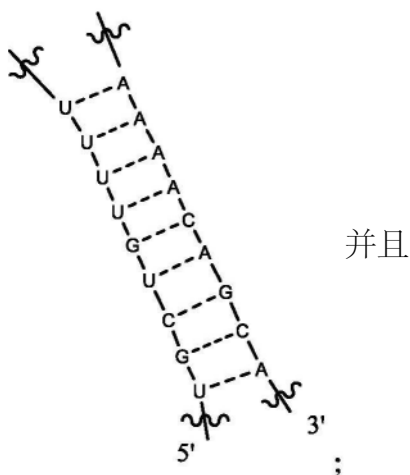
其中 $B_1$ 是胞嘧啶残基并且 $B_2$ 是鸟嘌呤残基;或其协变体。

实施方式192.根据实施方式190所述的指导分子,其中 $(N---N)_{u'}$ 具有下式:



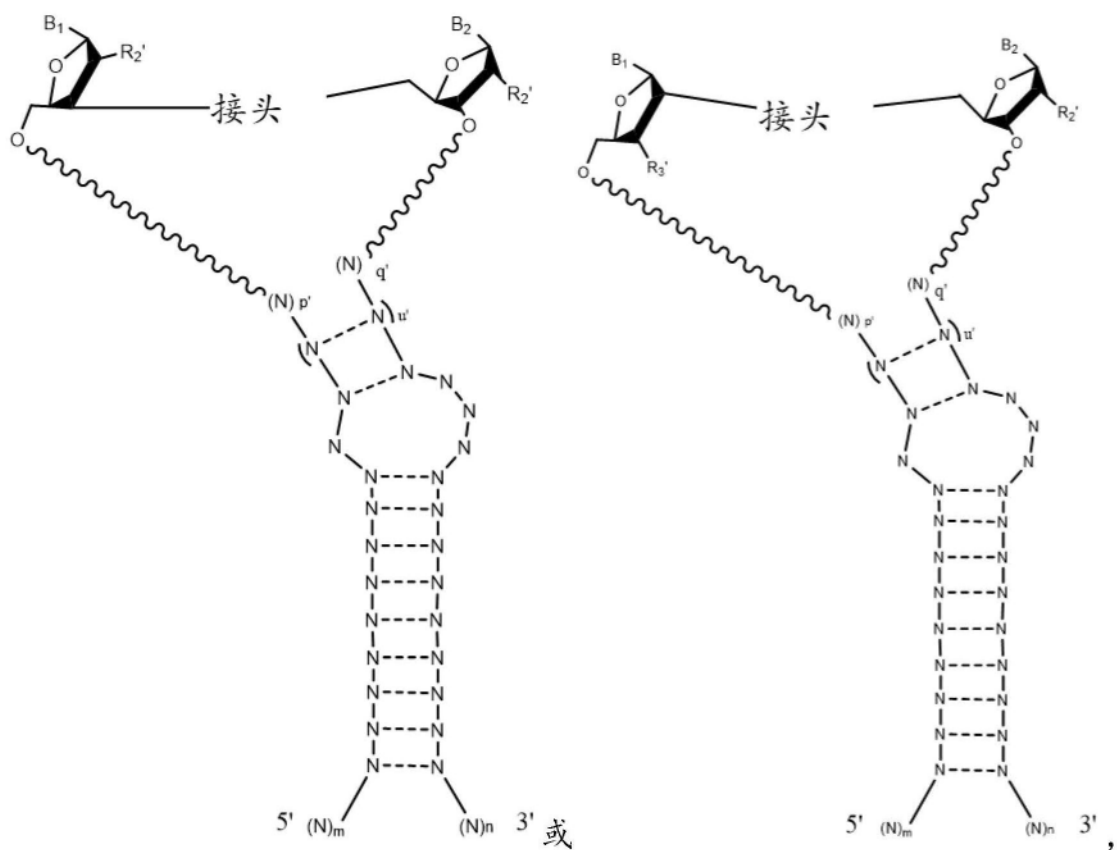
其中 $B_1$ 是鸟嘌呤残基并且 $B_2$ 是胞嘧啶残基;或其协变体。

实施方式193.根据实施方式190所述的指导分子,其中 $(N---N)_{u'}$ 具有下式:



其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式194.根据实施方式169-184中任一项所述的指导分子,其中所述指导分子具有下式:

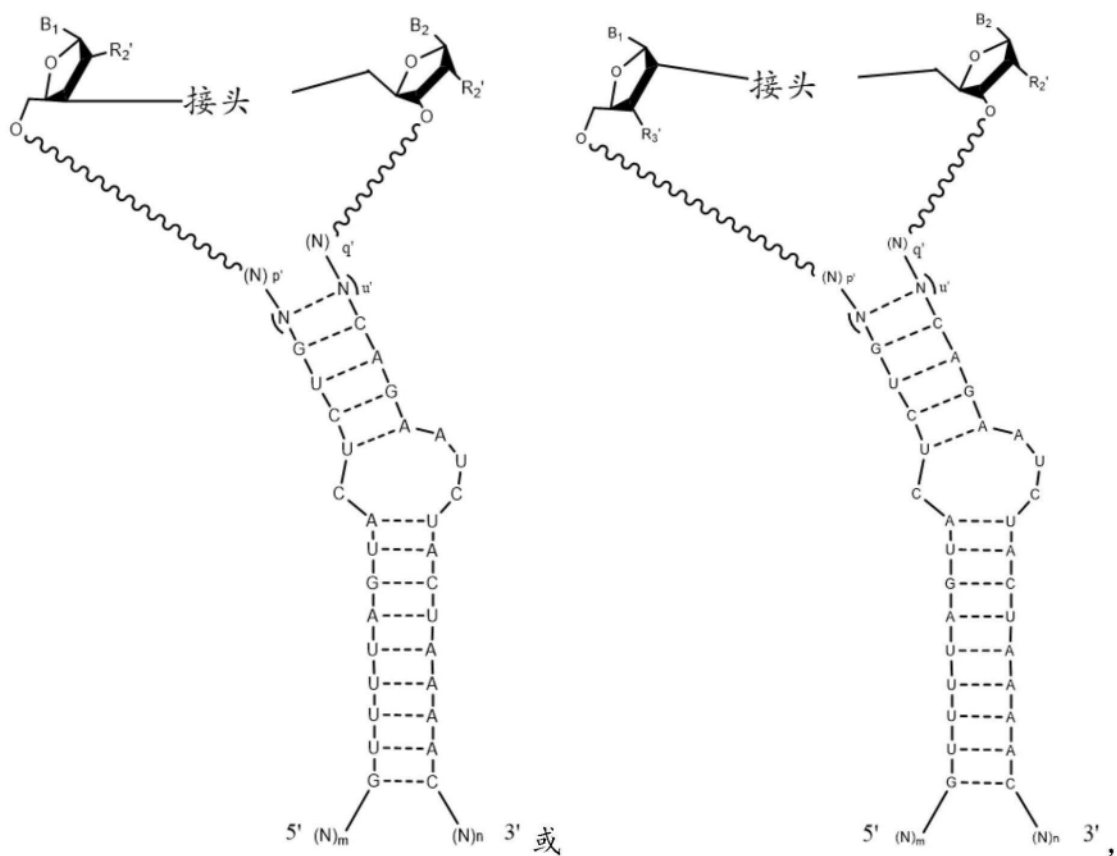


其中：

u'是2到22之间的整数,包括端值;并且

$p'$ 和 $q'$ 各为0。

实施方式195.根据实施方式194所述的指导分子,其中所述指导分子具有下式:



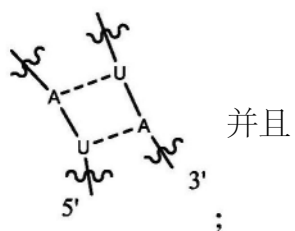
或其协变体,

其中：

u' 是0到19之间的整数,包括端值;并且

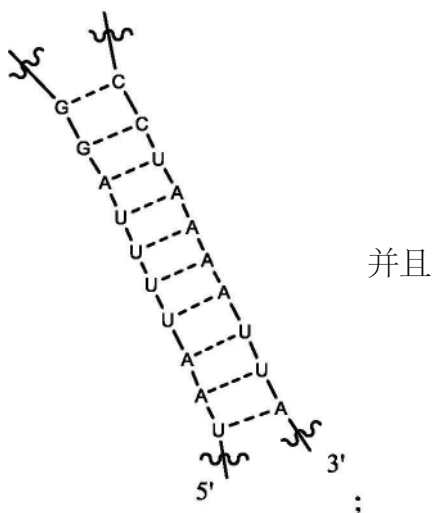
$p'$ 和 $q'$ 各为0。

实施方式196. 根据实施方式195所述的指导分子, 其中  $(N\cdots N)_u$ , 具有下式:



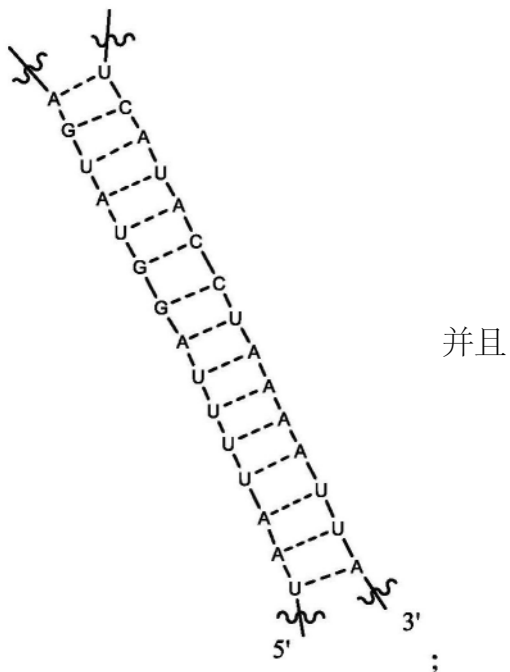
其中B<sub>1</sub>是腺嘌呤残基并且B<sub>2</sub>是尿嘧啶残基;或其协变体。

实施方式197. 根据实施方式195所述的指导分子, 其中  $(N\cdots N)_u$ , 具有下式:



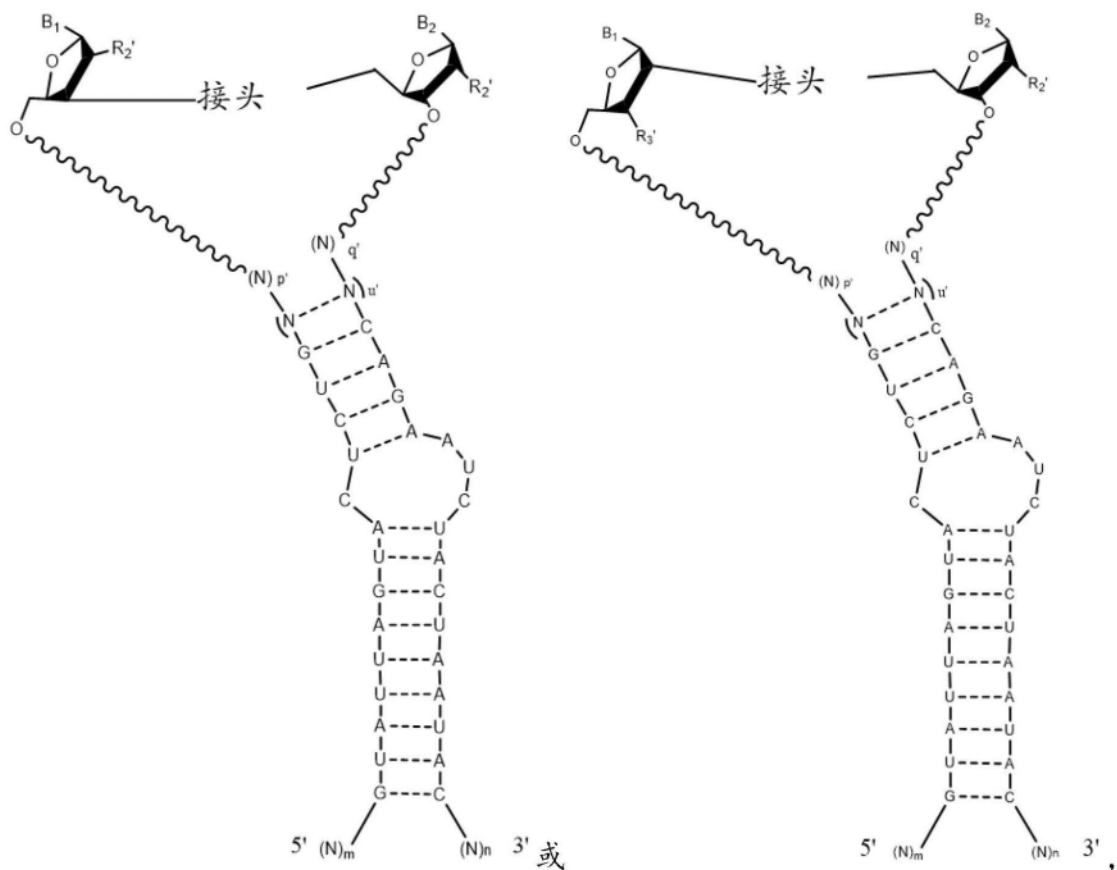
其中B<sub>1</sub>是尿嘧啶残基并且B<sub>2</sub>是腺嘌呤残基;或其协变体。

实施方式198. 根据实施方式195所述的指导分子, 其中 (N---N)<sub>u</sub> 具有下式:



其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式199.根据实施方式169-184中任一项所述的指导分子,其中所述指导分子具有下式:



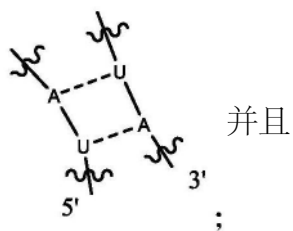
或其协变体,

其中:

$u'$  是0到19之间的整数,包括端值;并且

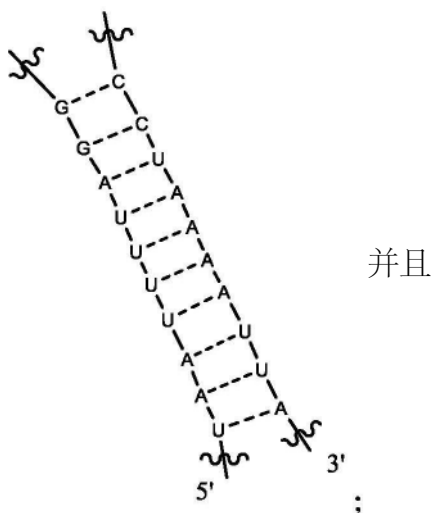
$p'$  和  $q'$  各为0。

实施方式200.根据实施方式199所述的指导分子,其中  $(N \cdots N)_{u'}$  具有下式:



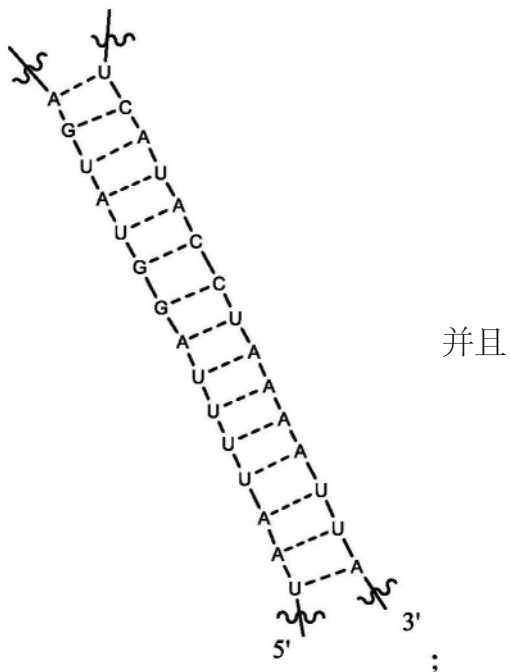
其中  $B_1$  是腺嘌呤残基并且  $B_2$  是尿嘧啶残基;或其协变体。

实施方式201.根据实施方式199所述的指导分子,其中  $(N \cdots N)_{u'}$  具有下式:



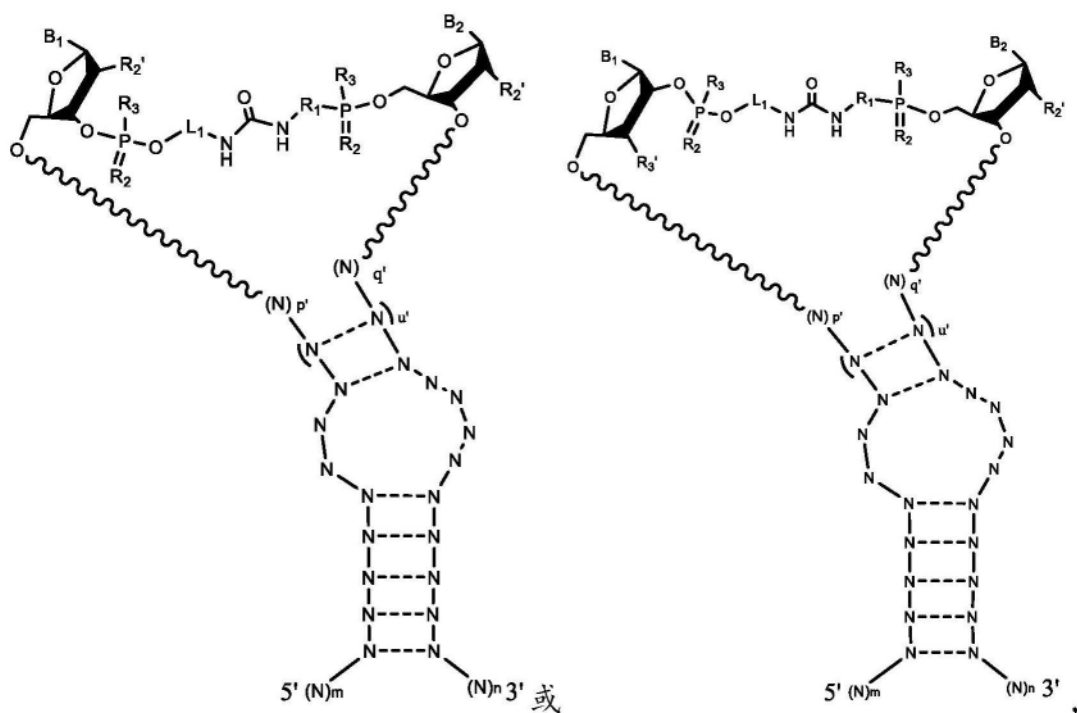
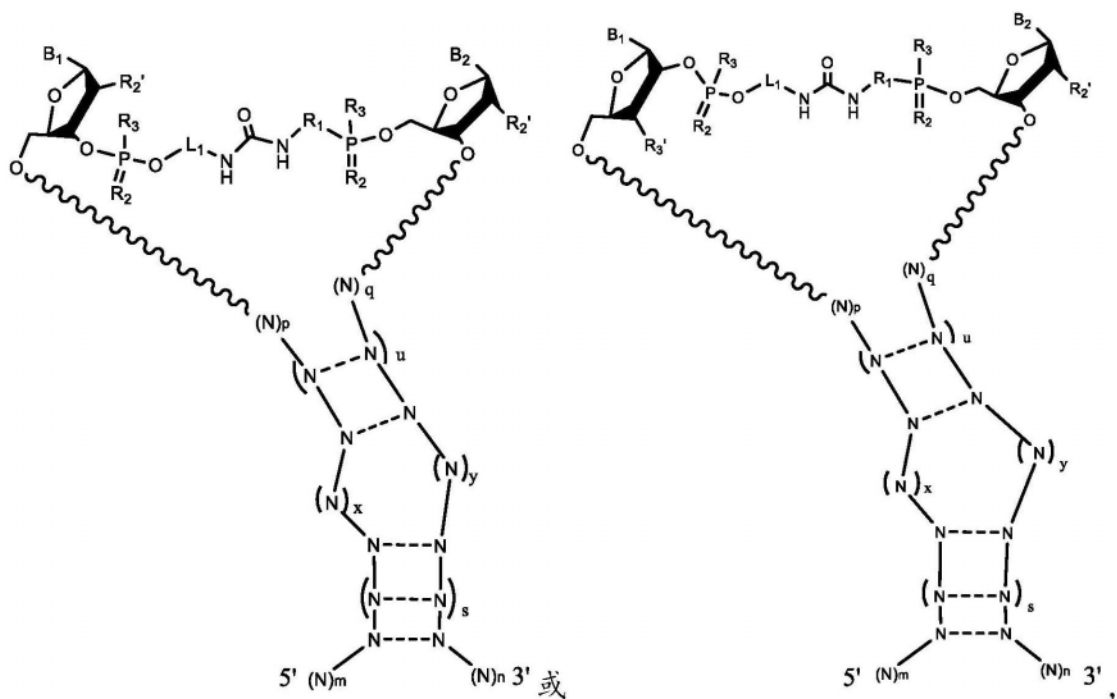
其中B<sub>1</sub>是尿嘧啶残基并且B<sub>2</sub>是腺嘌呤残基;或其协变体。

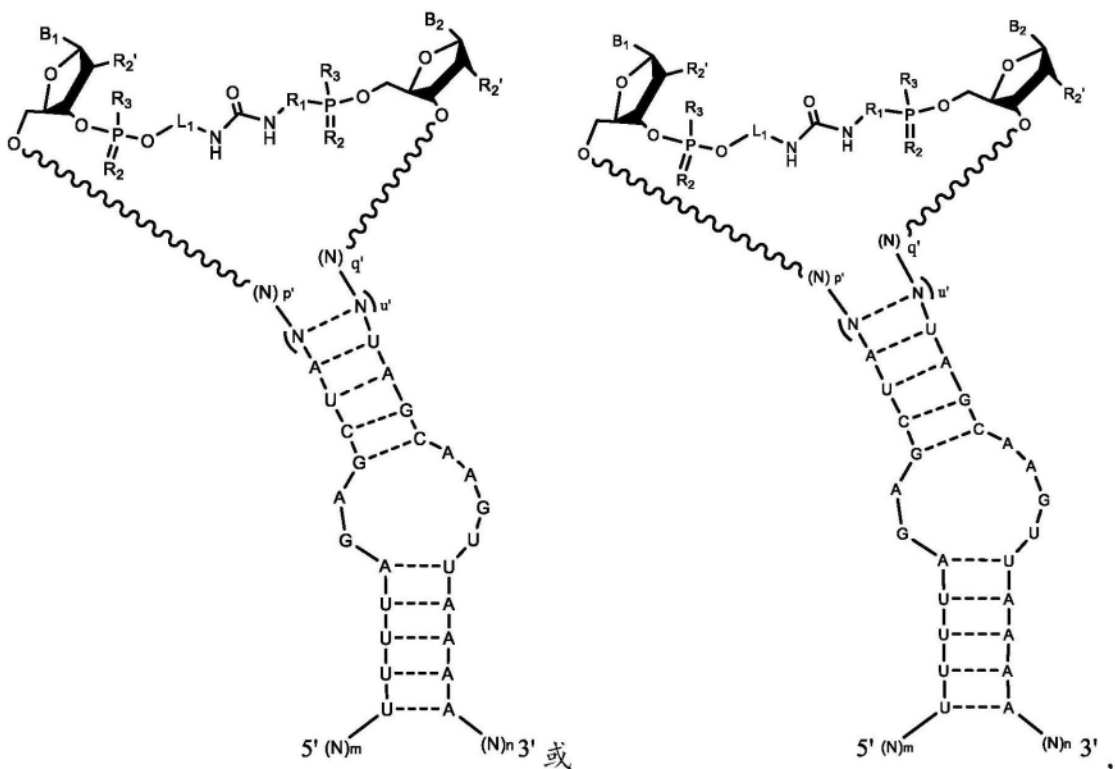
实施方式202. 根据实施方式199所述的指导分子,其中(N---N)<sub>u</sub>具有下式:



其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式203.根据实施方式169-184中任一项所述的指导分子,其中所述指导分子具有下式:



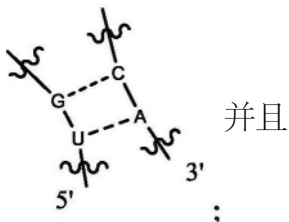


或其协变体,其中:

$p'$ 和 $q'$ 各为0;并且

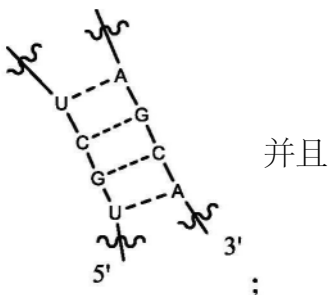
u' 是0到19之间的整数,包括端值。

实施方式206.根据实施方式205所述的指导分子,其中(N---N)<sub>n</sub>,具有下式:



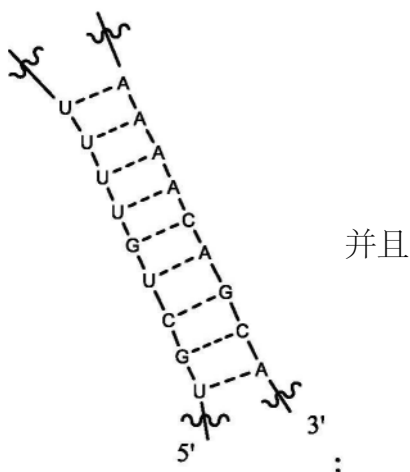
其中B<sub>1</sub>是胞嘧啶残基并且B<sub>2</sub>是鸟嘌呤残基;或其协变体。

实施方式207.根据实施方式205所述的指导分子,其中(N---N)<sub>n</sub>,具有下式:



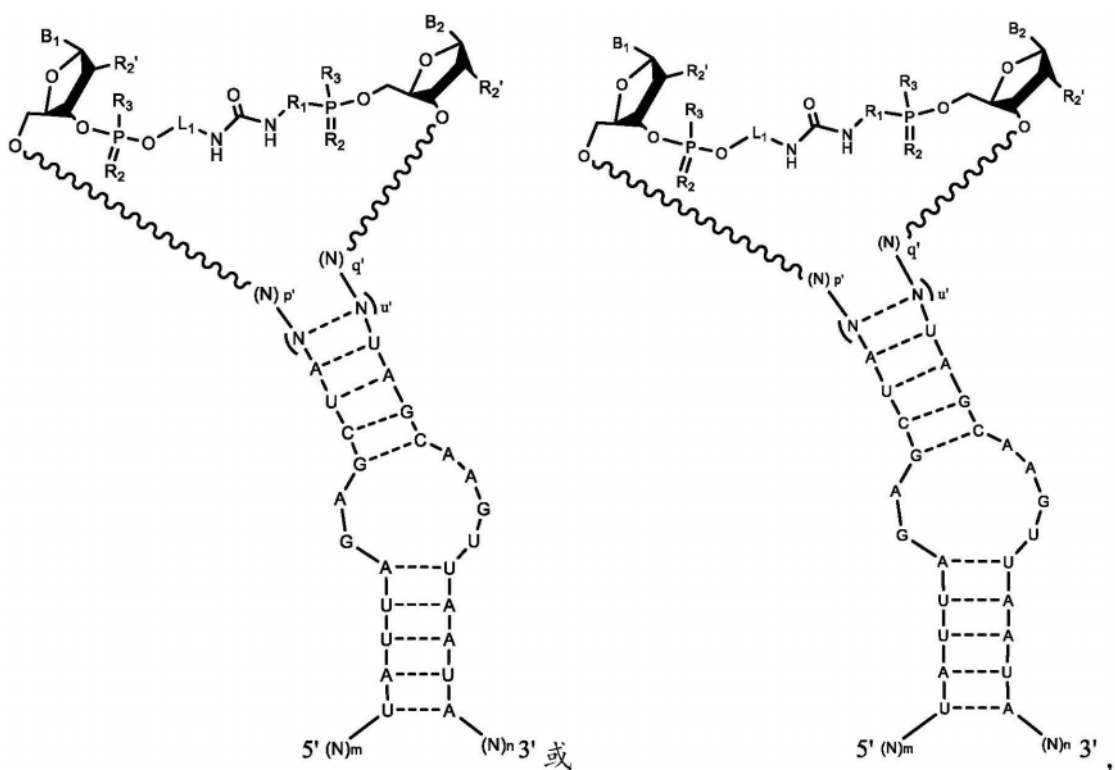
其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式208.根据实施方式205所述的指导分子,其中(N---N)<sub>n</sub>,具有下式:



其中 $B_1$ 是鸟嘌呤残基并且 $B_2$ 是胞嘧啶残基;或其协变体。

实施方式209. 根据实施方式203所述的指导分子, 其中所述指导分子具有下式:

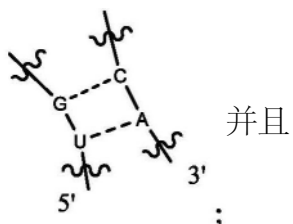


或其协变体, 其中:

$p'$ 和 $q'$ 各为0; 并且

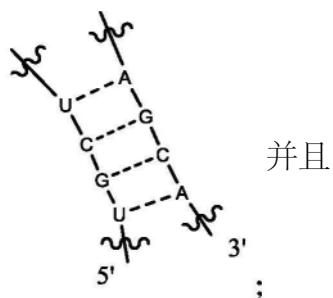
$u'$ 是0到19之间的整数, 包括端值。

实施方式210. 根据实施方式209所述的指导分子, 其中 $(N\cdots N)_{u'}$ 具有下式:



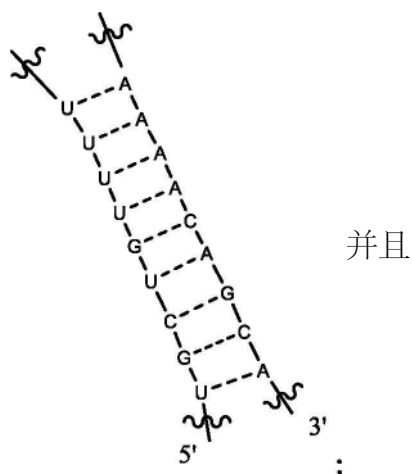
其中 $B_1$ 是胞嘧啶残基并且 $B_2$ 是鸟嘌呤残基;或其协变体。

实施方式211.根据实施方式209所述的指导分子,其中  $(N\cdots N)_U$  具有下式:



