



(12) 发明专利

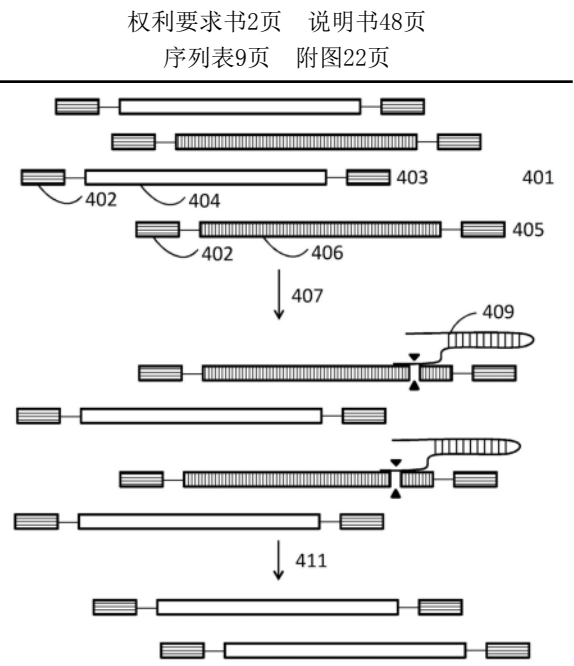
(10) 授权公告号 CN 110892485 B

(45) 授权公告日 2024. 03. 22

(21) 申请号 201880026589.9
(22) 申请日 2018.02.22
(65) 同一申请的已公布的文献号
 申请公布号 CN 110892485 A
(43) 申请公布日 2020.03.17
(30) 优先权数据
 62/462,284 2017.02.22 US
(85) PCT国际申请进入国家阶段日
 2019.10.22
(86) PCT国际申请的申请数据
 PCT/US2018/019268 2018.02.22
(87) PCT国际申请的公布数据
 W02018/156792 EN 2018.08.30
(73) 专利权人 特韦斯特生物科学公司
 地址 美国加利福尼亚州
(72) 发明人 比尔·詹姆士·佩克
 拉姆齐·易卜拉欣·泽图恩
(74) 专利代理机构 北京安信方达知识产权代理
 有限公司 11262
 专利代理师 韦昌金 郑霞

(51) Int.Cl.
 G16B 50/40 (2019.01)
 B82Y 10/00 (2011.01)
 G11C 13/00 (2006.01)
(56) 对比文件
 CN 105119717 A, 2015.12.02
 CN 105838711 A, 2016.08.10
 US 6013445 A, 2000.01.11
 WO 2013023220 A2, 2013.02.14
 US 2005053968 A1, 2005.03.10
 CN 104734848 A, 2015.06.24
 US 2015269313 A1, 2015.09.24
 WO 2016164779 A1, 2016.10.13
 CN 101495654 A, 2009.07.29
 Tyler D.P. Brunet. Aims and methods of
 biosteganography.《JOURNAL OF
 BIOTECHNOLOGY》.2016,第226卷第56-64页.(续)
 审查员 祝丽业

(54) 发明名称
 基于核酸的数据存储
(57) 摘要
 本文提供了用于生成和使用安全的基于生物分子的信息以供存储的组合物、装置、系统和方法。本文进一步描述了用于信息的生物加密或生物解密的组合物、装置、系统和方法。数字序列向基于核酸的序列的转化包括选择一种或多种生物加密方法的步骤。



CN 110892485 B

[接上页]

(56) 对比文件

AISLING O' DRISCOLL ET AL.Synthetic DNA : The next generation of big data storage.《BIOENGINEERED》.2013,第4卷(第3期),第123-125页.

SETH L. SHIPMAN ET AL.Molecular recordings by directed CRISPR spacer acquisition.SCIENCE.2016,第353卷(第6298期),第1-16页.

GEETHA V ET AL.Survey on security mechanisms for public cloud data.《2016 INTERNATIONAL CONFERENCE ON EMERGING TRENDS IN ENGINEERING》.2016,第1-8页.

YAZDI S M HOSSEIN TABATABAEI ET AL.DNA-Based Storage: Trends and Methods.《IEEE TRANSACTIONS ON MOLECULAR,

BIOLOGICAL AND MULTI-SCALE

COMMUNICATIONS》.2016,第1卷(第3期),第230-248页.

PAVANI YASHODHA DE SILVA ET AL.New Trends of Digital Data Storage in DNA.《BIOMED RESEARCH INTERNATIONAL》.2016,第2016卷第1-14页.

Dvorsky.Living bacteria can now store data.《<https://gizmodo.com/living-bacteria-can-now-store-data-1781773517>》.2016,第1-3页.

Leroy Hood.The digital code of DAN.《Nature》.2003,第421卷第444-448页.

FRANCIS CRICK.Central dogma of molecular biology.《Nature》.1970,第227卷第561-563页.

1. 一种用于存储信息的方法,该方法包括:
 - a) 以至少一个数字序列的形式通过计算机系统接收至少一项信息;
 - b) 通过计算机系统接收关于酶生物加密的指令;
 - c) 基于所述指令通过所述计算机系统将所述至少一个数字序列转化为多个寡核苷酸的数字序列,其中所述多个寡核苷酸的数字序列配置为用于通过将多个寡核苷酸暴露于核酸酶复合物来解密;
 - d) 在合成仪中合成所述多个寡核苷酸;以及
 - e) 存储所述多个寡核苷酸。
2. 根据权利要求1所述的方法,其中所述核酸酶复合物包括Cas核酸酶、锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶、Argonaute核酸酶、S1核酸酶或绿豆核酸酶。
3. 根据权利要求1所述的方法,其中所述核酸酶复合物包括DNA酶。
4. 根据权利要求1所述的方法,其中所述酶生物加密包括基于CRISPR/Cas的生物加密。
5. 根据权利要求1所述的方法,其中所述酶生物加密包括关于合成对选自以下的酶敏感的寡核苷酸的指令: AclI、HindIII、SspI、MluCI Tsp509I、PciI、AgeI、BspMI BfuAI、SexAI、MluI、BceAI、HpyCH4IV、HpyCH4III、BaeI、BsaXI、AflIII、SpeI、BsrI、BmrI、BglII、AfeI、AluI、StuI、ScaI、ClaI BspDI、PI-SceI、NsiI、AseI、SwaI、CspCI、MfeI、BssSI BssSαI、Nb.BssSI、BmgBI、PmlI、DraIII、AleI、EcoP15I、PvuII、AlwNI、BtsIMutI、TspRI、NdeI、NlaIII、CviAII、FatI、MslI、FspEI、XcmI、BstXI、PflMI、BccI、NcoI、BseYI、FauI、SmaI、XmaI TspMI、Nt.CviPII、LpnPI、AciI、SacII、BsrBI、MspI HpaII、ScrFI、BssKI StyD4I、BsaJI、BslI、BtgI、NciI、AvrII、MnlI、BbvCI、Nb.BbvCI、Nt.BbvCI、SbfI、Bpu10I、Bsu36I、EcoNI、HpyAV、BstNI、PspGI、StyI、BcgI、PvuI、BstUI、EagI、RsrII、BsiEI、BsiWI、BsmBI、Hpy99I、MspAII、MspJI、SgrAI、BfaI、BspCNI、XhoIPaeR7I TliI、EarI、AcuI、PstI、BpmI、DdeI、SfcI、AflII、BpuEI、SmlI、AvaI BsoBI、MboII、BbsI、XmnI、BsmI、Nb.BsmI、EcoRI、HgaI、AatII、ZraI、Tth111IPf1FI、PshAI、AhdI、DrdI、Eco53kI、SacI、BseRI、PleI、Nt.BstNBI、MlyI、HinfI、EcoRV、MboI Sau3AIDpnII BfuCI、DpnI、BsaBI、TfiI、BsrDI、Nb.BsrDI、BbvI、BtsI BtsαI、Nb.BtsI、BstAPI、SfaNI、SphI、SrfI、NmeAIII、NaeI、NgoMIV、BglI、AsiSI、BtgZI、HinPII、HhaI、BssHII、NotI、Fnu4HI、Cac8I、MwoI、NheI、BmtI、SapI BspQI、Nt.BspQI、BlpI、TseI ApeKI、Bsp1286I、AlwI、Nt.AlwI、BamHI、FokI、BtsCI、HaeIIIPhoI、FseI、SfiI、NarI、KasI、SfoI、PluTI、AscI、EciI、BsmFI、ApaI、Psp0MI、Sau96I、NlaIV、KpnI、Acc65I、BsaI、HphI、BstEII、AvaII、BanI、BaeGI、BsaHI、BanII、RsaI、CviQI、BstZ17I、BciVI、SalI、Nt.BsmAI、BsmAI BcoDI、ApaLI、BsgI、AccI、Hpy166II、Tsp45I、HpaI、PmeI、HincII、BsiHKAI、ApoI、NspI、BsrFI BsrFαI、BstYI、HaeII、CviKI-1、Eco0109I、PpuMI、I-CeuI、SnaBI、I-SceI、BspHI、BspEI、MmeI、TaqαI、NruI、Hpy188I、Hpy188III、XbaI、BclI、HpyCH4V、FspI、PI-PspI、MscI、BsrGI、MseI、PacI、PsiI、BstBI、DraI、PspXI、BsaWI、BsaAI和EaeI。
6. 根据权利要求1所述的方法,其中所述多个寡核苷酸包括至少100,000个寡核苷酸。
7. 根据权利要求1所述的方法,其中所述多个寡核苷酸包括至少100亿个寡核苷酸。
8. 一种用于存储信息的系统,该系统包含:
 - a) 接收单元,其用于接收关于至少一个数字序列的形式的至少一项信息的机器指令,以及关于酶生物加密的机器指令,其中所述酶生物加密配置为用于通过将多个寡核苷酸暴

露于核酸酶复合物来解密;

b) 处理器单元,其用于基于所述指令将所述至少一个数字序列自动转化为多个寡核苷酸的数字序列;

c) 合成仪单元,其用于从所述处理器单元接收关于合成所述多个寡核苷酸的机器指令;以及

d) 存储单元,其用于接收从所述合成仪单元沉积的所述多个寡核苷酸。

9. 根据权利要求8所述的系统,其中所述核酸酶复合物包括Cas核酸酶、锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶、Argonaute核酸酶、S1核酸酶或绿豆核酸酶。

10. 根据权利要求8所述的系统,其中所述核酸酶复合物包括DNA酶。

11. 根据权利要求8所述的系统,其中所述酶生物加密包括基于CRISPR/Cas的生物加密。

12. 根据权利要求8所述的系统,其中所述酶生物加密包括关于合成对选自以下的酶敏感的寡核苷酸的机器指令: AclI、HindIII、SspI、MluCI Tsp509I、PciI、AgeI、BspMI BfuAI、SexAI、MluI、BceAI、HpyCH4IV、HpyCH4III、BaeI、BsaXI、AflIII、SpeI、BsrI、BmrI、BglII、AfeI、AluI、StuI、ScaI、ClaI BspDI、PI-SceI、NsiI、AseI、SwaI、CspCI、MfeI、BssSI BssSαI、Nb.BssSI、BmgBI、PmlI、DraIII、AleI、EcoP15I、PvuII、AlwNI、BtsIMutI、TspRI、NdeI、NlaIII、CviAII、FatI、MslI、FspEI、XcmI、BstXI、PflMI、BccI、NcoI、BseYI、FauI、SmaI、XmaI TspMI、Nt.CviPII、LpnPI、AciI、SacII、BsrBI、MspI HpaII、ScrFI、BssKI StyD4I、BsaJI、BslI、BtgI、NciI、AvrII、MnlI、BbvCI、Nb.BbvCI、Nt.BbvCI、SbfI、Bpu10I、Bsu36I、EcoNI、HpyAV、BstNI、PspGI、StyI、BcgI、PvuI、BstUI、EagI、RsrII、BsiEI、BsiWI、BsmBI、Hpy99I、MspAII、MspJI、SgrAI、BfaI、BspCNI、XhoIPaeR7I TliI、EarI、AcuI、PstI、BpmI、DdeI、SfcI、AflII、BpuEI、SmlI、AvaI BsoBI、MboII、BbsI、XmnI、BsmI、Nb.BsmI、EcoRI、HgaI、AatII、ZraI、Tth111IPf1FI、PshAI、AhdI、DrdI、Eco53kI、SacI、BseRI、PleI、Nt.BstNBI、MlyI、HinfI、EcoRV、MboI Sau3AIDpnII BfuCI、DpnI、BsaBI、TfiI、BsrDI、Nb.BsrDI、BbvI、BtsI BtsαI、Nb.BtsI、BstAPI、SfaNI、SphI、SrfI、NmeAIII、NaeI、NgoMIV、BglI、AsiSI、BtgZI、HinPII、HhaI、BssHII、NotI、Fnu4HI、Cac8I、MwoI、NheI、BmtI、SapI BspQI、Nt.BspQI、BlpI、TseI ApeKI、Bsp1286I、AlwI、Nt.AlwI、BamHI、FokI、BtsCI、HaeIIIPhoI、FseI、SfiI、NarI、KasI、SfoI、PluTI、AscI、EciI、BsmFI、ApaI、Psp0MI、Sau96I、NlaIV、KpnI、Acc65I、BsaI、HphI、BstEII、AvaII、BanI、BaeGI、BsaHI、BanII、RsaI、CviQI、BstZ17I、BciVI、SalI、Nt.BsmAI、BsmAI BcoDI、ApaLI、BsgI、AccI、Hpy166II、Tsp45I、HpaI、PmeI、HincII、BsiHKAI、ApoI、NspI、BsrFI BsrFαI、BstYI、HaeII、CviKI-1、Eco0109I、PpuMI、I-CeuI、SnaBI、I-SceI、BspHI、BspEI、MmeI、TaqαI、NruI、Hpy188I、Hpy188III、XbaI、BclI、HpyCH4V、FspI、PI-PspI、MscI、BsrGI、MseI、PacI、PsiI、BstBI、DraI、PspXI、BsaWI、BsaAI和EaeI。

13. 根据权利要求8所述的系统,其中所述多个寡核苷酸包括至少100,000个寡核苷酸。

14. 根据权利要求8所述的系统,其中所述多个寡核苷酸包括至少100亿个寡核苷酸。

基于核酸的数据存储

[0001] 交叉引用

[0002] 本申请要求2017年2月22日提交的第62/462,284号美国临时申请的权益,该临时申请通过引用整体并入本文。

[0003] 序列表

[0004] 本申请含有以ASCII格式电子提交的序列表,并且其通过引用整体并入本文。创建于2018年2月20日的所述ASCII副本被命名为44854-738_601_SL.txt,大小为8,636个字节。

背景技术

[0005] 基于生物分子的信息存储系统(例如,基于DNA)随着时间的推移具有大存储容量和稳定性。然而,对用于信息存储的生物分子的可扩展的、自动化的、高度准确且高效的系统存在需求。另外,对保护这类信息的安全性也存在需求。

[0006] 援引并入

[0007] 本说明书中所提及的所有出版物、专利和专利申请均通过引用而并入本文,其程度犹如具体地和个别地指出每一单独的出版物、专利或专利申请均通过引用而并入。

发明内容

[0008] 本文提供了一种用于存储信息的方法,该方法包括:(a)以至少一个数字序列的形式接收至少一项信息;(b)接收关于选择至少一种生物加密格式的指令,其中所述生物加密格式是基于酶、电磁、化学或亲和力的生物加密;(c)基于所选择的生物加密格式将所述至少一个数字序列转化为多个寡核苷酸序列;(d)合成编码所述寡核苷酸序列的多个寡核苷酸;以及(e)存储所述多个寡核苷酸。本文进一步提供了用于存储信息的方法,其中所述基于酶的生物加密包括基于CRISPR/Cas的生物加密。本文进一步提供了用于存储信息的方法,其中所述基于酶的生物加密包括关于合成对表1所示的酶敏感的寡核苷酸的指令。本文进一步提供了用于存储信息的方法,其中所述基于电磁的生物加密包括关于合成对约0.01nm至约400nm的电磁波长敏感的寡核苷酸的指令。本文进一步提供了用于存储信息的方法,其中所述基于化学的生物加密包括关于合成对气态氨或甲胺施用敏感的寡核苷酸的指令。本文进一步提供了用于存储信息的方法,其中所述基于亲和力的生物加密包括关于合成对序列标签或亲和标签敏感的寡核苷酸的指令。本文进一步提供了用于存储信息的方法,其中所述亲和标签是生物素、洋地黄毒苷、Ni-次氨基三乙酸、脱硫生物素、组氨酸、聚组氨酸、myc、血凝素(HA)、FLAG、荧光标签、串联亲和纯化(TAP)标签、谷胱甘肽S转移酶(GST)、多核苷酸、适体、抗原或抗体。本文进一步提供了用于存储信息的方法,其中使用2、3、4或5种生物加密格式。本文进一步提供了用于存储信息的方法,其中所述多个寡核苷酸包括至少100,000个寡核苷酸。本文进一步提供了用于存储信息的方法,其中所述多个寡核苷酸包括至少100亿个寡核苷酸。

[0009] 本文提供了用于检索信息的方法,该方法包括:(a)从表面释放多个寡核苷酸;(b)对所述多个寡核苷酸施加基于酶、电磁、化学或亲和力的解密;(c)富集所述多个寡核苷酸;

(d) 对从所述多个寡核苷酸富集的寡核苷酸进行测序,以生成核酸序列;以及(e) 将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。本文进一步提供了用于检索信息的方法,其中所述多个寡核苷酸的解密包括将CRISPR/Cas复合物施加于所述多个寡核苷酸。本文进一步提供了用于检索信息的方法,其中所述基于酶的解密包括施加表1所示的酶。本文进一步提供了用于检索信息的方法,其中所述基于电磁的解密包括施加约0.01nm至约400nm的波长。本文进一步提供了用于检索信息的方法,其中所述基于化学的解密包括施加气态氨或甲胺施用。本文进一步提供了用于检索信息的方法,其中所述基于亲和力的解密包括施加序列标签或亲和标签。本文进一步提供了用于检索信息的方法,其中所述亲和标签是生物素、洋地黄毒苷、Ni-次氨基三乙酸、脱硫生物素、组氨酸、聚组氨酸、myc、血凝素(HA)、FLAG、荧光标签、串联亲和纯化(TAP)标签、谷胱甘肽S转移酶(GST)、多核苷酸、适体、抗原或抗体。本文进一步提供了用于检索信息的方法,其中使用2、3、4或5种形式的解密。

[0010] 本文提供了用于存储信息的系统,该系统包含:(a) 接收单元,其用于接收关于至少一个数字序列的形式的至少一项信息的机器指令,以及关于选择至少一种生物加密格式的机器指令,其中所述生物加密格式是基于酶、电磁、化学或亲和力的生物加密;(b) 处理器单元,其用于基于所选择的生物加密格式将所述至少一个数字序列自动转化为多个寡核苷酸序列;(c) 合成仪单元,其用于从所述处理器单元接收关于合成编码所述寡核苷酸序列的多个寡核苷酸的机器指令;(d) 存储单元,其用于接收从所述合成仪单元沉积的所述多个寡核苷酸。本文进一步提供了用于存储信息的系统,其中所述基于酶的生物加密包括基于CRISPR/Cas的生物加密。本文进一步提供了用于存储信息的系统,其中所述基于酶的生物加密包括关于合成对表1所示的酶敏感的寡核苷酸的机器指令。本文进一步提供了用于存储信息的系统,其中所述基于电磁的生物加密包括关于合成对约0.01nm至约400nm的电磁波长敏感的寡核苷酸的机器指令。本文进一步提供了用于存储信息的系统,其中所述基于化学的生物加密包括关于合成对气态氨或甲胺施用敏感的寡核苷酸的机器指令。本文进一步提供了用于存储信息的系统,其中所述基于亲和力的生物加密包括关于合成对序列标签或亲和标签敏感的寡核苷酸的指令。本文进一步提供了用于存储信息的系统,其中所述亲和标签是生物素、洋地黄毒苷、Ni-次氨基三乙酸、脱硫生物素、组氨酸、聚组氨酸、myc、血凝素(HA)、FLAG、荧光标签、串联亲和纯化(TAP)标签、谷胱甘肽S转移酶(GST)、多核苷酸、适体、抗原或抗体。本文进一步提供了用于存储信息的系统,其中所述多个寡核苷酸包括至少100,000个寡核苷酸。本文进一步提供了用于存储信息的系统,其中所述多个寡核苷酸包括至少100亿个寡核苷酸。

[0011] 本文提供了用于检索信息的系统,其包括:(a) 存储单元,其在表面上包含多个寡核苷酸;(b) 沉积单元,其用于将基于酶、电磁、化学或亲和力的生物加密施加于所述多个寡核苷酸;(c) 测序单元,其用于对所述多个寡核苷酸进行测序以获得核酸序列;(d) 处理器单元,其用于将所述核酸序列自动转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。本文进一步提供了用于检索信息的系统,其中所述沉积单元将CRISPR/Cas复合物施加于所述多个寡核苷酸。本文进一步提供了用于检索信息的系统,其中所述基于酶的生物加密包括施加表1所示的酶。本文进一步提供了用于检索信息的系统,其中所述基于电磁的生物加密包括施加约0.01nm至约400nm的波长。本文进一步提供了用于检索信

息的系统,其中所述基于化学的生物加密包括施加气态氨或甲胺施用。本文进一步提供了用于检索信息的系统,其中所述基于亲和力的生物加密包括序列标签或亲和标签。本文进一步提供了用于检索信息的系统,其中所述亲和标签是生物素、洋地黄毒苷、Ni-次氨基三乙酸、脱硫生物素、组氨酸、聚组氨酸、myc、血凝素(HA)、FLAG、荧光标签、串联亲和纯化(TAP)标签、谷胱甘肽S转移酶(GST)、多核苷酸、适体、抗原或抗体。

[0012] 本文提供了用于存储信息的方法,该方法包括:(a)以至少一个数字序列的形式接收至少一项信息;(b)接收关于至少一种形式的生物加密的指令;(c)将所述至少一个数字序列转化为多个生物加密的寡核苷酸序列;(d)合成所述多个生物加密的寡核苷酸序列;以及(e)存储所述多个寡核苷酸。

[0013] 本文提供了用于存储信息的方法,该方法包括:(a)以至少一个数字序列的形式接收至少一项信息;(b)接收关于基于酶、电磁、化学或亲和力的生物加密的指令;(c)将所述至少一个数字序列转化为多个生物加密的寡核苷酸序列;(d)合成所述多个生物加密的寡核苷酸序列;以及(e)存储所述多个寡核苷酸。

[0014] 本文提供了用于存储信息的方法,该方法包括:(a)以至少一个数字序列的形式接收至少一项信息;(b)将所述至少一个数字序列转化为多个生物加密的寡核苷酸序列,其中每个生物加密的寡核苷酸序列包含编码为被CRISPR/Cas复合物去除的附加序列;(c)合成所述多个生物加密的寡核苷酸序列;以及(d)存储所述多个寡核苷酸。

[0015] 本文提供了用于检索信息的方法,该方法包括:(a)从表面释放多个寡核苷酸;(b)对所述多个寡核苷酸施加至少一种形式的生物解密;(c)富集所述多个寡核苷酸,从而选择多个富集的寡核苷酸;(d)对所述富集的寡核苷酸进行测序以生成核酸序列;以及(e)将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

[0016] 本文提供了用于检索信息的方法,该方法包括:(a)从表面释放多个寡核苷酸;(b)对所述多个寡核苷酸施加基于酶、电磁、化学或亲和力的解密;(c)富集所述多个寡核苷酸,从而选择多个富集的寡核苷酸;(d)对所述富集的寡核苷酸进行测序以生成核酸序列;以及(e)将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

[0017] 本文提供了用于检索信息的方法,该方法包括:(a)从表面释放多个寡核苷酸;(b)将CRISPR/Cas复合物施加于所述多个寡核苷酸;(c)富集所述多个寡核苷酸,从而选择多个富集的寡核苷酸;(d)对所述富集的寡核苷酸进行测序以生成核酸序列;以及(e)将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

[0018] 本文提供了用于存储信息的系统,该系统包含:(a)接收单元,其用于接收关于至少一个数字序列的形式的至少一项信息的机器指令,以及关于至少一种形式的生物加密的机器指令;(b)处理器单元,其用于将所述至少一个数字序列转化为多个生物加密的寡核苷酸序列;(c)合成仪单元,其用于从所述处理器单元接收关于合成所述多个生物加密的寡核苷酸序列的机器指令;以及(d)存储单元,其用于接收从所述合成仪单元沉积的所述多个寡核苷酸。

[0019] 本文提供了用于存储信息的系统,该系统包含:(a)接收单元,其用于接收关于至少一个数字序列的形式的至少一项信息的机器指令,以及关于基于酶、电磁、化学或亲和力的生物加密的机器指令;(b)处理器单元,其用于将所述至少一个数字序列转化为多个生物

加密的寡核苷酸序列；(c) 合成仪单元,其用于从所述处理器单元接收关于合成所述多个生物加密的寡核苷酸序列的机器指令；以及(d) 存储单元,其用于接收从所述合成仪单元沉积的所述多个寡核苷酸。

[0020] 本文提供了用于存储信息的系统,该系统包含:(a) 接收单元,其接收关于至少一个数字序列的形式的至少一项信息的机器指令,以及关于由CRISPR/Cas复合物进行生物加密的机器指令；(b) 处理器单元,其用于将所述至少一个数字序列转化为多个生物加密的寡核苷酸序列；(c) 合成仪单元,其用于从所述处理器单元接收关于合成所述多个生物加密的寡核苷酸序列的机器指令；以及(d) 存储单元,其用于接收从所述合成仪单元沉积的所述多个寡核苷酸。

[0021] 本文提供了用于检索信息的系统,其包括:(a) 存储单元,其在表面上包含多个寡核苷酸；(b) 沉积单元,其用于将至少一种形式的生物解密施加于所述多个寡核苷酸；(c) 测序单元,其用于对所述多个寡核苷酸进行测序以获得核酸序列；以及(d) 处理器单元,其用于将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

[0022] 本文提供了用于检索信息的系统,该方法包括:(a) 存储单元,其在表面上包含多个寡核苷酸；(b) 沉积单元,其用于将至少基于酶、电磁、化学或亲和力的生物加密施加于所述多个寡核苷酸；(c) 测序单元,其用于对所述多个寡核苷酸进行测序以获得核酸序列；以及(d) 处理器单元,其用于将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

[0023] 本文提供了用于检索信息的系统,其包括:(a) 存储单元,其在表面上包含多个寡核苷酸；(b) 沉积单元,其用于将CRISPR/Cas复合物施加于所述多个寡核苷酸；(c) 测序单元,其用于对所述多个寡核苷酸进行测序以获得核酸序列；以及(d) 处理器单元,其用于将所述核酸序列转化为至少一个数字序列,其中所述至少一个数字序列编码至少一项信息。

附图说明

[0024] 在所附权利要求书中具体阐述了本发明的新颖特征。通过参考对在其中利用到本发明原理的说明性实施方案加以阐述的以下详细描述和附图,将会对本发明的特征和优点获得更好的理解,在附图中:

[0025] 图1示出了基于核酸的数据存储的示例性工作流程。

[0026] 图2示出了用于生物加密的存储的示例性工作流程。

[0027] 图3示出了在生物加密之后进行检索的示例性工作流程。

[0028] 图4A-4B描绘了使用Cas酶的生物加密方法。

[0029] 图5A-5C描绘了各种寡核苷酸序列设计方案。

[0030] 图6A-6C描绘了各种寡核苷酸序列设计方案。

[0031] 图7A-7B描绘了条形码设计方案。

[0032] 图8示出了被配置用于寡核苷酸合成的板,其包含24个区域或子域(每个具有256个簇的阵列)。

[0033] 图9示出了图8中具有 16×16 个簇(每个簇具有121个单独的座位)的子域的近视图。

- [0034] 图10示出了图8中的簇的详细视图,其中该簇具有121个座位。
- [0035] 图11A示出了具有多个通道的板的正视图。
- [0036] 图11B示出了具有多个通道的板的剖视图。
- [0037] 图12A-12B描绘了用于柔性结构的连续环和卷到卷(reel-to-reel)布置。
- [0038] 图13A-13C描绘了分别具有平坦特征(座位)、通道或孔的柔性结构的放大图。
- [0039] 图14A示出了本文所述结构上的特征的放大图。
- [0040] 图14B-14C示出了本文所述结构上的标记(markings)。
- [0041] 图15示出了寡核苷酸合成材料沉积装置。
- [0042] 图16示出了寡核苷酸合成工作流程。
- [0043] 图17示出了计算机系统的示例。
- [0044] 图18是示出计算机系统的架构的框图。
- [0045] 图19是说明网络的示图,该网络被配置用于并入多个计算机系统、多个蜂窝电话和个人数据助理,以及网络附加存储(NAS)。
- [0046] 图20是使用共享虚拟地址存储空间的多处理器计算机系统的框图。

具体实施方式

[0047] 定义

[0048] 除非另有定义,否则本文使用的所有技术和科学术语具有与这些发明所属领域普通技术人员通常理解的相同的含义。

[0049] 贯穿本公开内容,各个实施方案以范围格式给出。应当理解,范围格式的描述只是为了方便和简明,而不应被解释为对任何实施方案的范围的硬性限制。因此,除非上下文另有明确规定,否则对范围的描述应被认为明确公开了所有可能的子范围以及该范围内精确到下限单位十分之一的各个数值。例如,对诸如从1至6的范围的描述应被认为已经明确公开了诸如从1至3、从1至4、从1至5、从2至4、从2至6、从3至6等子范围,以及该范围内的各个值,例如,1.1、2、2.3、5和5.9。无论范围的宽度如何,这都是适用的。这些中间范围的上限和下限可独立地包括在更小的范围内,并且也被涵盖于本发明之中,但受制于所声称范围中的任何被明确排除的限值。除非上下文另有明确规定,否则当所声称的范围包括限值之一或全部两者时,排除了这些包括的限值之一或全部两者的范围也被包括在本发明中。

[0050] 本文使用的术语仅用于描述特定实施方案的目的,而非旨在限制任何实施方案。除非上下文另有明确规定,否则如本文所用的单数形式“一个”、“一种”和“该”也意欲包括复数形式。进一步应当理解,术语“包括”和/或“包含”在本说明书中使用指定所述特征、整体、步骤、操作、元件和/或组分的存在,但不排除存在或添加一个或多个其它特征、整体、步骤、操作、元件、组分和/或其群体。如本文所用的,术语“和/或”包括一个或多个相关所列项目的任何及所有组合。

[0051] 除非特别说明或从上下文中可以明显看出,否则如本文所用的,关于数字或数字范围的术语“约”应被理解为表示所述数字及其 $\pm 10\%$ 的数字,或者对于范围列出的值,表示低于所列下限的10%至高于所列上限的10%。

[0052] 如本文所用的术语“寡核苷酸”可与“寡核酸”互换使用。术语“寡核苷酸”和“寡核酸”涵盖双链或三链核酸以及单链分子。

[0053] 基于核酸的信息存储

[0054] 本文提供了用于基于核酸的信息(数据)存储的装置、组合物、系统和方法。图1中提供了示例性工作流程。在第一步中,接收编码信息项的数字序列(即,由计算机处理的二进制代码形式的数字信息)(101)。施加加密103方案以将数字序列从二进制代码转化为核酸序列105。选择用于核酸延伸的表面材料、用于核酸延伸的座位(也称为排列斑点)的设计和用于核酸合成的试剂(107)。准备结构的表面以用于核酸合成(108)。进行从头寡核苷酸合成(109)。存储合成的寡核苷酸(111)并且该合成的多核苷酸可全部或部分地用于随后的释放(113)。一旦释放,对寡核苷酸全部或部分地进行测序(115),并对其进行解密(117)以将该核酸序列转化回数字序列。然后装配数字序列(119)以获得原始信息项的排比编码。

