



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0094445  
(43) 공개일자 2017년08월17일

(51) 국제특허분류(Int. Cl.)  
C07H 21/02 (2006.01) C07H 1/00 (2006.01)  
(52) CPC특허분류  
C07H 21/02 (2013.01)  
C07H 1/00 (2013.01)  
(21) 출원번호 10-2017-7019616  
(22) 출원일자(국제) 2015년12월28일  
심사청구일자 없음  
(85) 번역문제출일자 2017년07월14일  
(86) 국제출원번호 PCT/US2015/067667  
(87) 국제공개번호 WO 2016/109423  
국제공개일자 2016년07월07일  
(30) 우선권주장  
62/098,037 2014년12월30일 미국(US)

(71) 출원인  
엑스-켄, 인크.  
미국 02453 매사추세츠주 왈탐 스위트 101 비버  
스트리트 100  
(72) 발명자  
키프, 앤서니, 디.  
미국 02140 매사추세츠주 케임브리지 #6 벨리스  
서클 9  
리토브치크, 알렉산더  
미국 01776 매사추세츠주 서드베리 더튼 로드 468  
(뒷면에 계속)  
(74) 대리인  
양영준, 김영

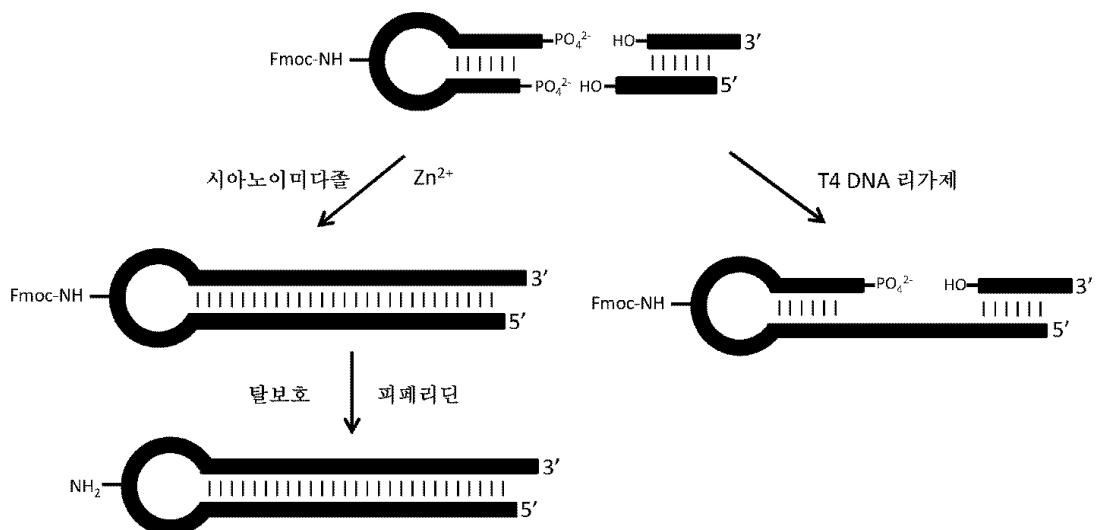
전체 청구항 수 : 총 22 항

(54) 발명의 명칭 DNA-코딩된 라이브러리에 태그부착하는 방법

(57) 요약

본 발명은 코딩된 화학적 엔티티를 생성하는 방법에 관한 것이다. 특히, 올리고뉴클레오타이드 및 방법은 화학적 라이게이션 기술을 통해 형성된 야생형 연결을 갖는 코딩된 화학적 엔티티를 포함할 수 있다. 본 발명은 상대적으로 높은 수율을 제공하는 시아노이미다졸 및  $Zn^{2+}$ 를 사용하여 5'-모노포스포 및 3'-히드록시 말단을 보유하는 올리고뉴클레오타이드 쌍으로부터 및 또한 5'-히드록시 및 3'-모노포스포 말단으로부터 유래된 야생형 연결 (예를 들어, 포스포디에스테르 연결)을 사용하는 올리고뉴클레오타이드 태그부착 전략에 관한 것이다. 이러한 화학적 라이게이션 방법은 주형-의존적이고 직교 3'- 및 5'-포스페이트의 사용을 허용하여, 이중-가닥 맥락으로 포함되는 코딩 올리고뉴클레오타이드의 순차적 라이게이션에 대한 낮은 비율의 오펜입 또는 미스코딩으로의 고도의 제어가 이루어진다.

대표도



(72) 발명자

**클락, 매튜**

미국 02421 매사추세츠주 렉싱턴 클렐랜드 로드 36

**와그너, 리차드 더블유.**

미국 02138 매사추세츠주 케임브리지 쿨리지 애비  
뉴 24

---

## 명세서

### 청구범위

#### 청구항 1

(a) 제1 관능기 및 제2 관능기를 포함하는 헤드피스를 제공하는 단계;

(b) 상기 헤드피스의 상기 제1 관능기를 화학적 엔티티의 성분에 결합시키는 단계이며, 여기서 상기 헤드피스는 상기 성분에 직접적으로 연결되거나 또는 상기 헤드피스는 상기 성분에 이관능성 스페이서에 의해 간접적으로 연결되는 것인 단계;

(c) 상기 헤드피스의 상기 제2 관능기를 화학적 라이게이션을 통해 제1 올리고뉴클레오타이드 태그에 라이게이션시켜 코딩된 화학적 엔티티를 형성하는 단계이며, 여기서 상기 화학적 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 것인 단계

를 포함하는, 코딩된 화학적 엔티티를 생성하는 방법이며;

여기서 단계 (b) 및 (c)는 임의의 순서로 수행될 수 있고, 여기서 상기 제1 올리고뉴클레오타이드 태그는 상기 단계 (b)의 결합 반응을 코딩하며,

이에 의해 코딩된 화학적 엔티티를 생성하는 것인 방법.

#### 청구항 2

제1항에 있어서, 상기 화학적 라이게이션이 포스포디에스테르 연결을 생성하는 것인 방법.

#### 청구항 3

제1항 또는 제2항에 있어서, 상기 헤드피스가 이중-가닥 올리고뉴클레오타이드, 단일-가닥 올리고뉴클레오타이드, 또는 헤어핀 올리고뉴클레오타이드를 포함하는 것인 방법.

#### 청구항 4

제3항에 있어서, 상기 헤드피스가 이중 가닥 올리고뉴클레오타이드 또는 헤어핀 올리고뉴클레오타이드를 포함하는 것인 방법.

#### 청구항 5

제4항에 있어서, 상기 헤드피스가 제3 관능기를 포함하는 것인 방법.

#### 청구항 6

제5항에 있어서, 상기 방법이 (d) 상기 헤드피스의 상기 제3 관능기를 화학적 라이게이션을 통해 제2 올리고뉴클레오타이드 태그에 라이게이션시키는 단계를 추가로 포함하며, 여기서 상기 화학적 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 것인 방법.

#### 청구항 7

제5항에 있어서, 상기 방법이 (d) 상기 헤드피스의 상기 제3 관능기를 제2 올리고뉴클레오타이드 태그에 라이게이션시키는 단계를 추가로 포함하며, 여기서 상기 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 화학적 라이게이션을 통하지 않는 것인 방법.

#### 청구항 8

제2항 내지 제7항 중 어느 한 항에 있어서, 상기 헤드피스가 5'-말단 및/또는 3'-말단에 포스페이트를 포함하는 것인 방법.

#### 청구항 9

제2항 내지 제8항 중 어느 한 항에 있어서, 상기 화학적 라이게이션이 상기 헤드피스 상의 5'- 또는 3'-포스페이트의 5'- 또는 3'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함하는 것인 방법.

#### 청구항 10

제9항에 있어서, 상기 화학적 라이게이션이 상기 헤드피스 상의 5'-포스페이트의 3'-히드록실 올리고뉴클레오타이드에 대한 및/또는 상기 헤드피스 상의 3'-포스페이트의 5'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함하는 것인 방법.

#### 청구항 11

제10항에 있어서, 상기 화학적 라이게이션이 상기 헤드피스 상의 5'-포스페이트의 3'-히드록실 올리고뉴클레오타이드에 대한 및 상기 헤드피스 상의 3'-포스페이트의 5'-히드록실 올리고뉴클레오타이드에 대한 동시 라이게이션을 포함하는 것인 방법.

#### 청구항 12

제8항 내지 제11항 중 어느 한 항에 있어서, 상기 화학적 라이게이션이 시아노이미다졸의 사용을 포함하는 것인 방법.

#### 청구항 13

제12항에 있어서, 상기 화학적 라이게이션이 2가 금속 공급원의 사용을 추가로 포함하는 것인 방법.

#### 청구항 14

제13항에 있어서, 상기 2가 금속 공급원이 가용성  $Zn^{2+}$  공급원인 방법.

#### 청구항 15

제14항에 있어서, 상기 가용성  $Zn^{2+}$  공급원이  $ZnCl_2$ 인 방법.

#### 청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 헤드피스가 상기 성분에 이관능성 스페이서에 의해 간접적으로 연결된 것인 방법.

#### 청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 헤드피스가 상기 성분에 직접적으로 연결된 것인 방법.

#### 청구항 18

제1항 내지 제17항 중 어느 한 항의 방법에 의해 생성된 1종 이상의 화학적 엔티티를 포함하는 라이브러리.

#### 청구항 19

제18항에 있어서, 복수의 헤드피스를 포함하는 라이브러리.

#### 청구항 20

제18항 또는 제19항에 있어서, 각각의 화학적 엔티티가 상이한 것인 라이브러리.

#### 청구항 21

(a) 표적을 제1항 내지 제16항 중 어느 한 항의 방법에 의해 제조된 코딩된 화학적 엔티티 및/또는 제17항 내지 제19항 중 어느 한 항의 라이브러리와 접촉시키는 단계; 및

(b) 대조군과 비교 시 상기 표적에 대한 미리 결정된 특징을 갖는 1종 이상의 코딩된 화학적 엔티티를 선택하며, 이에 의해 복수의 상기 화학적 엔티티를 스크리닝하는 단계

를 포함하는, 복수의 화학적 엔티티를 스크리닝하는 방법.

## 청구항 22

제21항에 있어서, 상기 미리 결정된 특징이 대조군과 비교 시 상기 표적에 대한 증가된 결합을 포함하는 것인 방법.

## 발명의 설명

### 배경 기술

- [0001] DNA-코딩된 화학적 라이브러리 구성원은 코딩 올리고뉴클레오타이드 태그의 조합과 회합된 조합 화학적 합성 프로세스에 의해 생성된 화학적 엔티티이다. 개별 라이브러리 구성원과 회합되는 태그의 조합을 결정하고 이를 사용하여 회합된 라이브러리 구성원의 화학적 합성 히스토리를 추론할 수 있다.
- [0002] 이러한 라이브러리를 생성하는 하나의 방법은 헤드피스 올리고뉴클레오타이드에 대한 올리고뉴클레오타이드 태그의 연속적인 화학적 라이게이션에 의한 것이며, 이 때 화학적으로-생성된 엔티티는 연속적인 분할-밋-혼합 단계에 의해 디스플레이된다. 각각의 분할 단계 동안 화학적 합성 단계는 올리고뉴클레오타이드 라이게이션 단계와 함께 수행된다.
- [0003] 효소-매개되기보다는, 화학적인 올리고뉴클레오타이드 라이게이션 단계는 용액 조건과 관련하여 더 큰 가요성을 허용하고, 잠재적으로 수천개의 저-부피 개별 분리된 구획을 위해 필요한 완충제 교환 단계를 감소시킬 수 있다.
- [0004] 그러나, 화학적 라이게이션 반응으로부터 생성된 대부분의 올리고뉴클레오타이드 연결 구조는 폴리머라제에 의해 전위될 수 없는 연결을 발생시킨다. 이는 이러한 연결이 서열분석과 같이 개별 라이브러리 구성원을 디코딩하는데 폴리머라제를 사용하는 프로세스에 직접적으로 이용될 수 없다는 것을 의미한다.
- [0005] 본 발명은 화학적 라이게이션 기술을 이용하여 DNA-코딩된 화학적 엔티티를 야생형 연결에 의해 태그부착하는 방법에 관한 것이다. 이는 폴리머라제에 의해 판독가능한 연결의 편의성을 유지하면서 화학적 라이게이션의 이점이 실현되도록 한다.

