



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년10월28일

(11) 등록번호 10-2723464

(24) 등록일자 2024년10월24일

(51) 국제특허분류(Int. Cl.)
G16B 50/40 (2019.01) *G06F 21/62* (2013.01)
G16B 50/30 (2019.01) *H04L 9/32* (2006.01)

(52) CPC특허분류
G16B 50/40 (2019.02)
G06F 21/6218 (2013.01)

(21) 출원번호 10-2019-7027305

(22) 출원일자(국제) 2018년02월22일

심사청구일자 2021년02월22일

(85) 번역문제출일자 2019년09월18일

(65) 공개번호 10-2019-0119107

(43) 공개일자 2019년10월21일

(86) 국제출원번호 PCT/US2018/019268

(87) 국제공개번호 WO 2018/156792

국제공개일자 2018년08월30일

(30) 우선권주장

62/462,284 2017년02월22일 미국(US)

(56) 선행기술조사문헌

J. Bornholt 외, "A DNA-Based Archival Storage System", ASPLOS '16, pp. 637-649. (2016.05.25.)*

S. M. Hosseine Tabatabaei Yazdi 외, "DNA-Based Storage: Trends and Methods", IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 1(3):230-248. (2015.09.)*

T. D. P. Brunet 외, "Aims and methods of biosteganography", Journal of Biotechnology, 226:56-64. (2016.03.25.)*

US20150032385 A1

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

트위스트 바이오사이언스 코퍼레이션

미국 94080 캘리포니아주 사우스 샌 프란시스코
게이트웨이 불러바드 681

(72) 발명자

팩 빌 제임스

미국 95051 캘리포니아주 산타 클라라 칼턴 플레
이스 3086

자이툰 램지 이브라힘

미국 94107 캘리포니아주 샌프란시스코 식스틴스
스트리트 1010 아파트먼트 608

(74) 대리인

김진희, 김태홍

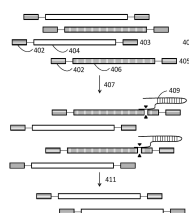
전체 청구항 수 : 총 10 항

심사관 : 김채연

(54) 발명의 명칭 **핵산 기반 데이터 저장****(57) 요약**

본원에는 저장을 위한 보증된 생체분자-기반 정보의 조성물, 디바이스, 시스템 및 생성 및 이용 방법이 제공된다. 또한, 본원에는 정보의 바이오암호화 또는 바이오복호화를 위한 조성물, 디바이스, 시스템 및 방법이 제공된다. 디지털 서열의 핵산 기반 서열로의 전환은 하나 이상의 바이오암호화 방법을 선택하는 단계를 포함한

(뒷면에 계속)

대표도 - 도4a

다.

(52) CPC특허분류

G16B 50/30 (2019.02)

H04L 9/3231 (2013.01)

명세서

청구범위

청구항 1

- a) 컴퓨터 시스템을 통하여, 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템을 수신하는 단계;
 - b) 컴퓨터 시스템을 통하여, 효소적 바이오암호화를 위한 명령을 수신하는 단계;
 - c) 컴퓨터 시스템을 통하여, 명령에 기반하여 적어도 하나의 디지털 서열을 복수의 올리고뉴클레오타이드의 디지털 서열로 전환하는 단계로서, 복수의 올리고뉴클레오타이드의 디지털 서열은 뉴클레아제 복합체에 대한 복수의 올리고뉴클레오타이드의 노출에 의한 복호화를 위해 구성되는 단계;
 - d) 합성기에서 복수의 올리고뉴클레오타이드를 합성하는 단계;
 - e) 복수의 올리고뉴클레오타이드를 저장하는 단계;
 - f) 효소적 복호화를 복수의 올리고뉴클레오타이드에 적용하는 단계;
 - g) 복수의 올리고뉴클레오타이드를 농축하는 단계;
 - h) 복수의 올리고뉴클레오타이드로부터의 농축된 올리고뉴클레오타이드를 시퀀싱하여 핵산 서열을 생성하는 단계; 및
 - i) 핵산 서열을 적어도 하나의 디지털 서열로 전환하는 단계로서, 상기 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이템을 코딩하는 것인 단계
- 를 포함하는 정보를 저장하는 방법.

청구항 2

제1항에 있어서, 뉴클레아제 복합체가 Cas 뉴클레아제(CRISPR 연관됨), 징크 핑거 뉴클레아제(ZFNs), 전사 활성화제-유사 이펙터 뉴클레아제, 알고너트 뉴클레아제, S1 뉴클레아제, 녹두 뉴클레아제 또는 DNase를 포함하는 것인 방법.

청구항 3

제1항에 있어서, 효소적 바이오암호화가 CRISPR/Cas 기반 바이오암호화를 포함하는 것인 방법.

청구항 4

제1항에 있어서, 효소적 바이오암호화가 표 1에 기재된 효소에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는 것인 방법.

청구항 5

제1항에 있어서, 복수의 올리고뉴클레오타이드가 적어도 100,000개의 올리고뉴클레오타이드를 포함하는 것인 방법.

청구항 6

제1항에 있어서, 복수의 올리고뉴클레오타이드가 적어도 백억개의 올리고뉴클레오타이드를 포함하는 것인 방법.

청구항 7

삭제

청구항 8

제1항에 있어서, 단계 (f) 전에 표면으로부터 복수의 올리고뉴클레오타이드를 방출하는 단계를 추가로 포함하는 방법.

청구항 9

제1항에 있어서, 적어도 하나의 정보 아이템이 텍스트, 오디오 또는 시각 정보를 포함하는 것인 방법.

청구항 10

제1항에 있어서, 적어도 하나의 정보 아이템이 복수의 파일을 포함하는 것인 방법.

청구항 11

제10항에 있어서, 복수의 파일이 적어도 1 기가바이트를 포함하는 것인 방법.

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

발명의 설명

배경 기술

- [0001] 상호 참조
- [0002] 본 출원은 2017년 2월 22일 출원된 미국 가출원 번호 62/462,284의 이익을 주장하며, 이는 전부 참조로 본원에 포함된다.
- [0003] 서열 목록
- [0004] 본 출원은 ASCII 포맷으로 전자적으로 제출되고 전부 참조로 본원에 포함되는 서열 목록을 포함한다. 2018년 2월 20일에 생성된 상기 ASCII 사본은 44854-738_601_SL.txt로 명명되며, 8,636 바이트의 크기이다.
- [0005] 배경
- [0006] 생체분자 기반 정보 저장 시스템, 예컨대 DNA-기반 정보 저장 시스템은 큰 저장 용량 및 시간 경과에 따른 안정성을 갖는다. 그러나, 정보 저장을 위한 생체분자에 대한 확장 가능하고 자동화되고 고도로 정밀하며 고도로 효율적인 시스템이 필요하다. 또한, 이러한 정보의 보안을 보호할 필요가 있다.
- [0007] 참조에 의한 인용
- [0008] 본 명세서에 기재된 모든 공개물, 특허 및 특허 출원은, 각각의 개별 공개물, 특허 또는 특허 출원이 구체적인 고 개별적으로 참조로 포함되는 것으로 표시되는 것과 동일한 정도로, 참조로 본원에 포함된다.

선행기술문헌

특허문헌

(특허문헌 0001) 미국 특허출원공개공보 US2005/0053968 (2005.03.10.)

(특허문헌 0002) 미국 특허출원공개공보 US2015/0269313 (2015.09.24.)

발명의 내용

[0009]

본원에는 하기 단계를 포함하는 정보를 저장하는 방법이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템을 수신하는 단계; b) 적어도 하나의 바이오암호화(bioencryption) 포맷을 선택하기 위한 명령을 수신하는 단계로서, 바이오암호화 포맷은 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화인 단계; c) 선택된 바이오암호화 포맷에 기반하여 적어도 하나의 디지털 서열을 복수의 올리고뉴클레오타이드 서열로 전환하는 단계; d) 올리고뉴클레오타이드 서열을 코딩하는 복수의 올리고뉴클레오타이드를 합성하는 단계; 및 e) 복수의 올리고뉴클레오타이드를 저장하는 단계. 또한, 본원에는 효소적 기반 바이오암호화가 CRISPR/Cas 기반 바이오암호화를 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 효소적 기반 바이오암호화가 표 1에 기재된 효소에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 전자기적 기반 바이오암호화가 약 0.01 nm 내지 약 400 nm의 전자기적 파장에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 화학적 기반 바이오암호화가 기체상 암모니아 또는 메틸아민 투여에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 친화성 기반 바이오암호화가 서열 태그(tag) 또는 친화성 태그에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 친화성 태그가 바이오틴, 디곡시게닌, Ni-니트릴로트리아세트산, 테스티오바이오틴, 히스티딘, 폴리히스티딘, myc, 혈구응집소(HA), FLAG, 형광 태그, 탠덤 친화성 정제(TAP) 태그, 글루타티온 S 트랜스퍼라제(GST), 폴리뉴클레오타이드, 압타머, 항원 또는 항체인, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 2, 3, 4 또는 5개의 바이오암호화 포맷이 이용되는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드가 적어도 100,000개의 올리고뉴클레오타이드를 포함하는, 정보를 저장하는 방법이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드가 적어도 백억개의 올리고뉴클레오타이드를 포함하는, 정보를 저장하는 방법이 제공된다.

[0010]

본원에는 하기 단계를 포함하는 정보를 인출하는 방법이 제공된다: a) 표면으로부터 복수의 올리고뉴클레오타이드를 방출하는 단계; b) 효소적, 전자기적, 화학적 또는 친화성 기반 복호화(decryption)를 복수의 올리고뉴클레오타이드에 적용하는 단계; c) 복수의 올리고뉴클레오타이드를 농축(enriching)하는 단계; d) 복수의 올리고뉴클레오타이드로부터의 농축된 올리고뉴클레오타이드를 시퀀싱하여 핵산 서열을 생성하는 단계; 및 e) 핵산 서열을 적어도 하나의 디지털 서열로 전환하는 단계로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이템을 코딩하는 것인 단계. 또한, 본원에는 복수의 올리고뉴클레오타이드의 복호화가 CRISPR/Cas 복합체를 복수의 올리고뉴클레오타이드에 적용하는 것을 포함하는, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 효소적 기반 복호화가 표 1에 기재된 효소를 적용하는 것을 포함하는, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 전자기적 기반 복호화가 약 0.01 nm 내지 약 400 nm의 파장을 적용하는 것을 포함하는, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 화학적 기반 복호화가 기체상 암모니아 또는 메틸아민 투여를 적용하는 것을 포함하는, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 친화성 기반 복호화가 서열 태그 또는 친화성 태그를 적용하는 것을 포함하는, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 친화성 태그가 바이오틴, 디곡시게닌, Ni-니트릴로트리아세트산, 테스티오바이오틴, 히스티딘, 폴리히스티딘, myc, 혈구응집소(HA), FLAG, 형광 태그, 탠덤 친화성 정제(TAP) 태그, 글루타티온 S 트랜스퍼라제(GST), 폴리뉴클레오타이드, 압타머, 항원 또는 항체인, 정보를 인출하는 방법이 제공된다. 또한, 본원에는 2, 3, 4 또는 5개 형태의 복호화가 이용되는, 정보를 인출하는 방법이 제공된다.

[0011]

본원에는 하기 유닛을 포함하는 정보를 저장하기 위한 시스템이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템에 대한 기계 명령(machine instruction) 및 적어도 하나의 바이오암호화 포맷의 선택에 대한 기계 명령을 수신하기 위한 수신 유닛으로서, 바이오암호화 포맷은 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화인 유닛; b) 선택된 바이오암호화 포맷에 기반하여 적어도 하나의 디지털 서열을 복수

의 올리고뉴클레오타이드 서열로 자동적으로 전환하기 위한 프로세서 유닛; c) 올리고뉴클레오타이드 서열을 코딩하는 복수의 올리고뉴클레오타이드를 합성하기 위해 프로세서 유닛으로부터 기계 명령을 수신하기 위한 합성기 유닛; 및 d) 합성기 유닛으로부터 침착된(deposited) 복수의 올리고뉴클레오타이드를 수신하기 위한 저장 유닛. 또한, 본원에는 효소적 기반 바이오암호화가 CRISPR/Cas 기반 바이오암호화를 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 효소적 기반 바이오암호화가 표 1에 기재된 효소에 민감한 올리고뉴클레오타이드를 합성하기 위한 기계 명령을 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 전자기적 기반 바이오암호화가 약 0.01 nm 내지 약 400 nm의 전자기적 파장에 민감한 올리고뉴클레오타이드를 합성하기 위한 기계 명령을 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 화학적 기반 바이오암호화가 기체상 암모니아 또는 메틸아민 투여에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 친화성 기반 바이오암호화가 서열 택 또는 친화성 택에 민감한 올리고뉴클레오타이드를 합성하기 위한 명령을 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 친화성 택이 바이오틴, 디곡시게닌, Ni-니트릴로트리아세트산, 데스티오바이오틴, 히스티딘, 폴리히스티딘, myc, 혈구응집소(HA), FLAG, 형광 택, 탠덤 친화성 정제(TAP) 택, 글루타티온 S 트랜스퍼라제(GST), 폴리뉴클레오타이드, 압타머, 항원 또는 항체인, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드가 적어도 100,000개의 올리고뉴클레오타이드를 포함하는, 정보를 저장하기 위한 시스템이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드가 적어도 백억개의 올리고뉴클레오타이드를 포함하는, 정보를 저장하기 위한 시스템이 제공된다.

[0012] 본원에는 하기 유닛을 포함하는 정보를 인출하기 위한 시스템이 제공된다: a) 표면 상에 복수의 올리고뉴클레오타이드를 포함하는 저장 유닛; b) 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화를 복수의 올리고뉴클레오타이드에 적용하기 위한 침착 유닛; c) 복수의 올리고뉴클레오타이드를 시퀀싱하여 핵산 서열을 획득하기 위한 시퀀싱 유닛; 및 d) 핵산 서열을 적어도 하나의 디지털 서열로 자동적으로 전환하기 위한 프로세서 유닛으로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이템을 코딩하는 것인 유닛. 또한, 본원에는 침착 유닛이 CRISPR/Cas 복합체를 복수의 올리고뉴클레오타이드에 적용하는, 정보를 인출하기 위한 시스템이 제공된다. 또한, 본원에는 효소적 기반 바이오암호화가 표 1에 기재된 효소를 적용하는 것을 포함하는, 정보를 인출하기 위한 시스템이 제공된다. 또한, 본원에는 전자기적 기반 바이오암호화가 약 0.01 nm 내지 약 400 nm의 파장을 적용하는 것을 포함하는, 정보를 인출하기 위한 시스템이 제공된다. 또한, 본원에는 화학적 기반 바이오암호화가 기체상 암모니아 또는 메틸아민 투여를 적용하는 것을 포함하는, 정보를 인출하기 위한 시스템이 제공된다. 또한, 본원에는 친화성 기반 바이오암호화가 서열 택 또는 친화성 택을 포함하는, 정보를 인출하기 위한 시스템이 제공된다. 또한, 본원에는 친화성 택이 바이오틴, 디곡시게닌, Ni-니트릴로트리아세트산, 데스티오바이오틴, 히스티딘, 폴리히스티딘, myc, 혈구응집소(HA), FLAG, 형광 택, 탠덤 친화성 정제(TAP) 택, 글루타티온 S 트랜스퍼라제(GST), 폴리뉴클레오타이드, 압타머, 항원 또는 항체인, 정보를 인출하기 위한 시스템이 제공된다.

[0013] 본원에는 하기 단계를 포함하는 정보를 저장하는 방법이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템을 수신하는 단계; b) 적어도 하나의 바이오암호화 형태에 대한 명령을 수신하는 단계; c) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오타이드 서열로 전환하는 단계; d) 복수의 바이오암호화된 올리고뉴클레오타이드 서열을 합성하는 단계; 및 e) 복수의 올리고뉴클레오타이드를 저장하는 단계.

[0014] 본원에는 하기 단계를 포함하는 정보를 저장하는 방법이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템을 수신하는 단계; b) 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화를 위한 명령을 수신하는 단계; c) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오타이드 서열로 전환하는 단계; d) 복수의 바이오암호화된 올리고뉴클레오타이드 서열을 합성하는 단계; 및 e) 복수의 올리고뉴클레오타이드를 저장하는 단계.

[0015] 본원에는 하기 단계를 포함하는 정보를 저장하는 방법이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이템을 수신하는 단계; b) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오타이드 서열로 전환하는 단계로서, 바이오암호화된 올리고뉴클레오타이드 서열 각각은 CRISPR/Cas 복합체에 의한 제거를 위해 코딩된 추가의 서열을 포함하는 것인 단계; c) 복수의 바이오암호화된 올리고뉴클레오타이드 서열을 합성하는 단계; 및 d) 복수의 올리고뉴클레오타이드를 저장하는 단계.

[0016] 본원에는 하기 단계를 포함하는 정보를 인출하는 방법이 제공된다: a) 표면으로부터 복수의 올리고뉴클레오타이드를 방출하는 단계; b) 적어도 하나의 바이오복호화(biodecryption) 형태를 복수의 올리고뉴클레오타이드에 적용하는 단계; c) 복수의 올리고뉴클레오타이드를 농축하여 복수의 농축된 올리고뉴클레오타이드를 선택하는 단계; d) 농축된 올리고뉴클레오타이드를 시퀀싱하여 핵산 서열을 생성하는 단계; 및 e) 핵산 서열을 적어도 하나의 디지털

서열로 전환하는 단계로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 단계.

[0017] 본원에는 하기 단계를 포함하는 정보를 인출하는 방법이 제공된다; a) 표면으로부터 복수의 올리고뉴클레오티드를 방출하는 단계; b) 효소적, 전자기적, 화학적 또는 친화성 기반 복호화를 복수의 올리고뉴클레오티드에 적용하는 단계; c) 복수의 올리고뉴클레오티드를 농축하여 복수의 농축된 올리고뉴클레오티드를 선택하는 단계; d) 농축된 올리고뉴클레오티드를 시퀀싱하여 핵산 서열을 생성하는 단계; 및 e) 핵산 서열을 적어도 하나의 디지털 서열로 전환하는 단계로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 단계.

[0018] 본원에는 하기 단계를 포함하는 정보를 인출하는 방법이 제공된다; a) 표면으로부터 복수의 올리고뉴클레오티드를 방출하는 단계; b) CRISPR/Cas 복합체를 복수의 올리고뉴클레오티드에 적용하는 단계; c) 복수의 올리고뉴클레오티드를 농축하여 복수의 농축된 올리고뉴클레오티드를 선택하는 단계; d) 농축된 올리고뉴클레오티드를 시퀀싱하여 핵산 서열을 생성하는 단계; 및 e) 핵산 서열을 적어도 하나의 디지털 서열로 전환하는 단계로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 단계.

[0019] 본원에는 하기를 포함하는 정보를 저장하기 위한 시스템이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이টে에 대한 기계 명령 및 적어도 하나의 바이오암호화 형태에 대한 기계 명령을 수신하기 위한 수신 유닛; b) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오티드 서열로 전환하기 위한 프로세서 유닛; c) 복수의 바이오암호화된 올리고뉴클레오티드 서열을 합성하기 위해 프로세서 유닛으로부터 기계 명령을 수신하기 위한 합성기 유닛; 및 d) 합성기 유닛으로부터 침착된 복수의 올리고뉴클레오티드를 수신하기 위한 저장 유닛.

[0020] 본원에는 하기를 포함하는 정보를 저장하기 위한 시스템이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이টে에 대한 기계 명령 및 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화에 대한 기계 명령을 수신하기 위한 수신 유닛; b) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오티드 서열로 전환하기 위한 프로세서 유닛; c) 복수의 바이오암호화된 올리고뉴클레오티드 서열을 합성하기 위해 프로세서 유닛으로부터 기계 명령을 수신하기 위한 합성기 유닛; 및 d) 합성기 유닛으로부터 침착된 복수의 올리고뉴클레오티드를 수신하기 위한 저장 유닛.

[0021] 본원에는 하기를 포함하는 정보를 저장하기 위한 시스템이 제공된다: a) 적어도 하나의 디지털 서열 형태의 적어도 하나의 정보 아이টে에 대한 기계 명령 및 CRISPR/Cas 복합체에 의한 바이오암호화에 대한 기계 명령을 수신하기 위한 수신 유닛; b) 적어도 하나의 디지털 서열을 복수의 바이오암호화된 올리고뉴클레오티드 서열로 전환하기 위한 프로세서 유닛; c) 복수의 바이오암호화된 올리고뉴클레오티드 서열을 합성하기 위해 프로세서 유닛으로부터 기계 명령을 수신하기 위한 합성기 유닛; 및 d) 합성기 유닛으로부터 침착된 복수의 올리고뉴클레오티드를 수신하기 위한 저장 유닛.

[0022] 본원에는 하기 유닛을 포함하는 정보를 인출하기 위한 시스템이 제공된다: a) 표면 상에 복수의 올리고뉴클레오티드를 포함하는 저장 유닛; b) 적어도 하나의 바이오복호화 형태를 복수의 올리고뉴클레오티드에 적용하기 위한 침착 유닛; c) 복수의 올리고뉴클레오티드 서열을 시퀀싱하여 핵산 서열을 획득하기 위한 시퀀싱 유닛; 및 d) 핵산 서열을 적어도 하나의 디지털 서열로 전환하기 위한 프로세서 유닛으로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 유닛.

[0023] 본원에는 하기 유닛을 포함하는 정보를 인출하기 위한 시스템이 제공된다: a) 표면 상에 복수의 올리고뉴클레오티드를 포함하는 저장 유닛; b) 적어도 효소적, 전자기적, 화학적 또는 친화성 기반 바이오암호화를 복수의 올리고뉴클레오티드에 적용하기 위한 침착 유닛; c) 복수의 올리고뉴클레오티드를 시퀀싱하여 핵산 서열을 획득하기 위한 시퀀싱 유닛; 및 d) 핵산 서열을 적어도 하나의 디지털 서열로 전환하기 위한 프로세서 유닛으로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 유닛.

[0024] 본원에는 하기 유닛을 포함하는 정보를 인출하기 위한 시스템이 제공된다: a) 표면 상에 복수의 올리고뉴클레오티드를 포함하는 저장 유닛; b) CRISPR/Cas 복합체를 복수의 올리고뉴클레오티드에 적용하기 위한 침착 유닛; c) 복수의 올리고뉴클레오티드를 시퀀싱하여 핵산 서열을 획득하기 위한 시퀀싱 유닛; 및 d) 핵산 서열을 적어도 하나의 디지털 서열로 전환하기 위한 프로세서 유닛으로서, 적어도 하나의 디지털 서열은 적어도 하나의 정보 아이টে을 코딩하는 것인 유닛.

도면의 간단한 설명

[0025] 본 발명의 신규한 특징은 첨부된 청구범위에 구체적으로 기재되어 있다. 본 발명의 특징 및 이점에 대한 더 나

은 이해는 본 발명의 원리를 이용하는 예시적인 실시양태를 기재하는 하기의 상세한 설명 및 첨부 도면을 참조함으로써 얻어질 것이다.

도 1은 핵산-기반 데이터 저장에 대한 예시적인 워크플로우를 도시한다.

도 2는 바이오암호화를 위한 저장에 대한 예시적인 워크플로우를 도시한다.

도 3은 바이오암호화 후 인출에 대한 예시적인 워크플로우를 도시한다.

도 4A-4B는 Cas 효소를 사용하는 바이오암호화 방법을 도시한다.

도 5A-5C는 다양한 올리고뉴클레오타이드 서열 디자인 스킴(scheme)을 도시한다.

도 6A-6C는 다양한 올리고뉴클레오타이드 서열 디자인 스킴을 도시한다.

도 7A-7B는 바코드 디자인 스킴을 도시한다.

도 8은 24개 영역 또는 서브-필드(각각 256개 클러스터의 어레이를 가짐)를 포함하는 올리고뉴클레오타이드 합성을 위해 구성된 플레이트를 도시한다.

도 9는 16 x16개의 클러스터(각각의 클러스터는 121개의 개별 로커스(locus)를 가짐)를 갖는 도 8의 서브-필드의 근접도를 도시한다.

도 10은 도 8의 클러스터의 상세도를 도시하며, 여기서 클러스터는 121개의 로커스를 갖는다.

도 11A는 복수의 채널을 갖는 플레이트의 정면도는 도시한다.

도 11B는 복수의 채널을 갖는 플레이트의 단면도를 도시한다.

도 12A-12B는 가요성 구조물을 위한 연속 루프 및 릴-투-릴(reel-to-reel) 배열을 도시한다.

도 13A-13C는 각각 플랫폼 피처(feature)(로커스), 채널 또는 웰을 갖는 가요성 구조물의 확대를 도시한다.

도 14A는 본원에 기재된 구조물 상의 피처의 확대를 도시한다.

도 14B-14C는 본원에 기재된 구조물 상의 마킹을 도시한다.

도 15는 올리고뉴클레오타이드 합성 물질 침착 디바이스를 도시한다.

도 16은 올리고뉴클레오타이드 합성 워크플로우를 도시한다.

도 17은 컴퓨터 시스템의 일례를 도시한다.

도 18은 컴퓨터 시스템의 아키텍처를 도시하는 블록 다이어그램이다.

도 19는 복수의 컴퓨터 시스템, 복수의 휴대전화 및 개인 데이터 어시스턴트, 및 네트워크 접속 저장장치(NAS)를 통합하도록 구성된 네트워크를 나타내는 다이어그램이다.

도 20은 공유된 가상 어드레스 메모리 공간을 이용하는 멀티프로세서 컴퓨터 시스템의 블록 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

정의

달리 정의되지 않는 한, 본원에 사용된 모든 기술 용어 및 과학 용어는 본 발명이 속하는 기술분야의 통상의 기술자가 일반적으로 이해하는 것과 동일한 의미를 갖는다.

본 개시 전반에 걸쳐, 다양한 실시양태가 범위 포맷으로 제시된다. 범위 포맷의 기재는 단지 편의와 간략성을 위한 것이며, 임의의 실시양태의 범위에 대한 확고한 제한으로서 해석되어서는 안된다는 것을 이해해야 한다. 따라서, 범위 기재는, 문맥이 명백하게 달리 지시하지 않는 한, 모든 가능한 하위범위 뿐만 아니라 그 범위 내의 개별 수치를 하한 단위의 10분의 1까지 구체적으로 개시하는 것으로 간주되어야 한다. 예컨대, 1 내지 6과 같은 범위 기재는 1 내지 3, 1 내지 4, 1 내지 5, 2 내지 4, 2 내지 6, 3 내지 6 등과 같은 하위범위 뿐만 아니라, 그 범위 내의 개별 값, 예컨대 1.1, 2, 2.3, 5 및 5.9를 구체적으로 개시하고 있는 것으로 간주되어야 한다. 이는 범위의 폭과는 관계 없이 적용된다. 이들 개재 범위의 상한 및 하한은 독립적으로 더 작은 범위에 포함될 수 있고, 기재된 범위 내에서 임의의 특별히 배제된 한계에 따라 본 발명 내에 포함되기도 한다. 기재된

범위가 한계들 중 하나 또는 둘 다를 포함하는 경우, 문맥이 명백하게 달리 지시하지 않는 한, 포함된 한계들 중 하나 또는 둘 다를 배제하는 범위도 본 발명 내에 포함된다.

[0029] 본원에 사용된 용어는 단지 특정 실시양태를 설명하기 위한 목적으로 사용된 것이고, 임의의 실시양태를 제한하려는 것이 아니다. 본원에서 사용되는 바와 같이, 단수형은, 문맥이 명백하게 달리 지시하지 않는 한, 복수형도 포함하도록 의도된다. 본 명세서에서 사용되는 용어 "포함하다" 및/또는 "포함하는"은 기재된 피처(feature), 정수, 단계, 조작, 요소 및/또는 성분의 존재를 명시하지만, 하나 이상의 다른 피처, 정수, 단계, 조작, 요소, 성분, 및/또는 이의 그룹의 존재 또는 첨가를 배제하는 것은 아니다. 본원에서 사용되는 바와 같이, 용어 "및/또는"은 하나 이상의 연관된 열거된 아이템의 임의의 모든 조합을 포함한다.

[0030] 문맥으로부터 구체적으로 기재되거나 명시적이지 않는 한, 본원에서 사용되는 바와 같이, 수치 또는 수치 범위와 관련하여 용어 "약"은 기재된 수치 및 이 수치의 $\pm 10\%$, 또는 범위로 기재된 값에 대해 기재된 하한치의 10% 하한 및 기재된 상한치의 10% 상한을 의미하는 것으로 이해된다.

[0031] 본원에서 사용되는 바와 같이 용어 "올리고뉴클레오타이드"는 "올리고핵산"과 상호교환적으로 사용된다. 용어 "올리고뉴클레오타이드" 및 "올리고핵산"은 단일-가닥 분자 뿐만 아니라 이중- 또는 삼중-가닥 핵산을 포함한다.

[0032] 핵산 기반 정보 저장

[0033] 본원에는 핵산-기반 정보(데이터) 저장을 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 예시적인 워크플로우가 도 1에 제공된다. 제1 단계에서, 정보 아이템을 코딩하는 디지털 서열(즉, 컴퓨터에 의한 프로세싱을 위한 2진 코드의 디지털 정보)가 수신된다(101). 암호화(encryption)(103) 스킴이 적용되어 2진 코드로부터의 디지털 서열을 핵산 서열로 전환한다(105). 핵산 연장을 위한 표면 물질, 핵산 연장을 위한 로커스(배열 스팟으로도 알려짐)에 대한 디자인 및 핵산 합성을 위한 시약이 선택된다(107). 핵산 합성을 위한 구조물의 표면이 제조된다(108). 드 노보 올리고뉴클레오타이드 합성이 수행된다(109). 합성된 올리고뉴클레오타이드가 저장되고(111), 전체 또는 일부가 후속 방출에 이용가능하다(113). 일단 방출되면, 올리고뉴클레오타이드 전체 또는 일부가 시퀀싱되고(115), 복호화되어(117) 핵산 서열이 다시 디지털 서열로 전환된다. 이어서, 디지털 서열은 어셈블링되어(119) 최초의 정보 아이템을 코딩하는 정렬을 수득한다.

[0034] 또한, 본원에는 적어도 하나의 정보 아이템을 코딩하는 하나 이상의 디지털 서열의 수신(201), 하나 이상의 디지털 서열의 핵산 서열로의 전환(203), 핵산 서열의 암호화(205) 및 암호화된 핵산 서열의 드 노보 올리고뉴클레오타이드 합성(207)을 포함하는 보증된 DNA-기반 정보 저장을 위한 방법 및 시스템이 제공된다. 도 2를 참조한다.

[0035] 본원에는, 디지털 서열로부터 핵산 서열로의 전환, 바이오암호화, 바이오복호화 또는 이들 단계의 임의의 것의 조합에 대한 기계 명령이 수신되는, 핵산-기반 정보 저장을 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 전환을 위한 목적하는 정보 아이템, 및 옵션 목록, 예컨대, 제한 없이, 효소적 기반(예컨대, CRISPR/Cas 복합체 또는 제한 효소 분해), 전자기 복사 기반(예컨대, 광분해 또는 광검출), 화학적 절단(예컨대, 티미딘-숙시닐 핵사미드 CED 포스포르아미다이트(ChemGenes로부터의 CLP-2244)를 절단하기 위한 기체상 암모니아 또는 메틸아민 처리), 및 친화성 기반(예컨대, 하이브리드화를 위한 서열 태그 또는 캡처 시약에 대해 증진된 친화성을 갖는 변형된 뉴클레오타이드의 통합) 형태의 바이오암호화로부터 선택되는 하나 이상의 유형의 바이오암호화에 대한 기계 명령이 수신될 수 있다. 특정 바이오암호화 선택의 수신 후, 올리고뉴클레오타이드의 드 노보 합성을 위해 합성 명령을 물질 침착 디바이스에 제공하기 전에, 프로그램 모듈은 정보 아이템을 핵산 서열로 전환하고 바이오암호화된 버전의 서열을 디자인하기 위한 디자인 명령을 적용하는 단계를 수행한다. 일부 경우에, 바이오암호화 카테고리 내의 하나 이상의 종을 선택하기 위한 기계 명령이 제공된다.

[0036] 또한, 본원에는 표면으로부터 올리고뉴클레오타이드의 방출(301), 목적하는 올리고뉴클레오타이드의 농축(303), 올리고뉴클레오타이드의 시퀀싱(305), 핵산 서열의 복호화(307), 및 정보 아이템을 코딩하는 하나 이상의 디지털 서열의 어셈블리(309)를 포함하는 보증된 DNA-기반 정보 인출을 위한 방법 및 시스템이 제공된다. 도 3을 참조한다.

[0037] 본원에 기재된 기계 명령은 또한 바이오복호화에 대해 제공될 수 있다. 바이오복호화는 기계 명령의 수신을 포함할 수 있다. 이러한 명령은 옵션 목록, 예컨대, 제한 없이, 효소적 기반(예컨대, CRISPR/Cas 복합체 또는 제한 효소 분해), 전자기 복사 기반(예컨대, 광분해 또는 광검출), 화학적 절단(예컨대, 티미딘-숙시닐 핵사미드 CED 포스포르아미다이트(ChemGenes로부터의 CLP-2244)를 절단하기 위한 기체상 암모니아 또는 메틸아민 처리), 및 친화성 기반(예컨대, 하이브리드화를 위한 서열 태그 또는 캡처 시약에 대해 증진된 친화성을 갖는 변형된 뉴

클레오티드의 통합) 형태의 올리고뉴클레오티드의 바이오복호화를 포함할 수 있다. 특정 바이오복호화 선택의 수신 후, 프로그램 모듈은 올리고뉴클레오티드 농축을 위한 조정제(들)를 방출하는 단계를 수행한다. 농축 후, 올리고뉴클레오티드는 시퀀싱되고, 임의적으로, 더 긴 핵산 서열과 정렬되고, 정보 아이템에 대응하는 디지털 서열로 전환된다. 일부 경우에, 바이오복호화 카테고리 내의 하나 이상의 종의 선택을 위한 기계 명령이 제공된다.