[0055] 本文进一步提供了用于安全的基于DNA的信息存储的方法和系统,其包括:接收编码至少一项信息的一个或多个数字序列201,将所述一个或多个数字序列转化为核酸序列203,对该核酸序列进行加密205,以及所述加密的核酸序列的从头寡核苷酸合成207。见图2。

[0056] 本文提供了用于基于核酸的信息存储的装置、组合物、系统和方法,其中接收关于从数字序列转化为核酸序列、生物加密、生物解密或这些步骤中任何步骤的組合的机器指令。可以接收关于用于转化的所需信息项以及关于一种或多种类型的生物加密的机器指令,该生物加密选项一系列选项,例如但不限于基于酶的(例如,CRISPR/Cas复合物或限制酶消化物)、基于电磁辐射的(例如,光解或光检测)、化学裂解(例如,气态氨或甲胺处理,以裂解胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244))和基于亲和力的(例如,用于杂交的序列标签,或对捕获试剂具有增强的亲和力的修饰核苷酸的掺入)形式的生物加密。在接收到特定的生物加密选择之后,程序模块执行以下步骤:在将合成指令提供给材料沉积装置以便从头合成寡核苷酸之前,将信息项转化为核酸序列并施加关于设计生物加密版本的序列的设计指令。在一些情况下,提供关于选择生物加密类别中的一个或多个种类的机器指令。

[0057] 本文进一步提供了用于安全的基于DNA的信息检索的方法和系统,其包括从表面释放寡核苷酸301,所需寡核苷酸的富集303,寡核苷酸的测序305,核酸序列的解密307,以及一个或多个编码信息项的数字序列的装配309。参见图3。

[0058] 还可以提供本文所述的机器指令用于生物解密。生物解密可以包括接收机器指令。这类指令可以包括选自一系列选项的一种或多种格式的生物解密,例如但不限于基于酶的(例如,CRISPR/Cas复合物或限制酶消化物)、基于电磁辐射的(例如,光解或光检测)、化学裂解(例如,气态氨或甲胺处理,以裂解胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244))和基于亲和力的(例如,用于杂交的核酸序列,或对捕获试剂具有增强的亲和力的修饰核苷酸的掺入)形式的生物解密。在接收到特定的生物解密选择之后,程序模块执行释放调节剂以富集寡核苷酸的步骤。富集后,对寡核苷酸进行测序,任选地与更长的核酸序列进行比对,并将其转化为与信息项相对应的数字序列。在一些情况下,提供关于选择生物解密类别中的一个或多个种类的机器指令。

[0059] 信息项

[0060] 任选地,本文公开的DNA数据存储过程的早期步骤包括获得或接收初始代码(例如,数字序列)形式的一项或多项信息。信息项包括但不限于文本、音频和视觉信息。信息项

的示例性来源包括但不限于书籍、期刊、电子数据库、病历、信件、表格、录音、动物记录、生物学谱、广播、电影、短视频、电子邮件、簿记电话记录、互联网活动日志、图纸、图画、版画、照片、像素化图形和软件代码。信息项的示例性生物学谱来源包括但不限于基因文库、基因组、基因表达数据和蛋白质活性数据。信息项的示例性格式包括但不限于.txt、.PDF、.doc、.docx、.ppt、.pptx、.xls、.xlsx、.rtf、.jpg、.gif、.psd、.bmp、.tiff、.png和.mpeg。以数字格式编码信息项的单个文件大小或编码信息项的多个文件的量包括但不限于最多1024个字节(等于1KB)、1024KB(等于1MB)、1024MB(等于1GB)、1024GB(等于1TB)、1024TB(等于1PB)、1艾字节(exabyte)、1泽字节(zettabyte)、1尧字节(yottabyte)、1xenottabyte或更多。在一些情况下,数字信息量为至少或约1吉字节(GB)。在一些情况下,数字信息量为至少或约1、2、3、4、5、6、7、8、9、10、20、50、100、200、300、400、500、600、700、800、900、1000或大于1000吉字节。在一些情况下,数字信息量为至少或约1太字节(TB)。在一些情况下,数字信息量为至少或约1、2、3、4、5、6、7、8、9、10、20、50、100、200、300、400、500、600、700、800、900、1000或大于1000太字节。在一些情况下,数字信息量为至少或约1拍字节(PB)。在一些情况下,数字信息量为至少或约1、2、3、4、5、6、7、8、9、10、20、50、100、200、300、400、500、600、700、800、900、1000或大于1000拍字节。

[0061] 加密

[0062] 生物加密和解密

[0063] 本文描述了装置、组合物、系统和方法,其包括在接收到编码信息项的数字序列之后进行生物学加密(也称为“生物加密”)。除了本文描述的生物加密和生物解密的单独形式之外,本文还提供了用于将掩蔽生物序列的一种或多种类别或种类的选择合并到关于信息存储和/或检索的工作流程中的过程。

[0064] 本文提供了从包含生物加密的较大核酸序列群体中对目的核酸序列进行目标富集的装置、组合物、系统和方法。在一些情况下,使用生物加密来富集来自噪声的目标信号。在一些情况下,该目标信号是目的核酸序列。在一些情况下,生物加密包括将目的核酸序列引入具有已知序列的较大核酸序列群体中。已知的核酸序列可以被称为加密核酸序列。在一些情况下,对加密核酸进行解密。在一些情况下,已知核酸序列的解密导致目的核酸序列的信噪比增加。

[0065] 本文提供了装置、组合物、系统和方法,其包括将生物分子加密合并到信息存储和/或检索工作流程中。生物加密和生物解密的示例性形式包括但不限于基于酶的、基于电磁辐射的、化学裂解的和基于亲和力的生物加密和生物解密。

[0066] 本文提供了装置、组合物、系统和方法,其包括施加基于核酸酶复合物活性的加密。示例性核酸酶包括但不限于Cas核酸酶(CRISPR关联的)、锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶、Argonaute核酸酶、S1核酸酶、绿豆核酸酶或DNA酶。示例性的Cas核酸酶包括但不限于Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx1S、Csf1、Csf2、Cs0、Csf4、Cpf1、c2c1和c2c3。在一些情况下,该Cas核酸酶是Cas9。在一些情况下,CRISPR/Cas复合物提供了一个或多个核酸序列的预定去除。在一些情况下,本文所述的富集步骤包括通过杂交消耗丰富序列(DASH)。在一些情况下,DASH包括施加核酸酶。例如,

核酸酶,如Cas9,当与包括指导RNA(“gRNA”)序列的CRISPR复合物结合时,诱导链断裂,使得更长形式的核酸序列不再完整。在一些情况下,切下的核酸在富集后不可用于随后的扩增。在一些情况下,gRNA将Cas9酶引导至特定的核酸段。在替代布置中,gRNA具有多个切割位点。基于gRNA的系统允许生成具有高度特异性和选择性的加密代码。例如,由于基于CRISPR/Cas9的系统使用20bp来识别待切割的序列,因此至少约 10^{12} 种不同的可能性可用于设计预定的gRNA序列,以供使用4碱基系统进行解密。除去外来(又称“垃圾”)DNA后,编码靶序列的预定寡核苷酸经历下游加工,例如扩增和测序,从而得到没有额外(垃圾)序列的最终序列。在一些情况下,设计所述多个寡核苷酸中的每个寡核苷酸以供在多个位置进行修饰(例如,切割、碱基交换、重组)。例如,合成所述多个寡核苷酸中的每个寡核苷酸,其具有用于结合约1、2、3、4、5、6、7、8、9、10个或更多个gRNA序列的互补区域。在这样的布置中,所述多个寡核苷酸中的每一个在1、2、3、4、5、6、7、8、9、10个或更多个位置处经历裂解、碱基交换、核酸酶(例如,CRISPR/Cas)复合物活性之后的重组。

[0067] 使用CRISPR/Cas9进行数据加密的目标富集的第一过程在图4A中示出。DNA序列群体401包含DNA信息403和加密的DNA 405。DNA信息403和加密的DNA 405分别包含衔接子序列402和DNA序列404、406。将指导RNA 409添加(407)到DNA序列群体401中。指导RNA409用于通过识别加密的DNA 405中加密的切割序列来去除加密的DNA 405。在添加指导RNA 409后,加密的DNA 405被切割,导致核酸序列不再完整。因此,例如,当包含不再完整的核酸序列的加密的DNA405不能被扩增时,将加密的DNA 405从该群体中去除411,留下DNA信息403。

[0068] 使用CRISPR/Cas9进行数据加密的目标富集的第二过程在图4B中示出。DNA序列群体421包含DNA信息423和加密的DNA 425。DNA信息423和加密的DNA 425分别包含衔接子序列422和DNA序列424、426。将指导RNA 429和供体DNA 431添加(427)到DNA序列群体421中。指导RNA 429识别加密的DNA 425中加密的切割位点,并生成用于插入供体DNA 431的切割位点。供体DNA 431的插入导致加密的DNA 425中的插入或移码。在一些情况下,供体DNA 431的插入导致序列标签的引入,以用于杂交或掺入对捕获试剂具有增强的亲和力的修饰核苷酸。例如,供体DNA 431被荧光探针识别。在一些情况下,供体DNA 431引入用于基于电磁辐射的(例如,光解或光检测)或基于化学裂解的(例如,气态氨或甲胺处理,以裂解胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244))生物加密和/或生物解密的序列。在一些情况下,加密的DNA 425不再被识别以扩增,而仅有DNA信息423得到扩增,导致DNA信息423的富集。

[0069] 如本文所述的包括施加基于核酸酶复合物活性的加密的装置、组合物、系统和方法可包括碱基交换或序列交换。例如,使用CRISPR/Cas的生物加密和生物解密包括碱基交换或序列交换。在一些情况下,生物加密包括CRISPR/dCas9,其中失效的或“死”的Cas9(“dCas9”)不再具有剪接功能,但是在添加另一种酶促活性的情况下,行使不同的靶分子修饰功能。例如,将胞苷脱氨酶拴系到dCas9上会将C-G DNA碱基对转化为T-A碱基对。在替代的dCas9方法中,拴系在dCas9上的不同的酶会导致靶DNA中的碱基C变为T,或者G变为A。

[0070] 本文提供了用于生物加密和生物解密的装置、组合物、系统和方法,其包括施加限制酶。在一些情况下,该限制酶靶向酶识别位点。在一些情况下,该酶识别位点是特定的核苷酸序列。在一些情况下,该限制酶在酶识别位点处或附近切割磷酸骨架。在一些情况下,该识别位点的切割导致非平端或平端。在一些情况下,该限制酶识别核苷酸(例如,A、T、G、

C、U)。在一些情况下,该限制酶识别修饰,例如但不限于甲基化、羟基化或糖基化。在一些情况下,该限制酶导致片段化。在一些情况下,片段化产生具有5' 突出端、3' 突出端、平端或其组合的片段。在一些情况下,例如基于大小选择片段。在一些情况下,由限制酶导致的片段化之后进行连接。例如,利用由限制酶导致的片段化留下可预测的突出端,然后与一个或多个包含与核酸片段上的可预测突出端互补的突出端的衔接子寡核苷酸连接。表1提供了示例性限制酶及其识别序列。

[0071] 表1.限制酶

[0072] 表1

[0073]

限制序列	酶
AA/CGTT	AclI

[0074]

A/AGCTT	HindIII
AAT/ATT	HindIII-HF®
/AATT	SspI SspI-HF®
A/CATGT	MluCI Tsp509I
A/CCGGT	PciI
ACCTGC(4/8)	AgeI AgeI-HF®
A/CCWGGT	BspMI BfuAI
A/CGCGT	SexAI
ACGGC(12/14)	MluI MluI-HF®
A/CGT	BceAI
ACN/GT	HpyCH4IV
(10/15)ACNNNNGTAYC(12/7) (SEQ ID NO: 5)	HpyCH4III
(9/12)ACNNNNNCTCC(10/7) (SEQ ID NO: 6)	BaeI
A/CRYGT	BsaXI
A/CTAGT	AflIII
ACTGG(1/-1)	SpeI SpeI-HF®
ACTGGG(5/4)	BsrI
A/GATCT	BmrI
AGC/GCT	BglII
AG/CT	AfeI
AGG/CCT	AluI
AGT/ACT	StuI
AT/CGAT	ScaI ScaI-HF®
ATCTATGTCGGGTGCGGAGAAAGAGGTAAT(-15/-19) (SEQ ID NO: 7)	Clal BspDI
ATGCA/T	PI-SceI
AT/TAAT	NsiI NsiI-HF®
ATTT/AAAT	AseI
(11/13)CAANNNNNGTGG(12/10) (SEQ ID NO: 8)	Swal
C/AATTG	CspCI
CACGAG(-5/-1)	MfeI MfeI-HF®
CACGAG	BssSI BssSaI
CACGTC(-3/-3)	Nb.BssSI
CAC/GTG	BmgBI
CACNNN/GTG	PmlI
CACNN/NNGTG (SEQ ID NO: 9)	DraIII DraIII-HF®
CAGCAG(25/27)	AleI
CAG/CTG	EcoP15I
CAGNNN/CTG	PvuII PvuII-HF®
CAGTG(2/0)	AlwNI
NNCASTGNN/	BtsIMutI
CA/TATG	TspRI
CATG/	NdeI
C/ATG	NlaIII
/CATG	CviAII
	FatI

[0075]

CAYNN/NNRTG (SEQ ID NO: 10)	MslI
CC(12/16)	FspEI
CCANNNNN/NNNNTGG (SEQ ID NO: 11)	XcmI
CCANNNNN/NTGG (SEQ ID NO: 12)	BstXI
CCANNNN/NTGG (SEQ ID NO: 13)	PflMI
CCATC(4/5)	BccI
C/CATGG	NcoI NcoI-HF®
CCCAGC(-5/-1)	BseYI
CCCGC(4/6)	FauI
CCC/GGG	SmaI
C/CCGGG	XmaI TspMI
(0/-1)CCD	Nt.CviPII
CCDG(10/14)	LpnPI
CCGC(-3/-1)	AciI
CCGC/GG	SacII
CCGCTC(-3/-3)	BsrBI
C/CGG	MspI HpaII
CC/NGG	ScrFI
/CCNGG	BssKI StyD4I
C/CNNGG	BsaJI
CCNNNNN/NNGG (SEQ ID NO: 14)	BslI
C/CRYGG	BtgI
CC/SGG	NciI
C/CTAGG	AvrII
CCTC(7/6)	MnlI
CCTCAGC(-5/-2)	BbvCI
CCTCAGC	Nb.BbvCI
CCTCAGC(-5/-7)	Nt.BbvCI
CCTGCA/GG	SbfI SbfI-HF®
CCTNAGC(-5/-2)	Bpu10I
CC/TNAGG	Bsu36I
CCTNN/NNNAGG (SEQ ID NO: 15)	EcoNI
CCTTC(6/5)	HpyAV
CC/WGG	BstNI
/CCWGG	PspGI
C/CWWGG	StyI StyI-HF®
(10/12)CGANNNNNNTGC(12/10) (SEQ ID NO: 16)	BcgI
CGAT/CG	PvuI PvuI-HF®
CG/CG	BstUI
C/GGCCG	EagI EagI-HF®
CG/GWCCG	RsrII
CGRY/CG	BsiEI
C/GTACG	BsiWI BsiWI-HF®
CGTCTC(1/5)	BsmBI
CGWCG/	Hpy99I

[0076]

CMG/CKG	MspAII
CNNR(9/13)	MspJI
CR/CCGGYG	SgrAI
C/TAG	BfaI
CTCAG(9/7)	BspCNI
C/TCGAG	XhoI PaeR7I TliI
CTCTTC(1/4)	EarI
CTGAAG(16/14)	AcuI
CTGCA/G	PstI PstI-HF®
CTGGAG(16/14)	BpmI
C/TNAG	DdeI
C/TRYAG	SfcI
C/TTAAG	AflII
CTTGAG(16/14)	BpuEI
C/TYRAG	SmlI
C/YCGRG	AvaI BsoBI
GAAGA(8/7)	MboII
GAAGAC(2/6)	BbsI BbsI-HF®
GAANN/NNTTC (SEQ ID NO: 17)	XmnI
GAATGC(1/-1)	BsmI
GAATGC	Nb.BsmI
G/AATTC	EcoRI EcoRI-HF®
GACGC(5/10)	HgaI
GACGT/C	AatII
GAC/GTC	ZraI
GACN/NNGTC	Tth111I PflFI
GACNN/NNGTC (SEQ ID NO: 18)	PshAI
GACNNN/NNGTC (SEQ ID NO: 19)	AhdI
GACNNNN/NNGTC (SEQ ID NO: 20)	DrdI
GAG/CTC	Eco53kI
GAGCT/C	SacI SacI-HF®
GAGGAG(10/8)	BseRI
GAGTC(4/5)	PleI
GAGTC(4/-5)	Nt.BstNBI
GAGTC(5/5)	MlyI
G/ANTC	HinfI
GAT/ATC	EcoRV EcoRV-HF®
/GATC	MboI Sau3AI DpnII BfuCI
GA/TC	DpnI
GATNN/NNATC (SEQ ID NO: 21)	BsaBI
G/AWTC	TfiI
GCAATG(2/0)	BsrDI
GCAATG	Nb.BsrDI
GCAGC(8/12)	BbvI

[0077]

GCAGTG(2/0)	BtsI BtsαI
GCAGTG	Nb.BtsI
GCANNNN/NTGC (SEQ ID NO: 22)	BstAPI
GCATC(5/9)	SfaNI
GCATG/C	SphI SphI-HF®
GCCC/GGGC	SrfI
GCCGAG(21/19)	NmeAIII
GCC/GGC	NaeI
G/CCGGC	NgoMIV
GCCNNNN/NGGC (SEQ ID NO: 23)	BglI
GCGAT/CGC	AsiSI
GCGATG(10/14)	BtgZI
G/CGC	HinPII
GCG/C	HhaI
G/CGCGC	BssHII
GC/GGCCGC	NotI NotI-HF®
GC/NGC	Fnu4HI
GCN/NGC	Cac8I
GCNNNNN/NGC (SEQ ID NO: 24)	MwoI
G/CTAGC	NheI NheI-HF®
GCTAG/C	BmtI BmtI-HF®
GCTCTTC(1/4)	SapI BspQI
GCTCTTC(1/-7)	Nt.BspQI
GC/TNAGC	BlpI
G/CWGC	TseI ApeKI
GDGCH/C	Bsp1286I
GGATC(4/5)	AlwI
GGATC(4/-5)	Nt.AlwI
G/GATCC	BamHI BamHI-HF®
GGATG(9/13)	FokI
GGATG(2/0)	BtsCI
GG/CC	HaeIII PhoI
GGCCGG/CC	FseI
GGCCNNNN/NGGCC (SEQ ID NO: 25)	SfiI
GG/CGCC	NarI
G/GCGCC	KasI
GGC/GCC	SfoI
GGCGC/C	PluTI
GG/CGCGCC	AscI
GGCGGA(11/9)	EciI
GGGAC(10/14)	BsmFI
GGGCC/C	ApaI
G/GGCCC	PspOMI
G/GNCC	Sau96I
GGN/NCC	NlaIV

[0078]

GGTAC/C	KpnI KpnI-HF®
G/GTACC	Acc65I
GGTCTC(1/5)	BsaI BsaI-HF®
GGTGA(8/7)	HphI
G/GTNACC	BstEII BstEII-HF®
G/GWCC	AvaII
G/GYRCC	BanI
GKGCM/C	BaeGI
GR/CGYC	BsaHI
GRGCY/C	BanII
GT/AC	RsaI
G/TAC	CviQI
GTA/TAC	BstZ17I
GTATAC	BstZ17I-HF®
GTATCC(6/5)	BciVI
G/TCGAC	SalI SalI-HF®
GTCTC(1/-5)	Nt.BsmAI
GTCTC(1/5)	BsmAI BcoDI
G/TGCAC	ApaLI
GTGCAG(16/14)	BsgI
GT/MKAC	AccI
GTN/NAC	Hpy166II
/GTSAC	Tsp45I
GTT/AAC	HpaI
GTTT/AAAC	PmeI
GTY/RAC	HincII
GWGCW/C	BsiHKA I
R/AATTY	ApoI ApoI-HF
RCATG/Y	NspI
R/CCGGY	BsrFI BsrFaI
R/GATCY	BstYI
RGCGC/Y	HaeII
RG/CY	CviKI-1
RG/GNCCY	EcoO109I
RG/GWCCY	PpuMI
TA ACTATAACGGTCCTAAGGTAGCGAA(-9/-13) (SEQ ID NO: 26)	I-CeuI
TAC/GTA	SnaBI
TAGGGATAACAGGGTAAT(-9/-13) (SEQ ID NO: 27)	I-SceI
T/CATGA	BspHI
T/CCGGA	BspEI
TCCRAC(20/18)	MmeI
T/CGA	TaqαI
TCG/CGA	NruI NruI-HF®
TCN/GA	Hpy188I

[0079]

TC/NNGA	Hpy188III
T/CTAGA	XbaI
T/GATCA	BclI
TG/CA	HpyCH4V
TGC/GCA	FspI
TGGCAAACAGCTATTATGGGTATTATGGGT(-13/-17) (SEQ ID NO: 28)	PI-PspI
TGG/CCA	MscI
T/GTACA	BsrGI BsrGI-HF®
T/TAA	MseI
TTAAT/TAA	PacI
TTA/TAA	PsiI
TT/CGAA	BstBI
TTT/AAA	DraI
VC/TCGAGB	PspXI
W/CCGGW	BsaWI
YAC/GTR	BsaAI
Y/GGCCR	EaeI

[0080] 本文提供了用于生物加密和生物解密的装置、组合物、系统和方法,其可以包括施加修复酶。在一些情况下,DNA修复酶来源于特定的生物体或病毒,或者是其非天然存在的变体。示例性的DNA修复酶包括但不限于大肠杆菌内切核酸酶IV、Tth内切核酸酶IV、人AP内切核酸酶、糖基化酶如UDG、大肠杆菌3-甲基腺嘌呤DNA糖基化酶(AIka)和人Aag、糖基化酶/裂合酶如大肠杆菌内切核酸酶III、大肠杆菌内切核酸酶VIII、大肠杆菌Fpg、人OGG1和T4PDG以及裂合酶。表2中列出了示例性的其它DNA修复酶。

[0081] 表2.DNA修复酶

[0082]

基因名称	活性	登录号
UNG	尿嘧啶-DNA 糖基化酶	NM_080911
SMUG1	尿嘧啶-DNA 糖基化酶	NM_014311
MBD4	去除 CpG 序列处 U 或 T 相对的 G	NM_003925
TDG	去除 U、T 或亚乙烯基 C 相对的 G	NM_003211
OGG1	去除 8-氧代 G 相对的 C	NM_016821
MUTYH (MYH)	去除 A 相对的 8-氧代 G	NM_012222
NTHL1 (NTH1)	去除环饱和的或片段化的嘧啶	NM_002528
MPG	去除 3-meA、亚乙烯基 A、次黄嘌呤	NM_002434
NEIL1	去除胸腺嘧啶二醇	NM_024608

[0083]

NEIL2	去除嘧啶的氧化产物	NM_145043
XPC	结合受损的 DNA 作为与 RAD23B、CETN2 的复合物	NM_004628
RAD23B (HR23B)	结合受损的 DNA 作为与 XPC、CETN2 的复合物	NM_002874
CETN2	结合受损的 DNA 作为与 XPC、RAD23B 的复合物	NM_004344
RAD23A (HR23A)	HR23B 的替代物	NM_005053
XPA	结合切开前复合物中受损的 DNA	NM_000380
RPA1	结合切开前复合物中的 DNA	NM_002945
RPA2	结合切开前复合物中的 DNA	NM_002946
RPA3	结合切开前复合物中的 DNA	NM_002947
ERCC5 (XPG)	3'切开	NM_000123
ERCC1	5'切开亚单位	NM_001983
ERCC4 (XPF)	5'切开亚单位	NM_005236
LIG1	DNA 接合	NM_000234
CKN1(CSA)	Cockayne 综合征; 转录耦合的 NER 所需的	NM_000082
ERCC6 (CSB)	Cockayne 综合征; 转录耦合的 NER 所需的	NM_000124
XAB2 (HCNP)	Cockayne 综合征; 转录耦合的 NER 所需的	NM_020196
DDB1	XP 组 E 缺陷的复合物	NM_001923
DDB2	DDB1, DDB2	NM_000107
MMS19L (MMS19)	转录和 NER	NM_022362
FEN1 (DNA 酶 IV)	Flap 内切核酸酶	NM_004111
SPO11	内切核酸酶	NM_012444
FLJ35220 (ENDOV)	次黄嘌呤和尿嘧啶 3'的切开	NM_173627
FANCA	参与 DNA 交联的耐受或修复	NM_000135
FANCB	参与 DNA 交联的耐受或修复	NM_152633
FANCC	参与 DNA 交联的耐受或修复	NM_000136
FANCD2	参与 DNA 交联的耐受或修复	NM_033084
FANCE	参与 DNA 交联的耐受或修复	NM_021922
FANCF	参与 DNA 交联的耐受或修复	NM_022725
FANCG (XRCC9)	参与 DNA 交联的耐受或修复	NM_004629
FANCL	参与 DNA 交联的耐受或修复	NM_018062
DCLRE1A (SNM1)	DNA 交联修复	NM_014881
DCLRE1B (SNM1B)	与 SNM1 有关	NM_022836
NEIL3	类似于 NEIL1 和 NEIL2	NM_018248
ATRIP (TREX1)	TREX1/ATRIP 基因的 ATR-相互作用蛋白 5'替代 ORF	NM_130384
NTH	除去受损的嘧啶	NP_416150.1
NEI	除去受损的嘧啶	NP_415242.1

[0084]	NFI	脱氧肌苷 3' 内切核酸酶	NP_418426.1
	MUTM	甲酰嘧啶 DNA 糖基化酶	NP_418092.1
	UNG	尿嘧啶-DNA 糖基化酶	NP_417075.1
	UVRA	DNA 切除修复酶复合物	NP_418482.1
	UVRB	DNA 切除修复酶复合物	NP_415300.1
	UVRC	DNA 切除修复酶复合物	NP_416423.3
	DENV	嘧啶二聚体糖基化酶	NP_049733.1

[0085] 本文提供了用于生物加密和/或生物解密的装置、组合物、系统和方法,其包括核酸修饰。在一些情况下,该核酸修饰影响测序反应中核酸序列的活性。例如,该核酸修饰防止加密的核酸序列被扩增。在一些情况下,该核酸修饰包括但不限于甲基化的碱基、PNA(肽核酸)核苷酸、LNA(锁定核酸)核苷酸和2'-O-甲基修饰的核苷酸。在一些情况下,该核酸修饰包含修饰的核碱基,该核碱基不是胞嘧啶、鸟嘌呤、腺嘌呤或胸腺嘧啶。非限制性的修饰的核碱基包括但不限于尿嘧啶、3-meA(3-甲基腺嘌呤)、次黄嘌呤、8-氧代G(7,8-二氢-8-氧代鸟嘌呤)、FapyG、FapyA、Tg(胸腺嘧啶二醇)、hoU(羟基尿嘧啶)、hmU(羟甲基尿嘧啶)、fU(甲酰基尿嘧啶)、hoC(羟基胞嘧啶)、fC(甲酰基胞嘧啶)、5-meC(5-甲基胞嘧啶)、6-meG(6-甲基鸟嘌呤)、7-meG(N7-甲基鸟嘌呤)、εC(亚乙烯基胞嘧啶)、5-caC(5-羧基胞嘧啶)、2-hA、εA(亚乙烯基腺嘌呤)、5-fU(5-氟尿嘧啶)、3-meG(3-甲基鸟嘌呤)和异径尿酸。

[0086] 本文提供了用于生物加密的装置、组合物、系统和方法,其包括使用核酸探针序列。在一些情况下,随后通过核酸酶去除与核酸序列的一部分互补的核酸探针序列。例如,该核酸酶是识别在核酸探针与核酸序列之间形成的双链核酸分子的双链体特异性核酸酶。在一些情况下,该核酸探针允许捕获并分离核酸序列。在一些情况下,该核酸探针在长度上包含至少约5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100个或超过100个核苷酸。

[0087] 在一些情况下,使用包含标记物的核酸探针鉴定核酸序列,该标记物例如是但不限于亲和标签,如生物素、洋地黄毒苷、Ni-次氨基三乙酸、脱硫生物素、组氨酸、聚组氨酸、myc、血凝素(HA)、FLAG、荧光标签、串联亲和纯化(TAP)标签、谷胱甘肽S转移酶(GST)、多核苷酸、适体、多肽(例如,抗原或抗体)或其衍生物。在一些情况下,通过光吸收、荧光、化学发光、电化学发光、质量或电荷来检测该标记物。荧光团的非限制性实例是Alexa-Fluor染料(例如,Alexa Fluor® 350、Alexa Fluor® 405、Alexa Fluor® 430、Alexa Fluor® 488、Alexa Fluor® 500、Alexa Fluor® 514、Alexa Fluor® 532、Alexa Fluor® 546、Alexa Fluor® 555、Alexa Fluor® 568、Alexa Fluor® 594、Alexa Fluor® 610、Alexa Fluor® 633、Alexa Fluor® 647、Alexa Fluor® 660、Alexa Fluor® 680、Alexa Fluor® 700和Alexa Fluor® 750)、APC、Cascade Blue、Cascade Yellow和R-藻红蛋白(PE)、DyLight 405、DyLight 488、DyLight 550、DyLight 650、DyLight 680、DyLight 755、DyLight 800、FITC、太平洋蓝(Pacific Blue)、PerCP、罗丹明(Rhodamine)、得克萨斯红(Texas Red)、Cy5、Cy5.5和Cy7。

[0088] 本文提供了用于生物加密和/或生物解密的装置、组合物、系统和方法,其包括基于核酸杂交的结合。可以使用包含亲和标签的核酸探针。在一些情况下,该亲和标签允许核酸序列被下拉。例如,亲和标签生物素缀合至与核酸序列互补的核酸探针,并使用链霉亲和

素将其下拉。在一些情况下,该亲和标签包含磁敏材料,例如磁体或磁敏金属。在一些情况下,使用诸如链霉亲和素的固体支持物将核酸序列下拉并固定在固体支持物上。在一些情况下,核酸序列在溶液中例如通过珠子被下拉。在一些情况下,核酸探针允许基于大小排除。例如,核酸探针产生具有与其它核酸序列不同的大小的核酸序列,从而通过基于大小的消耗来去除核酸序列。

[0089] 包括基于核酸杂交的结合的用于生物加密和/或生物解密的装置、组合物、系统和方法可以包括受控的扩增。在一些情况下,基于核酸杂交的结合策略涉及受控扩增,其中多个合成的寡核苷酸具有供正向引物结合的相似区域,但是反向引物区域不容易识别。在这样的情况下,将需要预定的反向引物。在第一示例性工作流程中,生成具有将与每个不同的合成寡核苷酸结合的预选区域的一组反向引物,并将其用于延伸扩增反应(例如,使用DNA聚合酶)中,以扩增寡核苷酸以供下游加工,例如,进一步扩增或DNA测序反应。任选地,每个反向引物包含衔接子区域,该衔接子区域包含共同序列以通过延伸扩增反应(例如,采用DNA聚合酶)掺入通用反向引物结合位点。在这样的布置中,由于仅需要单个正向或反向引物来对多个寡核苷酸进行扩增或测序,因此下游加工得到简化。在第二个示例性的工作流程中,合成多个寡核苷酸,每个寡核苷酸具有一个或两个包含杂交基序的区域,该杂交基序虽然有所不同,但具有足够的与共同引物杂交的能力,从而允许利用针对每个合成寡核苷酸的5'和3'区之一或两者的共同引物对多个寡核苷酸进行下游加工(例如,扩增或测序反应)。在一些情况下,寡核苷酸群体被设计为与共同引物的约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个核碱基杂交。