### 발명의 내용

- [0006] 태그 서열 및 회합 정보를 직접 복원하는 폴리머라제의 능력을 또한 보유하면서, 화학적 히스토리를 코딩하기 위한 수단으로서 화학적 라이게이션을 동시에 사용하는 이용가능한 하나의 전략은 야생형 포스포디에스테르 연결을 생성하는 방식으로 화학적 라이게이션을 수행하는 것이다. 이러한 방법은 일반적으로 이중-가닥 또는 주형화된 맥락으로 5'-포스페이트 및 3'-히드록실 올리고뉴클레오타이드와 함께 축합제 예컨대 브로민화시아노젠 등을 이용한다. 유사하게 브로민화시아노젠은 또한 5'-히드록실 및 3'-포스페이트인 기질 올리고뉴클레오타이드의 쌍을 화학적으로 라이게이션하는 것으로 제시된 바 있다. 그러나, 이들 방법은 DNA-코딩된 라이브러리를 태그 부착하는 것과 같은 반복적 프로세스에 그들을 사용하는 것을 부적합하게 하는 불량한 효율을 겪는다.
- [0007] 본 발명자들은 상대적으로 높은 수율로, 시아노이미다졸 및  $Zn^{2+}$ 를 사용하여 5'-모노포스포 및 3'-히드록시 말단을 보유하는 올리고뉴클레오타이드 쌍으로부터 및 또한 5'-히드록시 및 3'-모노포스포 말단으로부터 유래된 야생형 연결 (예를 들어, 포스포디에스테르 연결)을 사용하는 올리고뉴클레오타이드 태그부착 전략을 개발한 바 있다. 이러한 화학적 라이게이션 방법은 주형-의존성이고 직교 3'- 및 5'-포스페이트의 사용을 허용하기 때문에, 이중-가닥 맥락으로 포함되는 코딩 올리고뉴클레오타이드의 순차적 라이게이션에 대한 낮은 비율의 오펜업 또는 미스코딩으로의 고도의 제어가 발휘될 수 있다.
- [0008] 따라서, 제1 측면에서, 본 발명은 코딩된 화학적 엔티티를 생성하는 방법을 특색으로 한다. 이러한 방법은 (a) 제1 관능기 및 제2 관능기를 포함하는 헤드피스를 제공하는 단계; (b) 헤드피스의 제1 관능기를 화학적 엔티티의 성분에 결합시키는 단계이며, 여기서 헤드피스는 성분에 직접적으로 연결되거나 또는 헤드피스는 성분에 이 관능성 스페이서에 의해 간접적으로 연결되는 것인 단계; 및 (c) 헤드피스의 제2 관능기를 화학적 라이게이션을 통해 제1 올리고뉴클레오타이드 태그에 라이게이션시켜 코딩된 화학적 엔티티를 형성하는 단계이며, 여기서 화학적 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포티오에이트 연결을 생성하는 것인 단계를 포함하며; 여기서 단계 (b) 및 (c)는 임의의 순서로 수행될 수 있고, 여기서 제1 올리고뉴클레오타이드 태그는 단계

(b)의 결합 반응을 코딩하며, 이에 의해 코딩된 화학적 엔티티를 생성한다.

- [0009] 또 다른 측면에서, 본 발명은 코딩된 화학적 엔티티를 생성하는 추가의 방법을 특색으로 한다. 이러한 방법은 (a) 제1 관능기 및 제2 관능기를 포함하는 헤드피스를 제공하는 단계; (b) 헤드피스의 제1 관능기를 화학적 엔티티의 성분에 결합시키는 단계이며, 여기서 헤드피스는 성분에 직접적으로 연결되거나 또는 헤드피스는 성분에 이관능성 스페이서에 의해 간접적으로 연결되는 것인 단계; (c) 헤드피스의 제2 관능기를 화학적 라이게이션을 통해 제1 올리고뉴클레오타이드 태그에 라이게이션시켜 복합체를 형성하는 단계이며, 여기서 화학적 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 것인 단계; (d) 코딩된 화학적 엔티티의  $n_c$ 개의 추가의 성분을 결합시키는 단계이며, 여기서  $n_c$ 는 1 내지 10의 정수인 단계; 및 (e)  $n_t$ 개의 연결을 갖는  $n_t$ 개의 추가의 올리고뉴클레오타이드 태그를 라이게이션시켜 코딩된 화학적 엔티티를 형성하는 단계이며, 여기서  $n_t$ 는 1 내지 10의 정수이고, 여기서 각각의 연결은 2개의 인접한 태그 사이에 존재하고 각각의 태그는 성분 중 적어도 1종의 아이덴티티를 코딩하는 것인 단계를 포함하며; 여기서 단계 (b) 및 (c)는 임의의 순서로 수행될 수 있고, 여기서 제1 올리고뉴클레오타이드 태그는 단계 (b)의 결합 반응을 코딩하고; 여기서 단계 (d) 및 (e)는 임의의 순서로 수행될 수 있고, 여기서 각각의 추가의 태그는 단계 (d)의 각각의 추가의 성분의 결합 반응을 코딩하며, 이에 의해 코딩된 화학적 엔티티를 생성한다.
- [0010] 일부 실시양태에서,  $n_t$ 개의 연결 중 적어도 1개의 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 화학적 라이게이션을 통하지 않는다 (예를 들어,  $n_t$ 개의 연결 중 적어도 1개의 라이게이션은 판독가능하거나 또는 판독불가능한 연결을 생성하는 효소적 라이게이션 또는 화학적 라이게이션을 포함).
- [0011] 일부 실시양태에서,  $n_c$  및  $n_t$ 는 각각 독립적으로 1 내지 2, 1 내지 3, 1 내지 4, 1 내지 5, 1 내지 6, 1 내지 7, 1 내지 8, 1 내지 9, 1 내지 10, 2 내지 3, 2 내지 4, 2 내지 5, 2 내지 6, 2 내지 7, 2 내지 8, 2 내지 9, 2 내지 10, 3 내지 4, 3 내지 5, 3 내지 6, 3 내지 7, 3 내지 8, 3 내지 9, 3 내지 10, 4 내지 5, 4 내지 6, 4 내지 7, 4 내지 8, 4 내지 9, 4 내지 10의 정수이다. 특정 실시양태에서,  $n_c$ 는 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10이다. 일부 실시양태에서,  $n_t$ 는 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10이다.
- [0012] 일부 실시양태에서, 화학적 라이게이션은 포스포디에스테르 연결을 생성한다. 특정 실시양태에서, 화학적 라이게이션은 포스포네이트 연결을 생성한다. 일부 실시양태에서, 화학적 라이게이션은 포스포로티오에이트 연결을 생성한다.
- [0013] 일부 실시양태에서, 헤드피스는 이중-가닥 올리고뉴클레오타이드, 단일-가닥 올리고뉴클레오타이드 또는 헤어핀 올리고뉴클레오타이드를 포함한다. 특정 실시양태에서, 헤드피스는 이중 가닥 올리고뉴클레오타이드 또는 헤어핀 올리고뉴클레오타이드를 포함한다.
- [0014] 일부 실시양태에서, 헤드피스는 제3 관능기를 포함한다. 특정 실시양태에서, 방법은 (d) 헤드피스의 제3 관능기를 화학적 라이게이션을 통해 제2 올리고뉴클레오타이드 태그에 라이게이션시키는 단계를 추가로 포함하며, 여기서 화학적 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성한다.
- [0015] 일부 실시양태에서, 방법은 (d) 상기 헤드피스의 상기 제3 관능기를 제2 올리고뉴클레오타이드 태그에 라이게이션시키는 단계를 추가로 포함하며, 여기서 상기 라이게이션은 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 화학적 라이게이션을 통하지 않는다 (예를 들어, 라이게이션은 판독가능하거나 또는 판독불가능한 연결을 생성하는 효소적 라이게이션 또는 화학적 라이게이션임).
- [0016] 특정 실시양태에서, 헤드피스는 5'-말단 및/또는 3'-말단에 포스페이트를 포함한다 (예를 들어, 헤드피스는 5'-말단, 3'-말단에 포스페이트를 포함하거나, 또는 헤드피스가 이중 가닥 또는 헤어핀 올리고뉴클레오타이드인 경우에, 헤드피스는 임의로 5'-말단 및 3'-말단 둘 다에 포스페이트를 포함함).
- [0017] 일부 실시양태에서, 화학적 라이게이션은 상기 헤드피스 상의 5'- 또는 3'-포스페이트의 5'- 또는 3'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함한다. 일부 실시양태에서, 화학적 라이게이션은 헤드피스의 5'-말단의 포스페이트의 5'-히드록실 올리고뉴클레오타이드 또는 3'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함한다. 특정 실시양태에서, 화학적 라이게이션은 헤드피스의 3'-말단의 포스페이트의 5'-히드록실 올리고뉴클레오타이드 또는 3'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함한다.