[0038] 정보 아이템

[0039] 임의적으로, 본원에 개시된 DNA 데이터 저장 또는 공정의 초기 단계는 개시 코드(예컨대, 디지털 코드) 형태의 하나 이상의 정보 아이템을 얻거나 수신하는 것을 포함한다. 정보 아이템은 텍스트, 오디오 및 시각 정보를 포함하나, 이들로 제한되지 않는다. 정보 아이템에 대한 예시적인 공급원은 서적, 정기간행물, 전자적 데이터베이스, 의료 기록, 편지, 양식, 음성 기록, 동물 소리, 생물학적 프로파일, 방송, 영화, 짧은 비디오, 이메일, 부기 전화 기록, 인터넷 활동 기록, 도면, 그림, 프린트, 사진, 화소로 처리된 그래픽 및 소프트웨어 코드를 포함하나, 이들로 제한되지 않는다. 정보 아이템에 대한 예시적인 생물학적 프로파일 공급원은 유전자 라이브러리, 게놈, 유전자 발현 데이터 및 단백질 활성 데이터를 포함하나, 이들로 제한되지 않는다. 정보 아이템에 대한 예시적인 포맷은 .txt, .PDF, .doc, .docx, .ppt, .pptx, .xls, .xlsx, .rtf, .jpg, .gif, .psd, .bmp, .tiff, .png 및 .mpeg를 포함하나, 이들로 제한되지 않는다. 디지털 포맷의, 정보 아이템을 코딩하는 개별 파일 크기의 양, 또는 정보 아이템을 코딩하는 복수의 파일의 양은 최대 1024 바이트(1 KB와 같음), 1024 KB(1MB와 같음), 1024 MB(1 GB와 같음), 1024 GB(1 TB와 같음), 1024 TB(1 PB와 같음), 1 엑사바이트, 1 제타바이트, 1 요타바이트, 1 제노타바이트 또는 이의 초과를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1 기가바이트(GB)이다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 또는 1000 초과 기가바이트이다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1 테라바이트(TB)이다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 또는 1000 초과 테라바이트이다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1 페타바이트(PB)이다. 일부 경우에, 디지털 정보의 양은 적어도 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 또는 1000 초과 페타바이트이다.

[0040] 암호화

[0041] 생물학적 암호화 및 복호화

[0042] 본원에는 정보 아이템을 코딩하는 디지털 서열의 수신 후 생물학적 암호화("바이오암호화"로도 알려짐)를 포함하는 디바이스, 구성물, 시스템 및 방법이 제공된다. 본원에 기재된 바이오암호화 및 바이오복호화의 개별 형태에 추가하여, 본원에는 또한 생물학적 서열을 마스킹하는 하나 이상의 클래스 또는 종의 선택을 정보 저장 및/또는 인출을 위한 워크플로우로 통합하는 공정이 제공된다.

[0043] 본원에는 생물학적 암호화를 포함하는 큰 핵산 서열 집단으로부터 관심 핵산 서열을 표적 농축하는 디바이스, 구성물, 시스템 및 방법이 제공된다. 일부 경우에, 노이즈로부터 표적 시스템을 농축하는 생물학적 암호화가 이용된다. 일부 경우에, 표적 시그널은 관심 핵산 서열이다. 일부 경우에, 생물학적 암호화는 관심 핵산 서열을 공지된 서열을 갖는 더 큰 핵산 서열 집단으로 도입하는 것을 포함한다. 공지된 핵산 서열은 암호화 핵산 서열로 지칭될 수 있다. 일부 경우에, 암호화 핵산은 복호화된다. 일부 경우에, 공지된 핵산 서열의 복호화로 관심 핵산 서열의 시그널-대-노이즈 비율 증가가 초래된다.

[0044] 본원에는 생물학적 분자 암호화를 정보 저장 및/또는 인출 워크플로우로 통합하는 것을 포함하는 디바이스, 구성물, 시스템 및 방법이 제공된다. 바이오암호화 및 바이오복호화의 예시적인 형태는 효소적 기반, 전자기 복사 기반, 화학적 절단 및 친화성 기반 바이오암호화 및 바이오복호화를 포함하나, 이들로 제한되지 않는다.

[0045] 본원에는 뉴클레아제 복합체 활성 기반 암호화의 적용을 포함하는 디바이스, 구성물, 시스템 및 방법이 제공된다. 예시적인 뉴클레아제는 Cas 뉴클레아제(CRISPR 연관됨), 징크 핑거 뉴클레아제(ZFNs), 전사 활성화제-유사 이펙터 뉴클레아제, 알고너트 뉴클레아제, S1 뉴클레아제, 녹두 뉴클레아제 또는 DNase를 포함하나, 이들로 제한되지 않는다. 예시적인 Cas 뉴클레아제는 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 및 c2c3을 포함하나, 이들로 제한되지 않는다. 일부 경우에, Cas 뉴클레아제는

Cas9이다. 일부 경우에, CRISPR/Cas 복합체는 하나 이상의 핵산 서열의 미리결정된 제거를 가능하게 한다. 일부 경우에, 본원에 기재된 농축 단계는 하이브리드화에 의한 풍부 서열의 고갈(DASH)을 포함한다. 일부 경우에, DASH는 뉴클레아제의 적용을 포함한다. 예컨대, Cas9와 같은 뉴클레아제는 가이드 RNA("gRNA") 서열을 포함하는 CRISPR 복합체에 결합되는 경우 더 긴 형태의 핵산 서열이 더 이상 온전하지 않도록 가닥 파괴를 유도한다. 일부 경우에, 절제된 핵산은 농축 후 후속 증폭에 이용가능하지 않다. 일부 경우에, gRNA는 Cas9 효소를 특정 스트레치의 핵산으로 안내한다. 대안적 배열에서, gRNA는 절단을 위한 다중 부위를 갖는다. gRNA-기반 시스템은 높은 특이성 및 선택성으로 암호화 코드의 생성을 가능하게 한다. 예컨대, CRISPR/Cas9 기반 시스템은 절단할 서열을 확인하는 데 20 bp를 사용하므로, 4 염기 시스템을 사용하여 복호화를 위한 미리결정된 gRNA 서열을 디자인하는 데 적어도 약 10^{12} 개의 상이한 가능성이 이용가능하다. 관련 없는("정크"로도 알려짐) DNA를 제거한 후, 표적 서열을 코딩하는 미리결정된 올리고뉴클레오티드는 다운스트림 프로세싱, 예컨대 증폭 및 시퀀싱되어 관련 없는(정크) 서열 없이 최종 서열을 생성한다. 일부 경우에, 복수의 올리고뉴클레오티드의 각각의 올리고뉴클레오티드는 다중 위치에서 변형(예컨대, 절단, 염기 스와핑, 재조합)이 디자인된다. 예컨대, 복수의 올리고뉴클레오티드의 각각의 올리고뉴클레오티드는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 이를 초과하는 gRNA 서열에 결합하기 위한 상보적 영역과 함께 합성된다. 이러한 배열에서, 복수의 올리고뉴클레오티드 각각은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 이를 초과하는 위치에서 뉴클레아제(예컨대, CRISPR/Cas) 복합체 활성화 이후 절단, 염기 스와핑, 재조합된다.

[0046] CRISPR/Cas9를 사용하는 데이터 암호화를 위한 표적 농축에 대한 제1 공정은 도 4A에 도시된다. DNA 서열 집단(401)은 DNA 정보(403) 및 암호화된 DNA(405)를 포함한다. DNA 정보(403) 및 암호화된 DNA(405)는 각각 어댑터 서열(402) 및 DNA 서열(404, 406)을 포함한다. 가이드 RNA(409)가 DNA 서열 집단(401)에 첨가된다(407). 가이드 RNA(409)는 암호화된 DNA(405)에서 암호화된 절단 서열을 인식함으로써 암호화된 DNA(405)를 제거하는 데 사용된다. 가이드 RNA(409) 첨가 후, 암호화된 DNA(405)가 절단되어 핵산 서열은 더 이상 온전하지 않다. 따라서, 예컨대 더 이상 온전하지 않은 핵산 서열을 포함하는 암호화된 DNA(405)가 증폭될 수 없는 경우, 암호화된 DNA(405)는 집단으로부터 제거되어(411) DNA 정보(403)를 남긴다.

[0047] CRISPR/Cas9를 사용하는 데이터 암호화를 위한 표적 농축에 대한 제2 공정은 도 4B에 도시된다. DNA 서열 집단(421)은 DNA 정보(423) 및 암호화된 DNA(425)를 포함한다. DNA 정보(423) 및 암호화된 DNA(425)는 각각 어댑터 서열(422) 및 DNA 서열(424, 426)을 포함한다. 가이드 RNA(429) 및 도너 DNA(431)가 DNA 서열 집단(421)에 첨가된다(427). 가이드 RNA(429)는 암호화된 DNA(425)에서 암호화된 절단 서열을 인식하고 도너 DNA(431)의 삽입을 위한 절단 부위를 생성한다. 도너 DNA(431)의 삽입으로 암호화된 DNA(425)에 삽입 또는 프레임쉬프트가 초래된다. 일부 경우에, 도너 DNA(431)의 삽입으로 하이브리드화를 위한 서열 택의 도입 또는 캡처 시약에 대해 증진된 친화성을 갖는 변형된 뉴클레오티드의 통합이 초래된다. 예컨대, 도너 DNA(431)는 형광 프로브에 의해 인식된다. 일부 경우에, 도너 DNA(431)는 전자기 복사 기반(예컨대, 광분해 또는 광검출) 또는 화학적 절단 기반(예컨대, 티미딘-숙시닐 헥사미드 CED 포스포르아미다이트(ChemGenes로부터의 CLP-2244)를 절단하기 위한 기체상 암모니아 또는 메틸아민 처리) 바이오암호화 및/또는 바이오복호화를 위한 서열을 도입한다. 일부 경우에, 암호화된 DNA(425)는 증폭을 위해 더 이상 인식되지 않으며, DNA 정보(423)만이 증폭되어 DNA 정보(423)의 농축이 초래된다.

[0048] 본원에 기재된 뉴클레아제 복합체 활성화 기반 암호화의 적용을 포함하는 디바이스, 조성물, 시스템 및 방법은 염기 스와핑 또는 서열 스와핑을 포함할 수 있다. 예컨대, CRISPR/Cas를 사용하는 바이오암호화 및 바이오복호화는 염기 스와핑 또는 서열 스와핑을 포함한다. 일부 경우에, 바이오암호화는, 망가진(disabled) 또는 "죽은" Cas9("dCas9")는 더 이상 스플라이싱 기능을 갖지 못하나 또 다른 효소적 활성의 첨가와 함께 상이한 표적 분자 변형 기능을 수행하는, CRISPR/dCas9를 포함한다. 예컨대, 시티딘 디아미나제를 dCas9에 테더링하는 것은 C-G DNA 염기쌍을 T-A 염기쌍으로 전환한다. 대안적 dCas9 공정에서, dCas9에 테더링된 상이한 효소는 표적 DNA에서 염기 C의 T로의 변화 또는 G의 A로의 변화를 초래한다.

[0049] 본원에는 제한 효소의 적용을 포함하는 바이오암호화 및 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 일부 경우에, 제한 효소는 효소 인식 부위를 표적한다. 일부 경우에, 효소 인식 부위는 특정 뉴클레오티드 서열이다. 일부 경우에, 제한 효소는 효소 인식 부위에서 또는 근처에서 포스포이트 골격을 절단한다. 일부 경우에, 인식 부위의 절단으로 비-평활 또는 평활 말단이 초래된다. 일부 경우에, 제한 효소는 뉴클레오티드(예컨대, A, T, G, C, U)를 인식한다. 일부 경우에, 제한 효소는 변형, 예컨대, 제한 없이, 메틸화, 히드록실화 또는 당화를 인식한다. 일부 경우에, 제한 효소는 단편화를 초래한다. 일부 경우에, 단편화는 5' 돌출부, 3' 돌출부, 평활 말단 또는 이의 조합을 갖는 단편을 생성한다. 일부 경우에, 단편은 예컨대 크기에 기

반하여 선택된다. 일부 경우에, 제한 효소에 의한 단편화 후 결합이 이어진다. 예컨대, 제한 효소에 의한 단편화는 예측가능한 돌출부를 남기고, 이어서 핵산 단편 상의 예측가능한 돌출부에 상보적인 돌출부를 포함하는 하나 이상의 어댑터 올리고뉴클레오티드와의 연결에 사용된다. 예시적인 제한 효소 및 그들의 인식 서열이 표 1에 제공된다.

표 1

제한 효소

인식 서열	효소
AA/CGTT	AclI
A/AGCTT	HindIII HindIII-HF®
AAT/ATT	SspI SspI-HF®
/AATT	MluCI Tsp509I
A/CATGT	PciI
A/CCGGT	AgeI AgeI-HF®
ACCTGC(4/8)	BspMI BfuAI
A/CCWGGT	SexAI
A/CGCGT	MluI MluI-HF®
ACGGC(12/14)	BceAI
A/CGT	HpyCH4IV
ACN/GT	HpyCH4III
(10/15)ACNNNNGTAYC(12/7) (서열번호: 5)	BaeI
(9/12)ACNNNNCTCC(10/7) (서열번호: 6)	BsaXI
A/CRYGT	AflIII
A/CTAGT	SpeI SpeI-HF®
ACTGG(1/-1)	BsrI
ACTGGG(5/4)	BmrI
A/GATCT	BglII
AGC/GCT	AfeI
AG/CT	AluI
AGG/CCT	StuI
AGT/ACT	ScaI ScaI-HF®
AT/CGAT	ClaI BspDI
ATCTATGTCGGGTGCGGAGAAAGAGGTAAT(-15/-19) (서열번호: 7)	PI-SceI
ATGCA/T	NsiI NsiI-HF®
AT/TAAT	AseI
ATTT/AAAT	SwaI
(11/13)CAANNNNNGTGG(12/10) (서열번호: 8)	CspCI
C/AATTG	MfeI MfeI-HF®
CACGAG(-5/-1)	BssSI BssSI
CACGAG	Nb BssSI
CACGTC(-3/-3)	BmgBI
CAC/GTG	PmlI
CACNNN/GTG	DraIII DraIII-HF®

[0050]

CACNN/NGTG (서열번호: 9)	AleI
CAGCAG(25/27)	EcoP15I
CAG/CTG	PvuII PvuII-HF®
CAGNNN/CTG	AlwNI
CAGTG(2/0)	BtsIMutI
NNCASTGNN/	TspRI
CA/TATG	NdeI
CATG/	NlaIII
C/ATG	CviAII
/CATG	FatI
CAYNN/NNRTG (서열번호: 10)	MslI
CC(12/16)	FspEI
CCANNNNN/NNNTGG (서열번호: 11)	XcmI
CCANNNNN/NTGG (서열번호: 12)	BstXI
CCANNNN/NTGG (서열번호: 13)	PfiMI
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)	AcI
CCGC/GG	SacII
CCGCTC(-3/-3)	BsrBI
C/CGG	MspI HpaII
CC/NGG	ScrFI
/CCNGG	BssKI StyD4I
C/CNNGG	BsaII
CCNNNNN/NGG (서열번호: 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/NNAGG (서열번호: 15)	EcoNI
CCTTC(6/5)	HpyAV

[0051]

CC/WGG	BstNI
/CCWGG	PspGI
C/CWWGG	StyI StyI-HF®
(10/12)CGANNNNNNTGC(12/10) (서열번호: 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
CMG/CKG	MspAII
CNNR(9/13)	MspJI
CR/CCGGYG	SgrAI
C/TAG	BfaI
CTCAG(9/7)	BspCNI
C/TCGAG	XhoI PaeR7I ThI
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/NNITC (서열번호: 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/NGTC	Tth111I PfiFI
GACNN/NGTC (서열번호: 18)	PshAI
GACNNN/NGTC (서열번호: 19)	AhdI
GACNNNN/NGTC (서열번호: 20)	DrdI
GAG/CTC	Eco53kI
GAGCT/C	SacI SacI-HF®
GAGGAG(10/8)	BseRI

[0052]

GAGTC(4/5)	PleI
GAGTC(4/-5)	Nt.BstNBI
GAGTC(5/5)	MlyI
G/ANTC	Hinfl
GAT/ATC	EcoRV EcoRV-HF®
/GATC	MboI Sau3AI DpnII BfuCI
GA/TC	DpnI
GATNN/NNATC (서열번호: 21)	BsaBI
G/AWTC	TfiI
GCAATG(2/0)	BsrDI
GCAATG	Nb.BsrDI
GCAGC(8/12)	BbvI
GCAGTG(2/0)	BtsI BtsaI
GCAGTG	Nb.BtsI
GCANNNN/NTGC (서열번호: 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 (서열번호: 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 (서열번호: 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®

[0053]

GGATG(9/13)	FokI
GGATG(2/0)	BtsCI
GG/CC	HaeIII PhoI
GGCCGG/CC	FseI
GGCCNNNN/NGGCC (서열번호: 25)	SfiI
GG/CGCC	NarI
G/GCGCC	KasI
GGC/GCC	SfoI
GGCGC/C	PluII
GG/CGCGCC	AscI
GGCGGA(11/9)	EciI
GGGAC(10/14)	BsmFI
GGGCC/C	ApaI
G/GGCCC	PspOMI
G/GNCC	Sau96I
GGN/NCC	NlaIV
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®
GICTC(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	BsiHKAII
R/AATTY	ApoI ApoI-HF

[0054]

RCATG/Y	NspI
R/CCGGY	BsrFI BsrFol
R/GATCY	BstYI
RGCGC/Y	HaeII
RG/CY	CviKI-1
RG/GNCCY	EcoO109I
RG/GWCCY	PpuMI
TAACTATAACGGTCCTAAGGTAGCGAA(-9/-13) (서열번호: 26)	I-CeuI
TAC/GTA	SnaBI
TAGGGATAACAGGGTAAT(-9/-13) (서열번호: 27)	I-SceI
T/CATGA	BspHI
T/CCGGA	BspEI
TCCRAC(20/18)	MmeI
T/CGA	TaqoI
TCG/CGA	NruI NruI-HF®
TCN/GA	Hpy188I
TC/NNGA	Hpy188III
T/CTAGA	XbaI
T/GATCA	BclI
TG/CA	HpyCH4V
TGC/GCA	FspI
TGGCAAACAGCTATTATGGGTATTATGGGT(-13/-17) (서열번호: 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

[0055]

[0056]

본원에는 수복 효소의 적용을 포함하는 바이오암호화 및 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. DNA 수복 효소는, 일부 경우에, 특정 유기체 또는 바이러스로부터 유도되거나 이의 비-천연 발생 변이체이다. 예시적인 DNA 수복 효소는 이. 콜라이 엔도뉴클레아제 IV, Tth 엔도뉴클레아제 IV, 인간 AP 엔도뉴클레아제, 글리코실라제, 예컨대 UDG, 이. 콜라이 3-메틸아데닌 DNA 글리코실라제(AIkA) 및 인간 Aag, 글리코실라제/리아제, 예컨대 이. 콜라이 엔도뉴클레아제 III, 이. 콜라이 엔도뉴클레아제 VIII, 이. 콜라이 Fpg, 인간 OGG1 및 T4 PDG, 및 리아제를 포함하나, 이들로 제한되지 않는다. 예시적인 추가의 DNA 수복 효소가 표 2에 열거된다.

표 2

DNA 수복 효소

유전자명	활성	수탁 번호
UNG	우라실-DNA 글리코실라제	NM_080911
SMUG1	우라실-DNA 글리코실라제	NM_014311
MBD4	CpG 서열에서 G 반대편의 U 또는 T 제거	NM_003925
TDG	G 반대편의 U, T 또는 에테노 C 제거	NM_003211
OGG1	C 반대편의 8-옥소 G 제거	NM_016821
MUTYH (MYH)	8-옥소 G 반대편의 A 제거	NM_012222
NTHL1 (NTH1)	고리-포화된 또는 단편화된 피리미딘 제거	NM_002528
MPG	3-meA, 에테노 A, 히포크산틴 제거	NM_002434
NEIL1	티민 글리콜 제거	NM_024608
NEIL2	피리미딘의 산화 생성물 제거	NM_145043
XPC	RAD23B, CETN2 와 복합체로서 손상된 DNA 에 결합	NM_004628
RAD23B (HR23B)	XPC, CETN2 와 복합체로서 손상된 DNA 에 결합	NM_002874
CETN2	XPC, RAD23B 와 복합체로서 손상된 DNA 에 결합	NM_004344
RAD23A (HR23A)	HR23B 를 대체	NM_005053
XPA	프리절개 복합체에서 손상된 DNA 에 결합	NM_000380
RPA1	프리절개 복합체에서 DNA 에 결합	NM_002945
RPA2	프리절개 복합체에서 DNA 에 결합	NM_002946
RPA3	프리절개 복합체에서 DNA 에 결합	NM_002947
ERCC5 (XPG)	3' 절개	NM_000123
ERCC1	5' 절개 서브유닛	NM_001983
ERCC4 (XPF)	5' 절개 서브유닛	NM_005236
LIG1	DNA 접합	NM_000234
CKN1(CSA)	코케인 증후군; 전사-커플링된 NER 에 필요	NM_000082
ERCC6 (CSB)	코케인 증후군; 전사-커플링된 NER 에 필요	NM_000124

[0057]

XAB2 (HCNP)	코케인 증후군; 전사-커플링된 NER에 필요	NM_020196
DDB1	XP 그룹 E에 결합이 있는 복합체	NM_001923
DDB2	DDB1, DDB2	NM_000107
MMS19L (MMS19)	전사 및 NER	NM_022362
FEN1 (DNase IV)	플랩 엔도뉴클레아제	NM_004111
SPO11	엔도뉴클레아제	NM_012444
FLJ35220 (ENDOV)	히포크산틴 및 우라실의 3' 절개	NM_173627
FANCA	DNA 교차결합의 내성 또는 수복에 수반됨	NM_000135
FANCB	DNA 교차결합의 내성 또는 수복에 수반됨	NM_152633
FANCC	DNA 교차결합의 내성 또는 수복에 수반됨	NM_000136
FANCD2	DNA 교차결합의 내성 또는 수복에 수반됨	NM_033084
FANCE	DNA 교차결합의 내성 또는 수복에 수반됨	NM_021922
FANCF	DNA 교차결합의 내성 또는 수복에 수반됨	NM_022725
FANCG (XRCC9)	DNA 교차결합의 내성 또는 수복에 수반됨	NM_004629
FANCL	DNA 교차결합의 내성 또는 수복에 수반됨	NM_018062
DCLRE1A (SNM1)	DNA 교차결합 수복	NM_014881
DCLRE1B (SNM1B)	SNM1에 관련됨	NM_022836
NEIL3	NEIL1 및 NEIL2와 유사	NM_018248
ATRIP (TREX1)	TREX1 / ATRIP 유전자의 ATR-상호 작용 단백질 5' 대체 ORF	NM_130384
NTH	손상된 피리미딘을 제거	NP_416150.1
NEI	손상된 피리미딘을 제거	NP_415242.1
NFI	데옥시이노신 3' 엔도뉴클레아제	NP_418426.1
MUTM	포름아미도피리미딘 DNA 글리코실라제	NP_418092.1
UNG	우라실-DNA 글리코실라제	NP_417075.1
UVRA	DNA 절제 수복 효소 복합체	NP_418482.1
UVRB	DNA 절제 수복 효소 복합체	NP_415300.1
UVRC	DNA 절제 수복 효소 복합체	NP_416423.3
DENV	피리미딘 이량체 글리코실라제	NP_049733.1

[0058]

[0059]

본원에는 핵산 변형을 포함하는 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 일부 경우에, 핵산 변형은 시퀀싱 반응에서 핵산 서열의 활성화에 영향을 끼친다. 예컨대, 핵산 변형은 암호화된 핵산 서열이 증폭되는 것을 방지한다. 일부 경우에, 핵산 변형은 메틸화된 염기, PNA (펩티드 핵산) 뉴클레오타이드, LNA (잠금 핵산) 뉴클레오타이드 및 2'-O-메틸-변형된 뉴클레오타이드를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 핵산 변형은 시토신, 구아닌, 아데닌 또는 티민이 아닌 변형된 핵염기를 포함한다. 비제한적 핵염기는 우라실, 3-meA(3-메틸아데닌), 히포크산틴, 8-oxoG(7,8-디히드로-8-옥소구아닌), FapyG, FapyA, Tg(티민 글리콜), hoU(히드록시우라실), hmU(히드록시메틸우라실), fU(포르밀우라실), hoC(히드록시시토신), fC(포르밀시토신), 5-meC(5-메틸시토신), 6-meG(06-메틸구아닌), 7-meG(N7-메틸구아닌), εC(에테노시토신), 5-caC(5-카복실시토신), 2-hA, εA(에테노아데닌), 5-fU(5-플루오로우라실), 3-meG(3-메틸구아닌) 및 이소알루르산을 포함하나, 이들로 제한되지 않는다.

[0060]

본원에는 핵산 프로브 서열을 사용하는 것을 포함하는 바이오암호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 일부 경우에, 핵산 서열의 일부에 상보적인 핵산 프로브 서열은 이어서 뉴클레아제에 의해 제거된다. 예컨대, 뉴클레아제는 핵산 프로브와 핵산 서열 사이에 형성되는 이중 가닥 핵산 분자를 인식하는 듀플렉스 특이적 뉴클레아제이다. 일부 경우에, 핵산 프로브는 핵산 서열을 포획하고 단리하는 것을 가능하게 한다. 일부 경우에, 핵산 프로브는 적어도 약 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100개, 또는 100개 초과 뉴클레오타이드 길이를 포함한다.

[0061]

일부 경우에, 핵산 서열은 표지물, 예컨대, 이로 제한됨이 없이, 친화성 태그, 예컨대 바이오틴, 디곡시게닌, Ni-니트릴로트리아세트산, 테스티오바이오틴, 히스티딘, 폴리히스티딘, myc, 혈구응집소(HA), FLAG, 형광 태그, 탠덤

친화성 정제(TAP) 탭, 글루타티온 S 트랜스퍼라제(GST), 폴리뉴클레오티드, 압타머, 폴리펩티드(예컨대, 항원 또는 항체), 또는 이의 유도체를 포함하는 핵산 프로브를 사용하여 확인된다. 일부 경우에, 표지물은 광 흡수, 형광, 화학발광, 전기화학발광, 질량 또는 전하에 의해 검출된다. 형광단의 비제한적 예는 알렉사-플루오르 염료(예컨대, 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, 및 Alexa Fluor® 750), APC, 케스케이드 블루, 케스케이드 옐로우 및 R-피코에리트린(PE), DyLight 405, DyLight 488, DyLight 550, DyLight 650, DyLight 680, DyLight 755, DyLight 800, FITC, 퍼시픽 블루, PerCP, 로다민, 텍사스 레드, Cy5, Cy5.5, 및 Cy7이다.

[0062] 본원에는 핵산 하이브리드화 기반 결합을 포함하는 바이오암호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 친화성 탭을 포함하는 핵산 프로브가 사용될 수 있다. 일부 경우에, 친화성 탭은 핵산 서열이 풀다운 되도록 한다. 예컨대, 친화성 탭 바이오틴은 핵산 서열에 상보적인 핵산 프로브에 컨주게이션되고 스트레브아비딘을 사용하여 풀다운된다. 일부 경우에, 친화성 탭은 자기적으로 민감한 물질, 예컨대 자석 또는 자기적으로 민감한 금속을 포함한다. 일부 경우에, 핵산 서열은 고체 지지체, 예컨대 스트렙트아비딘을 사용하여 풀다운되고 고체 지지체 상에 고정된다. 일부 경우에, 핵산 서열은 용액 중에서 예컨대 비드를 통해 풀다운된다. 일부 경우에, 핵산 프로브는 크기에 기반한 배제를 가능하게 한다. 예컨대, 핵산 프로브는 핵산 서열이 크기-기반 고갈에 의해 제거되도록 다른 핵산 서열과 상이한 크기를 갖는 핵산 서열을 초래한다.

[0063] 핵산 하이브리드화 기반 결합을 포함하는 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법은 조절된 증폭을 포함할 수 있다. 일부 경우에, 핵산 하이브리드화 기반 결합 전략은 조절된 증폭에 관한 것으로서, 합성되는 복수의 올리고뉴클레오티드는 전방향 프라이머가 결합하기 위한 유사한 영역을 갖지만, 역방향 프라이머 영역은 쉽게 식별될 수 없다. 이러한 경우에, 미리결정된 역방향 프라이머가 필요할 것이다. 제1의 예시적인 워크플로우에서, 상이한 합성된 올리고뉴클레오티드 각각에 결합하는 미리선택된 영역을 갖는 역방향 프라이머의 풀이 생성되며, 다운스트림 프로세싱, 예컨대 추가의 증폭 또는 DNA 시퀀싱 반응을 위한 올리고뉴클레오티드를 증폭시키는 연장 증폭 반응(예컨대, DNA 폴리머라제를 사용)에 사용된다. 임의적으로, 역방향 프라이머 각각은 연장 증폭 반응(예컨대, DNA 폴리머라제를 사용)에 의해 범용 역방향 프라이머 결합 부위를 통합하기 위한 공통 서열을 포함하는 어댑터 영역을 포함한다. 이러한 배열에서, 복수의 올리고뉴클레오티드를 증폭 또는 시퀀싱하는 데 단지 단일 정방향 또는 역방향 프라이머가 필요하므로, 다운스트림 프로세싱은 단순화된다. 제2의 예시적인 워크플로우에서, 복수의 올리고뉴클레오티드가 합성되며, 각각은, 다양하지만, 합성되는 올리고뉴클레오티드 각각의 5' 및 3' 영역 중 하나 또는 둘 다에 대한 공통 프라이머를 사용하여 복수의 올리고뉴클레오티드의 다운스트림 프로세싱(예컨대, 증폭 또는 시퀀싱 반응)을 가능하게 하도록 공통 프라이머에 대해 충분한 하이브리드화 능력을 갖는, 하이브리드화 모티프를 포함하는 1 또는 2개의 영역을 갖는다. 일부 경우에, 올리고뉴클레오티드 집단은 공통 프라이머의 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 이를 초과하는 핵염기에 하이브리드화되도록 디자인된다.

[0064] 본원에는 전자기 복사(EMR)의 사용을 포함하는 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 일부 경우에, 전자기 복사는 핵산 서열의 절단 또는 이미지 캡처-기반 검출을 가능하게 한다. 일부 경우에, EMR은 약 100 nm 내지 약 400 nm, 약 100 nm 내지 약 300 nm, 또는 약 100 nm 내지 약 200 nm의 파장으로 표면을 향해 적용된다. 일부 경우에, EMR은 0.01 nm 미만의 파장으로 표면을 향해 적용된다. 일부 경우에, EMR은 약 10 nm 내지 약 400 nm, 약 400 nm 내지 약 700 nm, 또는 약 700 nm 내지 약 100,000 nm의 파장으로 표면을 향해 적용된다. 예컨대, EMR은 자외선(UV) 파장 또는 깊은 UV 파장에서 적용된다. 일부 경우에, 깊은 UV 광은 표면으로부터 결합된 체제를 절단하기 위해 약 172 nm의 파장으로 표면에 적용된다. 일부 경우에, EMR은 크세논 램프로 적용된다. 노출 거리는 램프와 표면 사이의 측정값이다. 일부 경우에, 노출 거리는 약 0.1 내지 5 cm이다. 일부 경우에, 노출 거리는 약 0.5 내지 2 cm이다. 일부 경우에, 노출 거리는 약 0.5, 1, 2, 3, 4 또는 5 cm이다. 일부 경우에, EMR은 레이저로 적용된다. 예시적인 레이저 및 그들의 파장은 Ar₂(126 nm), Kr₂(146 nm), F₂(157 nm), Xe₂(172 및 175 nm), ArF(193 nm)를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 핵산 서열은 특정 부위에서 광절단 가능한 핵염기를 포함한다. 일부 경우에, 핵산 서열은 광절단 가능한 변형된 핵염기를 포함한다. 일부 경우에, 핵산 서열은 특정 광 파장의 적용에 의해 광절단 가능하다. 일부 경우에, 핵산 서열은 다수의 광 파장의 적용에 의해 광절단이 가능하다.

[0065] 본원에는 화학적 분해의 이용을 포함하는 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법이 제공된다. 일부 경우에, 핵산 서열은 특정 부위에서 화학적으로 절단 가능한 핵염기를 포함한다. 일

부 경우에, 핵산 서열은 화학적으로 절단 가능한 변형된 핵염기를 포함한다. 일부 경우에, 변형된 핵염기는 화학적으로 절단 가능한 변형을 포함한다. 일부 경우에, 화학적 분해는 아민 시약을 사용하여 수행된다. 일부 경우에, 아민 시약은 액체, 기체, 수성 시약 또는 무수 시약이다. 아민 시약의 비제한적 예는 수산화암모늄, 암모니아 기체, C₁-C₆ 알킬아민 또는 메틸아민이다.

[0066] 본원에 기재된 바이오암호화를 위한 다바이스, 조성물, 시스템 및 방법은 디지털 서열을 핵산 서열로 전환하는 것을 포함할 수 있다. 일부 경우에, 핵산 서열은 DNA 서열이다. 일부 경우에, DNA 서열은 단일 가닥 또는 이중 가닥이다. 일부 경우에, 핵산 서열은 RNA 서열이다. 일부 경우에, RNA 서열은 단일 가닥 또는 이중 가닥이다. 일부 경우에, 핵산 서열은 더 큰 핵산 서열 집단에서 암호화된다. 일부 경우에, 더 큰 핵산 서열 집단은 균일 집단 또는 비균일 집단이다. 일부 경우에, 핵산 서열 집단은 DNA 서열을 포함한다. 일부 경우에, DNA 서열은 단일 가닥 또는 이중 가닥이다. 일부 경우에, 핵산 서열 집단은 RNA 서열이다. 일부 경우에, RNA 서열은 단일 가닥 또는 이중 가닥이다.

[0067] 다수의 핵산 서열이 암호화될 수 있다. 일부 경우에, 암호화되는 핵산 서열의 수는 약 10개 내지 약 1,000,000개 또는 이를 초과하는 서열이다. 일부 경우에, 암호화되는 핵산 서열의 수는 적어도 약 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, 1,000,000개 또는 1,000,000개를 초과하는 서열이다. 일부 경우에, 암호화되는 핵산 서열의 수는 1조개를 초과한다.