[0090] 本文提供了用于生物加密和/或生物解密的装置、组合物、系统和方法,其包括使用电磁辐射(EMR)。在一些情况下,电磁辐射提供对核酸序列的切割或基于图像捕获的检测。在一些情况下,以约100nm至约400nm、约100nm至约300nm或约100nm至约200nm的波长向表面施加EMR。在一些情况下,以小于0.01nm的波长向表面施加EMR。在一些情况下,以约10nm至约400nm、约400nm至约700nm或约700nm至约100,000nm的波长向表面施加EMR。例如,以紫外线(UV)波长或深紫外线波长施加EMR。在一些情况下,以约172nm的波长向表面施加深紫外线以从该表面上切割结合剂。在一些情况下,EMR与氙灯一起施加。暴露距离是指灯与表面之间的测量值。在一些情况下,暴露距离为约0.1至5cm。在一些情况下,暴露距离为约0.5至2cm。在一些情况下,暴露距离为约0.5、1、2、3、4或5cm。在一些情况下,用激光施加EMR。示例性激光器及其波长包括但不限于Ar₂(126nm)、Kr₂(146nm)、F₂(157nm)、Xe₂(172和175nm)、ArF(193nm)。在一些情况下,核酸序列包含在特定位点可光切割的核碱基。在一些情况下,核酸序列包含可光切割的修饰的核碱基。在一些情况下,通过施加特定波长的光,核酸序列是可光切割的。在一些情况下,通过施加多个波长的光,核酸序列是可光切割的。

[0091] 本文提供了用于生物加密和/或生物解密的装置、组合物、系统和方法,其包括使用化学裂解。在一些情况下,核酸序列包含在特定位点可化学裂解的核碱基。在一些情况下,核酸序列包含可化学裂解的修饰的核碱基。在一些情况下,该修饰的核碱基包含化学可裂解的修饰。在一些情况下,使用胺试剂进行化学裂解。在一些情况下,该胺试剂是液体、气体、水性试剂或无水试剂。胺试剂的非限制性实例是氢氧化铵、氨气、C₁-C₆烷基胺或甲胺。

[0092] 如本文所述的用于生物加密的装置、组合物、系统和方法可包括将数字序列转化为核酸序列。在一些情况下,该核酸序列是DNA序列。在一些情况下,该DNA序列是单链或双

链的。在一些情况下,该核酸序列是RNA序列。在一些情况下,该RNA序列是单链或双链的。在一些情况下,该核酸序列在较大的核酸序列群体中被加密。在一些情况下,该较大的核酸序列群体是均质群体或异质群体。在一些情况下,该核酸序列群体包含DNA序列。在一些情况下,该DNA序列是单链或双链的。在一些情况下,该核酸序列群体包含RNA序列。在一些情况下,该RNA序列是单链或双链的。

[0093] 许多核酸序列可以被加密。在一些情况下,被加密的核酸序列的数目为约10个序列至约1百万个或更多个序列。在一些情况下,被加密的核酸序列的数目至少为约10、50、100、200、500、1,000、2,000、4,000、8,000、10,000、25,000、30,000、35,000、40,000、45,000、50,000、55,000、60,000、65,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、一百万个或超过一百万个序列。在一些情况下,被加密的核酸序列的数目大于1万亿。

[0094] 在一些情况下,被加密的核酸序列在长度上包含至少10、25、50、75、100、125、150、175、200、225、250、300个或超过300个碱基。在一些情况下,被加密的核酸序列包含10个碱基至25碱基、10个碱基至50碱基、10个碱基至75碱基、10个碱基至100碱基、10个碱基至125碱基、10个碱基至150碱基、10个碱基至175碱基、10个碱基至200碱基、10个碱基至225碱基、10个碱基至250碱基、10个碱基至300碱基、25个碱基至50碱基、25个碱基至75碱基、25个碱基至100碱基、25个碱基至125碱基、25个碱基至150碱基、25个碱基至175碱基、25个碱基至200碱基、25个碱基至225碱基、25个碱基至250碱基、25个碱基至300碱基、50个碱基至75碱基、50个碱基至100碱基、50个碱基至125碱基、50个碱基至150碱基、50个碱基至175碱基、50个碱基至200碱基、50个碱基至225碱基、50个碱基至250碱基、50个碱基至300碱基、75个碱基至100碱基、75个碱基至125碱基、75个碱基至150碱基、75个碱基至175碱基、75个碱基至200碱基、75个碱基至225碱基、75个碱基至250碱基、75个碱基至300碱基、100个碱基至125碱基、100个碱基至150碱基、100个碱基至175碱基、100个碱基至200碱基、100个碱基至225碱基、100个碱基至250碱基、100个碱基至300碱基、125个碱基至150碱基、125个碱基至175碱基、125个碱基至200碱基、125个碱基至225碱基、125个碱基至250碱基、125个碱基至300碱基、150个碱基至175碱基、150个碱基至200碱基、150个碱基至225碱基、150个碱基至250碱基、150个碱基至300碱基、175个碱基至200碱基、175个碱基至225碱基、175个碱基至250碱基、175个碱基至300碱基、200个碱基至225碱基、200个碱基至250碱基、200个碱基至300碱基、225个碱基至250碱基、225个碱基至300碱基或250个碱基至300个碱基。

[0095] 在一些情况下,被加密的核酸序列导致目的核酸序列的至少10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或超过95%的富集。在一些情况下,被加密的核酸序列导致目的核酸序列的约10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或超过95%的富集。

[0096] 如本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法可以包括基于DNA或RNA的系统。规范DNA是具有可用的四种不同的核碱基——A、T、C或G(腺嘌呤、胸腺嘧啶、胞嘧啶和鸟嘌呤)的4碱基编码系统。因此,这4种碱基允许3碱基(使用少于全部碱基)或4碱基编码方案。此外,使用在RNA中发现的尿嘧啶(U)提供第五种碱基,并允许5碱基编码方案。另外,修饰的核碱基可用于编码大于4的核酸碱基编码。不是规范DNA核碱基的核碱基或修饰的核碱基包括但不限于尿嘧啶、3-meA(3-甲基腺嘌呤)、次黄嘌呤、8-氧代G(7,

8-二氢-8-氧代鸟嘌呤)、FapyG、FapyA、Tg(胸腺嘧啶二醇)、hoU(羟基尿嘧啶)、hmU(羟甲基尿嘧啶)、fU(甲酰基尿嘧啶)、hoC(羟基胞嘧啶)、fC(甲酰基胞嘧啶)、5-meC(5-甲基胞嘧啶)、6-meG(06-甲基鸟嘌呤)、7-meG(N7-甲基鸟嘌呤)、εC(亚乙烯基胞嘧啶)、5-caC(5-羧基胞嘧啶)、2-hA、εA(亚乙烯基腺嘌呤)、5-fU(5-氟尿嘧啶)、3-meG(3-甲基鸟嘌呤)和异径尿酸。本文进一步提供了这样的编码方案,其中机器指令将二进制序列形式的数字信息在最终被转化为最终核酸序列之前转化为中间代码。

[0097] 在一些情况下,为了将数据存储在DNA序列中,将信息从二进制代码的1和0转化为DNA的A、T、G和C碱基的代码。在一些情况下,首先以数字信息形式对信息项进行编码。在一些情况下,将数字信息的二进制代码转化为基于生物分子的(例如,基于DNA的)代码,同时保留该代码所代表的信息。就在本文公开的表面上沉积本文公开的生物分子而言,该转化的代码(数字二进制代码转化为生物分子代码)在本文中被称为导致“预定”序列。预定序列可以编码多个寡核苷酸的序列。

[0098] 二进制代码转化

[0099] 通常,初始代码是通常为计算机采用的二进制代码形式的数字信息。通用计算机是读取由数字“0”和“1”代表的“开”或“关”状态的电子设备。将这种二进制代码应用于计算机以读取多类型的信息项。在二进制算法中,数字2被写为数字10。例如,“10”表示“数字2的一次方加零”。数字“3”被写为“11”,表示“2的一次方加一”。数字“4”被写为“100”,数字“5”被写为“101”,“6”被写为“110”,诸如此类。表3中以小写和大写字母提供了用于二进制代码的美国标准代码II(ASCII)的实例。

[0100] 表3.二进制代码的美国标准代码II(ASCII)

[0101]

字母	ASCII 码	二进制	字母	ASCII 码	二进制	编号	ASCII 码	二进制
a	97	1100001	A	65	1000001	0	chr(0)	00000000
b	98	1100010	B	66	1000010	1	chr(1)	00000001
c	99	1100011	C	67	1000011	2	chr(2)	00000010
d	100	1100100	D	68	1000100	3	chr(3)	00000011
e	101	1100101	E	69	1000101	4	chr(4)	00000100
f	102	1100110	F	70	1000110	5	chr(5)	00000101
g	103	1100111	G	71	1000111	6	chr(6)	00000110
h	104	1101000	H	72	1001000	7	chr(7)	00000111
i	105	1101001	I	73	1001001	8	chr(8)	00001000
j	106	1101010	J	74	1001010	9	chr(9)	00001001
k	107	1101011	K	75	1001011	10	chr(10)	00001010
l	108	1101100	L	76	1001100	11	chr(11)	00001011
m	109	1101101	M	77	1001101	12	chr(12)	00001100
n	110	1101110	N	78	1001110	13	chr(13)	00001101
o	111	1101111	O	79	1001111	14	chr(14)	00001110
p	112	1110000	P	80	1010000	15	chr(15)	00001111
q	113	1110001	Q	81	1010001	16	chr(16)	00010000
r	114	1110010	R	82	1010010	17	chr(17)	00010001
s	115	1110011	S	83	1010011	18	chr(18)	00010010
t	116	1110100	T	84	1010100	19	chr(19)	00010011
u	117	1110101	U	85	1010101	20	chr(20)	00010100
v	118	1110110	V	86	1010110	21	chr(21)	00010101
w	119	1110111	W	87	1010111	22	chr(22)	00010110
x	120	1111000	X	88	1011000	23	chr(23)	00010111
y	121	1111001	Y	89	1011001	24	chr(24)	00011000
z	122	1111010	Z	90	1011010	25	chr(25)	00011001
						26	chr(26)	00011010
						27	chr(27)	00011011
						28	chr(28)	00011100
						29	chr(29)	00011101
						30	chr(30)	00011110

[0102] 本文提供了用于将第一代形式的信息(例如,二进制序列)转化为核酸序列的方法。该过程可以涉及从2碱基代码(即二进制)直接转化为更高级的碱基代码。示例性碱基代码包括2、3、4、5、6、7、8、9、10或更多。表4示出了各个碱基编号方案之间的示例性排比。接收机器转化指令的计算机可以自动将序列信息从一个代码转化为另一代码。

[0103] 表4.碱基编号方案的排比

[0104]

十进制	0	1	2	3	4	5	6	7	8	9
四进制	0	1	2	3	10	11	12	13	20	21
八进制	0	1	2	3	4	5	6	7	10	11
三进制	0	1	2	10	11	12	20	21	22	100
二进制	0	1	10	11	100	101	110	111	1000	1001

[0105] 核酸序列

[0106] 本文提供了用于设计本文所述的寡核苷酸的序列使得该核酸序列至少编码信息

项的一部分的方法。在一些情况下,每个寡核苷酸序列具有设计特征以便于在随后的装配步骤期间进行序列比对,并且也便于提供用于错误校正的手段。在一些布置中,设计寡核苷酸序列,使得每个寡核苷酸序列与群体中的另一寡核苷酸序列之间存在重叠。在一些情况下,每个寡核苷酸序列与仅一个其它寡核苷酸序列的一部分重叠,图5A。在替代的布置中,每个寡核苷酸序列区域与两个序列重叠,从而为单个寡核苷酸内的每个序列生成2个拷贝,图5B。在再一个其它布置中,每个寡核苷酸序列区域与超过两个序列重叠,从而为单个寡核苷酸内的每个序列生成3个拷贝,图5C。本文所述的寡核苷酸的序列可以编码10-2000、10-500、30-300、50-250或75-200个碱基的长度。在一些情况下,每个寡核苷酸序列的长度为至少10、15、20、25、30、50、100、150、200、500个或更多个碱基。

[0107] 本文提供了方法、系统和组合物,其中本文所述的每个寡核苷酸序列被设计为包含多个编码区和多个非编码区,图6A。在这样的布置中,每个编码区(例如,601、603、605)编码信息项的至少一部分。任选地,同一寡核苷酸中的每个编码区编码来自同一信息项的序列,并且如本文所述任选地采用重叠方案,图6B。在进一步的情况下,同一寡核苷酸中的每个编码区编码相同的序列,图6C。本文所述的寡核苷酸序列可编码至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个编码区。本文所述的寡核苷酸序列可编码至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个相同的编码区。在一些情况下,多个编码区中的每一个的长度为10-1000、20-500、30-300、50-250或75-200个碱基。在一些情况下,多个编码区中的每一个的长度为至少10、15、20、25、30、50、100、150、200个或更多个碱基。在一些情况下,每个寡核苷酸包含将分子连接到结构的表面602上的栓系区611。

[0108] 在多个编码序列存在于同一寡核苷酸中的布置中,切割区607任选地存在于每个编码区之间。切割区607可以存在于每个编码区之间的接点处,或者可以存在于在每个编码区之间具有一串序列的衔接子区内。切割区607可以编码一旦合成就将在施加切割信号之后从链上断裂下来的序列特征。切割区607可以编码限制酶识别位点、光敏性的且在施加电磁辐射时会断裂的修饰的核酸(例如,携带碱基敏感的S-新戊酰硫代乙基(t-Bu-SATE)磷酸三酯键的寡脱氧核苷酸杂聚物,其对>300nm的光波长敏感)或对施加某种化学物质(例如,胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244),其在施加氨气后断裂)敏感的修饰的核酸。因为从合成的寡核苷酸的测序中可能无法容易获知具有特定切割方案的序列的设计,所以切割方案提供了为由合成的核酸文库编码的序列增加安全性水平的手段。本文所述的寡核苷酸序列可编码至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个切割区。在一些情况下,每个切割区编码1-100、1-50、1-20、1-10、5-25或5-30个碱基的长度。在一些情况下,每个切割区编码至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、100个或更多个碱基。在一些布置中,对于每个寡核苷酸,每个编码区是相同的,并且每个编码区之间的每个切割区是不同的。例如,第一切割区607与第二切割区609不同。在一些布置中,最靠近表面602的切割区607与下一个远侧切割区607相同。

[0109] 条形码是允许鉴别与条形码关联的多核苷酸的一些特征的通常已知的核酸序列。图7A-图7B提供了说明性条形码布置。在图7A中,第一寡核苷酸701、第二寡核苷酸703和第三寡核苷酸705的每个编码区具有以下特征(从表面702向外):栓系区702、切割区707、第一

引物结合区701、条形码区703、编码区701、703、705以及第二引物结合区704。可以使用识别第一和/或第二引物结合区的引物扩增寡核苷酸。扩增可以发生在附接于表面或从表面释放的寡核苷酸上(即,经由切割区707处的切割)。在测序之后,条形码区703提供用于鉴别与编码区相关联的特性的指示。在一些情况下,条形码包含在与靶多核苷酸接合时用作衍生出靶多核苷酸的样品的标识物的核酸序列。条形码可以被设计为合适的长度以允许足够程度的鉴别,例如,至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55个或更多个碱基的长度。多个条形码,如2、3、4、5、6、7、8、9、10个或更多个条形码,可以在同一分子上使用,任选地被非条形码序列分隔开。在一些情况下,条形码的长度短于10、9、8、7、6、5或4个碱基。在一些情况下,与一些多核苷酸相关联的条形码在长度上不同于与其它多核苷酸相关联的条形码。通常,条形码具有足够的长度,并且包含足够不同从而允许基于与它们相关联的条形码来鉴别样品的序列。在一些布置中,可以在条形码序列中的一个或多个碱基的突变、插入或缺失,如1、2、3、4、5、6、7、8、9、10个或更多个碱基的突变、插入或缺失后,准确鉴别该条形码和与其相关的样品来源。在一些情况下,多个条形码中的每个条形码与所述多个条形码中的每个其它条形码至少有三个碱基位置如至少3、4、5、6、7、8、9、10个或更多个位置不同。本文提供的布置可包括条形码序列,该条形码序列指示编码数字序列的特定区域中的序列的核酸序列。例如,条形码序列可指示特定寡核苷酸序列在大文件中的编码位置。在一些情况下,条形码序列可指示特定寡核苷酸序列与哪个文件相关联。在一些情况下,条形码序列包含与特定序列的转化方案相关联的信息,从而提供额外的安全层。

[0110] 本文提供了寡核苷酸序列设计方案,其中寡核苷酸序列群体中的每个寡核苷酸序列被设计为在该群体中的寡核苷酸序列之间具有至少一个共同的区域。例如,同一群体中的所有寡核苷酸可包含一个或多个引物区。序列特异性引物区的设计允许从多个寡核苷酸的大文库中选择要在选定的批次中扩增的寡核苷酸。每个寡核苷酸序列可包含至少1、2、3、4、5、6、7、8、9、10个或更多个引物结合序列。寡核苷酸序列群体可包含至少1、2、3、4、5、6、7、8、9、10、25、50、100、200、500、1000、5000、10000、50000、100000个或更多个不相同的结合序列。引物结合序列可以在长度上包含5-100、10-75、7-60、8-60、10-50或10-40个碱基。

[0111] 用于寡核苷酸合成的结构

[0112] 本文提供了用于寡核苷酸合成的刚性或柔性结构,其与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用。在刚性结构的情况下,本文提供了具有用于生成寡核苷酸文库的结构(例如,板)的装置。图8中示出了示例性结构800,其中结构800具有与标准96孔板大致相同的大小尺寸:140mm×90mm。结构800包含在24个区域或子域805中分组的簇,每个子域805包含256个簇810的阵列。图9中示出了示例性子域805的展开视图。如图8和图9中所见的结构可以是基本平面的。在四个簇的展开视图中(图9),单个簇910的Y轴簇间距(相邻簇的中心到中心的距离)为1079.210μm或1142.694μm,并且X轴簇间距为1125μm。图10中描绘了说明性簇1010,其中Y轴座位间距(相邻座位的中心到中心的距离)为63.483μm,并且X轴座位间距为75μm。最长部分处的座位宽度(例如,圆形座位的直径)为50μm,并且座位之间的距离为24μm。图10中的示例性簇中座位1005的数目为121。座位(也称为“特征”)可以是平坦的、孔或通道。图11A-图11B中示出了示例性通道布置,其中示出了

包含主通道1110和连接到主通道1110的多个通道1115的板1105。主通道1110与多个通道1115之间的连接为从主通道1110到多个通道1115中的每个的流动路径提供流体连通。本文所述的板1105可包含多个主通道1110。多个通道1115共同形成主通道1110内的簇。

[0113] 在柔性结构的情况下,本文提供了这样的装置,其中柔性结构包含围绕一个或多个固定结构(例如,一对辊1203)缠绕的连续环1201或围绕单独的固定结构(例如,一对卷1205)缠绕的非连续柔性结构1207。参见图12A-图12B。本文提供了具有表面的柔性结构,该表面具有多个用于寡核苷酸延伸的特征(座位)。柔性结构1301的一部分中的每个特征可以是基本上平面的特征1303(例如,平坦的)、通道1305或孔1307。参见图13A-图13C。在一个示例性布置中,结构的每个特征的宽度约为10 μm ,并且每个结构的中心之间的距离约为21 μm 。参见图14A。特征可包含但不限于圆形的、矩形的、锥形的或弧形的形状。

[0114] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构可以包含通道。在一些情况下,本文所述的通道的宽度与深度(或高度)比为1至0.01,其中宽度是微通道最窄区段处宽度的测量值。在一些情况下,本文所述的通道的宽度与深度(或高度)比为0.5至0.01,其中宽度是微通道最窄区段处宽度的测量值。在一些情况下,本文所述的通道的宽度与深度(或高度)比为约0.01、0.05、0.1、0.15、0.16、0.2、0.5或1。

[0115] 本文提供了用于多核苷酸合成的结构,其包含多个用于多核苷酸合成的离散的座位、通道、孔或突起。本文描述的结构可包含多个簇,每个簇包含多个孔、基因座或通道。另外,本文描述了可包含孔、座位或通道的均匀排列的结构。在一些情况下,提供了本文描述的结构,其包含对应于簇内的多个特征(座位)的多个通道,其中通道的高度或深度为约5 μm 至约500 μm 、约5 μm 至约400 μm 、约5 μm 至约200 μm 、约5 μm 至约100 μm 、约5 μm 至约50 μm 或约10 μm 至约50 μm 。在一些情况下,通道的高度或深度小于100 μm 、小于80 μm 、小于60 μm 、小于40 μm 或小于20 μm 。在一些情况下,通道的高度或深度为约10、20、30、40、50、60、70、80、90、100、200、300、400、500 μm 或更高。在一些情况下,高度或深度为至少10、25、50、75、100、200、300、400、500、600、700、800、900、1000或超过1000nm。在一些情况下,高度或深度在约10nm至约1000nm、约25nm至约900nm、约50nm至约800nm、约75nm至约700nm、约100nm至约600nm或约200nm至约500的范围内。在一些情况下,高度或深度在约50nm至约1 μm 的范围内。

[0116] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构可以包含特征。在一些情况下,特征(例如,基本上平面的特征、孔、通道、座位或突起)的宽度为约0.1 μm 至约500 μm 、约0.5 μm 至约500 μm 、约1 μm 至约200 μm 、约1 μm 至约100 μm 、约5 μm 至约100 μm 或约0.1 μm 至约100 μm ,例如,约90 μm 、80 μm 、70 μm 、60 μm 、50 μm 、40 μm 、30 μm 、20 μm 、10 μm 、5 μm 、1 μm 或0.5 μm 。在一些情况下,特征(例如,微通道)的宽度小于约100 μm 、90 μm 、80 μm 、70 μm 、60 μm 、50 μm 、40 μm 、30 μm 、20 μm 或10 μm 。在一些情况下,特征的宽度为至少10、25、50、75、100、200、300、400、500、600、700、800、900、1000或高于1000nm。在一些情况下,特征的宽度在约10nm至约1000nm、约25nm至约900nm、约50nm至约800nm、约75nm至约700nm、约100nm至约600nm或约200nm至约500的范围内。在一些情况下,特征的宽度在约50nm至约1000nm的范围内。在一些情况下,两个相邻特征的中心之间的距离为约0.1 μm 至约500 μm 、0.5 μm 至约500 μm 、约1 μm 至约200 μm 、约1 μm 至约100 μm 、约5 μm 至约200 μm 、约5 μm 至约100 μm 、约5 μm 至约50 μm 或约5 μm 至约30 μm ,例如,约20 μm 。在一些情况下,特征的总宽

度为约5 μm 、10 μm 、20 μm 、30 μm 、40 μm 、50 μm 、60 μm 、70 μm 、80 μm 、90 μm 或100 μm 。在一些情况下,特征的总宽度为约1 μm 至100 μm ,30 μm 至100 μm 或50 μm 至70 μm 。在一些情况下,两个相邻特征的中心之间的距离为约0.5 μm 至约2 μm 、约0.5 μm 至约2 μm 、约0.75 μm 至约2 μm 、约1 μm 至约2 μm 、约0.2 μm 至约1 μm 、约0.5 μm 至约1.5 μm 、约0.5 μm 至约0.8 μm 或约0.5 μm 至约1 μm ,例如,约1 μm 。在一些情况下,特征的总宽度为约50nm、0.1 μm 、0.2 μm 、0.3 μm 、0.4 μm 、0.5 μm 、0.6 μm 、0.7 μm 、0.8 μm 、0.9 μm 、1 μm 、1.1 μm 、1.2 μm 、1.3 μm 、1.4 μm 或1.5 μm 。在一些情况下,特征的总宽度为约0.5 μm 至2 μm 、0.75 μm 至1 μm 或0.9 μm 至2 μm 。

[0117] 在一些情况下,每个特征支持与在另一特征上增长的寡核苷酸群体具有不同序列的寡核苷酸群体的合成。本文提供了包含至少10、100、256、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、11000、12000、13000、14000、15000、20000、30000、40000、50000个或更多个簇的表面。本文提供了包含超过2,000、5,000、10,000、20,000、30,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、5,000,000或10,000,000个或更多个不同特征的表面。在一些情况下,每个簇包含1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、120、130、150、200、500个或更多个特征。在一些情况下,每个簇包含50至500、50至200、50至150或100至150个特征。在一些情况下,每个簇包含100至150个特征。在示例性布置中,每个簇包含109、121、130或137个特征。

[0118] 本文提供了在最长区段处宽度为5至100 μm 的特征。在一些情况下,该特征在最长区段处的宽度为约30、35、40、45、50、55或60 μm 。在一些情况下,该特征是具有多区段的通道,其中每个区段的中心到中心的距离为5至50 μm 。在一些情况下,每个区段的中心到中心的距离分别为5、10、15、20或25 μm 。

[0119] 在一些情况下,在本文所述的结构的表面上合成的不同寡核苷酸的数目取决于基底中可用的不同特征的数目。在一些情况下,基底的簇内的特征密度为至少或约1个特征/ mm^2 、10个特征/ mm^2 、25个特征/ mm^2 、50个特征/ mm^2 、65个特征/ mm^2 、75个特征/ mm^2 、100个特征/ mm^2 、130个特征/ mm^2 、150个特征/ mm^2 、175个特征/ mm^2 、200个特征/ mm^2 、300个特征/ mm^2 、400个特征/ mm^2 、500个特征/ mm^2 、1,000个特征/ mm^2 或更大。在一些情况下,基底包含约10个特征/ mm^2 至约500个特征/ mm^2 、约25个特征/ mm^2 至约400个特征/ mm^2 、约50个特征/ mm^2 至约500个特征/ mm^2 、约100个特征/ mm^2 至约500个特征/ mm^2 、约150个特征/ mm^2 至约500个特征/ mm^2 、约10个特征/ mm^2 至约250个特征/ mm^2 、约50个特征/ mm^2 至约250个特征/ mm^2 、约10个特征/ mm^2 至约200个特征/ mm^2 或约50个特征/ mm^2 至约200个特征/ mm^2 。在一些情况下,簇内两个相邻特征中心之间的距离为约10 μm 至约500 μm 、约10 μm 至约200 μm 或约10 μm 至约100 μm 。在一些情况下,距相邻特征的两个中心之间的距离为大于约10 μm 、20 μm 、30 μm 、40 μm 、50 μm 、60 μm 、70 μm 、80 μm 、90 μm 或100 μm 。在一些情况下,距两个相邻特征的中心之间的距离为小于约200 μm 、150 μm 、100 μm 、80 μm 、70 μm 、60 μm 、50 μm 、40 μm 、30 μm 、20 μm 或10 μm 。在一些情况下,两个相邻特征的中心之间的距离小于约10000nm、8000nm、6000nm、4000nm、2000nm、1000nm、800nm、600nm、400nm、200nm、150nm、100nm、80nm、70nm、60nm、50nm、40nm、30nm、20nm或10nm。在一些实施方案中,本文所述结构的每平方米允许至少约 10^7 、 10^8 、 10^9 、 10^{10} 、 10^{11} 个特征,其中每个特征支持一个寡核苷酸。在一些情况下,在少于约6、5、4、3、2或1 m^2 的本文所述结构上支持 10^9 个寡核苷酸。

[0120] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构支持多个寡核苷酸的合成。在一些情况下,本文所述的结构为超过2,000、5,000、10,000、20,000、30,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、1,200,000、1,400,000、1,600,000、1,800,000、2,000,000、2,500,000、3,000,000、3,500,000、4,000,000、4,500,000、5,000,000、10,000,000个或更多个不相同的寡核苷酸的合成提供支持。在一些情况下,本文所述的结构为超过2,000、5,000、10,000、20,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、1,200,000、1,400,000、1,600,000、1,800,000、2,000,000、2,500,000、3,000,000、3,500,000、4,000,000、4,500,000、5,000,000、10,000,000个或更多个编码不相同的序列的寡核苷酸的合成提供支持。在一些情况下,至少一部分寡核苷酸具有相同的序列或被配置成用相同的序列合成。在一些情况下,该结构为具有至少约50、60、70、75、80、85、90、95、100、110、120、130、140、150、160、175、200、225、250、275、300、325、350、375、400、425、450、475、500个碱基或更多个碱基的寡核苷酸的增长提供表面环境。

[0121] 在一些情况下,在结构的不同特征上合成寡核苷酸,其中每个特征支持寡核苷酸群体的合成。在一些情况下,每个特征支持与在另一座位上增长的寡核苷酸群体具有不同序列的寡核苷酸群体的合成。在一些情况下,结构的特征位于多个簇内。在一些情况下,结构包含至少10、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、11000、12000、13000、14000、15000、20000、30000、40000、50000个或更多个簇。在一些情况下,结构包含超过2,000、5,000、10,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、1,100,000、1,200,000、1,300,000、1,400,000、1,500,000、1,600,000、1,700,000、1,800,000、1,900,000、2,000,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、1,200,000、1,400,000、1,600,000、1,800,000、2,000,000、2,500,000、3,000,000、3,500,000、4,000,000、4,500,000、5,000,000或10,000,000个或更多个不相同的特征。在一些情况下,每个簇包含1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、120、130、150个或更多个特征(座位)。在一些情况下,每个簇包含50至500、100至150或100至200个特征。在一些情况下,每个簇包含109、121、130或137个特征。在一些情况下,每个簇包含5、6、7、8、9、10、11或12个特征。在一些情况下,来自一个簇内不同特征的寡核苷酸具有在装配时编码预定序列的连续较长寡核苷酸的序列。

[0122] 结构大小

[0123] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构包括多种大小。在一些情况下,本文所述的结构为约标准96孔板的大小,例如约100至200mm乘约50至150mm。在一些情况下,本文所述的结构具有小于或等于约1000mm、500mm、450mm、400mm、300mm、250mm、200mm、150mm、100mm或50mm的直径。在一些情况下,基底的直径为约25mm至1000mm、约25mm至约800mm、约25mm至约600mm、约25mm至约500mm、约25mm至约400mm、约25mm至约300mm或约25mm至约200。基底大小的非限制性实例包括约300mm、200mm、150mm、130mm、100mm、76mm、51mm和25mm。在一些情况下,基底具有至少约100mm²、200mm²、500mm²、1,000mm²、2,000mm²、5,000mm²、10,000mm²、12,000mm²、15,000mm²、

20,000mm²、30,000mm²、40,000mm²、50,000mm²或更大的平面表面积。在一些情况下,基底的厚度为约50mm至约2000mm、约50mm至约1000mm、约100mm至约1000mm、约200mm至约1000mm或约250mm至约1000mm。基底厚度的非限制性实例包括275mm、375mm、525mm、625mm、675mm、725mm、775mm和925mm。在一些情况下,基底的厚度随直径而变化,并取决于基底的组成。例如,包含硅之外的材料的结构可以具有与相同直径的硅结构不同的厚度。结构厚度可以取决于所用材料的机械强度,并且该结构必须厚到足以在操作期间支撑其自身重量而不会破裂。在一些情况下,结构在任何一个维度上超过约1、2、3、4、5、10、15、30、40、50英寸。

[0124] 材料

[0125] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构可以由多种材料制成。在某些情况下,制造本公开的基底/固体支持物的材料表现出低水平的寡核苷酸结合。在一些情况下,可以采用对可见光和/或紫外线透明的材料。可以采用具有足够导电性的材料,例如能够在本文所述的全部或部分基底/固体支持物上形成均匀电场的那些材料。在一些情况下,这样的材料可以连接至电接地。在一些情况下,基底或固体支持物可以是导热的或隔热的。该材料可以是耐化学品且耐热的以支持化学或生物化学反应,如一系列寡核苷酸合成反应。对于柔性材料,感兴趣的材料可包括:改性尼龙和未改性尼龙、硝化纤维素、聚丙烯等。

[0126] 对于刚性材料,感兴趣的特定材料包括:玻璃;熔融石英;硅、塑料(例如,聚四氟乙烯、聚丙烯、聚苯乙烯、聚碳酸酯及其掺合物等);和金属(例如,金、铂等)。该结构可由选自硅、聚苯乙烯、琼脂糖、葡聚糖、纤维素聚合物、聚丙烯酰胺、聚二甲基硅氧烷(PDMS)和玻璃的材料制成。基底/固体支持物或者其中的显微结构、反应器可使用本文所列材料或本领域中已知的其它任何合适的材料的组合制成。