- [0018] 일부 실시양태에서, 화학적 라이게이션은 상기 헤드피스 상의 5'-포스페이트의 3'-히드록실 올리고뉴클레오타이드에 대한 및/또는 상기 헤드피스 상의 3'-포스페이트의 5'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함한다. 일부 실시양태에서, 화학적 라이게이션은 헤드피스의 5'-말단의 포스페이트의 3'-히드록실 올리고뉴클레오타이드에 대한 라이게이션 및 헤드피스의 3'-말단의 포스페이트의 5'-히드록실 올리고뉴클레오타이드에 대한 라이게이션을 포함한다.
- [0019] 특정 실시양태에서, 화학적 라이게이션은 상기 헤드피스 상의 5'-포스페이트의 3'-히드록실 올리고뉴클레오타이드에 대한 및 상기 헤드피스 상의 3'-포스페이트의 5'-히드록실 올리고뉴클레오타이드에 대한 동시 라이게이션을 포함한다.
- [0020] 일부 실시양태에서, 화학적 라이게이션은 시아노이미다졸의 사용을 포함한다. 특정 실시양태에서, 화학적 라이게이션은 2가 금속 공급원 (예를 들어, 가용성 2가 금속 공급원) 예컨대  $Zn^{2+}$  공급원 (예를 들어, 임의의 가용성  $Zn^{2+}$  공급원 예컨대  $ZnF_2$ ,  $ZnCl_2$ ,  $ZnBr_2$ ,  $ZnI_2$ ,  $Zn(NO_3)_2$ ,  $Zn(ClO_3)_2$ ,  $ZnSO_4$ , 또는  $Zn(O_2CCH_3)_2$  또는 계내 산화된 원소상 아연),  $Mn^{2+}$  공급원 (예를 들어, 임의의 가용성  $Mn^{2+}$  공급원 예컨대  $MnSO_4$ , 또는  $MnCl_2$ ) 또는  $Co^{2+}$  공급원 (예를 들어, 임의의 가용성  $Co^{2+}$  공급원 예컨대  $CoF_2$ ,  $CoCl_2$ ,  $CoBr_2$ , 또는  $CoI_2$ )의 사용을 추가로 포함한다.
- [0021] 일부 실시양태에서, 헤드피스는 화학적 엔티티의 성분에 이관능성 스페이스에 의해 간접적으로 연결된다 (예를 들어,  $C_{1-10}$  알킬, 1 내지 10개의 원자의 헤테로알킬,  $C_{2-10}$  알케닐,  $C_{2-10}$  알키닐,  $C_{5-10}$  아릴, 3 내지 20개의 원자의 시클릭 또는 폴리시클릭계, 포스포디에스테르, 펩티드, 올리고사카라이드, 올리고뉴클레오타이드, 올리고머, 중합체, 또는 폴리알킬 글리콜 예컨대 폴리에틸렌 글리콜, 예컨대  $-(CH_2CH_2O)_nCH_2CH_2-$  (여기서  $n$ 은 1 내지 50의 정수임)를 포함한 선형 또는 분지형 체).
- [0022] 특정 실시양태에서, 헤드피스는 코딩된 화학적 엔티티의 성분에 직접적으로 연결된다.
- [0023] 특정 실시양태에서, 화학적 엔티티는 1종 이상의 제1 라이브러리-확인 태그(들), 사용 태그(들), 및/또는 기원 태그(들)를 추가로 포함한다.
- [0024] 일부 실시양태에서, 화학적 엔티티는 2 내지 20개의 태그 (예를 들어, 2 내지 17개의 빌딩 블록 또는 스캐폴드 태그, 1개의 제1 라이브러리-확인 태그, 1개의 임의적인 사용 태그, 및 1개의 기원 태그)를 포함한다. 일부 실시양태에서, 각각의 태그는 1 내지 75개의 뉴클레오타이드로부터 포함한다 (예를 들어, 본원에 기재된 바와 같음, 예컨대 약 6 내지 12개의 뉴클레오타이드). 특정 실시양태에서, 개별 태그 세트 내의 각각의 태그는 거의 동일한 질량을 포함한다.
- [0025] 일부 실시양태에서, 코딩된 화학적 엔티티는 RNA, DNA, 변형된 DNA, 및/또는 변형된 RNA를 포함한다. 특정 실시양태에서, 변형된 DNA 또는 변형된 RNA는 동일한 올리고뉴클레오타이드 내의 PNA, LNA, GNA, TNA, 또는 그의 혼합물이다.
- [0026] 특정 실시양태에서, 코딩된 화학적 엔티티는 가역적 고정화를 위한 부위를 포함한다. 일부 실시양태에서, 가역적 고정화를 위한 부위는 적어도 1개의 결합 단계 후에 고정되고, 후속 결합 단계 전에 방출된다. 일부 실시양태에서, 가역적 고정화를 위한 부위는 복수의 결합된 단계 후에 고정되고, 후속 결합 단계 전에 방출된다.
- [0027] 일부 실시양태에서, 가역적 고정화를 위한 부위는 결합 쌍의 하나의 구성원, 예를 들어, 핵산, 예컨대 혼성화-적격 올리고뉴클레오타이드 (예를 들어, 혼성화-적격 단일-가닥 올리고뉴클레오타이드), 펩티드, 또는 소분자를 포함한다.
- [0028] 또 다른 측면에서, 본 발명은 임의의 상기 방법에 의해 생성된 1종 이상의 화학적 엔티티를 포함하는 라이브러리를 특색으로 한다.
- [0029] 특정 실시양태에서, 라이브러리는 복수의 헤드피스를 포함한다. 일부 실시양태에서, 복수의 헤드피스의 각각의 헤드피스는 동일한 서열 영역 (예를 들어, 프라이머-결합 영역) 및 상이한 코딩 영역 (예를 들어, 라이브러리의 사용, 라이브러리의 기원, 라이브러리의 아이덴티티, 라이브러리의 히스토리, 연결, 스페이스, 또는 제1 성분의 추가를 코딩하는 제1 태그; 또는 혼성화, 증폭, 클로닝 또는 서열분석 기술을 용이하게 하는 올리고뉴클레오타이드 서열)을 포함한다.



- [0030] 특정 실시양태에서, 라이브러리는 약  $10^2$  내지  $10^{20}$ 개의 화학적 엔티티 (예를 들어, 약  $10^2$  내지  $10^3$ ,  $10^2$  내지  $10^4$ ,  $10^2$  내지  $10^5$ ,  $10^2$  내지  $10^6$ ,  $10^2$  내지  $10^7$ ,  $10^2$  내지  $10^8$ ,  $10^2$  내지  $10^9$ ,  $10^2$  내지  $10^{10}$ ,  $10^2$  내지  $10^{11}$ ,  $10^2$  내지  $10^{12}$ ,  $10^2$  내지  $10^{13}$ ,  $10^2$  내지  $10^{14}$ ,  $10^2$  내지  $10^{15}$ ,  $10^2$  내지  $10^{16}$ ,  $10^2$  내지  $10^{17}$ ,  $10^2$  내지  $10^{18}$ ,  $10^2$  내지  $10^{19}$ ,  $10^4$  내지  $10^5$ ,  $10^4$  내지  $10^6$ ,  $10^4$  내지  $10^7$ ,  $10^4$  내지  $10^8$ ,  $10^4$  내지  $10^9$ ,  $10^4$  내지  $10^{10}$ ,  $10^4$  내지  $10^{11}$ ,  $10^4$  내지  $10^{12}$ ,  $10^4$  내지  $10^{13}$ ,  $10^4$  내지  $10^{14}$ ,  $10^4$  내지  $10^{15}$ ,  $10^4$  내지  $10^{16}$ ,  $10^4$  내지  $10^{17}$ ,  $10^4$  내지  $10^{18}$ ,  $10^4$  내지  $10^{19}$ ,  $10^4$  내지  $10^{20}$ ,  $10^5$  내지  $10^6$ ,  $10^5$  내지  $10^7$ ,  $10^5$  내지  $10^8$ ,  $10^5$  내지  $10^9$ ,  $10^5$  내지  $10^{10}$ ,  $10^5$  내지  $10^{11}$ ,  $10^5$  내지  $10^{12}$ ,  $10^5$  내지  $10^{13}$ ,  $10^5$  내지  $10^{14}$ ,  $10^5$  내지  $10^{15}$ ,  $10^5$  내지  $10^{16}$ ,  $10^5$  내지  $10^{17}$ ,  $10^5$  내지  $10^{18}$ ,  $10^5$  내지  $10^{19}$ , 또는  $10^5$  내지  $10^{20}$ 개의 복합체)를 포함한다. 라이브러리의 특정 실시양태에서, 각각의 화학적 엔티티는 상이하다.
- [0031] 또 다른 측면에서, 본 발명은 복수의 코딩된 화학적 엔티티를 스크리닝하는 방법을 특색으로 한다. 이 방법은 (a) 표적을 임의의 상기 방법에 의해 제조된 코딩된 화학적 엔티티 및/또는 임의의 상기 라이브러리와 접촉시키는 단계; 및 (b) 대조군과 비교 시 상기 표적에 대한 미리 결정된 특징을 갖는 1종 이상의 코딩된 화학적 엔티티를 선택하며, 이에 의해 복수의 코딩된 화학적 엔티티를 스크리닝하는 단계를 포함한다.
- [0032] 일부 실시양태에서, 미리 결정된 특징은 대조군과 비교 시 표적에 대한 증가된 결합을 포함한다. 특정 실시양태에서, 미리 결정된 특징은 대조군과 비교 시 표적의 증가된 억제력을 포함한다. 일부 실시양태에서, 미리 결정된 특징은 대조군과 비교 시 표적의 증가된 활성을 포함한다.
- [0033] 임의의 상기 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스, 제1 태그 및/또는 1종 이상의 추가의 태그, 존재하는 경우)는 라이브러리의 아이덴티티를 코딩한다. 일부 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스, 제1 태그 및/또는 1종 이상의 추가의 태그, 존재하는 경우)는 제1 라이브러리-확인 서열을 포함하며, 여기서 서열은 제1 라이브러리의 아이덴티티를 코딩한다. 특정한 실시양태에서, 올리고뉴클레오타이드는 제1 라이브러리-확인 태그이다. 일부 실시양태에서, 방법은 제1 라이브러리-확인 태그를 제공하는 단계이며, 여기서 태그는 제1 라이브러리를 코딩하는 서열을 포함하는 것인 단계, 및/또는 제1 라이브러리-확인 태그를 복합체에 결합시키는 단계를 포함한다. 일부 실시양태에서, 방법은 제2 라이브러리를 제공하는 단계 및 제1 라이브러리를 제2 라이브러리와 조합하는 단계를 포함한다. 추가 실시양태에서, 방법은 제2 라이브러리-확인 태그를 제공하는 단계를 포함하며, 여기서 태그는 제2 라이브러리를 코딩하는 서열을 포함한다. 일부 실시양태에서, 2종 초과 라이브러리가 조합된다 (예를 들어, 3, 4, 5, 6, 7, 8, 9, 10종 또는 그 초과 라이브러리).
- [0034] 임의의 상기 실시양태에서, 코딩된 정보는 1종 이상의 태그 또는 1종 초과 태그의 조합으로 제공된다. 일부 실시양태에서, 코딩된 정보는 1종 초과 태그 (예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10종 또는 그 초과 태그)에 의해 나타내어진다. 일부 실시양태에서, 코딩된 정보는 1종 초과 태그에 의해 나타내어지며, 여기서 모든 코딩 태그는 코딩 서열 내에 함유된다 (예를 들어, 정보를 코딩하기 위해 특정 태그 조합을 사용하는 것에 의함). 일부 실시양태에서, 코딩된 정보는 1종 초과 태그에 의해 나타내어지며, 여기서 모든 코딩 태그 미만이 코딩 서열 내에 함유된다 (예를 들어, 개별 코딩 서열 내에 코딩하기 위해 1종 초과 개별 태그의 한 세트로부터 1종의 태그를 사용하는 것에 의함). 일부 실시양태에서, 코딩된 정보는 직교형으로 나타내어지며, 여기서 코딩된 정보는 1종 초과 태그와 개별 라이브러리 구성원 내에 함유되어 있는 모든 코딩 정보 미만의 조합에 의해 나타내어지고, 따라서 1종 초과 상응하는 라이브러리 구성원은 코딩된 정보의 디컨볼루션을 위해 서열분석될 필요가 있다. 일부 실시양태에서, 1종 초과 화학적 빌딩 블록은 단일 태그에 의해 나타내어진다 (예를 들어, 라세미 빌딩 블록의 경우에, 예컨대 2, 3, 4, 5, 6, 7, 8, 9, 10종 또는 그 초과 빌딩 블록이 단일 태그에 의해 나타내어짐).
- [0035] 임의의 상기 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스 및/또는 1종 이상의 빌딩 블록)는 라이브러리의 구성원의 사용 (예를 들어, 본원에 기재된 바와 같은 선택 단계 또는 결합 단계에서의 사용)을 코딩한다. 일부 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스, 제1 태그 및/또는 1종 이상의 추가의 태그, 존재하는 경우)는 사용 서열을 포함하며, 여기서 서열은 1개 이상의 단계 (예를 들어, 선택 단계 및/또는 결합 단계)에서 라이브러리 내의 구성원의 하위세트의 사용을 코딩한다. 특정한 실시양태에서, 올리고뉴클레오타이드는 사용 서열을 포함하는 사용 태그이다. 일부 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스 및/또는 1종 이상의 올리고뉴클레오타이드 태그)는 라이브러리의 (예를 들어, 라이브러리의 특정한 일부 내



의) 구성원의 기원을 코딩한다. 일부 실시양태에서, 올리고뉴클레오타이드 (예를 들어, 헤드피스, 제1 태그 및/또는 1종 이상의 추가의 태그, 존재하는 경우)는 기원 서열 (예를 들어, 약 10, 9, 8, 7 또는 6개의 뉴클레오타이드의 길이를 갖는 무작위 또는 축중성 서열)을 포함하며, 여기서 서열은 달리 동일한 라이브러리 구성원의 동일하거나 상이한 경우로부터 유래된 증폭 생성물 사이의 식별을 가능하게 한다. 특정한 실시양태에서, 올리고뉴클레오타이드는 기원 서열을 포함하는 기원 태그이다. 일부 실시양태에서, 방법은 사용 태그 및/또는 기원 태그를 복합체에 연결, 결합 또는 작동가능하게 회합시키는 단계를 추가로 포함한다.