其中 $B_1$ 是鸟嘌呤残基并且 $B_2$ 是胞嘧啶残基;或其协变体。

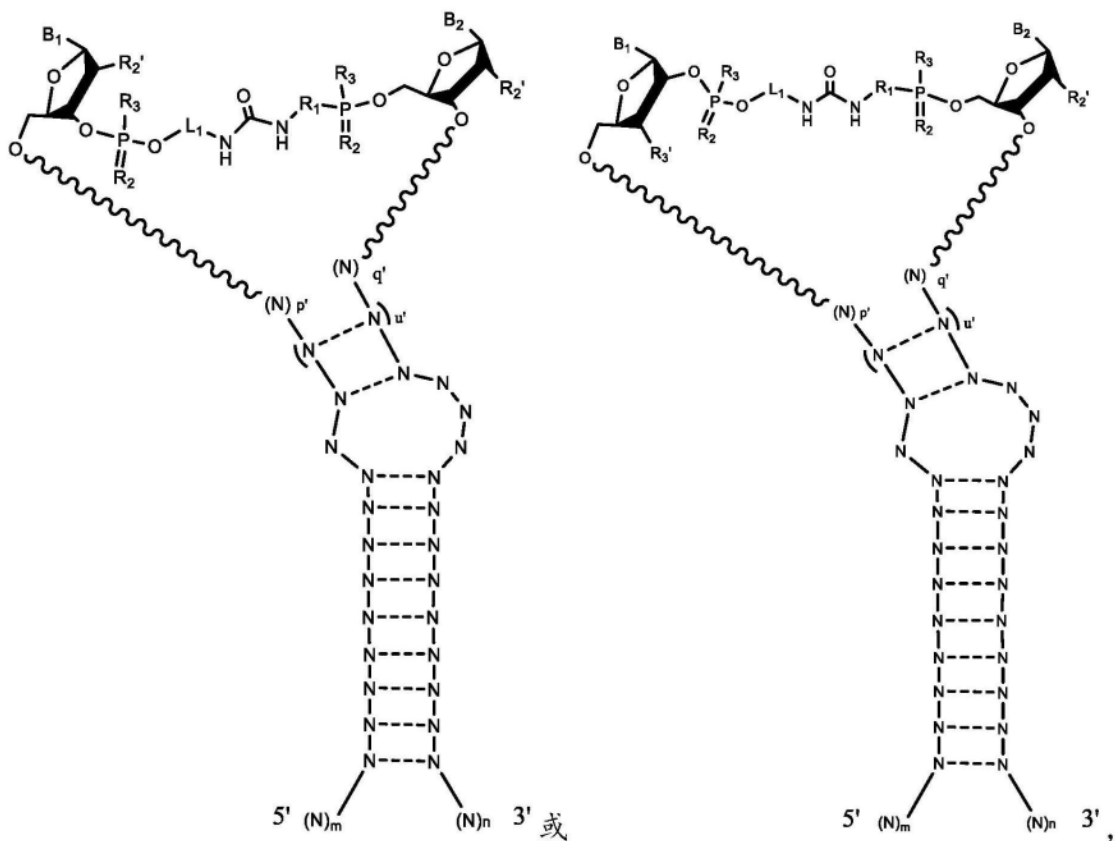
实施方式212.根据实施方式209所述的指导分子,其中  $(N\cdots N)_U$  具有下式:



其中 $B_1$ 是鸟嘌呤残基并且 $B_2$ 是胞嘧啶残基;或其协变体。

实施方式213.根据实施方式203中任一项所述的指导分子,其中所述指导分子具

有式

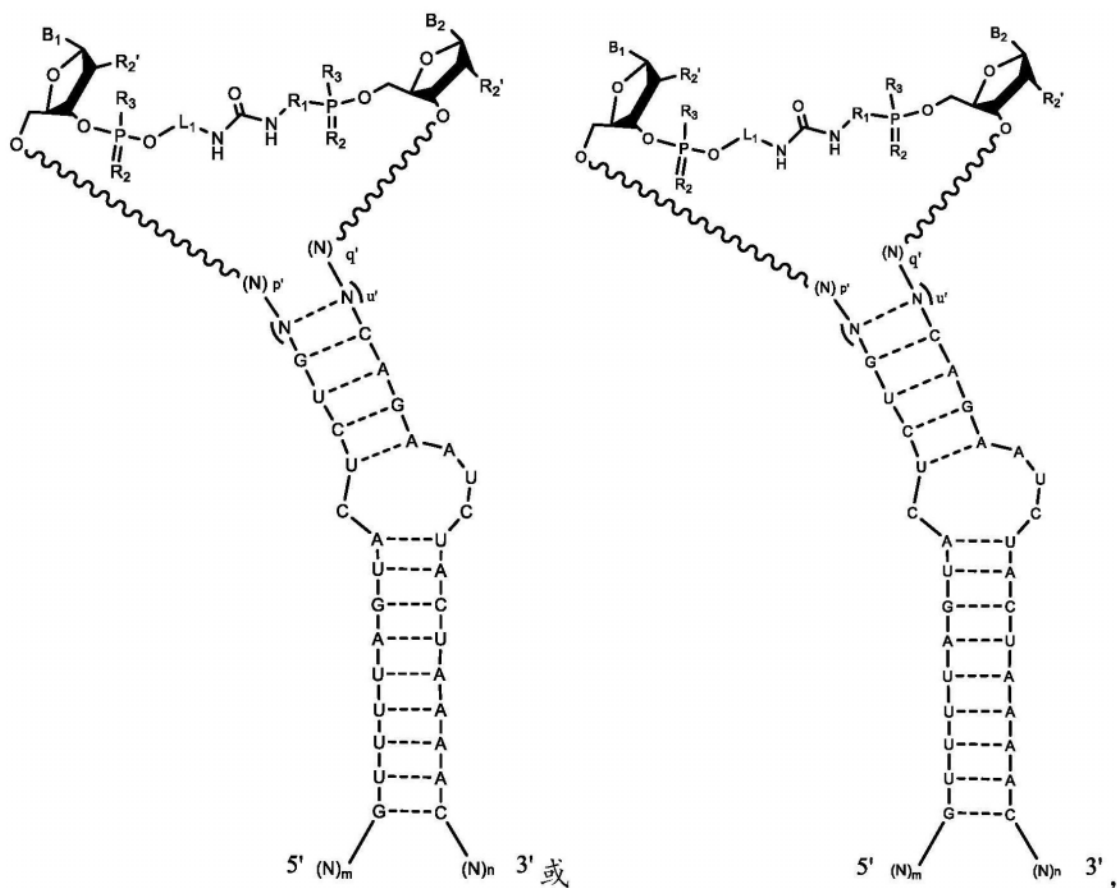


其中：

$u'$  是2到22之间的整数,包括端值;并且

$p'$  和  $q'$  各为0。

实施方式214. 根据实施方式213所述的指导分子,其中所述指导分子具有下式:



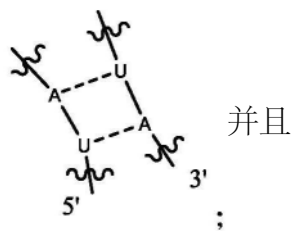
或其协变体,

其中：

u' 是0到19之间的整数,包括端值;并且

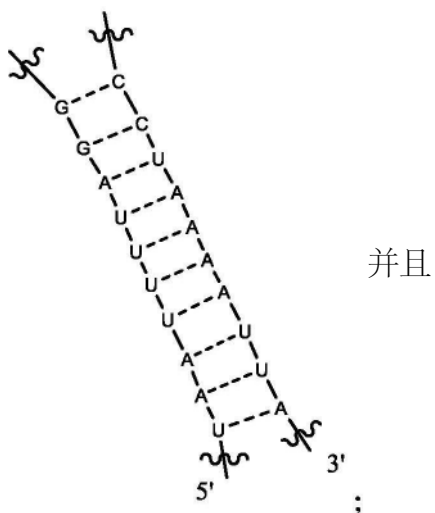
$p'$ 和 $q'$ 各为0。

实施方式215. 根据实施方式214所述的指导分子,其中(N---N)<sub>u</sub>具有下式:



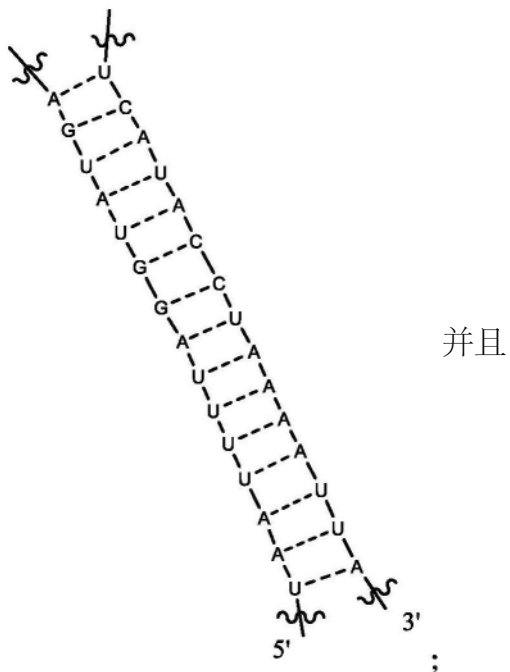
其中B<sub>1</sub>是腺嘌呤残基并且B<sub>2</sub>是尿嘧啶残基;或其协变体。

实施方式216. 根据实施方式214所述的指导分子, 其中  $(N\cdots N)_u$ , 具有下式:



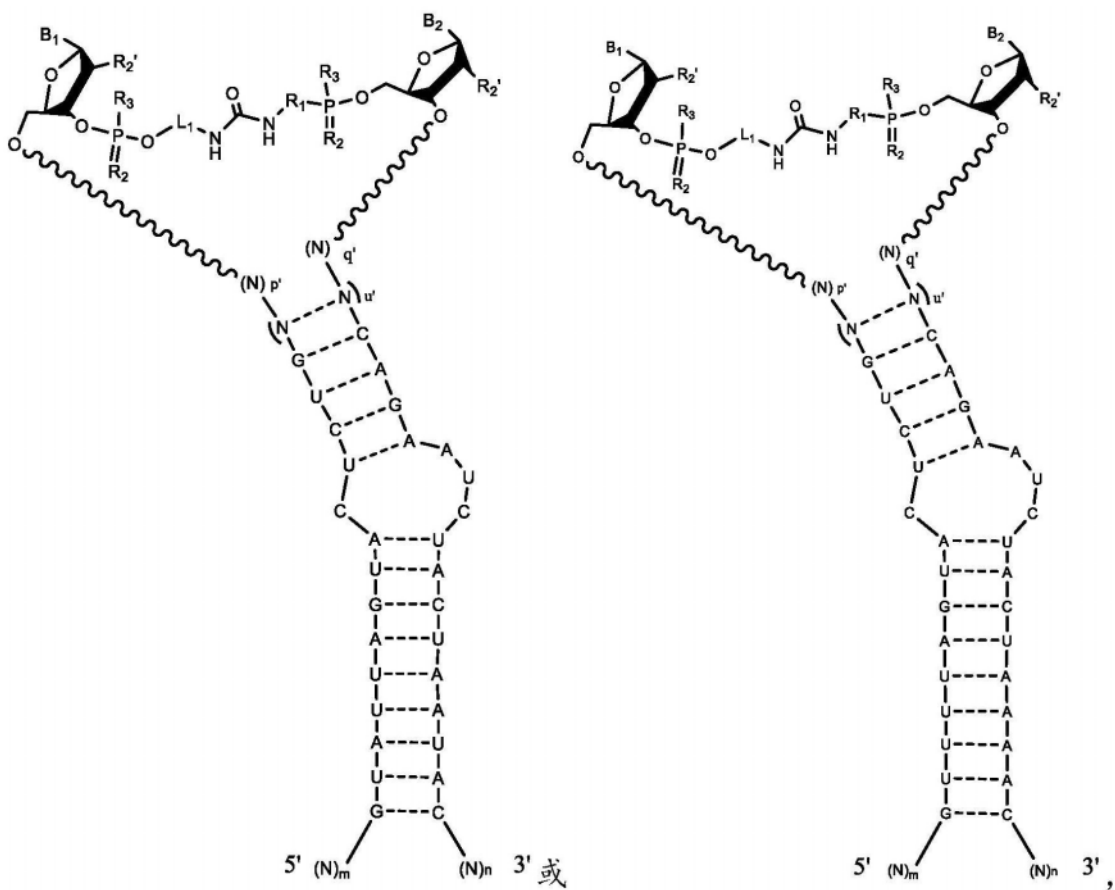
其中 $B_1$ 是尿嘧啶残基并且 $B_2$ 是腺嘌呤残基;或其协变体。

实施方式217. 根据实施方式214所述的指导分子, 其中 $(N\cdots N)_U$ 具有下式:



其中 $B_1$ 是鸟嘌呤残基并且 $B_2$ 是胞嘧啶残基;或其协变体。

实施方式218. 根据实施方式203所述的指导分子, 其中所述指导分子具有下式:



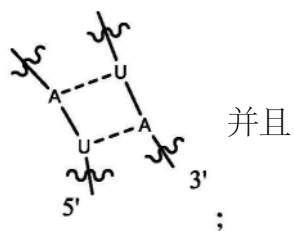
或其协变体,

其中:

u' 是0到19之间的整数,包括端值;并且

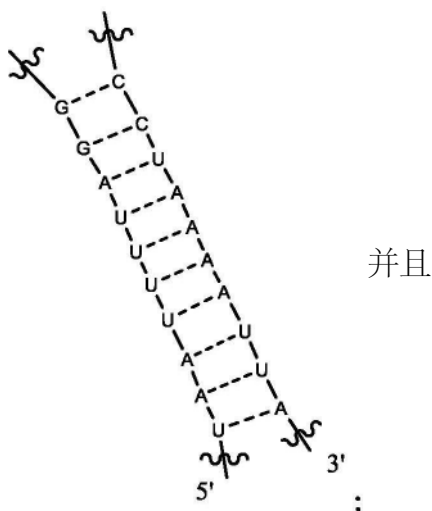
p' 和q' 各为0。

实施方式219. 根据实施方式218所述的指导分子,其中 (N---N)<sub>u'</sub> 具有下式:



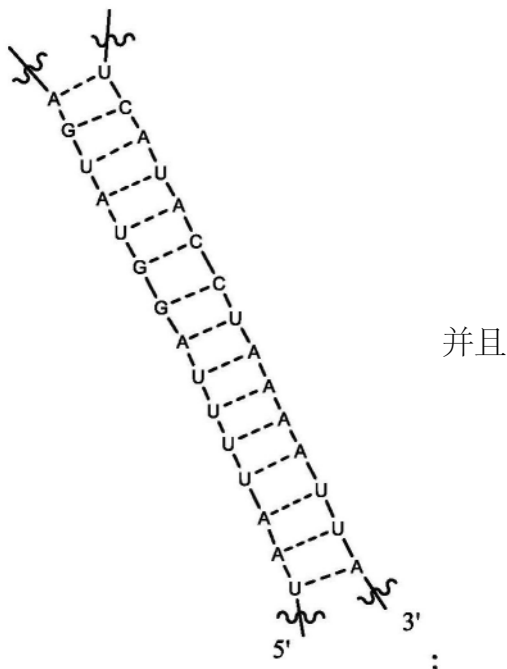
其中B<sub>1</sub>是腺嘌呤残基并且B<sub>2</sub>是尿嘧啶残基;或其协变体。

实施方式220. 根据实施方式218所述的指导分子,其中 (N---N)<sub>u'</sub> 具有下式:



其中B<sub>1</sub>是尿嘧啶残基并且B<sub>2</sub>是腺嘌呤残基;或其协变体。

实施方式221. 根据实施方式218所述的指导分子,其中(N---N)<sub>n</sub>,具有下式:



其中B<sub>1</sub>是鸟嘌呤残基并且B<sub>2</sub>是胞嘧啶残基;或其协变体。

实施方式222.一种组合物,其包含多个根据实施方式169-221中任一项所述的合成的指导分子,其中相对于参照指导分子序列,少于约10%的所述指导分子在5'末端包含截短。

实施方式223.根据实施方式222所述的组合物,其中至少约99%的所述指导分子包含以下5'序列,所述5'序列包含所述指导分子的与参照指导分子序列的相应5'序列100%相同的核苷酸1-20。

实施方式224.一种指导分子,其从5'至3'包含:

第一指导分子片段,其包含:

靶向结构域序列;

第一下茎序列;

第一凸起序列;  
第一上茎序列;  
非核苷酸化学键合;和  
第二指导分子片段,其包含:  
第二上茎序列;  
第二凸起序列;和  
第二下茎序列,

其中(a)所述第一下茎序列中的至少一个核苷酸与所述第二下茎序列中的核苷酸碱基配对,和(b)所述第一上茎序列中的至少一个核苷酸与所述第二上茎序列中的核苷酸碱基配对。

实施方式225.根据实施方式224所述的指导分子,其中(c)所述指导分子不包含所述第一和第二上茎序列之间的四环序列。

实施方式226.根据实施方式224或225所述的指导分子,其中所述第一和/或第二上茎序列包含数目从4至22的核苷酸,包括端值。

实施方式227.根据实施方式224-226中任一项所述的指导分子,其特征在于用于在所述第一和第二指导分子片段之间形成双链体的吉布斯自由能( $\Delta G$ )小于用于在两个第一指导分子片段之间形成双链体的 $\Delta G$ 。

实施方式228.根据实施方式227所述的指导分子,其中用于在所述第一和第二指导分子片段之间形成双链体的 $\Delta G$ 的特征在于(i)所述第一和第二上茎序列之间和(ii)所述第一和第二下茎序列之间各自的大于50%、60%、70%、80%、90%或95%碱基配对小于形成以下双链体的 $\Delta G$ ,所述双链体的特征在于(i)和(ii)之间少于50%、60%、70%、80%、90%或95%碱基配对。

实施方式229.根据实施方式224或225所述的指导分子,其中所述非核苷酸化学键合将所述第一上茎序列的3'末端处或其附近的第一核苷酸与所述第二上茎序列的5'末端处或其附近的第二核苷酸共价连接。

实施方式230.根据实施方式229所述的指导分子,其中所述第一和第二核苷酸碱基配对的。

实施方式231.根据实施方式229所述的指导分子,其中用于在所述第一和第二指导分子片段之间形成包含所述第一和第二核苷酸碱基配对的双链体的吉布斯自由能( $\Delta G$ )小于用于在所述第一和第二指导分子片段之间形成其中所述第一和第二核苷酸不碱基配对的双链体的 $\Delta G$ 。

实施方式232.根据实施方式224-231中任一项所述的指导分子,其中所述非核苷酸化学键包含脲。

实施方式233.一种组合物,其包含根据实施方式224-232中任一项所述的指导分子。

实施方式234.根据实施方式233所述的组合物,其特征在于所述组合物中多于90%的指导分子是全长指导分子。

实施方式235.根据实施方式233或234所述的组合物,其特征在于所述组合物中多于85%的指导分子包含相同的靶向结构域序列。

实施方式236.根据实施方式233-235中任一项所述的组合物,其中所述靶向结构域序列是预定的靶向序列。

实施方式237.根据实施方式233-236中任一项所述的组合物,其还包含Cas9蛋白,其中所述指导分子和所述Cas9蛋白形成能够与靶核酸相互作用的复合物,所述靶核酸包含(x)与所述靶向结构域序列互补的序列,和(y)被所述Cas9蛋白识别的原型间隔子邻近基序(PAM)序列。

实施方式238.根据实施方式237所述的组合物,其中所述复合物在所述靶核酸中形成单链或双链断裂。

实施方式239.根据实施方式237所述的组合物,其中所述复合物化学地修饰所述靶核酸或与所述靶核酸相关的蛋白质。

实施方式240.一种基因组编辑系统,其包含根据实施方式224-232中任一项所述的指导分子。

实施方式241.根据实施方式224-232中任一项所述的指导分子,根据实施方式233-239中任一项所述的组合物或根据实施方式240所述的基因组编辑系统,用于治疗。

实施方式242.根据实施方式224-232中任一项所述的指导分子,根据实施方式233-239中任一项所述的组合物或根据实施方式240所述的基因组编辑系统,用于制备药物。

实施方式243.根据实施方式224-232中任一项所述的指导分子,根据实施方式233-239中任一项所述的组合物或根据实施方式240所述的基因组编辑系统,用于修饰受试者的离体细胞。

实施方式244.根据实施方式224-232中任一项所述的指导分子,根据实施方式233-239中任一项所述的组合物或根据实施方式240所述的基因组编辑系统,用于修饰受试者的体内细胞。

实施方式245.一种改变细胞或受试者中的核酸的方法,所述方法包括向所述受试者施用根据实施方式1-42、169-221或224-232所述的指导分子,根据实施方式43-85、137-139、147-155、161-165、222、223或233-239所述的组合物。

实施方式246.一种组合物,其基本上由多个根据实施方式1-42、169-221或224-232中任一项所述的合成的分子指导分子组成。

实施方式247.一种组合物,其基本上由多个通过根据实施方式86-137中任一项所述的方法产生的指导分子和药学上可接受的载体组成。

实施方式248.根据实施方式86-137中任一项所述的方法,其中所述指导分子可以作为作用于RNA的酶的底物。

实施方式249.根据实施方式248所述的方法,其中所述酶是逆转录酶。

#### 实例

[0322] 通过以下非限制性实施例说明本披露的某些原理。

#### 实例1:胺官能化的指导分子片段与二琥珀酰亚胺基碳酸酯缀合的示例性方法

[0323] 如图1A所示,合成第一5'指导分子片段(例如,34mer),其在3'末端具有(C<sub>6</sub>)<sub>2</sub>-NH<sub>2</sub>接头,并且合成第二3'指导分子片段(例如66mer),其在5'末端具有TEG-NH<sub>2</sub>接头。将两种指导分子片段以1:1的摩尔比在包含10mM硼酸钠、150mM NaCl和5mM MgCl<sub>2</sub>的pH 8.5缓冲液中混

合。得到的指导分子浓度为约50至100 $\mu$ M。将两个指导分子片段退火,然后添加在DMF (2.5mM终浓度)中的二琥珀酰亚胺基碳酸酯(DSC)。将反应混合物短暂涡旋,然后在室温下混合1小时,然后除去过量的二琥珀酰亚胺基碳酸酯,并进行阴离子交换HPLC纯化。

#### 实例2:将硫醇官能化的指导分子片段与溴乙酰基官能化的指导分子片段缀合的示例性方法

[0324] 如图2A所示,合成第一5'指导分子片段(例如34mer),其在3'末端具有(C<sub>6</sub>)<sub>2</sub>-NH<sub>2</sub>接头。将其悬浮于pH 8.5的100mM硼酸盐缓冲液中。指导分子浓度为约100 $\mu$ M至1mM。将0.2体积的DMSO中的琥珀酰亚胺基-3-(溴乙酰胺基)丙酸酯(SBAP) (50当量)添加到指导分子溶液中。在室温下混合30分钟后,添加10体积的pH 7.0的100mM磷酸盐缓冲液。将混合物在10,000MW Amicon上浓缩10X或更多。通过(a)添加10体积的水,和(b)在10,000MW Amicon上浓缩10X或更多,进一步处理混合物。步骤(a)和(b)重复3次,得到第一5'指导分子片段(例如34mer),其在3'末端具有溴乙酰基部分。

[0325] 如图2B所示,合成第二3'指导分子片段(例如66mer),其在5'末端具有TEG-NH<sub>2</sub>接头。将其悬浮在含有1mM EDTA的pH 8.5的100mM硼酸盐缓冲液中。指导分子浓度为约100 $\mu$ M至1mM。将0.2体积的DMSO中的琥珀酰亚胺基-3-(2-吡啶基二硫代)丙酸酯(SPDP) (50当量)添加到指导分子溶液中。在室温下混合1小时后,添加在1 x PBS中的1M二硫苏糖醇(DTT)。混合物中DTT的最终浓度为20mM。在室温下混合30分钟后,添加5M NaCl,使混合物中终浓度为0.3M NaCl,然后添加3体积乙醇。混合物被进一步加工:(a)冷却至-20°C 15分钟;(b)以17,000g (优选4°C)离心5分钟;(c)除去上清液;(d)将残余物悬浮在0.3M NaCl中(用氩气喷射);以及(e)添加3倍体积的乙醇。步骤(a)-(e)重复3次。得到的沉淀(即在5'末端具有硫醇的第二3'指导分子片段)在真空下干燥。

[0326] 如图2C所示,将在5'末端具有硫醇的第二3'指导分子片段(例如66mer)悬浮于pH 8的100mM磷酸盐缓冲液中,所述缓冲液包含2mM EDTA(用氩气喷射)。指导分子浓度为约100 $\mu$ M至1mM。将在3'末端具有溴乙酰基部分的第一5'指导分子片段(例如34mer)悬浮在水中(相对于第二3'指导分子片段混合物的体积约0.1体积)。指导分子浓度为约100 $\mu$ M至1mM。将第一5'指导分子片段混合物加入第二3'指导分子片段混合物中(用氩气喷射)。将反应混合物在室温下混合过夜,然后进行阴离子交换HPLC纯化。

#### 实例3:用碳二亚胺使指导分子片段磷酸酯与3'羟基指导分子片段缀合的示例性方法

[0327] 如图3A和3B所示,使用标准亚磷酰胺化学合成第一5'指导分子片段(例如34mer)。还合成了包含5'-磷酸的第二3'指导分子片段(例如,66mer)。第一和第二指导分子片段以1:1的摩尔比在偶联缓冲液(100mM 2-(N-吗啉代)乙磺酸(MES) (pH 6)、150mM NaCl、5mM MgCl<sub>2</sub>和10mM ZnCl<sub>2</sub>)中混合。将两个指导分子片段退火,然后添加100mM 1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)和90mM咪唑。将反应混合物在4°C下混合1-5天,然后进行脱盐和阴离子交换HPLC纯化。

#### 实例4:评估指导分子在HEK293T细胞中的活性

[0328] 通过T7E1切割测定法在HEK293T细胞中评估根据实例2的方法缀合的指导分子的活性。为清楚起见,本实例中使用的所有指导分子含有相同的靶向结构域序列和基本相似的RNA骨架序列,如下表9所示。在该表中,靶向结构域序列由“N”表示为简并序列,而两个指

导分子片段之间的交联位置由[L]表示。

表9

指导分子或指导分子片段	SEQ ID NO.	序列
100mer gRNA	32	NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAUA GCAAGUUAUUUUUAAGGCUAGUCCGUUAUCAACUUGAAA AAGUGGCACCGAGUCGGUGCUUUU
34mer 5'gRNA 片段	33	NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGA
66mer 3' gRNA 片段	34	AAUAGCAAGUUAUUUUUAAGGCUAGUCCGUUAUCAACUU GAAAAAGUGGCACCGAGUCGGUGCUUUU
100mer 缀合的 gRNA	35	NNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGA [L]A AUAGCAAGUUAUUUUUAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCUUUU

[0329] 不同浓度的核糖核蛋白复合物通过脂质转染 (CRISPR-Max, 赛默飞科学公司 (Thermo Fisher Scientific, 沃尔瑟姆, 麻萨诸塞州) 被引入HEK293T细胞并且随后收获基因组DNA, 所述核糖核蛋白复合物包含由IVT产生的单分子指导分子、合成的单分子指导分子 (即, 未缀合地制备), 或通过实例2的溴乙酰基-硫醇方法缀合的合成的单分子指导分子。使用标准T7E1切割测定法, 使用商业试剂盒 (Surveyor™, 可从整合DNA系统公司 (Integrated DNA Systems), 科勒尔维尔 (Coralville), 爱荷华州购获得) 评估切割。结果如图4所示。

[0330] 结果显示, 缀合的指导分子以剂量依赖性方式支持HEK293细胞中的切割, 这与用IVT产生的单分子指导分子或合成的单分子指导分子观察到的一致。应该注意的是, 未缀合的退火的指导分子片段支持较低水平的切割, 但是以类似的剂量依赖性方式。这些结果表明, 根据本披露的方法缀合的指导分子以与用IVT产生的单分子指导分子或合成的单分子指导分子基本相同的方式支持高水平的DNA切割。

实例5:通过凝胶电泳和质谱法评估指导分子纯度

[0331] 通过总离子电流色谱法和质谱法将根据实例1的方法与脲接头缀合的指导分子的组合物的纯度与商业上制备 (即, 没有缀合地制备) 的合成的单分子指导分子的组合物的纯度进行比较。注入100pmol的分析物用于质量分析。通过LC-MS在配备有Waters ACQUITY UPLC系统的Bruker microTOF-QII质谱仪上实现分析。使用ThermoDNAPac C18柱进行分离。结果如图5所示。

[0332] 图5A显示代表性离子色谱, 图5B显示根据实例1的方法与脲接头缀合的离子交换纯化的指导分子的解卷积质谱。图5C显示代表性离子色谱, 图5D显示商业制备的合成的单分子指导分子的解卷积质谱。评估离子色谱中突出显示的峰的质谱。图5E显示了质谱的扩展版本。商业上制备的合成的单分子指导分子的质谱在左侧 (按总质量计, 34%纯度), 而根据实例1的方法与脲接头缀合的指导分子的质谱在右侧 (按总质量计, 72%纯度)。

实例6:通过序列分析评估指导分子纯度

[0333] 将如实例1中所述的用脲键合缀合的指导分子的组合物的纯度与商业制备 (即, 未缀合地制备) 的合成的单分子指导分子的组合物的纯度和如实例2中所述的用硫醚键合缀

合的指导分子的组合物的纯度进行比较。指导分子的所有组合物都基于相同的预定指导分子序列。

[0334] 图6A显示的图描绘从包含脲键合的合成的单分子指导分子产生的互补DNA (cDNA) 的5'末端的每个位置处发生个体碱基和长度变化的频率,并且图6B显示的图描绘从商业制备的合成的单分子指导分子(即没有缀合地制备)产生的cDNA的5'末端的每个位置处发生个体碱基和长度变化的频率。盒围绕指导分子的20bp靶向结构域。在该实施例中,与来自商业制备的合成的单分子指导分子的指导分子相比(其中少于10%的指导分子包括在任何给定位置的缺失,并且少于5%包括在任何给定位置的取代),包含脲键合的指导分子在靶向结构域中导致更高的序列保真度(即,少于1%的指导分子包括在任何给定位置的缺失,并且少于1%的指导分子包括在任何给定位置的取代)。