[0127] 术语“柔性”在本文中用于指能够弯曲、折叠或类似地操作而不破损的结构。在一些情况下,柔性结构围绕辊弯曲至少30度。在一些情况下,柔性结构围绕辊弯曲至少180度。在一些情况下,柔性结构围绕辊弯曲至少270度。在一些情况下,柔性结构围绕辊弯曲至少约360度。在一些情况下,辊的半径小于约10cm、5cm、3cm、2cm或1cm。在一些情况下,在20℃下,将该柔性结构在任一方向上反复弯曲和拉直至少100次而不会出问题(例如,破裂)或变形。在一些情况下,本文所述的柔性结构具有适合滚动的厚度。在一些情况下,本文所述的柔性结构的厚度小于约50mm、10mm、1mm或0.5mm。

[0128] 用于本文所述结构的示例性柔性材料包括但不限于尼龙(未改性尼龙、改性尼龙、透明尼龙),硝化纤维素,聚丙烯,聚碳酸酯,聚乙烯、聚氨酯,聚苯乙烯,缩醛,丙烯酸,丙烯腈,丁二烯苯乙烯(ABS),聚酯薄膜如聚对苯二甲酸乙二醇酯、聚甲基丙烯酸甲酯或其它丙烯酸,聚氯乙烯或其它乙烯基树脂,透明PVC箔、用于打印机的透明箔,聚(甲基丙烯酸甲酯)(PMMA),甲基丙烯酸酯共聚物,苯乙烯聚合物,高折射率聚合物,含氟聚合物,聚醚砜,含有脂环结构的聚酰亚胺,橡胶,织物,金属箔及其任何组合。各种增塑剂和改性剂可与聚合物基底材料一起使用以实现选定的柔性特性。

[0129] 本文所述的柔性结构可包含塑性材料,在一些情况下,该柔性结构包含热塑性材料。热塑性材料的非限制性实例包括丙烯酸、丙烯腈丁二烯苯乙烯、尼龙、聚乳酸、聚苯并咪唑、聚碳酸酯、聚醚砜、聚醚醚酮、聚醚酰亚胺、聚乙烯、聚苯醚、聚苯硫醚、聚丙烯、聚苯乙烯、聚氯乙烯和聚四氟乙烯。在一些情况下,所述基底包含聚芳醚酮(PEAK)家族中的热塑性

材料。PEAK热塑性塑料的非限制性实例包括聚醚酮(PEK)、聚醚酮酮(PEKK)、聚(醚醚酮酮)(PEEKK)、聚醚醚酮(PEEK)和聚醚酮醚酮酮(PEKEKK)。在一些情况下,该柔性结构包含与甲苯相容的热塑性材料。在一些情况下,通过添加增塑剂来增加塑性材料的柔性。增塑剂的实例为基于酯的增塑剂,如邻苯二甲酸酯。邻苯二甲酸酯增塑剂包括邻苯二甲酸二(2-乙基己基)酯(DEHP)、邻苯二甲酸二异壬酯(DINP)、邻苯二甲酸二正丁酯(DnBP,DBP)、邻苯二甲酸丁基苄酯(BBzP)、邻苯二甲酸二异癸酯(DIDP)、邻苯二甲酸二辛酯(DOP,DnOP)、邻苯二甲酸二异辛酯(DIOP)、邻苯二甲酸二乙酯(DEP)、邻苯二甲酸二异丁酯(DIBP)和邻苯二甲酸二正己酯。在一些情况下,通过共聚或通过聚合之前向单体添加非反应性侧链来修饰热塑性聚合物也增加了柔性。

[0130] 本文提供了可进一步包含氟弹性体的柔性结构。具有约80%氟弹性体的材料被命名为FKM。氟弹性体包含全氟弹性体(FFKM)和四氟乙烯/丙烯橡胶(FEPM)。氟弹性体具有五种已知类型。1型FKM由偏二氟乙烯(VDF)和六氟丙烯(HFP)组成,并且它们的氟含量通常为大约66重量%。2型FKM由VDF、HFP和四氟乙烯(TFE)组成,并且通常具有约68%至69%的氟。3型FKM由VDF、TFE和全氟甲基乙烯基醚(PMVE)组成,并且通常具有约62%至68%的氟。4型FKM由丙烯、TFE和VDF组成,并且通常具有约67%的氟。5型FKM由VDF、HFP、TFE、PMVE和乙烯组成。

[0131] 在一些情况下,本文公开的基底包含计算机可读材料。计算机可读材料包括但不限于磁性介质、卷到卷带、匣式磁带、盒式磁带、软盘、纸质介质、胶片、缩微胶片、连续带(例如,带)以及适用于存储电子指令的任何介质。在一些情况下,该基底包含磁性卷到卷带或磁性带。在一些情况下,该基底包含柔性印刷电路板。

[0132] 本文所述的结构对于可见光和/或紫外线可以是透明的。在一些情况下,本文所述的结构具有足够导电性以在整个或一部分结构上形成均匀电场。在一些情况下,本文所述的结构是导热的或隔热的。在一些情况下,该结构是耐化学品且耐热的以支持化学反应,如寡核苷酸合成反应。在一些情况下,该结构是磁性的。在一些情况下,该结构包含金属或金属合金。

[0133] 用于寡核苷酸合成的结构在任何维度上可以超过1、2、5、10、30、50英尺长或更长。在柔性结构的情况下,该柔性结构任选地以缠绕状态储存,例如以卷的状态储存。在大型刚性结构(例如,长度大于1英尺)的情况下,该刚性结构可以垂直或水平储存。

[0134] 结构表面上的加密密钥标记

[0135] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构可以包含加密标记。本文提供了具有标记1401的结构,其中该标记提供与源信息项相关的信息,该源信息项与附近的寡核苷酸群体、用于解密附近寡核苷酸群体的序列的加密方案、附近寡核苷酸群体的拷贝数或其任意组合相关联。参见例如图14B-图14C。标记可以是肉眼可见的,或者使用显微镜可以在放大的视图下可见。在一些情况下,表面上的标记仅在暴露标记的处理条件(例如热处理、化学处理或光处理(例如,用UV或IR线照射标记))之后可见。通过加热显色的示例性墨水包括但不限于氯化钴(加热时变为蓝色)。通过化学反应显色的示例性墨水包括但不限于酚酞、硫酸铜、硝酸铅(II)、氯化钴(II)和由硫酸锰和过氧化氢产生的草酸钴。

[0136] 表面准备

[0137] 与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用的用于寡核苷酸合成的结构可以包含用于寡核苷酸合成的表面。本文提供了支持将生物分子固定在基底上的方法,其中本文所述结构的表面包含材料并且/或者涂覆有促进与生物分子的偶联反应以用于附接的材料。为了制备用于生物分子固定的基底,可以利用表面修饰来通过加成工艺或减成工艺进行对基底表面的化学和/或物理改变,以改变基底表面或表面的选定位点或区域的一种或多种化学和/或物理性质。例如,表面修饰涉及:(1)改变表面的润湿性质;(2)对表面进行官能化,即,提供、修改或取代表面官能团;(3)对表面进行去官能化,即,移除表面官能团;(4)以其它方式例如通过刻蚀来改变表面的化学组成;(5)增大或减小表面粗糙度;(6)在表面上提供涂层,例如,展现出与表面的润湿性质不同的润湿性质的涂层;和/或(7)在表面上沉积微粒。在一些情况下,对结构的表面进行选择性的官能化以在结构上产生两个或更多个不同区域,其中至少一个区域具有与同一结构的另一个区域不同的表面或化学性质。这样的性质包括但不限于表面能、化学终止、化学部分的表面浓度等。

[0138] 在一些情况下,对本文公开的结构表面进行修饰以包含一个或多个活性官能化表面,其被配置为与基底表面和生物分子两者相结合,从而支持对表面的偶联反应。在一些情况下,表面还用不能有效结合生物分子的钝化材料进行官能化,从而防止在钝化官能化剂结合的位点处的生物分子附接。在一些情况下,表面包含仅限定用于生物分子支持的不同特征的活性层。

[0139] 在一些情况下,表面与有任何不同比率的官能化基团的混合物接触。在一些情况下,混合物包含至少2、3、4、5种或更多种不同类型的官能化剂。在一些情况下,混合物中的至少两种类型的表面官能化剂的比率为约1:1、1:2、1:5、1:10、2:10、3:10、4:10、5:10、6:10、7:10、8:10、9:10或用以达到两种基团的所需表面呈现的其它任何比率。在一些情况下,通过向基底表面提供合适比率的官能化剂来实现期望的表面张力、润湿性、水接触角和/或针对其它合适的溶剂的接触角。在一些情况下,混合物中的试剂选自合适的反应性和惰性部分,从而将反应性基团的表面密度稀释成用于下游反应的期望水平。在一些情况下,官能化剂的混合物包含一种或多种与生物分子结合的试剂和一种或多种不与生物分子结合的试剂。因此,试剂的调节允许控制在不同的官能化区域发生的生物分子结合的量。

[0140] 在一些情况下,用于基底官能化的方法包括将硅烷分子沉积到基底表面上。硅烷分子可以沉积在基底的高能表面上。在一些情况下,高表面能区域包括钝化官能化试剂。本文所述的方法提供硅烷基团以结合表面,而分子的剩余部分提供与表面的距离和在附接生物分子的末端处的游离羟基。在一些情况下,该硅烷为有机官能烷氧基硅烷分子。有机官能烷氧基硅烷分子的非限制性实例包括二甲基氯-十八烷基-硅烷、甲基二氯-十八烷基-硅烷、三氯-十八烷基-硅烷、三甲基-十八烷基-硅烷和三乙基-十八烷基-硅烷。在一些情况下,硅烷为氨基硅烷。氨基硅烷的实例包括但不限于11-乙酰氧基十一烷基三乙氧基硅烷、正癸基三乙氧基硅烷、(3-氨丙基)三甲氧基硅烷、(3-氨丙基)三乙氧基硅烷、缩水甘油基氧基丙基/三甲氧基硅烷和N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺。在一些情况下,硅烷包括11-乙酰氧基十一烷基三乙氧基硅烷、正癸基三乙氧基硅烷、(3-氨丙基)三甲氧基硅烷、(3-氨丙基)三乙氧基硅烷、缩水甘油基氧基丙基/三甲氧基硅烷、N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺或其任何组合。在一些情况下,活性官能化剂包括11-乙酰氧基十一

烷基三乙氧基硅烷。在一些情况下,活性官能化剂包括正癸基三乙氧基硅烷。在一些情况下,活性官能化剂包括缩水甘油基氧基丙基三乙氧基硅烷(GOPS)。在一些情况下,该硅烷为氟硅烷。在一些情况下,该硅烷为炔硅烷。在一些情况下,该硅烷为3-碘-丙基三甲氧基硅烷。在一些情况下,该硅烷为辛基氯硅烷。

[0141] 在一些情况下,通过与有机官能烷氧基硅烷分子自装配而在表面上进行硅烷化。有机官能烷氧基硅烷根据其有机官能来分类。硅烷官能化试剂的非限制性示例包括:羟烷基硅氧烷(甲硅烷基化表面,用乙硼烷官能化,并通过过氧化氢来氧化该醇)、二醇(二羟基烷基)硅氧烷(甲硅烷基化表面,并水解成二醇)、氨基烷基硅氧烷(胺不需要中间官能化步骤)、环氧丙氧基硅烷(3-环氧丙氧基丙基-二甲基-乙氧基硅烷、环氧丙氧基-三甲氧基硅烷)、巯基硅烷(3-巯基丙基-三甲氧基硅烷,3-4环氧环己基-乙基三甲氧基硅烷或3-巯基丙基-甲基-二甲氧基硅烷)、联环庚烯基-三氯硅烷、丁基-醛基-三甲氧基硅烷或二聚仲氨基烷基硅氧烷。示例性的羟烷基硅氧烷包括变成3-羟基丙基的烯丙基三氯氯硅烷或变成8-羟基辛基的7-辛-1-烯基三氯氯硅烷。二醇(二羟基烷基)硅氧烷包括缩水甘油基三甲氧基硅烷衍生的(2,3-二羟基丙氧基)丙基(GOPS)。氨基烷基硅氧烷包括变成3-氨基丙基的3-氨基三甲氧基硅烷(3-氨基丙基-三乙氧基硅烷、3-氨基丙基-二乙氧基-甲基硅烷、3-氨基丙基-二甲基-乙氧基硅烷或3-氨基丙基-三甲氧基硅烷)。在一些情况下,二聚仲氨基烷基硅氧烷为变成双(甲硅烷基氧基丙基)胺的双(3-三甲氧基甲硅烷基丙基)胺。

[0142] 活性官能化区域可以包含一种或多种不同种类的硅烷,例如,1、2、3、4、5、6、7、8、9、10种或更多种硅烷。在一些情况下,所述一种或多种硅烷中的一种以相比于另一种硅烷更高的量存在于官能化组合物中。例如,具有两种硅烷的混合硅烷溶液包含99:1、98:2、97:3、96:4、95:5、94:6、93:7、92:8、91:9、90:10、89:11、88:12、87:13、86:14、85:15、84:16、83:17、82:18、81:19、80:20、75:25、70:30、65:35、60:40、55:45的一种硅烷与另一种硅烷之比。在一些情况下,活性官能化剂包含11-乙酰氧基十一烷基三乙氧基硅烷和正癸基三乙氧基硅烷。在一些情况下,活性官能化剂包含比率为约20:80至约1:99、或约10:90至约2:98、或约5:95的11-乙酰氧基十一烷基三乙氧基硅烷和正癸基三乙氧基硅烷。

[0143] 在一些情况下,官能化包括通过任何沉积技术将官能化剂沉积到结构上,所述沉积技术包括但不限于化学汽相沉积(CVD)、原子层沉积(ALD)、等离子体增强CVD(PECVD)、等离子体增强ALD(PEALD)、金属有机CVD(MOCVD)、热丝CVD(HWCVD)、引发CVD(iCVD)、改进型CVD(MCVD)、汽相轴向沉积(VAD)、外汽相沉积(OVD)、物理汽相沉积(例如,溅射沉积、蒸发沉积)以及分子层沉积(MLD)。

[0144] 根据最终官能化的基底所需的性质,省略或改变以下官能化过程中的任何步骤或组分。在一些情况下,将额外的组分和/或过程步骤添加到本文所体现的过程工作流程中。在一些情况下,首先清洁基底,例如使用水虎鱼(piranha)溶液。清洁工艺的实例包括将基底在升高的温度(例如,120℃)下在水虎鱼溶液(例如,90% H_2SO_4 ,10% H_2O_2)中浸泡,以及洗涤(例如,水)并干燥该基底(例如,氮气)。该工艺任选地包括水虎鱼溶液后的处理,包括将经水虎鱼溶液处理的基底浸泡在碱性溶液(例如, NH_4OH)中,随后浸泡在水性洗液(例如,水)中。在一些情况下,任选地在水虎鱼溶液浸泡和可选的水虎鱼溶液后的处理之后,对结构的表面进行等离子体清洁。等离子体清洁工艺的实例包括氧等离子体蚀刻。在一些情况下,用活性官能化剂沉积表面,然后蒸发。在一些情况下,在清洁之前对基底进行活性官能

化,例如,通过水虎鱼处理和/或等离子体清洁。

[0145] 表面官能化的方法任选地包括抗蚀剂涂覆和抗蚀剂剥离。在一些情况下,在活性表面官能化之后,用抗蚀剂例如SPR™3612正性光致抗蚀剂旋涂基底。在各种情况下,表面官能化的过程包括具有经图案化的官能化的光刻。在一些情况下,在抗蚀剂涂覆之后进行光刻。在一些情况下,光刻后,目视检查表面的光刻缺陷。在一些情况下,表面官能化的过程包括清洁步骤,由此去除基底的残留物(例如,通过等离子体清洁或蚀刻)。在一些情况下,在光刻步骤之后的一些步骤进行等离子体清洁步骤。

[0146] 在一些情况下,例如在官能化之后和/或在光刻之后,处理涂覆有抗蚀剂的表面以去除抗蚀剂。在一些情况下,用溶剂,例如用包含N-甲基-2-吡咯烷酮的剥离溶液去除抗蚀剂。在一些情况下,抗蚀剂剥离包括声处理或超声处理。在一些情况下,涂覆和剥离抗蚀剂,然后对暴露区域进行活性官能化以产生所需的差异官能化图案。

[0147] 在一些情况下,本文所述的方法和组合物涉及施加光致抗蚀剂以在选择性区域中产生改性的表面性质,其中光致抗蚀剂的施加依赖于限定该光致抗蚀剂空间分布的表面的流体性质。不受理论束缚,与施加的流体相关的表面张力效应可以限定该光致抗蚀剂的流动。例如,表面张力和/或毛细作用效应可有助于在抗蚀剂溶剂蒸发之前以受控方式将光致抗蚀剂拉制成小结构。在一些情况下,抗蚀剂接触点通过尖锐边缘销连接,从而控制流体的前进。可以基于在制造和官能化过程期间用于施加光致抗蚀剂的所需的流动图案设计基础结构。溶剂蒸发后留下的固体有机层可用于继续进行制造过程的后续步骤。结构可以被设计为通过促进或抑制向相邻流体路径内的芯吸效应控制流体的流动。例如,设计结构以避免顶部边缘和底部边缘之间的重叠,这有利于将流体保持在顶部结构中,从而允许抗蚀剂的特定的布置。在替代的实例中,顶部边缘和底部边缘重叠,从而导致施加的流体芯吸到底部结构中。根据抗蚀剂的所需的施加,可以相应地选择适当的设计。

[0148] 在一些情况下,本文所述的结构具有包含厚度为至少或至少约0.1nm、0.5nm、1nm、2nm、5nm、10nm或25nm的材料的表面,所述材料包含能够结合核苷的反应性基团。示例性的表面包括但不限于玻璃和硅如二氧化硅和氮化硅。在一些情况下,示例性表面包括尼龙和PMMA。

[0149] 在一些情况下,紫外线形式的电磁辐射用于表面图案化。在一些情况下,灯用于表面图案化,并且掩模介导紫外线暴露于表面的位置。在一些情况下,激光器用于表面图案化,并且快门打开/关闭状态控制紫外线暴露于表面。激光器布置可以与能够移动的柔性结构结合使用。在这样的布置中,激光曝光和柔性结构移动的协调用于产生具有不同核苷偶联能力的一种或多种试剂的图案。

[0150] 材料沉积系统

[0151] 本文提供了用于在本文所述的结构上沉积和存储生物分子的系统和装置。在一些情况下,所述生物分子是将编码信息存储在其序列中的寡核苷酸。在一些情况下,该系统包括支持生物分子附接的结构表面和/或用于将生物分子施加至基底表面的装置。在一个实例中,用于施加生物分子的装置为寡核苷酸合成仪。在一些情况下,该系统包含用于采用流体处理基底的装置(例如,流动池)。在一些情况下,该系统包含用于在施加装置与处理装置之间移动基底的装置。对于其中基底为卷到卷带的情况,该系统可包含两个或更多个卷轴,其允许在不同时间将基底的的不同部分接入施加装置和可选的处理装置。

[0152] 图15中示出了用于寡核苷酸合成的寡核苷酸材料沉积系统的第一个实例。该系统包括在X-Y方向移动以与基底的位置对准的材料沉积装置。该材料沉积装置也可在Z方向移动以与基底密封,从而形成解析反应器。解析反应器被配置为允许将流体(包括寡核苷酸和/或试剂)从基底转移至加帽元件,和/或反之亦然。如图15所示,流体可穿过基底和加帽元件中的任一者或两者,并且包括但不限于偶联试剂、加帽试剂、氧化剂、解封闭剂、乙腈和氮气。能够进行高解析度小液滴沉积的装置的实例包括喷墨打印机和激光打印机的打印头。在本文所述的系统和方法中有用的装置实现约100个点/英尺(DPI)至约50,000DPI、约100DPI至约20,000DPI、约100DPI至约10,000DPI、约100DPI至约5,000DPI、约1,000DPI至约20,000DPI或约1,000DPI至约10,000DPI的解析度。在一些情况下,该装置具有至少约1,000、2,000、3,000、4,000、5,000、10,000、12,000DPI或20,000DPI的解析度。通过该装置进行的高解析度沉积与各个喷嘴的数目和密度(对应于基底的特征)相关。

[0153] 图16中示出了使用寡核苷酸合成仪在基底上从头合成寡核苷酸的示例性过程工作流程。将包含寡核苷酸合成试剂的小液滴以逐步的方式从材料沉积装置释放到基底,其中该材料沉积装置具有压电陶瓷材料和电极以将电信号转化为用于释放小液滴的机械信号。小液滴在基底表面上的特定位置上一次释放一个核碱基,以生成具有编码数据的预定序列的多个合成寡核苷酸。在一些情况下,将合成的寡核苷酸存储在基底上。核酸试剂可以以非连续或按需滴落方法沉积在基底表面上。在一些寡核苷酸合成方法中,核酸试剂以非连续或按需滴落方法沉积在基底表面上。这类方法的实例包括机电转移方法、电热转移方法和静电吸引方法。在机电转移方法中,由电脉冲变形的压电元件导致小液滴得以喷射。在电热转移方法中,在装置的腔室中生成气泡,并且气泡的膨胀力导致小液滴得以喷射。在静电吸引方法中,利用静电吸引力将小液滴喷射到基底上。在一些情况下,滴落频率为约5KHz至约500KHz;约5KHz至约100KHz;约10KHz至约500KHz;约10KHz至约100KHz;或约50KHz至约500KHz。在一些情况下,频率小于约500KHz、200KHz、100KHz或50KHz。

[0154] 分配的小液滴的大小与装置的解析度相关。在一些情况下,装置以约0.01p1至约20p1、约0.01p1至约10p1、约0.01p1至约1p1、约0.01p1至约0.5p1、约0.01p1至约0.01p1、或约0.05p1至约1p1的大小沉积试剂的小液滴。在一些情况下,小液滴的大小小于约1p1、0.5p1、0.2p1、0.1p1或0.05p1。由装置分配的小液滴的大小与沉积喷嘴的直径相关,其中每个喷嘴能够将试剂沉积到基底的特征上。在一些情况下,寡核苷酸合成仪的沉积装置包含约100至约10,000个喷嘴;约100至约5,000个喷嘴;约100至约3,000个喷嘴;约500至约10,000个喷嘴;或约100至约5,000个喷嘴。在一些情况下,沉积装置包含多于1,000、2,000、3,000、4,000、5,000或10,000个喷嘴。在一些情况下,每个材料沉积装置包含多个喷嘴,其中每个喷嘴任选地被配置为对应于基底上的特征。每个喷嘴可以沉积不同于另一个喷嘴的试剂组分。在一些情况下,每个喷嘴沉积覆盖基底的一个或多个特征的小液滴。在一些情况下,一个或多个喷嘴是成角度的。在一些情况下,多个沉积装置并排堆叠以实现通量的成倍增加。在一些实施方案中,增益为2x、4x、8x或更多。沉积装置的实例为Samba打印头(Fujifilm)。Samba打印头可以与Samba Web管理工具(SWAT)一起使用。

[0155] 通过在一定程度上使用并以剑角旋转相同的沉积装置可以增加沉积位点的数目。通过旋转沉积装置,每个喷嘴以对应于剑角的一定量的延迟时间进行喷射。这种不同步的喷射造成喷嘴之间的串扰。因此,当小液滴以不同于0度的某一剑角喷射时,来自喷嘴的小

液滴体积可以是不同的。

[0156] 在一些布置中,寡核苷酸合成系统的配置允许连续的寡核苷酸合成工艺,该工艺利用基底的柔性来以卷到卷式工艺行进。该合成工艺以连续生产线方式进行操作,其中使用一个或多个卷轴来旋转基底的位置使基底行进通过寡核苷酸合成的各个阶段。在示例性实施方案中,寡核苷酸合成反应包括滚动基底:通过在用于亚磷酰胺沉积的沉积装置下方的溶剂浴、通过氧化剂浴、通过乙腈洗涤浴,以及通过解封闭浴。任选地,带也穿过加帽浴。卷到卷式工艺允许包含合成的寡核苷酸的基底的最终产物容易地聚集在卷取卷轴上,在卷取卷轴处可将最终产物运输用于进一步处理或存储。

[0157] 在一些布置中,当连续柔性带沿着输送带系统输送时,寡核苷酸合成以连续工艺进行。类似于卷到卷式工艺,连续带上的寡核苷酸合成以生产线方式操作,其中基底在输送过程中行进通过寡核苷酸合成的各个阶段。然而,在输送带工艺中,连续带重新回到寡核苷酸合成步骤,而不需要滚动和展开带,如在卷到卷工艺中一样。在一些布置中,将寡核苷酸合成步骤划分为区段,并且连续带在循环中一次或多次输送通过每个区段。例如,寡核苷酸合成反应可包括(1)在循环中,将基底输送通过在用于亚磷酰胺沉积的沉积装置下方的溶剂浴、通过氧化剂浴、通过乙腈洗涤浴以及通过封闭浴;并随后(2)重复该循环以获得预定长度的合成的寡核苷酸。在寡核苷酸合成之后,将柔性基底从输送带系统移除,并且任选地将其卷起以便存储。可以围绕卷滚动以供存储。

[0158] 在示例性布置中,用核苷偶联剂涂覆包含热塑性材料的柔性基底。将涂层图案化成特征,使得每个特征具有约10 μ m的直径,两个相邻特征之间的中心到中心的距离为约21 μ m。在这种情况下,在寡核苷酸合成沉积步骤期间,特征大小足以容纳0.2p1的座滴体积。在一些情况下,特征密度为约22亿个特征/ m^2 (1个特征/ $441 \times 10^{-12} \text{m}^2$)。在一些情况下,4.5 m^2 的基底包含约100亿个特征,每个特征具有10 μ m的直径。

[0159] 本文所述的材料沉积装置可包含约2,048个喷嘴,每个喷嘴以1个核碱基/小液滴每秒沉积约100,000个小液滴。对于每个沉积装置,每天在基底上沉积至少约 1.75×10^{13} 个核碱基。在一些情况下,合成100至500个核碱基寡核苷酸。在一些情况下,合成200个核碱基寡核苷酸。任选地,在3天内,以每天约 1.75×10^{13} 个碱基的速率合成至少约 262.5×10^9 个寡核苷酸。

[0160] 在一些布置中,用于在合成反应期间将一种或多种试剂施加到基底的装置被配置成沉积用于基于核苷亚磷酰胺的合成的试剂和/或核苷酸单体。用于寡核苷酸合成的试剂包括用于寡核苷酸延伸的试剂和洗涤缓冲液。作为非限制性实例,该装置沉积清洁试剂、偶联试剂、加帽试剂、氧化剂、解封闭剂、乙腈、诸如氮气的气体及其任意组合。此外,该装置任选地沉积用于制备和/或维持基底完整性的试剂。在一些情况下,寡核苷酸合成仪以小于约1000、500、100、50或20p1的体积沉积直径小于约200 μ m、100 μ m或50 μ m的小液滴。在一些情况下,寡核苷酸合成仪每秒沉积约1至10000、1和5000、100和5000或1000至5000个小液滴。

[0161] 在一些布置中,在寡核苷酸合成过程中,将基底置于流动池内和/或密封在流动池内。流动池可提供液体的连续或不连续流动,诸如包含在基底内对于反应所必需的试剂(例如氧化剂和/或溶剂)的那些液体。流动池可提供气体(诸如氮气)的连续或不连续流动,以通常通过增强挥发性基底的蒸发来干燥基底。多种辅助装置可用来改善干燥并减少基底表面上的残留水分。这类辅助干燥装置的实例包括但不限于真空源、减压泵和真空罐。在一些

情况下,寡核苷酸合成系统包含一个或多个(如2、3、4、5、6、7、8、9、10或20个)流动池和一个或多个(如2、3、4、5、6、7、8、9、10或20个)基底。在一些情况下,流动池被配置为在合成反应的一个或多个步骤期间保持试剂并向基底提供试剂。在一些情况下,流动池包括在基底的顶部上滑动并且可被夹紧到位以在基底的边缘周围形成压力密封的盖子。足够的密封包括但不限于允许约1、2、3、4、5、6、7、8、9或10个大气压的密封。在一些情况下,打开流动池的盖子以允许接近施加装置如寡核苷酸合成仪。在一些情况下,寡核苷酸合成方法的一个或多个步骤在流动池内的基底上进行,而不需要进行基底的运输。

[0162] 在一些布置中,用于用流体处理基底的装置包括喷杆。用施加装置将核苷酸单体施加至基底表面上,随后喷杆使用该喷杆的喷嘴用一种或多种处理试剂喷射基底表面。在一些布置中,将喷嘴按顺序排列以与寡核苷酸合成过程中的不同处理步骤相关联。用于不同工艺步骤的化学物质可在喷杆中改变,以容易地适应合成方法或合成方法的步骤之间的变化。在一些情况下,当基底移动经过喷杆时,喷杆在基底表面上连续地喷射给定的化学物质。在一些情况下,喷杆沉积在大面积的基底上,非常像草坪洒水器中使用的喷杆。在一些情况下,喷杆喷嘴被定位成向基底的给定区域提供均匀的处理材料涂层。

[0163] 在一些情况下,寡核苷酸合成系统包含用于合成的寡核苷酸的下游加工的一个或多个元件。作为实例,该系统包含温度控制元件,如热循环装置。在一些情况下,温度控制元件与多个解析反应器一起使用以进行核酸装配如PCA和/或核酸扩增如PCR。

[0164] 从头寡核苷酸合成

[0165] 本文提供了用于短时间内在基底上进行高密度寡核苷酸的寡核苷酸合成的系统和方法,以供与本文所述的用于生物加密和/或生物解密的装置、组合物、系统和方法一起使用。在一些情况下,该基底是柔性基底。在一些情况下,一天内合成至少约 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 、 10^{14} 或 10^{15} 个碱基。在一些情况下,一天内合成至少约 10×10^8 、 10×10^9 、 10×10^{10} 、 10×10^{11} 或 10×10^{12} 个寡核苷酸。在一些情况下,每个合成的寡核苷酸包含至少约20、50、100、200、300、400或500个核碱基。在一些情况下,合成这些碱基的总平均错误率小于约1/100个碱基;1/200个碱基;1/300个碱基;1/400个碱基;1/500个碱基;1/1000个碱基;1/2000个碱基;1/5000个碱基;1/10000个碱基;1/15000个碱基;1/20000个碱基。在一些情况下,这些错误率是合成的寡核苷酸的至少50%、60%、70%、80%、90%、95%、98%、99%、99.5%或更多。在一些情况下,这些至少90%、95%、98%、99%、99.5%或更多的合成的寡核苷酸与其编码的预定序列无差异。在一些情况下,使用本文所述的方法和系统在基底上合成的寡核苷酸的错误率小于约1/200。在一些情况下,使用本文所述的方法和系统在基底上合成的寡核苷酸的错误率小于约1/1,000。在一些情况下,使用本文所述的方法和系统在基底上合成的寡核苷酸的错误率小于约1/2,000。在一些情况下,使用本文所述的方法和系统在基底上合成的寡核苷酸的错误率小于约1/3,000。在一些情况下,使用本文所述的方法和系统在基底上合成的寡核苷酸的错误率小于约1/5,000。各种类型的错误率包括在基底上合成的寡核苷酸的错配、缺失、插入和/或置换。术语“错误率”是指合成的寡核苷酸的总量与预定的多核苷酸序列的总和的比较。在一些情况下,本文公开的合成的寡核苷酸包含12至25个碱基的系链。在一些情况下,该系链包含10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个碱基。