[0036] 본원의 임의의 실시양태에서, 방법, 조성물 및 복합체는 테일피스를 임의로 포함하며, 여기서 테일피스는 본원에 기재된 바와 같은 라이브러리-확인 서열, 사용 서열 또는 기원 서열 중 1종 이상을 포함한다. 특정한 실시양태에서, 방법은 테일피스 (예를 들어, 라이브러리-확인 서열, 사용 서열 또는 기원 서열 중 1종 이상을 포함하는 것)를 복합체에 연결, 결합 또는 작동가능하게 회합시키는 단계를 추가로 포함한다.

[0037] 임의의 상기 실시양태에서, 방법, 조성물 및 복합체 또는 그의 부분 (예를 들어, 헤드피스, 제1 태그 및/또는 1종 이상의 추가의 태그, 존재하는 경우)은 반-, 감소된- 또는 비-수성 (예를 들어, 유기) 조건에서 용해도를 지지하는 변형을 포함할 수 있다. 일부 실시양태에서, 이관능성 스페이서, 헤드피스 또는 1종 이상의 태그는 유기 조건에서 상기 DNA-코딩된 화학적 라이브러리의 구성원의 용해도를 증가시키기 위해 변형된다. 일부 실시양태에서, 변형은 알킬 쇄, 폴리에틸렌 글리콜 단위, 양전하를 갖는 분지형 중 또는 소수성 고리 구조 중 1종 이상이다. 일부 실시양태에서, 변형은 소수성 모이어티를 갖는 1종 이상의 변형된 뉴클레오타이드 (예를 들어, T 또는 C 염기의 C5 위치에서 지방족 쇄로 변형된 것, 예컨대 5'-디메톡시트리틸-N4-디이소부틸아미노메틸리덴-5-(1-프로피닐)-2'-테옥시시딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 5'-디메톡시트리틸-5-(1-프로피닐)-2'-테옥시우리딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 5'-디메톡시트리틸-5-플루오로-2'-테옥시우리딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 및 5'-디메톡시트리틸-5-(피렌-1-일-에틸)-2'-테옥시우리딘 또는 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트 또는 소수성 모이어티를 갖는 삽입물 (예를 들어, 아조벤젠)을 포함한다. 일부 실시양태에서, 라이브러리의 구성원은 약 1.0 내지 약 2.5 (예를 들어, 약 1.0 내지 약 1.5, 약 1.0 내지 약 2.0, 약 1.3 내지 약 1.5, 약 1.3 내지 약 2.0, 약 1.3 내지 약 2.5, 약 1.5 내지 약 2.0, 약 1.5 내지 약 2.5 또는 약 2.0 내지 약 2.5)의 옥탄올:물 계수를 갖는다.

[0038] 임의의 상기 실시양태에서, 폴리머라제는 본원에 참조로 포함되는 국제 출원 번호 PCT/US13/50303에 기재된 바와 같이, 코딩된 화학적 엔티티의 연결 중 적어도 1개를 통해 감소된 판독 또는 전위 능력을 가질 수 있다. 일부 실시양태에서, 폴리머라제는 코딩된 화학적 엔티티의 연결을 통해 적어도 약 10% (예를 들어, 대조군과 비교 시 약 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 심지어 100%)의 감소된 판독 또는 전위 능력을 갖는다. 특정한 실시양태에서, 폴리머라제는 코딩된 화학적 엔티티의 연결을 통해 약 10% 내지 약 100%의 감소된 판독 또는 전위 능력을 갖는다 (예를 들어, 대조군과 비교 시 (예를 들어, 연결이 결여된 대조군 올리고뉴클레오타이드와 비교 시) 20% 내지 100%, 25% 내지 100%, 50% 내지 100%, 75% 내지 100%, 90% 내지 100%, 95% 내지 100%, 10% 내지 95%, 20% 내지 95%, 25% 내지 95%, 50% 내지 95%, 75% 내지 95%, 90% 내지 95%, 10% 내지 90%, 20% 내지 90%, 25% 내지 90%, 50% 내지 90%, 또는 75% 내지 90%).

[0039] 일부 실시양태에서, 코딩된 화학적 엔티티의 연결 중 약 10% 미만 (예를 들어, 약 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%)은 효소적 연결을 포함한다. 일부 실시양태에서, 코딩된 화학적 엔티티의 연결은 0% 내지 90%의 효소적 연결 (예를 들어, 약 0% 내지 40%, 0% 내지 45%, 0% 내지 50%, 0% 내지 55%, 0% 내지 60%, 0% 내지 65%, 0% 내지 70%, 0% 내지 75%, 0% 내지 80%, 0% 내지 85%, 0% 내지 90%, 0% 내지 95%, 0% 내지 96%, 0% 내지 97%, 0% 내지 98%, 0% 내지 99%, 5% 내지 40%, 5% 내지 45%, 5% 내지 50%, 5% 내지 55%, 5% 내지 60%, 5% 내지 65%, 5% 내지 70%, 5% 내지 75%, 5% 내지 80%, 5% 내지 85%, 5% 내지 90%, 5% 내지 95%, 5% 내지 96%, 5% 내지 97%, 5% 내지 98%, 5% 내지 99%, 10% 내지 40%, 10% 내지 45%, 10% 내지 50%, 10% 내지 55%, 10% 내지 60%, 10% 내지 65%, 10% 내지 70%, 10% 내지 75%, 10% 내지 80%, 10% 내지 85%, 10% 내지 90%, 10% 내지 95%, 10% 내지 96%, 10% 내지 97%, 10% 내지 98%, 10% 내지 99%, 15% 내지 40%, 15% 내지 45%, 15% 내지 50%, 15% 내지 55%, 15% 내지 60%, 15% 내지 65%, 15% 내지 70%, 15% 내지 75%, 15% 내지 80%, 15% 내지 85%, 15% 내지 90%, 15% 내지 95%, 15% 내지 96%, 15% 내지 97%, 15% 내지 98%, 15% 내지 99%, 20% 내지 40%, 20% 내지 45%, 20% 내지 50%, 20% 내지 55%, 20% 내지 60%, 20% 내지 65%, 20% 내지 70%, 20% 내지 75%, 20% 내지 80%, 20% 내지 85%, 20% 내지 90%, 20% 내지 95%, 20% 내지 96%, 20% 내지 97%, 20% 내지 98%, 또는 20% 내지 99%)을 포함한다.

- [0040] 일부 실시양태에서, 코딩된 화학적 엔티티의 연결 중 적어도 1개는 화학적 연결 (예를 들어, 화학적-반응성 기, 광-반응성 기, 삽입성 모이어티, 또는 가교 올리고뉴클레오타이드)을 포함한다. 특정한 실시양태에서, 적어도 1종 (예를 들어, 2, 3, 4, 5종 또는 그 초과)의 화학적-반응성 기, 광-반응성 기, 또는 삽입성 모이어티는 태그의 5'-말단에서의 또는 그에 근접한 5'-연결기 및/또는 태그의 3'-말단에서의 또는 그에 근접한 3'-연결기에 존재한다. 다른 실시양태에서, 5'-연결기의 적어도 1종의 서열은 인접한 3'-연결기의 서열에 대해 상보적이거나 또는 동일하거나 충분히 유사하여 상보적 올리고뉴클레오타이드에의 혼성화를 가능하게 한다. 일부 실시양태에서, 코딩된 화학적 엔티티의 연결 중 적어도 10% (예를 들어, 약 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 심지어 100%)는 화학적 연결이다. 다른 실시양태에서, 코딩된 화학적 엔티티의 연결 중 적어도 약 10% 내지 약 100% (예를 들어, 20% 내지 100%, 25% 내지 100%, 50% 내지 100%, 75% 내지 100%, 90% 내지 100%, 95% 내지 100%, 10% 내지 95%, 20% 내지 95%, 25% 내지 95%, 50% 내지 95%, 75% 내지 95%, 90% 내지 95%, 10% 내지 90%, 20% 내지 90%, 25% 내지 90%, 50% 내지 90%, 또는 75% 내지 90%)는 화학적 연결이다.
- [0041] 일부 실시양태에서, 화학적-반응성 기는 임의로 치환된 알킬닐 기 및 임의로 치환된 아지도 기의 쌍; 4  $\pi$ -전자계를 갖는 임의로 치환된 디엔 및 2  $\pi$ -전자계를 갖는 임의로 치환된 친디엔체 또는 임의로 치환된 친헥테로디엔체의 쌍; 친핵체 및 변형 헥테로시클릴 친전자체의 쌍; 임의로 치환된 아미노 기 및 알데히드 또는 케톤 기의 쌍; 임의로 치환된 아미노 기 및 카르복실산 기의 쌍; 임의로 치환된 히드라진 및 알데히드 또는 케톤 기의 쌍; 임의로 치환된 히드록실아민 및 알데히드 또는 케톤 기의 쌍; 친핵체 및 임의로 치환된 알킬 할라이드의 쌍; 백금 착물; 알킬화제; 또는 푸란-변형된 뉴클레오타이드로부터 선택된다.
- [0042] 일부 실시양태에서, 광-반응성 기는 삽입성 모이어티, 프소랄렌 유도체, 임의로 치환된 시아노비닐카르바졸 기 (예를 들어, 3-시아노비닐카르바졸 기, 예컨대 3-시아노비닐카르바졸-1'- $\beta$ -테옥시리보시드-5'-트리포스페이트), 임의로 치환된 비닐카르바졸 기 (예를 들어, 본원에 기재된 바와 같은 아미도비닐카르바졸 기, 카르복시비닐카르바졸 기 또는 C<sub>2-7</sub> 알콕시카르보닐비닐카르바졸 기), 임의로 치환된 시아노비닐 기, 임의로 치환된 아크릴아미드 기, 임의로 치환된 디아지린 기, 임의로 치환된 벤조페논 또는 임의로 치환된 아지드 기를 포함한다.
- [0043] 일부 실시양태에서, 삽입성 모이어티는 프소랄렌 유도체 (예를 들어, 프소랄렌, 8-메톡시프소랄렌 또는 4-히드록시메틸-4,5,8-트리메틸-프소랄렌 (HMT-프소랄렌)), 알칼로이드 유도체 (예를 들어, 베르베린, 팔마틴, 코랄린, 상귀나린 (예를 들어, 그의 이미늄 또는 알칸올아민 형태, 또는 아리스톨로라탐- $\beta$ -D-글루코시드), 에티뎀 양이온 (예를 들어, 브로민화에티뎀), 아크리딘 유도체 (예를 들어, 프로플라빈, 아크리플라빈 또는 암사크린), 안트라시클린 유도체 (예를 들어, 독소루비신, 에피루비신, 다우노루비신 (다우노마이신), 이다루비신 및 아클라루비신) 또는 탈리도미드이다.
- [0044] 일부 실시양태에서, 화학적 연결은 가교 올리고뉴클레오타이드를 포함하며, 여기서 가교 올리고뉴클레오타이드의 5'-말단에서의 적어도 5개의 뉴클레오타이드 서열은 1종 이상의 태그의 3'-말단에서의 적어도 5개의 뉴클레오타이드 서열에 대해 상보적이거나 또는 동일하거나 충분히 유사하여 상보적 올리고뉴클레오타이드에의 혼성화를 가능하게 하며, 가교 올리고뉴클레오타이드의 3'-말단에서의 적어도 5개의 뉴클레오타이드 서열은 1종 이상의 태그의 5'-말단에서의 적어도 5개의 뉴클레오타이드 서열에 대해 상보적이거나 또는 동일하거나 충분히 유사하여 상보적 올리고뉴클레오타이드에의 혼성화를 가능하게 한다. 특정한 실시양태에서, 1종 이상의 태그의 3'-말단은 3'-연결기를 포함한다. 특정한 실시양태에서, 1종 이상의 태그의 5'-말단은 5'-연결기를 포함한다.
- [0045] 일부 실시양태에서, 가교 올리고뉴클레오타이드의 5'-말단 및/또는 3'-말단은 가역적 공-반응성 기 (예를 들어, 본원에 기재된 바와 같은 시아노비닐카르바졸 기, 시아노비닐 기, 아크릴아미드 기, 티올 기 또는 비닐 술폰 기)를 포함한다.
- [0046] 일부 실시양태에서, 3'-연결기 및/또는 5'-연결기는 가역적 공-반응성 기 (예를 들어, 본원에 기재된 바와 같은 시아노비닐카르바졸 기, 시아노비닐 기, 아크릴아미드 기, 티올 기 또는 비닐 술폰 기)를 포함한다.
- [0047] 임의의 상기 실시양태에서, 헤드피스, 테일피스, 제1 태그, 1종 이상의 추가의 태그, 라이브러리-확인 태그, 사용 태그, 및/또는 기원 태그는, 존재하는 경우에, 약 5 내지 약 75개 뉴클레오타이드 (예를 들어, 5 내지 7개 뉴클레오타이드, 5 내지 8개 뉴클레오타이드, 5 내지 9개 뉴클레오타이드, 5 내지 10개 뉴클레오타이드, 5 내지 11개 뉴클레오타이드, 5 내지 12개 뉴클레오타이드, 5 내지 13개 뉴클레오타이드, 5 내지 14개 뉴클레오타이드, 5 내지 15개 뉴클레오타이드, 5 내지 16개 뉴클레오타이드, 5 내지 17개 뉴클레오타이드, 5 내지 18개 뉴클레오타이드, 5 내지 19개 뉴클레오타이드, 5 내지 20개 뉴클레오타이드, 5 내지 30개 뉴클레오타이드, 5 내지 40개 뉴클레오타이드, 5 내지 50개 뉴클