[0068] 일부 경우에, 암호화되는 핵산 서열은 적어도 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 300개 또는 300개 초과인 염기 길이를 포함한다. 일부 경우에, 암호화되는 핵산 서열은 10개 염기 내지 25개 염기, 10개 염기 내지 50개 염기, 10개 염기 내지 75개 염기, 10개 염기 내지 100개 염기, 10개 염기 내지 125개 염기, 10개 염기 내지 150개 염기, 10개 염기 내지 175개 염기, 10개 염기 내지 200개 염기, 10개 염기 내지 225개 염기, 10개 염기 내지 250개 염기, 10개 염기 내지 300개 염기, 25개 염기 내지 50개 염기, 25개 염기 내지 75개 염기, 25개 염기 내지 100개 염기, 25개 염기 내지 125개 염기, 25개 염기 내지 150개 염기, 25개 염기 내지 175개 염기, 25개 염기 내지 200개 염기, 25개 염기 내지 225개 염기, 25개 염기 내지 250개 염기, 25개 염기 내지 300개 염기, 50개 염기 내지 75개 염기, 50개 염기 내지 100개 염기, 50개 염기 내지 125개 염기, 50개 염기 내지 150개 염기, 50개 염기 내지 175개 염기, 50개 염기 내지 200개 염기, 50개 염기 내지 225개 염기, 50개 염기 내지 250개 염기, 50개 염기 내지 300개 염기, 75개 염기 내지 100개 염기, 75개 염기 내지 125개 염기, 75개 염기 내지 150개 염기, 75개 염기 내지 175개 염기, 75개 염기 내지 200개 염기, 75개 염기 내지 225개 염기, 75개 염기 내지 250개 염기, 75개 염기 내지 300개 염기, 100개 염기 내지 125개 염기, 100개 염기 내지 150개 염기, 100개 염기 내지 175개 염기, 100개 염기 내지 200개 염기, 100개 염기 내지 225개 염기, 100개 염기 내지 250개 염기, 100개 염기 내지 300개 염기, 125개 염기 내지 150개 염기, 125개 염기 내지 175개 염기, 125개 염기 내지 200개 염기, 125개 염기 내지 225개 염기, 125개 염기 내지 250개 염기, 125개 염기 내지 300개 염기, 150개 염기 내지 175개 염기, 150개 염기 내지 200개 염기, 150개 염기 내지 225개 염기, 150개 염기 내지 250개 염기, 150개 염기 내지 300개 염기, 175개 염기 내지 200개 염기, 175개 염기 내지 225개 염기, 175개 염기 내지 250개 염기, 175개 염기 내지 300개 염기, 200개 염기 내지 225개 염기, 200개 염기 내지 250개 염기, 200개 염기 내지 300개 염기, 225개 염기 내지 250개 염기, 225개 염기 내지 300개 염기, 또는 250개 염기 내지 300개 염기를 포함한다.

[0069] 일부 경우에, 암호화되는 핵산 서열은 관심 핵산 서열의 적어도 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 95% 초과인 농축을 초래한다. 일부 경우에, 암호화되는 핵산 서열은 관심 핵산 서열의 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 95% 초과인 농축을 초래한다.

[0070] 본원에 기재된 바이오암호화 및/또는 바이오복호화를 위한 다바이스, 조성물, 시스템 및 방법은 DNA 또는 RNA 기반 시스템을 포함할 수 있다. 표준 DNA는 이용가능한 4개의 상이한 핵염기: A, T, C 또는 G(아데닌, 티민, 시토신 및 구아닌)를 갖는 염기 4 코딩 시스템이다. 따라서, 이들 4개의 염기는 염기 3(전체 보다 적게 사용) 또는 4 염기 코딩 스킴을 허용한다. 또한, RNA에서 발견되는 우라실(U)의 사용은 제5 염기를 제공하고, 염기 5 코딩 스킴을 허용한다. 또한, 변형된 핵염기는 4보다 큰 핵산 염기 코딩에 사용될 수 있다. 표준 DNA 핵염기가 아닌 핵염기 또는 변형된 핵염기는 우라실, 3-meA(3-메틸아데닌), 히포크산틴, 8-oxoG(7,8-디히드로-8-옥소구아닌), FapyG, FapyA, Tg(티민 글리콜), hoU(히드록시우라실), hmU(히드록시메틸우라실), fU(포르밀우라실), hoC(히드록시시토신), fC(포르밀시토신), 5-meC(5-메틸시토신), 6-meG(6-메틸구아

닌), 7-meG(N7-메틸구아닌), ε C(에테노시토신), 5-caC(5-카복실시토신), 2-hA, ε A(에테노아데닌), 5-fU(5-플루오로우라실), 3-meG(3-메틸구아닌), hmC(히드록시메틸시토신) 및 이소디알루르산을 포함하나, 이들로 제한되지 않는다. 또한, 본원에는 최종적으로 최종 핵산 서열로 전환되기 전에 기계 명령이 2진 서열 형태의 디지털 정보를 중간 코드로의 전환을 가능하게 하는 코딩 스킴이 제공된다.

[0071] 일부 경우에, 데이터를 DNA 서열로 저장하기 위해, 정보는 1 및 0의 2진 코드로부터 DNA의 A, T, G 및 C 염기의 코드로 전환된다. 일부 경우에, 정보 아이템은 먼저 디지털 정보 형태로 코딩된다. 일부 경우에, 디지털 정보의 2진 코드는 코드가 나타내는 정보는 보존하면서 생체분자-기반(예컨대, DNA-기반) 코드로 전환된다. 이 전환된 코드(디지털 2진 코드 대 생체분자 코드)는 본원에 기재된 표면 상에 본원에 기재된 생체분자의 침착과 관련하여 "미리결정된" 서열을 초래하는 것으로 본원에서 언급된다. 미리결정된 서열은 복수의 올리고뉴클레오타이드에 대한 서열을 코딩할 수 있다.

[0072] 이전 코드 전환

[0073] 일반적으로, 개시 코드는 디지털 정보, 전형적으로 컴퓨터에 의해 사용되는 2진 코드 형태의 디지털 정보이다. 일반적인 목적의 컴퓨터는 숫자 "0" 및 "1"에 의해 표시되는 "온" 또는 "오프" 상태를 판독하는 전자 디바이스이다. 이 2진 코드는 컴퓨터가 여러 유형의 정보 아이템을 판독하기 위한 적용이다. 2진 연산에서, 숫자 2는 10으로 기록된다. 예컨대, "10"은 " $1 \times 2^1 + 0 \times 2^0$ "를 나타낸다. 숫자 "3"은 "11"로 기록되고, " $1 \times 2^1 + 1 \times 2^0$ "를 의미한다. 숫자 "4"는 "100"으로 기록되고, 숫자 "5"는 "101"로, 숫자 "6"은 "110"으로 기록된다. 2진 코드에 대한 미국 표준 코드 II(ASCII)의 예가 알파벳에 대해 표 3에서 제공된다.

표 3

2진 코드에 대한 미국 표준 코드 II(ASCII)

문자	ASCII 코드	2진수	문자	ASCII 코드	2진수	번호	ASCII 코드	2진수
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

[0074]

[0075] 본원에는 제1 코드 형태, 예컨대 2진 서열 형태의 정보를 핵산 서열로 전환하는 방법이 제공된다. 공정은 염기 2 코드(즉, 2진수)로부터 더 높은 염기 코드로의 직접 전환을 포함할 수 있다. 예시적인 염기 코드는 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 이의 초과를 포함한다. 표 4는 다양한 염기 넘버링 스킴 사이의 예시적인 정렬을 예시

한다. 전환에 대한 기계 명령을 수신하는 컴퓨터는 서열 정보를 한 코드로부터 다른 코드로 자동적으로 전환할 수 있다.

표 4

염기 넘버링 스킴의 정렬

10 진수	0	1	2	3	4	5	6	7	8	9
4 진수	0	1	2	3	10	11	12	13	20	21
8 진수	0	1	2	3	4	5	6	7	10	11
3 진수	0	1	2	10	11	12	20	21	22	100
2 진수	0	1	10	11	100	101	110	111	1000	1001

핵산 서열

본원에는 핵산 서열이 정보 아이템의 적어도 일부분을 코딩하도록 본원에 기재된 올리고뉴클레오타이드에 대한 서열을 디자인하는 방법이 제공된다. 일부 경우에, 각각의 올리고뉴클레오타이드 서열은 후속 어셈블리 단계 동안 서열 정렬을 용이하게 하고, 또한 에러 보정을 위한 수단을 제공하는 디자인 피처를 갖는다. 일부 배열에서, 올리고뉴클레오타이드 서열은 각각의 올리고뉴클레오타이드 서열과 집단 내의 다른 올리고뉴클레오타이드 서열 사이에 중첩이 존재하도록 디자인된다. 일부 경우에, 각각의 올리고뉴클레오타이드 서열은 단지 하나의 다른 올리고뉴클레오타이드 서열의 일부와 중첩된다(도 5A). 대안적인 배열에서, 각각의 올리고뉴클레오타이드 서열 영역은 단일 올리고뉴클레오타이드 내의 각각의 서열에 대해 2개의 카피가 생성되도록 2개의 서열과 중첩된다(도 5B). 또 다른 배열에서, 각각의 올리고뉴클레오타이드 서열 영역은 단일 폴리뉴클레오타이드 내의 각각의 서열에 대해 3개의 카피가 생성되도록 2개 초과 서열과 중첩된다(도 5C). 본원에 기재된 올리고뉴클레오타이드에 대한 서열은 10-2000, 10-500, 30-300, 50-250 또는 75-200개 염기 길이를 코딩할 수 있다. 일부 경우에, 각각의 올리고뉴클레오타이드 서열은 적어도 10, 15, 20, 25, 30, 50, 100, 150, 200, 500개 또는 이를 초과하는 염기 길이이다.

본원에는 본원에 기재된 각각의 올리고뉴클레오타이드 서열이 복수의 코딩 영역 및 복수의 비-코딩 영역을 포함하도록 디자인되는 방법, 시스템 및 조성이 제공된다(도 6A). 이러한 배열에서, 각각의 코딩 영역(예컨대, 601, 603, 605)은 정보 아이템의 적어도 일부를 코딩한다. 임의적으로, 동일한 올리고뉴클레오타이드 내의 각각의 코딩 영역은 동일한 정보 아이템으로부터의 서열을 코딩하고, 중첩 스킴이 본원에 기재된 바와 같이 임의적으로 사용된다(도 6B). 추가의 경우에서, 동일한 올리고뉴클레오타이드 내의 각각의 코딩 영역은 동일한 서열을 코딩한다(도 6C). 본원에 기재된 올리고뉴클레오타이드에 대한 서열은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 이를 초과하는 코딩 영역을 코딩할 수 있다. 본원에 기재된 올리고뉴클레오타이드에 대한 서열은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 이를 초과하는 동일한 코딩 영역을 코딩할 수 있다. 일부 경우에, 다수의 코딩 영역 각각은 10-1000, 20-500, 30-300, 50-250 또는 75-200개 염기 길이이다. 일부 경우에, 다수의 코딩 영역 각각은 적어도 10, 15, 20, 25, 30, 50, 100, 150, 200개 또는 이를 초과하는 염기 길이이다. 일부 경우에, 각각의 올리고뉴클레오타이드는 구조물의 표면(602)에 분자를 연결하는 데터 영역(611)을 포함한다.

다수의 코딩 서열이 동일한 올리고뉴클레오타이드 내에 존재하는 배열에서, 절단 영역(607)은 각각의 코딩 영역 사이에 임의적으로 존재한다. 절단 영역(607)은 각각의 코딩 영역 사이의 접합부에 존재할 수 있거나, 각각의 코딩 영역 사이에 일련의 서열을 갖는 어댑터 영역 내에 존재할 수 있다. 절단 영역(607)은 일단 합성되면, 절단 신호의 적용 후 가닥으로부터 파괴되는 서열 피처를 코딩할 수 있다. 절단 영역(607)은 제한 효소 인식 부위, 감광성이고 전자기 복사의 적용 하에 파괴되는 변경된 핵산(예컨대, >300 nm 파장의 광에 민감성인 염기 민감성 S-피발로일티오에틸(t-Bu-SATE)를 보유하는 올리고데옥시뉴클레오타이드 헤테로중합체), 또는 특정 화합물 질, 예컨대 암모니아 기체의 적용 후 파괴되는 티미딘-숙시닐 헥사미드 CED 포스포라미다이트(ChemGenes의 CLP-2244)의 적용에 민감한 변경된 핵산을 코딩할 수 있다. 특정 절단 스킴을 갖는 서열의 디자인은 합성된 올리고뉴클레오타이드의 시퀀싱으로부터 쉽게 명백하지 않을 수 있기 때문에, 절단 스킴은 합성된 핵산 라이브러리에 의해 코딩되는 서열에 일정 수준의 보안을 부가하는 수단을 제공한다. 본원에 기재된 올리고뉴클레오타이드에 대한 서열은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 이를 초과하는 절단 영역을 코딩할 수 있다. 일부 경우에, 각각의 절단 영역은 1-100, 1-50, 1-20, 1-10, 5-25 또는 5-30개 염기 길이를 코딩한다. 일부 경우에, 각각의 코딩 영역은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,

12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50 100개 또는 이를 초과하는 염기를 코딩한다. 일부 배열에서, 각각의 올리고뉴클레오타이드에 대해, 각각의 코딩 영역은 동일하고, 각각의 코딩 영역 사이의 각각의 절단 영역은 상이하다. 예컨대, 제1 절단 영역(607)은 제2 절단 영역(609)과 상이하다. 일부 배열에서, 표면(602)에 가장 근접한 절단 영역(607)은 다음 원위 절단 영역(607)과 동일하다.

[0081] 바코드는, 바코드가 연관된 올리고뉴클레오타이드의 일부 피처의 확인을 가능하게 하는 전형적으로 공지된 핵산 서열이다. 도 7A-7B는 예시적인 바코드 배열을 제공한다. 도 7A에서, 제1 올리고뉴클레오타이드(701), 제2 올리고뉴클레오타이드(703) 및 제3 올리고뉴클레오타이드(705)에 대한 각각의 코딩 영역은(표면(702)으로부터 외측으로) 하기 피처: 테더 영역(702), 절단 영역(707), 제1 프라이머 결합 영역(701), 바코드 영역(703), 코딩 영역(701, 703, 705), 및 제2 프라이머 결합 영역(704)을 갖는다. 올리고뉴클레오타이드는 제1 및/또는 제2 프라이머 결합 영역을 인식하는 프라이머를 사용하여 증폭될 수 있다. 증폭은 표면에 부착되거나 표면으로부터 방출되는(즉, 절단 영역(707)에서 절단에 의해) 올리고뉴클레오타이드에 대해 발생할 수 있다. 시퀀싱 후, 바코드 영역(703)은 코딩 영역과 연관된 특징을 확인하기 위한 표시자를 제공한다. 일부 경우에, 바코드는 표적 올리고뉴클레오타이드에 결합되는 경우 표적 올리고뉴클레오타이드가 유도되는 샘플의 확인자로서 기능하는 핵산 서열을 포함한다. 바코드는 충분한 정도의 확인을 가능하게 하는 적합한 길이, 예컨대 적어도 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, 55개, 또는 이를 초과하는 염기 길이로 디자인될 수 있다. 다수의 바코드, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 이를 초과하는 바코드가, 임의적으로 비-바코드 서열에 의해 분리되는, 동일한 분자에 사용될 수 있다. 일부 경우에, 바코드는 10, 9, 8, 7, 6, 5 또는 4개 염기 길이보다 더 짧다. 일부 경우에, 일부 폴리뉴클레오타이드와 연관된 바코드는 다른 폴리뉴클레오타이드와 연관된 바코드와는 길이가 상이하다. 일반적으로, 바코드는 길이가 충분하고, 서열들이 연관되는 바코드에 기초하여 샘플의 확인을 가능하게 하기에 충분히 상이한 서열을 포함한다. 일부 배열에서, 바코드 및 이와 연관된 샘플 공급원은 바코드 서열 내의 하나 이상의 염기의 돌연변이, 삽입 또는 결실, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 이를 초과하는 염기의 돌연변이, 삽입 또는 결실 후에 정확하게 확인될 수 있다. 일부 경우에, 복수의 바코드 중의 각각의 바코드는 적어도 3개의 염기 위치, 예컨대 적어도 3, 4, 5, 6, 7, 8, 9, 10개, 또는 이를 초과하는 위치에 의해 복수의 바코드 내의 모든 다른 바코드와 상이하다. 본원에 제공된 배열은 디지털 서열의 특정 영역의 서열을 코딩하는 핵산 서열을 나타내는 바코드 서열을 포함할 수 있다. 예컨대, 바코드 서열은, 큰 파일에서 특정 올리고뉴클레오타이드 서열이 코딩되는 위치를 나타낼 수 있다. 일부 경우에, 바코드 서열은 특정 올리고뉴클레오타이드 서열이 연관된 파일을 나타낼 수 있다. 일부 경우에, 바코드 서열은 특정 서열에 대한 전환 스킴과 연관된 정보를 포함하여 추가의 보안 층을 제공한다.

[0082] 본원에는 올리고뉴클레오타이드 서열 내의 각각의 올리고뉴클레오타이드가 그 집단 내의 올리고뉴클레오타이드 서열 중에서 공통인 적어도 하나의 영역을 갖도록 디자인되는 올리고뉴클레오타이드 서열 디자인 스킴이 제공된다. 예컨대, 동일한 집단 내의 모든 올리고뉴클레오타이드는 하나 이상의 프라이머 영역을 포함할 수 있다. 서열-특이적인 프라이머 영역의 디자인은 다수의 올리고뉴클레오타이드의 큰 라이브러리로부터 선택된 배치에서 증폭될 올리고뉴클레오타이드의 선택을 가능하게 한다. 각각의 올리고뉴클레오타이드 서열은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 이를 초과하는 프라이머 결합 서열을 포함할 수 있다. 올리고뉴클레오타이드 서열의 집단은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25, 50, 100, 200, 500, 1000, 5000, 10000, 50000, 100000개 또는 이를 초과하는 동일하지 않은 결합 서열을 포함할 수 있다. 프라이머 결합 서열은 5-100, 10-75, 7-60, 8-60, 10-50 또는 10-40개 염기 길이를 포함할 수 있다.

[0083] **올리고뉴클레오타이드 합성을 위한 구조물**

[0084] 본원에는 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위해 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 강성 또는 가요성 구조물이 제공된다. 강성 구조물의 경우, 본원에는 올리고뉴클레오타이드의 라이브러리 생성을 위한 구조물(예컨대, 플레이트)을 갖는 디바이스가 제공된다. 예시적인 구조물(800)이 도 8에 도시되며, 구조물(800)은 표준 96웰 플레이트와 대략 동일한 크기의 치수를 갖는다: 140 mm x 90 mm. 구조물(800)은 24개의 영역 또는 서브필드(805) 내에 그룹화된 클러스터를 포함하며, 각각의 서브필드(805)는 256개의 클러스터(810)의 어레이를 포함한다. 예시적인 서브필드(805)의 확대도가 도 9에 도시된다. 구조물은 도 8 및 도 9에 도시되는 바와 같이 실질적으로 편평할 수 있다. 4개의 클러스터의 확대도(도 9)에서, 단일 클러스터(910)는 1079.210 um 또는 1142.694 um um의 Y축 클러스터 피치(인접한 클러스터의 중심으로부터 중심까지의 거리), 및 1125 um의 X축 클러스터 피치를 갖는다. 예시적인 클러스터(1010)가 도 10에 도시되며, 여기서 Y축 로커스 피치(인접한 로커스의 중심으로부터 중심까지의 거리)는

63.483 μm 이고, X축 로커스 피치는 75 μm 이다. 가장 긴 부분에서의 로커스 너비, 예컨대 원형 로커스의 경우 직경은 50 μm 이고, 로커스 사이의 거리는 24 μm 이다. 도 10의 예시적인 클러스터에서 로커스(1005)의 수는 121이다. 로커스(또한 "피쳐"로 지칭됨)는 플랫폼, 웰 또는 채널일 수 있다. 예시적인 채널 배열이 도 11A-11B에 도시되며, 여기서 주 채널(1110) 및 주 채널(1110)에 연결된 복수의 채널(1115)을 포함하는 플레이트(1105)가 도시된다. 주 채널(1110)과 복수의 채널(1115) 사이의 연결은 주 채널(1110)로부터 각각의 복수의 채널(1115)까지의 유로를 위한 유체 소통을 가능하게 한다. 본원에 기재된 플레이트(1105)는 다수의 주 채널(1110)을 포함할 수 있다. 복수의 채널(1115)은 주 채널(1110) 내에 클러스터를 집합적으로 형성한다.

[0085] 가요성 구조물의 경우, 본원에는 가요성 구조물이 하나 이상의 고정 구조물, 예컨대 한 쌍의 롤러(1203) 주위를 둘러싸는 연속 루프(1201) 또는 별개의 고정 구조물, 예컨대 한 쌍의 릴(1205) 주위를 둘러싸는 불연속 가요성 구조물(1207)을 포함하는 디바이스가 제공된다. 도 12A-12B를 참조한다. 본원에는 올리고뉴클레오타이드 연장을 위한 복수의 피쳐(로커스)를 갖는 표면을 갖는 가요성 구조물이 제공된다. 가요성 구조물(1301)의 일부에서 각각의 피쳐는 실질적으로 편평한 피쳐(1303)(예컨대, 플랫폼), 채널(1305) 또는 웰(1307)일 수 있다. 도 13A-13C를 참조한다. 하나의 예시적인 배열에서, 구조물의 각각의 피쳐는 너비는 약 10 μm 이고, 각각의 구조물의 중심 사이의 거리는 약 21 μm 이다. 도 14A를 참조한다. 피쳐는 원형, 직사각형, 끝이 가늘어지는(tapered), 또는 둥근 형상을 포함할 수 있으나, 이들로 제한되지 않는다.

[0086] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 채널을 포함할 수 있다. 일부 경우에, 본원에 기재된 채널은 1 대 0.01의 너비 대 깊이(또는 높이) 비를 가지며, 여기서 너비는 마이크로채널의 가장 좁은 세그먼트에서의 너비 측정치이다. 일부 경우에, 본원에 기재된 채널은 0.5 대 0.01의 너비 대 깊이(또는 높이) 비를 가지며, 여기서 너비는 마이크로채널의 가장 좁은 세그먼트에서의 너비 측정치이다. 일부 경우에, 본원에 기재된 채널은 약 0.01, 0.05, 0.1, 0.15, 0.16, 0.2, 0.5 또는 1의 너비 대 깊이(또는 높이) 비를 갖는다.

[0087] 본원에는 폴리고뉴클레오타이드 합성을 위한 복수의 구별되는 로커스, 채널 또는 웰을 포함하는 구조물이 제공된다. 본원에 기재된 구조물은 복수의 클러스터를 포함할 수 있으며, 각각의 클러스터는 복수의 웰, 로커스 또는 채널을 포함한다. 대안적으로, 본원에는 웰, 로커스 또는 채널의 균일 배열을 포함할 수 있는 구조물이 기재된다. 일부 경우에, 본원에 기재된 구조물은 클러스터 내에 복수의 피쳐(로커스)에 대응하는 복수의 채널을 포함하며, 여기서 채널의 높이 또는 깊이는 약 5 μm 내지 약 500 μm , 약 5 μm 내지 약 400 μm , 약 5 μm 내지 약 300 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm 또는 약 10 μm 내지 약 50 μm 이다. 일부 경우에, 채널의 높이 또는 깊이는 100 μm 미만, 80 μm 미만, 60 μm 미만, 40 μm 미만 또는 20 μm 미만이다. 일부 경우에, 채널의 높이 또는 깊이는 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 μm 또는 이를 초과한다. 일부 경우에, 채널의 높이 또는 깊이는 적어도 10, 25, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 nm, 또는 1000 nm를 초과한다. 일부 경우에, 채널의 높이 또는 깊이는 약 10 nm 내지 약 1000 nm, 약 25 nm 내지 약 900 nm, 약 50 nm 내지 약 800 nm, 약 75 nm 내지 약 700 nm, 약 100 nm 내지 약 600 nm 또는 약 200 nm 내지 약 500 nm의 범위이다. 일부 경우에, 높이 또는 깊이는 약 50 nm 내지 약 1 μm 의 범위이다.

[0088] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 피쳐를 포함할 수 있다. 일부 경우에, 피쳐(예컨대, 실질적으로 편평한 피쳐, 웰, 채널, 로커스 또는 돌출물)의 너비는 약 0.1 μm 내지 약 500 μm , 약 0.5 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 100 μm , 또는 약 0.1 μm 내지 약 100 μm , 예컨대 약 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm , 10 μm , 5 μm , 1 μm 또는 0.5 μm 이다. 일부 경우에, 피쳐(예컨대, 마이크로채널)의 너비는 약 100 μm , 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 미만이다. 일부 경우에, 피쳐의 너비는 적어도 110, 25, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 nm, 또는 1000 nm 초과이다. 일부 경우에, 피쳐의 너비는 약 10 nm 내지 약 1000 nm, 약 25 nm 내지 약 900 nm, 약 50 nm 내지 약 800 nm, 약 75 nm 내지 약 700 nm, 약 100 nm 내지 약 600 nm, 또는 약 200 nm 내지 약 500 nm의 범위이다. 일부 경우에, 피쳐의 너비는 약 50 nm 내지 약 1000 nm의 범위이다. 일부 경우에, 2개의 인접한 피쳐의 중심 사이의 거리는 약 0.1 μm 내지 약 500 μm , 0.5 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm , 또는 약 5 μm 내지 약 30 μm , 예컨대 약 20 μm 이다. 일부 경우에, 피쳐의 총 너비는 약 5 μm , 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm , 또는 100 μm 이다. 일부 경우에, 피쳐의 총 너비는 약 1 μm 내지 100 μm , 30 μm 내지 100 μm , 또는 50 μm 내지 70 μm 이다. 일부

경우에, 2개의 인접한 피처의 중심 사이의 거리는 약 0.5 μm 내지 약 2 μm , 약 0.5 μm 내지 약 2 μm , 약 0.75 μm 내지 약 2 μm , 약 1 μm 내지 약 2 μm , 약 0.2 μm 내지 약 1 μm , 약 0.5 μm 내지 약 1.5 μm , 약 0.5 μm 내지 약 0.8 μm , 또는 약 0.5 μm 내지 약 1 μm , 예컨대, 약 1 μm 이다. 일부 경우에, 피처의 총 너비는 약 50 nm, 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 , 또는 1.5 μm 이다. 일부 경우에, 피처의 총 너비는 약 0.5 μm 내지 2 μm , 0.75 μm 내지 1 μm , 또는 0.9 μm 내지 2 μm 이다.

[0089] 일부 경우에, 각각의 피처는 또 다른 피처에서 성장하는 올리고뉴클레오티드의 집단과는 상이한 서열을 갖는 올리고뉴클레오티드의 집단의 합성을 지지한다. 본원에서 적어도 10, 100, 256, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000개 또는 이를 초과하는 클러스터를 포함하는 표면이 제공된다. 본원에는 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; 또는 10,000,000개 초과 또는 이를 초과하는 구별되는 로커스를 포함하는 표면이 제공된다. 일부 경우에, 각각의 클러스터는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 500개 또는 이를 초과하는 피처를 포함한다. 일부 경우에, 각각의 클러스터는 50 내지 500, 50 내지 200, 50 내지 150 또는 100 내지 150개 피처를 포함한다. 일부 경우에, 각각의 클러스터는 100 내지 150개 피처를 포함한다. 예시적인 배열에서, 각각의 클러스터는 109, 121, 130 또는 137개 피처를 포함한다.

[0090] 본원에는 가장 긴 세그먼트에서의 너비가 5 내지 100 μm 인 피처가 제공된다. 일부 경우에, 피처는 가장 긴 세그먼트에서 약 30, 35, 40, 45, 50, 55 또는 60 μm 의 너비를 갖는다. 일부 경우에, 피처는 다수의 세그먼트를 갖는 채널이고, 여기서 각각의 세그먼트는 5 내지 50 μm 이격된 중심 대 중심 거리를 갖는다. 일부 경우에, 각각의 세그먼트의 이격된 중심 대 중심 거리는 약 5, 10, 15, 20 또는 25 μm 이다.

[0091] 일부 경우에, 본원에 기재된 구조물의 표면 상에서 합성되는 구별되는 올리고뉴클레오티드의 수는 기관에서 이용가능한 구별되는 피처의 수에 의존한다. 일부 경우에, 기관의 클러스터 내 피처의 밀도는 적어도 또는 약 1개 피처/ mm^2 , 10개 피처/ mm^2 , 25개 피처/ mm^2 , 50개 피처/ mm^2 , 65개 피처/ mm^2 , 75개 피처/ mm^2 , 100개 피처/ mm^2 , 130개 피처/ mm^2 , 150개 피처/ mm^2 , 175개 피처/ mm^2 , 200개 피처/ mm^2 , 300개 피처/ mm^2 , 400개 피처/ mm^2 , 500개 피처/ mm^2 , 1,000개 피처/ mm^2 또는 이를 초과한다. 일부 경우에, 기관은 약 10개 피처/ mm^2 내지 약 500 mm^2 , 약 25개 피처/ mm^2 내지 약 400 mm^2 , 약 50개 피처/ mm^2 내지 약 500 mm^2 , 약 100개 피처/ mm^2 내지 약 500 mm^2 , 약 150개 피처/ mm^2 내지 약 500 mm^2 , 약 10개 피처/ mm^2 내지 약 250 mm^2 , 약 50개 피처/ mm^2 내지 약 250 mm^2 , 약 10개 피처/ mm^2 내지 약 200 mm^2 , 또는 약 50개 피처/ mm^2 내지 약 200 mm^2 를 포함한다. 일부 경우에, 클러스터 내의 2개의 인접한 피처의 중심 사이의 거리는 약 10 μm 내지 약 500 μm , 약 10 μm 내지 약 200 μm 또는 약 10 μm 내지 약 100 μm 이다. 일부 경우에, 인접한 피처의 2개의 중심 사이의 거리는 약 10 μm , 20 μm , 30 μm , 40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm 또는 100 μm 보다 크다. 일부 경우에, 2개의 인접한 피처의 중심 사이의 거리는 약 200 μm , 150 μm , 100 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 미만이다. 일부 경우에, 2개의 인접한 피처의 중심 사이의 거리는 약 10000 nm, 8000 nm, 6000 nm, 4000 nm, 2000 nm, 1000 nm, 800 nm, 600 nm, 400 nm, 200 nm, 150 nm, 100 nm, 80 nm, 70 nm, 60 nm, 50 nm, 40 nm, 30 nm, 20 nm 또는 10 nm 미만이다. 일부 경우에, 본원에 기재된 구조물의 각각의 제곱 미터는 적어도 약 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} 개 피처를 허용하며, 여기서 각각의 피처는 하나의 올리고뉴클레오티드를 지지한다. 일부 경우에, 10^9 개 올리고뉴클레오티드는 본원에 기재된 구조물의 약 6, 5, 4, 3, 2 또는 1 m^2 미만 상에서 지지된다.

[0092] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오티드 합성을 위한 구조물은 다수의 올리고뉴클레오티드의 합성을 지지한다. 일부 경우에, 본원에 기재된 구조물은 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; 10,000,000개 초과 또는 이를 초과하는 동일하지 않은 올리고뉴클레오티드의 합성을 위한 지지체를 제공한다. 일부 경우에, 구조물은 구별되는 서열을 코딩하는 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;

10,000,000개 초과 또는 이를 초과하는 올리고뉴클레오타이드의 합성을 위한 지지체를 제공한다. 일부 경우에, 올리고뉴클레오타이드의 적어도 일부는 동일한 서열을 갖거나 동일한 서열로 합성되도록 구성된다. 일부 경우에, 구조물은 적어도 약 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, 500개 염기 또는 이를 초과하는 염기를 갖는 올리고뉴클레오타이드의 성장을 위한 표면 환경을 제공한다.

[0093] 일부 경우에, 올리고뉴클레오타이드는 구조물의 구별되는 피쳐 상에서 합성되며, 여기서 각각의 피쳐는 올리고뉴클레오타이드 집단의 합성을 지지한다. 일부 경우에, 각각의 피쳐는 또 다른 로커스에서 성장되는 올리고뉴클레오타이드 집단과는 상이한 서열을 갖는 올리고뉴클레오타이드 집단의 합성을 지지한다. 일부 경우에, 구조물의 피쳐는 복수의 클러스터 내에 위치된다. 일부 경우에, 구조물은 적어도 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000개 또는 이를 초과하는 클러스터를 포함한다. 일부 경우에, 구조물은 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; 또는 10,000,000개 초과 또는 이를 초과하는 구별되는 피쳐를 포함한다. 일부 경우에, 각각의 클러스터는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150개 또는 이를 초과하는 피쳐(로커스)를 포함한다. 일부 경우에, 각각의 클러스터는 50 내지 500, 100 내지 150, 또는 100 내지 200개 피쳐를 포함한다. 일부 경우에, 각각의 클러스터는 109, 121, 130 또는 137개 피쳐를 포함한다. 일부 경우에, 각각의 클러스터는 5, 6, 7, 8, 9, 10, 11 또는 12개 피쳐를 포함한다. 일부 경우에, 하나의 클러스터 내의 구별되는 위치로부터의 올리고뉴클레오타이드는, 어셈블링되는 경우, 미리결정된 서열의 인접하는 더 긴 올리고뉴클레오타이드를 코딩하는 서열을 갖는다.