[0166] 本公开的在基底上合成多核苷酸的适当方法是亚磷酰胺方法,其包括在亚磷酰胺结构单元与结合到基底的核苷之间形成亚磷酸三酯键的偶联步骤中将亚磷酰胺结构单元(即核苷亚磷酰胺)受控添加至增长的寡核苷酸链中。在一些情况下,将核苷亚磷酰胺提供给活化的基底。在一些情况下,将核苷亚磷酰胺提供给具有活化剂的基底。在一些情况下,核苷亚磷酰胺以相对于与基底结合的核苷1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50、60、70、80、90、100倍或更多倍的过量来提供给基底。在一些情况下,核苷亚磷酰胺的添加在无水环境中(例如,在无水乙腈中)进行。在偶联步骤中添加和连接核苷亚磷酰胺后,任选地洗涤该基底。在一些情况下,偶联步骤重复一次或额外多次,任选地在向基底添加核苷亚磷酰胺之间进行洗涤步骤。在一些情况下,本文使用的寡核苷酸合成方法包括1、2、3个或更多个连续的偶联步骤。在许多情况下,在偶联之前,与基底结合的核苷通过去除保护基团来脱保护,其中该保护基团起到防止聚合的作用。常见的保护基团为4,4'-二甲氧基三苯甲基(DMT)。

[0167] 偶联后,亚磷酰胺寡核苷酸合成方法任选地包括加帽步骤。在加帽步骤中,用加帽剂处理增长的寡核苷酸。加帽步骤通常用来在偶联后封闭未反应的与基底结合的5'-OH基团以防止进一步链延伸,从而防止形成具有内部碱基缺失的寡核苷酸。此外,用1H-四唑活化的亚磷酰胺通常在很小的程度上与鸟苷的O6位置反应。不受理论的束缚,在用I₂/水氧化后,该副产物(可能经由O6-N7迁移)经历脱嘌呤。无嘌呤位点可终止在寡核苷酸的最终脱保护过程中被切割,从而降低全长产物的产率。O6修饰可通过在用I₂/水氧化之前用加帽试剂处理而去除。在一些情况下,与没有加帽的合成相比,在寡核苷酸合成过程中包括加帽步骤会降低错误率。作为实例,加帽步骤包括用乙酸酐和1-甲基咪唑的混合物处理与基底结合的寡核苷酸。在加帽步骤之后,任选地洗涤所述基底。

[0168] 在添加核苷亚磷酰胺之后,并且任选地在加帽和一个或多个洗涤步骤之后,对与基底结合的增长的核酸进行氧化。氧化步骤包括将亚磷酸三酯氧化成四配位的磷酸三酯——天然存在的磷酸二酯核苷间连接的受保护的前体。在一些情况下,增长的寡核苷酸的氧化通过任选地在弱碱(如吡啶、二甲基吡啶、三甲吡啶)的存在下用碘和水处理来实现。氧化有时在无水条件下采用叔丁基过氧化氢或(1S)-(+)-(10-樟脑磺酰基)-氧杂吡丙啉(CSO)进行。在一些方法中,在氧化之后进行加帽步骤。第二个加帽步骤允许基底干燥,因为可能持续存在的来自氧化的残余水可以抑制随后的偶联。氧化后,任选地洗涤基底和增长的寡核苷酸。在一些情况下,氧化步骤用硫化步骤来代替,以获得寡核苷酸硫代磷酸,其中任何加帽步骤均可在硫化之后进行。许多试剂能够进行有效的硫转移,包括但不限于3-(二甲基氨基亚甲基)氨基)-3H-1,2,4-二噻唑-3-硫酮、DDTT、3H-1,2-苯并二噻戊环-3-酮1,1-二氧化物(也被称为Beaucage试剂)和N,N,N'-四乙基秋兰姆二硫化物(TETD)。

[0169] 为了使后续核苷掺入循环通过偶联而发生,必须除去与基底结合的增长的寡核苷酸的受保护的5'末端,使得伯羟基可与下一个核苷亚磷酰胺反应。在一些情况下,保护基团为DMT,并且用在二氯甲烷中的三氯乙酸进行解封闭。进行延长时间的脱三苯甲基化或者使用比推荐的酸溶液更强的酸溶液进行脱三苯甲基化可导致与固体支持物结合的寡核苷酸的脱嘌呤增加,因此降低了所需全长产物的产率。本文所述的方法和组合物提供了受控的解封闭条件,从而限制不希望的脱嘌呤反应。在一些情况下,与基底结合的寡核苷酸在解封闭后洗涤。在一些情况下,解封闭后的有效洗涤有助于以低错误率合成寡核苷酸。

[0170] 本文所述的在基底上合成寡核苷酸的方法通常包括一系列迭代的以下步骤:将受保护的单体施加至基底特征的表面以与该表面、连接体或与预先脱保护的单体连接;使所施加的单体脱保护,使其可与随后施加的受保护的单体反应;以及施加另一种受保护的单体以供连接。一个或多个中间步骤包括氧化或硫化。在一些情况下,在一个或全部步骤之前或之后有一个或多个洗涤步骤。

[0171] 在一些情况下,合成具有光不稳定的保护基团的寡核苷酸,其中表面上生成的羟基被光不稳定的保护基团封闭。当该表面例如通过光刻掩模暴露于紫外线时,可在表面上生成游离羟基的图案。这些羟基可按照亚磷酰胺化学法与光保护的核苷亚磷酰胺反应。可采用第二光刻掩模,且表面可暴露于紫外线以生成羟基的第二图案,随后与5'-光保护的核苷亚磷酰胺偶联。同样,可生成图案并可使寡聚物链延伸。不受理论的束缚,可光切割的基团的不稳定性取决于波长和所采用的溶剂的极性,并且光切割的速率可受暴露的持续时间和光的强度影响。该方法可利用许多因素,例如,掩模对准的准确性、光保护基团去除的效率和亚磷酰胺偶联步骤的产率。此外,不期望的光向邻近位点的泄露可被最小化。每个斑点合成寡聚物的密度可通过调整合成表面上的先导核苷的负载来监测。

[0172] 为寡核苷酸合成提供支持的基底的表面可以被化学修饰,以允许从表面上切割合成的寡核苷酸链。在一些情况下,在寡核苷酸脱保护的同时切割寡核苷酸链。在一些情况下,在寡核苷酸链脱保护之后切割寡核苷酸链。在示例性方案中,三烷氧基甲硅烷基胺如 $(\text{CH}_3\text{CH}_2\text{O})_3\text{Si}-(\text{CH}_2)_2-\text{NH}_2$ 与基底的表面 SiOH 基团反应,然后与琥珀酸酐和胺反应以产生支持核酸链增长的酰胺键和游离 OH 。切割包括用氨或甲胺进行气体切割。在一些情况下,一旦从表面释放,寡核苷酸就被装配成更大的核酸,对其进行测序并解码以提取存储的信息。

[0173] 寡核苷酸可以被设计为共同跨越编码信息的预定序列的大区域。在一些情况下,通过连接反应连接合成的寡核苷酸来生成较大的寡核苷酸。连接反应的一个实例是聚合酶链装配(PCA)。在一些情况下,至少一部分寡核苷酸被设计为包含作为通用引物结合的基底的附加区域。对于PCA反应,预先合成的寡核苷酸包括彼此的重叠(例如,具有重叠序列的4、20、40个或更多个碱基)。在聚合酶循环过程中,寡核苷酸与互补片段退火,然后被聚合酶补平。根据哪些寡核苷酸彼此发现,每个循环随机增加各个片段的长度。片段之间的互补性允许形成完整的大跨度的双链DNA。在一些情况下,在PCA反应完成之后,使用错配修复检测酶来进行错误校正步骤以去除序列中的错配。一旦生成较大的靶序列片段,就可以对该片段进行扩增。例如,在一些情况下,包含5'和3'末端衔接子序列的靶序列在聚合酶链反应(PCR)中进行扩增,该PCR包括与衔接子序列杂交的修饰的引物。在一些情况下,修饰的引物包含一个或多个尿嘧啶碱基。使用修饰的引物允许通过集中于靶向修饰的碱基和/或由从片段切割修饰的碱基对的酶所留下的缺口的酶促反应去除引物。剩下的是缺少衔接子序列残余物的双链扩增产物。以这种方式,多种扩增产物可以与同一组引物平行生成,以生成不同的双链DNA片段。

[0174] 可以对合成的寡核苷酸和/或装配的产物进行错误校正。用于错误校正的示例性策略涉及通过重叠延伸PCR进行定点诱变以纠正错误,其任选地与两轮或更多轮克隆和测序相结合。在某些情况下,选择性地从正确合成的核酸群体中去除具有错配、凸起和小环、化学改变的碱基和/或其它异源双链体的双链核酸。在一些情况下,使用识别并结合或紧挨着双链核酸内的错配或未配对的碱基的蛋白质/酶进行错误校正,以产生单链或双链断裂

或启动链转移转座事件。用于错误校正的蛋白质/酶的非限制性实例包括内切核酸酶(T7内切核酸酶I、大肠杆菌内切核酸酶V、T4内切核酸酶VII、绿豆核酸酶、细胞大肠杆菌内切核酸酶IV、UVDE)、限制酶、糖基化酶、核糖核酸酶、错配修复酶、解离酶、解旋酶、连接酶、错配特异性抗体及其变体。特异性错误校正酶的实例包括T4内切核酸酶7、T7内切核酸酶1、S1、绿豆内切核酸酶、MutY、MutS、MutH、MutL、切割酶(cleavase)、CELI和HINF1。在一些情况下,DNA错配结合蛋白MutS(水生栖热菌(*Thermus aquaticus*))用于从合成产物群体中去除失败产物。在一些情况下,使用校正酶(Correctase)进行错误校正。在一些情况下,使用SURVEYOR内切核酸酶(Transgenomic)——一种扫描异源双链体DNA中的已知和未知的突变以及多态性的错配特异性DNA内切核酸酶,来进行错误校正。

[0175] 释放、提取和装配

[0176] 本文提供了用于可复制信息存储的方法和装置。在一些情况下,合成相同编码区的多个拷贝、多核苷酸、相同簇、包含寡核苷酸的结构的部分或包含寡核苷酸的整个结构。当合成相同寡核苷酸的多个拷贝时,每个寡核苷酸可以附接到表面的不同区域。可以通过断裂或切割分离不同的区域。或者,每个寡核苷酸可以以斑点、孔或通道的形式存在于特征上并且是可单独访问的。例如,使特征与切割试剂接触然后与水接触将使寡核苷酸的一个拷贝游离,同时保留其它拷贝完整。类似地,在整个区域或整个板上切割寡核苷酸允许访问一部分重复群体。重复群体可以存在于分离的卷、板、带等中。在柔性材料(如带)的情况下,可以切割复制区域并且可以将带的其余区域拼接在一起。或者,可以通过使用引物和DNA聚合酶扩增附接于结构表面的寡核苷酸来获得合成和存储的寡核苷酸的核酸信息。

[0177] 在一些情况下,将水性或气态转移介质沉积在结构中的一个或多个通道上,以将寡核苷酸从该结构转移到接收单元。例如,转移介质可以穿过结构中的通道以粘附寡核苷酸、收集寡核苷酸并将寡核苷酸从结构中的通道转移到接收单元。在一些情况下,采用电荷传导特征和施加的电压将转移介质吸引或排斥到结构中的通道或通过结构中的通道。在一些情况下,采用滑动件将转移介质引导至结构中的通道。在一些情况下,采用压力释放将转移介质引导至结构中或穿过结构中的通道。在一些情况下,使用喷嘴来形成高压的局部区域,其迫使转移介质进入或通过结构中的通道。在一些情况下,采用销将寡核苷酸从结构中的通道转移至容器至接收单元。在这样的情况下,销可包括促进转移介质粘附的试剂。在一些情况下,通过在导电特征与结构之间形成电压电势,采用电荷传导特征将转移介质吸引或排斥到结构中的通道或通过结构中的通道。在一些情况下,使用移液管尖端或其它诱导毛细管流的结构经由毛细管流转移流体和寡核苷酸。在一些情况下,容器包含一个或多个隔室,每个隔室接收从单个相应通道发射的一部分转移介质及其中的一个或多个寡核苷酸。在一些情况下,容器包含单个隔室,其接收从一个或多个结构通道发射的转移介质的一个或多个部分(每个部分中包含一个或多个寡核苷酸)。

[0178] 测序

[0179] 在从结构表面提取和/或扩增寡核苷酸后,可以采用合适的测序技术对寡核苷酸进行测序。在一些情况下,在基底上或在结构的特征内读取DNA序列。在一些情况下,提取存储在基底上的寡核苷酸并将其任选地装配成更长的核酸,然后对其进行测序。

[0180] 在本文所述结构上合成并存储的寡核苷酸编码可通过读取合成的寡核苷酸的序列并将该序列转化为计算机可读的二进制代码进行解读的数据。在一些情况下,序列需要

装配,并且装配步骤可能需要处于核酸序列阶段或数字序列阶段。

[0181] 本文提供了检测系统,其包含能够直接在结构上和/或在从主结构移除之后对存储的寡核苷酸进行测序的装置。在结构是卷到卷带的柔性材料的情况下,该检测系统包含用于保持结构并推进结构通过检测位置的装置和安置在检测位置附近用于检测源自该带的一部分(当该部分处于检测位置时)的信号的检测器。在一些情况下,该信号指示存在寡核苷酸。在一些情况下,该信号指示寡核苷酸的序列(例如,荧光信号)。在一些情况下,当将带连续传送通过可操作地连接到计算机的检测器时,该计算机读取在连续带上的寡核苷酸内编码的信息。在一些情况下,检测系统包含计算机系统,其包含寡核苷酸测序装置、用于存储和检索与寡核苷酸序列有关的数据的数据库、用于将寡核苷酸序列的DNA代码转化为二进制代码的软件、用于读取二进制代码的计算机或其任意组合。

[0182] 计算机系统

[0183] 在各个方面,本文描述的任何系统均可操作地连接至计算机,并且任选地本地或远程地通过计算机进行自动化。在各种情况下,本公开的方法和系统进一步包括计算机系统上的软件程序及其使用。因此,对于分配/抽真空/再填充功能的同步(如编排和同步材料沉积装置运动、分配动作和真空致动)的计算机化控制处于本公开内容的范围内。计算机系统可被编程为在用户指定的碱基序列与材料沉积装置的位置之间接合,以将正确的试剂递送至基底的指定区域。

[0184] 图17中示出的计算机系统1700可被理解为能够从介质1711和/或网络端口1705读取指令的逻辑设备,其可任选地连接至具有固定介质1712的服务器1709。诸如图17示出的系统可包括CPU 1701、磁盘驱动器1703、可选的输入设备如键盘1715和/或鼠标1716以及可选的监视器1707。可通过示出的通信媒介实现与本地或远程位置处的服务器的数据通信。通信媒介可包括传输和/或接收数据的任何手段。例如,通信媒介可以是网络连接、无线连接或因特网连接。这样的连接可提供经由万维网的通信。可以预期有关本公开内容的数据可经过这样的网络或连接而传输,以便由用户方1722接收和/或审阅。

[0185] 图18是示出可与本公开的示例实施方案结合使用的计算机系统1800的第一示例架构的框图。如图18所示,该示例计算机系统可包括用于处理指令的处理器1802。处理器的非限制性实例包括: Intel Xeon™处理器、AMD Opteron™处理器、Samsung 32-位RISC ARM 1176JZ (F) -S v1.0™处理器、ARM Cortex-A8 Samsung S5PC100™处理器、ARM Cortex-A8 Apple A4™处理器、Marvell PXA 930™处理器或功能上等效的处理器。多个执行线程可用于并行处理。在一些情况下,也可以使用多个处理器或具有多个核的处理器,无论是在单一计算机系统中,在集群中,还是通过包含多个计算机、蜂窝电话和/或个人数据助理设备的网络跨系统分布。

[0186] 如图18所示,高速缓冲存储器1804可连接至或并入处理器1802,以提供由处理器1802新近或频繁使用的指令或数据的高速存储器。处理器1802通过处理器总线1808连接至北桥1806。北桥1806通过存储器总线1812连接至随机存取存储器(RAM) 1810,并管理处理器1802对RAM 1810的访问。北桥1806还通过芯片集总线1816连接至南桥1814。南桥1814又连接至外围总线1818。外围总线可以是例如PCI、PCI-X、PCI Express或其它外围总线。北桥和南桥通常被称为处理器芯片集,并管理在处理器、RAM与外围总线1818上的外围组件之间的数据传送。在一些备选的架构中,北桥的功能性可以并入处理器中,而不是使用单独的北桥

芯片。

[0187] 在一些情况下,系统1800可包括附接至外围总线1818的加速器卡1822。加速器可包括现场可编程门阵列 (FPGA) 或用于加速某个处理的其它硬件。例如,加速器可用于适应性数据重建或用来评价在扩展集处理中使用的代数表达式。

[0188] 软件和数据存储在外部存储器1824中,并可加载至RAM 1810和/或高速缓冲存储器1804中,以供处理器使用。系统1800包括用于管理系统资源的操作系统;操作系统的非限制性实例包括:Linux、Windows™、MACOS™、iOS™和其它功能上等价的操作系统,以及在操作系统顶部运行的、用于根据本公开内容的示例实例管理数据存储和优化的应用软件。

[0189] 在该实例中,系统1800还包括与外围总线连接的网络接口卡 (NIC) 1820和1821,以提供与外部存储如网络附加存储 (NAS) 和可用于分布式并行处理的其它计算机系统的网络接口。

[0190] 图19是显示了具有多个计算机系统1902a和1902b、多个蜂窝电话和个人数据助理1902c以及网络附加存储 (NAS) 1904a和1904b的网络1900的示图。在示例实例中,系统1902a、1902b和1902c可管理数据存储并优化对存储在网络附加存储 (NAS) 1904a和1904b中的数据的数据访问。数学模型可用于该数据,并使用跨计算机系统1902a和1902b和蜂窝电话以及个人数据助理系统1902c的分布式并行处理进行评价。计算机系统1902a和1902b和蜂窝电话以及个人数据助理系统1902c也可提供对存储在网络附加存储 (NAS) 1904a和1904b中的数据的适应性数据重建的并行处理。图19仅示出了一个实例,而多种多样的其它计算机架构和系统可与本公开的多个实施方案一起使用。例如,刀片式服务器可用来提供并行处理。处理器刀片可通过背板连接,以提供并行处理。存储还可通过单独的网络接口连接至背板或作为网络附加存储 (NAS) 。

[0191] 在一些示例实施方案中,处理器可维持单独的存储空间,并通过网络接口、背板或其它连接器传输数据以便由其它处理器并行处理。在其它情况下,部分或全部处理器可使用共享的虚拟地址存储空间。

[0192] 图20是根据示例实施方案使用共享虚拟地址存储空间的多处理器计算机系统2000的框图。该系统包括可访问共享的存储器子系统2004的多个处理器2002a-f。该系统中并入存储器子系统2004中的多个可编程硬件存储算法处理器 (MAP) 2006a-f。MAP 2006a-f中的每一个可包括存储器2008a-f和一个或多个现场可编程门阵列 (FPGA) 2010a-f。MAP提供可配置的功能单元,并且可向FPGA 2010a-f提供特定算法或算法的部分,以便与各自的处理器密切协调处理。例如,在示例实施方案中,MAP可用来评价与数据模型相关的代数表达式以及用来进行适应性数据重建。在该示例中,每个MAP可被用于这些目的的所有处理器全局访问。在一种配置中,每个MAP可使用直接存储器访问 (DMA) 来访问相关联的存储器2008a-f,使其独立于且异步于各自的微处理器2002a-f而执行任务。在这一配置中,MAP可将结果直接馈送至另一MAP以用于流水处理和并行执行算法。

[0193] 以上计算机架构和系统仅为实例,并且多种多样的其它计算机、蜂窝电话和个人数据助理架构和系统可与示例实施方案结合使用,包括使用通用处理器、协处理器、FPGA和其它可编程逻辑设备、芯片上系统 (SOC)、专用集成电路 (ASIC) 和其它处理和逻辑元件的任何组合的系统。在一些情况下,全部或部分计算机系统可用软件或硬件来实现。任何种类的数据存储介质可与示例实例结合使用,包括随机存取存储器、硬盘驱动器、闪速存储器、磁

带驱动器、磁盘阵列、网络附加存储 (NAS) 和它的本地或分布式数据存储设备和系统。

[0194] 在示例实施方案中,计算机系统可使用在任何上述或其它计算机架构和系统上执行的软件模块来实现。在其它实例中,该系统的功能可部分或完全地在固件、可编程逻辑设备如现场可编程门阵列 (FPGA)、芯片上系统 (SOC)、专用集成电路 (ASIC) 或其它处理和逻辑元件中实现。例如,集处理器 (Set Processor) 和优化器 (Optimizer) 可通过使用硬件加速器卡如加速器卡用硬件加速方式实现。

[0195] 本文提供了用于存储信息的方法,其包括:将至少一个数字序列形式的信息项转化为至少一个核酸序列;提供具有表面的柔性结构;合成共同编码所述至少一个核酸序列的具有预定序列的多个寡核苷酸,其中所述多个寡核苷酸包含至少约100,000个寡核苷酸,并且其中所述多个寡核苷酸从所述柔性结构的表面延伸;以及存储所述多个寡核苷酸。本文进一步提供了这样的方法,其中合成包括:在预定位置的表面上沉积核苷;以及移动所述柔性结构的至少一部分穿过浴或来自喷杆的排放物。本文进一步提供了这样的方法,其中所述浴或来自喷杆的排放物将所述结构的表面暴露于氧化剂或解封闭剂。本文进一步提供了这样的方法,其中合成进一步包括对沉积在所述表面上的核苷进行加帽。本文进一步提供了这样的方法,其中所述核苷包括核苷亚磷酰胺。本文进一步提供了这样的方法,其中所述柔性结构包括卷到卷带或连续带。本文进一步提供了这样的方法,其中所述柔性结构包括热塑性材料。本文进一步提供了这样的方法,其中所述热塑性材料包括聚芳醚酮。本文进一步提供了这样的方法,其中所述聚芳醚酮为聚醚酮、聚醚酮酮、聚(醚醚酮酮)、聚醚醚酮或聚醚酮醚酮酮。本文进一步提供了这样的方法,其中所述柔性结构包括尼龙、硝化纤维素、聚丙烯、聚碳酸酯、聚乙烯、聚氨酯、聚苯乙烯、缩醛、丙烯酸、丙烯腈、丁二烯苯乙烯、聚对苯二甲酸乙二醇酯、聚甲基丙烯酸甲酯、聚氯乙烯、透明PVC箔、聚(甲基丙烯酸甲酯)、苯乙烯聚合物、含氟聚合物、聚醚砜或聚酰亚胺。本文进一步提供了这样的方法,其中所述多个寡核苷酸中的每个寡核苷酸在长度上包含50至500个碱基。本文进一步提供了这样的方法,其中所述多个寡核苷酸包含至少约100亿个寡核苷酸。本文进一步提供了这样的方法,其中在24小时内合成至少约 1.75×10^{13} 个核碱基。本文进一步提供了这样的方法,其中在72小时内合成至少约 262.5×10^9 个寡核苷酸。本文进一步提供了这样的方法,其中所述信息项为文本信息、音频信息或视觉信息。本文进一步提供了这样的方法,其中所述核苷包括核苷亚磷酰胺。

[0196] 本文提供了用于存储信息的方法,其包括:将至少一个数字序列形式的信息项转化为至少一个核酸序列;提供具有表面的结构;合成共同编码所述至少一个核酸序列的具有预定序列的多个寡核苷酸,其中所述多个寡核苷酸包含至少约100,000个寡核苷酸,其中所述多个寡核苷酸从所述结构的表面延伸,并且其中合成包括:清洁所述结构的表面;在预定位置处将核苷沉积在所述表面上;对沉积在所述表面上的核苷进行氧化、解封闭和任选地加帽;其中清洁、氧化、解封闭和加帽包括移动所述柔性结构的至少一部分穿过浴或来自喷杆的排放物;以及存储所述多个寡核苷酸。本文进一步提供了其中所述核苷包括核苷亚磷酰胺的方法。

[0197] 阐述以下实施例是为了向本领域技术人员更清楚地说明本文所公开的实施方案的原理和实践,而不应解释为限制任何请求保护的实施方案的范围。除非另有说明,否则所有份数和百分比均以重量计。

[0198] 实施例

[0199] 实施例1:装置表面的官能化

[0200] 将装置进行官能化以支持寡核苷酸文库的附着和合成。首先使用包含90% H_2SO_4 和10% H_2O_2 的水虎鱼溶液 (piranha solution) 将装置表面润湿清洗20分钟。将该装置在含有去离子水的数个烧杯中冲洗,在去离子水鹅颈旋塞下保持5min,并用 N_2 干燥。随后将该装置在 NH_4OH (1:100; 3mL:300mL) 中浸泡5min,使用手持式喷枪 (handgun) 用去离子水冲洗,在连续三个含有去离子水的烧杯中各浸泡1min,然后再使用手持式喷枪用去离子水冲洗。然后通过将装置表面暴露于 O_2 来等离子体清洗该装置。使用SAMCO PC-300仪器在下游模式下以250瓦进行 O_2 等离子体蚀刻1min。

[0201] 使用具有以下参数的YES-1224P气相沉积烘箱系统,用包含N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺的溶液对清洁的装置表面进行活性官能化:0.5至1托,60min,70℃,135℃汽化器。使用Brewer Science 200X旋涂仪对装置表面进行抗蚀剂涂覆。将 SPR^{TM} 3612光致抗蚀剂以2500rpm旋涂在装置上40sec。该装置在Brewer热板上以90℃预烘30min。使用Karl Suss MA6掩模对准仪对装置进行光刻。将该装置暴露2.2sec并在MSF 26A中显影1min。剩余的显影剂用手持式喷枪冲洗,并将装置在水中浸泡5min。该装置在烘箱中以100℃烘烤30min,随后使用Nikon L200目视检查光刻缺陷。采用残余去除工艺利用SAMCO PC-300仪器以250瓦进行 O_2 等离子体蚀刻1min来去除残余抗蚀剂。

[0202] 用与10 μL 轻质矿物油混合的100 μL 全氟辛基三氯硅烷溶液对装置表面进行钝化官能化。将该装置放置于腔室中,泵送10min,随后关闭通往泵的阀门并静置10min。使该腔室排气。该装置通过在70℃下在500mL NMP中进行两次5min浸泡并同时以最大功率(在Crest系统上的9)进行超声波处理来剥离抗蚀剂。然后将该装置在室温下在500mL异丙醇中浸泡5min,同时以最大功率进行超声波处理。将该装置浸入300mL的200标准酒精度 (proof) 的乙醇中并用 N_2 吹干。活化该官能化表面以充当寡核苷酸合成的支持物。

[0203] 实施例2:在寡核苷酸合成装置上合成50-聚体序列

[0204] 将二维寡核苷酸合成装置装配至流动池中,其与流动池 (Applied Biosystems (ABI394DNA合成仪)) 连接。该二维寡核苷酸合成装置用N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺 (Gelest) 均匀地官能化,并用来使用本文所述的寡核苷酸合成方法合成50bp的示例性寡核苷酸 (“50-聚体寡核苷酸”)。

[0205] 所述50-聚体的序列如SEQ ID NO.:1所述。5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT###TTTTTTTTTT3' (SEQ ID NO.:1),其中#表示胸苷-琥珀酰基己酰胺CED亚磷酰胺 (来自ChemGenes的CLP-2244),它是允许在脱保护过程中从表面上释放寡核苷酸的可切割的连接体。

[0206] 根据表5中的方案和ABI合成仪,使用标准DNA合成化学法 (偶联、加帽、氧化和解封闭) 完成合成。

[0207] 表5:合成方案

[0208]

表 5		
通用 DNA 合成 工艺名称	工艺步骤	时间(sec)
洗涤(乙腈洗液流)	乙腈系统冲洗	4
	乙腈至流动池	23
	N2 系统冲洗	4
	乙腈系统冲洗	4
DNA 碱基添加(亚磷酸胺+活化剂流)	活化剂歧管冲洗	2
	活化剂至流动池	6
	活化剂+亚磷酸胺至流动池	6
	活化剂至流动池	0.5
	活化剂+亚磷酸胺至流动池	5
	活化剂至流动池	0.5
	活化剂+亚磷酸胺至流动池	5
	活化剂至流动池	0.5
	活化剂+亚磷酸胺至流动池	5
	孵育 25sec	25
洗涤(乙腈洗液流)	乙腈系统冲洗	4
	乙腈至流动池	15
	N2 系统冲洗	4
	乙腈系统冲洗	4
DNA 碱基添加(亚磷酸胺+活化剂流)	活化剂歧管冲洗	2
	活化剂至流动池	5
	活化剂+亚磷酸胺至流动池	18
	孵育 25sec	25
洗涤(乙腈洗液流)	乙腈系统冲洗	4
	乙腈至流动池	15
	N2 系统冲洗	4
	乙腈系统冲洗	4
加帽(帽 A+B, 1:1, 流)	帽 A+B 至流动池	15
洗涤(乙腈洗液流)	乙腈系统冲洗	4
	乙腈至流动池	15
	乙腈系统冲洗	4
氧化(氧化剂流)	氧化剂至流动池	18
洗涤(乙腈洗液流)	乙腈系统冲洗	4
	N2 系统冲洗	4
	乙腈系统冲洗	4

	表 5	
	通用 DNA 合成 工艺名称	工艺步骤 时间(sec)
[0209]		乙腈至流动池
		乙腈系统冲洗
		乙腈至流动池
		N2 系统冲洗
		乙腈系统冲洗
		乙腈至流动池
		N2 系统冲洗
		乙腈系统冲洗
	解封闭(解封闭流)	解封闭剂至流动池
	洗涤(乙腈洗液流)	乙腈系统冲洗
		N2 系统冲洗
		乙腈系统冲洗
		乙腈至流动池
		N2 系统冲洗
		乙腈系统冲洗
		乙腈至流动池

[0210] 亚磷酰胺/活化剂组合以类似于本体试剂通过流动池递送的方式进行递送。当在全部时间内保持环境被试剂“润湿”时,不进行干燥步骤。

[0211] 从ABI 394合成仪中去除限流器,以使得能够更快速流动。在没有限流器的情况下,酰胺类(amidites)(在ACN中0.1M)、活化剂(在ACN中的0.25M苯甲酰基硫基四唑(“BTT”;来自GlenResearch的30-3070-xx))和Ox(在20%吡啶、10%水和70%THF中的0.02M I2)的流速大致为约100uL/sec,乙腈(“ACN”)和加帽试剂(帽A和帽B的1:1混合物,其中帽A是在THF/吡啶中的乙酸酐,帽B是在THF中的16%1-甲基咪唑(1-methylimidazole))的流速大致为约200uL/sec,而解封闭剂(在甲苯中的3%二氯乙酸)的流速大致为约300uL/sec(相比之下,在有限流器的情况下,所有试剂的流速均为约50uL/sec)。观测完全排出氧化剂的时间,相应地调节化学品流动时间的时间选择,并在不同的化学品之间引入额外的ACN洗涤。在寡核苷酸合成后,将芯片在75psi下在气态氨中脱保护过夜。将五滴水施加到表面上以装配寡核苷酸。然后在BioAnalyzer小RNA芯片上分析所装配的寡核苷酸(数据未示出)。

[0212] 实施例3:在寡核苷酸合成装置上合成100-聚体序列

[0213] 使用实施例2中描述的用于合成50-聚体序列的相同过程,在两个不同的硅芯片上合成100-聚体寡核苷酸(“100-聚体寡核苷酸”;5’CGGGATCCTTATCGTCATCGTCGTACAGATCCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTT3’,其中#表示胸苷-琥珀酰基己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244);SEQ ID NO.:2),第一个用N-(3-三乙氧基甲硅烷基丙基)-4-羟基丁酰胺均匀地官能化,而第二个用11-乙酰氧基十一烷基三乙氧基硅烷和正癸基三乙氧基硅烷的5/95混合物官能化,并在BioAnalyzer仪器上分析从表面提取的寡核苷酸(数据未示出)。

[0214] 使用下列热循环程序,在50uL PCR混合物(25uL NEB Q5主混合物,2.5uL 10uM正向引物,2.5uL 10uM反向引物,1uL从表面提取的寡核苷酸,用水加至50uL)中使用正向引物