레오티드, 5 내지 60개 뉴클레오티드, 5 내지 70개 뉴클레오티드, 6 내지 7개 뉴클레오티드, 6 내지 8개 뉴클레오티드, 6 내지 9개 뉴클레오티드, 6 내지 10개 뉴클레오티드, 6 내지 11개 뉴클레오티드, 6 내지 12개 뉴클레오티드, 6 내지 13개 뉴클레오티드, 6 내지 14개 뉴클레오티드, 6 내지 15개 뉴클레오티드, 6 내지 16개 뉴클레오티드, 6 내지 17개 뉴클레오티드, 6 내지 18개 뉴클레오티드, 6 내지 19개 뉴클레오티드, 6 내지 20개 뉴클레오티드, 7 내지 8개 뉴클레오티드, 7 내지 9개 뉴클레오티드, 7 내지 10개 뉴클레오티드, 7 내지 11개 뉴클레오티드, 7 내지 12개 뉴클레오티드, 7 내지 13개 뉴클레오티드, 7 내지 14개 뉴클레오티드, 7 내지 15개 뉴클레오티드, 7 내지 16개 뉴클레오티드, 7 내지 17개 뉴클레오티드, 7 내지 18개 뉴클레오티드, 7 내지 19개 뉴클레오티드, 7 내지 20개 뉴클레오티드, 8 내지 9개 뉴클레오티드, 8 내지 10개 뉴클레오티드, 8 내지 11개 뉴클레오티드, 8 내지 12개 뉴클레오티드, 8 내지 13개 뉴클레오티드, 8 내지 14개 뉴클레오티드, 8 내지 15개 뉴클레오티드, 8 내지 16개 뉴클레오티드, 8 내지 17개 뉴클레오티드, 8 내지 18개 뉴클레오티드, 8 내지 19개 뉴클레오티드, 8 내지 20개 뉴클레오티드, 9 내지 10개 뉴클레오티드, 9 내지 11개 뉴클레오티드, 9 내지 12개 뉴클레오티드, 9 내지 13개 뉴클레오티드, 9 내지 14개 뉴클레오티드, 9 내지 15개 뉴클레오티드, 9 내지 16개 뉴클레오티드, 9 내지 17개 뉴클레오티드, 9 내지 18개 뉴클레오티드, 9 내지 19개 뉴클레오티드, 9 내지 20개 뉴클레오티드, 10 내지 11개 뉴클레오티드, 10 내지 12개 뉴클레오티드, 10 내지 13개 뉴클레오티드, 10 내지 14개 뉴클레오티드, 10 내지 15개 뉴클레오티드, 10 내지 16개 뉴클레오티드, 10 내지 17개 뉴클레오티드, 10 내지 18개 뉴클레오티드, 10 내지 19개 뉴클레오티드, 10 내지 20개 뉴클레오티드, 10 내지 30개 뉴클레오티드, 10 내지 40개 뉴클레오티드, 10 내지 50개 뉴클레오티드, 10 내지 60개 뉴클레오티드, 10 내지 70개 뉴클레오티드, 10 내지 75개 뉴클레오티드, 11 내지 12개 뉴클레오티드, 11 내지 13개 뉴클레오티드, 11 내지 14개 뉴클레오티드, 11 내지 15개 뉴클레오티드, 11 내지 16개 뉴클레오티드, 11 내지 17개 뉴클레오티드, 11 내지 18개 뉴클레오티드, 11 내지 19개 뉴클레오티드, 11 내지 20개 뉴클레오티드, 12 내지 13개 뉴클레오티드, 12 내지 14개 뉴클레오티드, 12 내지 15개 뉴클레오티드, 12 내지 16개 뉴클레오티드, 12 내지 17개 뉴클레오티드, 12 내지 18개 뉴클레오티드, 12 내지 19개 뉴클레오티드, 12 내지 20개 뉴클레오티드, 13 내지 14개 뉴클레오티드, 13 내지 15개 뉴클레오티드, 13 내지 16개 뉴클레오티드, 13 내지 17개 뉴클레오티드, 13 내지 18개 뉴클레오티드, 13 내지 19개 뉴클레오티드, 13 내지 20개 뉴클레오티드, 14 내지 15개 뉴클레오티드, 14 내지 16개 뉴클레오티드, 14 내지 17개 뉴클레오티드, 14 내지 18개 뉴클레오티드, 14 내지 19개 뉴클레오티드, 14 내지 20개 뉴클레오티드, 15 내지 16개 뉴클레오티드, 15 내지 17개 뉴클레오티드, 15 내지 18개 뉴클레오티드, 15 내지 19개 뉴클레오티드, 15 내지 20개 뉴클레오티드, 16 내지 17개 뉴클레오티드, 16 내지 18개 뉴클레오티드, 16 내지 19개 뉴클레오티드, 16 내지 20개 뉴클레오티드, 17 내지 18개 뉴클레오티드, 17 내지 19개 뉴클레오티드, 17 내지 20개 뉴클레오티드, 18 내지 19개 뉴클레오티드, 18 내지 20개 뉴클레오티드, 19 내지 20개 뉴클레오티드, 20 내지 30개 뉴클레오티드, 20 내지 40개 뉴클레오티드, 20 내지 50개 뉴클레오티드, 20 내지 60개 뉴클레오티드, 20 내지 70개 뉴클레오티드, 20 내지 75개 뉴클레오티드, 30 내지 40개 뉴클레오티드, 30 내지 50개 뉴클레오티드, 30 내지 60개 뉴클레오티드, 30 내지 70개 뉴클레오티드, 40 내지 50개 뉴클레오티드, 40 내지 60개 뉴클레오티드, 40 내지 70개 뉴클레오티드, 40 내지 75개 뉴클레오티드, 50 내지 60개 뉴클레오티드, 50 내지 70개 뉴클레오티드, 50 내지 75개 뉴클레오티드, 60 내지 70개 뉴클레오티드, 60 내지 75개 뉴클레오티드 및 70 내지 75개 뉴클레오티드)를 포함할 수 있다. 특정한 실시양태에서, 헤드피스, 제1 태그, 제2 태그, 1종 이상의 추가의 태그, 라이브러리-확인 태그, 사용 태그 및/또는 기원 태그는, 존재하는 경우에, 20개 뉴클레오티드 미만 (예를 들어, 19개 뉴클레오티드 미만, 18개 뉴클레오티드 미만, 17개 뉴클레오티드 미만, 16개 뉴클레오티드 미만, 15개 뉴클레오티드 미만, 14개 뉴클레오티드 미만, 13개 뉴클레오티드 미만, 12개 뉴클레오티드 미만, 11개 뉴클레오티드 미만, 10개 뉴클레오티드 미만, 9개 뉴클레오티드 미만, 8개 뉴클레오티드 미만 또는 7개 뉴클레오티드 미만)의 길이를 갖는다.

[0048] 임의의 상기 실시양태에서, 코딩 서열 (예를 들어, 헤드피스, 테일피스, 제1 태그, 1종 이상의 추가의 태그, 라이브러리-확인 태그, 사용 태그 및/또는 기원 태그, 존재하는 경우)은 20개 초과와 뉴클레오티드 (예를 들어, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 또는 75개 초과와 뉴클레오티드)를 포함할 수 있다.

[0049] 정의

[0050] "약"은 언급된 값의 +/- 10%를 의미한다.

[0051] "이관능성"은 2개의 화학적 모이어티의 결합을 가능하게 하는 2개의 반응성 기를 갖는다는 것을 의미한다.

[0052] "이관능성 스페이서"는 화학적 엔티티 및 복합체의 코딩 정보의 결합을 가능하게 하는 2개의 반응성 기를 갖는 스페이싱 모이어티를 의미한다. 하나의 비제한적 예에서, 이관능성 스페이서는 화학적 엔티티와 태그 사이에 제공된다. 또 다른 비제한적 예에서, 이관능성 스페이서는 화학적 엔티티와 헤드피스 사이에 제공된다. 예시

적인 이관능성 스페이서가 본원에 제공된다.