[0335] 图6C显示的图描绘从包含硫醚键合的合成的单分子指导分子产生的cDNA的5'末端的每个位置处发生个体碱基和长度变化的频率。如图6C所示,观察到高水平的5'序列保真度,证明了具有高水平序列保真度和纯度的指导分子的组合物的产生。图6A(脲键合)和图6C(硫醚键合)中的比对也显示在键合位点(位置34)具有相对高频率的错配/插入缺失的区。这些数据表明,与市售指导分子相比,通过本披露的方法合成的指导分子表现出降低的缺失和取代频率。

[0336] 图7A和7B描绘从包含脲键合的合成的单分子指导分子(图7A)产生的cDNA和从商业制备(即,没有缀合地制备)的合成的单分子指导分子(图7B)的5'末端的前41个位置处的内部序列长度变化(+5至-5)。如图所示,相对于商业制备的合成的单分子指导分子(即,没有缀合地制备),包含脲键合的指导分子具有插入/缺失的频率和长度的减少。

#### 实例7:评估指导分子在CD34<sup>+</sup>细胞中的活性。

[0337] 通过下一代测序技术在CD34<sup>+</sup>细胞中评估根据实例1的方法缀合有脲键合的指导分子的活性。本实例中讨论的指导分子含有三种靶向结构域序列之一和各种指导分子骨架序列,如下表10和图8A-H、9A-D和10A-D所示。两个指导分子片段之间的脲键合的位置在表10和图8A-H、9A-D和10A-D中用[UR]表示。具有前两个靶向结构域序列的指导分子(表示为gRNA 1后跟字母或gRNA 2后跟字母)基于化脓链球菌gRNA骨架,而具有第三靶向结构域序列的指导分子(表示为gRNA 3后跟字母)基于金黄色葡萄球菌gRNA骨架。

[0338] 将缀合的指导分子重悬于pH 7.5缓冲液中,熔化并再退火,然后加入化脓链球菌Cas9悬浮液中产生55 $\mu$ M完全复合的核糖核蛋白的溶液。

[0339] 计数人CD34<sup>+</sup>细胞,离心沉淀并重悬于P3 Nucleofection缓冲液中,然后分配到预先填充有人HSC培养基(StemSpan<sup>TM</sup>无血清扩增培养基,干细胞技术公司(StemCell Technologies),温哥华,不列颠哥伦比亚省,加拿大)的96孔Nucleocuvette板的每个孔中以产生50,000个细胞/孔。将如上所述的完全复合的核糖核蛋白溶液加入Nucleocuvette板中的每个孔中,然后温和混合。在Amaxa Nucleofector System(龙沙公司(Lonza),巴塞尔(Basel),瑞士)上进行核转染。将经核转染的细胞在37°C和5% CO<sub>2</sub>下孵育72小时,以允许编辑至平台期。然后根据制造商的说明使用DNAadvance DNA分离试剂盒从核转染的细胞中提取基因组DNA。使用下一代测序技术评估切割以相对于野生型人参照序列定量%插入和缺失(indel)。表10中在CD34<sup>+</sup>细胞中测试的gRNA的结果显示在图11中。

[0340] 如图11中的结果所示,根据实例1产生的连接的指导分子支持CD34<sup>+</sup>细胞中的DNA

切割。发现%indel随着茎环长度的增加而增加,但是在茎环序列附近掺入U-A交换(参见gRNA 1E,gRNA 1F和gRNA 2D)减轻了所述作用。这些数据表明化学缀合的合成的单分子指导分子具有较长的茎环特征导致细胞中更高水平的DNA切割。此外,DNA切割活性与连接效率无关,必须凭经验确定。

表10

指导分子	SEQ ID NO.	序列
gRNA 1A	36	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAGA [UR] AAUAGCAAGUAAAAUAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 1B	37	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUA [UR] UAGCAAGUAAAAUAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 1C	38	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAGG [UR] CCUAGCAAGUAAAAUAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 1D	39	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAUGC [UR] GCAUAGCAAGUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 1E	40	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAUGCUGUUUUG [UR] CAAAACAGCA UAGCAAGUAAUAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCG GUGCUUUU
gRNA 1F	41	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAUGCUG [UR] CAGCAUAGCAAGUUA AUAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
gRNA 1G	42	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAUGCUGUUUUG [UR] CAAAACAGCA UAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCG GUGCUUUU

<b>gRNA 1H</b>	43	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAUGCUG [UR] CAGCAUAGCAAGUUA AAAAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 1I</b>	44	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAAAAGA [UR] AAUUUAGCAAGUUA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 1J</b>	45	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAAA [UR] UUUAGCAAGUUA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 1K</b>	46	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAAAAGGGA [UR] AACUUUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 1L</b>	47	GUAACGGCAGACUUCUCCUCGUUUUAGAGCUAGdA [UR] AAUAGCAAGUUA AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 2A</b>	48	CUAACAGUUGCUUUUAUCACGUUUUAGAGCUAUGC [UR] GCAUAGCAAGUUA AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 2B</b>	49	CUAACAGUUGCUUUUAUCACGUUUUAGAGCUAUGCUG [UR] CAGCAUAGCAAGUUA AAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 2C</b>	50	CUAACAGUUGCUUUUAUCACGUUUUAGAGCUAUGCUGUUUUG [UR] CAAAACAGCA UAGCAAGUUA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCG GUGCUUUU
<b>gRNA 2D</b>	51	CUAACAGUUGCUUUUAUCACGUUUUAGAGCUAUGCUGUUUUG [UR] CAAAACAGCA UAGCAAGUUA AUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCG GUGCUUUU
<b>gRNA 2E</b>	52	CUAACAGUUGCUUUUAUCACGUUUUAGAGCUAGA [UR] AAUAGCAAGUUA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
<b>gRNA 3A</b>	53	GUAACGGCAGACUUCUCCUCGUUUUAGUACUCUG [UR] CAGAAUCUACUAAAA GGCAAAUAGCCGUGUUUAUCUCGUCACUUGUUGGCGAGAUUUU
<b>gRNA 3B</b>	54	GUAACGGCAGACUUCUCCUCGUUUUAGUACUCUGUAA [UR] UUAACAGAAUCUACU AAACAAGGCAAAUAGCCGUGUUUAUCUCGUCACUUGUUGGCGAGAUUUU
<b>gRNA 3C</b>	55	GUAACGGCAGACUUCUCCUCGUUUUAGUACUCUGUAAUUUUAGGU [UR] ACCUAAA AUUACAGAAUCUACUAAAA CAAGGCAAAUAGCCGUGUUUAUCUCGUCACUUGUUG GCGAGAUUUU
<b>gRNA 3D</b>	56	GUAACGGCAGACUUCUCCUCGUUUUAGUACUCUGUAAUUUUAGGUUAGAG [UR] CU CAUACCUAAAAUUAACAGAAUCUACUAAAA CAAGGCAAAUAGCCGUGUUUAUCUCGU CAACUUGUUGGCGAGAUUUU

#### 实例8:连接效率计算模型的评估

[0341] 实例1中描述的反应的连接效率是特定指导分子结构的适合性的一种度量。由于实例1中第一和第二指导分子片段的反应性官能团是相同的(胺),竞争性同型偶联是潜在的副产物。使用位于<http://www.idtdna.com/calc/analyzer>的OligoAnalyzer 3.1工具,该实施例评估是否可以通过同型偶联反应( $\Delta G_1$ )的自由能差相比于异型偶联反应( $\Delta G_2$ )的自由能差的计算模型预测预测连接效率(即,反应产物中异型偶联产物的%)。该分析的结果如表11所示。

表11

指导分子	SEQ ID NO.	连接效率	$\Delta G_1$ (kcal/mol)	$\Delta G_2$ (kcal/mol)	$\Delta G_2 - \Delta G_1$ (kcal/mol)
gRNA 1A	36	~55%	-6.90	-10.93	-4.03
gRNA 1C	38	18%	-6.90	-10.93	-4.03
gRNA 1D	39	50%	-6.90	-12.27	-5.37
gRNA 1E	40	50%	-6.34	-24.95	-18.61
gRNA 1F	41	31%	-6.34	-15.82	-9.48
gRNA 1G	42	12%	-6.90	-24.95	-18.05
gRNA 1H	43	60%	-6.90	-15.82	-8.92
gRNA 1I	44	~50%	-6.90	-10.93	-4.03
gRNA 1J	45	~50%	-6.90	-10.93	-4.03
gRNA 1K	46	~55%	-6.90	-10.93	-4.03
gRNA 2A	48	18%	-6.34	-12.27	-5.93
gRNA 2C	50	48%	-6.84	-24.95	-18.11
gRNA 2D	51	45%	-6.84	-24.95	-18.11
gRNA 2E	52	5%	-6.34	-8.64	-2.30

[0342] 如表11中所示,对于大多数具有对应于更有利的连接效率的更负的  $\Delta G_2 - \Delta G_1$  值的序列(例如,比较gRNA 2A和2C),很好地预测了连接效率(如通过凝胶分析后的密度测定法测量)。然而,形成某些指导分子的连接效率并不总是与  $\Delta G_2 - \Delta G_1$  值相关(例如,参见gRNA 1G,其中更负的  $\Delta G_2 - \Delta G_1$  值不会导致更高的连接效率),这表明修饰和实验可能是结合某些指导分子片段所需的。例如,通过在下茎的序列中实施U-A交换(比较gRNA 1G与gRNA 1E的连接效率)来改善gRNA 1G的连接效率,其中U-A交换被设计为防止连接前两个指导分子片段的交错退火。

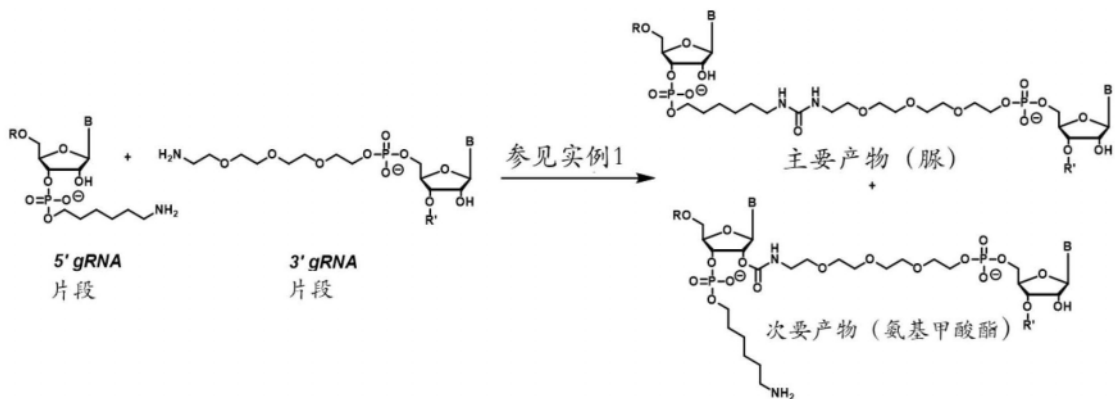
实例9:用质谱法表征脲键合

[0343] 通过质谱法表征含有脲键合并如实例1中所述合成的化学缀合的指导分子。在合成、化学连接和纯化后,使用T1内切核酸酶将gRNA 1A(参见表10)在一级序列中每个G核苷酸的3'末端切割成片段。使用LC-MS分析这些片段。特别是,在保留时间为4.50分钟时检测到含有脲键合A-[UR]-AAUAG(A34:G39)的片段, $m/z=1190.7$ (图12A和图12B)。该前体离子的LC/MS-MS分析显示碰撞诱导的解离片段离子与gRNA 1A中的脲键合一致。

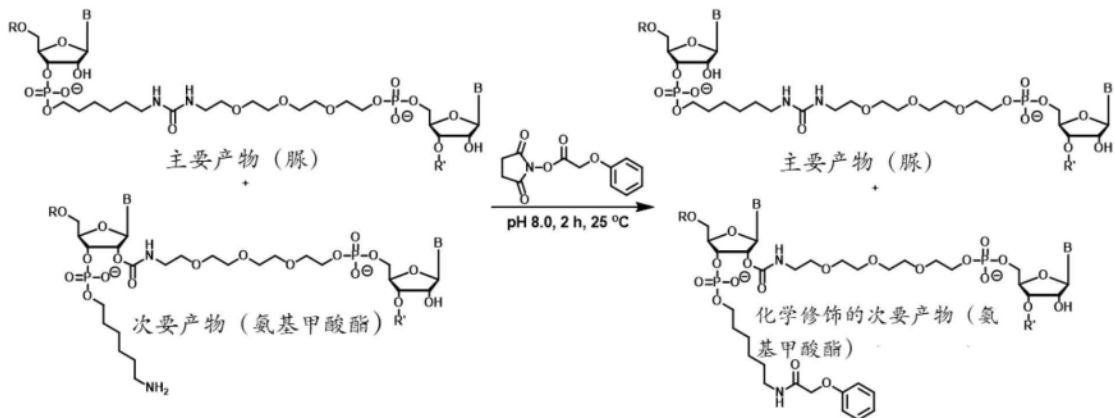
### 实例10:氨基甲酸酯副产物的表征

[0344] 图13A显示了脲连接的指导分子的未纯化的组合物的LC-MS数据,其中存在主要产物(A-2,保留时间为3.25min)和次要产物(A-1,保留时间为3.14min)。我们注意到,为了说明的目的,通过组合来自阴离子交换纯化的级分(其包含更高百分比的氨基甲酸酯次要产物)来富集图13A中的次要产物(A-1)。在根据实例1的方法合成指导分子时,通常以高达10%的产率检测到副产物。通过质谱分析每个峰表明两种产物具有相同的分子量(参见图13B和图13C)。

[0345] 鉴于此,我们假设次要产物是由3'指导分子片段的5'末端的5'-NH<sub>2</sub>与5'指导分子片段的3'末端的2'-OH之间的反应产生的氨基甲酸酯副产物,如下:



[0346] 为了进一步证实氨基甲酸酯副产物的分配,用苯氧基乙酸N-羟基琥珀酰亚胺酯进行化学修饰。基本化学原理预测只有次要产物(氨基甲酸酯)具有反应性亲核中心(游离胺),因此只有次要产物将被化学官能化。因此,在脲连接的指导分子的粗混合物中加入苯氧基乙酸N-羟基琥珀酰亚胺酯应该产生主要产物(脲)和化学修饰的次要产物(氨基甲酸酯)的混合物:



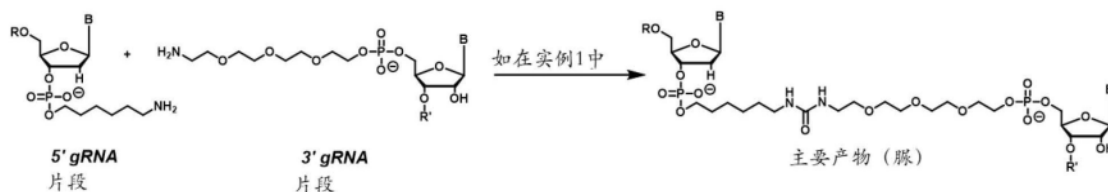
[0347] 图14A显示化学修饰后指导分子组成的LC-MS数据。主要产物(B-1,脲)具有与原始分析相同的保留时间(3.26min,图13A),而次要产物(B-1,氨基甲酸酯)的保留时间已经转

移至3.86min,与游离胺部分的化学官能化相一致。此外,在3.86min (M+134) 处的峰的质谱分析表明已经发生预测的官能化(参见图14B)。这些结果表明次要产物确实是氨基甲酸酯副产物。

[0348] 为了进一步证实氨基甲酸酯副产物的特性,将主要产物(脒)和化学修饰的次要产物(氨基甲酸酯)的混合物用核糖核酸酶A进行消化(参见实例9),其在一级序列中每个G核苷酸的3'末端处切割指导分子。然后通过LC-MS分析片段,并检测脒键合(G35-[UR]-C36)和化学修饰的氨基甲酸酯键(G35-[CA+PAA]-C36)。图15A显示了片段混合物的LC-MS迹线,其中脒键合的保留时间是4.31min并且化学修饰的氨基甲酸酯键的保留时间是5.77min。图15B显示了在4.31min处的峰的质谱,其中 $m/z=532.1$ 被分配给 $[M-2H]^{2-}$ ,而图15C显示了在5.77min处的峰的质谱,其中 $m/z=599.1$ 被分配给 $[M-2H]^{2-}$ 。使用LC-MS/MS技术进一步分析质谱。在 $m/z532.1$ ,  $[M-2H]^{2-}$  处的脒连接产物的LC-MS/MS谱(图15D)包含在寡核苷酸碰撞诱导解离(CID)实验中观察到的典型a-d和x-z离子。此外,观察到UR键合的从5'末端( $m/z=487.1$ 和 $461.1$ )和3'末端( $m/z=603.1$ 和 $577.1$ )的任一侧的MS/MS片段离子。相比之下,在 $m/z599.1$ ,  $[M-2H]^{2-}$ , 化学修饰的氨基甲酸酯连接产物的LC-MS/MS谱图(图15E)中仅观察到两种产物离子,包括来自氨基甲酸酯键合的5'末端( $m/z=595.2$ )的和来自氨基甲酸酯键合的3'末端( $m/z=603.1$ )的MS/MS片段离子。

#### 实例11:用于单一产物形成的核苷酸修饰

[0349] 我们假设可以通过在5'指导分子片段的3'末端的核苷酸中的策略性2'-修饰来防止如实例10中所述的氨基甲酸酯副产物的形成。例如,在用2'-H替换5'指导分子片段的3'末端的核苷酸中的2'-OH情况下,假设按照实例1的方法合成脒连接的指导分子产生单一的脒连接产物,不含氨基甲酸酯副产物:



[0350] 图16A显示与2'-H修饰的5'指导分子片段(上部光谱)反应的粗反应混合物的LC-MS数据,相比于与未修饰形式的相同5'指导分子反应的粗反应混合物(下部光谱)。用2'-H修饰的5'指导分子片段(上部光谱)观察不到氨基甲酸酯副产物形成。相反,用于与未修饰形式的相同5'指导分子片段反应的粗反应混合物(下部光谱)包括主要脒连接产物(A-2)和次要氨基甲酸酯副产物(A-1)的混合物。我们注意到,与实例10不同,氨基甲酸酯副产物没有富集,因此检测到的水平比实例10的图13A中的低得多。此外,质谱分析与2'-H修饰的5'指导分子片段(B)的反应的产物得到M-16(与A-2(主要的未经修饰的脒连接产物)相比),如对于其中2'-OH已经被2'-H代替的分子所预期(参见图16B和图16C)。

[0351] 使用表10的gRNA 1L(其含有相同的2'-H修饰)进行类似实验。通过T1核酸内切酶消化,然后进行质谱分析,确认2'-H修饰的脒连接的指导分子的形成(参见实例9)。含有脒键合的片段(2'-H-A)-[UR]-AAUAG(A34:G39)在保留时间4.65min时检测到(图17A),  $m/z=1182.7$ (图17B)。该前体离子的LC-MS/MS分析显示在与2'-H修饰的核苷酸的反应中与脒键合一致的片段离子。

[0352] 这些结果表明,通过5'指导分子片段的3'末端的核苷酸中的2'-OH修饰,可以避免

氨基甲酸酯副产物的形成。因此,脲连接的指导分子以高纯度合成,这简化了制备缀合的指导分子的整个过程。

通过引用并入

[0353] 本文提及的所有出版物、专利和专利申请都通过引用以其全文而特此结合,如同每一单独的出版物、专利或专利申请具体且单独地指明通过引用而结合一样。在有冲突的情况下,以本申请(包括本文的任何定义)为准。