(5'ATGCGGGGTTCTCATCATC3';SEQ ID NO.:3)和反向引物(5'CGGGATCCTTATCGTCATCG3';SEQ ID NO.:4)进一步PCR扩增来自两个芯片的全部十个样品:

[0215] 98°C,30sec

[0216] 98°C,10sec;63°C,10sec;72°C,10sec;重复12个循环

[0217] 72°C,2min

[0218] PCR产物还在BioAnalyzer上运行(数据未示出),在100-聚体位置处显示出尖锐峰。然后,对PCR扩增的样品进行克隆,并进行Sanger测序。表6总结了从来自芯片1的斑点1-5采集的样品和从来自芯片2的斑点6-10采集的样品的Sanger测序结果。

[0219] 表6:测序结果

[0220]	斑点	错误率	循环效率
	1	1/763 bp	99.87%
	2	1/824 bp	99.88%
	3	1/780 bp	99.87%
[0221]	斑点	错误率	循环效率
	4	1/429 bp	99.77%
	5	1/1525 bp	99.93%
	6	1/1615 bp	99.94%
	7	1/531 bp	99.81%
	8	1/1769 bp	99.94%
	9	1/854 bp	99.88%
	10	1/1451 bp	99.93%

[0222] 因此,合成的寡核苷酸的高质量和均匀性在具有不同表面化学的两个芯片上重现。总体上,89%,相当于被测序的262个100-聚体中的233个,是没有错误的完美序列。

[0223] 表7总结了从来自斑点1-10的寡核苷酸样品中获得的序列的错误特征。

[0224] 表7:错误特征

[0225]

样品编号/ 斑点号	OSA_0046/1	OSA_0047/2	OSA_0048/3	OSA_0049/4	OSA_0050/5	OSA_0051/6	OSA_0052/7	OSA_0053/8	OSA_0054/9	OSA_0055/10
全序列	32	32	32	32	32	32	32	32	32	32
测序质量	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28
寡核苷酸质量	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI 匹配计数	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 突变	2	2	1	3	1	0	2	1	2	1
ROI 多碱基缺失	0	0	0	0	0	0	0	0	0	0
ROI 小插入	1	0	0	0	0	0	0	0	0	0
ROI 单碱基缺失	0	0	0	0	0	0	0	0	0	0
大缺失计数	0	0	1	0	0	1	1	0	0	0
突变: G>A	2	2	1	2	1	0	2	1	2	1
突变: T>C	0	0	0	1	0	0	0	0	0	0
ROI 错误计数	3	2	2	3	1	1	3	1	2	1
ROI 错误率	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约	错误: 约

[0226]

	1/834	1/1350	1/1282	1/708	1/2500	1/2667	1/876	1/2900	1/1400	
ROI-引物错误率	MP 错误: 约 1/763	MP 错误: 约 1/824	MP 错误: 约 1/780	MP 错误: 约 1/429	MP 错误: 约 1/1525	MP 错误: 约 1/1615	MP 错误: 约 1/531	MP 错误: 约 1/1769	MP 错误: 约 1/854	MP 错误: 约 1/1451

[0227] 实施例4:高度准确的基于DNA的信息存储和装配

[0228] 以二进制数据的形式选择数字信息,总计约0.2GB,包含超过100种语言的“世界人权宣言(Universal Declaration of Human Rights)”的内容、Project Guttenberg的前100本书和种子数据库。将数字信息加密成基于核酸的序列并划分为字符串。以类似于实施例2中所述的方式,在刚性硅表面上合成超过1千万个不同的寡核苷酸,每个对应于一个字符串。每个不同的寡核苷酸的长度等于或小于200个碱基。收集合成的寡核苷酸并对其进行测序,并解码回数字代码,与初始的至少一个数字序列相比,其对于源数字信息具有100%的准确度。

[0229] 实施例5:将数字信息转化为核酸序列

[0230] 计算机txt文件包括文本信息。通用计算机使用具有根据接收到的指令将序列转化为3、4或5碱基序列的机器指令的软件程序。3碱基中的每个数字被指定核酸(例如,A=0,T=1,C=2)。4碱基中的每个数字被指定核酸(例如,A=0,T=1,C=2,G=3)。或者,使用5碱基五进制序列,其中5碱基中的每个数字被指定核酸(例如,A=0,T=1,C=2,G=3,U=4)。如表8所示生成序列。然后提供关于从头合成编码该核酸序列的寡核苷酸的机器指令。

[0231] 表8.序列转化

[0232]	文本	Jack went up the hill.
	二进制序列	01001010011000010110001101101011001000000111011101100101011011100111010000100000011101010111000000100000011101000110100001100101001000000110100001101001011010001101000010100000110100001010
	三进制序列	10101020110002210101002110201222120010111220221000212200221020001112212102011201021112122200101110001002001022002222211002222112
[0233]	四进制序列	10221201120312230200131312111232131002001311130002001310122012110200122012211230123002320031002200310022
	五进制序列	33221433013301230301312300103024443334330043122410302032021020112342341100431241100334213

[0234] 实施例6:具有高密度特征的柔性表面

[0235] 用核苷偶联试剂涂覆包含热塑性材料的柔性结构。对涂层剂进行图案化以获得高密度的特征。图14A中示出了柔性表面的一部分。每个特征具有10 μ m的直径,且两个相邻特征之间的中心距为21 μ m。在寡核苷酸合成沉积步骤期间,特征大小足以容纳0.2p1的座滴体积。小特征尺寸允许在基底表面上合成高密度的寡核苷酸。特征密度为22亿个特征/ m^2 (1个特征/ $441 \times 10^{-12} \text{m}^2$)。4.5 m^2 的基底被制造成具有100亿个特征,每个特征具有10 μ m的直径。将柔性结构任选地放置在连续环路系统(图12A)或卷到卷系统(图12B)中用于寡核苷酸合成。

[0236] 实施例7:柔性结构上的寡核苷酸合成

[0237] 制备在热塑性柔性材料上包含多个特征的柔性结构。该结构用作使用包含沉积装置的寡核苷酸合成装置合成寡核苷酸的支持物。该柔性结构为柔性介质的形式,非常像磁性卷到卷带。

[0238] 从头合成以连续生产线方式进行,其中结构行进通过溶剂浴,随后行进至一堆打印头下方,在此处将亚磷酸胺印刷到该结构的表面上。使表面上沉积有座滴的柔性结构滚入氧化剂浴,然后带从氧化浴中涌出并浸入乙腈洗涤浴中,然后浸入解封闭浴中。任选地,该带穿过加帽浴。在替代的工作流程中,柔性结构从氧化浴中涌出并在洗涤步骤中用乙腈进行喷射。

[0239] 或者,使用喷杆代替液体浴。在该工艺中,仍然采用喷墨装置将核苷酸沉积在表面上,但充溢步骤现在在具有喷嘴的腔室中完成。例如,沉积装置具有2,048个喷嘴,每个喷嘴以1个核碱基/小液滴每秒沉积100,000个小液滴。存在顺序排列的喷嘴,以模拟标准亚磷酸胺化学法中的充溢步骤的顺序。这种技术可容易地改变喷杆中装载的化学物质,以适应不同的工艺步骤。以与如实施例2中所述的方式相同的方式将寡核苷酸脱保护或切割。

[0240] 对于每个沉积装置,每天在结构上沉积超过 1.75×10^{13} 个核碱基。合成超过200个核碱基寡核苷酸。在3天内,以每天 1.75×10^{13} 个碱基的速率合成 262.5×10^9 个寡核苷酸。

[0241] 实施例8:选择生物加密

[0242] 接收关于供转化的所需信息项以及关于一种或多种类别的生物加密的机器指令,该生物加密选自基于酶的(例如,CRISPR/Cas复合物和限制酶消化物)、基于电磁辐射的(例如,光解和光检测)、化学裂解(例如,气态氨或甲胺处理,以裂解胸苷-琥珀酰己酰胺CED亚

磷酸胺(来自ChemGenes的CLP-2244))和基于亲和力的(例如,用于杂交的序列标签,或对捕获试剂具有增强的亲和力的修饰核苷酸的掺入)形式的生物加密。在接收到特定的生物加密选择之后,程序模块执行以下步骤:将信息项转化为核酸序列,并施加关于设计生物加密版本的序列的设计指令。选择生物加密类别中的特定加密子类型。然后将合成指令提供给材料沉积装置以供从头合成寡核苷酸。

[0243] 实施例9:选择的生物解密

[0244] 提供关于施加一种或多种类别的生物解密的机器指令,该生物解密选自基于酶的(例如,CRISPR/Cas复合物或限制酶消化物)、基于电磁辐射的(例如,光解或光检测)、基于化学裂解的(例如,气态氨或甲胺处理,以裂解胸苷-琥珀酰己酰胺CED亚磷酸胺(来自ChemGenes的CLP-2244))和基于亲和力的(例如,用于杂交的序列标签,或对捕获试剂具有增强的亲和力的修饰核苷酸的掺入)形式的生物解密。在接收到特定的生物解密选择之后,程序模块执行释放调节剂以富集寡核苷酸的步骤。富集后,对寡核苷酸进行测序,任选地与更长的核酸序列进行比对,并将其转化为与信息项相对应的数字序列。

[0245] 实施例10:用CRISPR/Cas9对DNA序列的生物加密和解密

[0246] 接收编码信息项的数字序列。然后将该数字序列转化为核酸序列。该核酸序列在较大的核酸序列群体中被加密。加密过程包括添加“垃圾”区域,以通过CRISPR/Cas9复合物进行检测和去除。核酸序列如实施例2-3中那样合成。

[0247] 将包含加密的核酸序列的核酸序列群体与Cas9和gRNA在Cas9缓冲液中混合,并在37°C下孵育2小时。然后使Cas9失活并通过纯化除去。然后通过下一代测序分析所纯化的样品。

[0248] 实施例11:用CRISPR/Cas9对DNA序列的生物加密和解密,包括序列交换

[0249] 接收编码信息项的数字序列,将该数字序列转化为核酸序列。通过使用CRISPR/Cas9系统和指导RNA序列添加特定序列来加密该核酸序列。核酸序列如实施例2-3中那样合成。

[0250] 然后将核酸序列与荧光标记的与交换序列互补的探针混合。从群体中去除由荧光标记的探针鉴定的核酸序列。

[0251] 实施例12:使用限制酶消化物对DNA序列的生物加密和解密

[0252] 接收编码信息项的数字序列,并将该数字序列转化为核酸序列。通过添加被限制酶EcoRI识别的特定序列来加密核酸序列群体。核酸序列如实施例2-3中那样合成,并储存。

[0253] 将核酸序列与EcoRI一起孵育。切割包含EcoRI识别位点的加密的核酸序列。在切割加密的核酸序列后,将具有互补突出端的序列与释放的DNA杂交并连接。然后分离连接的复合物,对纯化的样品进行测序,并装配原始的数字信息。

[0254] 实施例13:使用光解对DNA序列的生物加密和解密

[0255] 接收编码信息项的数字序列,并将该数字序列转化为核酸序列。核酸序列群体被设计为包括可光切割的核碱基。核酸序列如实施例2-3中那样合成,并储存。

[0256] 将280nm的UV-B辐射施加至核酸序列。切割并去除包含可光切割位点的加密核酸序列。然后收集核酸序列并进行测序。或者,核酸序列例如通过氨气裂解而从结构的表面上释放,然后暴露于电磁辐射以提供核苷酸序列的断裂。例如通过使用结合有互补捕获探针的珠子的下拉试验,使用被选择为仅扩增靶序列的引物的PCR,或大小排阻色谱法,来富集

群体的部分。然后对富集的核酸进行测序,将其转化为数字序列,并接收信息项。

[0257] 实施例14:使用化学富集对DNA序列的生物加密和解密

[0258] 接收编码信息项的数字序列,并将该数字序列转化为核酸序列。通过添加可被氨气化学裂解的特定序列(例如,胸苷-琥珀酰己酰胺CED亚磷酰胺(来自ChemGenes的CLP-2244))对核酸序列群体进行加密。核酸序列如实施例2-3中那样合成。

[0259] 将氨气施加到核酸序列。使用本文所述的富集方法,从群体中释放并富集包含化学可裂解序列的加密核酸序列。然后对富集的核酸进行测序,将其转化为数字序列,并接收信息项。

[0260] 实施例15:使用包含生物素的核酸探针对DNA序列的生物加密和解密

[0261] 接收编码信息项的数字序列,并将该数字序列转化为核酸序列。通过将预定的残基设计为包含生物素的核碱基对核酸序列群体进行加密。核酸序列如实施例2-3中那样合成。

[0262] 从结构上切下核酸序列,并与含有链霉亲和素的珠子混合。然后将核酸序列与链霉亲和素磁珠一起孵育。包含生物素的核酸序列被磁珠拉下。然后对富集的核酸进行测序,将其转化为数字序列,并接收信息项。

[0263] 实施例16:使用光检测对DNA序列的生物加密和解密

[0264] 接收编码信息项的数字序列,并将该数字序列转化为核酸序列。通过设计为包括被Alexa488标记的核酸探针所识别的特定序列对核酸序列群体进行加密。核酸序列如实施例2-3中那样合成。

[0265] 从结构上释放核酸序列,并与Alexa488标记的核酸探针混合。然后根据荧光强度对核酸序列进行分类。进一步分析了用Alexa488标记的核酸探针所标记的核酸序列。然后对探针结合的核酸进行测序,将其转化为数字序列,并接收信息项。

[0266] 实施例17:使用修饰的核苷酸对DNA序列的生物加密和解密

[0267] 接收编码信息项的数字序列,将该数字序列转化为核酸序列。通过针对在预定位置添加包含肽核酸(PNA)的预定核碱基进行设计以及设计限制酶识别位点以切除包含PNA的部分,对核酸序列群体进行加密。核酸序列如实施例2-3中那样合成。

[0268] 释放核酸序列,进行限制酶消化,然后通过PCR进行扩增。包含PNA的核酸序列无法扩增。然后对富集、扩增的核酸进行测序,将其转化为数字序列,并接收信息项。

[0269] 实施例18:使用CRISPR/Cas9和化学裂解对DNA序列的生物加密和解密

[0270] 接收编码信息项的数字序列,将该数字序列转化为核酸序列。通过使用CRISPR/Cas9和指导RNA序列添加特定序列,对核酸序列群体进行加密。CRISPR/Cas9系统在核酸序列的预选位置引入化学可裂解位点。核酸序列如实施例2-3中那样合成。

[0271] 将氨气施加到核酸序列。切割并通过大小排阻纯化法去除包含化学可裂解位点的加密核酸序列,并通过下一代测序进行分析。

[0272] 虽然本文已经示出并描述了本公开内容的优选实施方案,但对于本领域技术人员明显的是,这些实施方案仅通过示例的方式提供。本领域技术人员在不脱离本发明的情况下将会想到许多变化、改变和替代。应当理解,可在实施本发明时采用本文描述的本发明实施方案的各种替代方案。旨在以所附权利要求限定本发明的范围,并且由此涵盖这些权利要求范围内的方法和结构及其等同物。

- [0001] 序列表
- [0002] <110> 特韦斯特生物科学公司
- [0003] <120> 基于核酸的数据存储
- [0004] <130> 44854-738.601
- [0005] <140>
- [0006] <141>
- [0007] <150> 62/462,284
- [0008] <151> 2017-02-22
- [0009] <160> 28
- [0010] <170> PatentIn version 3.5
- [0011] <210> 1
- [0012] <211> 62
- [0013] <212> DNA
- [0014] <213> 人工序列
- [0015] <220>
- [0016] <223> 人工序列的描述:合成寡核苷酸
- [0017] <400> 1
- [0018] agacaatcaa ccatttgagg tggacagcct tgacctctag acttcggcat tttttttttt 60
- [0019] tt 62
- [0020] <210> 2
- [0021] <211> 112
- [0022] <212> DNA
- [0023] <213> 人工序列
- [0024] <220>
- [0025] <223> 人工序列的描述:合成多核苷酸
- [0026] <400> 2
- [0027] cgggatcctt atcgatcatg tcgtacagat cccgacccat ttgctgtcca ccagtcgatgc 60
- [0028] tagccatacc atgatgatga tgatgatgag aaccccgcat tttttttttt tt 112
- [0029] <210> 3
- [0030] <211> 19
- [0031] <212> DNA
- [0032] <213> 人工序列
- [0033] <220>
- [0034] <223> 人工序列的描述:合成引物
- [0035] <400> 3
- [0036] atgcgggggtt ctcatcatc 19
- [0037] <210> 4
- [0038] <211> 20

[0039]	<212> DNA
[0040]	<213> 人工序列
[0041]	<220>
[0042]	<223> 人工序列的描述:合成引物
[0043]	<400> 4
[0044]	cgggatacctt atcgatcatcg 20
[0045]	<210> 5
[0046]	<211> 11
[0047]	<212> DNA
[0048]	<213> 人工序列
[0049]	<220>
[0050]	<223> 人工序列的描述:合成寡核苷酸
[0051]	<220>
[0052]	<221> 修饰的碱基
[0053]	<222> (3) .. (6)
[0054]	<223> a、c、t、g、未知或其它
[0055]	<400> 5
[0056]	acnnnngtay c 11
[0057]	<210> 6
[0058]	<211> 11
[0059]	<212> DNA
[0060]	<213> 人工序列
[0061]	<220>
[0062]	<223> 人工序列的描述:合成寡核苷酸
[0063]	<220>
[0064]	<221> 修饰的碱基
[0065]	<222> (3) .. (7)
[0066]	<223> a、c、t、g、未知或其它
[0067]	<400> 6
[0068]	acnnnnnctc c 11
[0069]	<210> 7
[0070]	<211> 30
[0071]	<212> DNA
[0072]	<213> 人工序列
[0073]	<220>
[0074]	<223> 人工序列的描述:合成寡核苷酸
[0075]	<400> 7
[0076]	atctatgtcg ggtgcggaga aagaggtaat 30
[0077]	<210> 8

[0078]	<211> 12
[0079]	<212> DNA
[0080]	<213> 人工序列
[0081]	<220>
[0082]	<223> 人工序列的描述:合成寡核苷酸
[0083]	<220>
[0084]	<221> 修饰的碱基
[0085]	<222> (4) .. (8)
[0086]	<223> a、c、t、g、未知或其它
[0087]	<400> 8
[0088]	caannnnngt gg 12
[0089]	<210> 9
[0090]	<211> 10
[0091]	<212> DNA
[0092]	<213> 人工序列
[0093]	<220>
[0094]	<223> 人工序列的描述:合成寡核苷酸
[0095]	<220>
[0096]	<221> 修饰的碱基
[0097]	<222> (4) .. (7)
[0098]	<223> a、c、t、g、未知或其它
[0099]	<400> 9
[0100]	cacnnnnngtg 10
[0101]	<210> 10
[0102]	<211> 10
[0103]	<212> DNA
[0104]	<213> 人工序列
[0105]	<220>
[0106]	<223> 人工序列的描述:合成寡核苷酸
[0107]	<220>
[0108]	<221> 修饰的碱基
[0109]	<222> (4) .. (7)
[0110]	<223> a、c、t、g、未知或其它
[0111]	<400> 10
[0112]	caynnnnrntg 10
[0113]	<210> 11
[0114]	<211> 15
[0115]	<212> DNA
[0116]	<213> 人工序列

[0117]	<220>
[0118]	<223> 人工序列的描述:合成寡核苷酸
[0119]	<220>
[0120]	<221> 修饰的碱基
[0121]	<222> (4) .. (12)
[0122]	<223> a、c、t、g、未知或其它
[0123]	<400> 11
[0124]	ccannnnnnnn nntgg 15
[0125]	<210> 12
[0126]	<211> 12
[0127]	<212> DNA
[0128]	<213> 人工序列
[0129]	<220>
[0130]	<223> 人工序列的描述:合成寡核苷酸
[0131]	<220>
[0132]	<221> 修饰的碱基
[0133]	<222> (4) .. (9)
[0134]	<223> a、c、t、g、未知或其它
[0135]	<400> 12
[0136]	ccannnnnnt gg 12
[0137]	<210> 13
[0138]	<211> 11
[0139]	<212> DNA
[0140]	<213> 人工序列
[0141]	<220>
[0142]	<223> 人工序列的描述:合成寡核苷酸
[0143]	<220>
[0144]	<221> 修饰的碱基
[0145]	<222> (4) .. (8)
[0146]	<223> a、c、t、g、未知或其它
[0147]	<400> 13
[0148]	ccannnnntg g 11
[0149]	<210> 14
[0150]	<211> 11
[0151]	<212> DNA
[0152]	<213> 人工序列
[0153]	<220>
[0154]	<223> 人工序列的描述:合成寡核苷酸
[0155]	<220>

- [0156] <221> 修饰的碱基
[0157] <222> (3) .. (9)
[0158] <223> a、c、t、g、未知或其它
[0159] <400> 14
[0160] ccnnnnnnng g 11
[0161] <210> 15
[0162] <211> 11
[0163] <212> DNA
[0164] <213> 人工序列
[0165] <220>
[0166] <223> 人工序列的描述:合成寡核苷酸
[0167] <220>
[0168] <221> 修饰的碱基
[0169] <222> (4) .. (8)
[0170] <223> a、c、t、g、未知或其它
[0171] <400> 15
[0172] cctnnnnnag g 11
[0173] <210> 16
[0174] <211> 12
[0175] <212> DNA
[0176] <213> 人工序列
[0177] <220>
[0178] <223> 人工序列的描述:合成寡核苷酸
[0179] <220>
[0180] <221> 修饰的碱基
[0181] <222> (4) .. (9)
[0182] <223> a、c、t、g、未知或其它
[0183] <400> 16
[0184] cgannnnnnt gc 12
[0185] <210> 17
[0186] <211> 10
[0187] <212> DNA
[0188] <213> 人工序列
[0189] <220>
[0190] <223> 人工序列的描述:合成寡核苷酸
[0191] <220>
[0192] <221> 修饰的碱基
[0193] <222> (4) .. (7)
[0194] <223> a、c、t、g、未知或其它

[0195]	<400> 17
[0196]	gaannnnnttc 10
[0197]	<210> 18
[0198]	<211> 10
[0199]	<212> DNA
[0200]	<213> 人工序列
[0201]	<220>
[0202]	<223> 人工序列的描述:合成寡核苷酸
[0203]	<220>
[0204]	<221> 修饰的碱基
[0205]	<222> (4) .. (7)
[0206]	<223> a、c、t、g、未知或其它
[0207]	<400> 18
[0208]	gacnnnnngtc 10
[0209]	<210> 19
[0210]	<211> 11
[0211]	<212> DNA
[0212]	<213> 人工序列
[0213]	<220>
[0214]	<223> 人工序列的描述:合成寡核苷酸
[0215]	<220>
[0216]	<221> 修饰的碱基
[0217]	<222> (4) .. (8)
[0218]	<223> a、c、t、g、未知或其它
[0219]	<400> 19
[0220]	gacnnnnngt c 11
[0221]	<210> 20
[0222]	<211> 12
[0223]	<212> DNA
[0224]	<213> 人工序列
[0225]	<220>
[0226]	<223> 人工序列的描述:合成寡核苷酸
[0227]	<220>
[0228]	<221> 修饰的碱基
[0229]	<222> (4) .. (9)
[0230]	<223> a、c、t、g、未知或其它
[0231]	<400> 20
[0232]	gacnnnnnng tc 12
[0233]	<210> 21

[0234] <211> 10
[0235] <212> DNA
[0236] <213> 人工序列
[0237] <220>
[0238] <223> 人工序列的描述:合成寡核苷酸
[0239] <220>
[0240] <221> 修饰的碱基
[0241] <222> (4) .. (7)
[0242] <223> a、c、t、g、未知或其它
[0243] <400> 21
[0244] gatnnnnatc 10
[0245] <210> 22
[0246] <211> 11
[0247] <212> DNA
[0248] <213> 人工序列
[0249] <220>
[0250] <223> 人工序列的描述:合成寡核苷酸
[0251] <220>
[0252] <221> 修饰的碱基
[0253] <222> (4) .. (8)
[0254] <223> a、c、t、g、未知或其它
[0255] <400> 22
[0256] gcannnnntg c 11
[0257] <210> 23
[0258] <211> 11
[0259] <212> DNA
[0260] <213> 人工序列
[0261] <220>
[0262] <223> 人工序列的描述:合成寡核苷酸
[0263] <220>
[0264] <221> 修饰的碱基
[0265] <222> (4) .. (8)
[0266] <223> a、c、t、g、未知或其它
[0267] <400> 23
[0268] gccnnnnngg c 11
[0269] <210> 24
[0270] <211> 11
[0271] <212> DNA
[0272] <213> 人工序列

- [0273] <220>
[0274] <223> 人工序列的描述:合成寡核苷酸
[0275] <220>
[0276] <221> 修饰的碱基
[0277] <222> (3) .. (9)
[0278] <223> a、c、t、g、未知或其它
[0279] <400> 24
[0280] gcnnnnnnng c 11
[0281] <210> 25
[0282] <211> 13
[0283] <212> DNA
[0284] <213> 人工序列
[0285] <220>
[0286] <223> 人工序列的描述:合成寡核苷酸
[0287] <220>
[0288] <221> 修饰的碱基
[0289] <222> (5) .. (9)
[0290] <223> a、c、t、g、未知或其它
[0291] <400> 25
[0292] ggccnnnnng gcc 13
[0293] <210> 26
[0294] <211> 27
[0295] <212> DNA
[0296] <213> 人工序列
[0297] <220>
[0298] <223> 人工序列的描述:合成寡核苷酸
[0299] <400> 26
[0300] taactataac ggtcctaagg tagcgaa 27
[0301] <210> 27
[0302] <211> 18
[0303] <212> DNA
[0304] <213> 人工序列
[0305] <220>
[0306] <223> 人工序列的描述:合成寡核苷酸
[0307] <400> 27
[0308] tagggataac agggtaat 18
[0309] <210> 28
[0310] <211> 30
[0311] <212> DNA

-
- [0312] <213> 人工序列
- [0313] <220>
- [0314] <223> 人工序列的描述:合成寡核苷酸
- [0315] <400> 28
- [0316] tggcaaacag ctattatggg tattatgggt 30