- [0053] "결합"은 공유 결합 또는 비공유 결합에 의한 부착을 의미한다. 비공유 결합은 반 데르 발스 힘, 수소 결합, 이온 결합, 포획 또는 물리적 캡슐화, 흡수, 흡착 및/또는 다른 분자간 힘에 의해 형성된 것을 포함한다. 결합은 임의의 유용한 수단에 의해, 예컨대 효소적 결합 (예를 들어, 효소적 연결을 제공하는 효소적 라이게이션)에 의해 또는 화학적 결합 (예를 들어, 화학적 연결을 제공하는 화학적 라이게이션)에 의해 유발될 수 있다. "라이게이션"은 공유 결합에 의한 부착을 의미한다.
- [0054] "빌딩 블록"은 화학적 엔티티의 구조 단위를 의미하며, 여기서 단위는 다른 화학적 구조 단위에 직접적으로 연결되거나 또는 스캐폴드를 통해 간접적으로 연결된다. 화학적 엔티티가 중합체 또는 올리고머인 경우에, 빌딩 블록은 중합체 또는 올리고머의 단량체 단위이다. 빌딩 블록은 1종 이상의 다른 빌딩 블록 또는 스캐폴드의 추가를 가능하게 하는 1종 이상의 다양성 노드를 가질 수 있다. 대부분의 경우에, 각각의 다양성 노드는 1종 이상의 빌딩 블록 또는 스캐폴드와 반응하여 화학적 엔티티를 형성할 수 있는 관능기이다. 일반적으로, 빌딩 블록은 적어도 2개의 다양성 노드 (또는 반응성 관능기)를 갖지만, 일부 빌딩 블록은 1개의 다양성 노드 (또는 반응성 관능기)를 가질 수 있다. 대안적으로, 코딩된 화학적 또는 결합 단계는 여러 화학적 성분 (예를 들어, 다-성분 축합 반응 또는 다-단계 프로세스)을 포함할 수 있다. 2개의 상이한 빌딩 블록 상의 반응성 기는 상보적이어야 하며, 즉 함께 반응하여 공유 또는 비공유 결합을 형성할 수 있어야 한다.
- [0055] "화학적 엔티티"는 1종 이상의 빌딩 블록, 1종 이상의 스캐폴드, 또는 가역적 고정화를 위한 부위를 포함하는 화합물을 의미한다. 화학적 엔티티는 1종 이상의 목적하는 특징, 예를 들어 생물학적 표적에 결합하는 능력, 용해도, 수소 결합 공여자 및 수용자의 이용가능성, 결합의 회전 자유도, 양전하, 음전하, 또는 가역적 고정화를 위한 부위를 갖도록 설계 또는 구축된 임의의 소분자, 펩티드, 핵산, 펩티드 약물 또는 약물 후보일 수 있다. 특정 실시양태에서, 화학적 엔티티는 이관능성 또는 삼관능성 (또는 그 초과) 엔티티로서 추가로 반응할 수 있다.
- [0056] "화학적-반응성 기"는 모듈 반응에 참여하여 연결을 생성하는 반응성 기를 의미한다. 예시적인 반응 및 반응성 기는 본원에 기재된 바와 같은, 임의로 치환된 알킬닐 기 및 임의로 치환된 아지도 기의 쌍과의 휘스겐 1,3-쌍극자 고리화첨가 반응; 4  $\pi$ -전자계를 갖는 임의로 치환된 디엔 및 2  $\pi$ -전자계를 갖는 임의로 치환된 친디엔체 또는 임의로 치환된 친헥테로디엔체의 쌍과의 딜스-아더 반응; 친핵체 및 변형 헥테로시클릴 친전자체와의 개환 반응; 포스포로티오에이트 기 및 아이오도 기와의 스플린트 라이게이션 반응; 및 알데히드 기 및 아미노 기와의 환원성 아미노화 반응으로부터 선택되는 것을 포함한다.
- [0057] "상보적"은 본원에 기재된 바와 같이 혼성화하여 2차 구조 (핵산 분자의 듀플렉스 또는 이중-가닥 부분)를 형성할 수 있는 서열을 의미한다. 상보성은 완전할 필요는 없지만, 1, 2, 3개 또는 그 초과 뉴클레오타이드에서 1개 이상의 미스매치를 포함할 수 있다. 예를 들어, 상보적 서열은 왓슨-크릭 염기쌍 규칙 (예를 들어, G와 C, A와 T 또는 A와 U) 또는 다른 수소 결합 모티프 (예를 들어, 디아미노퓨린과 T, 5-메틸 C와 G, 2-티오티미딘과 A, 이노신과 C, 슈도이소시토신과 G)에 따라 수소 결합을 형성할 수 있는 핵염기를 함유할 수 있다. 서열 및 그의 상보적 서열은 동일한 올리고뉴클레오타이드 내 또는 상이한 올리고뉴클레오타이드 내에 존재할 수 있다.
- [0058] "복합체" 또는 "라이게이션된 복합체"는 공유 결합 또는 비공유 결합에 의해 화학적 엔티티 및/또는 1종 이상의 올리고뉴클레오타이드 태그와 작동가능하게 회합된 헤드피스를 의미한다. 복합체는 화학적 엔티티와 헤드피스 사이에 이관능성 스페이서를 임의로 포함할 수 있다.
- [0059] 화학적 엔티티의 "성분"은 스캐폴드 또는 빌딩 블록을 의미한다.
- [0060] 올리고뉴클레오타이드 태그의 "연결기"는 고정된 서열을 갖는 5'- 또는 3'-말단에서의 또는 그에 근접한 태그의 부분을 의미한다. 5'-연결기는 올리고뉴클레오타이드의 5'-말단에 또는 그에 근접하여 위치하고, 3'-연결기는 올리고뉴클레오타이드의 3'-말단에 또는 그에 근접하여 위치한다. 복합체 내에 존재하는 경우에, 각각의 5'-연결기는 동일하거나 상이할 수 있으며, 각각의 3'-연결기는 동일하거나 상이할 수 있다. 1종 초과 태그를 갖는 예시적인 비제한적 복합체에서, 각각의 태그는 5'-연결기 및 3'-연결기를 포함할 수 있으며, 여기서 각각의 5'-연결기는 동일한 서열을 갖고 각각의 3'-연결기는 동일한 서열을 갖는다 (예를 들어, 여기서 5'-연결기의 서열은 3'-연결기의 서열과 동일하거나 상이할 수 있음). 또 다른 예시적인 비제한적 복합체에서, 5'-연결기의 서열은 본원에 기재된 바와 같이 3'-연결기의 서열에 상보적으로 설계된다 (예를 들어, 5'- 및 3'-연결기 사이의 혼성화가 가능하도록). 연결기는 연결 (예를 들어, 폴리머라제가 감소된 판독 또는 전위 능력을 갖게 하는 연결, 예컨대 화학적 연결)을 가능하게 하는 1종 이상의 기를 임의로 포함할 수 있다.



- [0061] "불변" 또는 "고정된 불변" 서열은 정보를 코딩하지 않는 올리고뉴클레오타이드 서열을 의미한다. 불변 서열을 갖는 비제한적인 예시적 복합체의 부분은 프라이머-결합 영역, 5'-연결기 또는 3'-연결기를 포함한다. 본 발명의 헤드피스는 정보를 코딩할 수 있거나 (따라서, 태그) 또는 대안적으로 정보를 코딩할 수 없다 (따라서, 불변 서열). 유사하게, 본 발명의 테일피스는 정보를 코딩할 수 있거나 또는 코딩할 수 없다.
- [0062] "가교 올리고뉴클레오타이드"는 본원에 기재된 바와 같이 복합체 내의 2개의 인접한 태그 사이의 특정한 접합부에 작동가능하게 회합된 올리고뉴클레오타이드를 의미한다. 비제한적 예에서, 가교 올리고뉴클레오타이드의 하나의 말단은 제1 태그의 3'-연결기에 혼성화하고, 가교 올리고뉴클레오타이드의 다른 말단은 제1 태그에 인접한 제2 태그의 5'-연결기에 혼성화한다. 가교 올리고뉴클레오타이드의 예시적인 비제한적 실시양태는 인접한 태그 또는 인접한 태그의 연결기와 작동가능하게 회합된 1종 이상의 반응성 기 (예를 들어, 화학적-반응성 기, 광-반응성 기, 삽입성 모이어티 또는 가역적 공-반응성 기 또는 본원에 기재된 임의의 것)를 갖는 것을 포함한다.
- [0063] "다양성 노드"는 또 다른 빌딩 블록의 추가를 가능하게 하는 스캐폴드 또는 빌딩 블록 내의 위치의 관능기를 의미한다.
- [0064] "헤드피스"는 제1 화학적 엔티티의 성분, 태그, 예를 들어 출발 올리고뉴클레오타이드, 및 가역적 고정화를 위한 부위를 포함하는 제2 화학적 엔티티에 작동가능하게 연결된 라이브러리 합성을 위한 화학 구조를 의미한다. 임의로 헤드피스는 뉴클레오타이드를 거의 또는 전혀 함유하지 않을 수 있지만, 작동가능하게 회합될 수 있는 지점을 제공할 수 있다. 임의로, 이관능성 스페이서는 헤드피스를 성분에 연결시킨다.
- [0065] "혼성화하다"는 다양한 엄격성 조건 하에서 상보적 올리고뉴클레오타이드들 또는 그의 부분들 사이에 쌍을 이루어 이중-가닥 분자를 형성하는 것을 의미한다. (예를 들어, 문헌 [Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507] 참조.) 예를 들어, 고 엄격성 혼성화는 통상적으로 약 750 mM NaCl 및 75 mM 시트르산삼나트륨 미만, 약 500 mM NaCl 및 50 mM 시트르산삼나트륨 미만, 또는 약 250 mM NaCl 및 25 mM 시트르산삼나트륨 미만의 염 농도에 의해 수득될 수 있다. 저 엄격성 혼성화는 유기 용매, 예를 들어 포름아미드의 부재 하에 수득될 수 있는 반면에, 고 엄격성 혼성화는 적어도 약 35% 포름아미드 또는 적어도 약 50% 포름아미드의 존재 하에 수득될 수 있다. 고 엄격성 혼성화 온도 조건은 통상적으로 적어도 약 30°C, 37°C 또는 42°C의 온도를 포함할 것이다. 추가의 파라미터, 예컨대 혼성화 시간, 세제, 예를 들어 소듐 도데실 술페이트 (SDS)의 농도, 및 캐리어 DNA의 포함 또는 배제를 변화시키는 것은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다. 다양한 수준의 엄격성이 이들 다양한 조건을 필요에 따라 조합함으로써 달성된다. 한 실시양태에서, 혼성화는 30°C에서, 750 mM NaCl, 75 mM 시트르산삼나트륨 및 1% SDS 중에서 이루어질 것이다. 대안적 실시양태에서, 혼성화는 37°C에서, 500 mM NaCl, 50 mM 시트르산삼나트륨, 1% SDS, 35% 포름아미드 및 100 µg/ml 변성 연어 정자 DNA (ssDNA) 중에서 이루어질 것이다. 추가의 대안적 실시양태에서, 혼성화는 42°C에서, 250 mM NaCl, 25 mM 시트르산삼나트륨, 1% SDS, 50% 포름아미드, 및 200 µg/ml의 ssDNA 중에서 이루어질 것이다. 이들 조건에 대한 유용한 변경은 관련 기술분야의 통상의 기술자에게 용이하게 이해될 것이다.
- [0066] 대부분의 적용의 경우에, 혼성화 후의 세척 단계도 또한 엄격성을 변화시킬 것이다. 세척 엄격성 조건은 염 농도 및 온도에 의해 규정될 수 있다. 상기와 같이, 세척 엄격성은 염 농도를 감소시킴으로써 또는 온도를 증가시킴으로써 증가될 수 있다. 예를 들어, 세척 단계에 대한 고 엄격성 염 농도는, 예를 들어 약 30 mM NaCl 및 3 mM 시트르산삼나트륨 미만, 또는 약 15 mM NaCl 및 1.5 mM 시트르산삼나트륨 미만일 수 있다. 세척 단계를 위한 고 엄격성 온도 조건은 통상적으로, 예를 들어 적어도 약 25°C, 42°C 또는 68°C의 온도를 포함할 것이다. 한 실시양태에서, 세척 단계는 25°C에서, 30 mM NaCl, 3 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어질 것이다. 대안적 실시양태에서, 세척 단계는 42°C에서, 15 mM NaCl, 1.5 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어질 것이다. 추가의 대안적 실시양태에서, 세척 단계는 68°C에서, 15 mM NaCl, 1.5 mM 시트르산삼나트륨 및 0.1% SDS 중에서 이루어질 것이다. 이들 조건에 대한 추가의 변형은 관련 기술분야의 통상의 기술자에게 용이하게 명백할 것이다. 혼성화 기술은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있고, 예를 들어 문헌 [Benton and Davis (*Science* 196:180, 1977); Grunstein and Hogness (*Proc. Natl. Acad. Sci., USA* 72:3961, 1975); Ausubel et al. (*Current Protocols in Molecular Biology*, Wiley Interscience, New York, 2001); Berger and Kimmel (*Guide to Molecular Cloning Techniques*, 1987, Academic Press, New York); 및 Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York]에 기재되어 있다.
- [0067] "삽입성 모이어티"는 2종 이상의 뉴클레오타이드 사이에 모이어티를 포함시키는 반응성 기를 의미한다. 비제한적