等效物

[0354] 本领域技术人员仅使用常规实验就将认识到或能够确定本文所述的具体实施例的许多等效物。此类等效物旨在由以下权利要求涵盖。

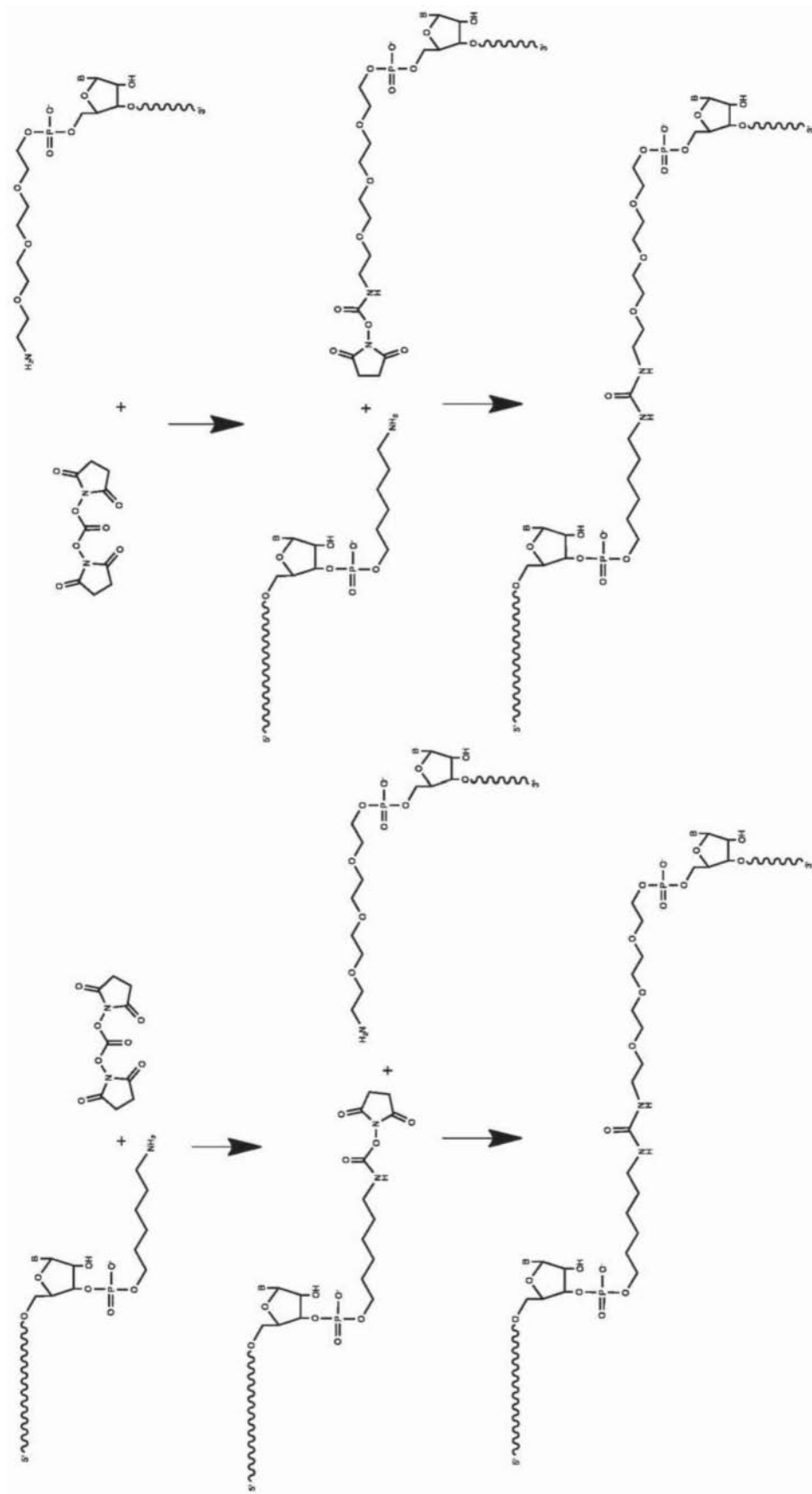


图1A

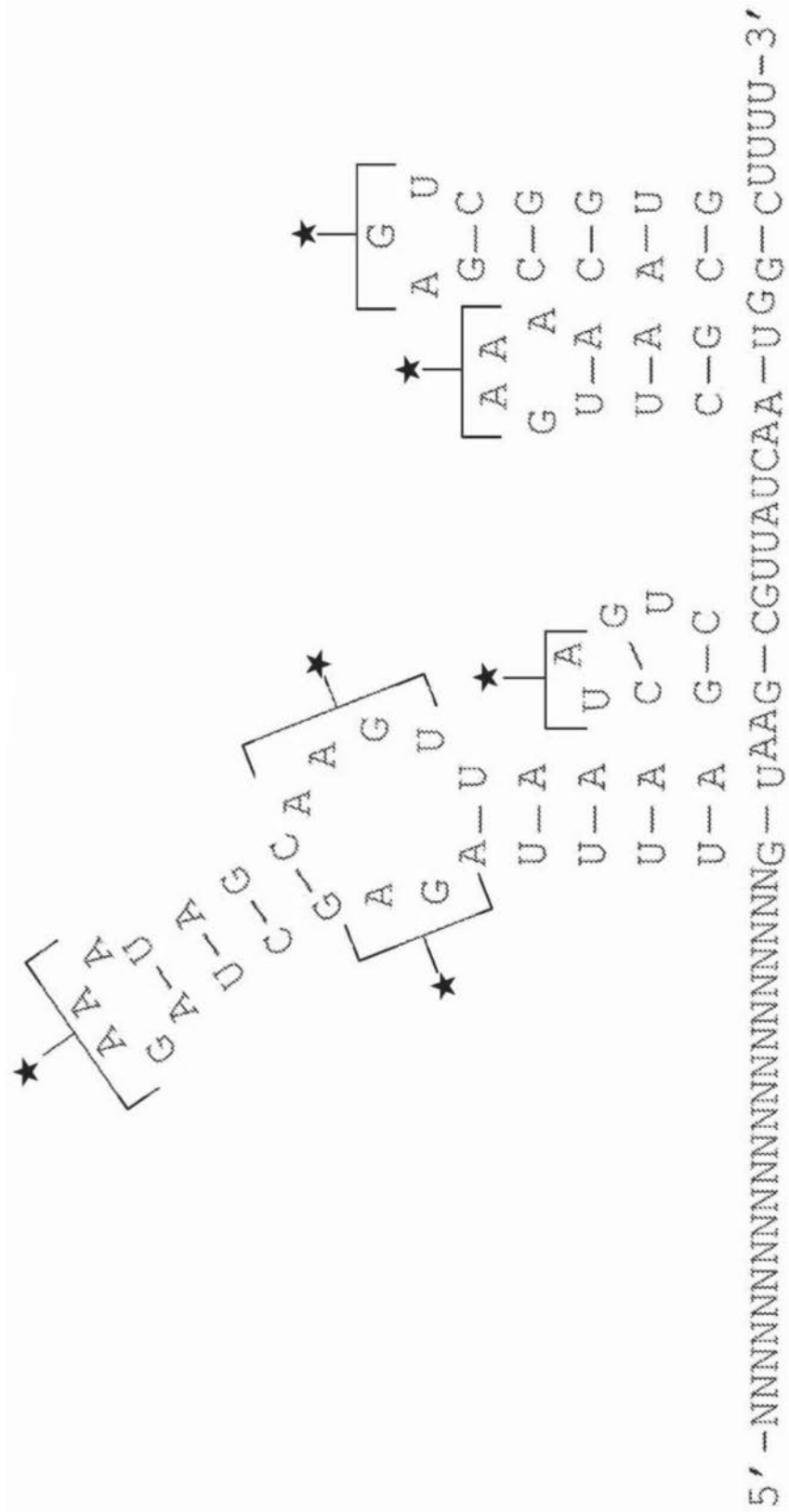


图1B

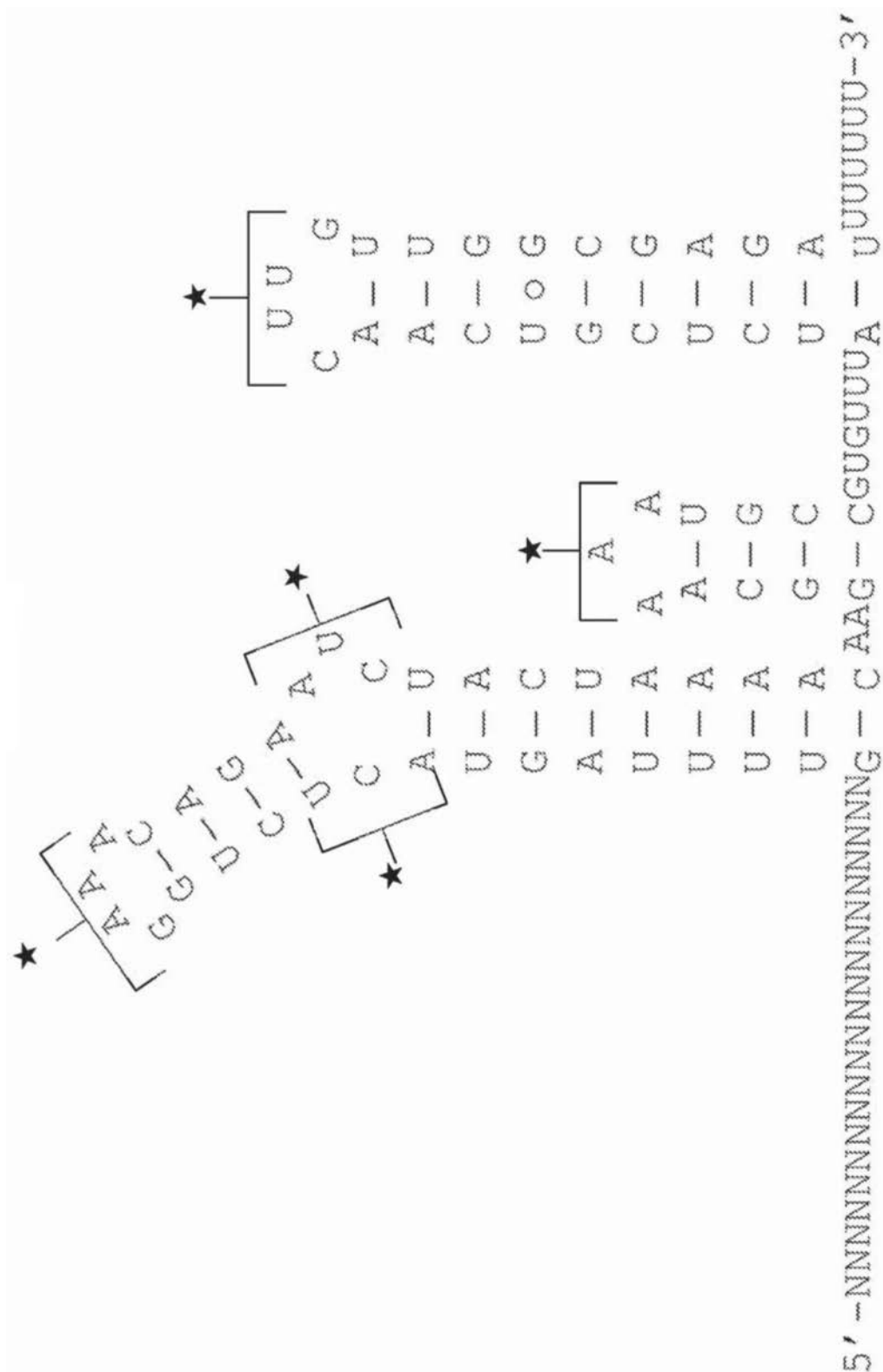


图1C

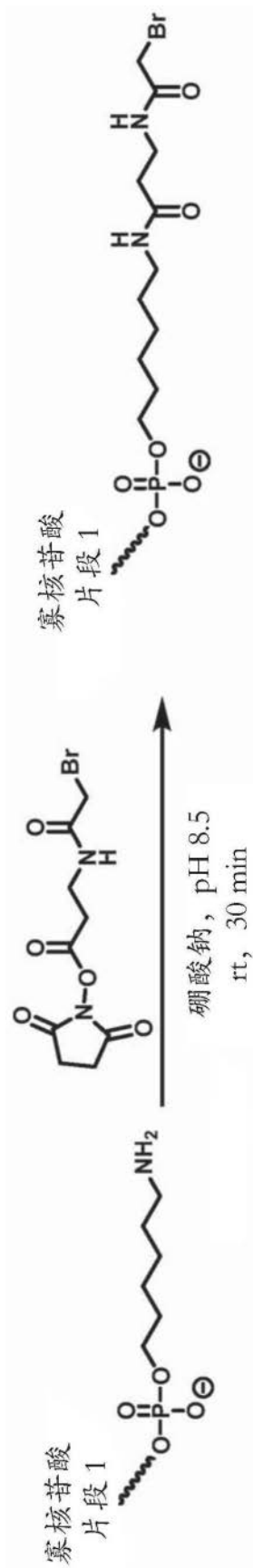


图2A

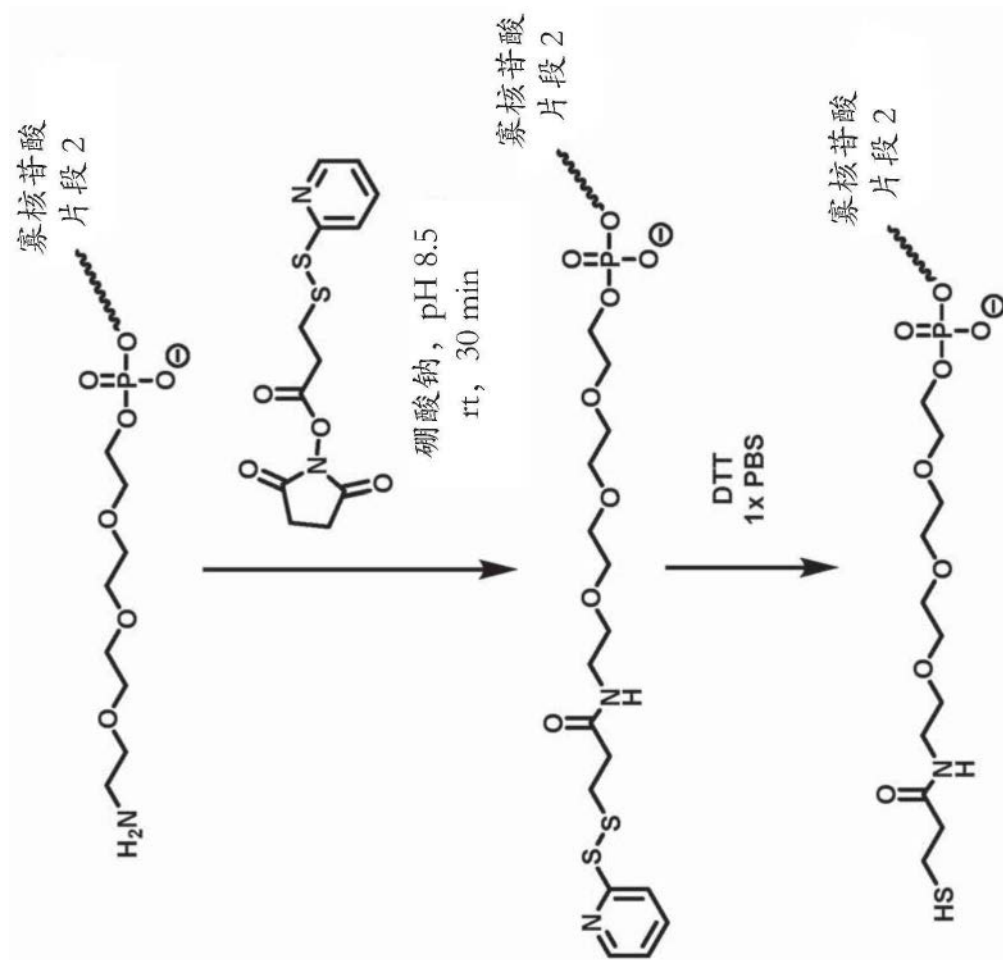


图2B

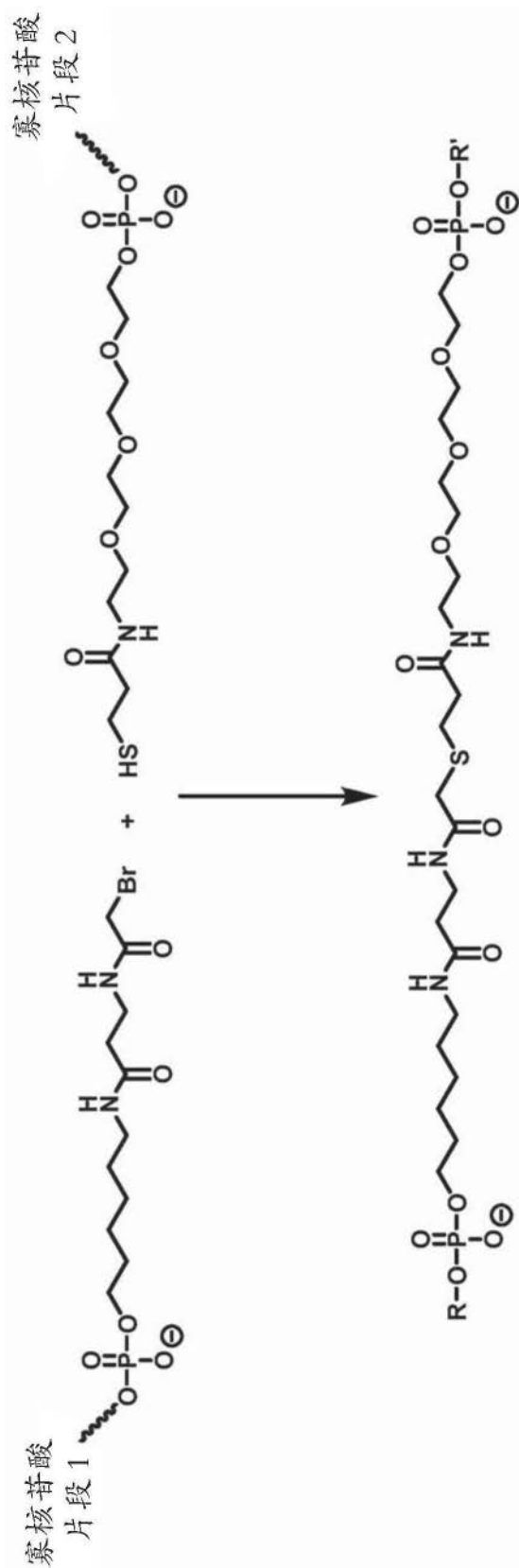


图2C

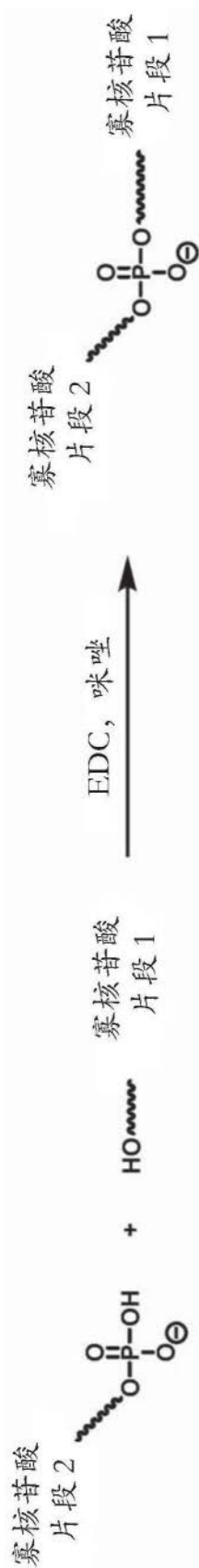


图3A

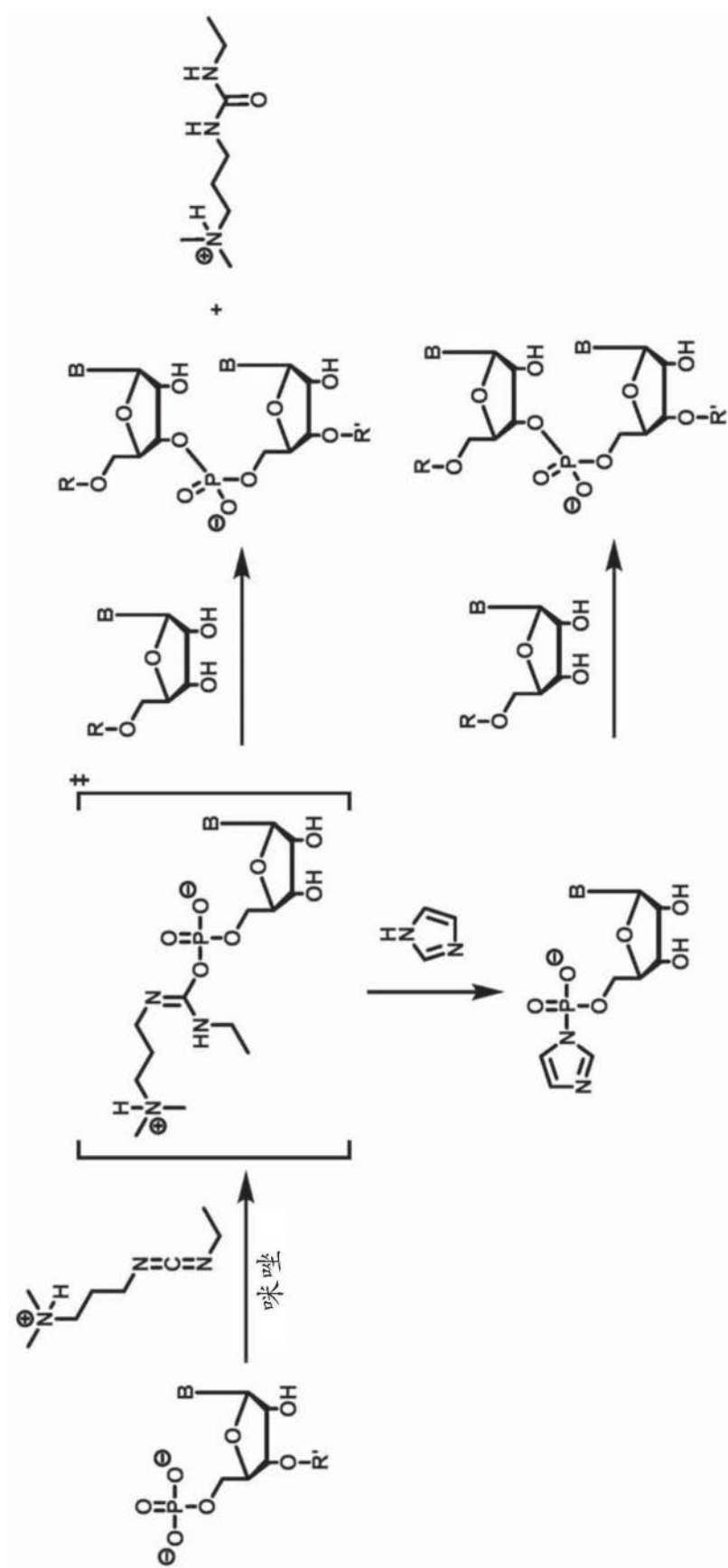


图3B

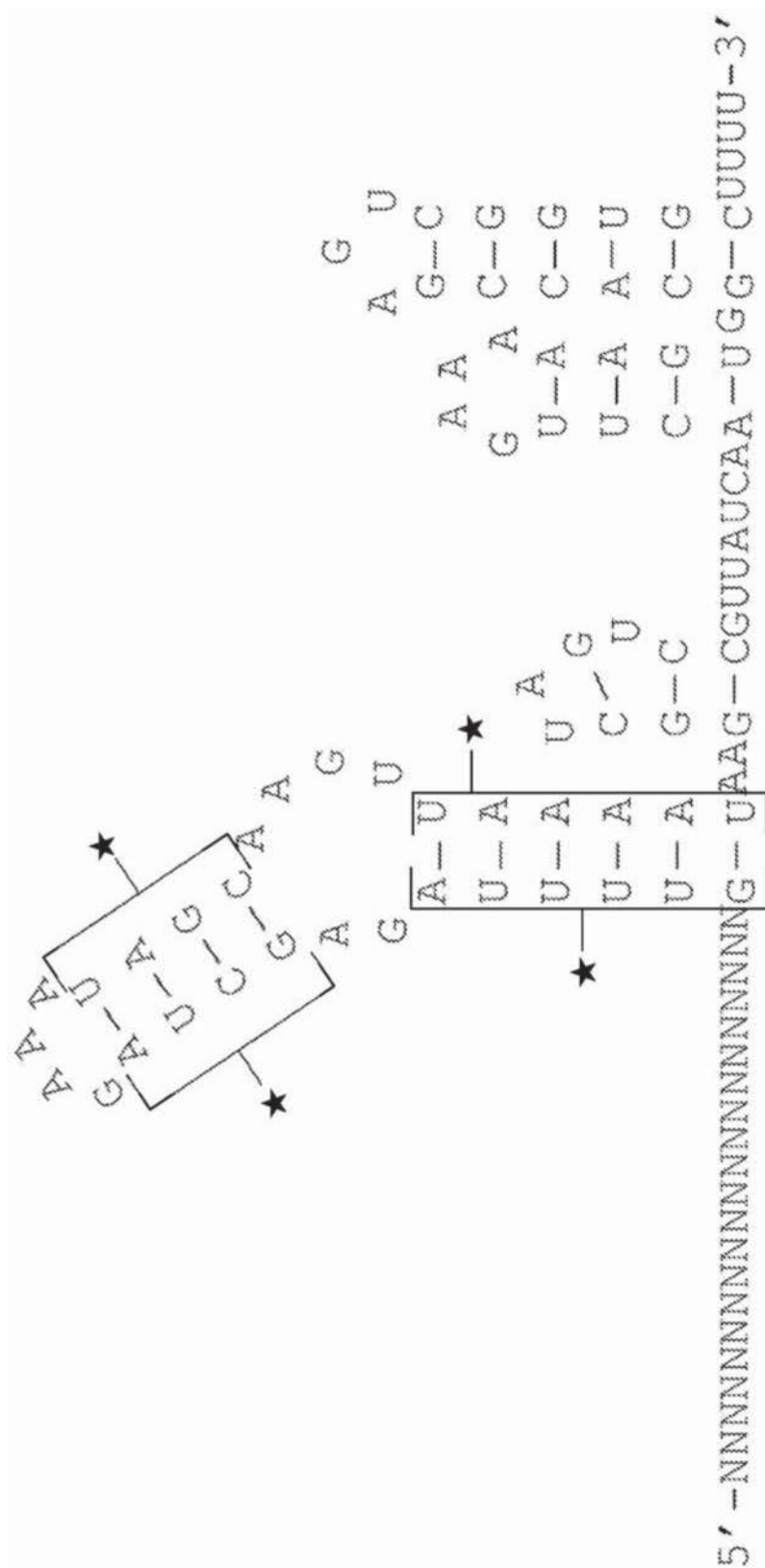


图3C