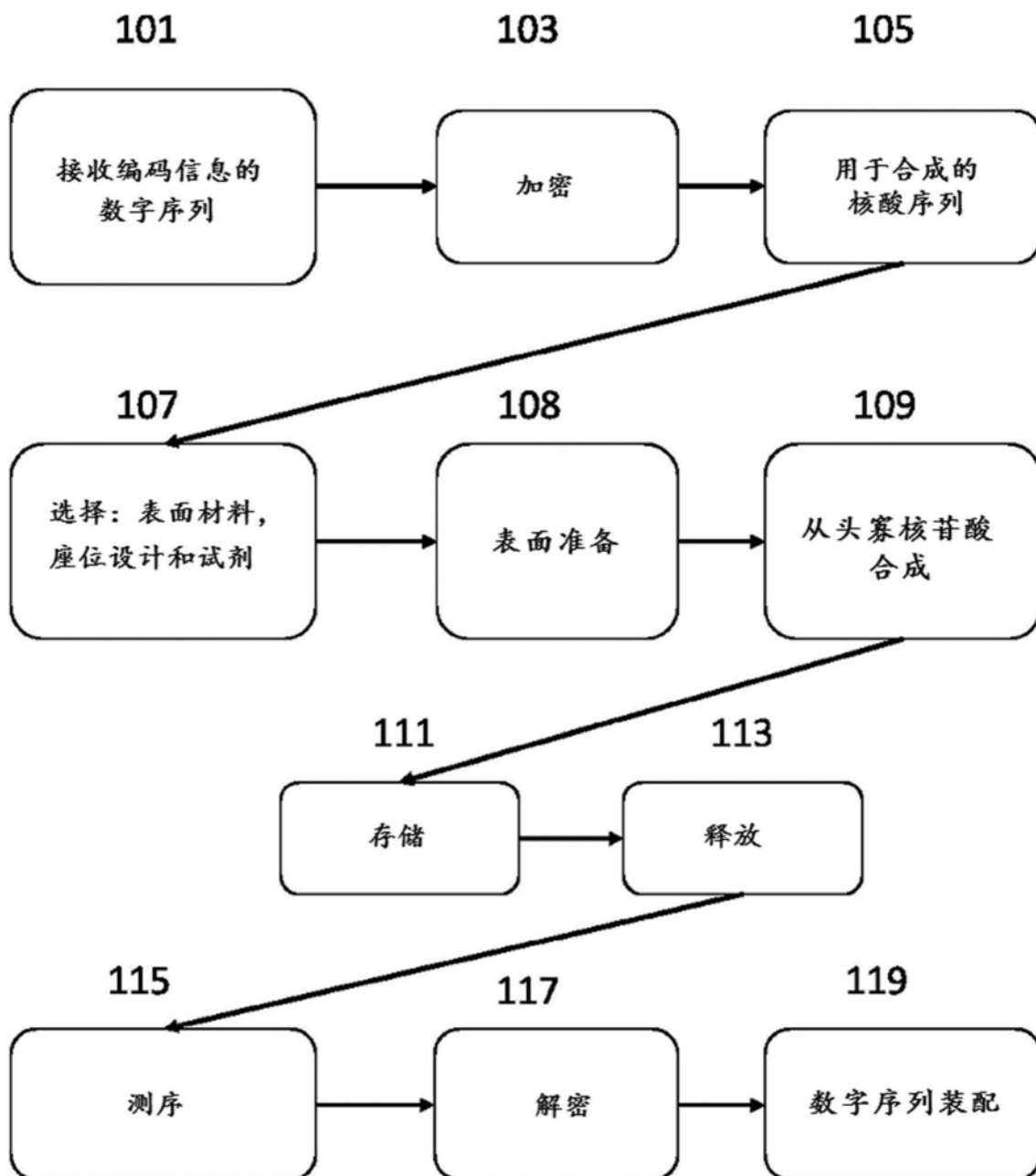


图1

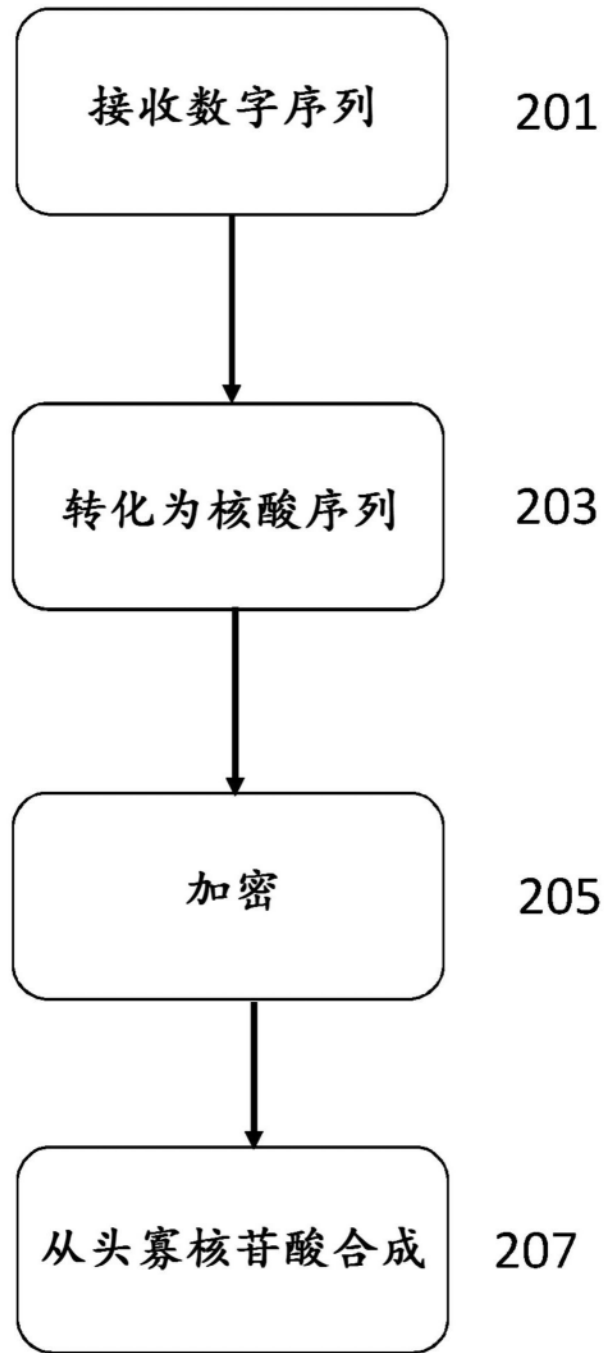


图2

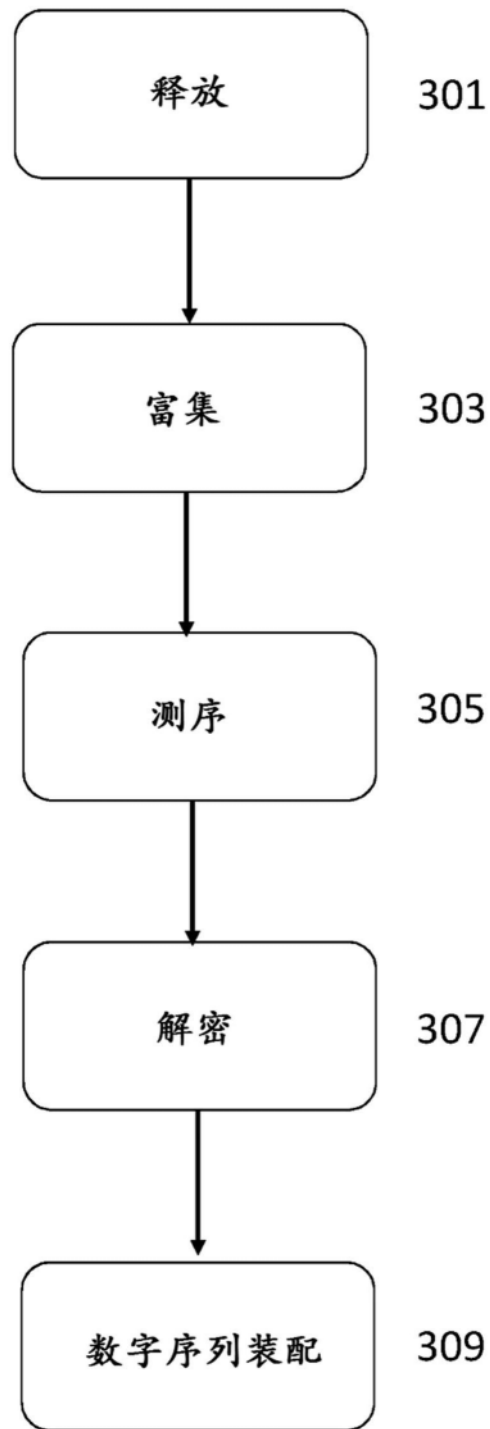


图3

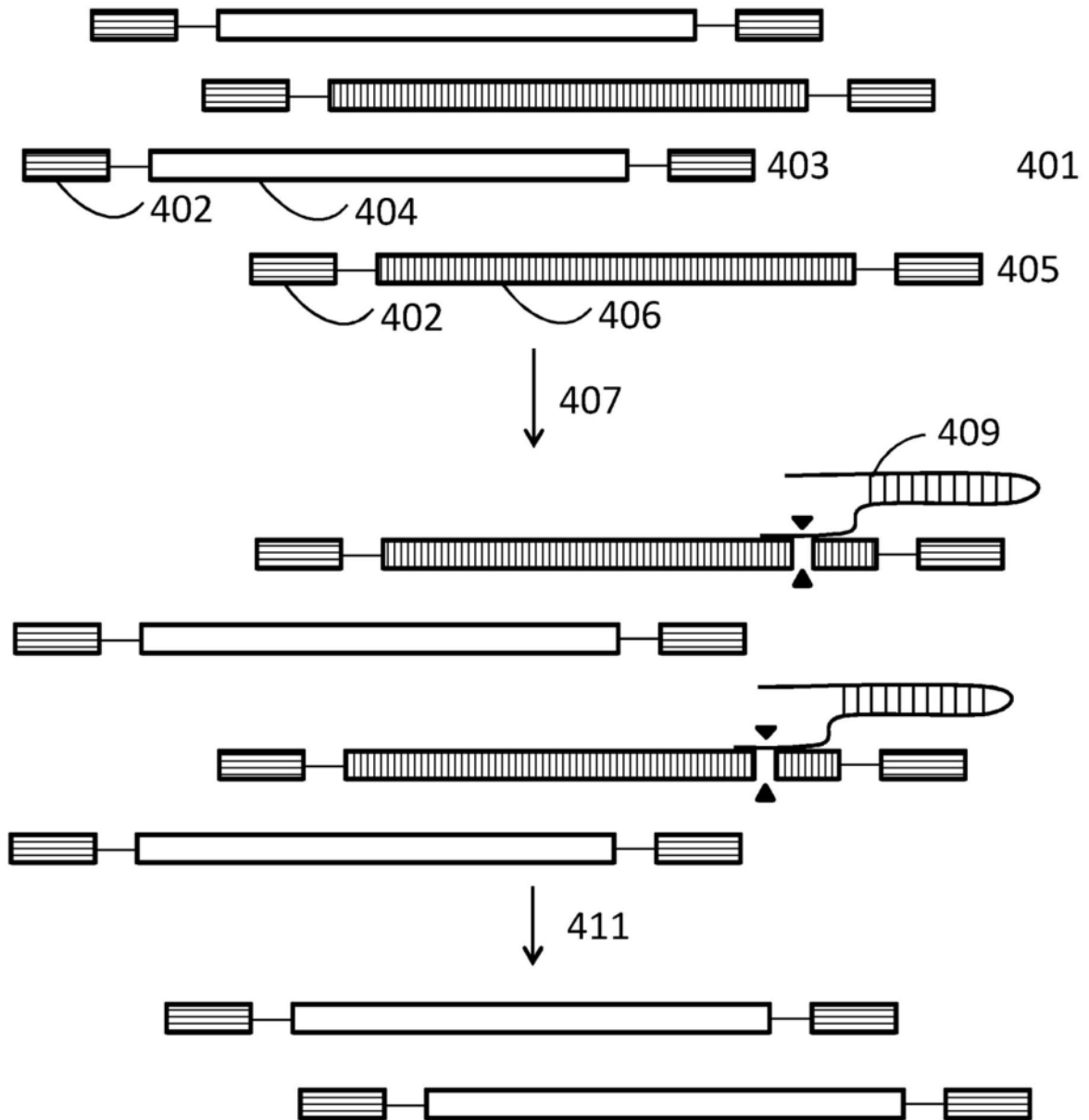


图4A

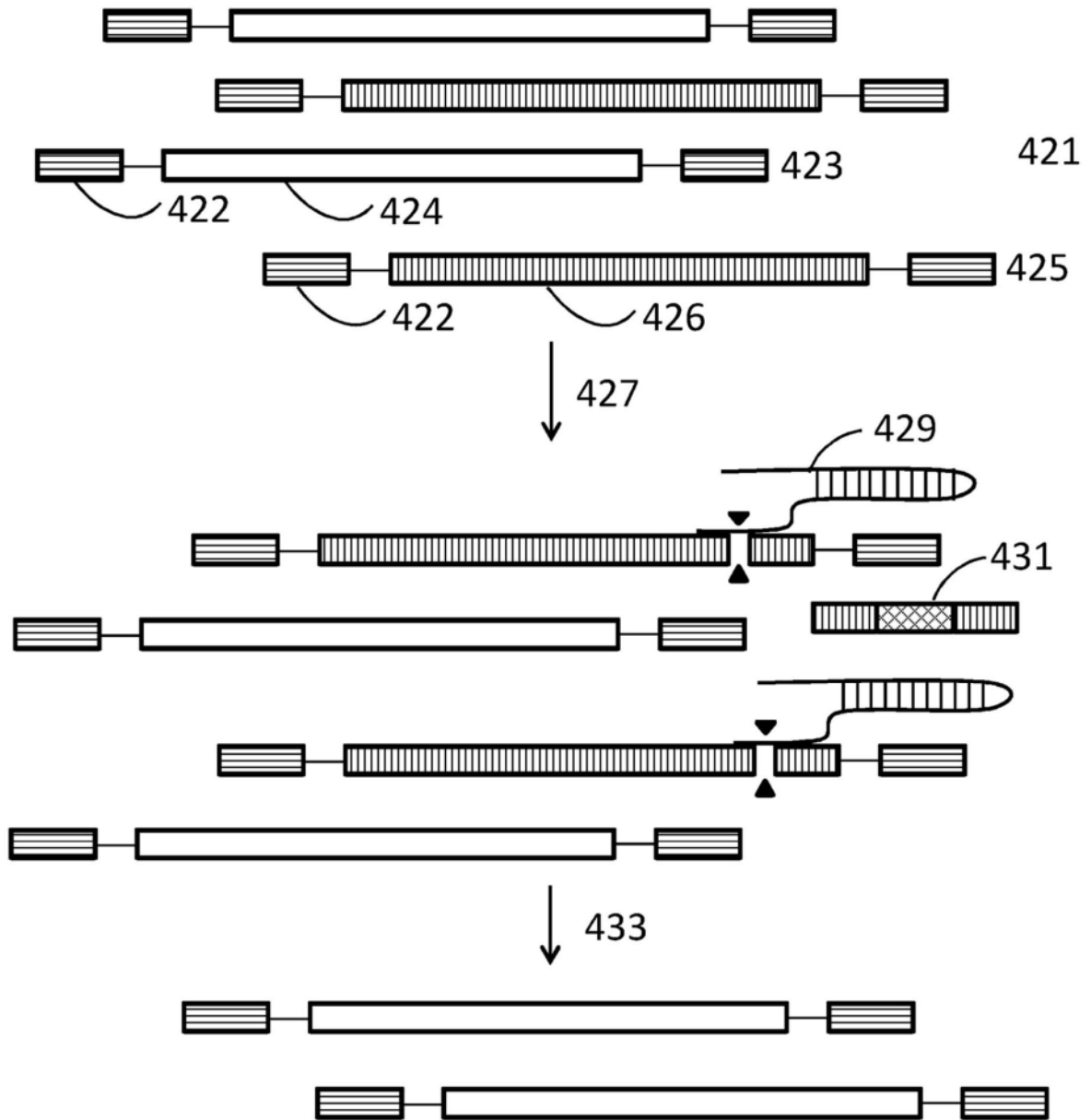


图4B

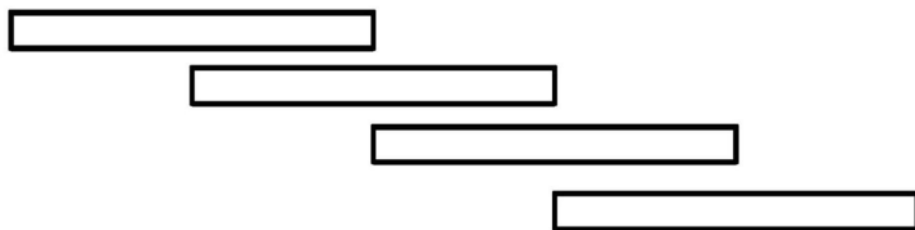


图5A



图5B

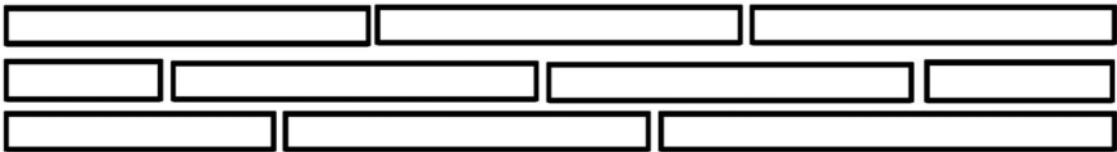


图5C



图6A

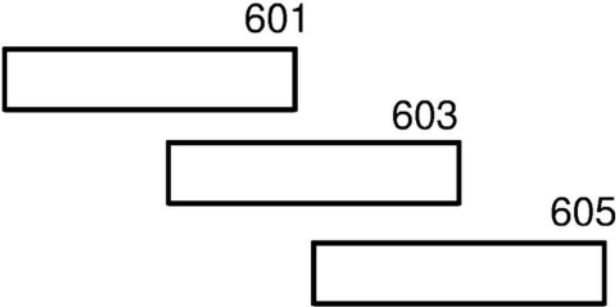


图6B

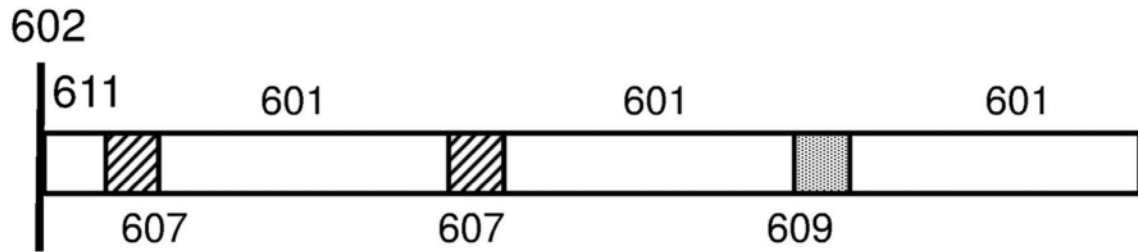


图6C

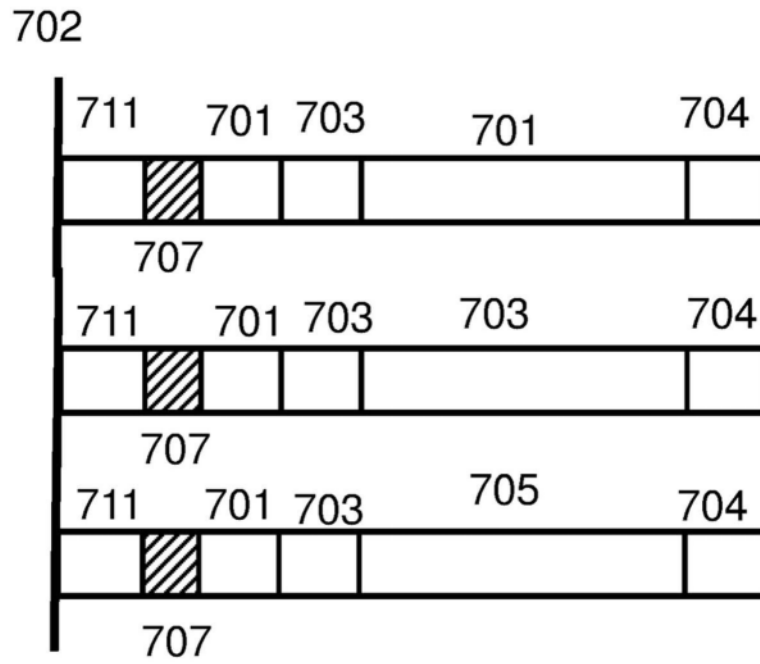


图7A

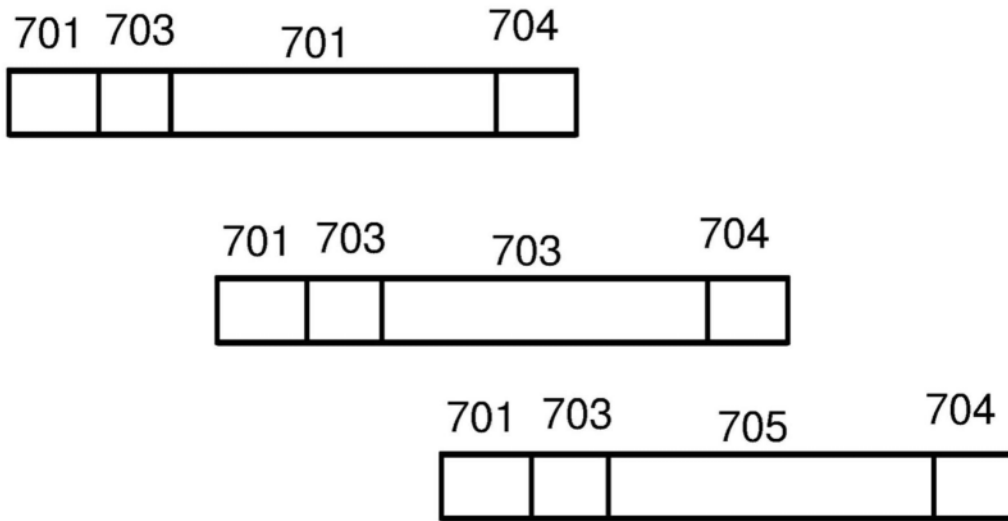


图7B

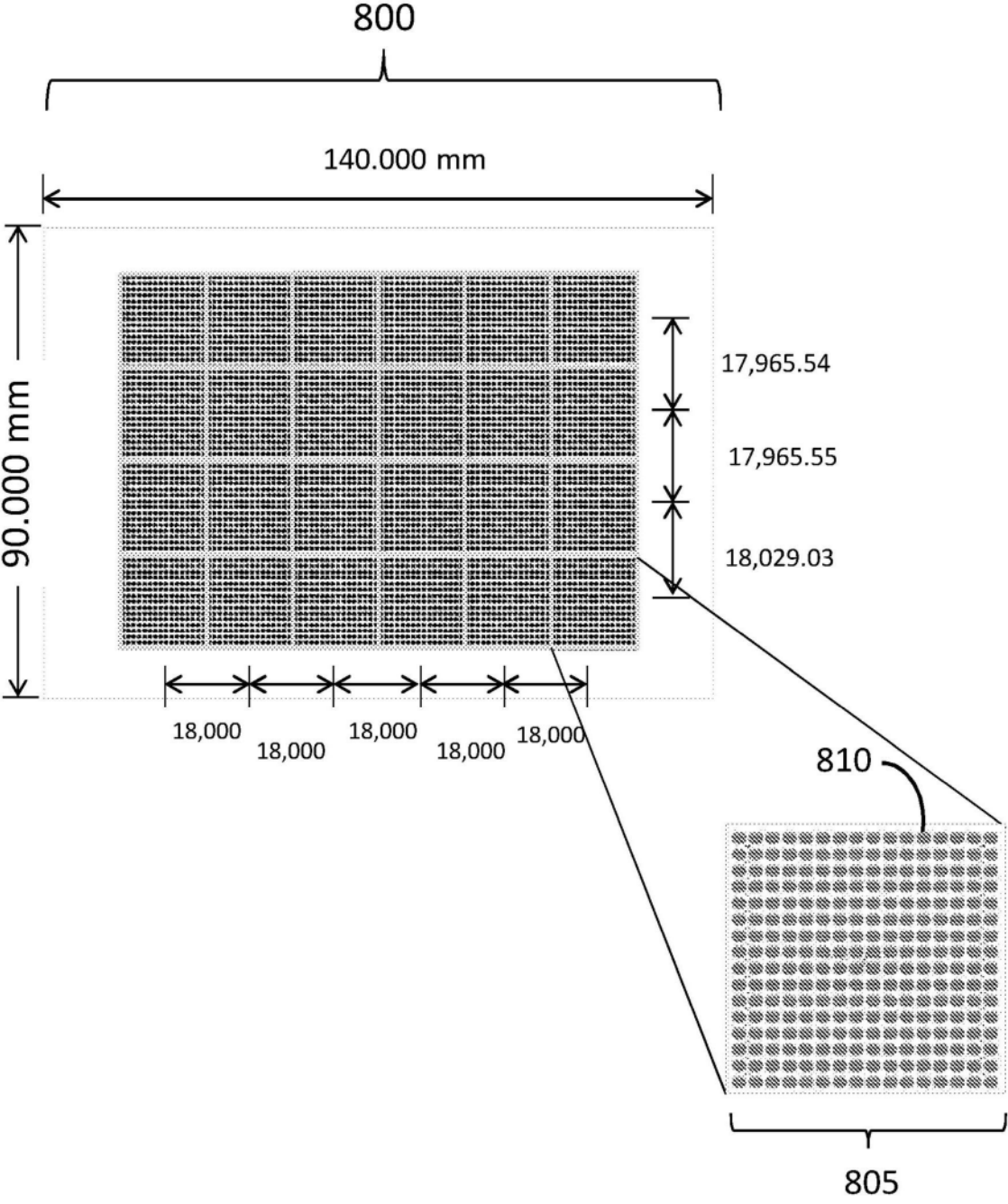


图8

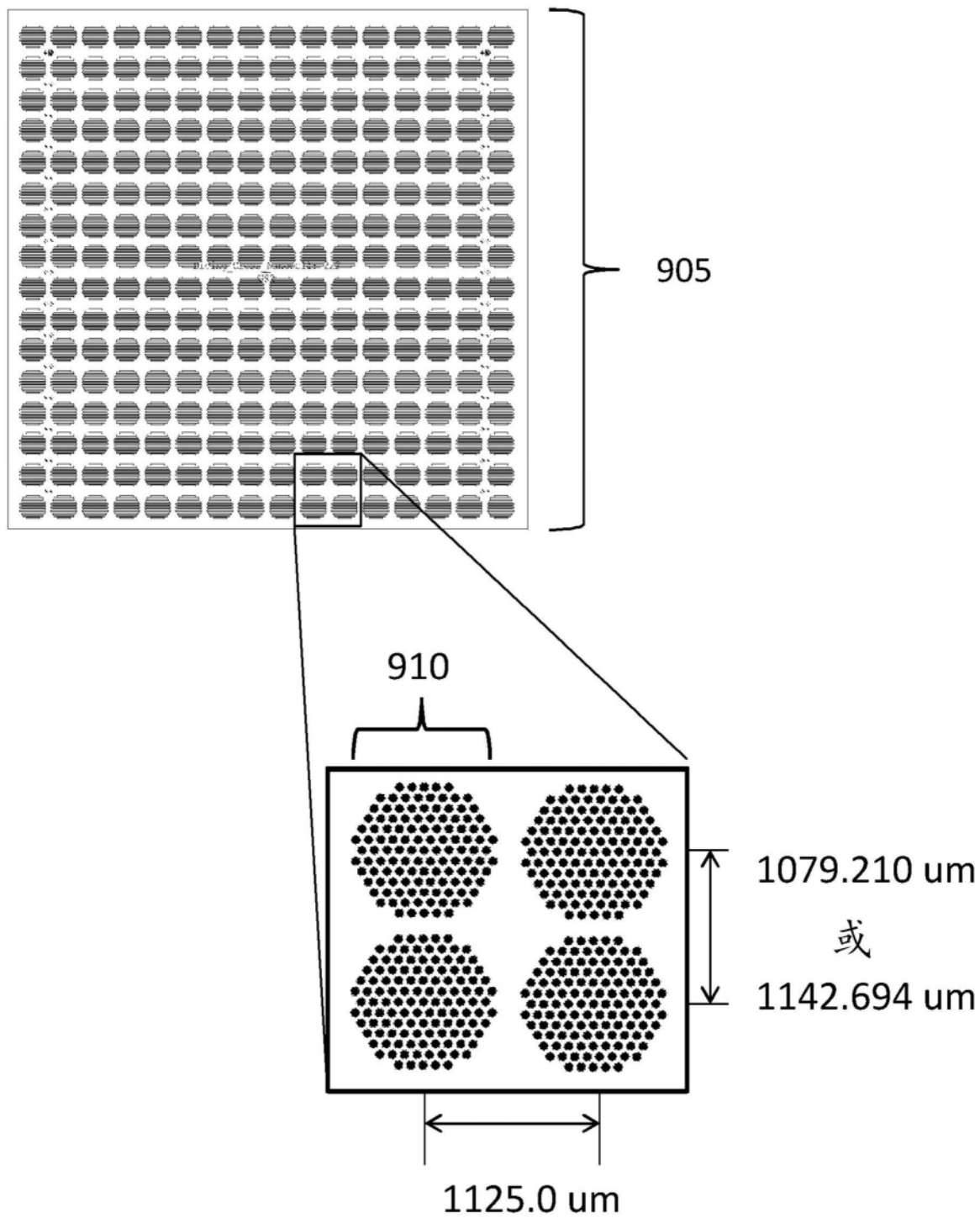


图9

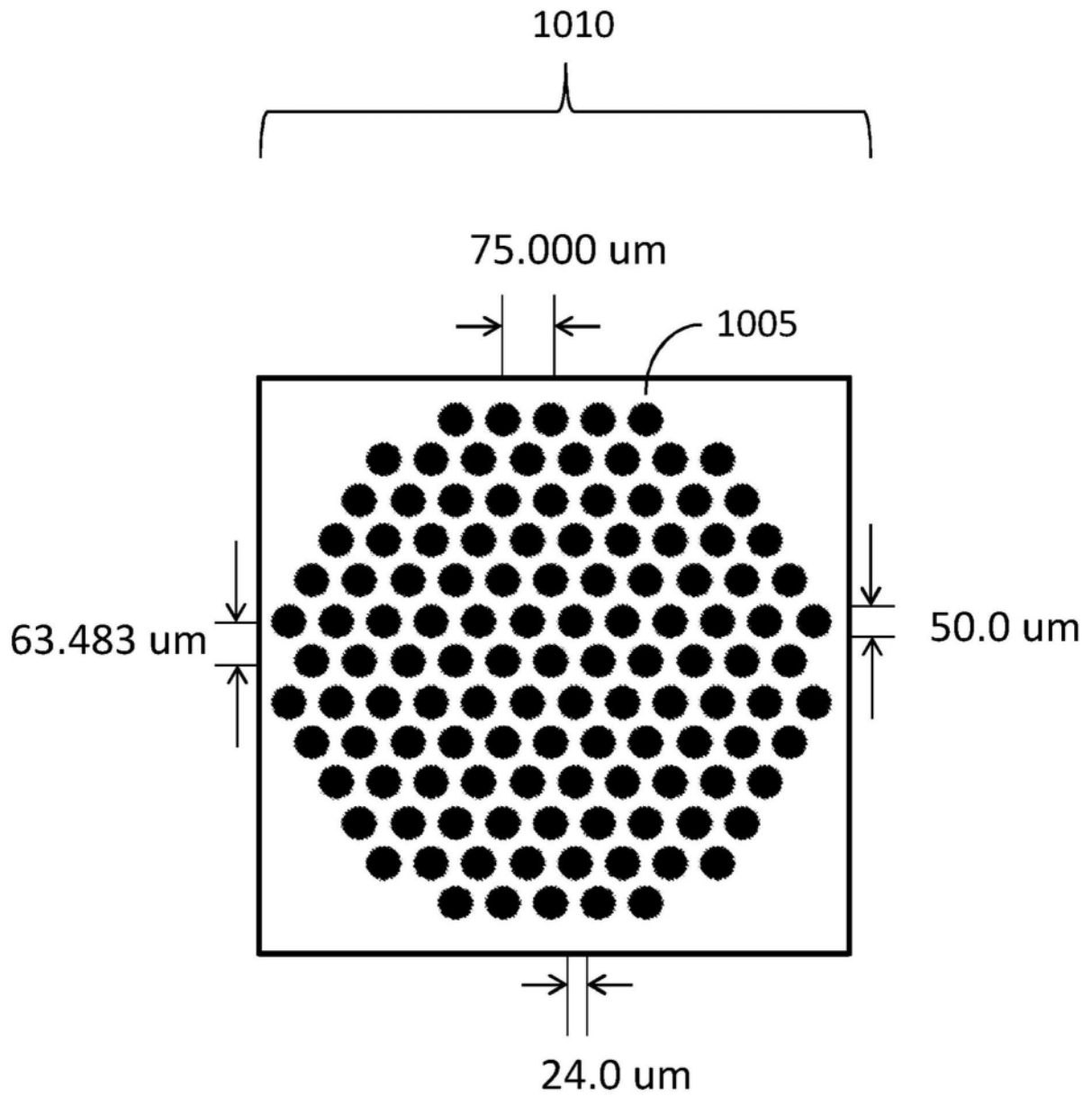


图10

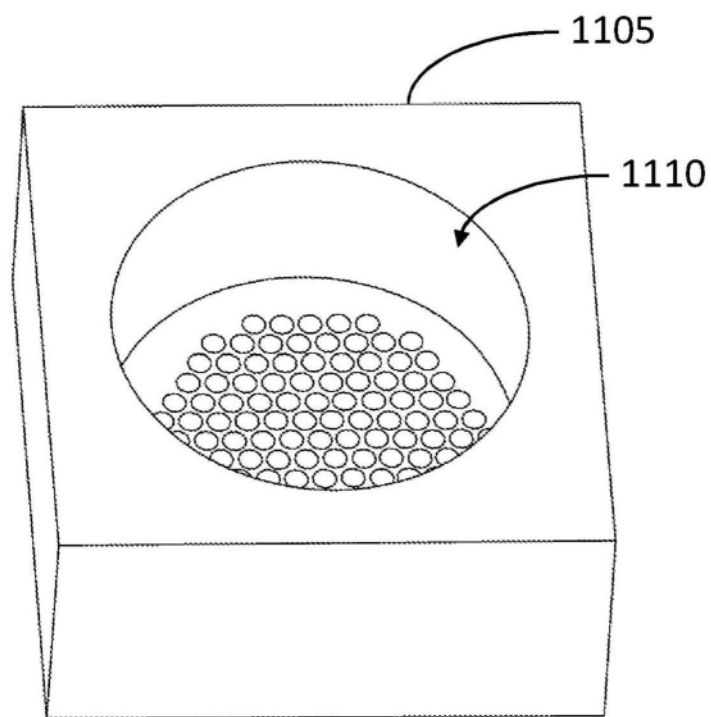


图11A