예에서, 삽입성 모이어티는 1종 이상의 뉴클레오타이드와 반응하여 듀플렉스 또는 트리플렉스 올리고뉴클레오타이드 사이에 가닥간 또는 가닥내 가교를 형성한다. 예시적인 비제한적 삽입성 모이어티가 본원에 기재되어 있다.

[0068] "접합부"는 복합체 내의 2개의 인접한 태그 사이의 닉 (뉴클레오타이드내 결합의 결여) 또는 겹 (1개 이상의 뉴클레오타이드의 결여)을 의미한다. 접합부는 또한 2개의 인접한 태그 내에 존재하는 2개의 인접한 연결기 사이 (예를 들어, 제1 태그의 3'-연결기 및 제1 태그에 인접한 제2 태그의 5'-연결기 사이)일 수 있다.

[0069] "라이브리리"는 분자 또는 화학적 엔티티의 집합을 의미한다. 임의로, 분자 또는 화학적 엔티티는 분자 또는 화학적 엔티티의 부분을 코딩하는 1종 이상의 올리고뉴클레오타이드에 결합된다.

[0070] "연결"은 2종 이상의 화학 구조를 작동가능하게 회합시키는 것을 가능하게 하는 화학적 연결 엔티티를 의미하며, 여기서 연결은 헤드피스와 1종 이상의 태그 사이, 2종의 태그 사이, 또는 태그와 테일피스 사이에 존재한다. 화학적 연결 엔티티는 2개의 관능기 사이의 비-공유 결합 (예를 들어, 본원에 기재된 바와 같음), 공유 결합 또는 반응 생성물일 수 있다. "화학적 연결"은 2개의 관능기, 예컨대 모노포스페이트 및 히드록실기 사이의 비-효소적, 화학적 반응에 의해 형성된 연결을 의미한다. 예시적인 비제한적 관능기는 화학적-반응성기, 광-반응성기, 삽입성 모이어티, 또는 가교 올리고뉴클레오타이드 (예를 들어, 본원에 기재된 바와 같음)를 포함한다. "효소적 연결"은 효소에 의해 형성된 뉴클레오타이드간 또는 뉴클레오시드간 연결을 의미한다. 예시적인 비제한적 효소는 키나제, 폴리머라제, 리가제 또는 그의 조합을 포함한다. "폴리머라제가 감소된 판독 또는 전위 능력을 갖게 하는" 연결은, 올리고뉴클레오타이드 주형 내에 존재하는 경우에, 연결이 결여된 대조군 올리고뉴클레오타이드와 비교 시, 폴리머라제에 의해 감소된 양의 신장 및/또는 증폭된 생성물을 제공하는 연결을 의미한다. 이러한 연결을 결정하는 예시적인 비제한적 방법은 PCR 분석 (예를 들어, 정량적 PCR), RT-PCR 분석, 액체 크로마토그래피-질량 분광측정법, 서열 집단통계, 또는 다른 방법에 의해 평가되는 바와 같은 프라이머 연장을 포함한다. 예시적인 비제한적 폴리머라제는 DNA 폴리머라제 및 RNA 폴리머라제, 예컨대 DNA 폴리머라제 I, DNA 폴리머라제 II, DNA 폴리머라제 III, DNA 폴리머라제 VI, Taq DNA 폴리머라제, 딥 벤트알(Deep VentR)<sup>TM</sup> DNA 폴리머라제 (고-충실도 고온성 DNA 폴리머라제, 뉴잉글랜드 바이오랩스(New England Biolabs)로부터 입수가가능함), T7 DNA 폴리머라제, T4 DNA 폴리머라제, RNA 폴리머라제 I, RNA 폴리머라제 II, RNA 폴리머라제 III 또는 T7 RNA 폴리머라제를 포함한다.

[0071] "다가 양이온"은 1종 초과 리간드 또는 음이온과 1개 초과 결합을 형성할 수 있는 양이온을 의미한다. 다가 양이온은 이온 복합체 또는 배위 복합체를 형성할 수 있다. 예시적인 다가 양이온은 알칼리 토금속 (예를 들어, 마그네슘) 및 전이 금속 (예를 들어, 망가니즈 (II) 또는 코발트 (III))으로부터의 것, 및 1개 이상의 음이온 및/또는 1개 이상의 1가 또는 여러자리 리간드에 임의로 결합하는 것, 예컨대 클로라이드, 아민 및/또는 에틸렌디아민을 포함한다.

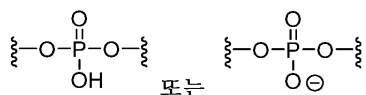
[0072] "올리고뉴클레오타이드"는 5'-말단, 3'-말단, 및 5'-말단과 3'-말단 사이의 내부 위치에 1개 이상의 뉴클레오타이드를 갖는 뉴클레오타이드의 중합체를 의미한다. 올리고뉴클레오타이드는, 합성되어 염기-쌍 인식에 사용될 수 있는, 관련 기술분야에 공지되어 있는 DNA, RNA 또는 그의 임의의 유도체를 포함할 수 있다. 올리고뉴클레오타이드는 인접한 염기를 가져야 하는 것은 아니지만, 링커 모이어티에 의해 산재될 수 있다. 올리고뉴클레오타이드 중합체 및 뉴클레오타이드 (예를 들어, 변형된 DNA 또는 RNA)는 천연 염기 (예를 들어, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신, 데옥시시티딘, 이노신 또는 디아미노 퓨린), 염기 유사체 (예를 들어, 2'-아미노아데노신, 2'-티오티미딘, 이노신, 피콜로-피리미딘, 3-메틸 아데노신, C5-프로피닐시티딘, C5-프로피닐우리딘, C5-브로모우리딘, C5-플루오로우리딘, C5-아이오도우리딘, C5-메틸시티딘, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, 0(6)-메틸구아닌 및 2'-티오시티딘), 변형된 염기 (예를 들어, 2'-치환된 뉴클레오타이드, 예컨대 2'-O-메틸화 염기 및 2'-플루오로 염기), 삽입된 염기, 변형된 당 (예를 들어, 2'-플루오로리보스; 리보스; 2'-데옥시리보스; 아라비노스; 핵소스; 안히드로핵시톨; 알트리톨; 만니톨; 시클로핵사닐; 시클로핵세닐; 또한 포스포르아미데이트 백본을 갖는 모르폴리노; 잠금 핵산 (예를 들어, 리보스의 2'-히드록실이 C<sub>1-6</sub> 알킬렌 또는 C<sub>1-6</sub> 헤테로알킬렌 가교에 의해 동일한 리보스 당의 4'-탄소에 연결되며, 여기서 예시적인 가교는 메틸렌, 프로필렌, 에테르 또는 아미노 가교를 포함하는 것인 LNA); 글리콜 핵산 (리보스가 포스포디에스테르 결합에 부착된 글리콜 단위에 의해 대체된 것인 GNA, 예를 들어 R-GNA 또는 S-GNA); 트레오스 핵산 (리보스가 α-L-트레오파라노실-(3'→2')로 대체된 것인 TNA); 및/또는 리보스 내 산소의 대체 (예를 들어, S, Se 또는 알킬렌, 예컨대 메틸렌 또는 에틸렌에 의함), 변형된 백본 (예를 들어, 2'-아미노-에틸-글리신 연결이 리보스 및 포스포디에스테르 백본을 대체한 것인 펩티드 핵산 (PNA)) 및/또는 변형된 포스페이트 기 (예를 들어, 포스포로티오에이트, 5'-N-포스포르아미다이트, 포스포로셀레네이트, 보라노포스

페이트, 보라노포스페이트 에스테르, 수소 포스포네이트, 포스포르아미데이트, 포스포로디아미데이트, 알킬 또는 아릴 포스포네이트, 포스포트리에스테르, 가교된 포스포르아미데이트, 가교된 포스포로티오에이트 및 가교된 메틸렌-포스포네이트)를 포함할 수 있다. 올리고뉴클레오타이드는 단일-가닥 (예를 들어, 헤어핀), 이중-가닥일 수 있거나, 또는 다른 2차 또는 3차 구조 (예를 들어, 줄기-루프 구조, 이중 나선, 트리플렉스, 퀴드로플렉스 등)를 보유할 수 있다. 올리고뉴클레오타이드는 또한 1개 이상의 3'-3' 또는 5'-5' 연결, 또는 1개 이상의 역전된 뉴클레오타이드를 포함할 수 있다. 이는 이들이 2개의 3'-말단 또는 2개의 5'-말단을 함유한다는 것을 의미할 수 있다. 올리고뉴클레오타이드는 또한 1회 이상 분지화될 수 있으며, 여기서 이들은 2개 초과 말단을 함유할 수 있다. 올리고뉴클레오타이드는 또한 원형화될 수 있으며, 여기서 이들은 2개 미만의 말단을 함유할 수 있고, 말단을 전혀 함유하지 않을 수 있다.

[0073] "결합 쌍의 하나의 구성원"은 가역적 고정화를 위해 또 다른 상보적 화학적 엔티티와 쌍을 형성할 수 있는 화학적 엔티티를 의미한다 (예를 들어, 핵산, 펩티드, 또는 소분자).

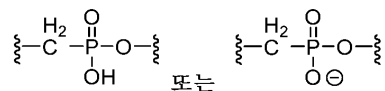
[0074] "작동가능하게 연결된" 또는 "작동가능하게 회합된"은 2종 이상의 화학 구조가 직접적으로 또는 간접적으로 함께 연결됨으로써, 일어날 것으로 예상되는 다양한 조작을 통해서도 연결된 상태로 유지되는 것을 의미한다. 전형적으로, 화학적 엔티티 및 헤드피스는 간접적인 방식으로 (예를 들어, 적절한 스페이서를 통해 공유적으로) 작동가능하게 회합된다. 예를 들어, 스페이서는 화학적 엔티티에 대한 부착 부위 및 헤드피스에 대한 부착 부위를 갖는 이관능성 모이어티일 수 있다.