[0094] **구조물 크기**

[0095] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 다양한 크기를 포함한다. 일부 경우에, 본원에 기재된 구조물은 약 표준 96웰 플레이트의 크기, 예컨대 약 100 내지 200 mm x 약 50 내지 150 mm이다. 일부 경우에, 본원에 기재된 구조물은 약 1000 mm, 500 mm, 450 mm, 400 mm, 300 mm, 250 mm, 200 mm, 150 mm, 100 mm 또는 50 mm 미만 또는 이와 동일한 직경을 갖는다. 일부 경우에, 기관의 직경은 약 25 mm 내지 1000 mm, 약 25 mm 내지 약 800 mm, 약 25 mm 내지 약 600 mm, 약 25 mm 내지 약 500 mm, 약 25 mm 내지 약 400 mm, 약 25 mm 내지 약 300 mm, 또는 약 25 mm 내지 약 200이다. 기관 크기의 비제한적인 예는 약 300 mm, 200 mm, 150 mm, 130 mm, 100 mm, 76 mm, 51 mm 및 25 mm를 포함한다. 일부 경우에, 기관은 적어도 약 100 mm²; 200 mm²; 500 mm²; 1,000 mm²; 2,000 mm²; 5,000 mm²; 10,000 mm²; 12,000 mm²; 15,000 mm²; 20,000 mm²; 30,000 mm²; 40,000 mm²; 50,000 mm² 또는 이를 초과하는 평면 표면적을 갖는다. 일부 경우에, 기관의 두께는 약 50 mm 내지 약 2000 mm, 약 50 mm 내지 약 1000 mm, 약 100 mm 내지 약 1000 mm, 약 200 mm 내지 약 1000 mm, 또는 약 250 mm 내지 약 1000 mm이다. 기관의 두께의 비제한적인 예는 275 mm, 375 mm, 525 mm, 625 mm, 675 mm, 725 mm, 775 mm 및 925 mm를 포함한다. 일부 경우에, 기관의 두께는 직경에 따라 달라지고, 기관의 조성에 의존적이다. 예컨대, 규소 이외의 다른 물질을 포함하는 구조물은 동일한 직경의 규소 구조물과는 상이한 두께를 가질 수 있다. 구조물 두께는 사용되는 물질의 기계적 강도에 의해 결정될 수 있고, 구조물은 취급 중에 균열되지 않고 자신의 무게를 지탱하기에 충분히 두꺼워야 한다. 일부 경우에, 구조물은 어떠한 한 차원에서도 약 1, 2, 3, 4, 5, 10, 15, 30, 40, 50 피트를 초과한다.

[0096] **물질**

[0097] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 다양한 물질로 제작될 수 있다. 특정 경우에, 본 개시의 기관/고체 지지체가 제작되는 물질은 낮은 수준의 올리고뉴클레오타이드 결합을 나타낸다. 일부 상황에서, 가시광 및/또는 UV 광에 투명한 물질(들)이 사용될 수 있다. 충분히 전도성인 물질, 예컨대 본원에 기재된 기관/고형 지지체의 전부 또는 부분을 가로질러 균일한 전기장을 형성할 수 있는 물질이 사용될 수 있다. 일부 경우에, 이러한 물질은 전기 접지에 연결될 수 있다. 일부 경우에, 기관 또는 고형 지지체는 열전도성이거나 단열될 수 있다. 물질은 화학적 또는 생화학적 반응, 예컨대 일련의 올리고뉴클레오타이드 합성 반응을 지지하도록 내화학

성 및 내열성일 수 있다. 가요성 물질의 경우, 관심 물질은 변형된 나일론, 변형되지 않은 나일론, 니트로셀룰로스, 폴리프로필렌 등을 포함할 수 있다.

[0098] 강성 물질의 경우, 관심의 특정 물질은 유리; 융합 실리카; 규소, 플라스틱(예컨대, 폴리테트라플루오로에틸렌, 폴리프로필렌, 폴리스티렌, 폴리카보네이트 및 이의 블렌드 등); 금속(예컨대, 금, 백금 등)을 포함한다. 구조물은 규소, 폴리스티렌, 아가로스, 텍스트란, 셀룰로스 중합체, 폴리아크릴아미드, 폴리디메틸실록산(PDMS) 및 유리로 이루어진 군으로부터 선택되는 물질로 제작될 수 있다. 기관/고형 지지체 또는 마이크로구조물, 그 내의 리액터는 본원에 열거된 물질의 조합 또는 당해 기술분야에 공지된 임의의 다른 적합한 물질로 제작될 수 있다.

[0099] 본원에서 용어 "가요성"은 구부러지거나, 접거나 또는 파괴 없이 유사하게 조작될 수 있는 구조물을 지칭하기 위해 사용된다. 일부 경우에, 가요성 구조물은 롤러 주위에서 적어도 30도 구부러진다. 일부 경우에, 가요성 구조물은 롤러 주위에서 적어도 180도 구부러진다. 일부 경우에, 가요성 구조물은 롤러 주위에서 적어도 270도 구부러진다. 일부 경우에, 가요성 구조물은 롤러 주위에서 적어도 360도 구부러진다. 일부 경우에, 롤러는 반경이 약 10 cm, 5 cm, 3 cm, 2 cm 또는 1 cm 미만이다. 일부 경우에, 가요성 구조물은 고장(예컨대, 균열) 또는 20 °C에서 변형 없이 적어도 100회 어느 방향으로든 반복해서 구부러지고 곧게 펴진다. 일부 경우에, 본원에 기재된 가요성 구조물은 회전될 수 있는 두께를 갖는다. 일부 경우에, 본원에 기재된 가요성 구조물의 두께는 약 50 mm, 10 mm, 1 mm 또는 0.5 mm 미만이다.

[0100] 본원에 기재된 구조물에 대한 예시적인 가요성 물질은 나일론(변형되지 않은 나일론, 변형된 나일론, 투명 나일론), 니트로셀룰로스, 폴리프로필렌, 폴리카보네이트, 폴리에틸렌, 폴리우레탄, 폴리스티렌, 아세탈, 아크릴, 아크릴로니트릴, 부타디엔 스티렌(ABS), 폴리에스테르 필름, 예컨대 폴리에틸렌 테레프탈레이트, 폴리메틸 메타크릴레이트 또는 다른 아크릴, 폴리비닐 클로라이드 또는 다른 비닐 수지, 투명 PVC 호일, 프린터용 투명 호일, 폴리(메틸 메타크릴레이트)(PMMA), 메타크릴레이트 공중합체, 스티렌 중합체, 높은 굴절 지수 중합체, 불소 함유 중합체, 폴리에테르설폰, 폴리이미드 함유 지환식 구조물, 고무, 식물, 금속 호일, 및 이들의 임의의 조합을 포함하나, 이들로 제한되지 않는다. 다양한 가소제 및 변형제가 선택된 가요성 특징을 달성하기 위해 중합체성 기관 물질과 함께 사용될 수 있다.

[0101] 본원에 기재된 가요성 구조물은 플라스틱 물질을 포함할 수 있다. 일부 경우에, 가요성 구조물은 열가소성 물질을 포함할 수 있다. 열가소성 물질의 비제한적인 예는 아크릴, 아크릴로니트릴 부타디엔 스티렌, 나일론, 폴리락트산, 폴리벤즈이미다졸, 폴리카보네이트, 폴리에테르 설폰, 폴리에테르에테르 케톤, 폴리에테르이미드, 폴리에틸렌, 폴리페닐렌 옥사이드, 폴리페닐렌 설파이드, 폴리프로필렌, 폴리스티렌, 폴리비닐 클로라이드 및 폴리테트라플루오로에틸렌을 포함한다. 일부 경우에, 기관은 폴리아릴에테르케톤(PEAK) 패밀리의 열가소성 물질을 포함한다. PEAK 열가소성 물질의 비제한적인 예는 폴리에테르케톤(PEK), 폴리에테르케톤케톤(PEKK), 폴리(에테르 에테르 케톤 케톤)(PEEKK), 폴리에테르에테르 케톤(PEEK) 및 폴리에테르케톤에테르케톤케톤(PEKEKK)을 포함한다. 일부 경우에, 가요성 구조물은 톨루엔과 상용성인 열가소성 물질을 포함한다. 일부 경우에, 플라스틱 물질의 가요성은 가소제의 첨가에 의해 증가된다. 가소제의 예는 에스테르계 가소제, 예컨대 프탈레이트이다. 프탈레이트 가소제는 비스(2-에틸헥실)프탈레이트(DEHP), 디이소노닐 프탈레이트(DINP), 디-n-부틸 프탈레이트(DnBP, DBP), 부틸 벤질 프탈레이트(BBzP), 디이소데실 프탈레이트(DIDP), 디옥틸 프탈레이트(DOP, DnOP), 디이소옥틸 프탈레이트(DIOP), 디에틸 프탈레이트(DEP), 디이소부틸프탈레이트(DIBP), 및 디-n-헥실 프탈레이트를 포함한다. 일부 경우에, 공중합을 통하거나 또는 중합 이전에 단량체에 비-반응성 측쇄의 첨가를 통한 열가소성 중합체의 변형은 가요성을 증가시킨다.

[0102] 본원에는 플루오로엘라스토머를 추가로 포함할 수 있는 가요성 구조물이 제공된다. 약 80%의 플루오로엘라스토머를 갖는 물질은 FKM으로 지정된다. 플루오로엘라스토머는 퍼플루오로-엘라스토머(FFKM) 및 테트라플루오로에틸렌/프로필렌 고무(FEPM)를 포함한다. 플루오로엘라스토머는 5종의 공지된 유형을 갖는다. 1형 FKM은 비닐리덴 플루오라이드(VDF) 및 헥사플루오로프로필렌(HFP)로 구성되고, 그들의 불소 함량은 전형적으로 약 66 중량%이다. 2형 FKM은 VDF, HFP 및 테트라플루오로에틸렌(TFE)으로 구성되고, 전형적으로 약 68% 내지 69%의 불소를 갖는다. 3형 FKM은 VDF, TFE 및 퍼플루오로메틸비닐에테르(PMVE)로 구성되고, 약 62% 내지 68%의 불소를 갖는다. 4형 FKM은 프로필렌, TFE 및 VDF로 구성되고, 전형적으로 약 67%의 불소를 갖는다. 5형 FKM은 VDF, HFP, TFE, PMVE 및 에틸렌으로 구성된다.

[0103] 일부 경우에, 본원에 개시된 기관은 컴퓨터 판독 가능한 물질을 포함한다. 컴퓨터 판독 가능한 물질은 자기 매체, 릴-투-릴 테이프, 카트리지 테이프, 카세트 테이프, 가요성 디스크, 종이 매체, 필름, 마이크로피시, 연속 테이프(예컨대, 벨트) 및 전자적 명령을 저장하기에 적합한 임의의 매체를 포함하나, 이들로 제한되지 않는다.

일부 경우에, 기관은 자기 릴-투-릴 테이프 또는 자기 벨트를 포함한다. 일부 경우에, 기관은 가요성 인쇄 회로판을 포함한다.

[0104] 본원에 기재된 구조물은 가시광 및/또는 UV 광에 투명할 수 있다. 일부 경우에, 본원에 기재된 구조물은 구조물의 전부 또는 일부를 가로질러 균일한 전기장을 형성하기에 충분히 전도성이다. 일부 경우에, 본원에 기재된 구조물은 열전도성이거나 단열된다. 일부 경우에, 구조물은 화학적 반응, 예컨대 올리고뉴클레오타이드 합성 반응을 지지하도록 내화학적 및 내열성이다. 일부 경우에, 구조물은 자성이다. 일부 경우에, 구조물은 금속 또는 금속 합금을 포함한다.

[0105] 올리고뉴클레오타이드 합성을 위한 구조물은 임의의 차원에서 1, 2, 5, 10, 30, 50 피트 이상 또는 이를 초과하는 길이일 수 있다. 가요성 구조물의 경우, 가요성 구조물은 임의적으로 감긴 상태, 예컨대 릴로 저장된다. 큰, 예컨대 1피트 초과 길이의 강성 구조물의 경우, 강성 구조물은 수직으로 또는 수평으로 저장될 수 있다.

[0106] 구조물 표면 상의 암호화 키 마킹

[0107] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 암호화 마킹을 포함할 수 있다. 본원에는 인접 올리고뉴클레오타이드 집단과 연관된 정보의 공급원 아이템, 인접 올리고뉴클레오타이드 집단의 서열을 복호화하기 위한 암호화 스킴, 인접 올리고뉴클레오타이드 집단에 대한 카피 수 또는 이들의 임의의 조합과 관련된 정보를 제공하는 마킹(1401)을 갖는 구조물이 제공된다. 예컨대 도 14B-14C를 참조한다. 마킹은 육안으로 볼 수 있거나 현미경으로 확대하여 볼 수 있다. 일부 경우에, 표면 상의 마킹은 단지 마킹을, 예컨대 열, 화학물질 또는 광 처리(예컨대, 마킹을 발광시키기 위한 UV 또는 IR 광)에 노출시키는 처리 조건 후에만 볼 수 있다. 열에 의해 현상되는 예시적인 잉크는 염화코발트(가열되는 경우 청색으로 변함)를 포함하나, 이로 제한되는 것은 아니다. 화학 반응에 의해 현상되는 예시적인 잉크는 페놀프탈레인, 황산구리, 질산납(II), 염화코발트(II), 및 황산망간과 과산화수소에 의해 현상되는 옥살산세륨을 포함하나, 이들로 제한되지 않는다.

[0108] 표면 제조

[0109] 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한 올리고뉴클레오타이드 합성을 위한 구조물은 올리고뉴클레오타이드 합성을 위한 표면을 포함할 수 있다. 본원에는 기관 상에 생체분자의 고정화를 지지하는 방법이 제공되며, 여기서 본원에 기재된 구조물의 표면은 부착을 위한 생체분자와의 커플링 반응을 용이하게 하는 물질을 포함하고/하거나 이 물질로 코팅된다. 생체분자 고정화를 위한 구조물을 제조하기 위해, 기관 표면 또는 표면의 선택된 부위 또는 영역의 하나 이상의 화학적 및/또는 물리적 특성을 변화시키는 부가 또는 공제 공정에 의해 기관 표면을 화학적으로 및/또는 물리적으로 변경시키는 표면 변형이 이용될 수 있다. 예컨대, 표면 변형은 (1) 표면의 습윤 특성을 변화시키는 것, (2) 표면을 작용화, 즉 표면 작용기를 제공하거나, 변경하거나, 치환하는 것, (3) 표면을 탈작용화, 즉 표면 작용기를 제거하는 것, (4) 달리, 예컨대 에칭을 통해 표면의 화학 조성을 변경시키는 것, (5) 표면 조도를 증가 또는 감소시키는 것, (6) 표면 상에 코팅, 예컨대 표면의 습윤 특성과는 상이한 습윤 특성을 나타내는 코팅을 제공하는 것, 및/또는 (7) 표면 상에 미립자를 침착시키는 것을 포함한다. 일부 경우에, 구조물의 표면은 구조물 상에서 2개 이상의 구별되는 영역을 생성하기 위해 선택적으로 작용화되며, 여기서 적어도 하나의 영역은 동일한 구조물의 다른 영역과는 상이한 표면 또는 화학적 특성을 갖는다. 이러한 특성은 표면 에너지, 화학적 말단화, 화학적 모이어티의 표면 농도 등을 포함하나, 이들로 제한되지 않는다.

[0110] 일부 경우에, 본원에 개시된 구조물의 표면은 기관의 표면 및 생체분자 둘 다에 결합하도록 구성된 하나 이상의 능동적으로 작용화된 표면을 포함하도록 변형됨으로써 표면에 대한 커플링 반응을 지지한다. 일부 경우에, 표면은 또한 생체분자를 효율적으로 결합하지 않은 수동적(passive) 물질로 작용화됨으로써 수동적 작용화제가 결합된 부위에서 생체분자의 부착을 방지한다. 일부 경우에, 표면은 생체분자 지지를 위한 구별되는 피처만을 규정하는 활성층을 포함한다.

[0111] 일부 경우에, 표면은 임의의 상이한 비율의 작용화 그룹의 혼합물과 접촉된다. 일부 경우에, 혼합물은 적어도 2, 3, 4, 5개 또는 이를 초과하는 상이한 유형의 작용화제를 포함한다. 일부 경우에, 혼합물 중의 적어도 2개 유형의 표면 작용화제의 비율은 약 1:1, 1:2, 1:5, 1:10, 2:10, 3:10, 4:10, 5:10, 6:10, 7:10, 8:10, 9:10, 또는 2개의 그룹의 목적하는 표면 표현을 달성하기 위한 임의의 다른 비율이다. 일부 경우에, 목적하는 표면 장력, 습윤성, 수 접촉각 및/또는 다른 적합한 용매에 대한 접촉각은 기관 표면에 적합한 비율의 작용화제를 제공함으로써 달성된다. 일부 경우에, 혼합물 중의 작용화제는 적합한 반응성 모이어티 및 불활성 모이어티로부터

선택되어, 반응성 그룹의 표면 밀도를 다운스트림 반응을 위한 목적하는 수준으로 희석한다. 일부 경우에, 작용화 시약의 혼합물은 생체분자에 결합하는 하나 이상의 시약 및 생체분자에 결합하지 않은 하나 이상의 시약을 포함한다. 따라서, 시약의 조성은 구별되는 작용화 구역에서 발생하는 생체분자의 결합량의 조절을 가능하게 한다.

[0112] 일부 경우에, 기관 작용화를 위한 방법은 기관의 표면 상에 실란 분자의 침착을 포함한다. 실란 분자는 기관의 고 에너지 표면 상에 침착될 수 있다. 일부 경우에, 고 표면 에너지 영역은 수동적 작용화 시약을 포함한다. 본원에 기재된 방법은 실란기가 표면에 결합하는 것을 가능하게 하는 한편, 분자의 나머지는 표면으로부터 거리를 제공하고 생체분자가 부착되는 말단에 유리 히드록실기를 제공한다. 일부 경우에, 실란은 유기작용성 알콕시실란 분자이다. 유기작용성 알콕시실란의 비제한적인 예는 디메틸클로로-옥토데실-실란, 메틸디클로로-옥토데실-실란, 트리클로로-옥토데실-실란, 트리메틸-옥토데실-실란 및 트리에틸-옥토데실-실란을 포함한다. 일부 경우에, 실란은 아미노실란이다. 아미노실란의 예는 11-아세톡시운데실트리에톡시실란, n-데실트리에톡시실란, (3-아미노프로필)트리메톡시실란, (3-아미노프로필)트리에톡시실란, 글리시딜옥시프로필/트리메톡시실란 및 N-(3-트리에톡시실릴프로필)-4-히드록시부티르아미드를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 실란은 11-아세톡시운데실트리에톡시실란, n-데실트리에톡시실란, (3-아미노프로필)트리메톡시실란, (3-아미노프로필)트리에톡시실란, 글리시딜옥시프로필/트리메톡시실란, N-(3-트리에톡시실릴프로필)-4-히드록시부티르아미드, 또는 이들의 임의의 조합을 포함한다. 일부 경우에, 능동적 작용화제는 11-아세톡시운데실트리에톡시실란을 포함한다. 일부 경우에, 능동적 작용화제는 n-데실트리에톡시실란을 포함한다. 일부 경우에, 능동적 작용화제는 글리시딜옥시프로필트리에톡시실란(GOPS)을 포함한다. 일부 경우에, 실란은 플루오로실란이다. 일부 경우에, 실란은 탄화수소 실란이다. 일부 경우에, 실란은 3-요오도-프로필트리메톡시실란이다. 일부 경우에, 실란은 옥틸클로로실란이다.

[0113] 일부 경우에, 실란화는 유기작용성 알콕시실란 분자와의 자가 어셈블리를 통해 표면 상에서 수행된다. 유기작용성 알콕시실란은 그들의 유기 작용에 따라 분류된다. 실록산 작용화 시약의 비제한적인 예는 히드록시알킬실록산(표면을 실릴화, 디보란으로 작용화 및 과산화수소에 의한 알코올의 산화), 디올(디히드록시알킬)실록산(표면의 실릴화 및 디올로의 가수분해), 아미노알킬 실록산(아민은 중간 작용화 단계를 필요로 하지 않음), 글리시독시실란(3-글리시독시프로필-디메틸-에톡시실란, 글리시독시-트리메톡시실란), 머캅토실란(3-머캅토프로필-트리메톡시실란, 3-4 에폭시시클로헥실-에틸트리메톡시실란 또는 3-머캅토프로필-메틸-디메톡시실란), 비시클로헵테닐-트리클로로실란, 부틸-알데히드-트리메톡시실란, 또는 이량체성 2차 아미노알킬 실록산을 포함한다. 예시적인 히드록시알킬 실록산은 3-히드록시프로필로 전환하는 알릴 트리클로로클로로실란, 또는 8-히드록시옥틸로 전환하는 7-옥트-1-에닐 트리클로로클로로실란을 포함한다. 디올(디히드록시알킬)실록산은 글리시딜트리메톡시실란-유도된 (2,3-디히드록시프로필옥시)프로필(GOPS)을 포함한다. 아미노알킬실록산은 3-아미노프로필(3-아미노프로필-트리에톡시실란, 3-아미노프로필-디에톡시-메틸실란, 3-아미노프로필-디메톡시-에톡시실란, 또는 3-아미노프로필-트리메톡시실란)으로 전환하는 3-아미노프로필 트리메톡시실란을 포함한다. 일부 경우에, 이량체성 제2 아미노알킬 실록산은 비스(실릴옥실프로필)아민으로 전환하는 비스(3-트리메톡시실릴프로필)아민을 포함한다.

[0114] 능동적 작용화 구역은 하나 이상의 상이한 종의 실란, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 이를 초과하는 실란을 포함할 수 있다. 일부 경우에, 하나 이상의 실란 중 하나는 또 다른 실란보다 더 많은 양으로 작용화 조성물 중에 존재한다. 예컨대, 2개의 실란을 갖는 혼합된 실란 용액은 하나의 실란 대 또 다른 실란의 비율이 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, 55:45이다. 일부 경우에, 능동적 작용화제는 11-아세톡시운데실트리에톡시실란 및 n-데실트리에톡시실란을 포함한다. 일부 경우에, 능동적 작용화제는 11-아세톡시운데실트리에톡시실란 및 n-데실트리에톡시실란을 약 20:80 내지 약 1:99, 또는 약 10:90 내지 약 2:98, 또는 약 5:95의 비율로 포함한다.

[0115] 일부 경우에, 작용화는 임의의 침착 기술에 의해 작용화제를 기관에 침착시키는 것을 포함하며, 침착 기법은 화학적 증착(CVD), 원자층 침착(ALD), 플라즈마 강화된 CVD(PECVD), 플라즈마 강화된 ALD(PEALD), 금속 유기 CVD(MOCVD), 핫 와이어 CVD(HWCVD), 개시된 CVD(iCVD), 변형된 CVD(MCVD), 기상 축 침착(VAD), 외부 증착(OVD), 물리적 증착(예컨대, 스퍼터 침착, 증발 침착), 및 분자층 침착(MLD)을 포함하나, 이들로 제한되지 않는다.

[0116] 후술하는 작용화 공정에서 임의의 단계 또는 성분은 최종 작용화된 기관의 목적하는 특성에 따라 생략되거나 변경된다. 일부 경우에, 추가의 성분 및/또는 공정 단계가 본원에 예시된 공정 워크플로우에 추가된다. 일부 경우에, 기관은 먼저, 예컨대 피라냐(piranha) 용액을 사용하여 세정된다. 세정 공정의 예는 승온(예컨대, 120℃)에

서 피라냐 용액(예컨대, 90% H_2SO_4 , 10% H_2O_2) 중에 기판을 침지하고, 기판을 세척하고(예컨대, 물), 및 기판을 건조시킨다(예컨대, 질소 기체). 공정은 임의적으로 염기성 용액(예컨대, NH_4OH)에 피라냐 처리된 기판을 침지하고, 이어서 수성 세척(예컨대, 물)하는 것을 포함하는 피라냐 후 처리를 포함한다. 일부 경우에, 구조물의 표면은 플라즈마 세정되고, 임의적으로 이어서 피라냐 침지되고 및 임의의 피라냐 후 처리된다. 플라즈마 세정 공정의 예는 산소 플라즈마 에칭을 포함한다. 일부 경우에, 표면은 능동적 작용화제로 침착되고, 이어서 증발된다. 일부 경우에, 기판은 세정 이전에, 예컨대 피라냐 처리 및/또는 플라즈마 세정에 의해 능동적으로 작용화된다.

[0117] 표면 작용화를 위한 공정은 임의적으로 레지스트 코트 및 레지스트 스트립을 포함한다. 일부 경우에, 능동적 표면 작용화 후, 기판은 레지스트, 예컨대 SPRTM 3612 포지티브 포토레지스트로 스핀 코팅된다. 다양한 경우에서, 표면 작용화를 위한 공정은 패턴화된 작용화를 갖는 리소그래피를 포함한다. 일부 경우에, 레지스트 코팅후 포토리소그래피가 수행된다. 일부 경우에, 리소그래피 후, 표면은 리소그래피 결함에 대해 육안 검사된다. 일부 경우에, 표면 작용화를 위한 공정은 세정 단계를 포함함으로써, 기판의 잔류물은, 예컨대 플라즈마 세정 또는 에칭에 의해 제거된다. 일부 경우에, 플라즈마 세정 단계는 리소그래피 단계 후 일부 단계에서 수행된다.

[0118] 일부 경우에, 레지스트로 코팅된 표면은, 예컨대 작용화 후 및/또는 리소그래피 후에 레지스트를 제거하기 위해 처리된다. 일부 경우에, 레지스트는 용매, 예컨대 N-메틸-2-피롤리돈을 포함하는 스트립핑 용액으로 제거된다. 일부 경우에, 레지스트 스트립핑은 음파처리 또는 초음파처리를 포함한다. 일부 경우에, 레지스트는 코팅 및 스트립핑되고, 이어서 노출된 구역의 능동적 작용화에 의해 목적하는 차등 작용화 패턴을 생성한다.

[0119] 일부 경우에, 본원에 기재된 방법 및 조성물은 선택적 구역에서 변형된 표면 특성 생성을 위한 포토레지스트의 적용에 관한 것이며, 여기서 포토레지스트의 적용은 포토레지스트의 공간적인 분포를 규정하는 표면의 유체 특성에 의존한다. 이론에 구애됨이 없이, 적용된 유체와 관련된 표면 장력 효과는 포토레지스트의 유동을 규정할 수 있다. 예컨대, 표면 장력 및/또는 모세관 작용 효과는 레지스트 용매가 증발되기 전에 조절된 방식으로 작은 구조물로 포토레지스트를 당기는 것을 용이하게 할 수 있다. 일부 경우에, 레지스트 접촉점은 날카로운 에지에 의해 고정되어 유체의 진행을 조절한다. 기반 구조물은 제작 및 작용화 공정 동안 포토레지스트를 적용하기 위해 사용되는 목적하는 유동 패턴에 기반하여 디자인될 수 있다. 용매 증발 후에 잔존하는 고정 유기층은 제작 공정의 후속 단계를 수행하기 위해 사용될 수 있다. 구조물은 이웃하는 유체 경로로의 위킹 효과(wicking effect)를 용이하게 하거나 또는 억제함으로써 유체의 유동을 조절하도록 디자인될 수 있다. 예컨대, 구조물은 상부 에지 및 하부 에지 사이의 중첩을 회피하도록 디자인되며, 이는 상부 구조물 내에 유체의 보유를 가능하게 하여 레지스트의 특정 침착을 허용 한다. 대안적인 예에서, 상부 및 하부 에지는 중첩되어 적용된 유체의 하부 구조물로의 위킹을 이끈다. 따라서, 적합한 디자인은 목적하는 레지스트의 적용에 따라 선택될 수 있다.

[0120] 일부 경우에, 본원에 기재된 구조물은 뉴클레오시드를 결합할 수 있는 반응성 기를 포함하는 적어도 또는 적어도 약 0.1 nm, 0.5 nm, 1 nm, 2 nm, 5 nm, 10 nm 또는 25 nm의 두께를 갖는 물질을 포함하는 표면을 갖는다. 예시적인 표면은 유리 및 규소, 예컨대 이산화규소 및 질화규소를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 예시적인 표면은 나일론 및 PMMA를 포함한다.

[0121] 일부 경우에, UV 광 형태의 전자기 복사가 표면 패턴화에 사용된다. 일부 경우에, 램프가 표면 패턴화에 사용되고, 마스크는 표면에 대한 UV 광의 노출 위치를 조정한다. 일부 경우에, 레이저가 표면 패턴화에 사용되고, 셔터 개방/폐쇄 상태가 표면에 대한 UV 광의 노출을 조절한다. 레이저 배열은 이동할 수 있는 가요성 구조물과 함께 사용될 수 있다. 이러한 배열에서, 레이저 노출 및 가요성 구조물 이동의 조정이 상이한 뉴클레오시드 커플링 능력을 갖는 하나 이상의 제제의 패턴을 생성하는 데 이용된다.

[0122] 물질 침착 시스템

[0123] 본원에는 본원에 기재된 구조물 상에 생체분자의 침착 및 저장을 위한 시스템 및 디바이스가 제공된다. 일부 경우에, 생체분자는 그들의 서열에 코딩된 정보를 저장하는 올리고뉴클레오티드이다. 일부 경우에, 시스템은 생체분자 부착을 지지하는 구조물의 표면 및/또는 기판의 표면에 생체분자의 적용을 위한 디바이스를 포함한다. 한 예에서, 생체분자 적용을 위한 디바이스는 올리고뉴클레오티드 합성기이다. 일부 경우에, 시스템은 유체로 기판을 처리하기 위한 디바이스, 예컨대 플로우 셀을 포함한다. 일부 경우에, 시스템은 적용 디바이스와 처리 디바이스 사이에서 기판을 이동시키기 위한 디바이스를 포함한다. 예컨대, 기판이 릴-투-릴 테이프인 경우, 시스템은 상이한 시간에 적용 및 임의의 처리 디바이스에 대한 기판의 상이한 부분의 접근을 허용하는 2개 이상의 릴을 포함할 수 있다.

- [0124] 올리고뉴클레오타이드 합성을 위한 올리고뉴클레오타이드 물질 침착 시스템의 제1 예가 도 15에 도시된다. 시스템은 기관의 위치를 따라 정렬되도록 X-Y 방향으로 이동하는 물질 침착 디바이스를 포함한다. 물질 침착 디바이스는 또한 Z 방향으로 이동하여 기관과 밀봉되어 리졸브드(resolved) 리액터를 형성할 수 있다. 리졸브드 리액터는 기관으로부터 캡핑 부재로 및/또는 그 반대로 올리고뉴클레오타이드 및/또는 시약을 포함하는 유체의 전달이 가능하도록 구성된다. 도 15에 도시되는 바와 같이, 유체는 기관 및 캡핑 부재 중 하나 또는 둘 다를 통해 통과할 수 있고, 커플링 시약, 캡핑 시약, 산화제, 디블록킹제, 아세트오니트릴 및 질소 기체를 포함하나, 이들로 제한되지 않는다. 고해상도의 액적 침착이 가능한 디바이스의 예는 잉크젯 프린터 및 레이저 프린터의 프린트헤드를 포함한다. 본원에 기재된 시스템 및 방법에 유용한 디바이스는 약 100 도트 퍼 인치(DPI) 내지 약 50,000 DPI; 약 100 DPI 내지 약 20,000 DPI; 약 100 DPI 내지 약 10,000 DPI; 약 100 DPI 내지 약 5,000 DPI; 약 1,000 DPI 내지 약 20,000 DPI; 또는 약 1,000 DPI 내지 약 10,000 DPI의 해상도를 달성한다. 일부 경우에, 디바이스는 적어도 약 1,000; 2,000; 3,000; 4,000; 5,000; 10,000; 12,000 DPI 또는 20,000 DPI의 해상도를 갖는다. 디바이스에 의해 수행되는 고해상도의 침착은 기관의 피처에 대응하는 각각의 노즐의 수 및 밀도와 관련된다.
- [0125] 올리고뉴클레오타이드 합성기를 사용하여 기관 상에서 올리고뉴클레오타이드를 드 노보 합성하는 예시적인 공정 워크플로우가 도 16에 도시된다. 올리고뉴클레오타이드 합성 시약을 포함하는 액적은 물질 침착 디바이스로부터 단계적 방식으로 기관으로 방출되며, 여기서 물질 침착 디바이스는 전기 신호를 액적 방출을 위한 기계적 신호로 전환하는 피에조 세라믹 물질 및 전극을 갖는다. 액적은 데이터를 코딩하는 미리결정된 서열을 갖는 복수의 합성된 올리고뉴클레오타이드를 생성하기 위해 한 번에 하나의 핵염기를 기관의 표면 상의 특정 위치에 방출한다. 일부 경우에, 합성된 올리고뉴클레오타이드는 기관 상에 저장된다. 핵산 시약은 불연속적인 방법 또는 주문형 드롭 방법으로 기관 표면 상에 침착될 수 있다. 이러한 방법의 예는 전자기계적 전달 방법, 전기적 열 전달 방법 및 정전기적 인력 방법을 포함한다. 전자기계적 전달 방법에서, 전기 펄스에 의해 변형된 압전 부재는 액적의 방출을 야기한다. 전기적 열 전달 방법에서, 버블이 디바이스의 챔버에 생성되며, 버블의 팽창력은 액적의 방출을 야기한다. 정전기적 유인 방법에서, 정전기적 인력이 기관 상에 액적을 방출하는 데 사용된다. 일부 경우에, 드롭 주파수는 약 5 KHz 내지 약 500 KHz; 약 5 KHz 내지 약 100 KHz; 약 10 KHz 내지 약 500 KHz; 약 10 KHz 내지 약 100 KHz; 또는 약 50 KHz 내지 약 500 KHz이다. 일부 경우에, 주파수는 약 500 KHz, 200 KHz, 100 KHz 또는 50 KHz 미만이다.
- [0126] 분배되는 액적 크기는 디바이스의 해상도와 관련된다. 일부 경우에, 디바이스는 약 0.01 pL 내지 약 20 pL, 약 0.01 pL 내지 약 10 pL, 약 0.01 pL 내지 약 1 pL, 약 0.01 pL 내지 약 0.5 pL, 약 0.01 pL 내지 약 0.01 pL, 또는 약 0.05 pL 내지 약 1 pL 크기로 시약의 액적을 침착한다. 일부 경우에, 액적의 크기는 약 1 pL, 0.5 pL, 0.2 pL, 0.1 pL 또는 0.05 pL 미만이다. 디바이스에 의해 분배되는 액적의 크기는 침착 노즐의 직경과 관련되며, 여기서 각각의 노즐은 기관의 피처 상에 시약을 침착시킬 수 있다. 일부 경우에, 올리고뉴클레오타이드 합성기의 침착 디바이스는 약 100 내지 약 10,000개의 노즐; 약 100 내지 약 5,000개의 노즐; 약 100 내지 약 3,000개의 노즐; 약 500 내지 약 10,000개의 노즐; 또는 약 100 내지 약 5,000개의 노즐을 포함한다. 일부 경우에, 침착 디바이스는 1,000; 2,000; 3,000; 4,000; 5,000; 또는 10,000개 초과 노즐을 포함한다. 일부 경우에, 각각의 물질 침착 디바이스는 복수의 노즐을 포함하며, 여기서 각각의 노즐은 임의적으로 기관 상의 피처에 대응되도록 구성된다. 각각의 노즐은 또 다른 노즐과는 상이한 시약 성분을 침착할 수 있다. 일부 경우에, 각각의 노즐은 기관의 하나 이상의 피처를 커버하는 액적을 침착한다. 일부 경우에, 하나 이상의 노즐은 비스듬할 수 있다. 일부 경우에, 다수의 침착 디바이스는 처리량을 배수의 증가를 달성하도록 나란히 적층된다. 일부 경우에, 증가는 2x, 4x, 8x 또는 이를 초과한다. 침착 디바이스의 예는 삼바 프린트헤드(Samba Printhead)(Fujifilm)이다. 삼바 프린트헤드는 SWAT(Samba Web Administration Tool)와 함께 사용될 수 있다.
- [0127] 침착 부위의 수는 동일한 침착 디바이스 이용하고 특정 각도 또는 사브르 각(saber angle)로 회전시킴으로써 증가될 수 있다. 침착 디바이스를 회전시킴으로써, 각각의 노즐은 사브르 각에 대응되는 특정 양의 지연 시간으로 분사된다. 이 비동기식 분사는 노즐 간에 크로스 토크를 생성한다. 따라서, 액적이 0도와는 상이한 특정 사브르 각으로 분사되는 경우, 노즐로부터의 액적 부피는 상이할 수 있다.
- [0128] 일부 배열에서, 올리고뉴클레오타이드 합성 시스템의 구성은 릴-투-릴 유형 공정으로 이동하기 위해 기관의 가요성을 이용하는 연속적인 올리고뉴클레오타이드 합성 공정을 가능하게 한다. 이 합성 공정은 기관의 위치를 회전시키기 위해 하나 이상의 릴을 사용하여 다양한 올리고뉴클레오타이드 합성 단계를 통해 이동하는 기관을 이용하는 연속 생산 라인 방식에서 구동된다. 예시적인 실시양태에서, 올리고뉴클레오타이드 합성 반응은, 용매 조를 통해, 포스포라미다이트 침착을 위한 침착 디바이스 아래로, 산화제 조를 통해, 아세트오니트릴 세척 조를 통해, 및 디블록 조를 통해 기관을 회전시키는 것을 포함한다. 임의적으로, 테이프는 또한 캡핑 조를 통해 횡단된다. 릴-

투-릴 유형 공정은 합성된 올리고뉴클레오타이드를 포함하는 기관의 최종 생성물이 테이크-업 릴 상에 용이하게 수집될 수 있게 하며, 이는 추가 프로세싱 또는 저장을 위해 수송될 수 있다.