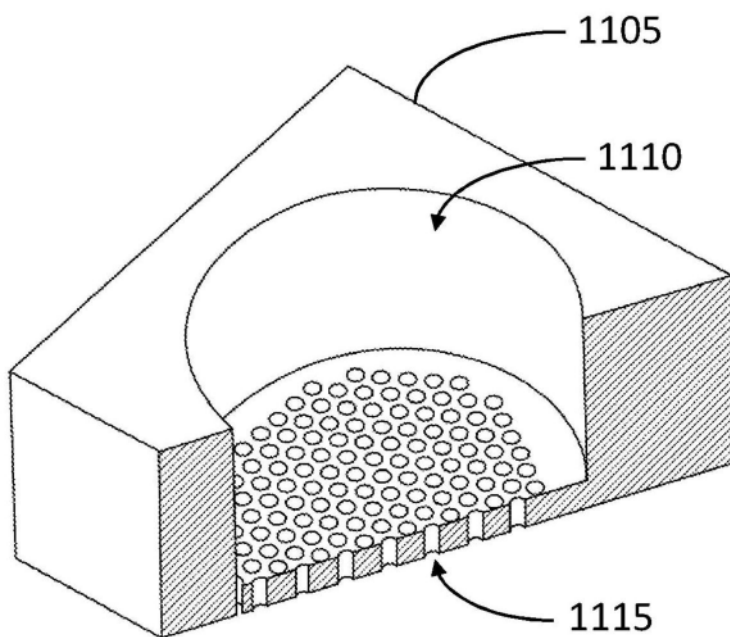


图11B

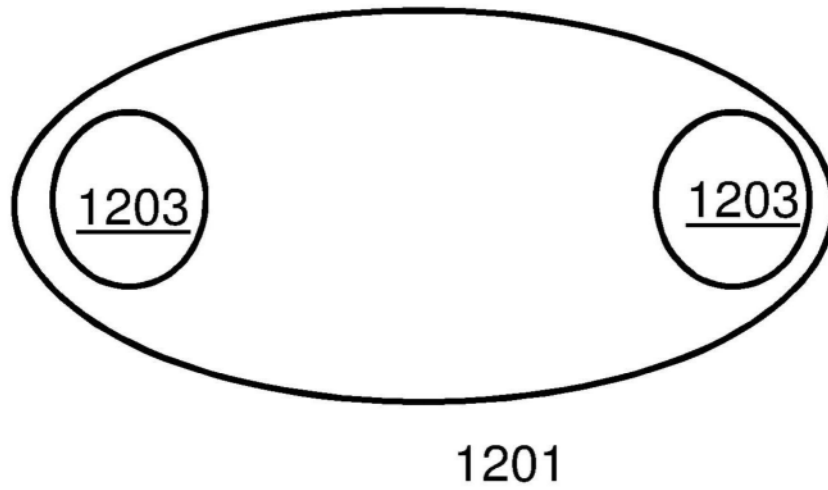


图12A

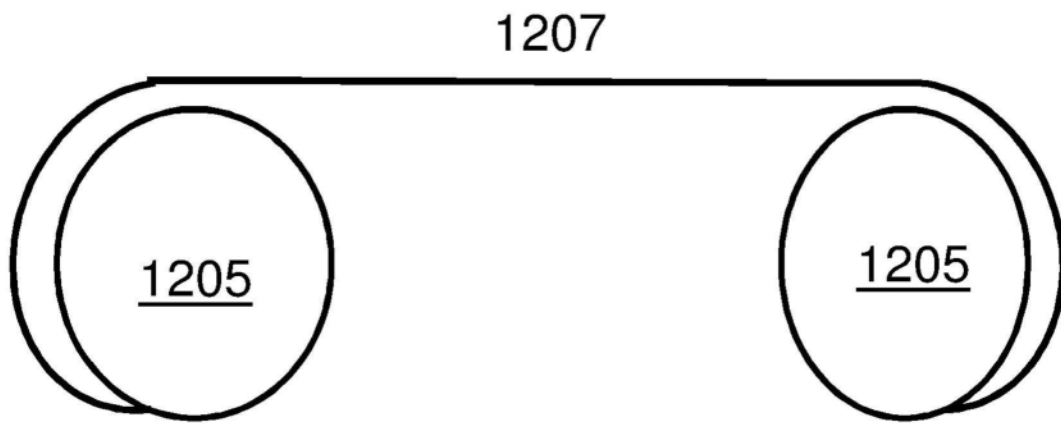


图12B

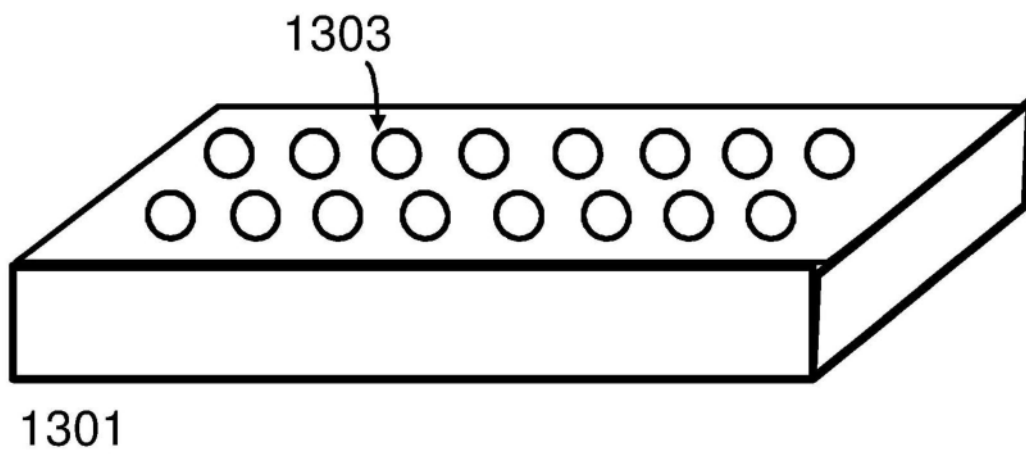


图13A

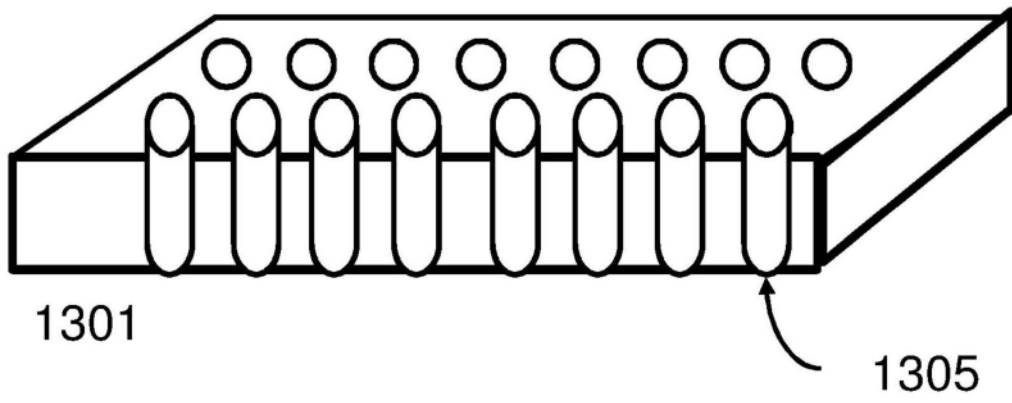


图13B

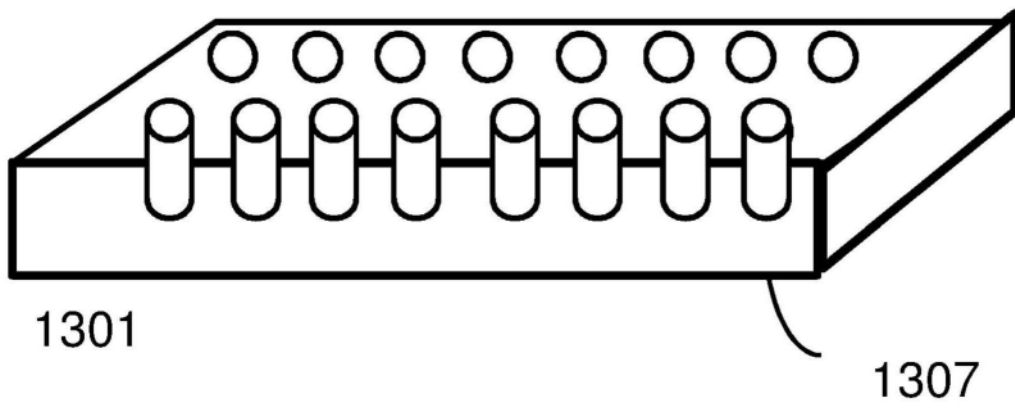


图13C

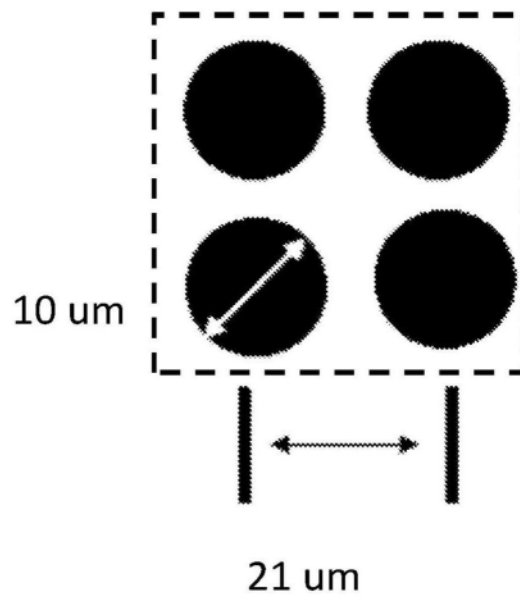


图14A

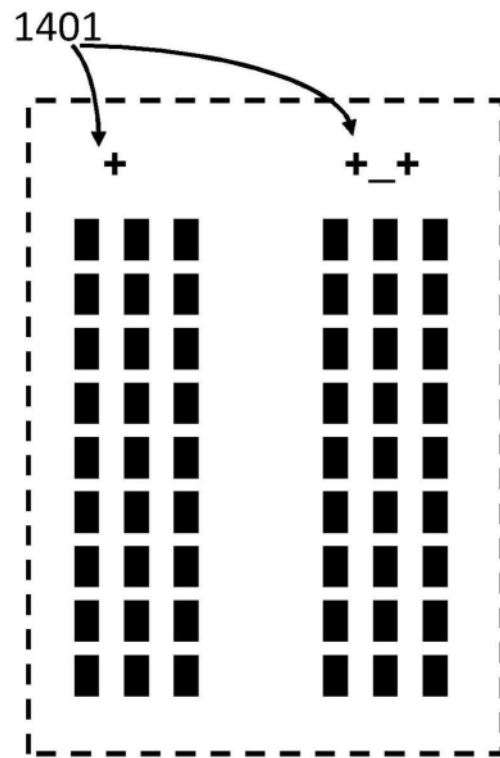


图14B

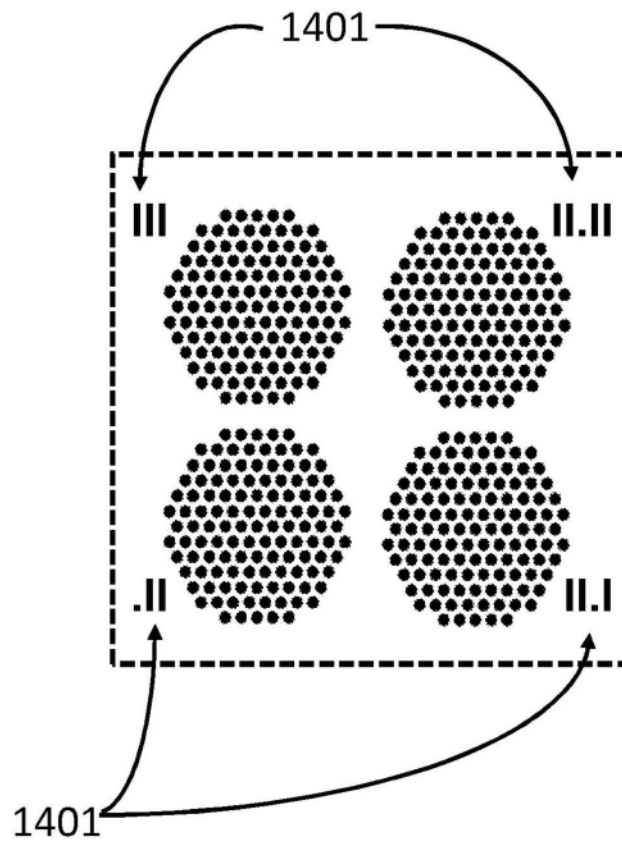


图14C

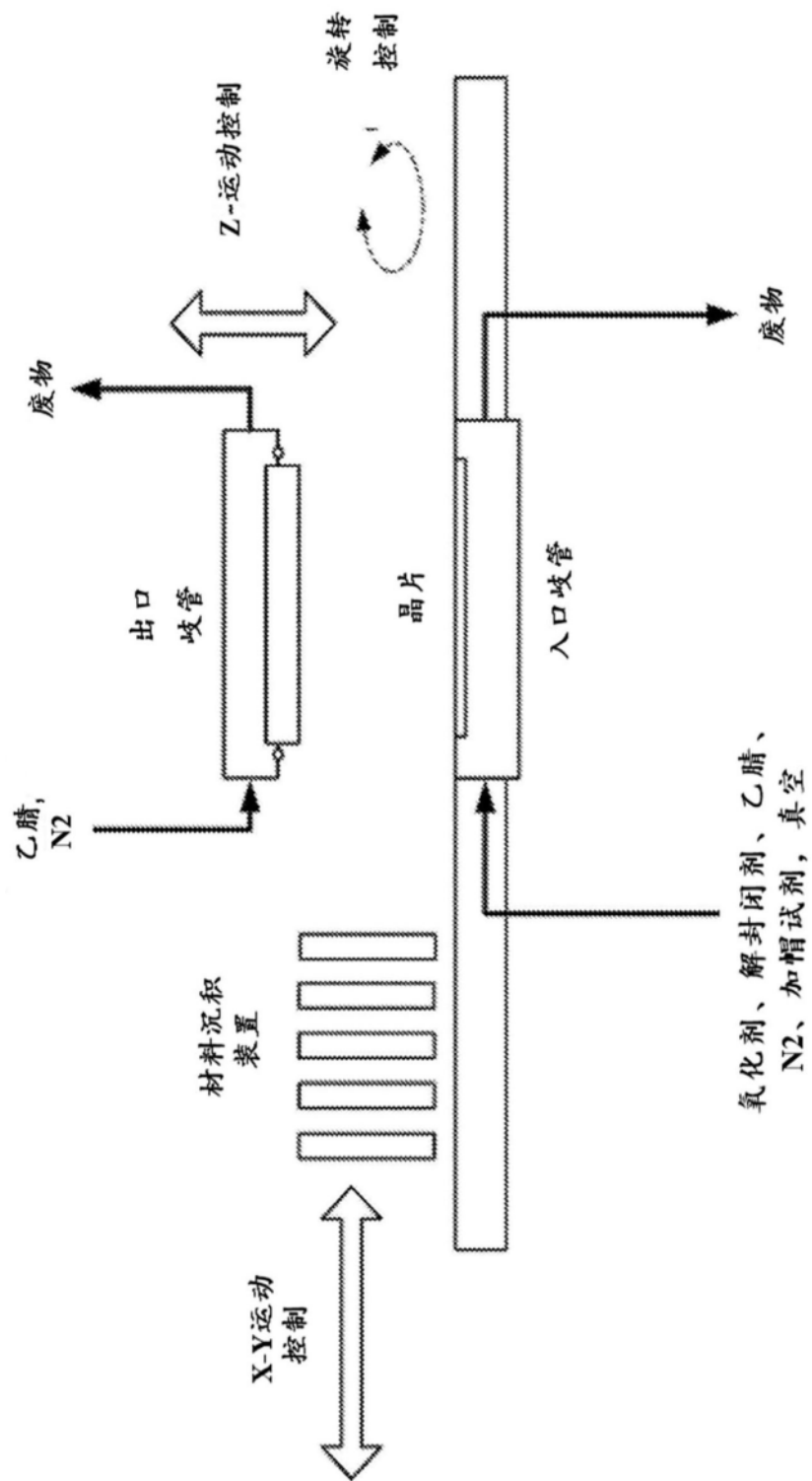


图15

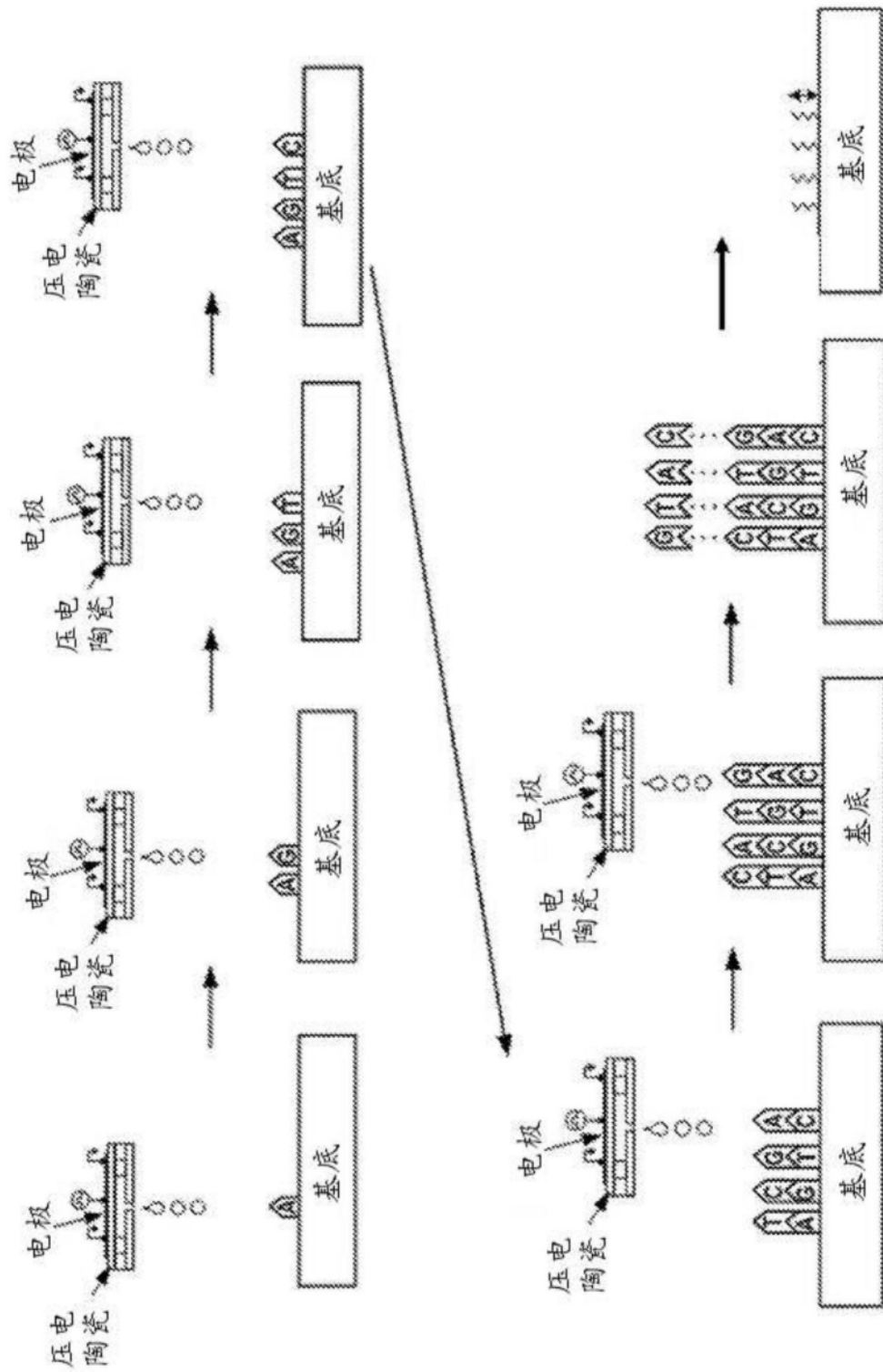


图16

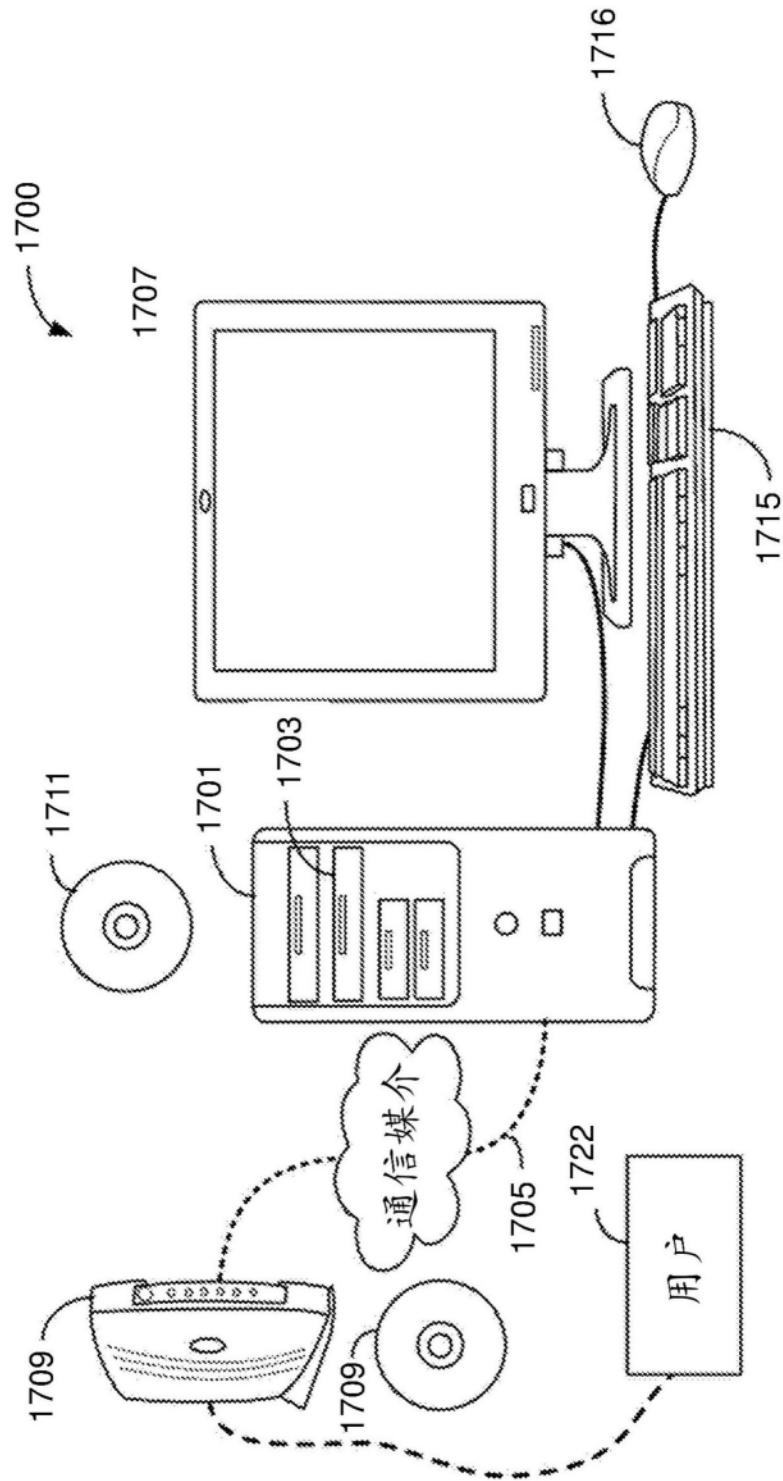


图17

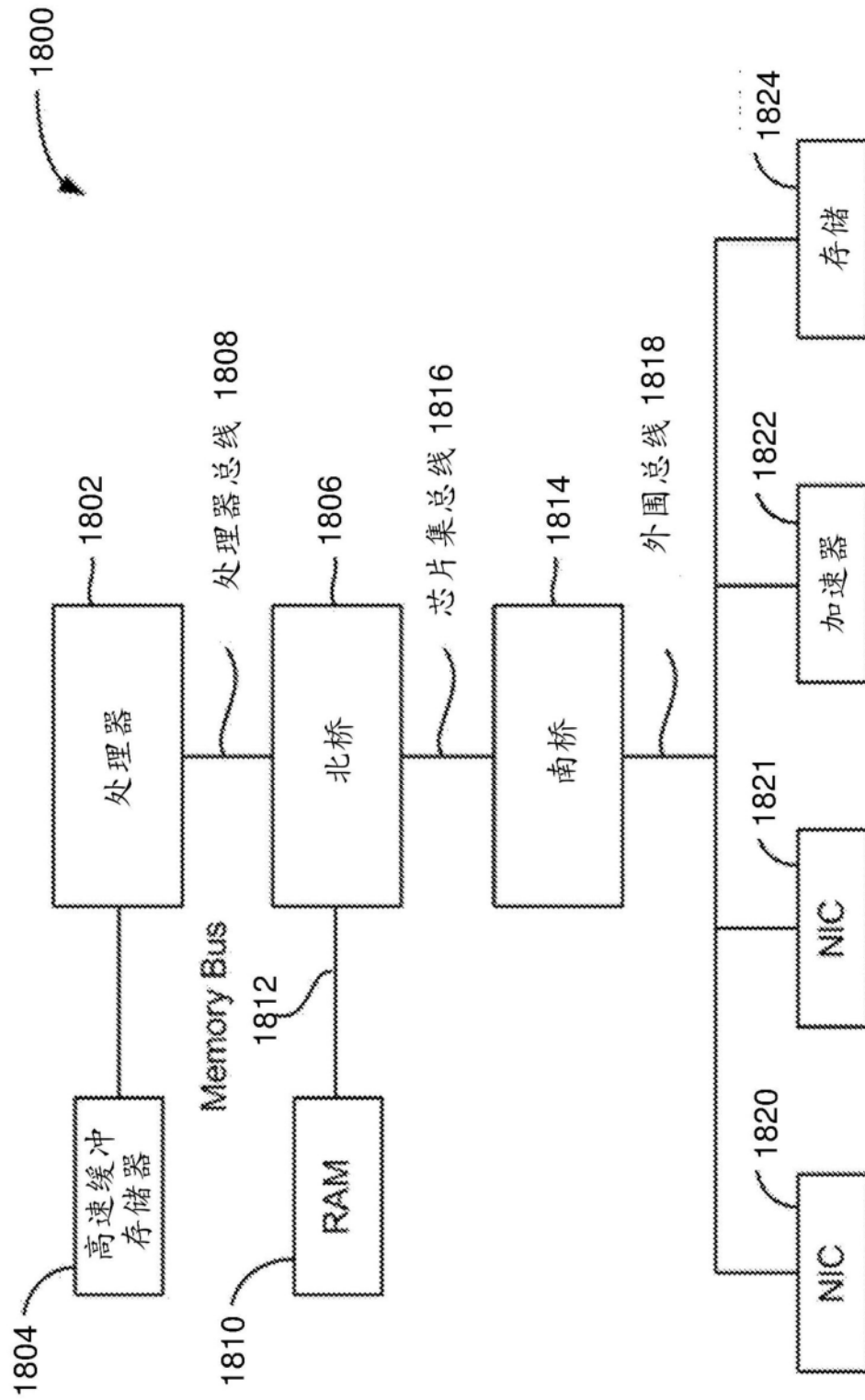


图18

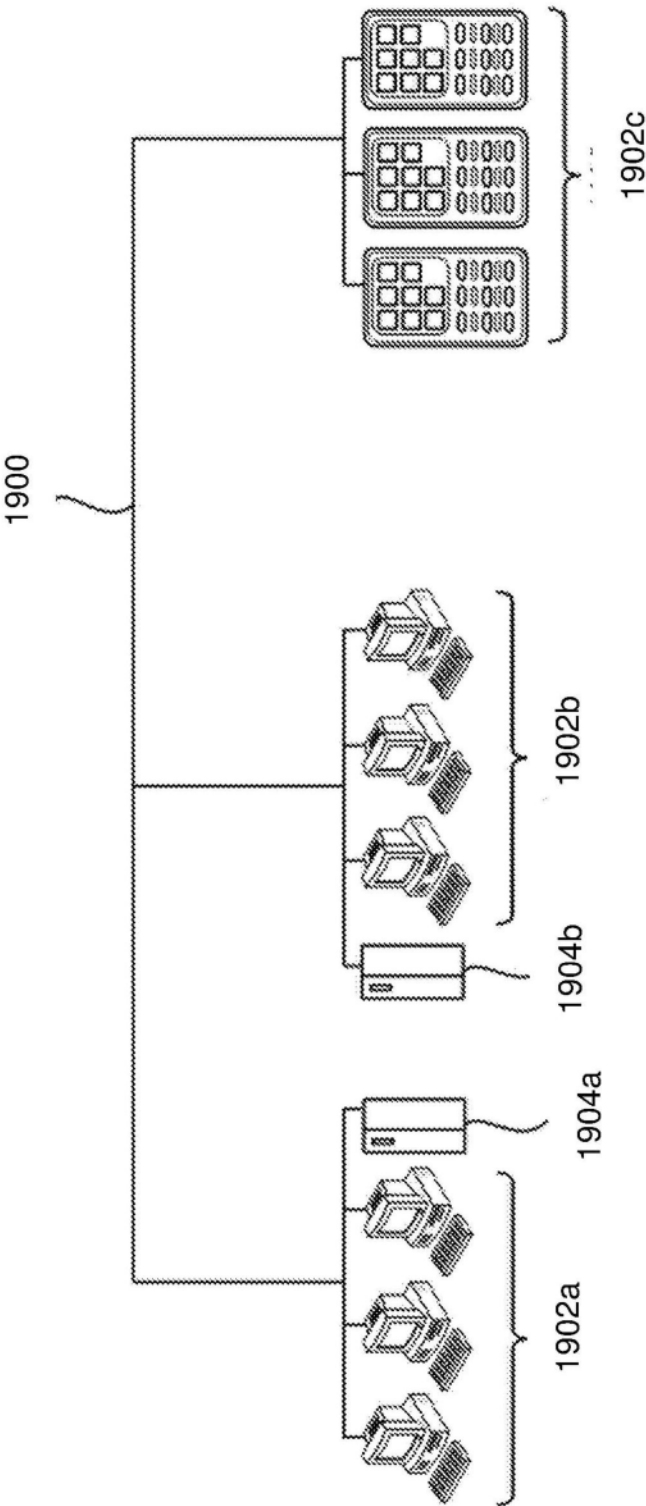


图19

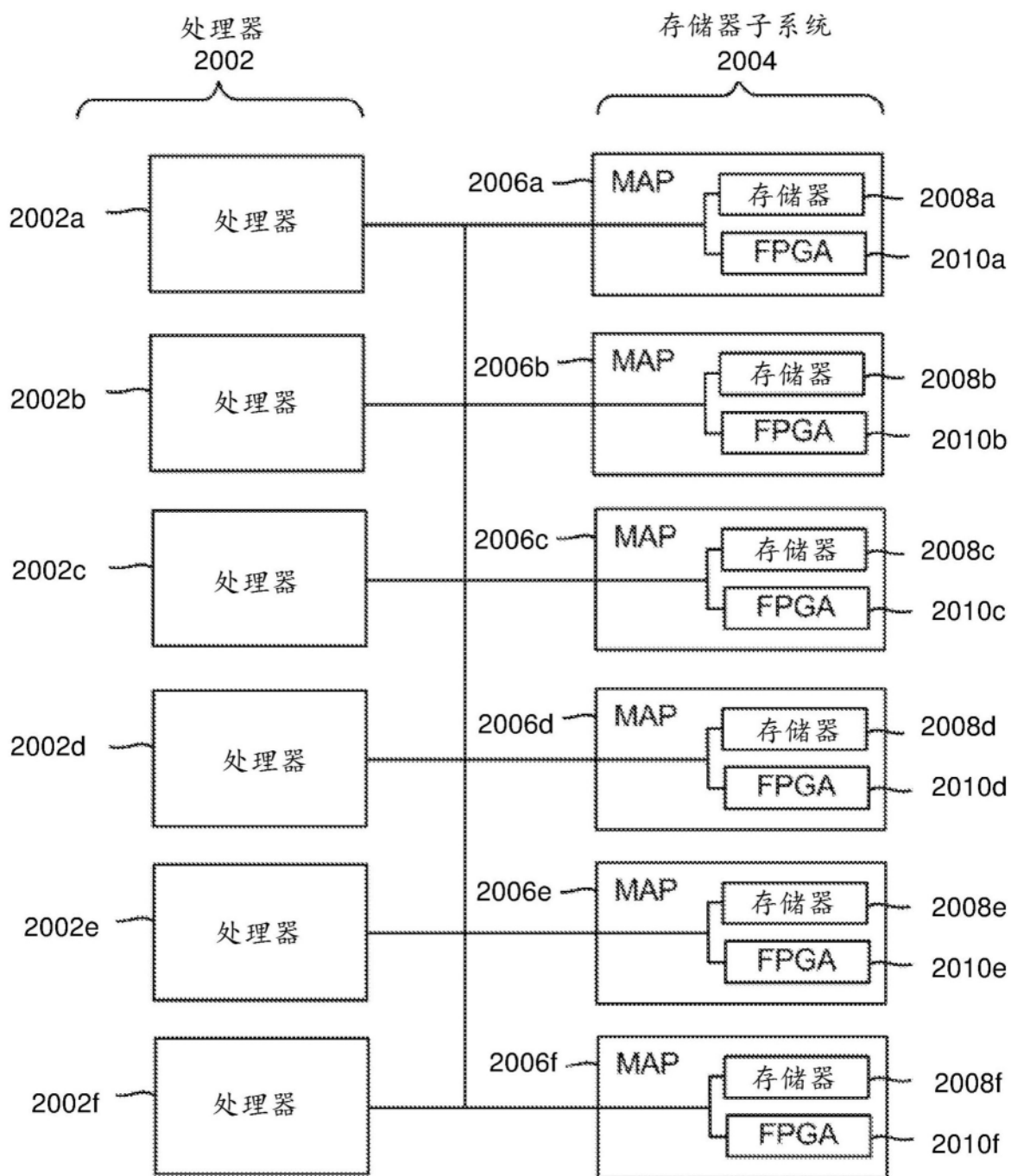


图20