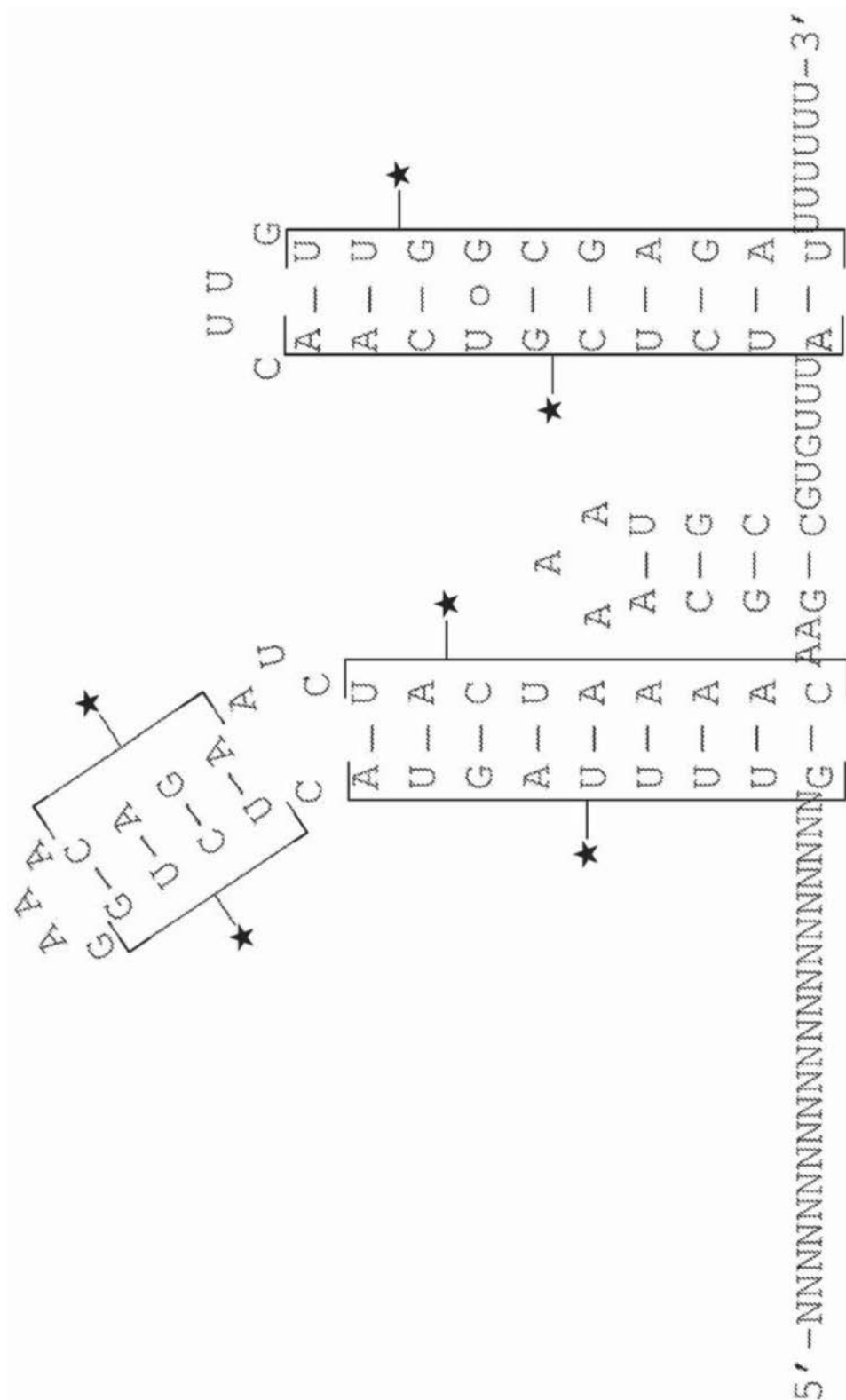


图3D

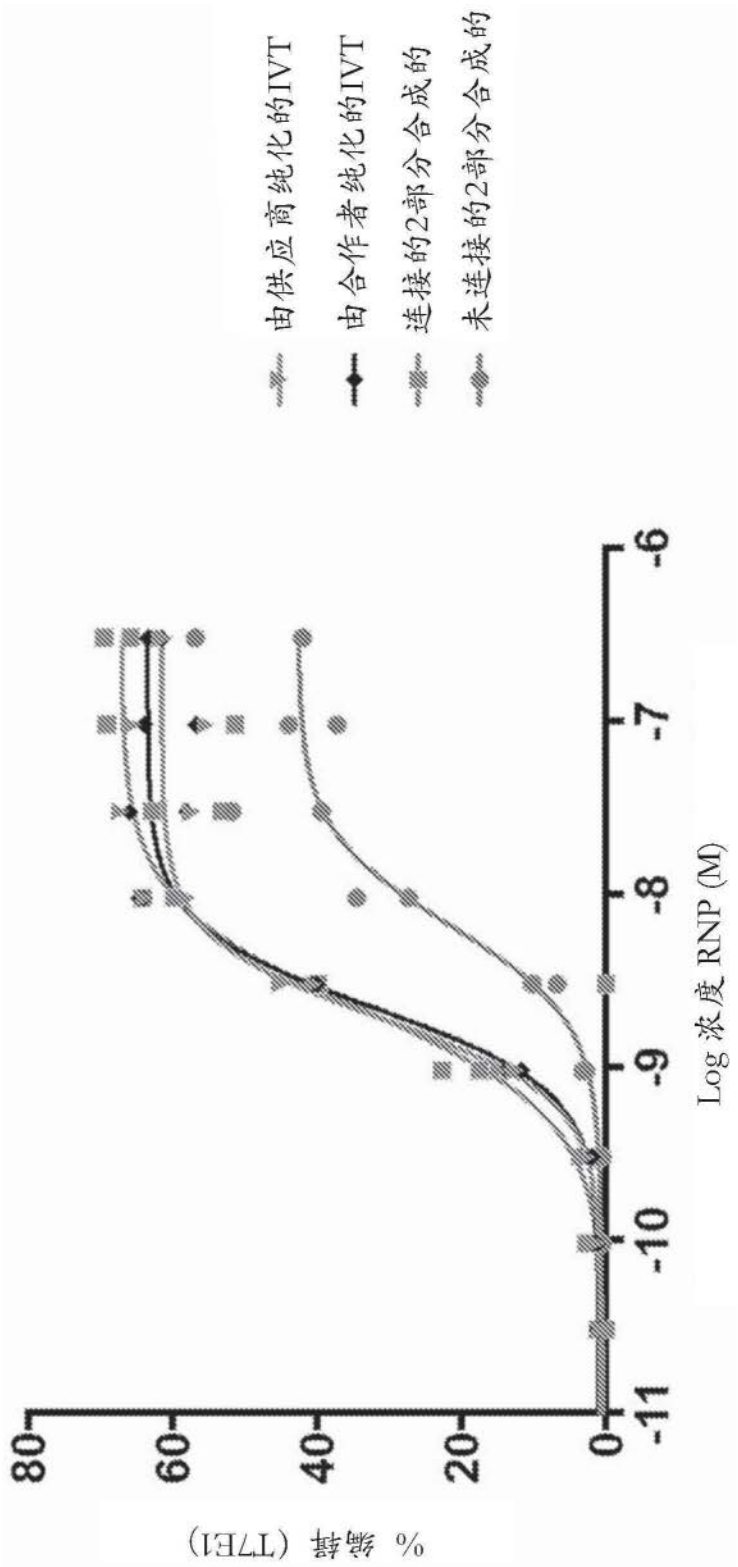


图4

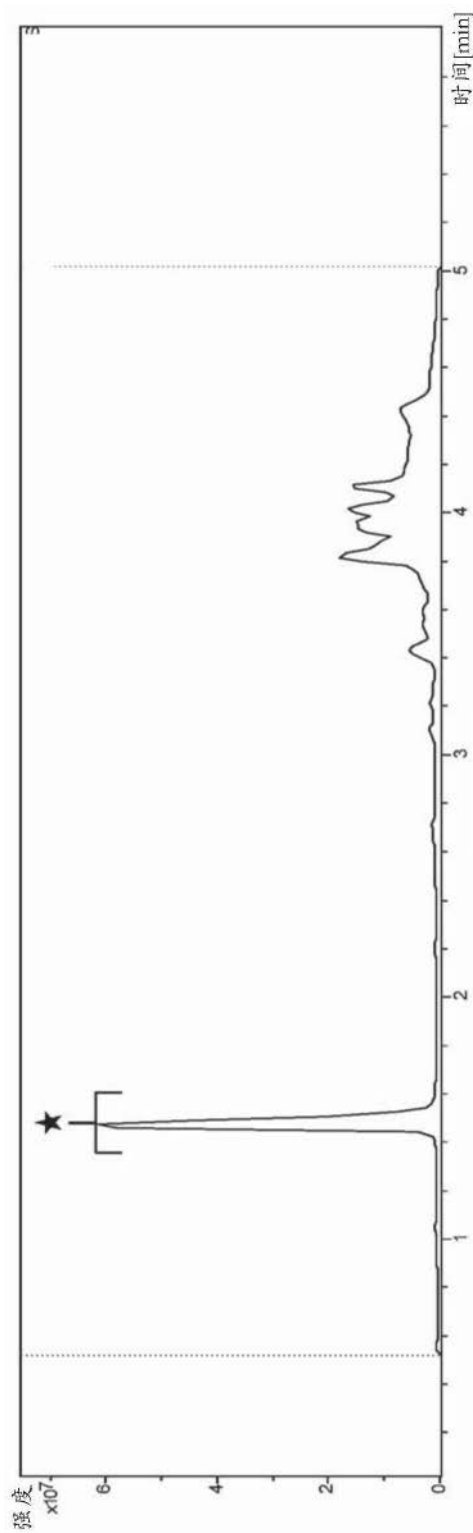


图5A

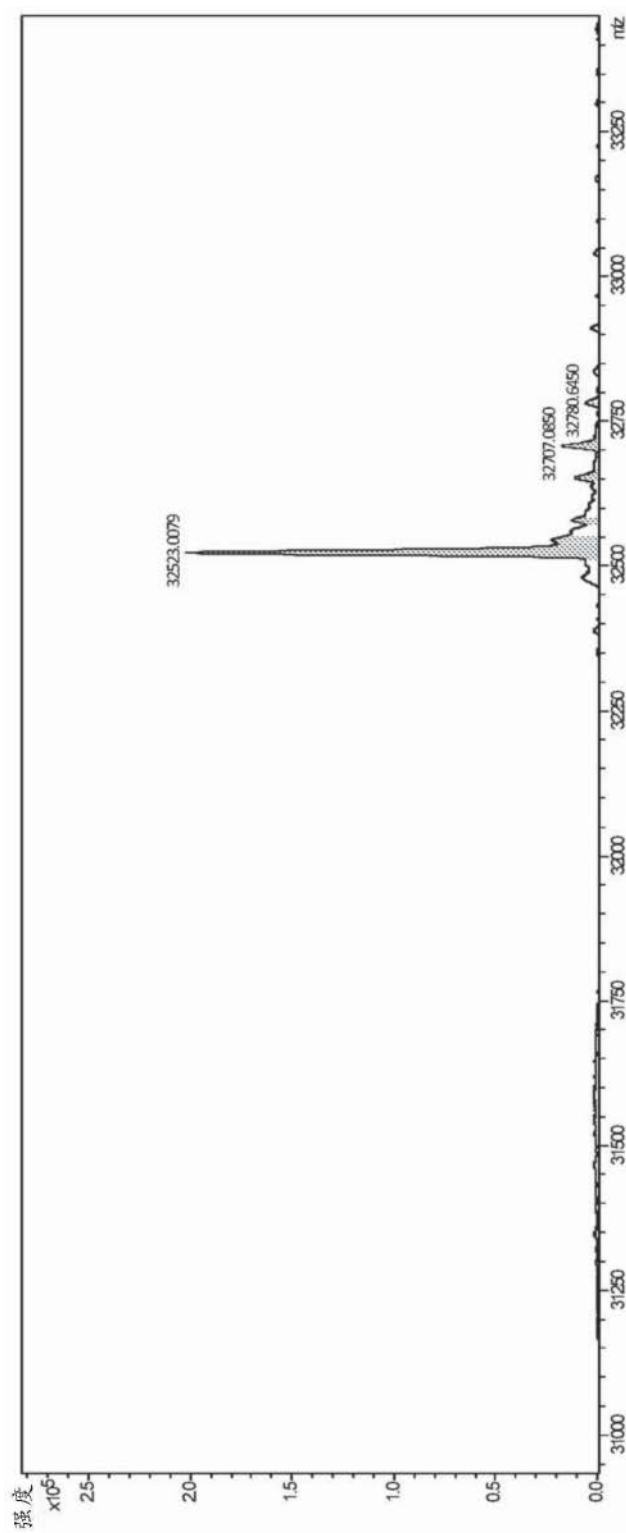


图5B

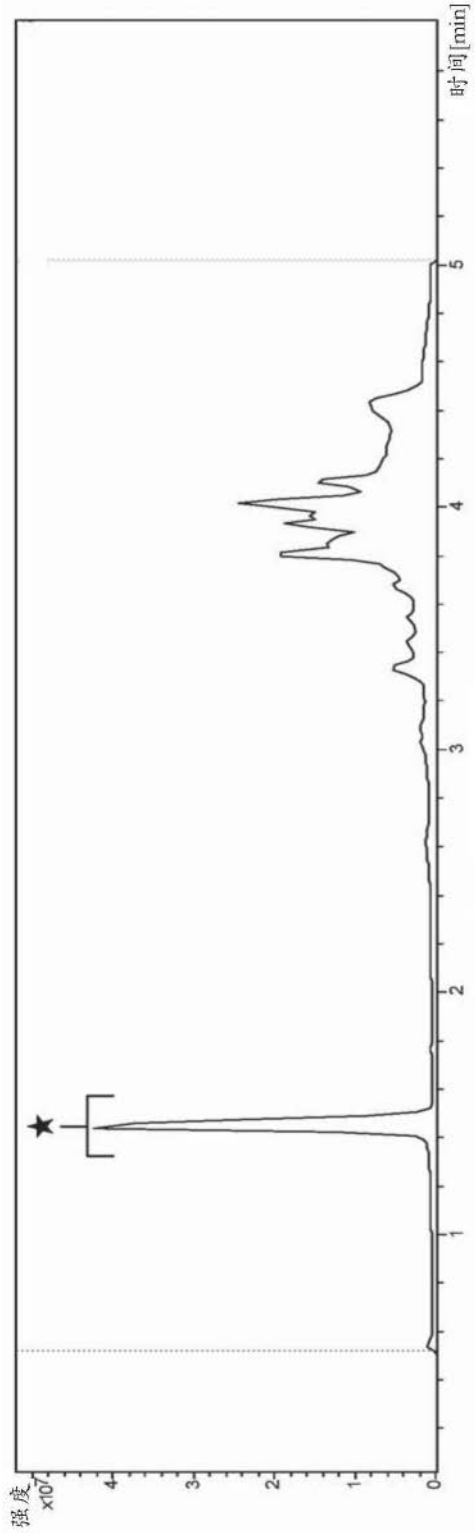


图5C

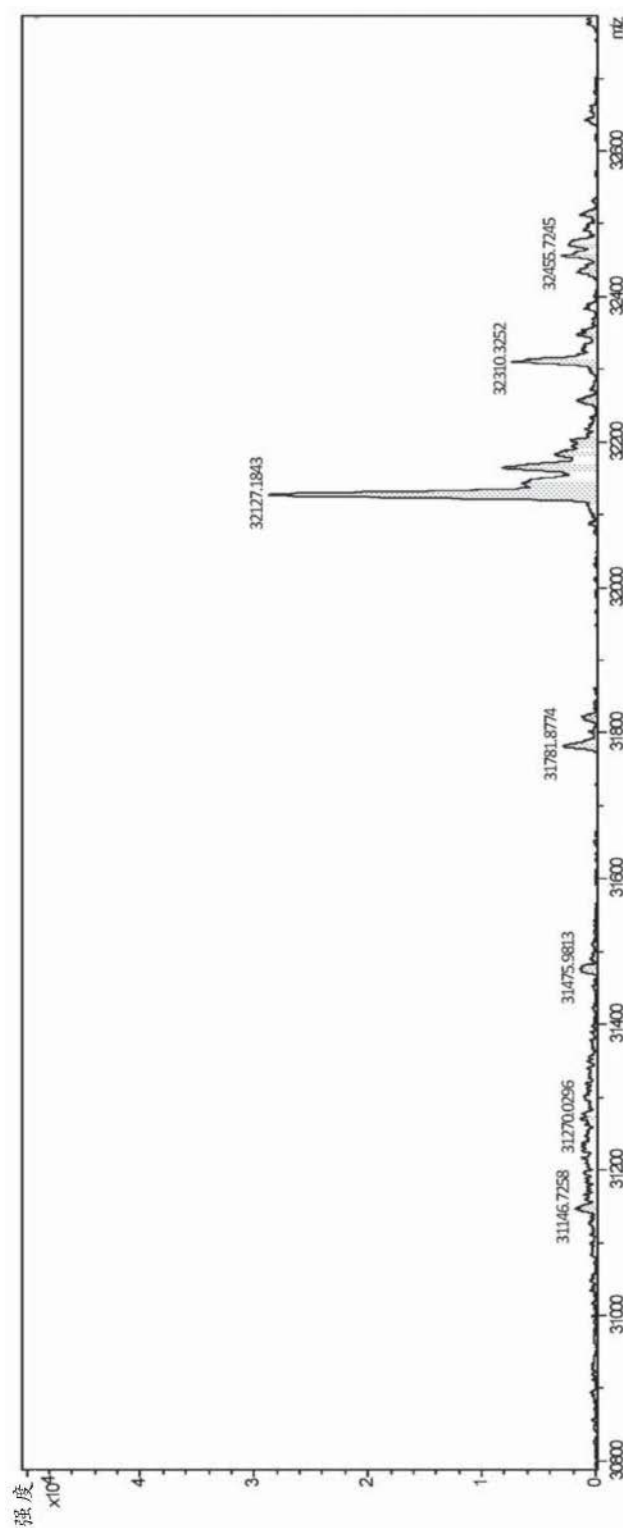


图5D

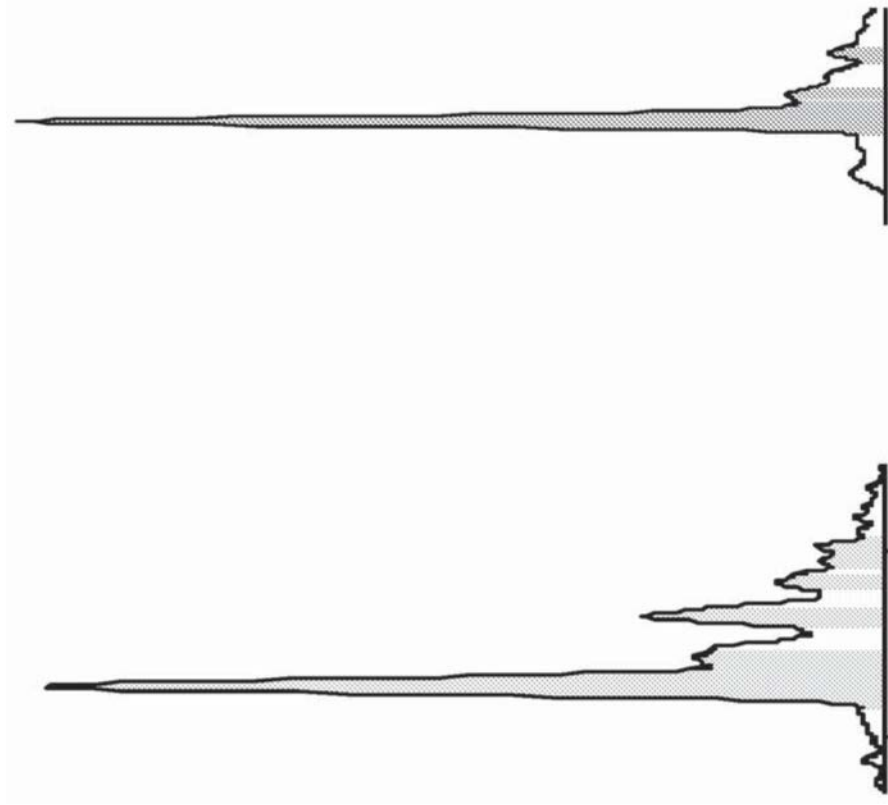


图5E

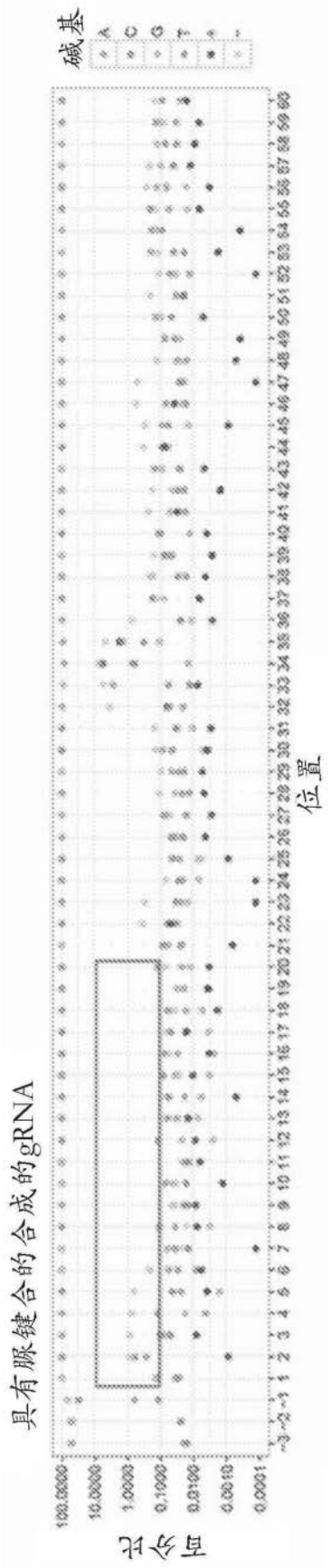


图6A

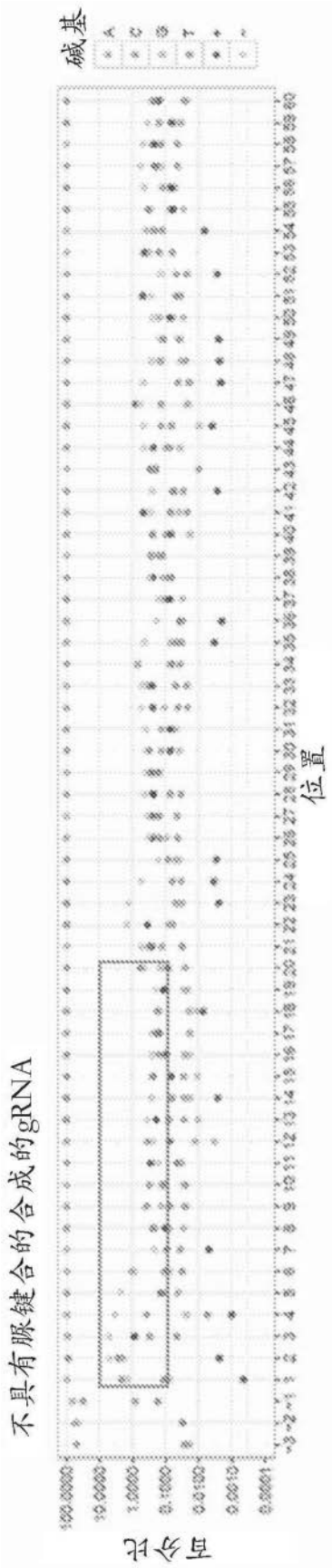


图6B

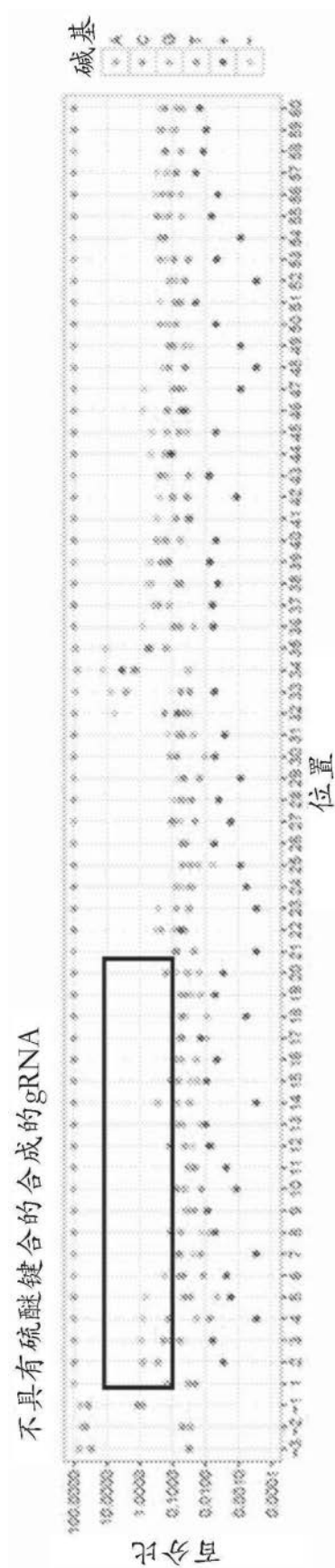


图6C

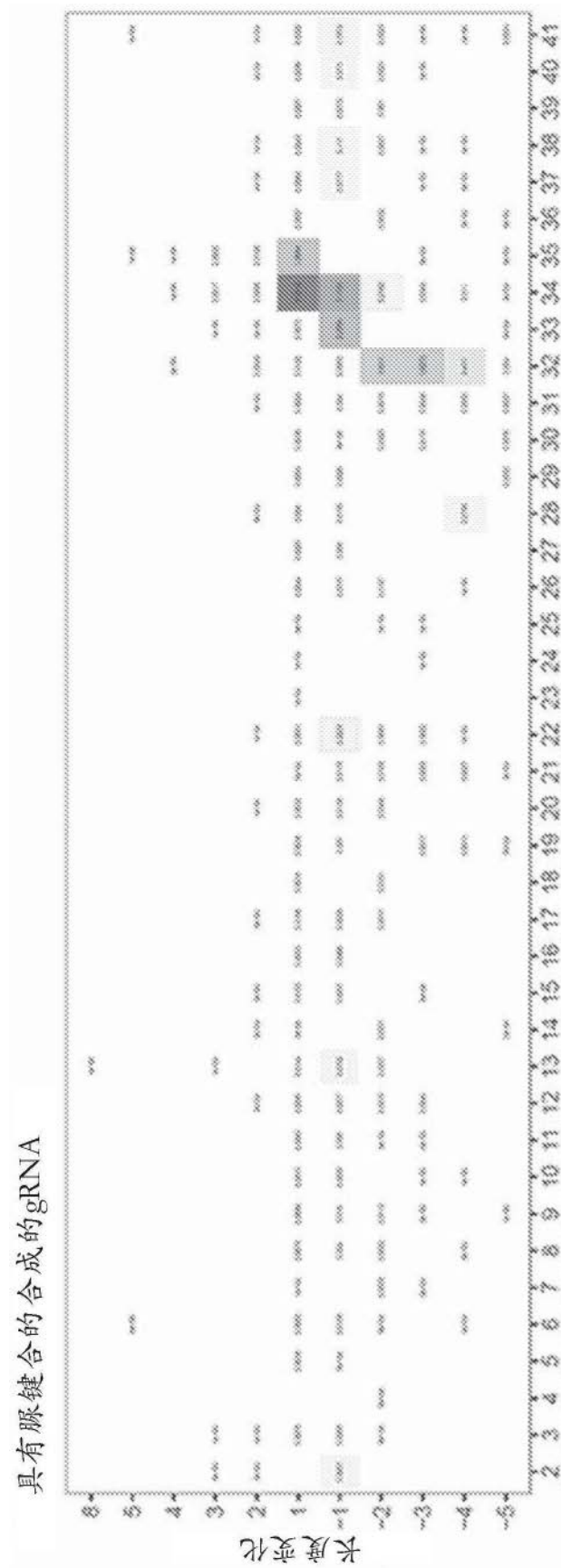


图7A