[0129] 일부 배열에서, 올리고뉴클레오타이드 합성은 연속적인 가요성 테이프가 컨베이어 벨트 시스템을 따라 운반되는 연속 공정으로 진행된다. 릴-투-릴 유형 공정과 유사하게, 연속적인 테이프 상에서의 올리고뉴클레오타이드 합성은, 기관이 운반 동안 다양한 올리고뉴클레오타이드 합성 단계를 통해 이동하는 생산라인 방식으로 구동된다. 그러나, 컨베이어 벨트 공정에서, 연속적인 테이프는 릴-투-릴 공정에서와 같은 테이프의 롤링 및 폴립 없이 올리고뉴클레오타이드 합성 단계를 다시 수행한다. 일부 배열에서, 올리고뉴클레오타이드 합성 단계는 대역으로 분할되고, 연속적인 테이프는 한 사이클에서 1회 이상 각각의 대역을 통해 운반된다. 예컨대, 올리고뉴클레오타이드 합성 반응은 (1) 한 사이클에서 용매 조를 통해, 포스포르아미다이트 침착을 위한 침착 디바이스 아래로, 산화제 조를 통해, 아세트니트릴 세척 조를 통해, 및 블록 조를 통해 기관을 운반하는 것; 및 (2) 미리결정된 길이의 합성된 올리고뉴클레오타이드를 수득하기 위해 사이클을 반복하는 것을 포함할 수 있다. 올리고뉴클레오타이드 합성 후, 가요성 기관은 컨베이어 벨트 시스템으로부터 제거되고, 임의적으로, 저장을 위해 롤링된다. 롤링은 저장을 위해 릴 주위에서 수행된다.

[0130] 예시적인 배열에서, 열가소성 물질을 포함하는 가요성 기관은 뉴클레오시드 커플링 시약으로 코팅된다. 코팅은 각각의 피처가 약 10 μm 의 직경을 갖고 2개의 인접한 피처 사이의 중심 대 중심 거리가 약 21 μm 이도록 피처로 패턴화된다. 이 경우, 피처 크기는 올리고뉴클레오타이드 합성 침착 단계 동안 0.2 pL의 정적 부피를 수용하기에 충분하다. 일부 경우에, 피처 밀도는 약 22억개의 피처/ m^2 (1 피처/ $441 \times 10^{-12} \text{ m}^2$)이다. 일부 경우에, 4.5 m^2 기관은 약 백억개의 피처를 포함하고, 각각의 피처의 직경은 10 μm 이다.

[0131] 본원에 기재된 물질 침착 디바이스는 각각의 노즐이 액적당 1 핵염기로 초당 약 100,000개의 액적을 침착하는 약 2,048개의 노즐을 포함할 수 있다. 각각의 침착 디바이스에서, 일당 적어도 약 1.75×10^{13} 개의 핵염기가 기관 상에 침착된다. 일부 경우에, 100 내지 500개의 핵염기 올리고뉴클레오타이드가 합성된다. 일부 경우에, 200개의 핵염기 뉴클레오타이드가 합성된다. 임의적으로, 3일에 걸쳐, 약 1.75×10^{13} 개 염기/일의 속도로, 적어도 약 262.5×10^9 개의 올리고뉴클레오타이드가 합성된다.

[0132] 일부 배열에서, 합성 반응 동안 기관에 하나 이상의 시약을 적용하기 위한 디바이스는 뉴클레오시드 포스포르아미다이트 기반 합성을 위한 시약 및/또는 뉴클레오타이드 단량체를 침착하도록 구성된다. 올리고뉴클레오타이드 합성을 위한 시약은 올리고뉴클레오타이드 연장을 위한 시약 및 세척 완충액을 포함한다. 비제한적인 예로서, 디바이스는 세정 시약, 커플링 시약, 캡핑 시약, 산화제, 디블록킹제, 아세트니트릴, 기체, 예컨대 질소 기체, 및 이들의 임의의 조합을 침착한다. 또한, 디바이스는 임의적으로 기관 통합성의 제조 및/또는 유지를 위한 시약을 침착한다. 일부 경우에, 올리고뉴클레오타이드 합성기는 약 1000, 500, 100, 50 또는 20 pL 미만의 부피로 직경이 약 200 μm , 100 μm 또는 50 μm 미만인 드롭을 침착한다. 일부 경우에, 올리고뉴클레오타이드 합성기는 초당 약 1 내지 10000, 1 내지 5000, 100 내지 5000 또는 1000 내지 5000개의 액적을 침착한다.

[0133] 일부 배열에서, 올리고뉴클레오타이드 합성 동안, 기관은 플로우 셀 내에 위치되고/거나 밀봉된다. 플로우 셀은 액체, 예컨대 기관 내에서 반응에 필요한 시약, 예컨대 산화제 및/또는 용매를 포함하는 액체의 연속적인 또는 불연속적인 유동을 제공할 수 있다. 플로우 셀은 전형적으로 휘발성 기관의 강화된 증발을 통해 기관을 건조시키기 위한 기체, 예컨대 질소의 연속적인 또는 불연속적인 유동을 제공할 수 있다. 다양한 보조 디바이스가 기관의 표면 상에서 건조 및 잔류 수분 감소를 개선하는 데 유용하다. 이러한 보조 건조 디바이스의 예는 진공 공급원, 탈압 펌프 및 진공 탱크를 포함하나, 이들로 제한되지 않는다. 일부 경우에, 올리고뉴클레오타이드 합성 시스템은 하나 이상, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 20개의 플로우 셀 및 하나 이상, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 20개의 기관을 포함한다. 일부 경우에, 플로우 셀은 합성 반응에서 하나 이상의 단계 동안 기관에 시약을 유지 및 제공하도록 구성된다. 일부 경우에, 플로우셀은 기관의 상부 위로 미끄러지고 기관의 에지 주위에 가압 기밀을 형성하도록 제자리에서 클램핑될 수 있는 리드를 포함한다. 적합한 밀봉은 약 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10 대기 압력을 허용하는 밀봉을 포함하나, 이들로 제한되지 않는다. 일부 경우에, 플로우 셀의 리드는 적용 디바이스, 예컨대 올리고뉴클레오타이드 합성기에 접근 가능하도록 개방된다. 일부 경우에, 올리고뉴클레오타이드 합성 방법의 하나 이상의 단계는 기관의 수송 없이 플로우 셀 내의 기관 상에서 수행된다.

[0134] 일부 배열에서, 유체로 기관을 처리하기 위한 디바이스는 분무 바를 포함한다. 뉴클레오타이드 단량체는 기관 표면 상에 적용되고, 그 후 분무 바는 분무 바의 분무 노즐을 사용하여 하나 이상의 처리 시약을 기관 표면에 분

무한다. 일부 배열에서, 분무 노즐은 올리고뉴클레오타이드 합성 동안 상이한 처리 단계와 상호관련되도록 순차적으로 주문된다. 상이한 공정 단계에서 사용되는 화학물질은 합성 방법 및 합성 방법의 단계들 사이에서의 변화를 용이하게 수용하도록 분무 바 내에서 변화될 수 있다. 일부 경우에, 분무 바는 기관이 분무 바를 지나서 이동함에 따라 기관의 표면 상에 소정의 화학물질을 연속적으로 분무한다. 일부 경우에, 분무 바는 잔디 스프링클러에서 사용되는 분무 바와 매우 유사하게 기관의 넓은 구역에 걸쳐 침착한다. 일부 경우에, 분무 바 노즐은 기관의 소정의 구역에 처리 물질의 균일한 코트를 제공하도록 배치된다.

[0135] 일부 경우에, 올리고뉴클레오타이드 합성 시스템은 합성되는 올리고뉴클레오타이드의 다운스트림 프로세싱에 유용한 하나 이상의 부재를 포함한다. 한 예로서, 시스템은 온도 조절 부재, 예컨대 열 사이클링 디바이스를 포함한다. 일부 경우에, 온도 조절 부재는 핵산 어셈블리, 예컨대 PCA 및 핵산 증폭, 예컨대 PCR을 수행하기 위한 복수의 리졸브드 리액터와 함께 사용된다.

[0136] 드 노보 올리고뉴클레오타이드 합성

[0137] 본원에는 본원에서 기재되는 바와 같은 바이오암호화 및/또는 바이오복호화를 위한 디바이스, 조성물, 시스템 및 방법과 함께 사용하기 위한, 짧은 시간에 기관 상에서 고밀도의 올리고뉴클레오타이드를 합성하기 위한 시스템 및 방법이 제공된다. 일부 경우에, 기관은 가요성 기관이다. 일부 경우에, 적어도 약 10^{10} , 10^{11} , 10^{12} , 10^{13} , 10^{14} 또는 10^{15} 개의 염기가 하루에 합성된다. 일부 경우에, 적어도 약 10×10^8 , 10×10^9 , 10×10^{10} , 10×10^{11} 또는 10×10^{12} 개의 올리고뉴클레오타이드가 하루에 합성된다. 일부 경우에, 합성된 각각의 올리고뉴클레오타이드는 적어도 약 20, 50, 100, 200, 300, 400 또는 500개의 핵염기를 포함한다. 일부 경우에, 이들 염기는 100; 200; 300; 400; 500; 1000; 2000; 5000; 10000; 15000; 20000개의 염기 중 약 1개 미만의 전체 평균 에러율로 합성된다. 일부 경우에, 이들 에러율은 합성되는 올리고뉴클레오타이드의 적어도 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99.5%, 또는 이를 초과하는 것에 대한 것이다. 일부 경우에, 합성되는 올리고뉴클레오타이드의 이들 적어도 90%, 95%, 98%, 99%, 99.5%, 또는 이를 초과하는 이들이 코딩하는 미리결정된 서열과 상이하지 않다. 일부 경우에, 본원에 기재된 방법 및 시스템을 이용하는, 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 에러율은 200개의 염기 중 약 1개 미만이다. 일부 경우에, 본원에 기재된 방법 및 시스템을 이용하는, 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 에러율은 1,000개의 염기 중 약 1개 미만이다. 본원에 기재된 방법 및 시스템을 이용하는, 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 에러율은 2,000개의 염기 중 약 1개 미만이다. 본원에 기재된 방법 및 시스템을 이용하는, 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 에러율은 3,000개의 염기 중 약 1개 미만이다. 본원에 기재된 방법 및 시스템을 이용하는, 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 에러율은 5,000개의 염기 중 약 1개 미만이다. 에러율의 개별 유형은 기관 상에서 합성되는 올리고뉴클레오타이드에 대한 미스매치, 결실, 삽입 및/또는 치환을 포함한다. 용어 "에러율"은 합성된 올리고뉴클레오타이드의 총량을 미리결정된 올리고뉴클레오타이드 서열의 합계와 비교하는 것을 지칭한다. 일부 경우에, 본원에 개시된 합성된 올리고뉴클레오타이드는 12 내지 25개 염기의 테더를 포함한다. 일부 경우에, 테더는 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, 50개 또는 이를 초과하는 염기를 포함한다.

[0138] 본 개시의 기관 상에서 올리고뉴클레오타이드 합성을 위한 적절한 방법은, 포스포르아미다이트 빌딩 블록과 기관에 결합된 뉴클레오시드 사이에 포스포이트 트리아스테르 결합을 형성하는 커플링 단계에서 성장하는 올리고뉴클레오타이드에 대한 포스포르아미다이트 빌딩 블록, 즉 뉴클레오시드 포스포르아미다이트의 조절된 첨가를 포함하는 포스포르아미다이트 방법이다. 일부 경우에, 뉴클레오시드 포스포르아미다이트는 활성화된 기관에 제공된다. 일부 경우에, 뉴클레오시드 포스포르아미다이트는 기관-결합된 뉴클레오시드에 비해 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, 100배 과량 또는 이를 초과하게 기관에 제공된다. 일부 경우에, 뉴클레오시드 포스포르아미다이트의 첨가는 무수 환경, 예컨대 무수 아세토니트릴 중에서 수행된다. 커플링 단계에서 뉴클레오시드 포스포르아미다이트의 첨가 및 결합 후, 기관은 임의적으로 세척된다. 일부 경우에, 커플링 단계는 1회 이상 추가로 반복되며, 임의적으로 기관에 뉴클레오시드 포스포르아미다이트를 첨가하는 사이에 세척 단계를 수행한다. 일부 경우에, 본원에서 사용되는 올리고뉴클레오타이드 합성 방법은 1, 2, 3회 또는 이를 초과하는 순차적인 커플링 단계를 포함한다. 많은 경우에, 커플링 이전에 기관에 결합된 뉴클레오시드는 보호기의 제거에 의해 탈보호되며, 여기서 보호기는 중합을 방지하는 작용을 한다. 통상적인 보호기는 4,4'-디메톡시트리틸(DMT)이다.

[0139] 커플링 후, 포스포르아미다이트 올리고뉴클레오타이드 합성 방법은 임의적으로 캡핑 단계를 포함한다. 캡핑 단계

에서, 성장하는 올리고뉴클레오타이드는 캡핑제로 처리된다. 일반적으로, 캡핑 단계는 커플링 후 미반응 기관-결합된 5'-OH 기가 추가로 쇠 연장되지 않도록 하여 내부 염기 결실을 가진 올리고뉴클레오타이드의 형성을 방지한다. 또한, 1H-테트라졸로 활성화된 포스포르아미다이트는 종종 소량으로 구아노신의 O6 위치와 반응한다. 이론에 구애됨이 없이, I₂/물로 산화 시, 이 부산물은 가능하게는 O6-N7 이동에 의해 탈퓨린화를 겪는다. 탈퓨린 부위는 올리고뉴클레오타이드의 최종 탈보호의 과정 중에 절단되어 전장 생성물의 수율을 감소시킬 수 있다. O6 변형은 I₂/물에 의한 산화 이전에 캡핑 시약으로 처리함으로써 제거될 수 있다. 일부 경우에, 올리고뉴클레오타이드 합성 동안 캡핑 단계가 포함되면 캡핑이 없는 합성에 비해 어려움이 감소된다. 한 예로서, 캡핑 단계는 기관-결합된 올리고뉴클레오타이드를 아세트산 무수물과 1-메틸이미다졸의 혼합물로 처리하는 것을 포함한다. 캡핑 단계 후, 기관은 임의적으로 세척된다.

[0140] 뉴클레오시드 포스포르아미다이트의 첨가에 이어서, 임의적으로 캡핑 및 1회 이상의 세척 단계 후에, 기관 결합된 성장하는 핵산은 산화될 수 있다. 산화 단계는 포스파이트 트리에스테르를, 천연적으로 발생하는 포스페이트 디에스테르 뉴클레오시드간 결합의 보호된 전구체인, 4배위결합된 포스페이트 트리에스테르로 산화시키는 것을 포함한다. 일부 경우에, 성장하는 올리고뉴클레오타이드의 산화는, 임의적으로 약염기, 예컨대 피리딘, 루티딘 또는 콜리딘의 존재 하에서, 요오드 및 물로 처리함으로써 달성된다. 산화는 때때로 tert-부틸 히드록시옥시드 또는 (1S)-(+)-(10-캄포르술폰닐)-옥사지리딘(CSO)을 사용하여 무수 조건 하에서 수행된다. 일부 방법에서, 캡핑 단계는 산화 후에 수행된다. 제2 캡핑 단계는, 지속될 수 있는 산화로부터의 잔류수가 후속의 커플링을 억제할 수 있으므로, 기관 건조를 허용한다. 산화 후, 기관 및 성장하는 올리고뉴클레오타이드는 임의적으로 세척된다. 일부 경우에, 산화 단계는 올리고뉴클레오타이드 포스포르티오에이트를 수득하기 위해 황화 단계로 대체되며, 여기서 임의의 캡핑 단계가 황화 후에 수행될 수 있다. 3-(디메톡시아미노메틸리덴)아미노-3H-1,2,4-디티아졸-3-티온, DDTT, 뷰케이지(Beaucage) 시약으로도 공지된 3H-1,2-벤조디티올-3-온 1,1-디옥사이드 및 N,N,N',N'-테트라에틸티우람 디설파이드(TETD)를 포함하나, 이들로 제한되지 않는, 많은 시약이 효율적인 황 전달을 수행할 수 있다.

[0141] 뉴클레오시드 통합의 후속 사이클이 커플링을 통해 일어나기 위해서는, 1급 히드록실기가 다음의 뉴클레오시드 포스포르아미다이트와 반응할 수 있도록 기관 결합된 성장하는 올리고뉴클레오타이드의 보호된 5' 말단은 제거되어야 한다. 일부 경우에, 보호기는 DMT이고, 디블록킹은 디클로로메탄 중의 트리클로로아세트산에 의해 일어난다. 연장된 시간 동안, 또는 권장되는 산 용액보다 더 강하게, 탈트리틸화를 수행하는 것은 고체 지지체-결합된 올리고뉴클레오타이드의 탈퓨린화의 증가를 유도할 수 있고, 이에 따라 목적하는 전장 생성물의 수율을 감소시킬 수 있다. 본원에 기재된 방법 및 조성물은 바람직하지 않은 탈퓨린화 반응을 제한하는 제어된 디블록킹 조건을 제공한다. 일부 경우에, 기관 결합된 올리고뉴클레오타이드는 디블록킹 후에 세척된다. 일부 경우에, 디블록킹 후의 효율적인 세척은 어려움이 낮은 합성된 올리고뉴클레오타이드에 기여한다.

[0142] 본원에 기재된 기관 상에서 올리고뉴클레오타이드의 합성 방법은 전형적으로 하기 단계의 반복 순서를 포함한다: 기관 피처의 표면에 보호된 단량체를 적용하여 표면, 링커, 또는 사전에 탈보호된 단량체와 연결시키는 단계; 이후에 적용되는 보호된 단량체와 반응할 수 있도록 적용된 단량체를 탈보호하는 단계; 결합을 위해 또 다른 보호된 단량체를 적용하는 단계. 하나 이상의 중간 단계는 산화 및/또는 황화를 포함한다. 일부 경우에, 하나 이상의 세척 단계는 상기 단계들 중 하나 또는 모두에 선행하거나 후행한다.

[0143] 일부 경우에, 올리고뉴클레오타이드는 광불안정성 보호기로 합성되며, 여기서 표면 상에서 생성되는 히드록실기는 광불안정성 보호기에 의해 블록킹된다. 표면이 예컨대 포토리소그래피 마스크를 통해서 UV 광에 노출될 때, 표면 상에 유리 히드록실기의 패턴이 생성될 수 있다. 이들 히드록실기는 포스포르아미다이트 화학에 따라 광보호된 뉴클레오시드 포스포르아미다이트와 반응할 수 있다. 제2 포토리소그래피 마스크를 적용하고 표면을 UV 광에 노출시켜 히드록실기의 제2 패턴을 생성시킨 후, 5'-광보호된 뉴클레오시드 포스포르아미다이트와 커플링할 수 있다. 마찬가지로, 패턴이 생성될 수 있고, 올리고머 쇠가 연장될 수 있다. 이론에 구애됨이 없이, 광절단 가능한 기의 불안정성은 사용되는 용매의 파장 및 극성에 따라 달라지며, 광절단 속도는 노출 지속 시간 및 광 강도에 의해 영향을 받을 수 있다. 이 방법은 다수의 인자, 예컨대 마스크 정렬의 정확도, 광 보호기의 제거 효율 및 포스포르아미다이트 커플링 단계의 수율과 같은 다수의 인자를 활용할 수 있다. 또한, 이웃 부위로의 의도하지 않은 광의 누출을 최소화할 수 있다. 스팟당 합성되는 올리고머의 밀도는 합성 표면 상의 리더 뉴클레오시드의 로딩을 조절함으로써 모니터링될 수 있다.

[0144] 올리고뉴클레오타이드 합성을 위한 지지체를 제공하는 기관의 표면은 합성된 올리고뉴클레오타이드 쇠가 표면으로부터 절단되도록 화학적으로 변형될 수 있다. 일부 경우에, 올리고뉴클레오타이드 쇠는 올리고뉴클레오타이드가 탈보

호되는 것과 동일한 시간에 절단된다. 일부 경우에, 올리고뉴클레오티드 쇠는, 올리고뉴클레오티드가 탈보호된 후, 절단된다. 예시적인 스킴에서, 트리알콕시실릴아민, 예컨대 $(CH_3CH_2O)_3Si-(CH_2)_2-NH_2$ 가 기관의 $SiOH$ 기와 반응되고, 이어서 아민과 함께 숙신산 무수물과 반응되어 아마이드 결합 및 핵산 쇠 성장이 지지되는 유리 OH를 생성한다. 절단은 암모니아 또는 메틸아민을 이용하는 기체상 절단을 포함한다. 일부 경우에, 올리고뉴클레오티드는 표면으로부터 일단 방출되면, 저장된 정보를 추출하기 위해 시퀀싱되고 해독되는 더 긴 핵산으로 어셈블링된다.

[0145] 올리고뉴클레오티드는 정보를 코딩하는 미리결정된 서열의 넓은 영역을 집합적으로 포괄하도록 디자인될 수 있다. 일부 경우에, 더 큰 올리고뉴클레오티드는 합성된 올리고뉴클레오티드를 합치는 연결 반응을 통해 생성된다. 연결 반응의 한 예는 폴리머라제 쇠 어셈블리(PCA)이다. 일부 경우에, 올리고뉴클레오티드의 적어도 일부는 범용 프라이머 결합을 위한 기체인 추가된 영역을 포함하도록 디자인된다. PCA 반응의 경우, 예비합성된 올리고뉴클레오티드는 서로 중첩되는 부분(예컨대, 중첩 서열을 갖는 4, 20, 40개 또는 이를 초과하는 염기)을 포함한다. 폴리머라제 사이클 동안, 올리고뉴클레오티드는 상보적 단편에 어닐링된 다음 폴리머라제에 의해 채워진다. 따라서 각각의 사이클은 올리고뉴클레오티드가 서로를 발견하는나에 따라 무작위적으로 다양한 단편의 길이를 증가시킨다. 단편 사이의 상보성은 완전한 큰 범위의 이중 가닥 DNA의 형성을 가능하게 한다. 일부 경우에, PCA 반응이 완료된 후, 서열 내의 미스매치를 제거하는 미스매치 수복 검출 효소를 사용하여 에러 보정 단계가 수행된다. 일단 표적 서열의 더 큰 단편이 생성되면, 이들은 증폭될 수 있다. 예컨대, 일부 경우에, 5' 및 3' 말단 어댑터 서열을 포함하는 표적 서열은 어댑터 서열에 하이브리드화되는 변형된 프라이머를 포함하는 폴리머라제 연쇄 반응(PCR)으로 증폭된다. 일부 경우에, 변형된 프라이머는 하나 이상의 우라실 염기를 포함한다. 변형된 프라이머의 사용은, 변형된 염기 및/또는 단편으로부터의 변형된 염기쌍을 절단하는 효소에 의해 남겨진 갭을 표적하는 것에 집중하는 효소 반응을 통한 프라이머의 제거를 가능하게 한다. 잔류하는 것은 어댑터 서열의 잔재가 없는 이중 가닥 증폭 생성물이다. 이 방식으로, 동일한 프라이머 세트와 병행하여 다수의 증폭 생성물이 생성되어 이중 가닥 DNA의 상이한 단편을 생성할 수 있다.

[0146] 에러 보정이 합성된 올리고뉴클레오티드 및/또는 어셈블링된 생성물에 대해 수행될 수 있다. 에러 보정을 위한 예시적 전략은 에러를 보정하기 위한 중첩 연장 PCR에 의한 부위-지시된 돌연변이유발을 포함하며, 이는 임의적으로 2회 이상의 복제 및 시퀀싱 라운드와 커플링된다. 특정한 경우에, 미스매치, 벌지 및 작은 루프, 화학적으로 변경된 염기 및/또는 다른 헤테로듀플렉스를 갖는 이중 가닥 핵산은 정확하게 합성된 핵산의 집단으로부터 선택적으로 제거된다. 일부 경우에, 이중 가닥 핵산 내의 미스매치 또는 쌍을 이루지 않은 염기를 인식하고 이에 또는 옆에 결합하는 단백질/효소를 사용하여 단일 또는 이중-가닥 파괴를 생성하거나 가닥 전달 전위 이벤트를 개시하는 에러 보정이 수행된다. 에러 보정을 위한 단백질/효소의 비제한적인 예는 엔도뉴클레아제(T7 엔도뉴클레아제 I, 이. 콜라이 엔도뉴클레아제 V, T4 엔도뉴클레아제 VII, 녹두 뉴클레아제, CelI, 이. 콜라이 엔도뉴클레아제 IV, UVDE), 제한 효소, 글리코실라아제, 리보뉴클레아제, 미스매치 수복 효소, 리졸바아제, 헬리케이스, 리가아제, 미스매치에 특이적인 항체, 및 이들의 변이체를 포함한다. 특정 에러 보정 효소의 예는 T4 엔도뉴클레아제 7, T7 엔도뉴클레아제 1, S1, 녹두 엔도뉴클레아제, MutY, MutS, MutH, MutL, 클리바아제, CELI 및 HINF1을 포함한다. 일부 경우에, DNA 미스매치-결합 단백질 MutS(써머스 아쿠아티쿠스(*Thermus aquaticus*))를 사용하여 합성된 생성물 집단으로부터 불량 생성물을 제거한다. 일부 경우에, 효소 코렉타아제를 사용하여 에러 보정이 수행된다. 일부 경우에, 헤테로듀플렉스 DNA에 대해 알려진 및 알려지지 않은 돌연변이 및 다형성을 검사하는 미스매치-특이적 DNA 엔도뉴클레아제인 SURVEYOR 엔도뉴클레아제(Transgenomic)를 사용하여 에러 보정을 수행한다.

[0147] 방출, 추출 및 어셈블리

[0148] 본원에는 복제 가능한 정보 저장을 위한 방법 및 디바이스가 제공된다. 일부 경우에, 동일한 코딩 영역, 올리고뉴클레오티드, 동일한 클러스터, 올리고뉴클레오티드를 포함하는 구조물의 동일한 부분, 또는 올리고뉴클레오티드를 포함하는 전체 구조물의 다수 카피가 합성된다. 동일한 올리고뉴클레오티드의 다수 카피가 합성되는 경우, 각각의 올리고뉴클레오티드는 표면의 구별되는 영역에 부착될 수 있다. 구별되는 영역은 파괴 또는 절단에 의해 분리될 수 있다. 대안적으로, 각각의 올리고뉴클레오티드는 스팟, 웰 또는 채널의 형태로 피처에 존재할 수 있고, 개별적으로 접근 가능하다. 예컨대, 피처의 절단 시약 및 이어서 물과의 접촉은 올리고뉴클레오티드의 하나의 카피는 유리시키면서 다른 카피는 그대로 남겨둔다. 유사하게, 전체 영역 내에서 또는 전체 플레이트에 걸쳐 올리고뉴클레오티드의 절단은 복제 집단의 분획에 접근할 수 있게 한다. 복제 집단은 분리된 릴, 플레이트, 벨트 등에 존재할 수 있다. 가요성 물질, 예컨대 테이프의 경우, 복제 영역은 절단될 수 있고, 테이프의 나머지 영역들은 함께 다시 스플라이싱될 수 있다. 대안적으로, 합성되고 저장된 올리고뉴클레오티드의 핵산 정보는 프

라이머 및 DNA 폴리머라제를 사용하여 구조물의 표면에 부착된 올리고뉴클레오티드의 증폭을 수행함으로써 얻을 수 있다.

[0149] 일부 경우에, 수성 또는 기체상 전달 매체는 구조물 내의 하나 또는 복수의 채널 상에 침착되어 올리고뉴클레오티드를 구조물로부터 수신 유닛으로 전달한다. 예컨대, 전달 매체는 올리고뉴클레오티드에 부착되고, 이를 수집하고, 이를 구조물의 채널로부터 수신 유닛으로 전달하기 위해 구조물 내의 채널을 통해 통과할 수 있다. 일부 경우에, 전하 전도성 피쳐 및 인가된 전압은 구조물 내의 채널에 또는 채널을 통해 전달 매체를 당기거나 밀어내는 데 사용된다. 일부 경우에, 슬립은 구조물 내의 채널 내로 전달 매체를 유도하는 데 사용된다. 일부 경우에, 압력 방출은 구조물 내의 채널 내로 또는 채널을 통해 전달 매체를 유도하는 데 사용된다. 일부 경우에, 노즐은 구조물 내의 채널 내로 또는 채널을 통해 전달 매체를 미는 고압의 국소화된 구역을 형성하는 데 사용된다. 일부 경우에, 핀은 올리고뉴클레오티드를 구조물 내의 채널로부터 용기로 수신 유닛으로 전달하는 데 사용된다. 이러한 경우에, 핀은 전달 매체 부착을 용이하게 하게 하는 제제를 포함할 수 있다. 일부 경우에, 전하 전도성 피쳐는 전도성 피쳐와 구조물 사이에 전압 전위를 형성함으로써 구조물 내의 채널로 또는 채널을 통해 전달 매체를 당기거나 밀어내기 위해 사용된다. 일부 경우에, 피켓 팁 또는 다른 모세관 유동 유도 구조물은 모세관 유동에 의해 유체 및 올리고뉴클레오티드를 전달하는 데 사용된다. 일부 경우에, 용기는 하나 이상의 구획을 포함하며, 각각의 구획은 그 내부에 각각의 단일 채널로부터 방출되는, 전달 매체의 일부 및 하나 이상의 올리고뉴클레오티드를 수신한다. 일부 경우에, 용기는 전달 매체의 하나 이상의 부분을 수신하는 단일 구획을 포함하며, 각각은 그 내부에 하나 이상의 구조물 채널로부터 방출되는, 하나 이상의 올리고뉴클레오티드를 함유한다.

[0150] 시퀀싱

[0151] 구조물의 표면으로부터 올리고뉴클레오티드의 추출 및/또는 증폭 후, 적합한 시퀀싱 기술을 이용하여 올리고뉴클레오티드를 시퀀싱할 수 있다. 일부 경우에, DNA 서열은 기관 상에서 또는 구조물의 피쳐 내에서 판독된다. 일부 경우에, 기관 상에 저장된 올리고뉴클레오티드는 추출되고, 임의적으로 더 긴 핵산 서열로 어셈블링되며, 이어서 시퀀싱된다.

[0152] 본원에 기재된 구조물 상에서 합성되고 저장된 올리고뉴클레오티드는, 합성된 올리고뉴클레오티드의 서열을 판독하고 서열을 컴퓨터에 의해 판독가능한 2진 코드로 변환함으로써 해석될 수 있는 데이터를 코딩한다. 일부 경우에, 서열은 어셈블리가 필요하고, 및 어셈블리 단계는 핵산 서열 단계 또는 디지털 서열 단계에서 필요할 수 있다.

[0153] 본원에는 구조물 상에서 직접적으로 및/또는 주 구조물로부터 제거 후에 저장된 올리고뉴클레오티드를 시퀀싱할 수 있는 디바이스를 포함하는 검출 시스템이 제공된다. 구조물이 가요성 물질의 릴-투-릴 테이프인 경우, 검출 시스템은 검출 위치를 통해 구조물을 유지 및 진행시키기 위한 디바이스 및 섹션이 검출 위치에 있는 경우 테이프의 섹션으로부터 유래하는 신호를 검출하기 위한 검출 위치에 인접하게 배치된 검출기를 포함한다. 일부 경우에, 신호는 올리고뉴클레오티드의 존재를 나타낸다. 일부 경우에, 신호는 올리고뉴클레오티드의 서열을 나타낸다(예컨대, 형광 신호). 일부 경우에, 테이프는 컴퓨터에 작동 가능하게 연결된 검출기를 통해 연속적으로 운반되므로, 연속 테이프 상에서 올리고뉴클레오티드 내에 코딩된 정보는 컴퓨터에 의해 판독된다. 일부 경우에, 검출 시스템은 올리고뉴클레오티드 시퀀싱 디바이스, 올리고뉴클레오티드 서열과 관련된 데이터의 저장 및 회수를 위한 데이터베이스, 올리고뉴클레오티드 서열의 DNA 코드를 2진 코드로 변환하기 위한 소프트웨어, 2진 코드를 판독하기 위한 컴퓨터, 또는 이들의 임의의 조합을 포함한다.