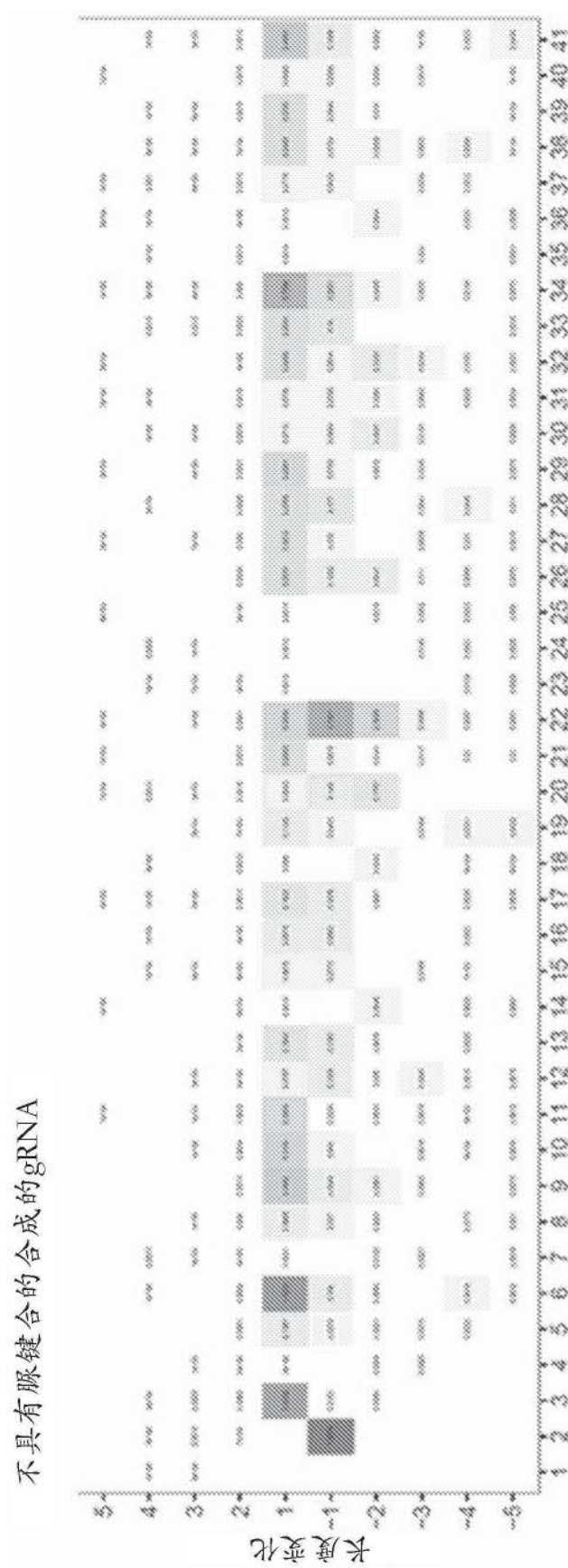


图7B

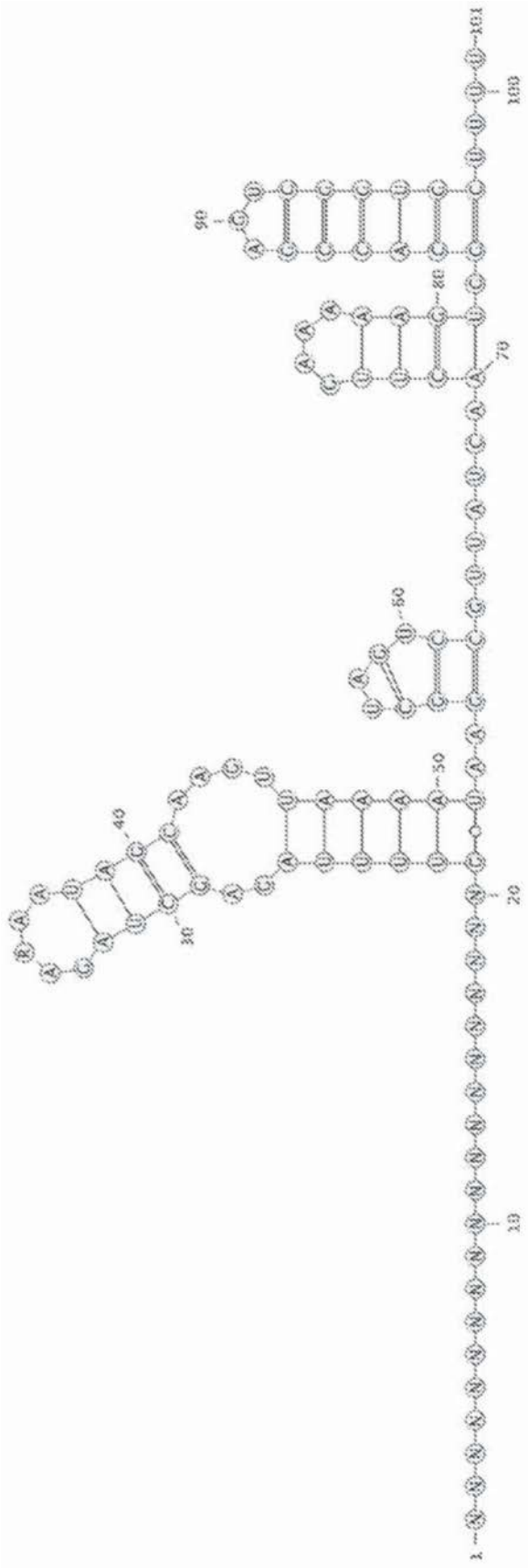


图8A

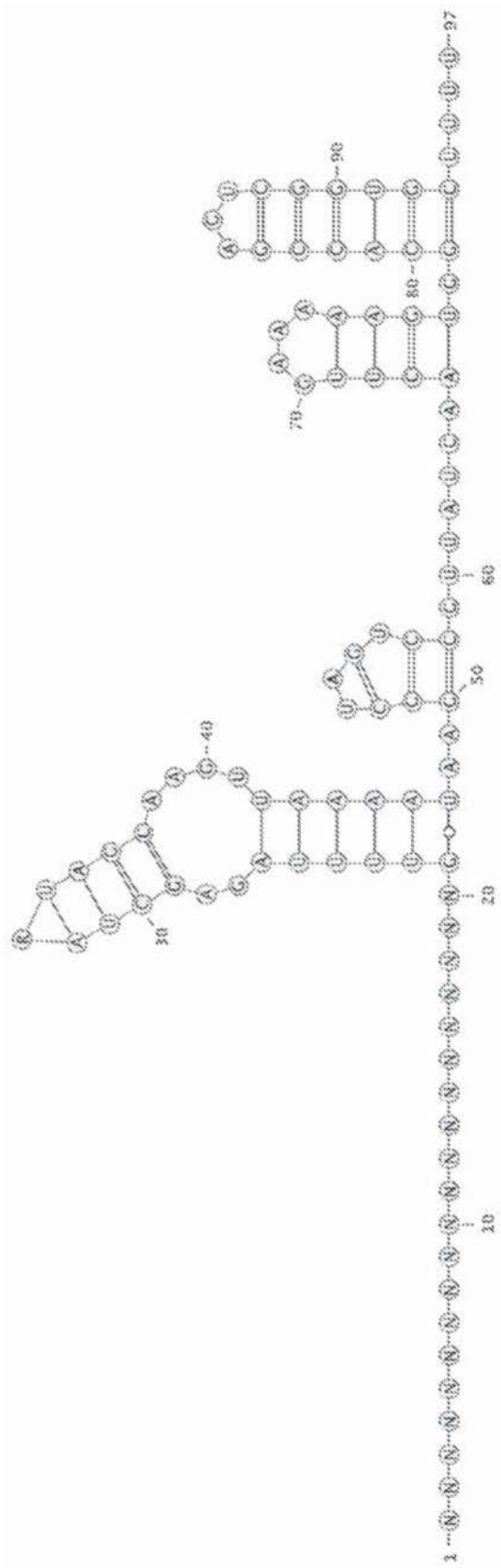


图8B

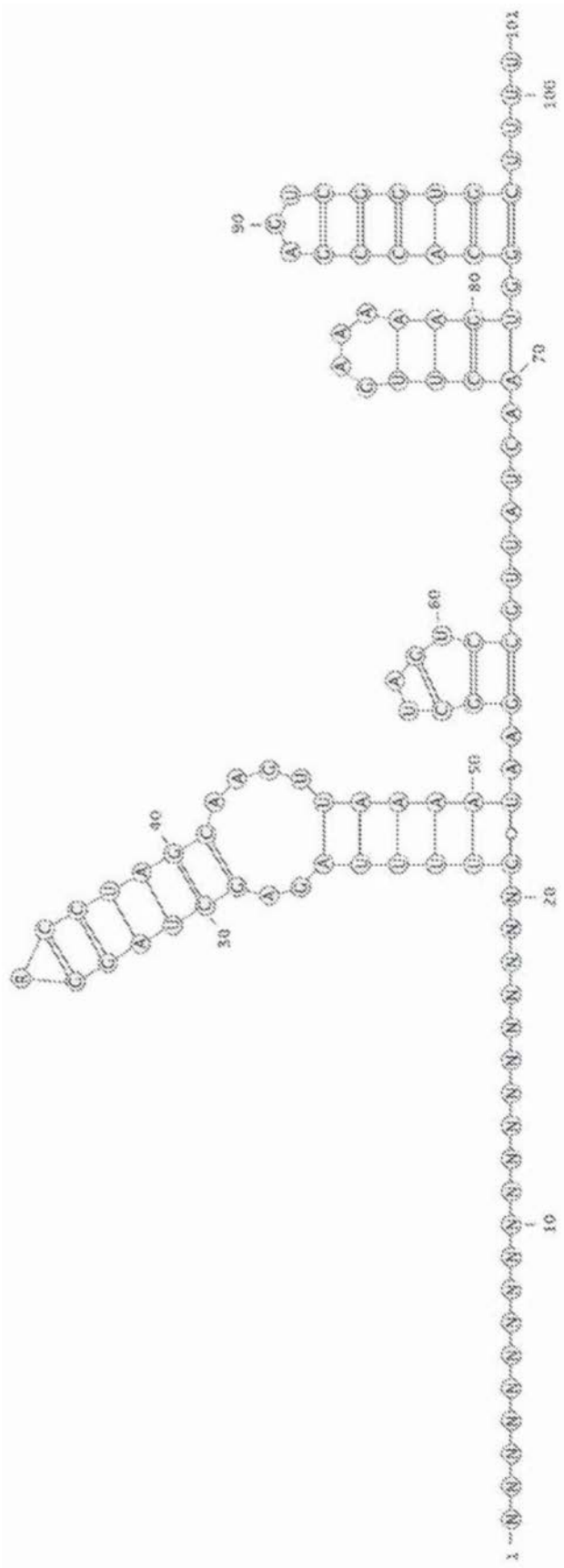


图8C

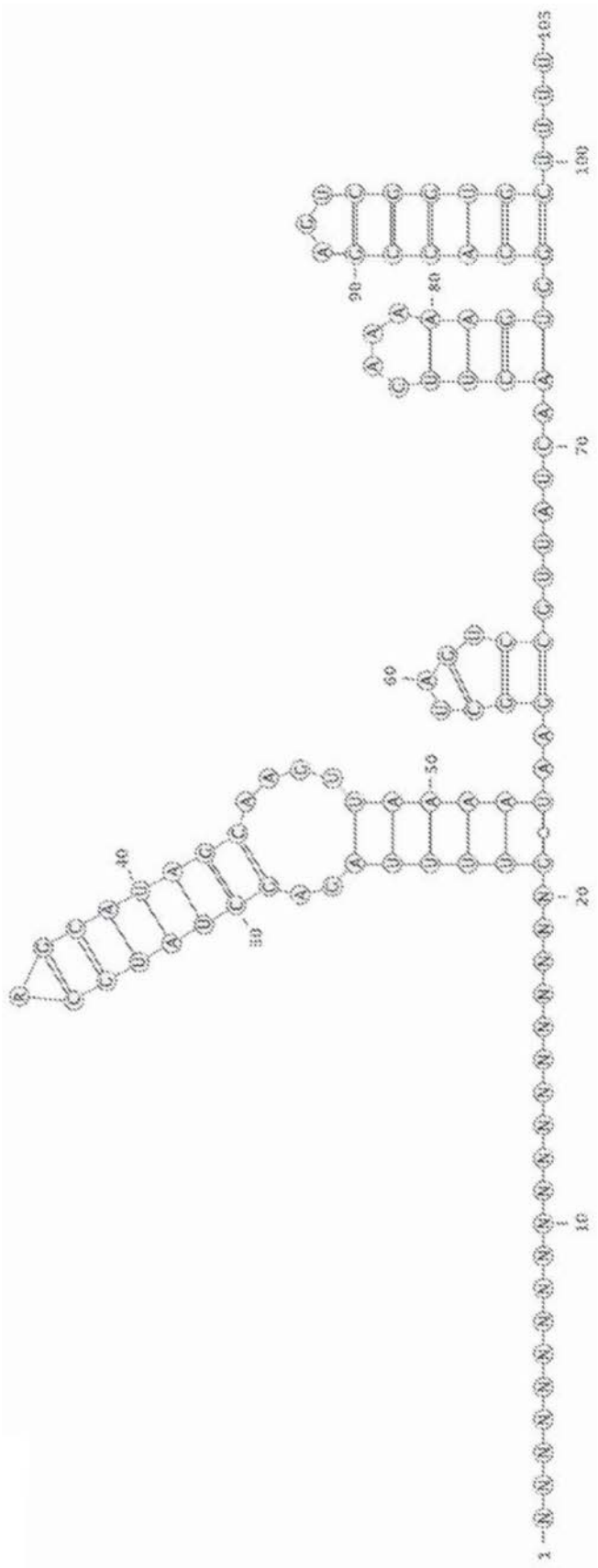


图8D

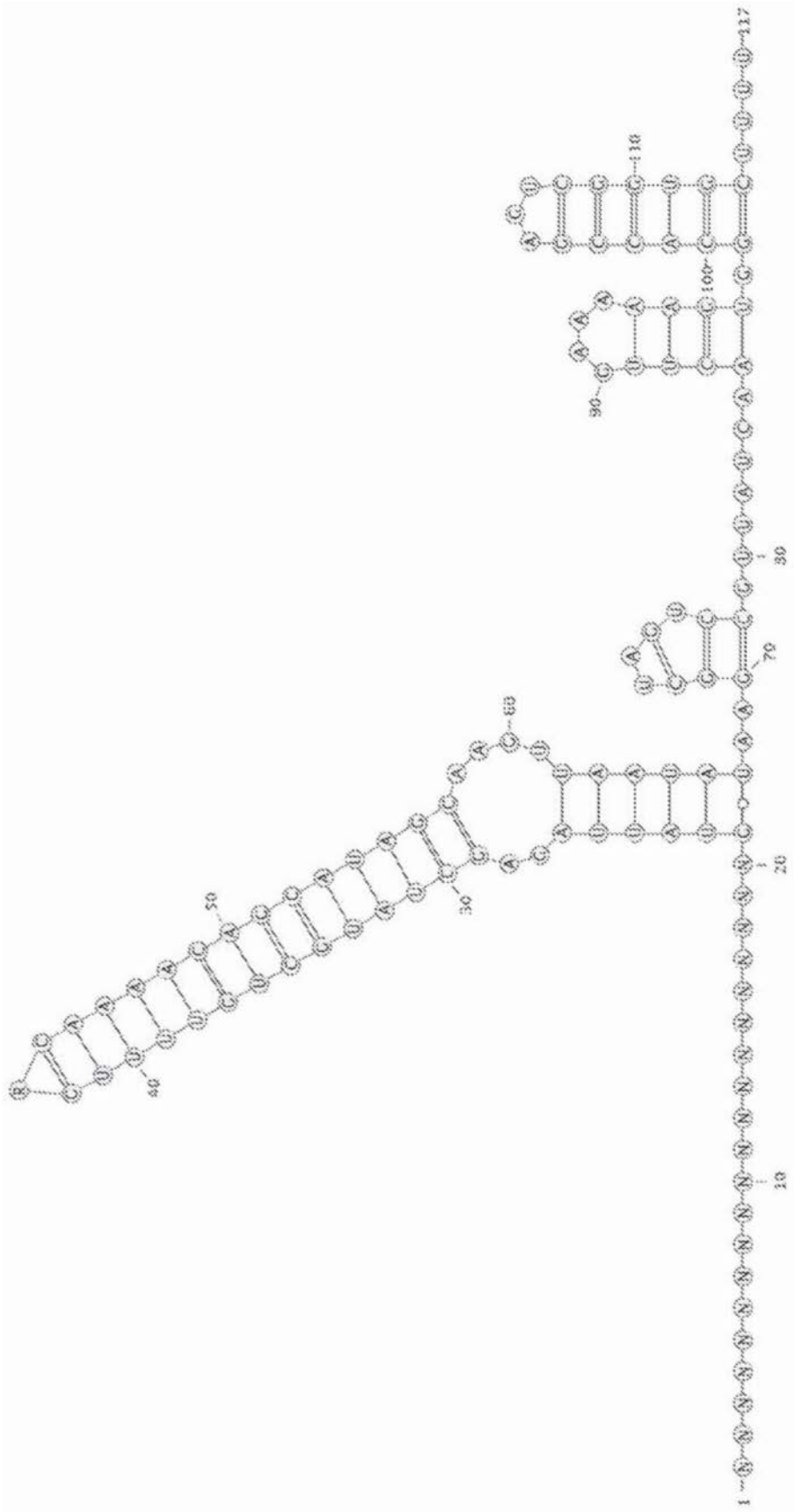


图8E



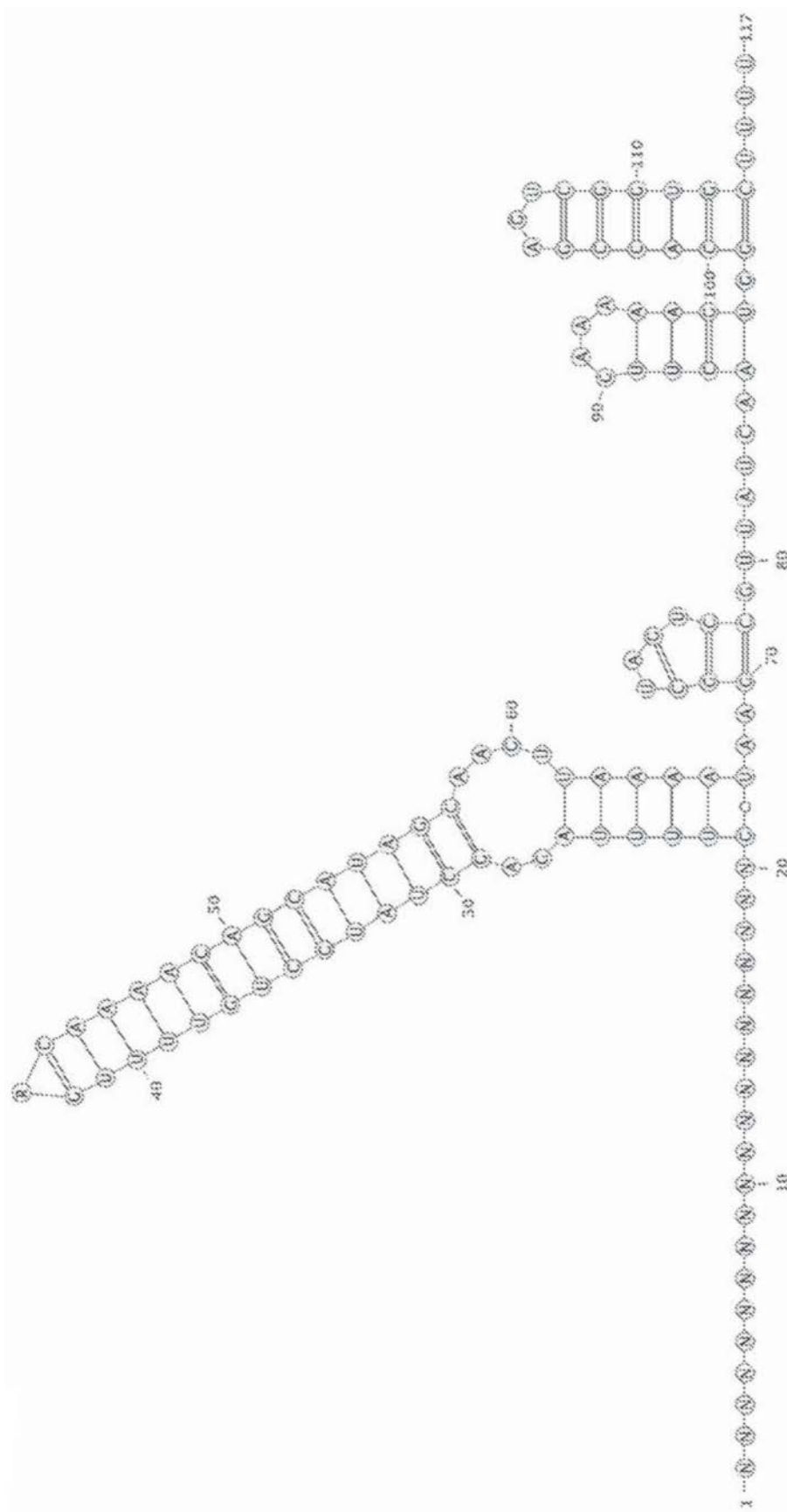


图8G

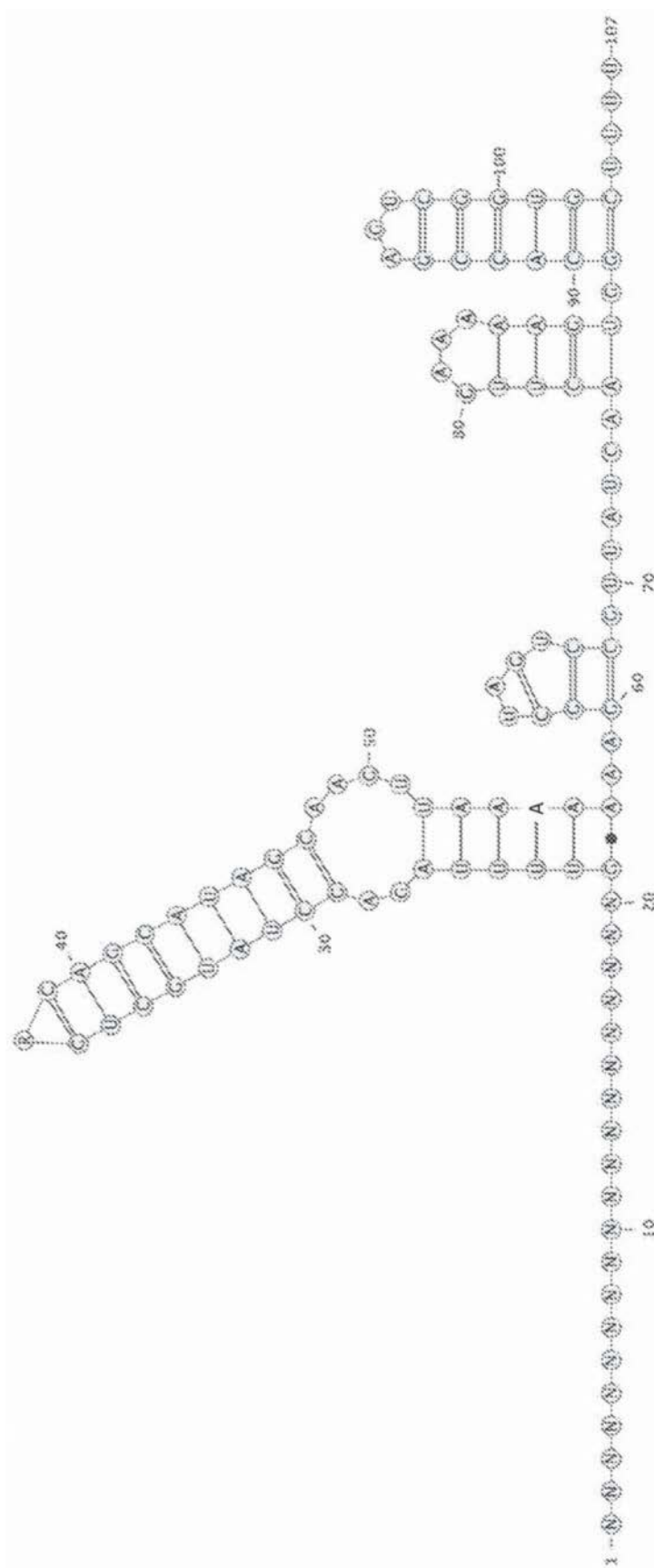


图8H

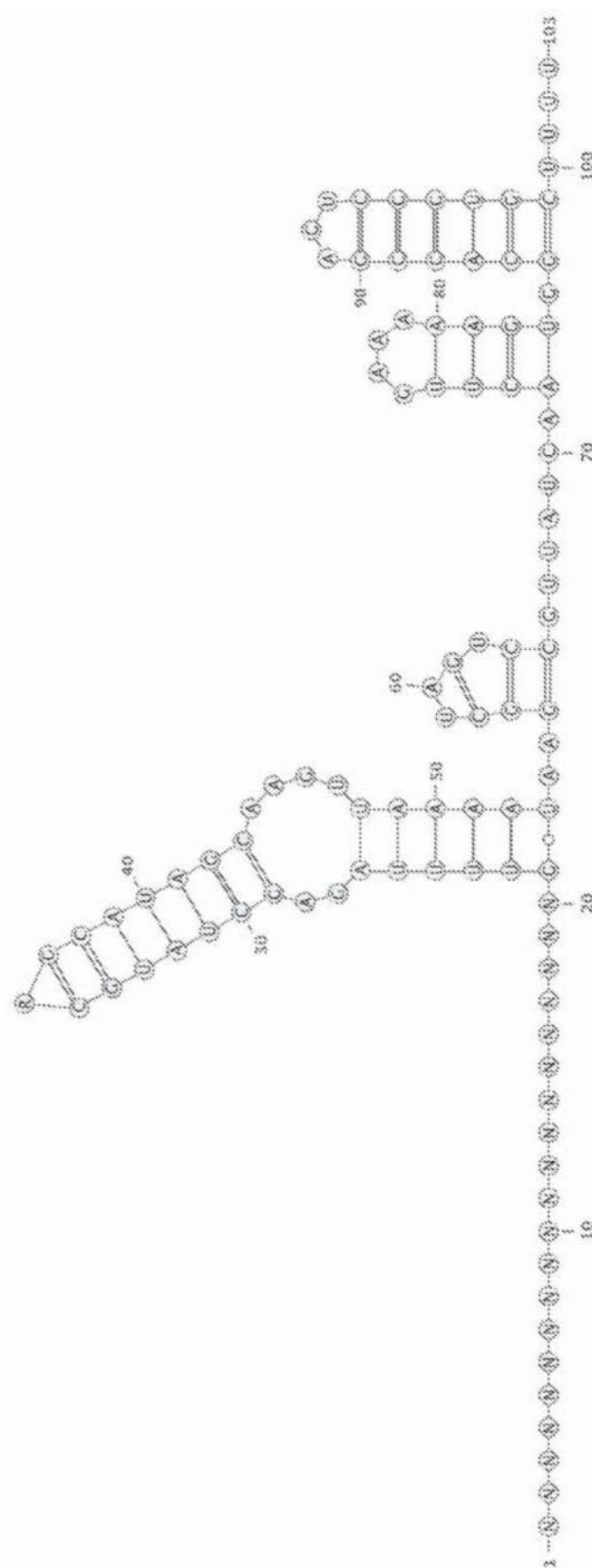


图9A

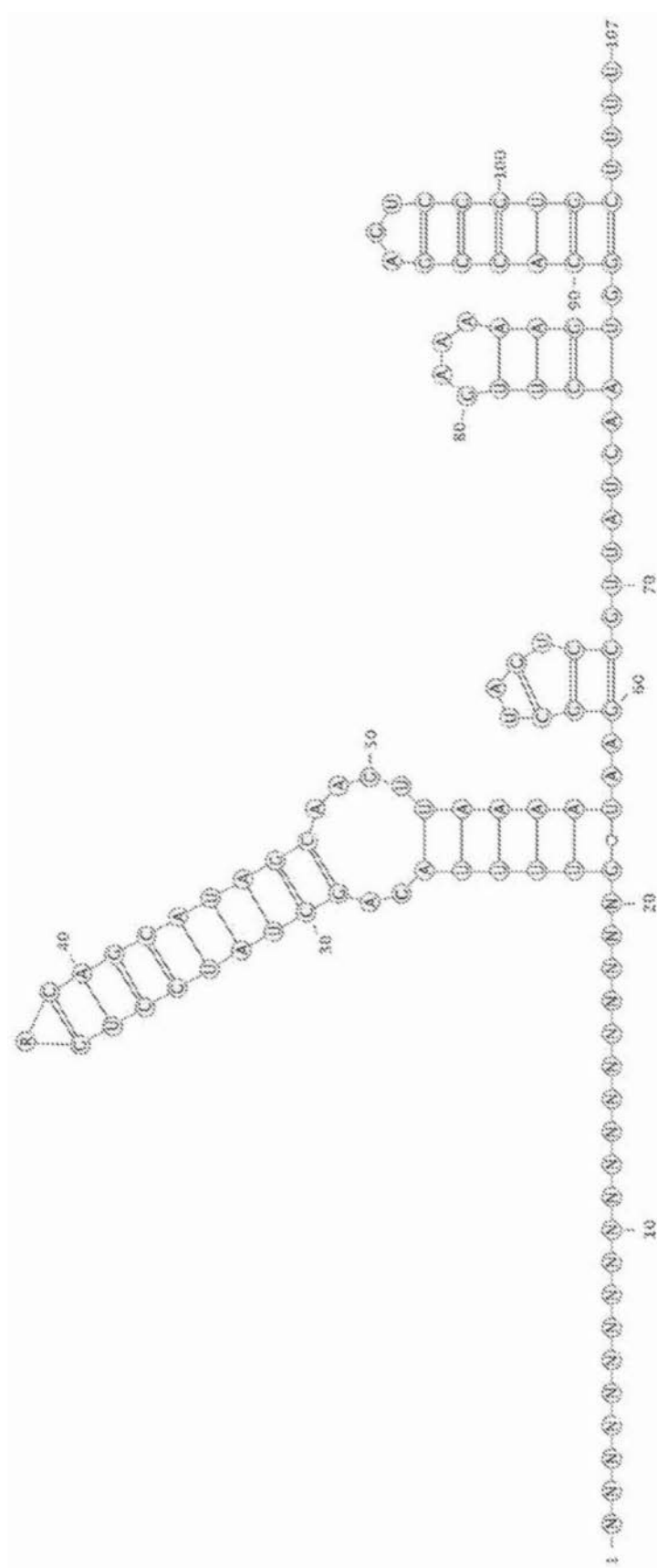


图9B

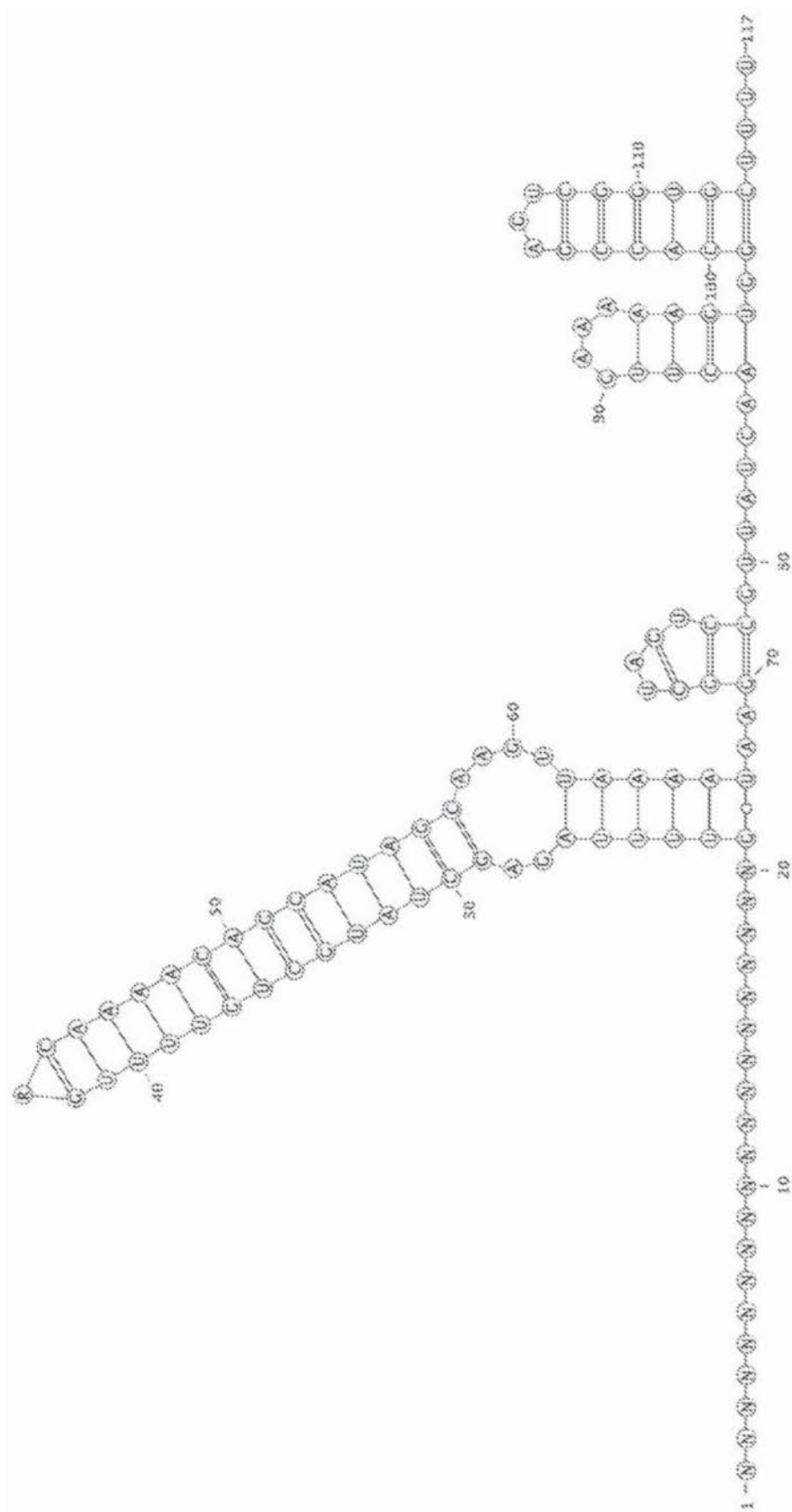