[0154] 컴퓨터 시스템

[0155] 다양한 측면에서, 본원에 기재된 임의의 시스템은 컴퓨터에 작동 가능하게 연결되고, 임의적으로 로컬 또는 원격으로 컴퓨터를 통해 자동화된다. 다양한 경우에서, 본 개시의 방법 및 시스템은 추가로 컴퓨터 시스템에 대한 소프트웨어 프로그램 및 이의 사용을 포함한다. 따라서, 분배/진공/재충전 기능의 동기화, 예컨대 물질 침착 디바이스 이동, 분배 작용 및 진공 작동을 조화 및 동기화를 위한 컴퓨터화된 조절은 본 개시의 범위 내에 있다. 일부 경우에, 컴퓨터 시스템은 사용자가 특정한 염기 서열과 기관의 특정 영역으로 정확한 시약을 전달하는 물질 침착 디바이스의 위치 사이의 인터페이스로 프로그래밍된다.

[0156] 도 17에 도시된 컴퓨터 시스템(1700)은, 임의적으로 고정된 매체(1712)를 갖는 서버(1709)에 연결될 수 있는, 매체(1711) 및/또는 네트워크 포트(1705)로부터의 명령을 판독할 수 있는 논리적인 장치로서 이해될 수 있다. 도 17에 도시된 바와 같은 시스템은 CPU(1701), 디스크 드라이브(1703), 임의의 입력 디바이스, 예컨대 키보드

(1715) 및/또는 마우스(1716) 및 임의의 모니터(1707)를 포함할 수 있다. 데이터 통신은 로컬 또는 원격 위치에 서 서버에 대한 지시된 통신 매체를 통해 달성될 수 있다. 통신 매체는 데이터를 전송 및/또는 수신하기 위한 임의의 수단을 포함할 수 있다. 예컨대, 통신 매체는 네트워크 연결, 무선 연결 또는 인터넷 연결일 수 있다. 이러한 연결은 월드 와이드 웹 상에서의 통신을 제공할 수 있다. 본 개시와 관련된 데이터는 사용자(1722)에 의한 접수 및/또는 검토를 위해 이러한 네트워크 또는 연결로 전송될 수 있는 것으로 고려된다.

[0157] 도 18는 본 개시의 예시적인 실시양태와 함께 사용될 수 있는 컴퓨터 시스템(1800)의 제1 예시의 아키텍처를 도시하는 블록 다이어그램이다. 도 18에 도시되는 바와 같이, 예시적 컴퓨터 시스템은, 명령을 프로세싱하기 위한 프로세서(1802)를 포함할 수 있다. 프로세서의 비제한적인 예는 인텔 Xeon™ 프로세서, AMD Opteron™ 프로세서, 삼성 32-bit ISC ARM 1176JZ(F)-S v1.0™ 프로세서, ARM Cortex-A8 삼성 S5PC100™ 프로세서, ARM Cortex-A8 애플 A4™ 프로세서, 마벨 PXA 930™ 프로세서, 또는 기능적으로 균등한 프로세서를 포함한다. 병렬 프로세싱에 다수의 실행 스레드를 이용할 수 있다. 일부 경우에, 복수의 컴퓨터, 휴대전화, 및/또는 개인 데이터 어시스턴트 디바이스를 포함하는 네트워크를 통해 단일 컴퓨터 시스템에서든, 클러스터에서든, 또는 시스템에 분산되어 있든, 다수의 프로세서 또는 다수의 코어를 갖는 프로세서가 또한 사용될 수 있다.

[0158] 도 18에 도시된 바와 같이, 고속 캐시(1804)는 최근에 프로세서(1802)에 의해 사용되었거나 자주 사용되는 명령 또는 데이터를 위한 고속 메모리를 제공하기 위해 프로세서(1802)에 연결되거나 통합될 수 있다. 프로세서(1802)는 프로세서 버스(1806)에 의해 노스 브리지(1806)에 연결된다. 노스 브리지(1806)는 메모리 버스(1812)에 의해 랜덤 접근 메모리(RAM)(1810)에 접속되고, 프로세서(1802)에 의해 RAM(1810)에 대한 접근을 관리한다. 노스 브리지(1806)는 또한 칩셋 버스(1816)에 의해 사우스 브리지(1814)에 연결된다. 사우스 브리지(1814)는 차례로 주변장치 버스(1818)에 접속된다. 주변장치 버스는, 예컨대 PCI, PCI-X, PCI Express 또는 다른 주변장치 버스일 수 있다. 노스 브리지 및 사우스 브리지는 종종 프로세서 칩셋으로 지칭되고, 프로세서, RAM 및 주변장치 버스(1818) 상의 주변장치 구성요소 사이의 데이터 전송을 관리한다. 일부 대안적 아키텍처에서, 별도의 노스 브리지 칩 대신에, 노스 브리지의 기능을 프로세서에 통합시킬 수 있다.

[0159] 일부 경우에, 시스템(1800)은 주변장치 버스(1818)에 부착된 가속기 카드(1822)를 포함할 수 있다. 가속기는 필드 프로그램 가능한 게이트 어레이(FPGA) 또는 특정 프로세싱을 가속화하기 위한 다른 하드웨어를 포함할 수 있다. 예컨대, 가속기는 적응 데이터 재구성, 또는 확장된 세트 프로세싱에 사용되는 대수 표현식을 평가하는 데 사용될 수 있다.

[0160] 소프트웨어 및 데이터는 외부 저장장치(1824)에 저장되고, 프로세서에 의한 사용을 위해 RAM(1810) 및/또는 캐시(1804)로 로딩될 수 있다. 시스템(1800)은 시스템 자원을 관리하기 위한 운영 시스템을 포함하고, 운영 체계의 비제한적인 예는, Linux, Windows™, MACOSTM, BlackBerry OSTM, iOSTM 및 다른 기능적으로 동등한 운영 체계 뿐만 아니라, 본 개시의 예시적인 실시양태에 따라 데이터 저장 및 최적화를 관리하기 위한 운영 체계의 정상부에서 실행되는 응용 소프트웨어도 포함한다.

[0161] 이 예에서, 시스템(1800)은 또한 외부 저장장치, 예컨대 네트워크 접속 저장장치(NAS)에 네트워크 인터페이스를 제공하기 위해 주변장치 버스에 연결된 네트워크 인터페이스 카드(NIC)(1820 및 1821) 및 분산 병렬 프로세싱에 사용될 수 있는 다른 컴퓨터 시스템을 포함한다.

[0162] 도 19는 복수의 컴퓨터 시스템(1902a 및 1902b), 복수의 휴대전화 및 개인 데이터 어시스턴트(1902c), 및 네트워크 접속 저장장치(NAS)(1904a 및 1904b)를 갖는 네트워크(1900)를 도시하는 다이어그램이다. 예시적인 실시양태에서, 시스템(1902a, 1902b 및 1902c)은 네트워크 접속 저장장치(NAS)(1904a 및 1904b)에 저장된 데이터에 대한 데이터 저장을 관리하고 데이터 접근을 최적화할 수 있다. 컴퓨터 시스템(1902a 및 1902b), 휴대전화 및 개인 데이터 어시스턴트 시스템(1902c)을 거쳐, 분산된 병렬 프로세싱을 이용하여 수학적 모델이 데이터에 사용 및 평가될 수 있다. 컴퓨터 시스템(1902a 및 1902b), 휴대전화 및 개인 데이터 어시스턴트 시스템(1902c)은 또한 네트워크 접속 저장장치(NAS)(1904a 및 1904b)에 저장된 데이터의 적응 데이터 재구성을 위한 병렬 프로세싱을 제공할 수 있다. 도 19는 단지 예시를 도시하는 것이며, 다양한 다른 컴퓨터 아키텍처 및 시스템이 본 발명의 다양한 실시양태와 연관되어 사용될 수 있다. 예컨대, 블레이드 서버를 사용하여 병렬 프로세싱을 제공할 수 있다. 프로세서 블레이드는 백 플레인을 통해 연결되어 병렬 프로세싱을 제공할 수 있다. 저장장치는 또한 별도의 네트워크 인터페이스를 통해 백 플레인에 또는 네트워크 접속 저장장치(NAS)로 연결될 수 있다.

[0163] 일부 예시적인 실시양태에서, 프로세서는 다른 프로세서에 의한 병렬 프로세싱을 위해 별도의 메모리 공간을 유지하고, 네트워크 인터페이스, 백 플레인 또는 다른 커넥터를 통해 데이터를 전송할 수 있다. 다른 경우에, 프로세서의 일부 또는 전부는 공유된 가상 주소 메모리 공간을 사용할 수 있다.

- [0164] 도 20은 예시적인 실시양태에 따른 공유된 가상 주소 메모리 공간을 이용하는 멀티프로세서 컴퓨터 시스템(2000)의 블록 다이어그램이다. 시스템은 공유된 메모리 서브시스템(2004)에 접근할 수 있는 복수의 프로세서(2002a-f)를 포함한다. 시스템은 메모리 서브시스템(2004)에 복수의 프로그램 가능한 하드웨어 메모리 알고리즘 프로세서(MAP)(2006a-f)를 통합한다. 각각의 MAP(2006a-f)는 메모리(2008a-f) 및 하나 이상의 필드 프로그램 가능한 게이트 어레이(FPGA)(2010a-f)를 포함할 수 있다. MAP는 구성 가능한 기능 유닛을 제공하며, 각각의 프로세서와 긴밀히 협조하여 프로세싱하기 위해, 특정 알고리즘 또는 알고리즘의 일부가 FPGA(2010a-f)에 제공될 수 있다. 예컨대, MAP는 데이터 모델에 관한 대수 표현을 평가하고 예시적인 실시양태에서 적응 데이터 채구성을 수행하는 데 사용될 수 있다. 이 예에서, 각각의 MAP는 이들 목적을 위해 모든 프로세서에 의해 전역으로 접근가능하다. 한 구성에서, 각각의 MAP는 직접 메모리 액세스(DMA)를 사용하여 관련 메모리(2008a-f)에 접근할 수 있으며, 이는 각각의 마이크로프로세서(2002a-f)와 독립적으로 그리고 비동기적으로 작업을 실행할 수 있게 한다. 이 구성에서, MAP는 파이프라이닝 및 알고리즘의 병렬 실행을 위해 또 다른 MAP에 직접적으로 결과를 제공할 수 있다.
- [0165] 상기 컴퓨터 아키텍처 및 시스템은 단지 예시일 뿐이며, 일반적인 프로세서, 코-프로세서, FPGA 및 기타 프로그램 가능한 논리 디바이스, 시스템 온 칩(SOC), 응용 주문형 집적 회로(ASIC) 및 기타 프로세싱 및 논리 소자를 포함하는, 다양한 다른 컴퓨터, 휴대전화 및 개인 데이터 어시스턴트 아키텍처 및 시스템이 예시적인 실시양태와 관련하여 사용될 수 있다. 일부 경우에, 컴퓨터 시스템의 전부 또는 일부는 소프트웨어 또는 하드웨어로 구현될 수 있다. 랜덤 액세스 메모리, 하드 드라이브, 플래시 메모리, 테이프 드라이브, 디스크 어레이, 네트워크 접속 저장장치(NAS) 및 다른 로컬 또는 분산 데이터 저장장치 및 시스템을 포함하는, 임의의 다양한 데이터 저장 매체가, 예시적인 실시양태와 관련하여 사용될 수 있다.
- [0166] 예시적인 실시양태에서, 컴퓨터 시스템은 상기 또는 다른 컴퓨터 아키텍처 및 시스템 중 임의의 것에서 실행하는 소프트웨어 모듈을 사용하여 구현될 수 있다. 다른 경우에, 시스템의 기능은 펌웨어, 프로그램 가능한 논리 디바이스, 예컨대 필드 프로그램 가능한 게이트 어레이(FPGA), 시스템 온 칩(SOC), 응용 주문형 집적 회로(ASIC) 또는 다른 프로세싱 및 논리 소자에서 부분적으로 또는 완전히 구현될 수 있다. 예컨대, Set 프로세서 및 Optimizer는 하드웨어 가속기 카드, 가속기 카드의 사용을 통한 하드웨어 가속에 의해 구현될 수 있다.
- [0167] 본원에는 정보를 저장하기 위한 방법이 제공되며, 상기 방법은 적어도 하나의 디지털 서열 형태의 정보 아이টে를 적어도 하나의 핵산 서열로 전환하는 단계; 표면을 갖는 가요성 구조물을 제공하는 단계; 적어도 하나의 핵산 서열을 집합적으로 코딩하는 미리결정된 서열을 갖는 복수의 올리고뉴클레오타이드를 합성하는 단계로서, 복수의 올리고뉴클레오타이드는 적어도 약 100,000개의 올리고뉴클레오타이드를 포함하고, 복수의 올리고뉴클레오타이드는 가요성 구조물의 표면으로부터 연장되는 것인 단계; 및 복수의 올리고뉴클레오타이드를 저장하는 단계를 포함한다. 또한, 본원에는 합성이 미리결정된 위치에서 표면 상에 뉴클레오타이드를 침착시키는 단계; 및 조 또는 분무 바로부터의 방출을 통해 가요성 구조물의 적어도 일부를 이동시키는 단계를 포함하는 방법이 제공된다. 또한, 본원에는 조 또는 분무 바로부터의 방출이 구조물의 표면을 산화 시약 또는 디블록킹 시약에 노출시키는 방법이 제공된다. 또한, 본원에는 합성이 표면 상에 침착된 뉴클레오타이드를 캡핑하는 것을 추가로 포함하는 방법이 제공된다. 또한, 본원에는 뉴클레오타이드가 뉴클레오타이드 포스포르아미다이트를 포함하는 방법이 제공된다. 또한, 본원에는 가요성 구조물이 릴-투-릴 테이프 또는 연속 테이프를 포함하는 방법이 제공된다. 또한, 본원에는 가요성 구조물이 열가소성 물질을 포함하는 방법이 제공된다. 또한, 본원에는 열가소성 물질이 폴리아릴에테르케톤을 포함하는 방법이 제공된다. 또한, 본원에는 폴리아릴에테르케톤이 폴리에테르케톤, 폴리에테르케톤케톤, 폴리(에테르 에테르 케톤 케톤), 폴리에테르 에테르 케톤 또는 폴리에테르케톤에테르케톤케톤인 방법이 제공된다. 또한, 본원에는 가요성 구조물이 나일론, 니트로셀룰로스, 폴리프로필렌, 폴리카보네이트, 폴리에틸렌, 폴리우레탄, 폴리스티렌, 아세탈, 아크릴, 아크릴로니트릴, 부타디엔 스티렌, 폴리에틸렌 테레프탈레이트, 폴리메틸 메타크릴레이트, 폴리비닐 클로라이드, 투명 PVC 호일, 폴리(메틸 메타크릴레이트), 스티렌 중합체, 불소 함유 중합체, 폴리에테르설폰 또는 폴리이미드를 포함하는 방법이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드의 각각의 올리고뉴클레오타이드가 50 내지 500개 염기 길이를 포함하는 방법이 제공된다. 또한, 본원에는 복수의 올리고뉴클레오타이드가 적어도 약 백억개의 올리고뉴클레오타이드를 포함하는 방법이 제공된다. 또한, 본원에는 적어도 약 1.75×10^{13} 개의 핵염기가 24시간 내에 합성되는 방법이 제공된다. 또한, 본원에는 적어도 약 262.5×10^9 개의 올리고뉴클레오타이드가 72시간 내에 합성되는 방법이 제공된다. 또한, 본원에는 정보 아이টে이 텍스트 정보, 청각 정보 또는 시각 정보인 방법이 제공된다. 또한, 본원에는 뉴클레오타이드가 뉴클레오타이드 포스포르아미다이트를 포함하는 방법이 제공된다.

- [0168] 본원에는 정보를 저장하기 위한 방법이 제공되며, 상기 방법은 적어도 하나의 디지털 서열 형태의 정보 아이템을 적어도 하나의 핵산 서열로 전환하는 단계; 표면을 갖는 구조물을 제공하는 단계; 적어도 하나의 핵산 서열을 집합적으로 코딩하는 미리결정된 서열을 갖는 복수의 올리고뉴클레오타이드를 합성하는 단계로서, 여기서 복수의 올리고뉴클레오타이드는 적어도 약 100,000개의 올리고뉴클레오타이드를 포함하고, 여기서 복수의 올리고뉴클레오타이드는 구조물의 표면으로부터 연장되고, 여기서 합성은 구조물의 표면을 세척하고; 미리결정된 위치에서 표면 상에 뉴클레오타이드를 침착하고; 표면 상에 침착된 뉴클레오타이드를 산화시키고, 디블록킹하고, 임의적으로 캡핑하는 것을 포함하고, 여기서 세정, 산화, 디블록킹 및 캡핑은 조 또는 분무 바로부터의 방출을 통해 가요성 구조물의 적어도 일부를 이동시키는 것을 포함하는 단계; 및 복수의 올리고뉴클레오타이드를 저장하는 단계를 포함한다. 또한, 본원에는 뉴클레오타이드가 뉴클레오타이드 포스포라미다이트를 포함하는 방법이 제공된다.
- [0169] 하기의 실시예는 당업자에게 본원에 개시된 실시양태의 원리 및 실시를 보다 명확하게 예시하기 위해 기재된 것으로, 임의의 청구된 실시양태의 범위를 제한하는 것으로 해석되어서는 안 된다. 달리 지시되지 않은 한, 모든 부 및 백분율은 중량 기준이다.
- [0170] **실시예**
- [0171] **실시예 1: 디바이스 표면의 작용화**
- [0172] 올리고뉴클레오타이드 라이브러리의 부착 및 합성을 지지하기 위해 디바이스를 작용화하였다. 디바이스 표면을 먼저 90% H_2SO_4 및 10% H_2O_2 를 포함하는 피라냐 용액을 사용하여 20분 동안 습윤 세정하였다. 디바이스를 DI 수로 수개의 비이커에서 린스하고, DI 수 거위목형 수도꼭지 아래에 5분 동안 위치시키고, N_2 로 건조시켰다. 이어서, 디바이스를 NH_4OH (1:100; 3 mL:300 mL) 중에 5분 동안 침지시키고, 핸드건을 사용하여 DI 수로 린스하고, DI 수를 갖는 3개의 연속 비이커에 각각 1분 동안 침지시키고, 이어서 핸드건을 사용하여 DI 수로 다시 린스하였다. 이어서, 디바이스 표면을 O_2 에 노출시킴으로써 디바이스를 플라즈마 세정하였다. SAMCO PC-300 장비를 사용하여 250 watt에서 1분 동안 다운스트림 모드로 플라즈마 에칭하였다.
- [0173] 세정된 디바이스 표면을 하기 파라미터를 갖는 YES-I224P 증착 오븐 시스템을 사용하여 N-(3-트리에톡시실릴프로필)-4-히드록시부티르아미드를 포함하는 용액으로 능동적으로 작용화하였다: 0.5 내지 1 torr, 60분, 70°C, 135°C 증발기. 디바이스 표면을 브루어 사이언스 200X 스핀 코터(Brewer Science 200X spin coater)를 사용하여 레지스트 코팅하였다. SPR™ 3612 포토레지스트를 2500 rpm으로 40초 동안 디바이스 상에 스핀 코팅하였다. 디바이스를 브루어 열판 상에서 90°C에서 30분 동안 예비-베이킹(pre-baking)하였다. 디바이스를 칼 수스 MA6 마스크 얼라이너(Karl Suss MA6 mask aligner) 장비를 사용하여 포토리소그래피를 수행하였다. 디바이스를 2.2초 동안 노출하고, 1분 동안 MSF 26A에서 현상하였다. 잔류 현상액을 핸드건으로 린스하고, 디바이스를 5분 동안 물에 침지시켰다. 디바이스를 100°C 오븐에서 30분 동안 베이킹하고, 이어서 니콘(Nikon) L200을 사용하여 리소그래피 결함을 육안 검사하였다. SAMCO PC-300 장비를 사용하여 250 watt에서 1분 동안 O_2 플라즈마 에칭하는 스크 제거 공정을 이용하여 잔류 레지스트를 제거하였다.
- [0174] 디바이스 표면을 경무기질유 10 μL 와 혼합된 퍼플루오로옥틸트리클로로실란 용액 100 μL 로 수동적으로 작용화하였다. 디바이스를 챔버에 넣고, 10분 동안 펌핑하고, 이어서 밸브를 닫고, 10분 동안 정치시켰다. 챔버를 환기시켰다. 최대 출력(크레스트(Crest) 시스템에서 9)에서의 초음파 처리와 함께 70°C에서 500 mL NMP 중에서 5분 동안 2회 침지를 수행함으로써 디바이스를 레지스트 스트립핑하였다. 이어서, 디바이스를 최대 출력으로의 초음파 처리와 함께 실온에서 500 mL 이소프로판올 중에 5분 동안 침지시켰다. 디바이스를 200 프루프 에탄올 300 mL 중에 담그고, N_2 로 블로운(blowed) 건조시켰다. 작용화된 표면을 올리고뉴클레오타이드 합성을 위한 지지체로서 기능하도록 활성화시켰다.
- [0175] **실시예 2: 올리고뉴클레오타이드 합성 디바이스 상에서 50-mer 서열의 합성**
- [0176] 2차원 올리고뉴클레오타이드 합성 디바이스를 플로우셀(Applied Biosystems (ABI394 DNA 합성기))에 연결된 플로우셀로 어셈블링하였다. 2차원 올리고뉴클레오타이드 합성 디바이스를 N-(3-트리에톡시실릴프로필)-4-히드록시부티르아미드(Gelast)로 균일하게 작용화하고, 본원에 기재된 올리고뉴클레오타이드 합성 방법을 이용하여 50 bp의 예시적인 올리고뉴클레오타이드("50-mer 올리고뉴클레오타이드")를 합성하는 데 사용하였다.
- [0177] 50-mer의 서열은 서열번호 1에 나타낸 바와 같았다. 5' AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTTTTT3' (서열번호: 1), 여기서 #는 티미딘-숙시

닐 핵사미드 CED 포스포르아미다이트(ChemGenes로부터의 CLP-2244)를 나타내며, 이는 탈보호 동안 표면으로부터 올리고뉴클레오타이드의 방출을 가능하게 하는 절단 가능한 링커이다.

[0178] 표 5의 프로토콜에 따른 표준 DNA 합성 화학(커플링, 캡핑, 산화 및 디블록킹) 및 ABI 합성기를 사용하여 합성을 수행하였다.

표 5

합성 프로토콜

표 5		
일반적인 DNA 합성 공정 명칭	공정 단계	시간(초)
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	23
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
DNA 염기 첨가 (포스포르아미다이트 + 활성화제 유동)	활성화제 매니폴드 플러시	2
	플로우셀로의 활성화제	6
	플로우셀로의 활성화제 + 포스포르아미다이트	6
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포르아미다이트	5
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포르아미다이트	5
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포르아미다이트	5
	25 초 동안 인큐베이션	25
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
DNA 염기 첨가 (포스포르아미다이트 + 활성화제 유동)	활성화제 매니폴드 플러시	2
	플로우셀로의 활성화제	5
	플로우셀로의 활성화제 + 포스포르아미다이트	18
	25 초 동안 인큐베이션	25
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
캡핑 (CapA+B, 1:1, 유동)	플로우셀로의 CapA+B	15
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	아세트오니트릴 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
산화 (산화제 유동)	플로우셀로의 산화제	18
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	아세트오니트릴 시스템 플러시	4

[0179]

표 5		
일반적인 DNA 합성 공정 명칭	공정 단계	시간(초)
	플로우셀로의 아세트오니트릴	15
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	23
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
디블록킹 (디블록 유동)	플로우셀에 대한 디블록	36
세척 (아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	18
	N2 시스템 플러시	4.13
	아세트오니트릴 시스템 플러시	4.13
	플로우셀로의 아세트오니트릴	15

[0180]

[0181] 포스포르아미다이트/활성화제 조합을 플로우셀을 통해 벌크 시약의 전달과 유사하게 전달하였다. 환경이 전체

[0192] 표 7은 스왑 1-10으로부터의 올리고뉴클레오타이드 샘플로부터 얻은 서열에 대한 에러 특성을 요약한 것이다.

표 7

에러 특성

샘플 ID/스왑 번호	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
중 서열	32	32	32	32	32	32	32	32	32	32
시퀀싱 품질	28 중 25	27 중 27	30 중 26	23 중 21	26 중 25	30 중 29	31 중 27	31 중 29	29 중 28	28 중 25
올리고 품질	25 중 23	27 중 25	26 중 22	21 중 18	25 중 24	29 중 25	27 중 22	29 중 28	28 중 26	25 중 20
ROI 매칭 카운트	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 돌연변이	2	2	1	3	1	0	2	1	2	1
ROI 다중 염기 결실	0	0	0	0	0	0	0	0	0	0
ROI 소규모 삽입	1	0	0	0	0	0	0	0	0	0
ROI 단일 염기 결실	0	0	0	0	0	0	0	0	0	0
대규모 결실 카운트	0	0	1	0	0	1	1	0	0	0
돌연변이: G>A	2	2	1	2	1	0	2	1	2	1
돌연변이: T>C	0	0	0	1	0	0	0	0	0	0
ROI 에러 카운트	3	2	2	3	1	1	3	1	2	1
ROI 에러율	Err: 834 중 ~1	Err: 1350 중 ~1	Err: 1282 중 ~1	Err: 708 중 ~1	Err: 2500 중 ~1	Err: 2667 중 ~1	Err: 876 중 ~1	Err: 2900 중 ~1	Err: 1400 중 ~1	Err: 2349 중 ~1
ROI 마이너스 프라이머 에러율	MP Err: 763 중 ~1	MP Err: 824 중 ~1	MP Err: 780 중 ~1	MP Err: 429 중 ~1	MP Err: 1525 중 ~1	MP Err: 1615 중 ~1	MP Err: 531 중 ~1	MP Err: 1769 중 ~1	MP Err: 854 중 ~1	MP Err: 1451 중 ~1

[0193]

[0194] 실시예 4: 고정밀 DNA-기반 정보 저장 및 어셈블리

[0195] 디지털 정보를 100개 초과 언어의 세계인권선언, 프로젝트 구텐베르크(Project Gutenberg)의 상위 도서 100권 및 시드 데이터베이스에 대한 내용을 포함하는 총 약 0.2 GB의 2진 데이터의 형태로 선택하였다. 디지털 정보를 핵산-기반 서열로 암호화하고, 스트링으로 분할하였다. 각각 스트링에 대응되는, 천만개 이상의 동일하지 않은 올리고뉴클레오타이드를 실시예 2에 기재된 것과 유사한 방식으로 강성 실리콘 표면 상에서 합성하였다. 각각의 동일하지 않은 올리고뉴클레오타이드 길이는 200개 이하의 염기였다. 합성된 올리고뉴클레오타이드를 수집하고, 시퀀싱하고, 다시 디지털 코드로 디코딩하였으며, 초기의 적어도 하나의 디지털 서열과 비교하여 공급원 디지털 정보에 대해 100% 정확도를 가졌다.

[0196] 실시예 5: 디지털 정보의 핵산 서열로의 전환

[0197] 컴퓨터 txt 파일은 텍스트 정보를 포함한다. 범용 컴퓨터는 수신된 명령에 따라 서열을 염기 3, 4 또는 5 서열로 전환하기 위한 기계 명령을 갖는 소프트웨어 프로그램을 이용한다. 염기 3의 각각의 숫자에 핵산이 지정된다(예컨대, A=0, T=1, C=2). 염기 4의 각각의 숫자에 핵산이 지정된다(예컨대, A=0, T=1, C=2, G=3). 대안적으로, 염기 5의 5진법 서열이 사용되고, 염기 5의 각각의 숫자에 핵산이 지정된다(예컨대, A=0, T=1, C=2, G=3, U=4). 도 8에 도시되는 바와 같이 서열이 생성된다. 이어서, 핵산 서열을 코딩하는 올리고뉴클레오타이드의 드 노보 합성을 위한 기계 명령이 제공된다.

표 8

서열 전환

텍스트	책은 언덕에 올라갔다.
2 진법 서열	010010100110000101100011011010110010000001110111011001010110111001 110100001000000111010101110000001000000111010001101000011001010010 000001101000011010010110110001101100001011100000110100001010000011 0100001010
3 진법 서열	101010201100022101010021102012221200101112202210002122002210200011 1122121020112010211121222001011100010020010220022222110022222112
4 진법 서열	102212011203122302001313121112321310020013111300020013101220121102 00122012211230123002320031002200310022
5 진법 서열	332214330133012303013123001030244433343300431224103020320210201123 42341100431241100334213

[0198]

[0199]

실시예 6: 고밀도의 피처를 갖는 가요성 표면

[0200]

열가소성 물질을 포함하는 가요성 구조물을 뉴클레오타이드 커플링 시약으로 코팅한다. 고밀도 피처를 위해 코팅제를 패터닝한다. 가요성 표면의 일부가 도 14A에 도시된다. 각각의 피처의 직경은 10 μm 이고, 2개의 인접한 피처 사이의 중심-대-중심 거리는 21 μm 이다. 피처 크기는 올리고뉴클레오타이드 합성 침착 단계 동안 0.2 pL의 정적 부피를 수용하기에 충분하다. 작은 피처 크기는 기관의 표면 상에서 고밀도의 올리고뉴클레오타이드의 합성을 가능하게 한다. 피처 밀도는 22억개 피처/ m^2 (1 피처/ $441 \times 10^{-12} \text{m}^2$)이다. 백억개의 피처를 갖는 45 m^2 기관이 제조되며, 각각의 피처의 직경은 10 μm 이다. 가요성 구조물은 임의적으로 올리고뉴클레오타이드 합성을 위해 연속적인 루프 시스템(도 12A) 또는 릴-투-릴 시스템(도 12B)에 위치된다.

[0201]

실시예 7: 가요성 구조물 상에서 올리고뉴클레오타이드 합성

[0202]

열가소성의 가요성 물질 상에 복수의 피처를 포함하는 가요성 구조물을 제조한다. 구조물은 침착 디바이스를 포함하는 올리고뉴클레오타이드 합성 디바이스를 사용하는 올리고뉴클레오타이드 합성을 위한 지지체의 역할을 한다. 가요성 구조물은 자기 릴-투-릴 테이프와 매우 유사한 가요성 매체 형태이다.

[0203]

드 노보 합성은 구조물이 용매 조를 통과하고 이어서 포스포르아미다이트가 구조물의 표면 상에 프린트되는 프린트헤드의 스택 아래로 이동하는 연속 생산 라인 방식으로 작동한다. 표면 상에 침착된 정적을 갖는 가요성 구조물은 산화제 조로 롤링되며, 이어서 테이프는 산화 조를 벗어나 아세트니트릴 세척조에 잠기고, 이어서 디블록킹조에 침지된다. 임의적으로, 테이프는 캡핑 조를 통해 횡단한다. 대안적인 워크플로우에서, 가요성 구조물은 산화 조로부터 벗어나 세척 단계에서 아세트니트릴로 분무된다.

[0204]

대안적으로, 분무 바가 액체 조 대신에 사용된다. 이 공정에서, 뉴클레오타이드는 잉크젯 디바이스를 사용하여 표면 상에 여전히 침착되지만, 플러드(flood) 단계는 분무 노즐을 사용하여 챔버 내에서 수행된다. 예컨대, 침착 디바이스는 2,048개의 노즐을 가지며, 각각의 노즐은 액적당 1 핵염기로 초당 100,000개의 액적을 침착한다. 표준 포스포르아미다이트 화학에서 플러드 단계의 순서를 모방하는 분무 노즐의 순차적인 순서가 존재한다. 이 기술은 상이한 공정 단계를 수용하도록 분무 바에 로딩되는 화학물질을 용이하게 변경할 수 있도록 한다. 올리고뉴클레오타이드는 실시예 2에 기재된 바와 동일한 방식으로 탈보호되거나 절단된다.

[0205]

각각의 침착 디바이스에 대해, 일당 1.75×10^{13} 개 초과 핵염기가 구조물 상에 침착된다. 복수의 200개 핵염기 올리고뉴클레오타이드가 합성된다. 3일에, 1.75×10^{13} 개 염기/일의 속도로 262.5×10^9 개 올리고뉴클레오타이드가 합성된다.

[0206]

실시예 8: 바이오암호화의 선택

[0207]

전환을 위한 목적하는 정보 아이템, 및 효소적 기반(예컨대, CRISPR/Cas 복합체 및 제한 효소 분해), 전자기 복사 기반(예컨대, 광분해 및 광검출), 화학적 절단(예컨대, 티미딘-숙시닐 핵사미드 CED 포스포르아미다이트 (ChemGenes로부터의 CLP-2244)를 절단하기 위한 기체상 암모니아 또는 메틸아민 처리), 및 친화성 기반(예컨대, 하이브리드화를 위한 서열 태그 또는 캡처 시약에 대해 증진된 친화성을 갖는 변형된 뉴클레오타이드의 통합) 형태

의 바이오암호화로부터 선택되는 바이오암호화의 하나 이상의 카테고리에 대한 기계 명령이 수신된다. 특정 바이오암호화 선택의 수신 후, 프로그램 모듈은 정보 아이템을 핵산 서열로 전환하고 바이오암호화된 버전의 서열을 디자인하기 위한 디자인 명령을 적용하는 단계를 수행한다. 바이오암호화 카테고리 내의 구체적인 암호화 서브유형이 선택된다. 이어서, 올리고뉴클레오타이드의 드 노보 합성을 위해 합성 명령이 물질 칩작 디바이스에 제공된다.