图9C

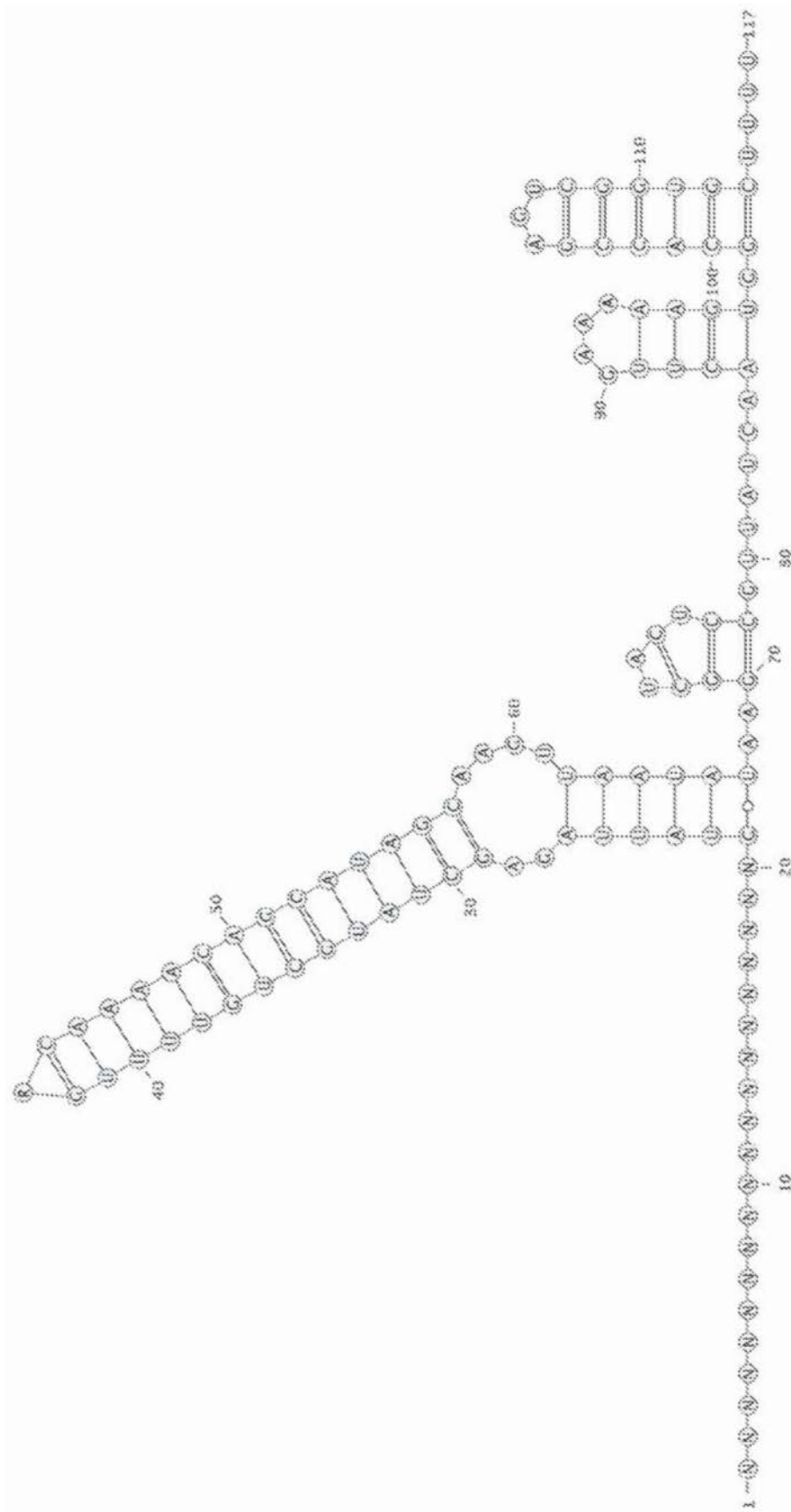


图9D

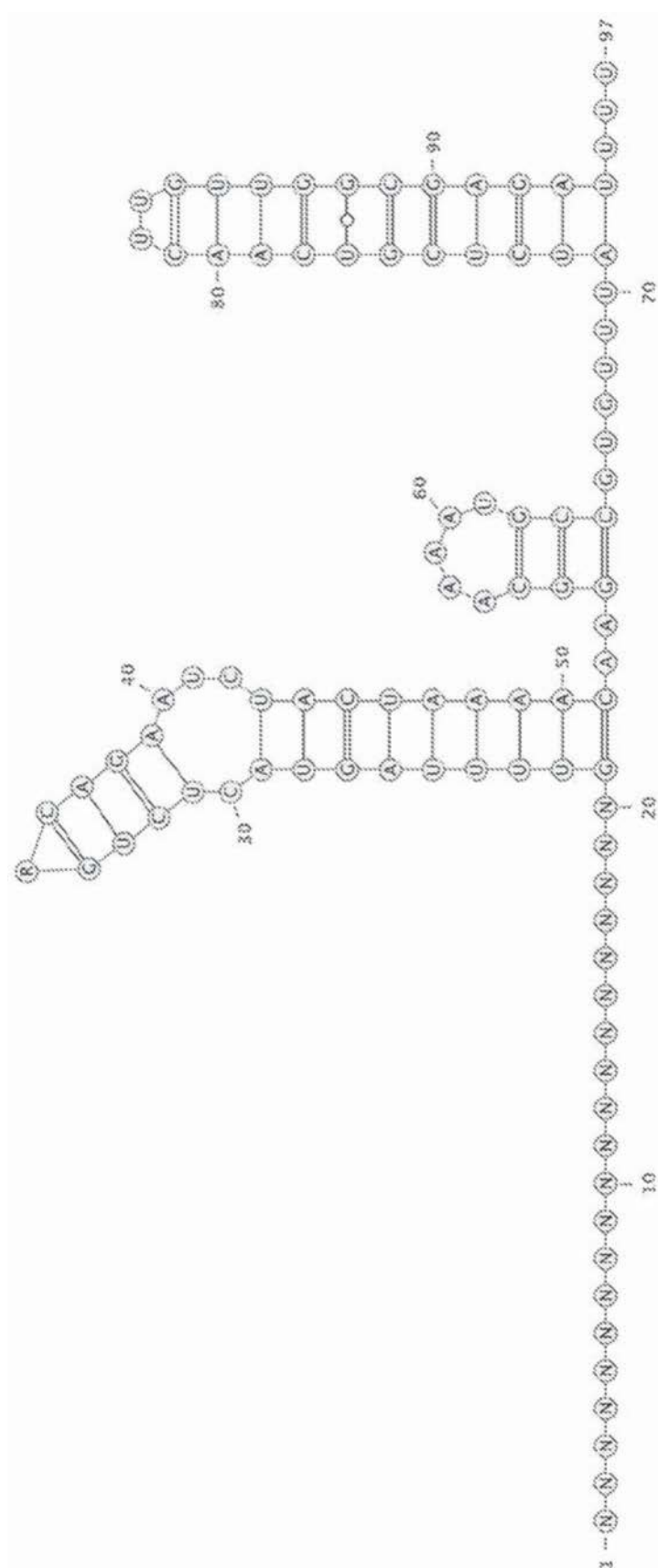


图10A

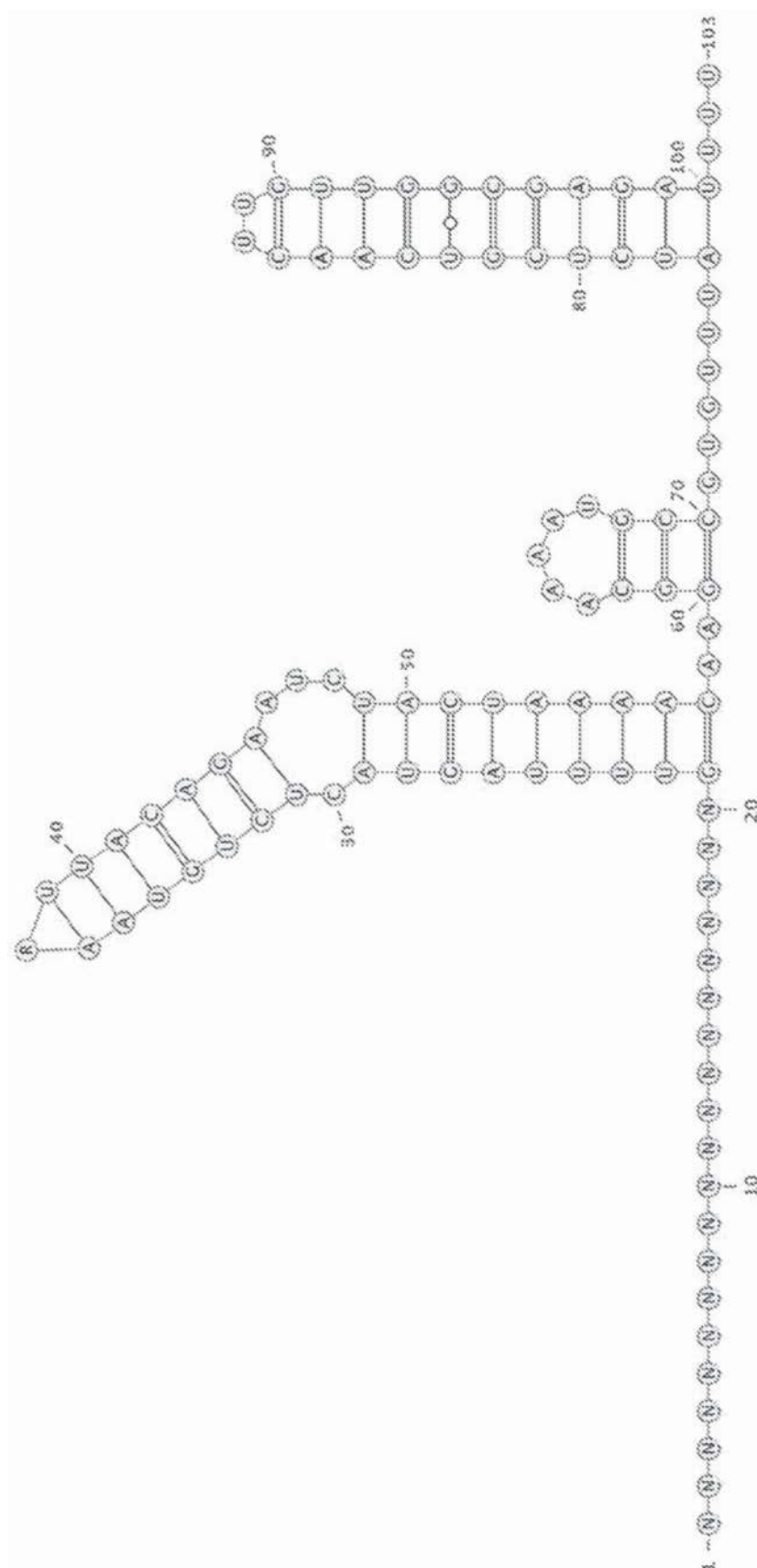


图10B

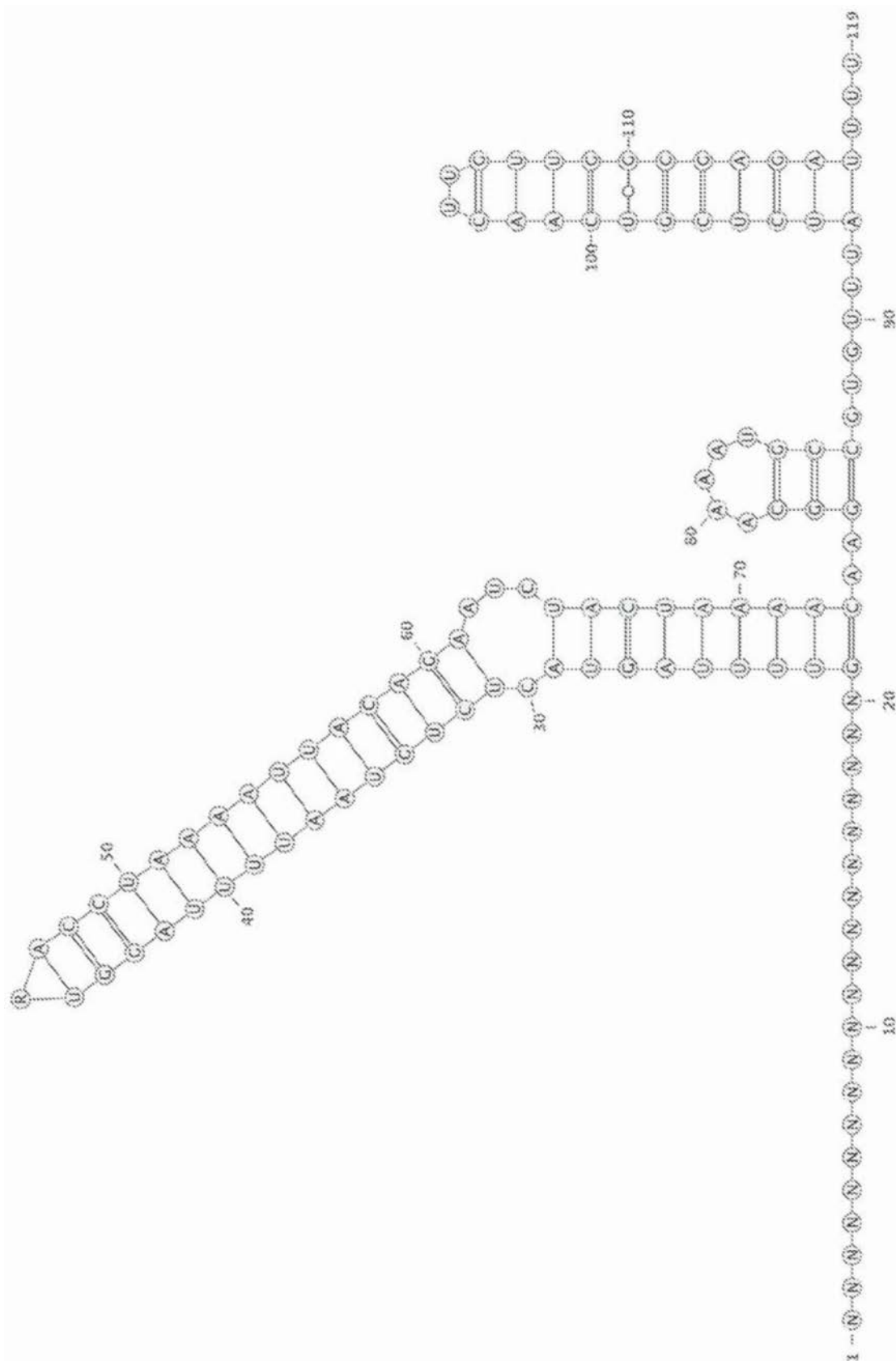


图10C

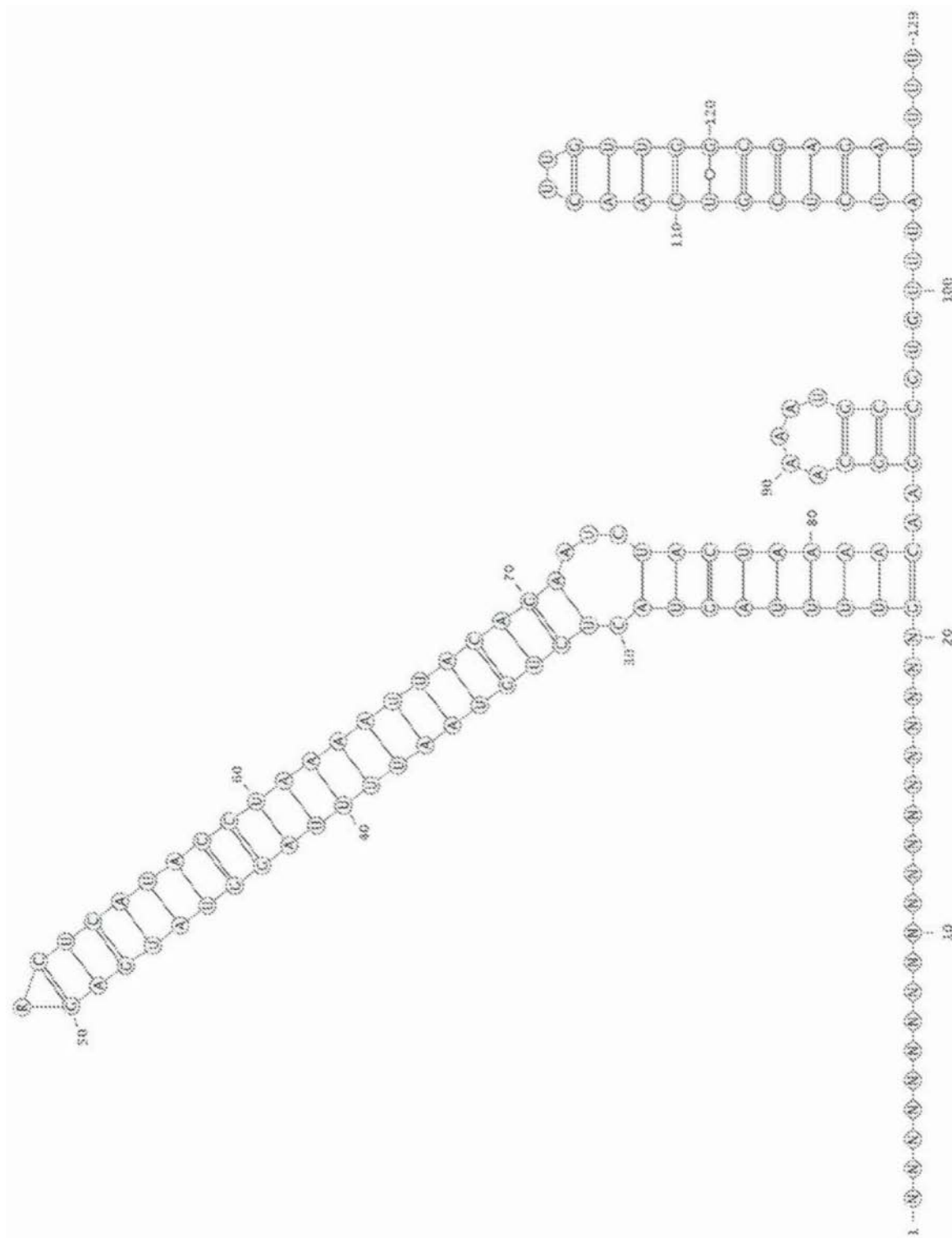


图10D

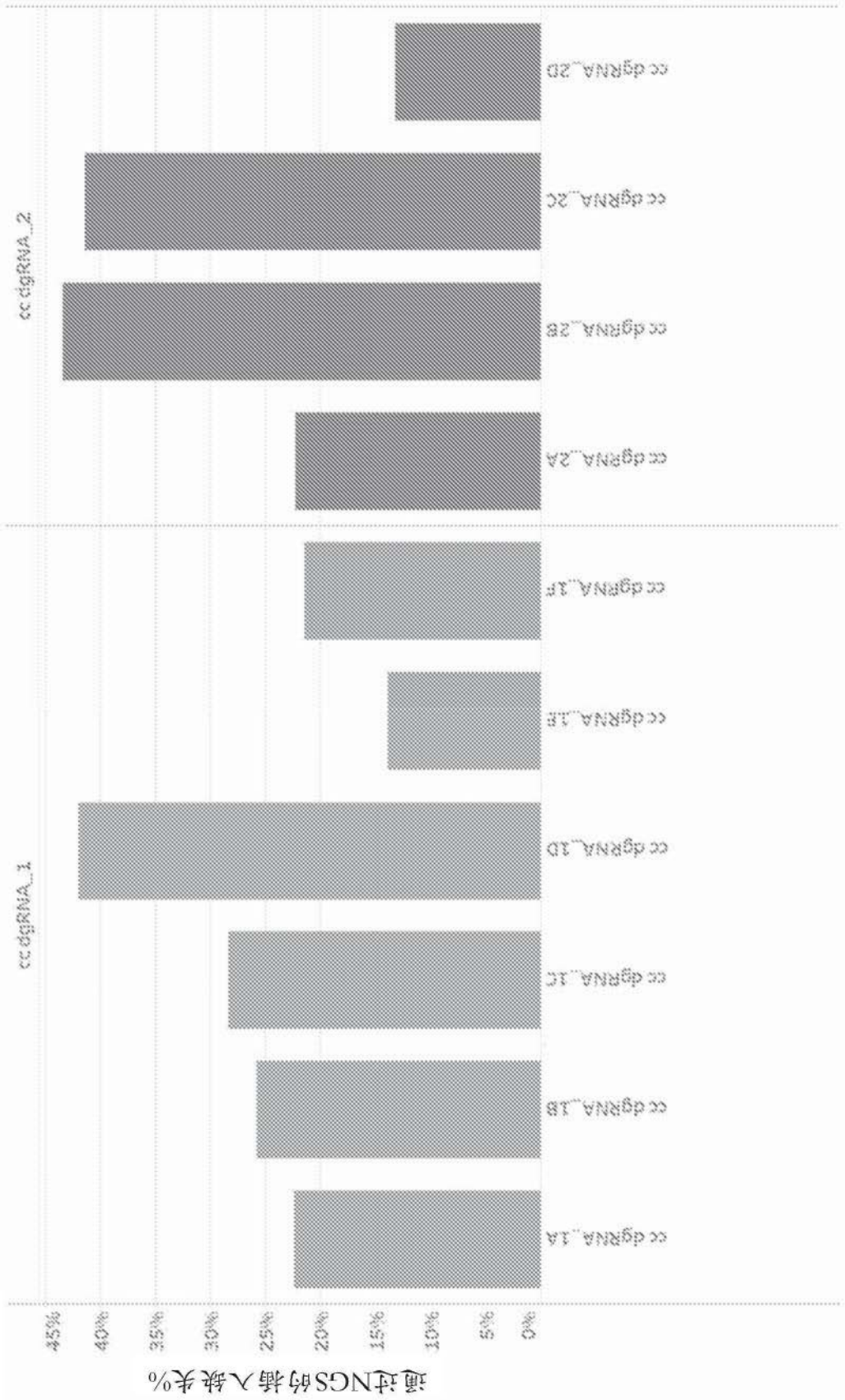


图11

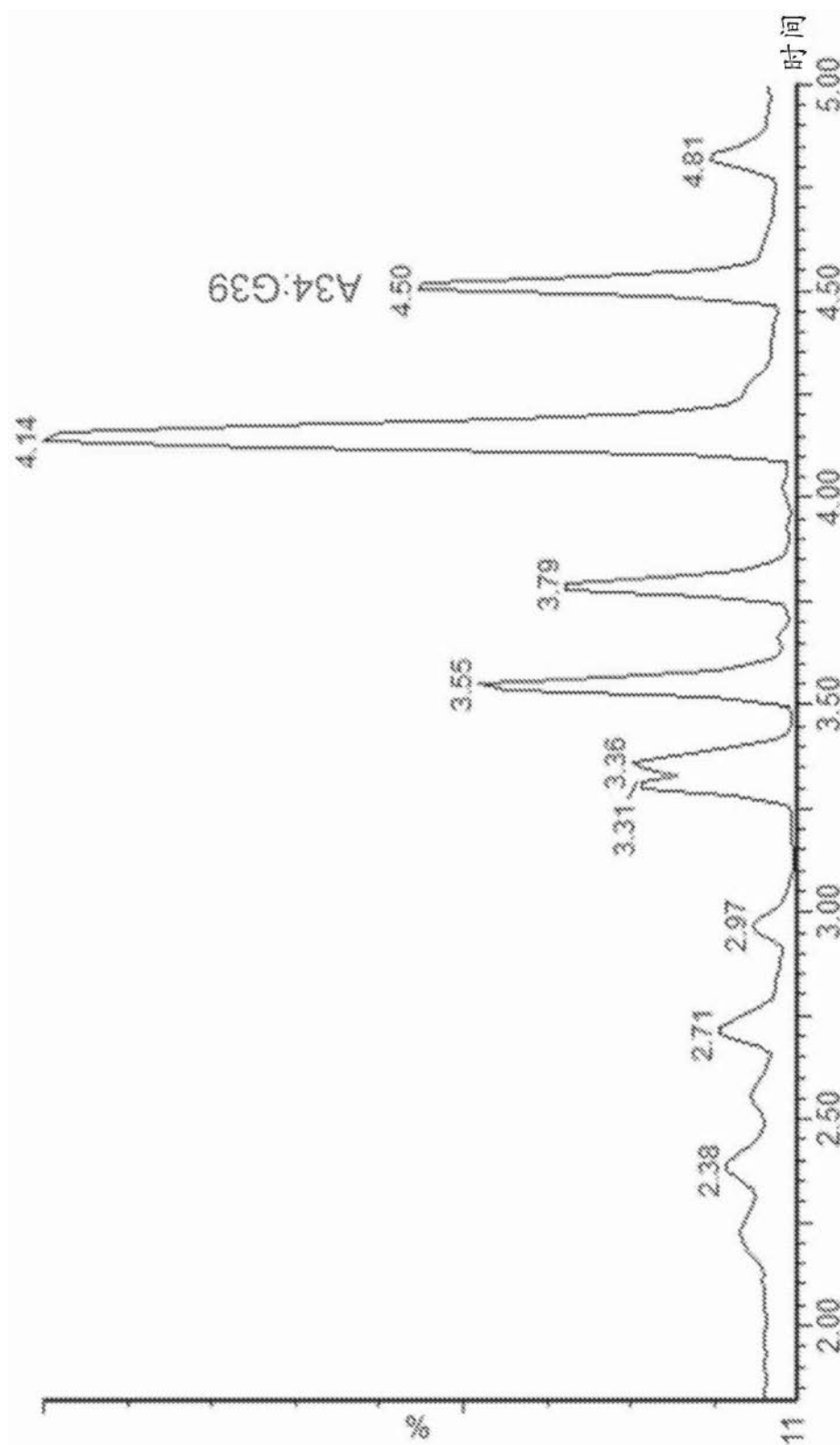


图12A

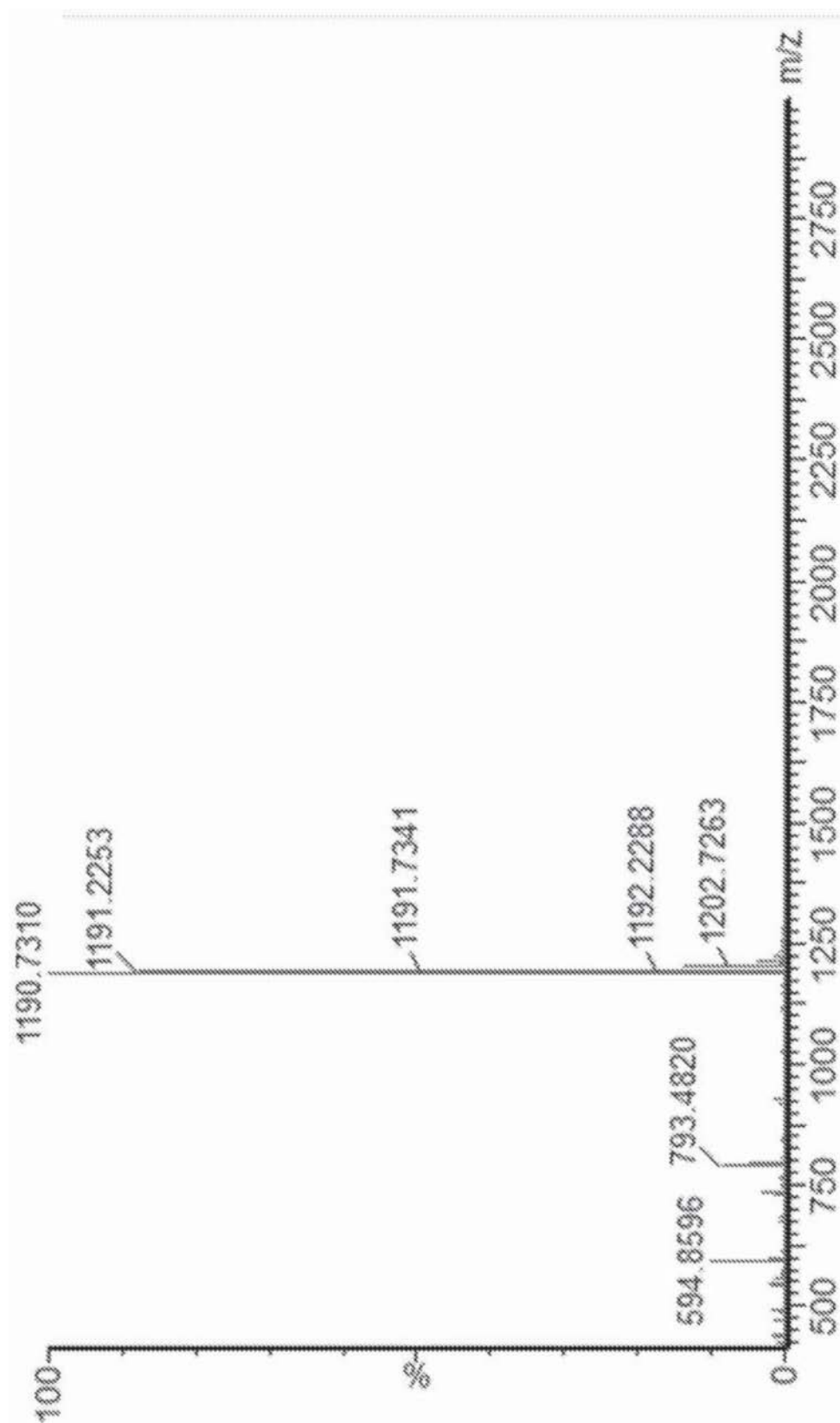


图12B

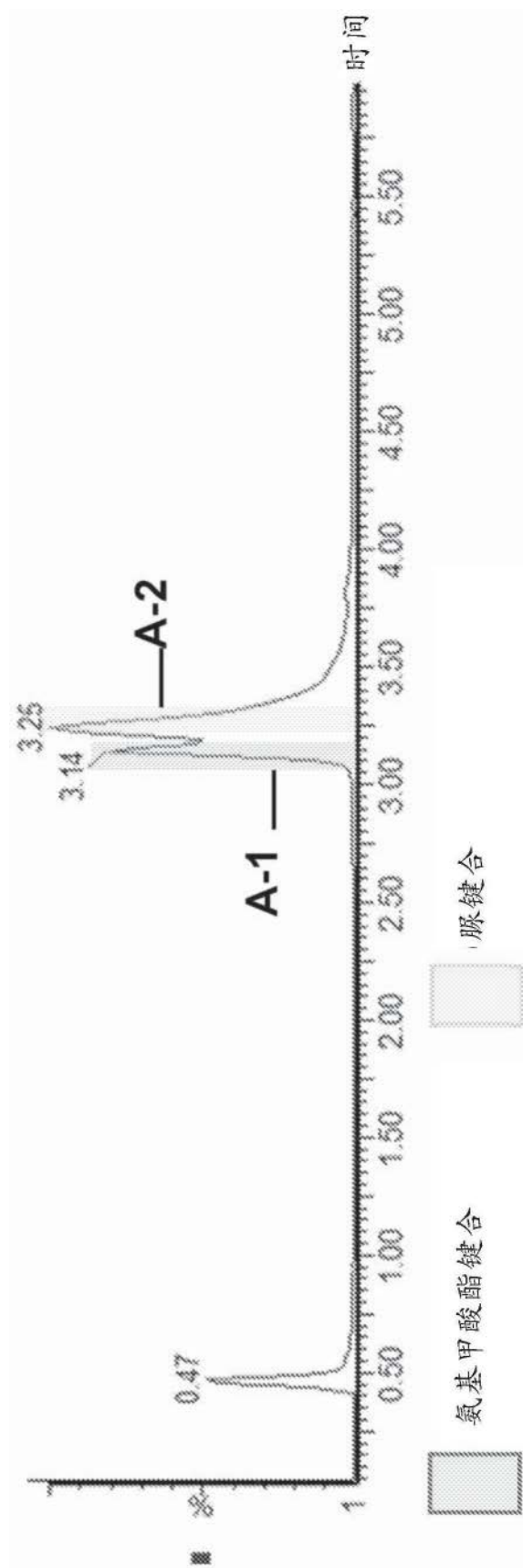


图13A

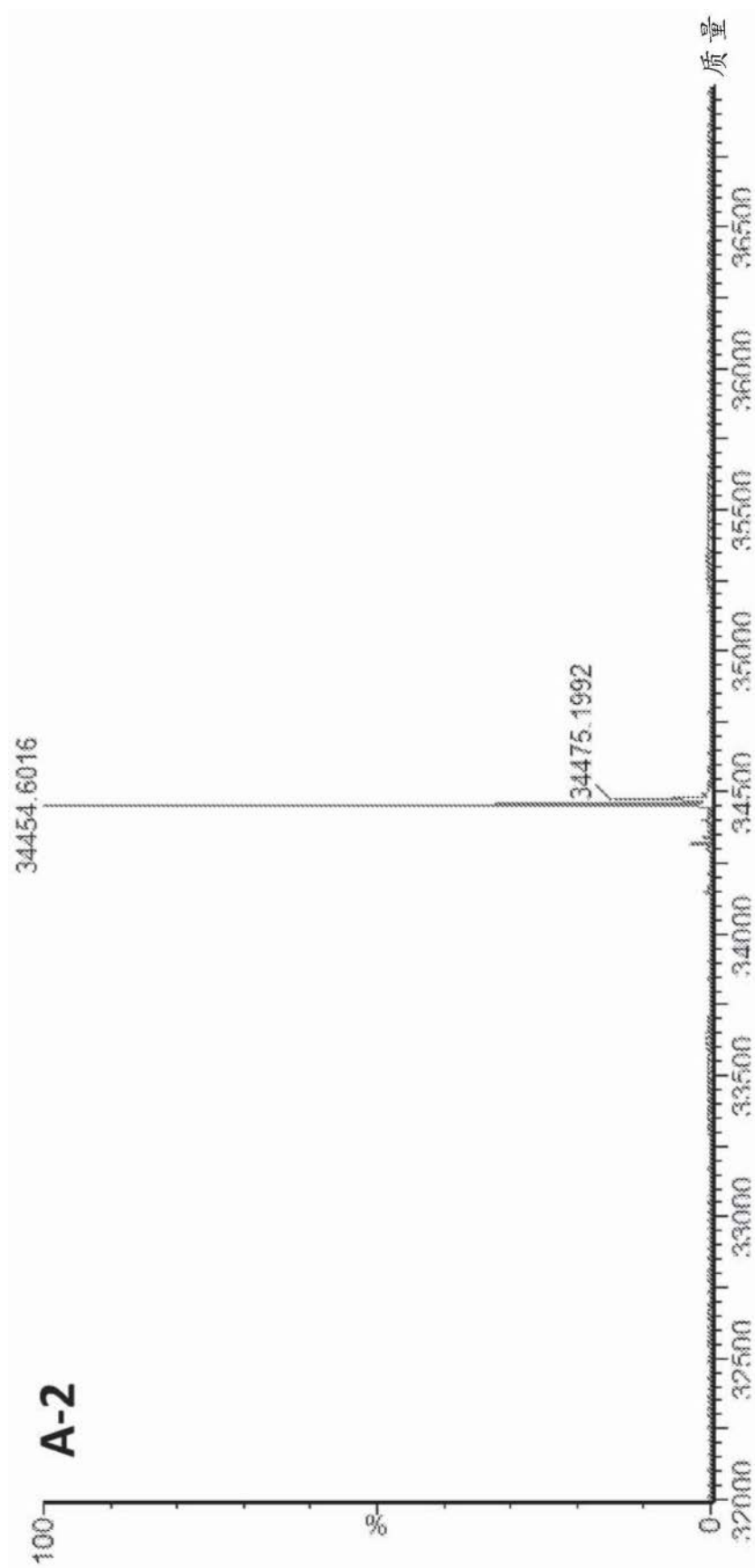


图13B

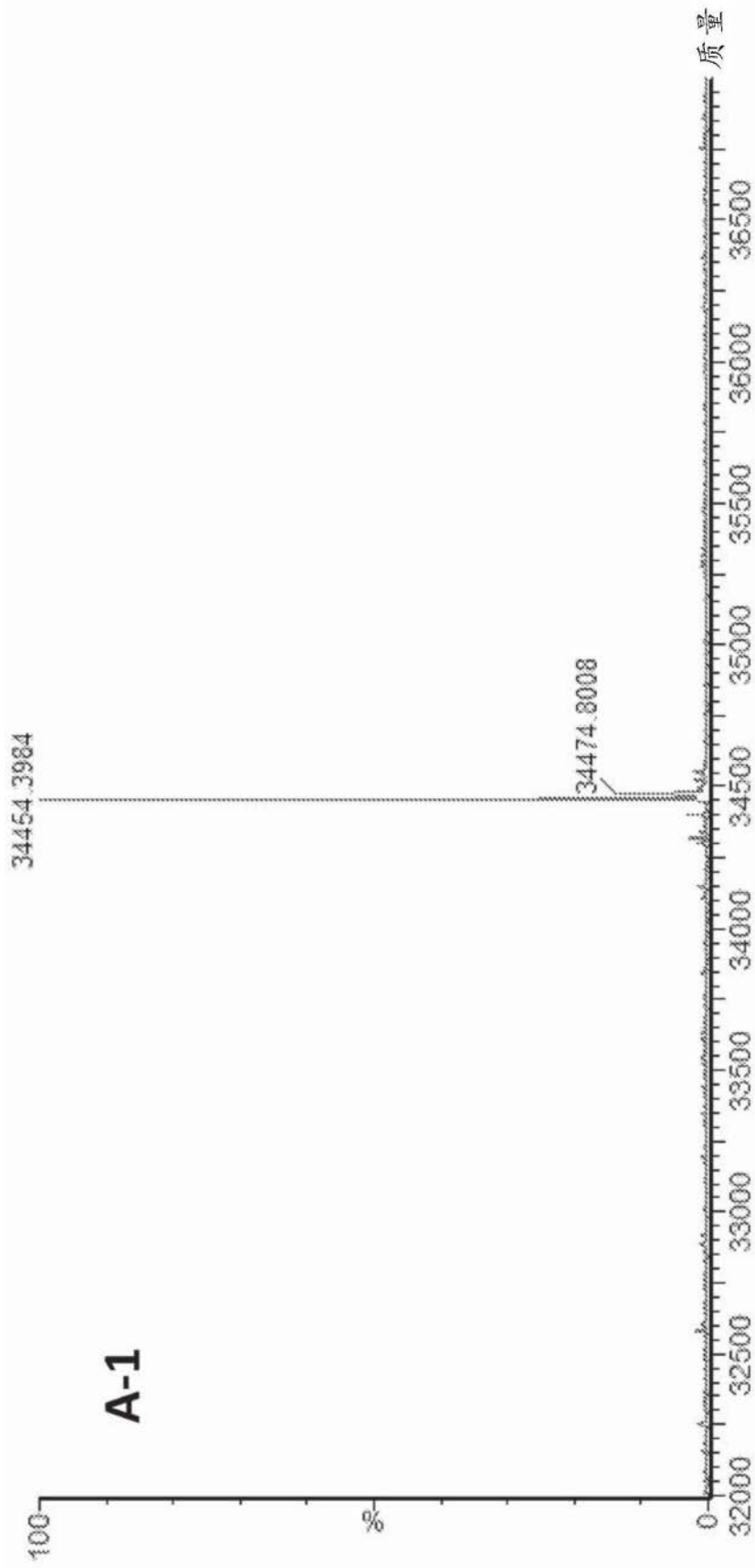