[0208] **실시예 9: 선택된 바이오복호화**

[0209] 효소적 기반(예컨대, CRISPR/Cas 복합체 또는 제한 효소 분해), 전자기 복사 기반(예컨대, 광분해 또는 광검출), 화학적 절단(예컨대, 티미딘-숙시닐 헥사미드 CED 포스포프아미다이트(ChemGenes로부터의 CLP-2244)를 절단하기 위한 기체상 암모니아 또는 메틸아민 처리), 및 친화성 기반(예컨대, 하이브리드화를 위한 서열 태크 또는 캡처 시약에 대해 증진된 친화성을 갖는 변형된 뉴클레오타이드의 통합) 바이오복호화로부터 선택되는 바이오복호화의 하나 이상의 카테고리의 적용을 위한 기계 명령이 제공된다. 특정 바이오복호화 선택의 수신 후, 프로그램 모듈은 올리고뉴클레오타이드 농축을 위한 조정제(들)를 방출하는 단계를 수행한다. 농축 후, 올리고뉴클레오타이드가 시퀀싱되고, 임의적으로 보다 긴 핵산 서열에 대해 정렬되고, 정보 아이템에 대응하는 디지털 서열로 전환된다.

[0210] **실시예 10: CRISPR/Cas9를 사용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0211] 정보 아이템을 코딩하는 디지털 서열이 수신된다. 이어서, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열은 더 큰 핵산 서열 집단에서 암호화된다. 암호화 공정은 CRISPR/Cas9 복합체에 의한 검출 및 제거를 위한 "정크(junk)" 영역을 첨가하는 것을 수반한다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0212] 암호화된 핵산 서열을 포함하는 핵산 서열 집단이 Cas9 버퍼에서 Cas9 및 gRNA와 혼합되고, 37°C에서 2시간 동안 인큐베이션된다. 이어서, Cas9가 불활성화되며, 정제에 의해 제거된다. 이어서, 정제된 샘플은 차세대 시퀀싱에 의해 분석된다.

[0213] **실시예 11: 서열 스와핑을 포함하는 CRISPR/Cas9를 사용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0214] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열은 CRISPR/Cas9 시스템 및 가이드 RNA 서열을 사용하면서 특정 서열을 첨가함으로써 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0215] 이어서, 핵산 서열은 스와핑된 서열에 상보적인 형광-태깅된 프로브와 혼합된다. 형광-태깅된 프로브에 의해 확인되는 핵산 서열은 집단으로부터 제거된다.

[0216] **실시예 12: 제한 효소 분해를 이용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0217] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 제한 효소 EcoRI에 의해 인식되는 특정 서열을 첨가함으로써 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성되고 저장된다.

[0218] 핵산 서열은 EcoRI와 함께 인큐베이션된다. EcoRI 인식 부위를 포함하는 암호화된 핵산 서열이 절단된다. 암호화된 핵산 서열의 절단 후, 상보성 돌출부를 갖는 서열이 하이브리드화되고, 방출된 DNA에 결합된다. 이어서, 결합된 복합체가 분리되고, 정제된 샘플이 시퀀싱되며, 최초 디지털 정보가 어셈블링된다.

[0219] **실시예 13: 광분해를 이용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0220] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 광분해가능한 핵염기를 포함하도록 디자인된다. 핵산 서열은 실시예 2-3에서와 같이 합성되고 저장된다.

[0221] 280 nm의 UV-B 조사가 핵산 서열에 적용된다. 광분해가능한 부위를 포함하는 암호화된 핵산 서열이 절단되고 제거된다. 이어서, 핵산 서열이 수집되고 시퀀싱된다. 대안적으로, 핵산 서열은 구조물의 표면으로부터, 예컨대 암모니아 기체 절단에 의해 방출되고, 이어서 뉴클레오타이드 서열에 파괴를 제공하는 전자기 복사에 노출된다. 집단의 일부가, 예컨대 결합된 상보적 캡처 프로브를 갖는 비드를 사용하는 풀 다운 어세이, 표적 서열만을 증폭시키도록 선택되는 프라이머를 사용하는 PCR, 또는 크기 배제 크로마토그래피에 의해 농축된다. 이어서, 농축된 핵산은 시퀀싱되고, 디지털 서열로 전환되며, 정보 아이템이 수신된다.

[0222] **실시예 14: 화학적 농축을 이용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0223] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 암

모니아 기체에 의해 화학적으로 절단가능한 특정 서열(예컨대, 티미딘-숙시닐 헥사미드 CED 포스포르아미다이트 (ChemGenes로부터의 CLP-2244))의 첨가에 의해 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0224] 암모니아 기체가 핵산 서열에 적용된다. 화학적으로 절단가능한 서열을 포함하는 암호화된 핵산 서열이 방출되고, 본원에 기재된 농축 방법을 이용하여 집단으로부터 농축된다. 이어서, 농축된 핵산은 시퀀싱되고, 디지털 서열로 전환되며, 정보 아이템이 수신된다.

[0225] **실시예 15: 바이오틴을 포함하는 핵산 프로브를 사용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0226] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 바이오틴 함유 핵염기를 포함하도록 미리결정된 잔기의 디자인에 의해 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0227] 핵산 서열이 구조물로부터 절단되고, 스트랩트아비딘 함유 비드와 혼합된다. 이어서, 핵산 서열은 스트랩트아비딘 자기 비드와 함께 인큐베이션된다. 바이오틴을 포함하는 핵산 서열은 자기 비드에 의해 풀다운된다. 이어서, 농축된 핵산은 시퀀싱되고, 디지털 서열로 전환되며, 정보 아이템이 수신된다.

[0228] **실시예 16: 광검출을 이용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0229] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 Alexa488-태깅된 핵산 프로브에 의해 인식되는 특정 서열을 포함하도록 디자인에 의해 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0230] 핵산 서열이 구조물로부터 방출되고, Alexa488-태깅된 핵산 프로브와 혼합된다. 이어서, 핵산 서열은 형광 강도에 의해 분류된다. Alexa488-태깅된 핵산 프로브로 태깅된 핵산 서열은 추가로 분석된다. 이어서, 프로브 결합된 핵산은 시퀀싱되고, 디지털 서열로 전환되며, 정보 아이템이 수신된다.

[0231] **실시예 17: 변형된 뉴클레오티드를 사용하는 DNA 서열의 생물학적 암호화 및 복호화**

[0232] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 미리결정된 위치에 펩티드 핵산(PNA)을 포함하는 미리결정된 핵염기의 첨가 및 PNA 함유 섹션을 절제하는 제한 효소 인식 크기의 디자인을 위한 디자인에 의해 암호화된다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0233] 핵산 서열이 방출되고, 제한 효소 분해되며, 이어서 PCR에 의해 증폭된다. PNA를 포함하는 핵산 서열은 증폭될 수 없다. 이어서, 농축되고 증폭된 핵산은 시퀀싱되고, 디지털 서열로 전환되며, 정보 아이템이 수신된다.

[0234] **실시예 18: CRISPR/Cas9 및 화학적 절단을 이용하는 DNA 서열의 생물학적 암호화 및 복호화**

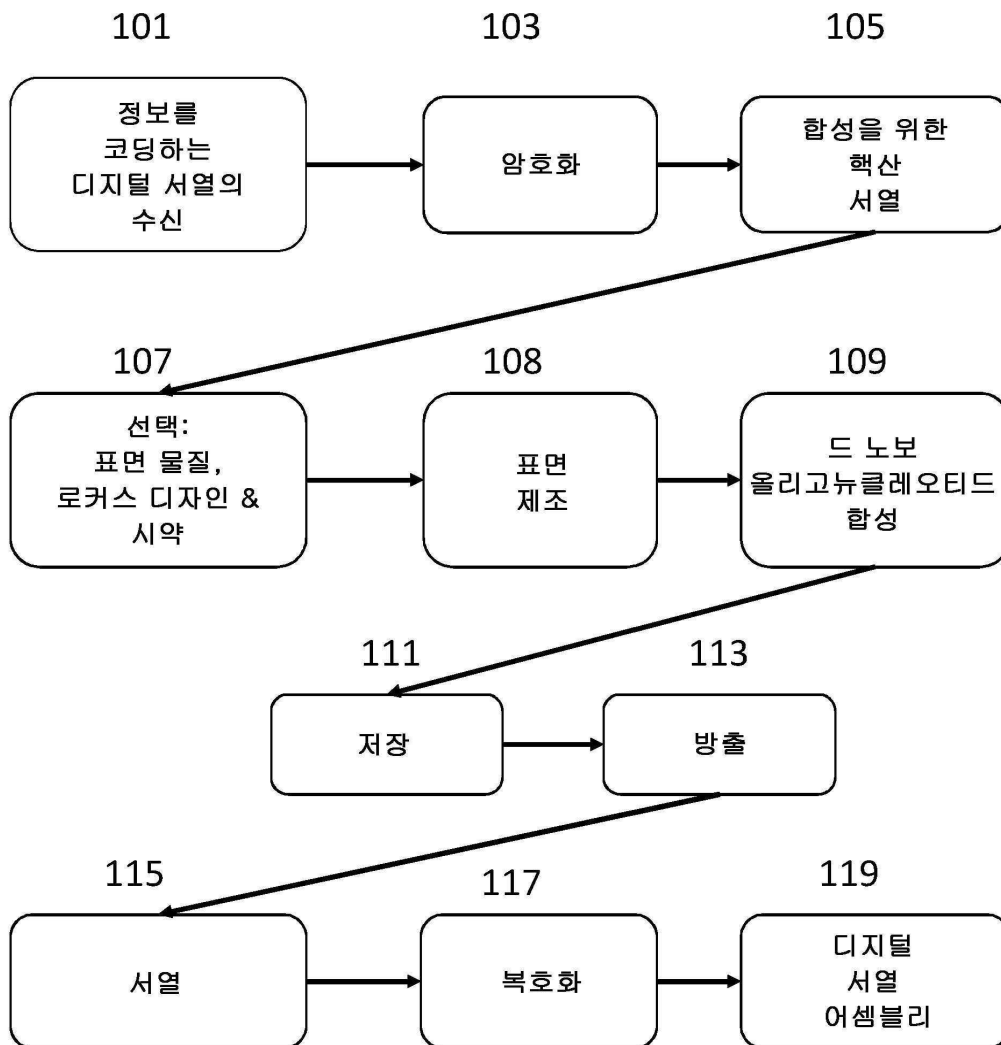
[0235] 정보 아이템을 코딩하는 디지털 서열이 수신되고, 디지털 서열은 핵산 서열로 전환된다. 핵산 서열의 집단은 CRISPR/Cas9 및 가이드 RNA 서열을 사용하여 특정 서열을 첨가함으로써 암호화된다. CRISPR/Cas9 시스템은 미리 선택된 위치에서 핵산 서열에 화학적으로 절단가능한 부위를 도입한다. 핵산 서열은 실시예 2-3에서와 같이 합성된다.

[0236] 암모니아 기체가 핵산 서열에 적용된다. 화학적으로 절단가능한 부위를 포함하는 암호화된 핵산 서열이 절단되고, 크기 배제 정제에 의해 제거되며, 차세대 시퀀싱에 의해 분석된다.

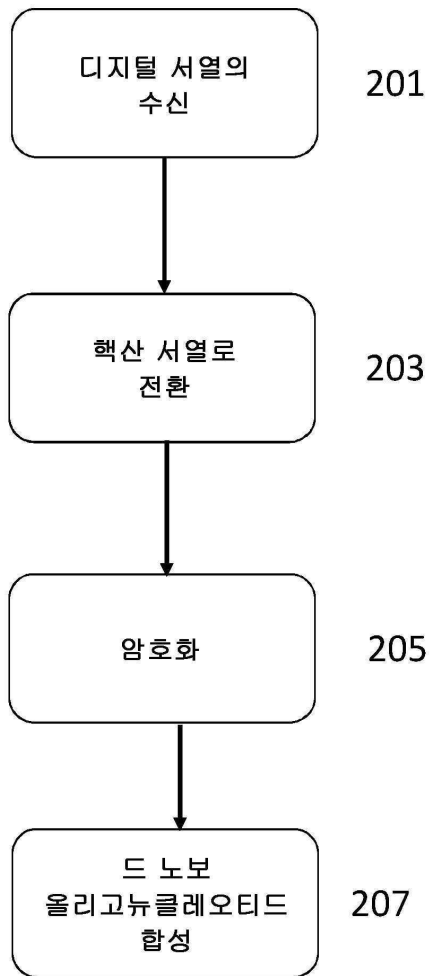
[0237] 본 발명의 바람직한 실시양태를 본원에 제시하고 기술하였으나, 이러한 실시양태는 단지 예시로 제공된다는 것이 당업자에게 명백할 것이다. 다수의 변화, 변경 및 치환이 본 발명을 벗어나지 않고 당업자에게 발생할 것이다. 본원에 기재된 본 발명의 실시양태에 대한 다양한 대안이 본 발명을 실시하는 데 사용될 수 있음을 이해하여야 한다. 하기 청구범위는 본 발명의 범위를 정의하며, 그에 따라 이들 청구범위 내의 방법 및 구조물 및 이들의 등가물이 포함되는 것으로 의도된다.