图13C

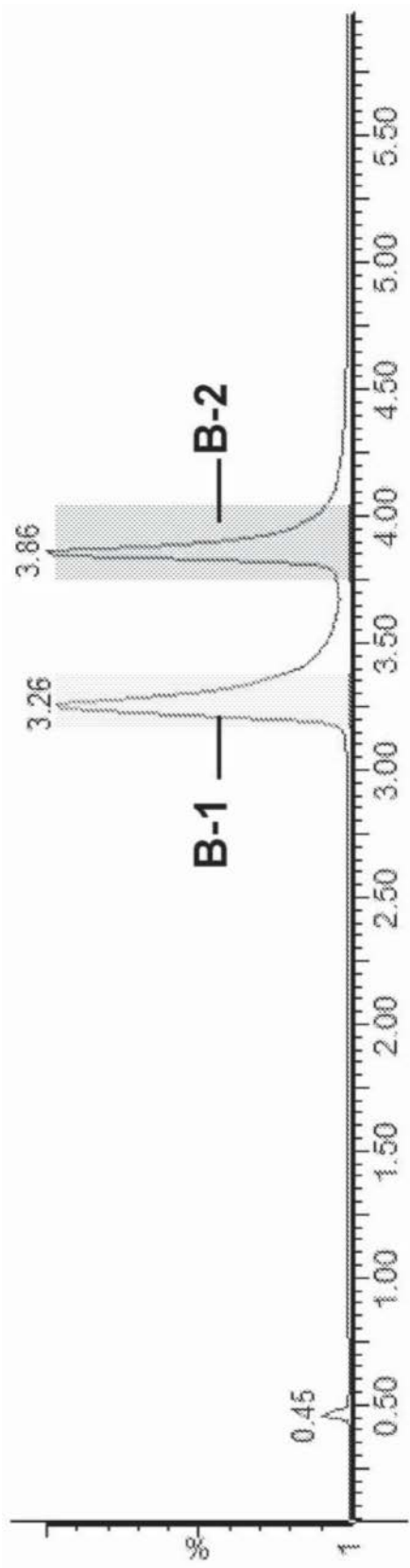


图14A

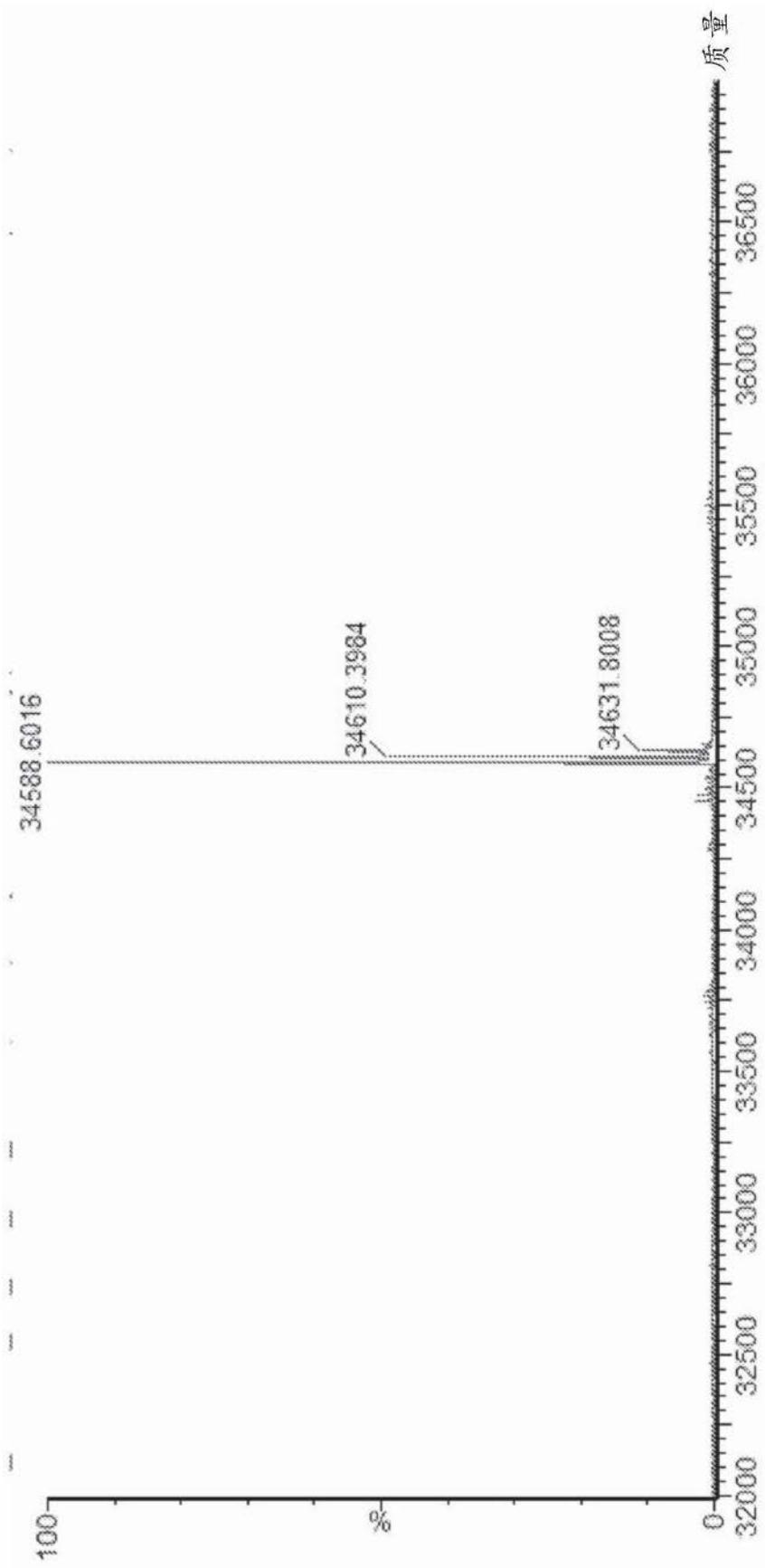


图14B

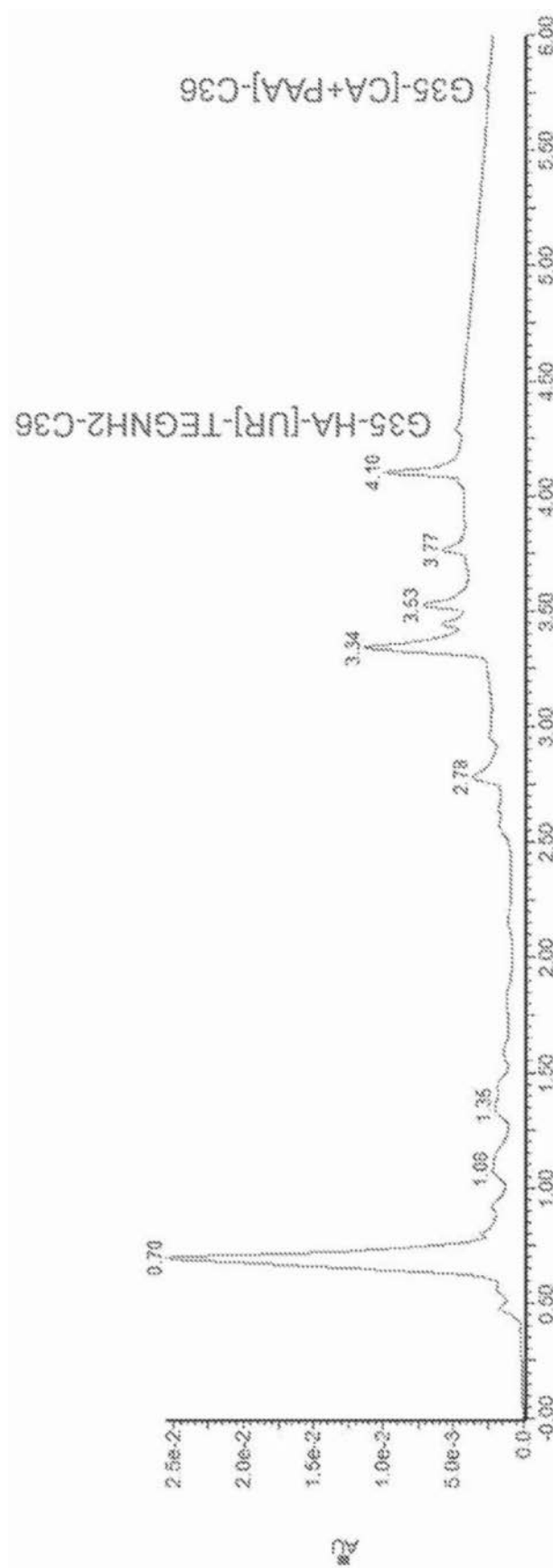


图15A



图15B

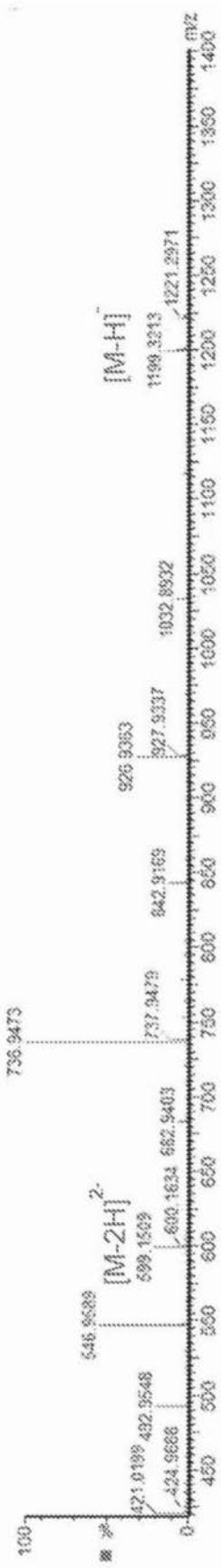


图15C



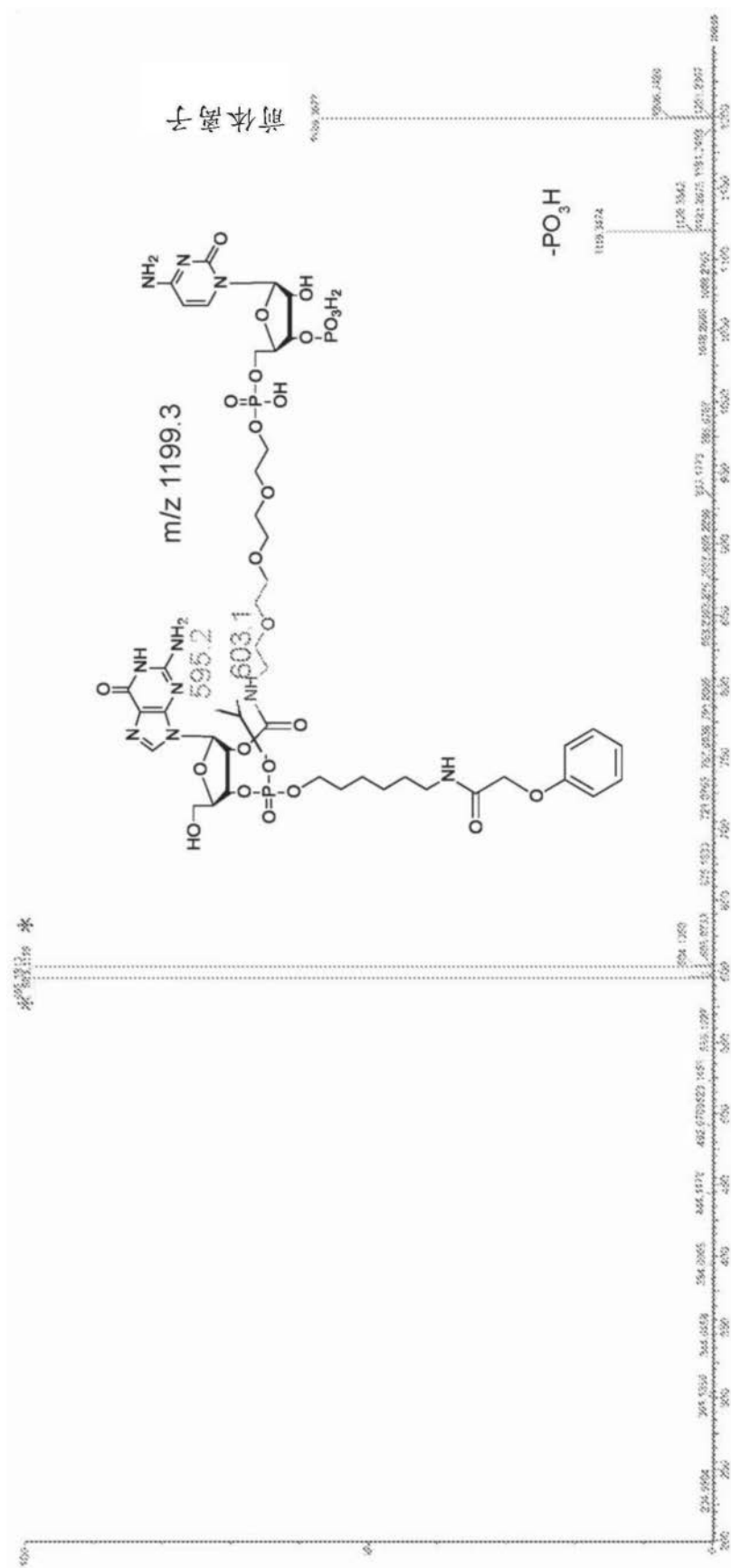


图15E

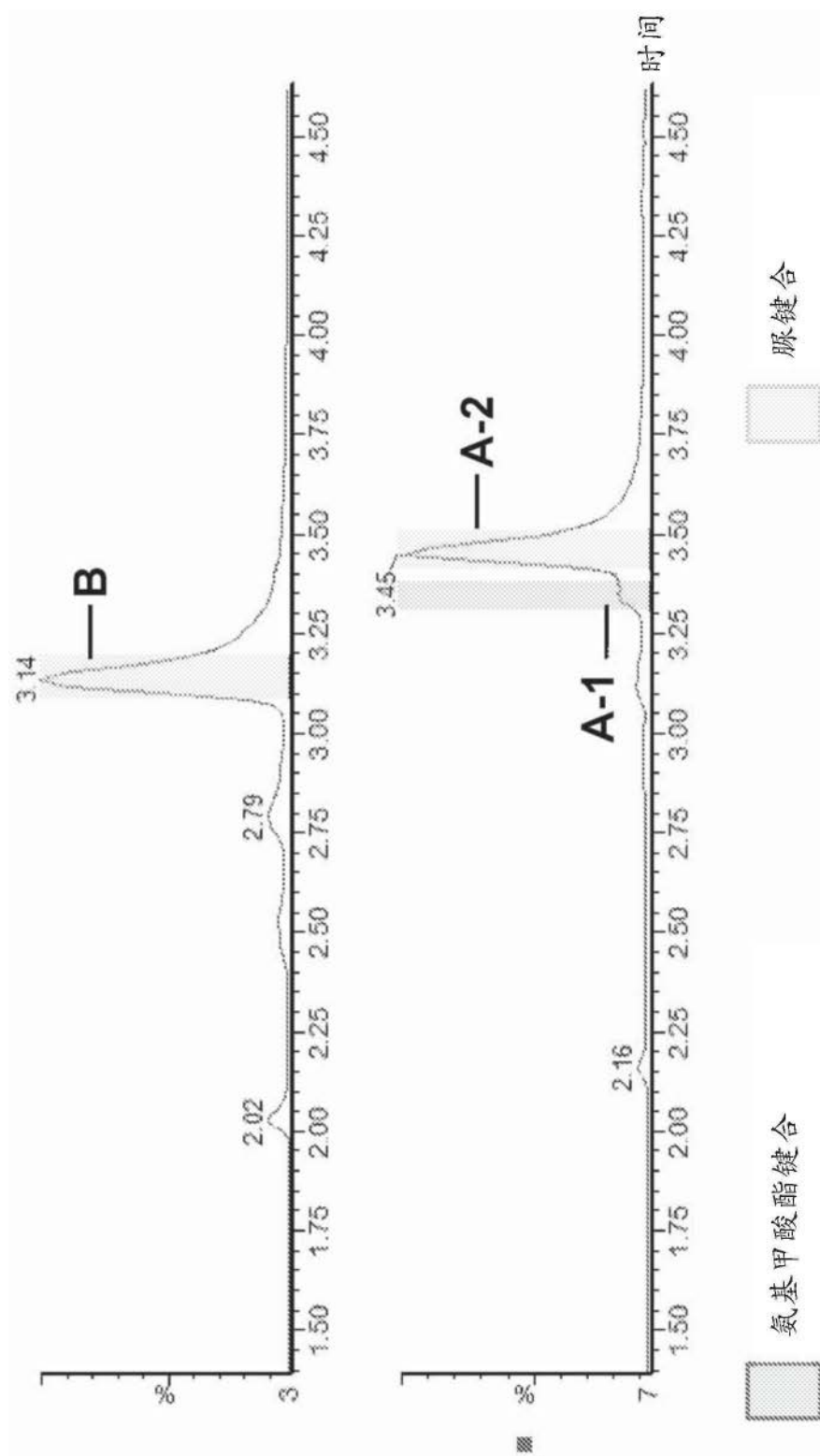


图16A

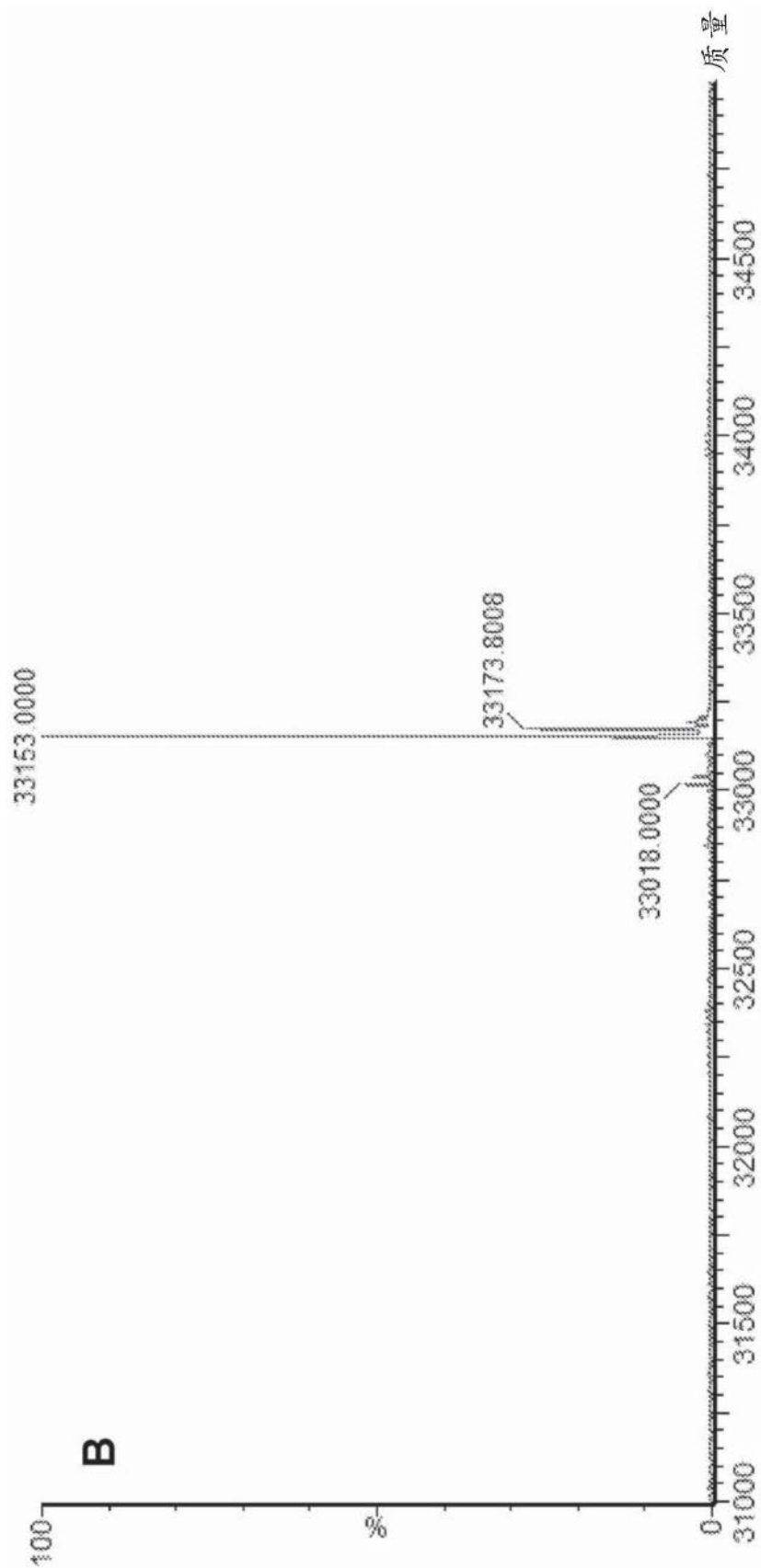


图16B

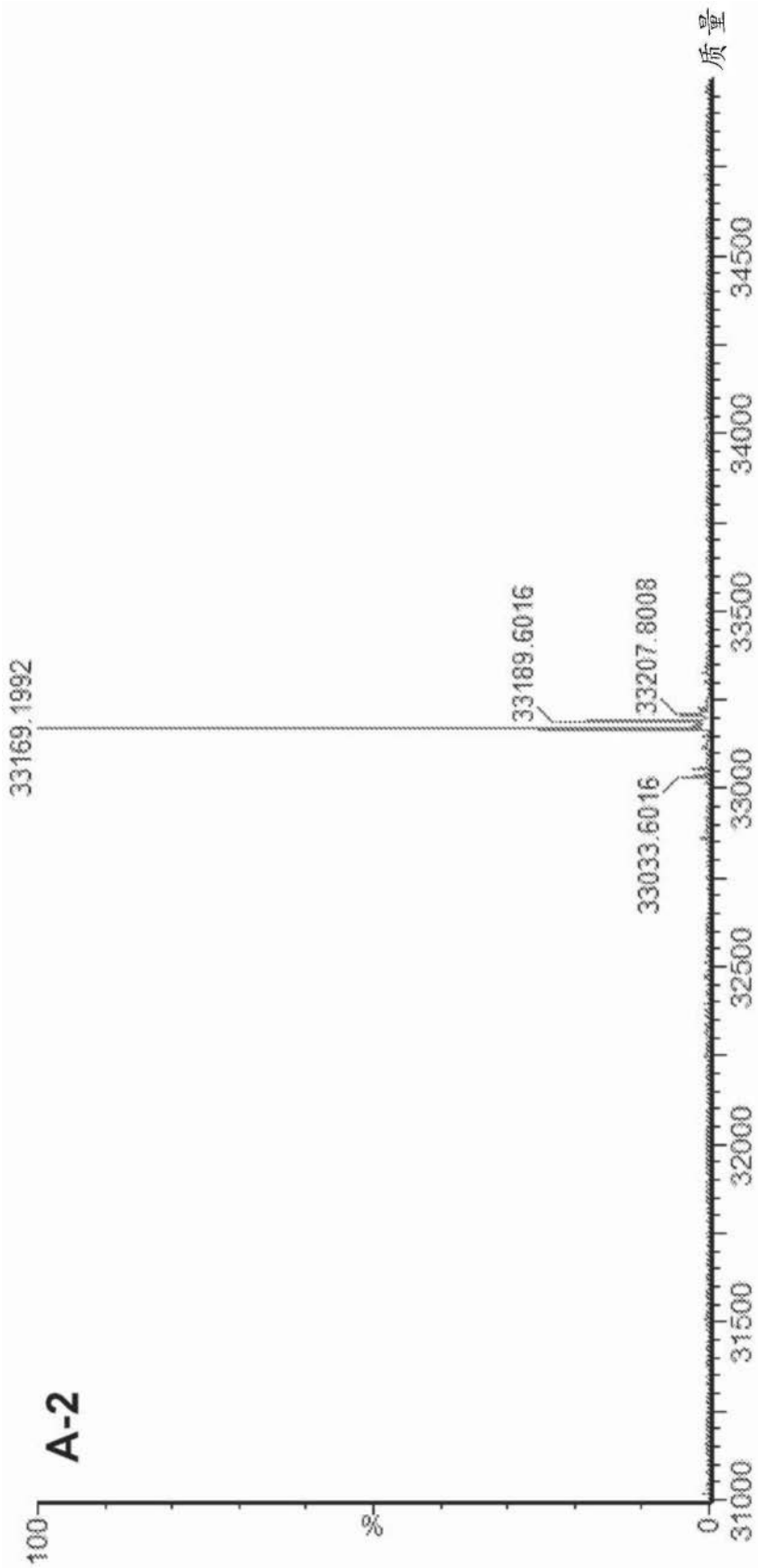


图16C

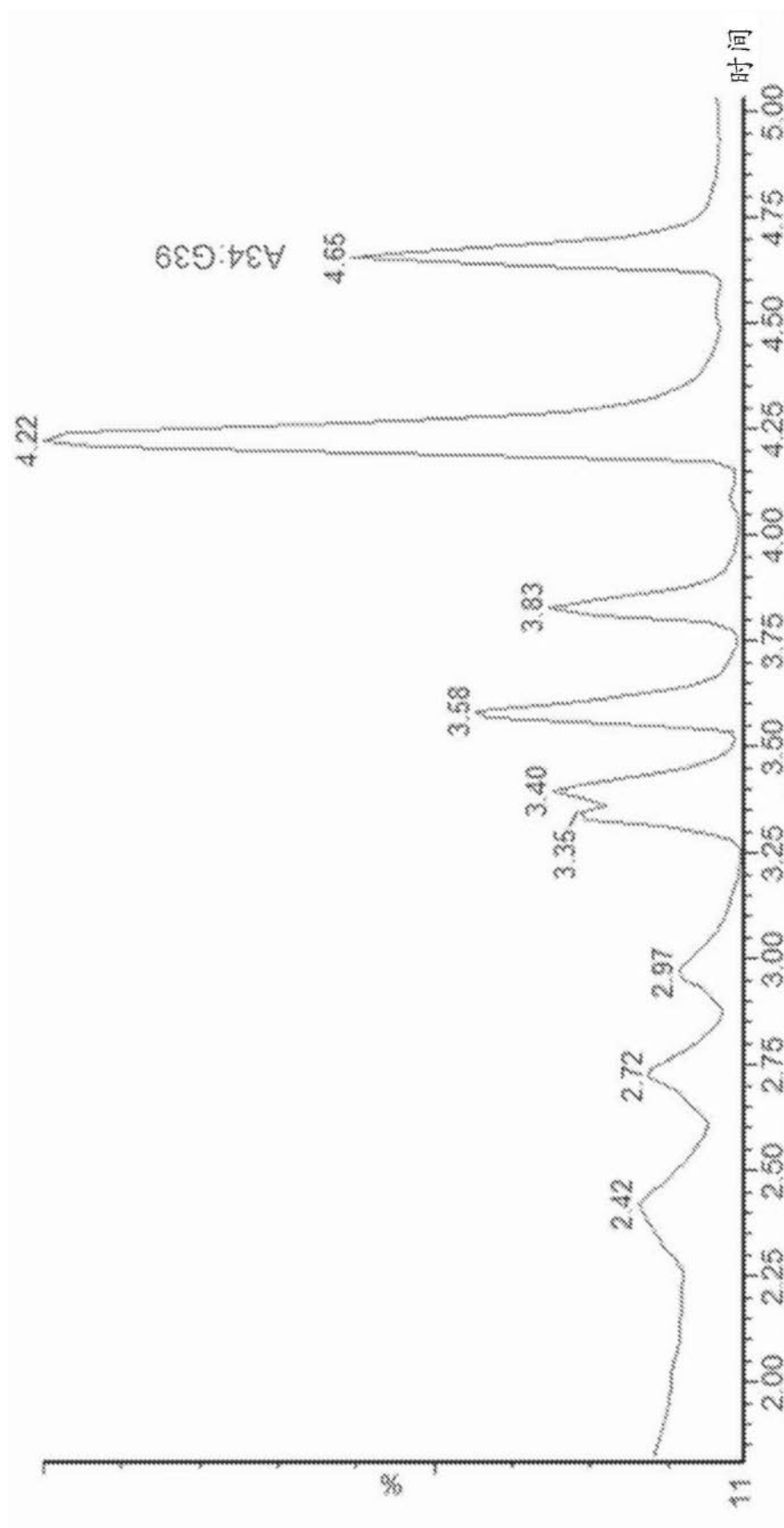


图17A

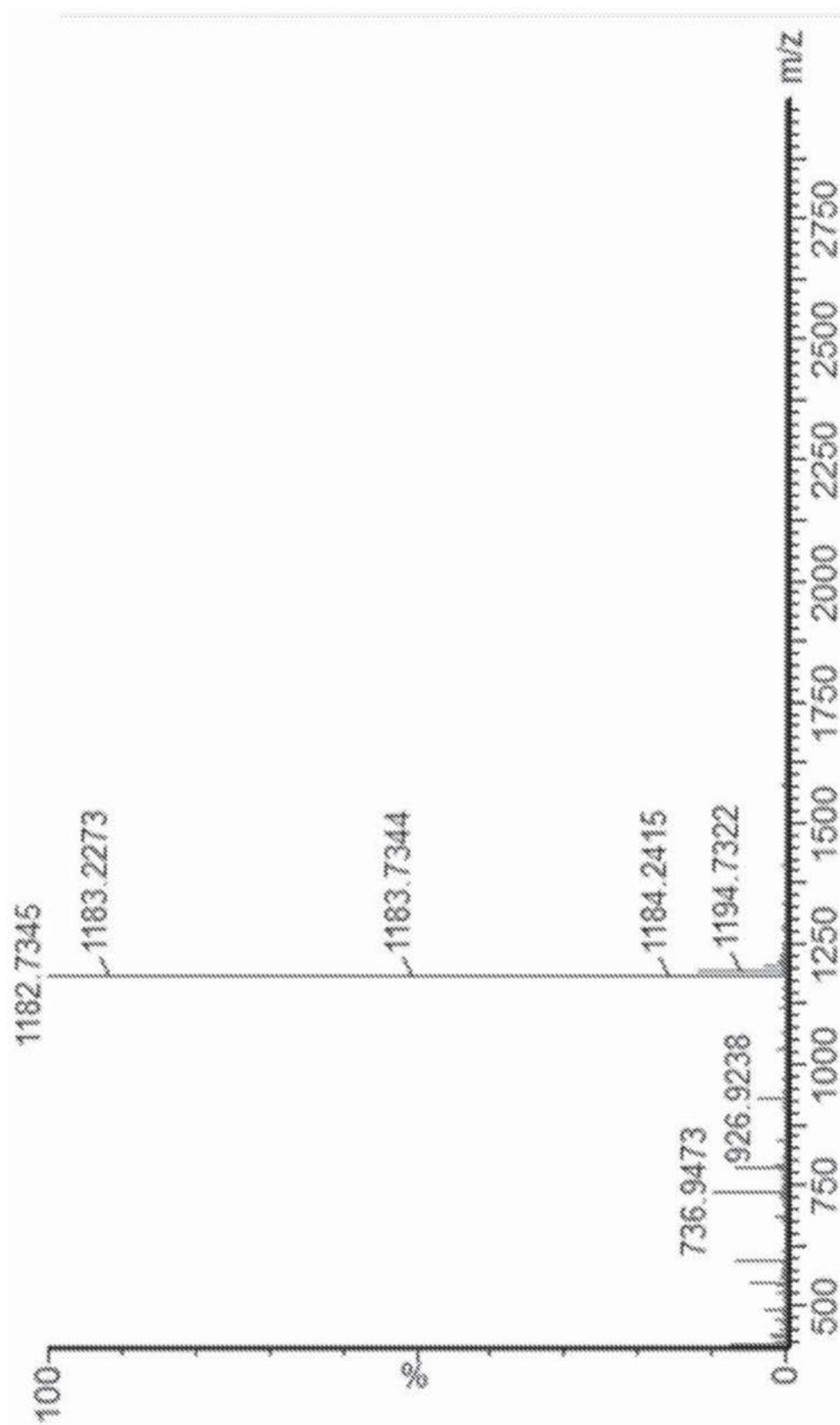


图17B