도면

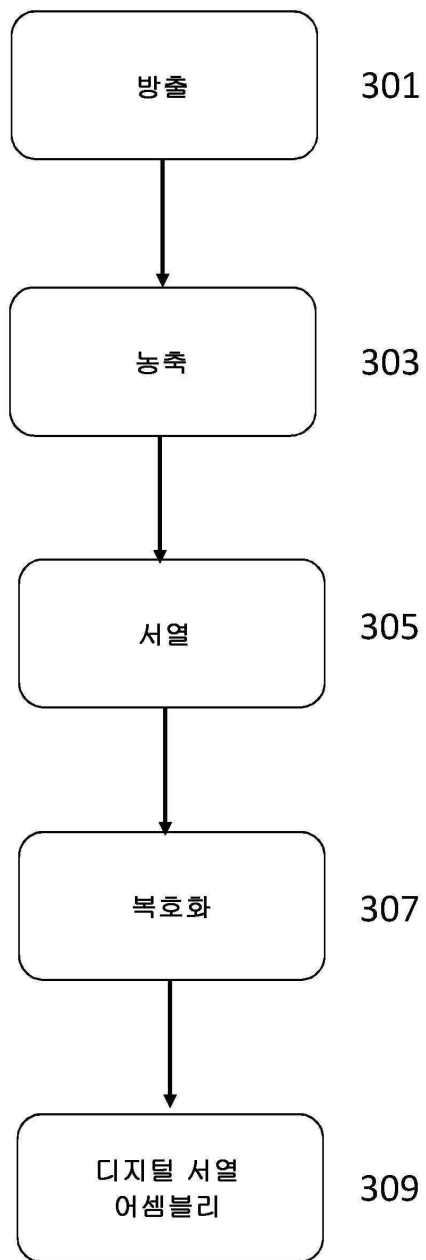
도면1



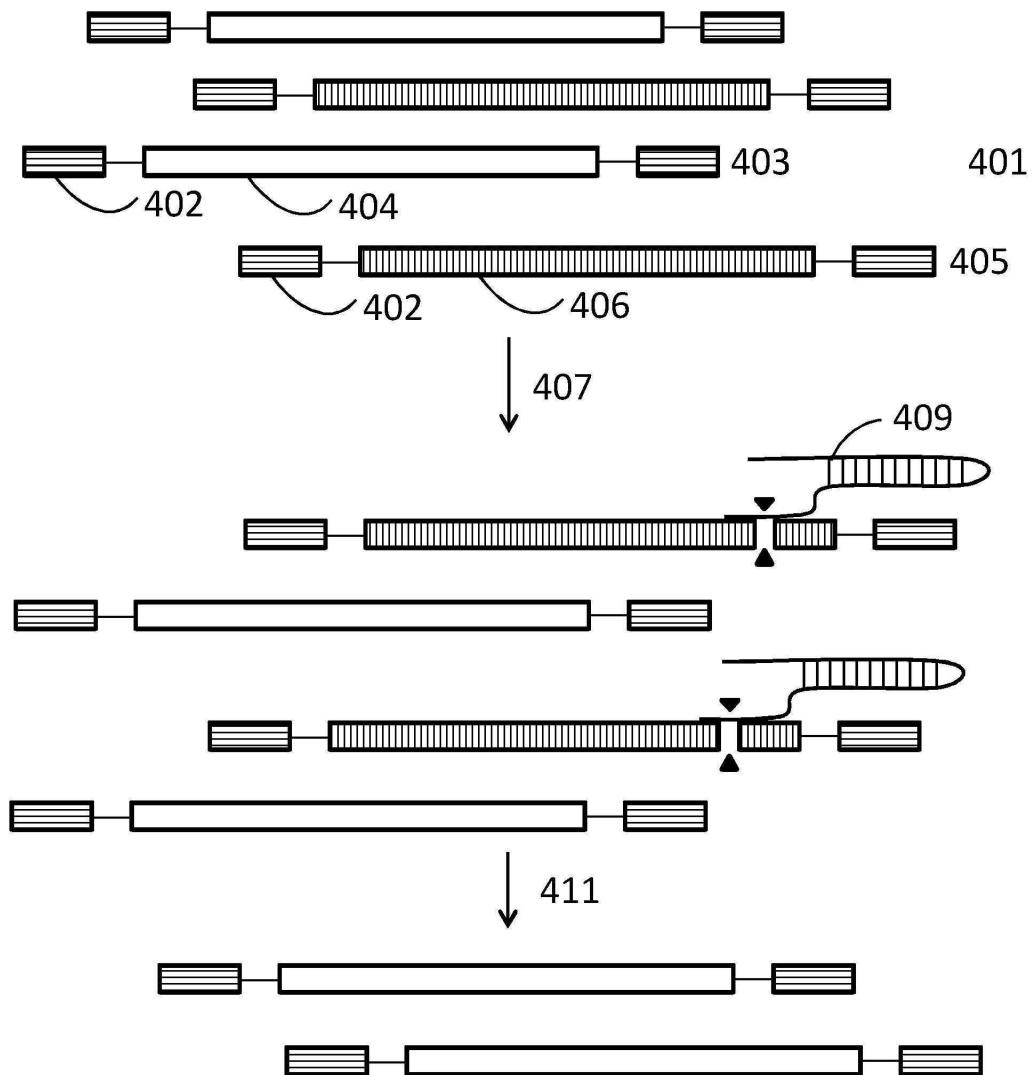
도면2



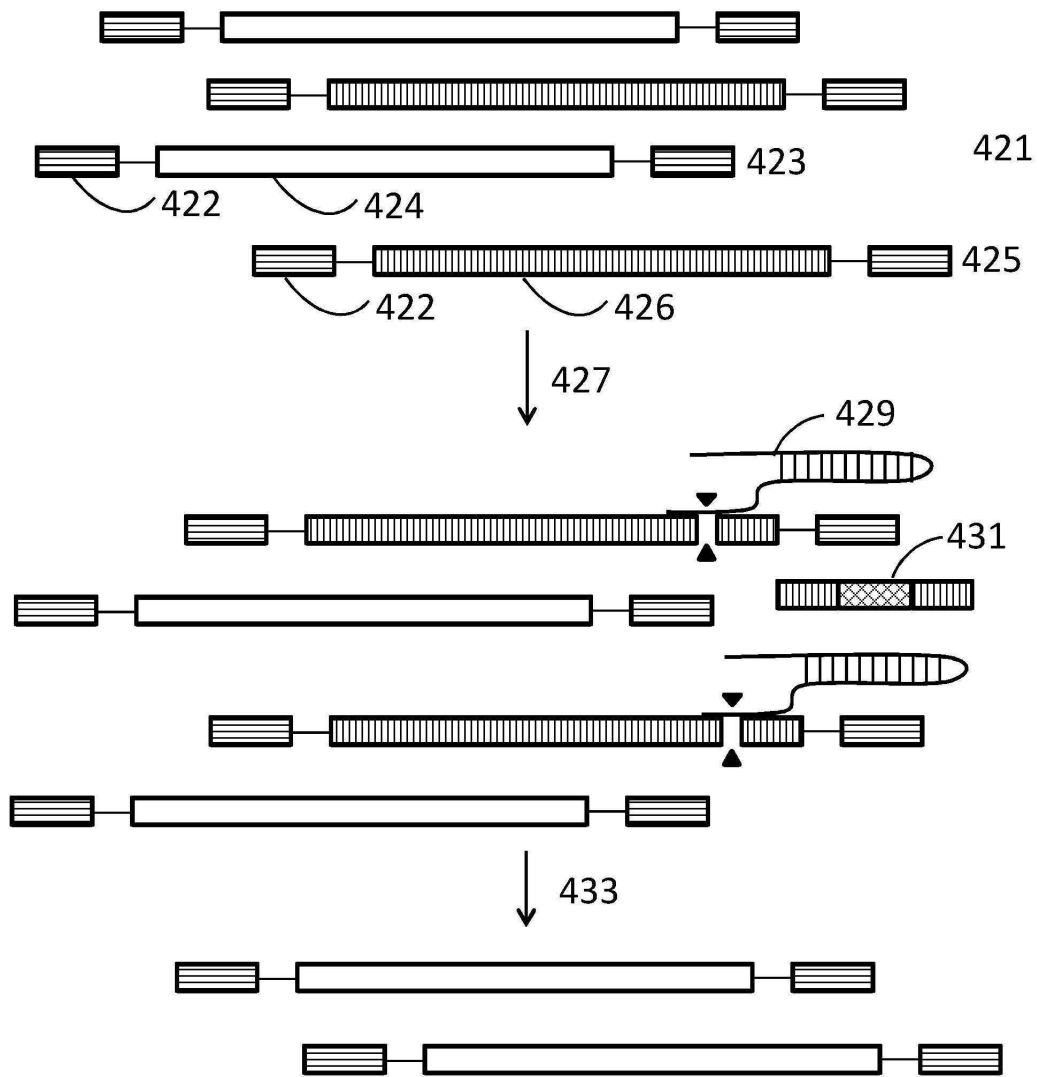
도면3



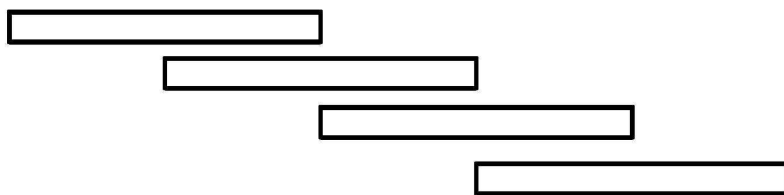
도면4a



도면4b



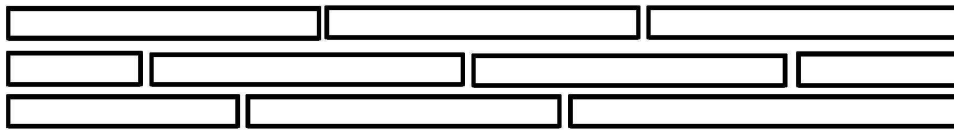
도면5a



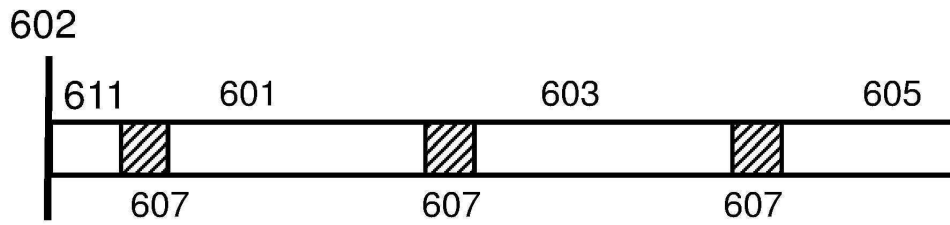
도면5b



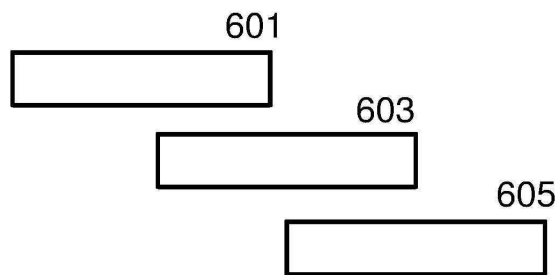
도면5c



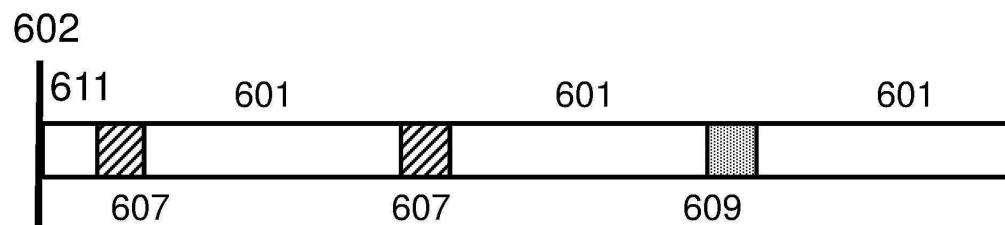
도면6a



도면6b

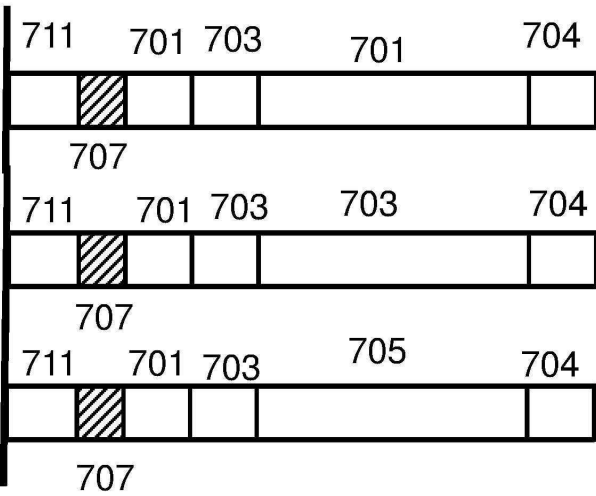


도면6c

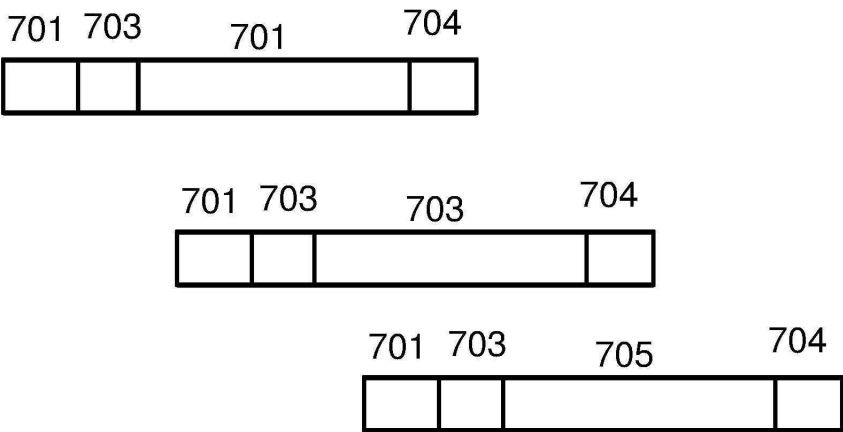


도면7a

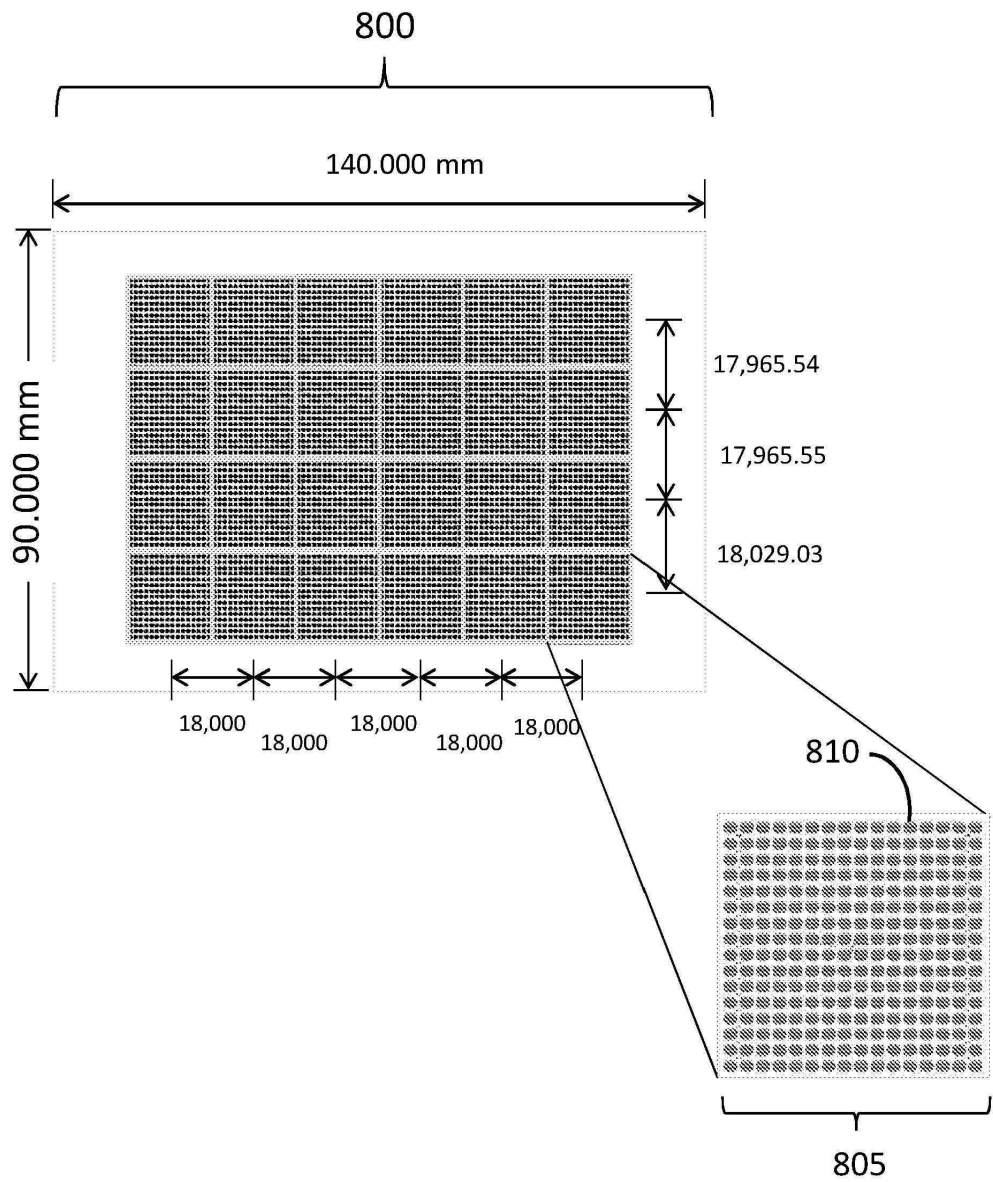
702



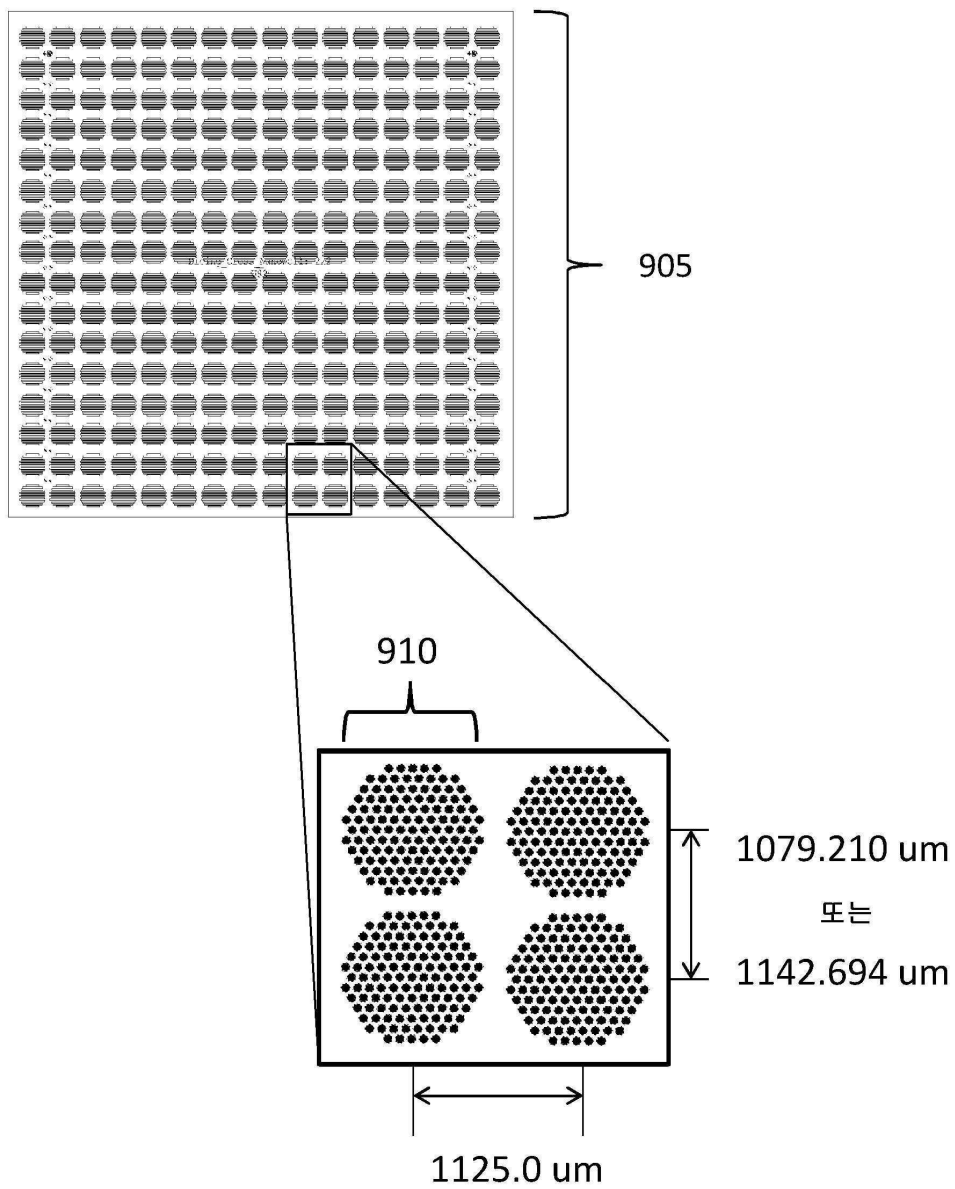
도면7b



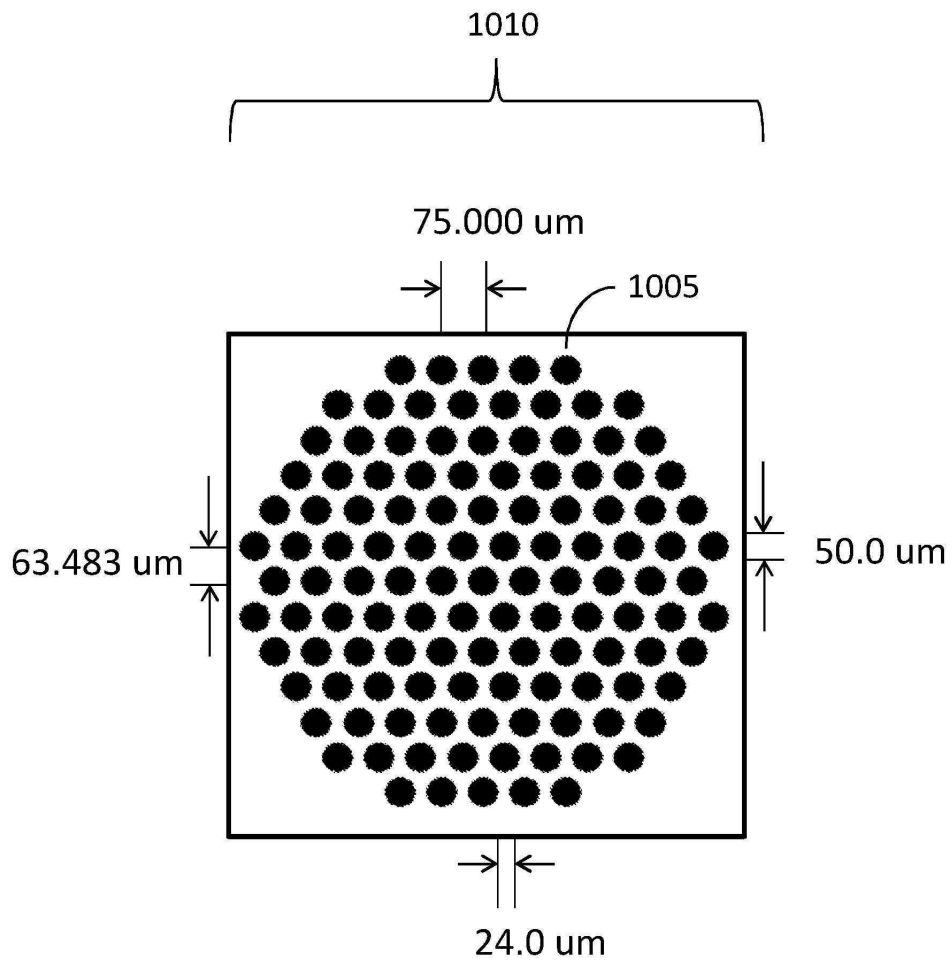
도면8



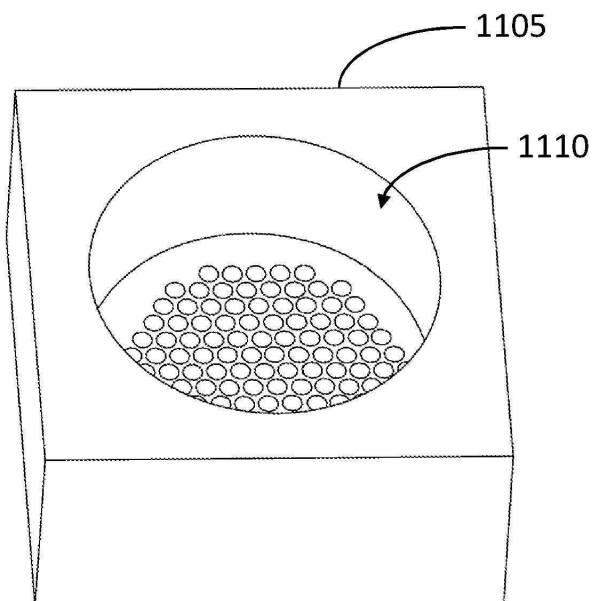
도면9



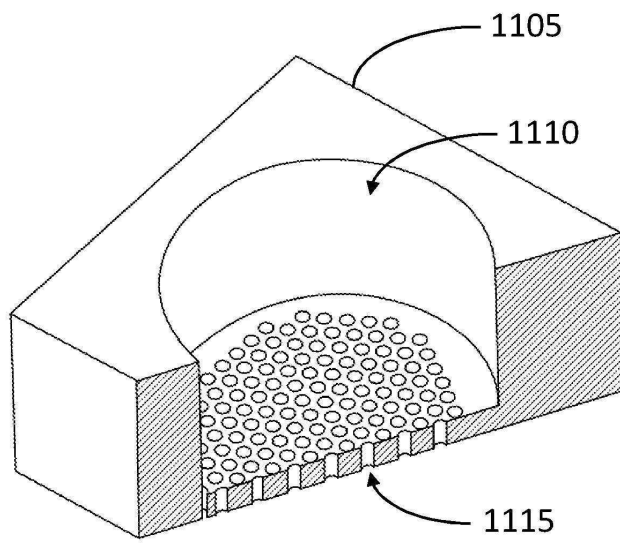
도면10



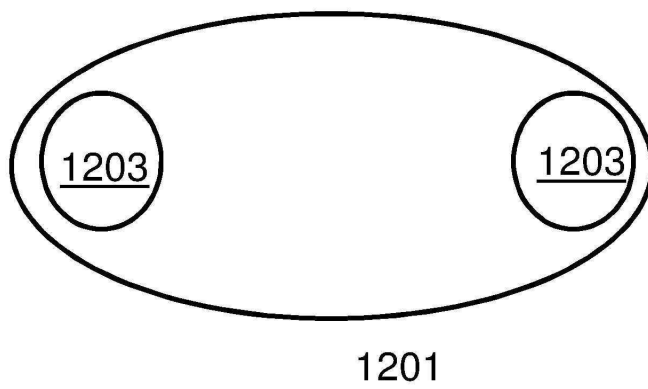
도면11a



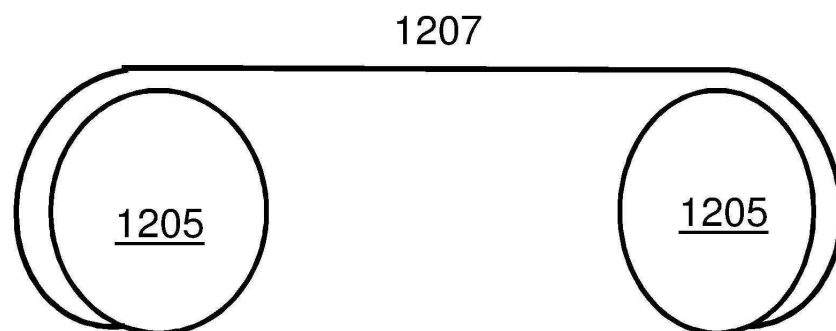
도면11b



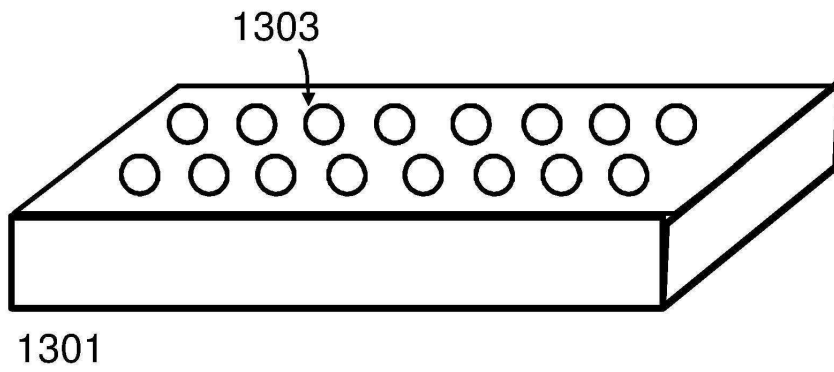
도면12a



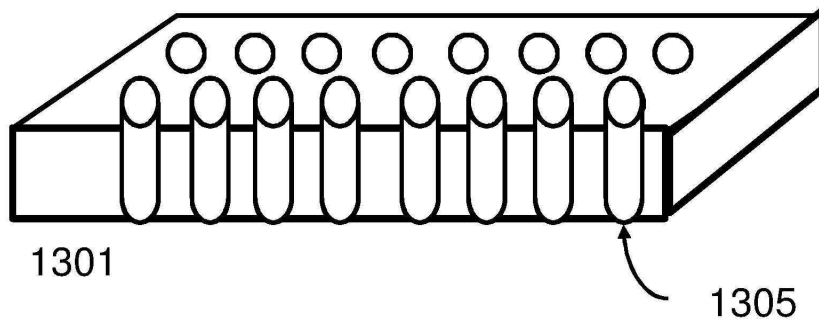
도면12b



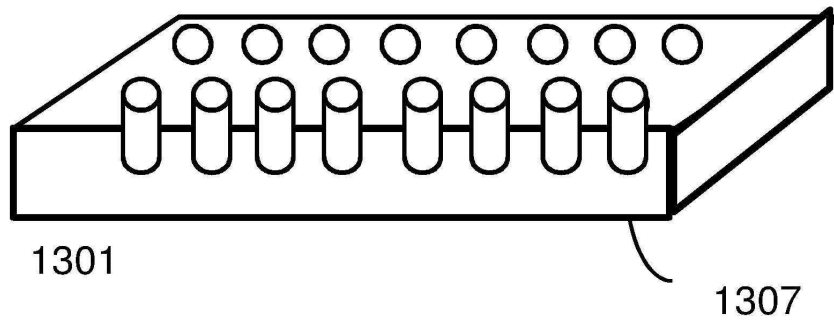
도면13a



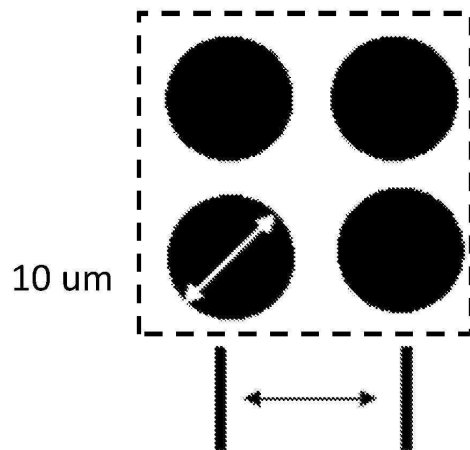
도면13b



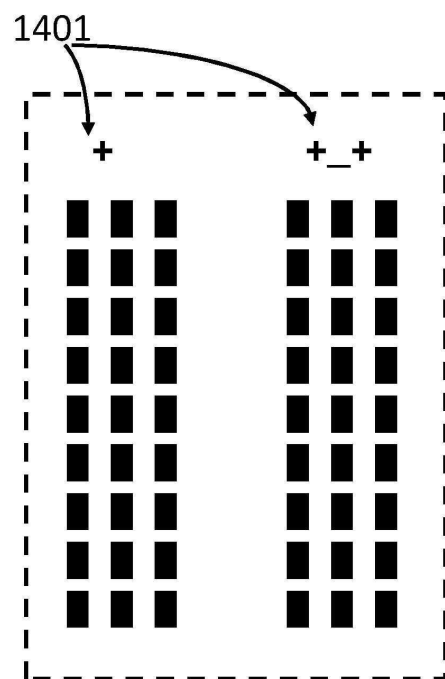
도면13c



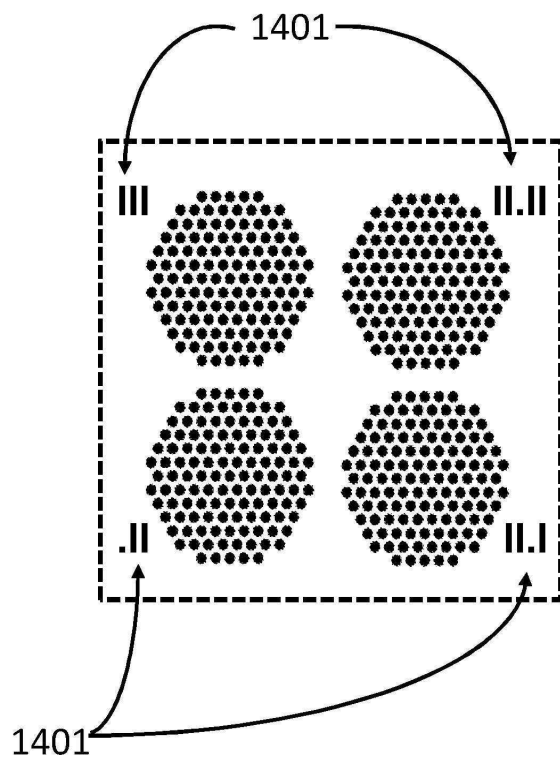
도면14a



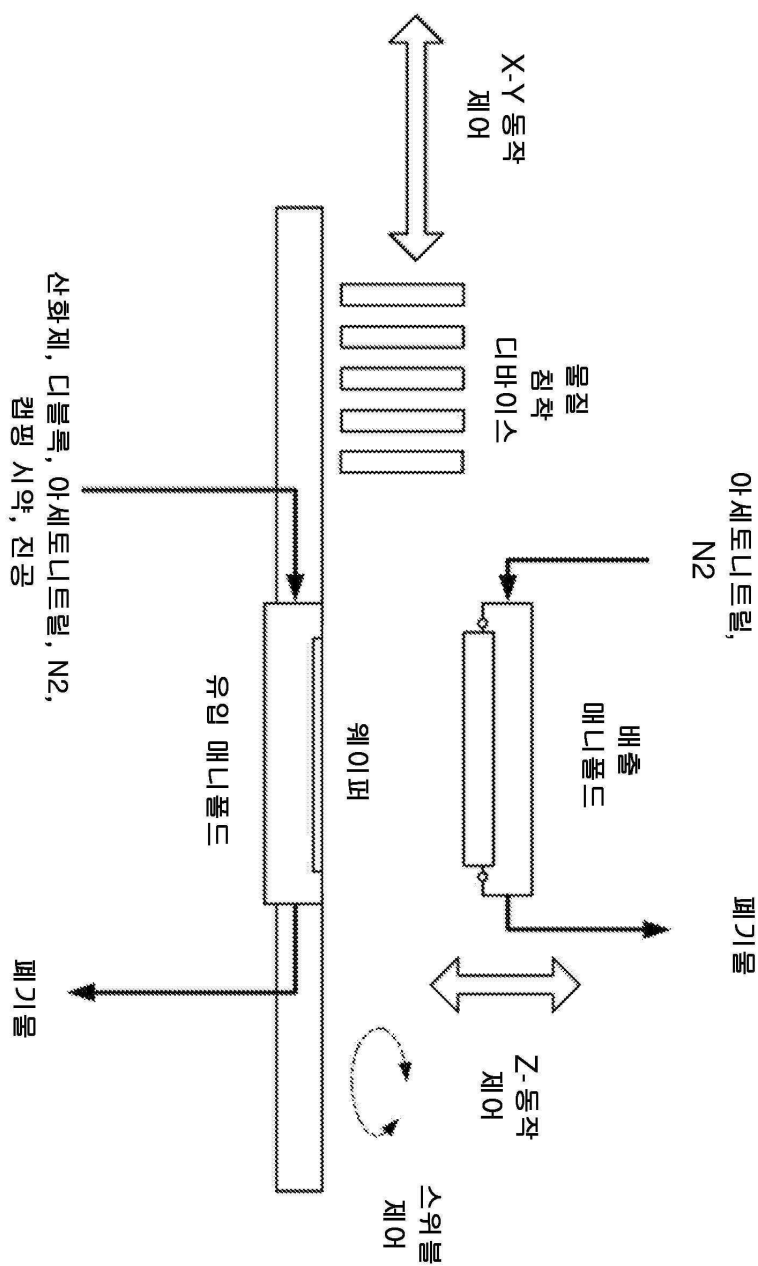
도면14b



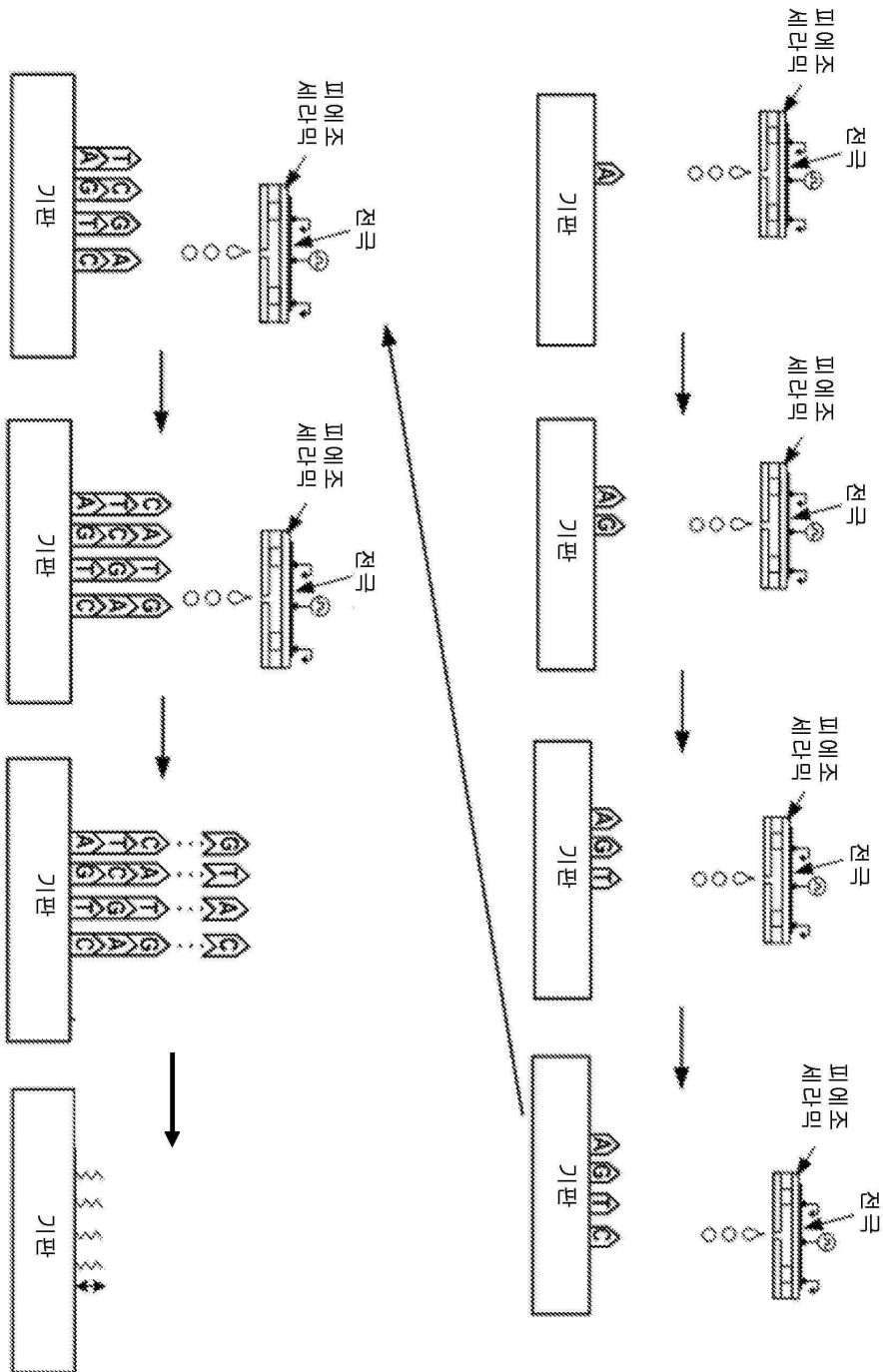
도면14c



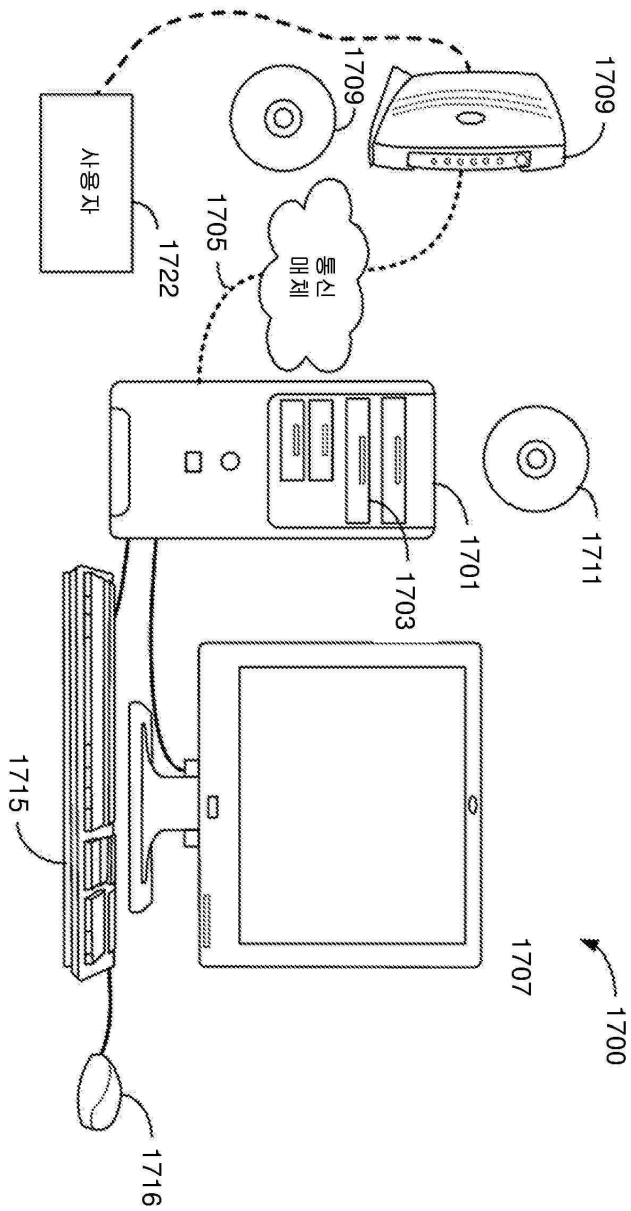
도면15



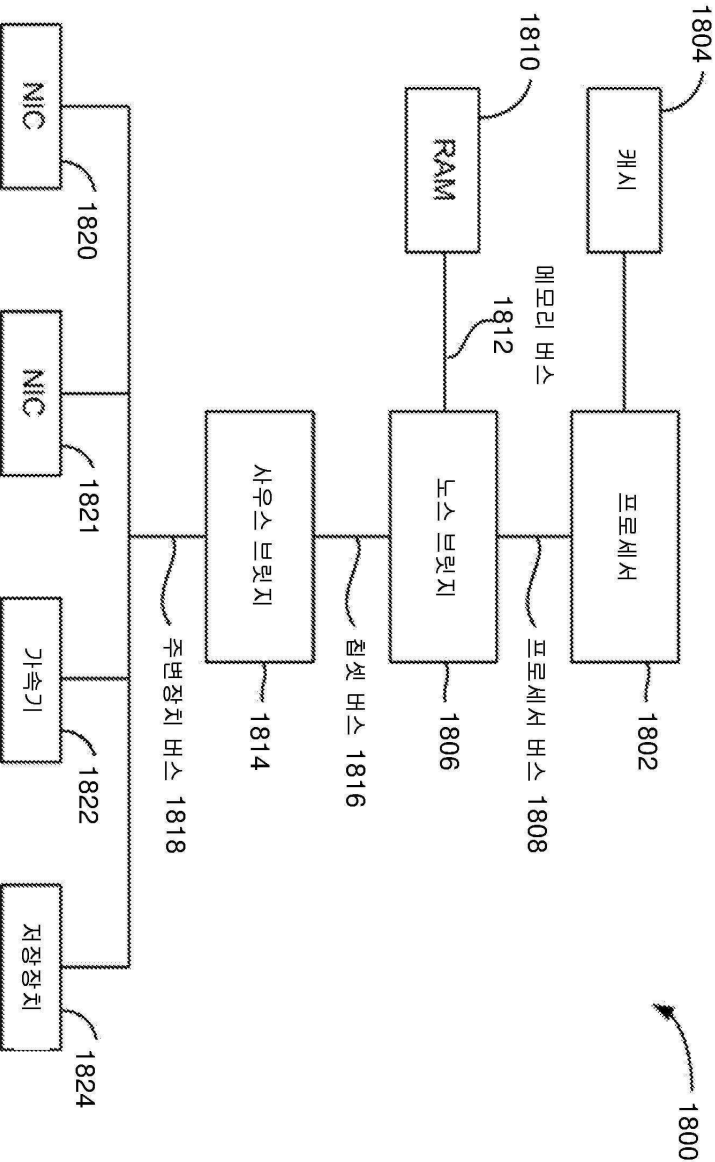
도면16



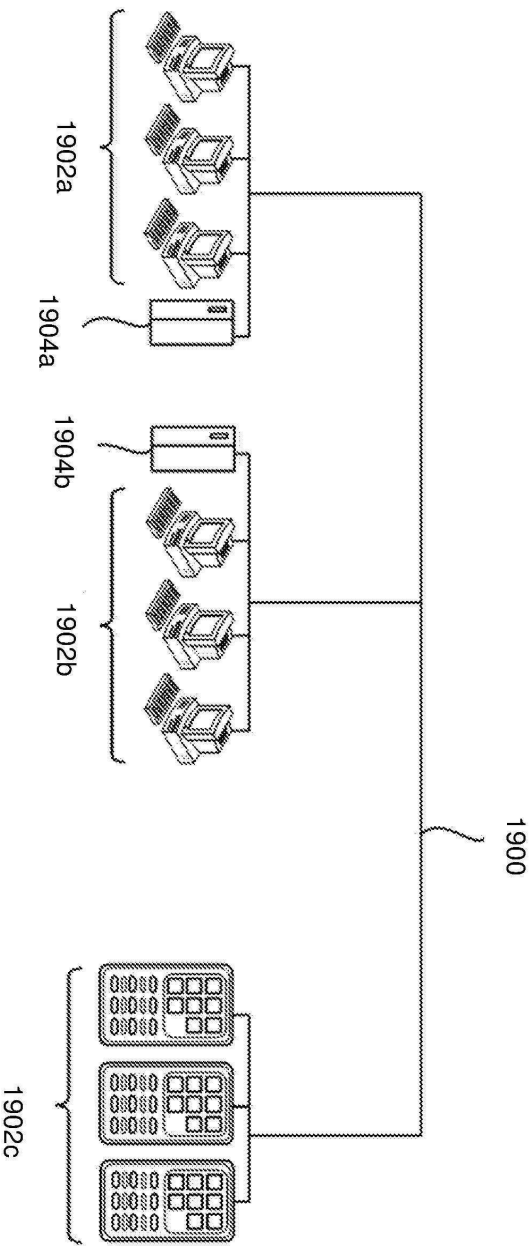
도면17



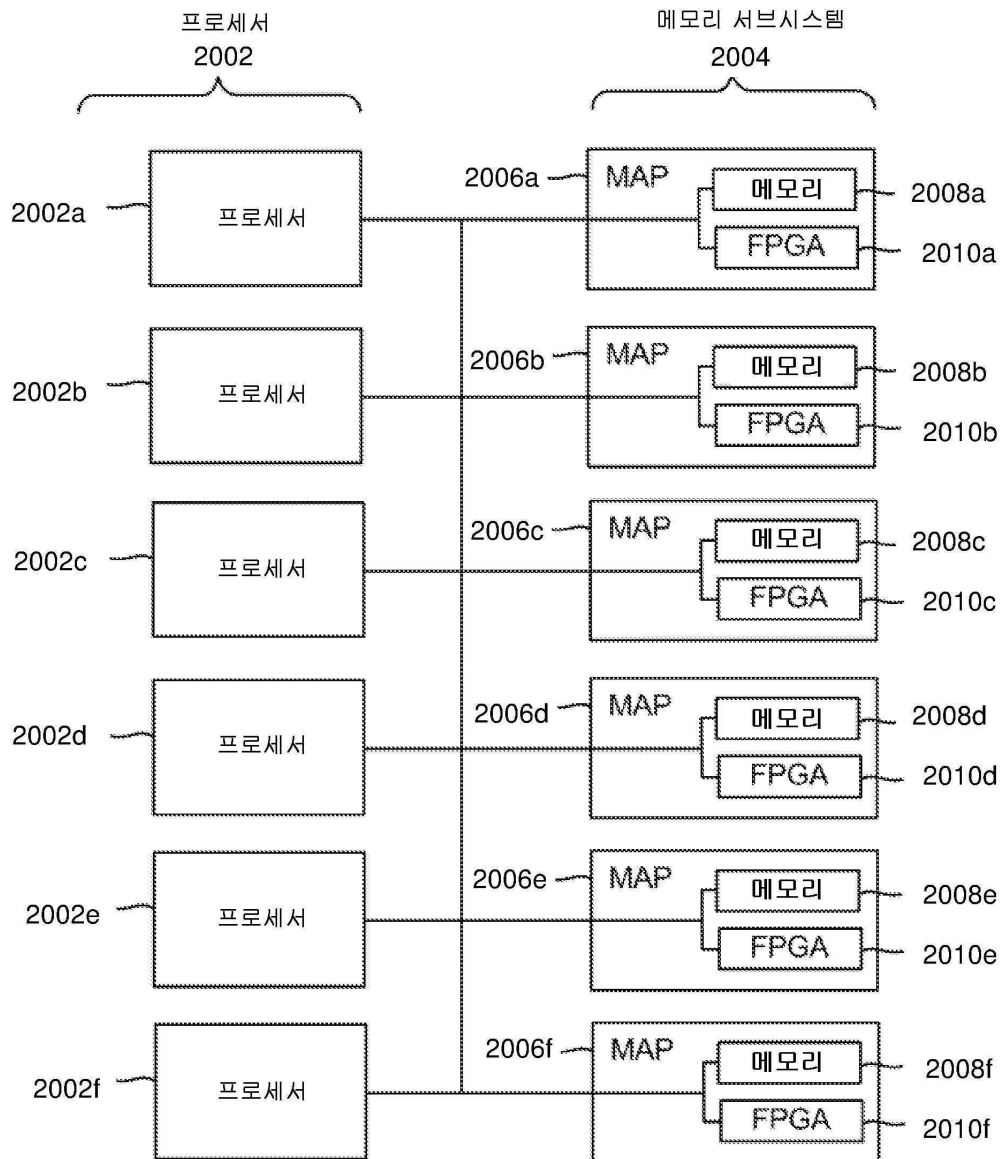
도면18



도면19



도면20



서열 목록

SEQUENCE LISTING

<110> TWIST BIOSCIENCE CORPORATION

<120> NUCLEIC ACID BASED DATA STORAGE

<130> 44854-738.601

<140><141><150> 62/462,284

<151> 2017-02-22

<160> 28

<170> PatentIn version 3.5

<210> 1

<211> 62
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 1
 agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat tttttttttt 60
 tt 62

<210> 2
 <211> 112
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide
 <400> 2
 cgggatacctt atcgatcatcg tcgtacagat cccgacccat ttgctgtcca ccagtcatgc 60
 tagccataacc atgatgatga tgatgatgag aaccccgcat tttttttttt tt 112

<210> 3
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 3
 atgcgggggtt ctcatcatc 19

<210> 4
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 4
 cgggatacctt atcgatcatcg 20

<210> 5
 <211> 11
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220><221> modified_base
 <222> (3)..(6)
 <223> a, c, t, g, unknown or other
 <400> 5
 acnnnnngtay c 11

<210> 6
 <211> 11
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220><221> modified_base
 <222> (3)..(7)
 <223> a, c, t, g, unknown or other
 <400> 6
 acnnnnnctc c 11

<210> 7
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 7
 atctatgtcg ggtgcggaga aagaggtaat 30

<210> 8
 <211> 12
 <212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 8

caannnnngt gg 12

<210> 9

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(7)

<223> a, c, t, g, unknown or other

<400> 9

cacnnnnngtg 10

<210> 10

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(7)

<223> a, c, t, g, unknown or other

<400> 10

caynnnnrtg 10

<210> 11

<211> 15

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(12)

<223> a, c, t, g, unknown or other

<400> 11

ccannnnnnn nntgg

15

<210> 12

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(9)

<223> a, c, t, g, unknown or other

<400> 12

ccannnnnnt gg

12

<210>

13

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 13

ccannnnntg g

11

<210> 14

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (3)..(9)

<223> a, c, t, g, unknown or other

<400> 14

ccnnnnnnng g

11

<210> 15

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 15

cctnnnnnag g

11

<210> 16

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(9)

<223> a, c, t, g, unknown or other

<400> 16

cgannnnnnt gc

12

<210> 17

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(7)

<223> a, c, t, g, unknown or other

<400> 17

gaannnnnttc

10

<210> 18

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(7)

<223> a, c, t, g, unknown or other

<400> 18

gacnnnnngtc

10

<210> 19

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 19

gacnnnnngt c

11

<210> 20

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(9)

<223> a, c, t, g, unknown or other

<400> 20

gacnnnnnng tc

12

<210> 21

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(7)

<223> a, c, t, g, unknown or other

<400> 21

gatnnnnatc

10

<210> 22

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 22

gcannnnntg c

11

<210> 23

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (4)..(8)

<223> a, c, t, g, unknown or other

<400> 23

gccnnnnngg c

11

<210> 24

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (3)..(9)

<223> a, c, t, g, unknown or other

<400> 24

gcnnnnnnng c

11

<210> 25

<211> 13

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><221> modified_base

<222> (5)..(9)

<223> a, c, t, g, unknown or other

<400> 25

ggccnnnnng gcc

13

<210> 26

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 26

taactataac ggtcctaagg tagcgaa

27

<210> 27

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 27

tagggataac agggtaat

18

<210> 28

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 28

tggcaaacag ctattatggg tattatgggt

30