[0075] "포스포디에스테르 연결"은 하기 구조를 포함하는 연결을 의미한다:



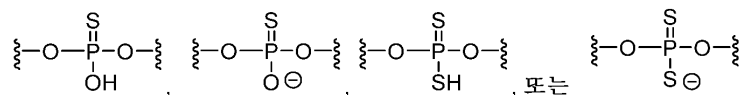
[0076]

[0077] "포스포네이트 연결"은 하기 구조를 포함하는 연결을 의미한다:



[0078]

[0079] "포스포로티오에이트 연결"은 하기 구조를 포함하는 연결을 의미한다:



[0080]

[0081] "광-반응성 기"는 자외, 가시 또는 적외 방사선의 흡수에 의해 유발되는 반응에 참여하여 연결을 생성하는 반응성 기를 의미한다. 예시적인 비제한적 광-반응성 기가 본원에 기재되어 있다.

[0082] "보호기"는 올리고뉴클레오타이드-코딩된 라이브러리를 제조, 태그부착 또는 사용하는 하나 이상의 결합 단계 동안 바람직하지 않은 반응에 대해 올리고뉴클레오타이드의 3'-말단 또는 5'-말단을 보호하거나, 또는 화학적 엔티티, 스캐폴드 또는 빌딩 블록의 1종 이상의 관능기를 보호하도록 의도되는 기를 의미한다. 통상적으로 사용되는 보호기는 문헌 [Greene, "Protective Groups in Organic Synthesis," 4<sup>th</sup> Edition (John Wiley & Sons, New York, 2007)]에 개시되어 있고, 이는 본원에 참조로 포함된다. 올리고뉴클레오타이드에 대한 예시적인 보호기는 비가역적 보호기, 예컨대 디테옥시뉴클레오타이드 및 디테옥시뉴클레오시드 (ddNTP 또는 ddN), 및 보다 바람직하게는, 히드록실 기에 대한 가역적 보호기, 예컨대 에스테르 기 (예를 들어, 0-( $\alpha$ -메톡시에틸)에스테르, 0-이소발레릴 에스테르 및 0-레불리닐 에스테르), 트리틸 기 (예를 들어, 디메톡시트리틸 및 모노메톡시트리틸), 크산테닐 기 (예를 들어, 9-페닐크산텐-9-일 및 9-(p-메톡시페닐)크산텐-9-일), 아실 기 (예를 들어, 페녹시아세틸 및 아세틸) 및 실릴 기 (예를 들어, t-부틸디메틸실릴)를 포함한다. 화학적 엔티티, 스캐폴드 및 빌딩 블록에 대한 예시적인 비제한적 보호기는 합성 절차 동안 바람직하지 않은 반응에 대해 아미노 기를 보호하기 위한 N-보호기 (예를 들어, 아실; 아릴로일; 카르바미, 예컨대 포르밀, 아세틸, 프로피오닐, 피발로일, t-부틸아세틸, 2-클로로아세틸, 2-브로모아세틸, 트리플루오로아세틸, 트리클로로아세틸, 프탈릴, o-니트로페녹시아세틸,  $\alpha$ -클로로부틸, 벤조일, 4-클로로벤조일, 4-브로모벤조일, 4-니트로벤조일, 및 키랄 보조기, 예컨대 보호 또는 탈보호된 D, L 또는 D, L-아미노산, 예컨대 알라닌, 류신, 페닐알라닌; 술포닐-함유 기, 예컨대 벤젠술포닐,



p-톨루엔술포닐; 카르바메이트 형성기, 예컨대 벤질옥시카르보닐, p-클로로벤질옥시카르보닐, p-메톡시벤질옥시카르보닐, p-니트로벤질옥시카르보닐, 2-니트로벤질옥시카르보닐, p-브로모벤질옥시카르보닐, 3,4-디메톡시벤질옥시카르보닐, 3,5-디메톡시벤질옥시카르보닐, 2,4-디메톡시벤질옥시카르보닐, 4-메톡시벤질옥시카르보닐, 2-니트로-4,5-디메톡시벤질옥시카르보닐, 3,4,5-트리메톡시벤질옥시카르보닐, 1-(p-비페닐릴)-1-메틸에톡시카르보닐, α, α-디메틸-3,5-디메톡시벤질옥시카르보닐, 벤즈히드릴옥시카르보닐, t-부틸옥시카르보닐, 디이소프로필메톡시카르보닐, 이소프로필옥시카르보닐, 에톡시카르보닐, 메톡시카르보닐, 알릴옥시카르보닐, 2,2,2,-트리클로로에톡시카르보닐, 페녹시카르보닐, 4-니트로페녹시카르보닐, 플루오레닐-9-메톡시카르보닐, 시클로펜틸옥시카르보닐, 아다만틸옥시카르보닐, 시클로헥실옥시카르보닐, 페닐티오키카르보닐; 알킬아릴 기, 예컨대 벤질, 트리페닐메틸, 벤질옥시메틸; 및 실릴 기, 예컨대 트리메틸실릴; 여기서 바람직한 N-보호기는 포르밀, 아세틸, 벤조일, 피발로일, t-부틸아세틸, 알라닐, 페닐술포닐, 벤질, t-부틸옥시카르보닐 (Boc), 및 벤질옥시카르보닐 (Cbz)임; 합성 절차 동안 바람직하지 않은 반응에 대해 히드록실 기를 보호하기 위한 O-보호기 (예를 들어, 알킬카르보닐 기, 예컨대 아실, 아세틸, 피발로일; 임의로 치환된 아릴카르보닐 기, 예컨대 벤조일; 실릴 기, 예컨대 트리메틸실릴 (TMS), tert-부틸디메틸실릴 (TBDMS), 트리-이소-프로필실릴옥시메틸 (TOM), 트리아이소프로필실릴 (TIPS); 이러한 메틸, 메톡시메틸, 테트라히드로피라닐, 벤질, p-메톡시벤질, 트리틸의 히드록실을 갖는 에테르-형성기; 알콕시카르보닐, 예컨대 메톡시카르보닐, 에톡시카르보닐, 이소프로폭시카르보닐, n-이소프로폭시카르보닐, n-부틸옥시카르보닐, 이소부틸옥시카르보닐, sec-부틸옥시카르보닐, t-부틸옥시카르보닐, 2-에틸헥실옥시카르보닐, 시클로헥실옥시카르보닐, 메틸옥시카르보닐; 알콕시알콕시카르보닐 기, 예컨대 메톡시메톡시카르보닐, 에톡시메톡시카르보닐, 2-메톡시에톡시카르보닐, 2-에톡시에톡시카르보닐, 2-부톡시에톡시카르보닐, 2-메톡시에톡시메톡시카르보닐, 알릴옥시카르보닐, 프로파르길옥시카르보닐, 2-부텐옥시카르보닐, 3-메틸-2-부텐옥시카르보닐; 할로알콕시카르보닐, 예컨대 2-클로로에톡시카르보닐, 2-클로로에톡시카르보닐, 2,2,2-트리클로로에톡시카르보닐; 임의로 치환된 아릴알콕시카르보닐 기, 예컨대 벤질옥시카르보닐, p-메톡시벤질옥시카르보닐, p-메톡시벤질옥시카르보닐, p-니트로벤질옥시카르보닐, 2,4-디니트로벤질옥시카르보닐, 3,5-디메틸벤질옥시카르보닐, p-클로로벤질옥시카르보닐, p-브로모벤질옥시카르보닐; 및 임의로 치환된 아릴옥시카르보닐 기, 예컨대 페녹시카르보닐, p-니트로페녹시카르보닐, o-니트로페녹시카르보닐, 2,4-디니트로페녹시카르보닐, p-메틸-페녹시카르보닐, m-메틸페녹시카르보닐, o-브로모페녹시카르보닐, 3,5-디메틸페녹시카르보닐, p-클로로페녹시카르보닐, 2-클로로-4-니트로페녹시카르보닐); 카르보닐-보호기 (예를 들어, 아세탈 및 케탈 기, 예컨대 디메틸 아세탈, 1,3-디옥솔란; 아시달 기; 및 디티안 기, 예컨대 1,3-디티안, 1,3-디티올란); 카르복실산-보호기 (예를 들어, 에스테르 기, 예컨대 메틸 에스테르, 벤질 에스테르, t-부틸 에스테르, 오르토 에스테르; 실릴 기, 예컨대 트리메틸실릴, 뿐만 아니라 본원에 기재된 임의의 것; 및 옥사졸린 기); 및 포스페이이트-보호기 (예를 들어, 임의로 치환된 에스테르 기, 예컨대 메틸 에스테르, 이소프로필 에스테르, 2-시아노에틸 에스테르, 알릴 에스테르, t-부틸 에스테르, 벤질 에스테르, 플루오레닐메틸 에스테르, 2-(트리메틸실릴)에틸 에스테르, 2-(메틸술포닐)에틸 에스테르, 2,2,2-트리클로로에틸 에스테르, 3',5'-디메톡시벤조인 에스테르, p-히드록시페나실 에스테르)를 포함한다.

- [0083] 올리고뉴클레오타이드의 말단에 "근접" 또는 "근접한"은 다른 나머지 말단보다 언급된 말단에 인접하거나 가까운 것을 의미한다. 예를 들어, 올리고뉴클레오타이드의 3'-말단에 근접한 모이어티 또는 기는 5'-말단보다 3'-말단에 인접하거나 가깝다. 특정한 실시양태에서, 올리고뉴클레오타이드의 3'-말단에 근접한 모이어티 또는 기는 3'-말단으로부터의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15개 또는 그 초과 뉴클레오타이드이다. 다른 실시양태에서, 올리고뉴클레오타이드의 5'-말단에 근접한 모이어티 또는 기는 5'-말단으로부터의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15개 또는 그 초과 뉴클레오타이드이다.
- [0084] "정제하는 것"은 연속 단계에서 사용하고자 하는 화학적 또는 생물학적 작용제의 활성을 감소시킬 수 있는, 반응 혼합물 중에 존재하는 임의의 미반응 생성물 또는 임의의 작용제를 제거하는 것을 의미한다. 정제하는 것은 제거하고자 하는 미반응 생성물 또는 시약의 크로마토그래피 분리, 전기영동 분리, 및 침전 중 1종 이상을 포함할 수 있다. 정제하는 것은 또한 용매의 제거를 포함할 수 있다.
- [0085] "가역적 공-반응성 기"는 가역 반응에 참여하는 반응성 기를 의미한다. 예시적인 비제한적 반응성 기는, 특정한 흡수 방사선에 노출된 경우에 광-반응성 기 사이에 연결을 생성하고, 상이한, 특정한 흡수 방사선에 노출된 경우에 형성된 연결의 절단을 생성하는 광-반응성 기 (예를 들어, 시아노비닐카르바졸 기, 시아노비닐 기 및 아크릴아미드 기)를 포함한다. 또 다른 예시적인 비제한적 반응성 기는 가역적으로 환원 또는 산화될 수 있는 산화환원-반응성 기 (예를 들어, 티올 기)를 포함한다.
- [0086] "가역적 고정화"는 온화한 조건 하에 지지체로부터의 탈착을 가능하게 하는 방식으로 복합체의 고정화를 의미

한다 (예를 들어, 흡착, 이온 결합, 친화도 결합, 킬레이트화, 디설피드 결합 형성, 올리고뉴클레오타이드 혼성화, 소분자-소분자 상호작용, 가역적 화학, 단백질-단백질 상호작용 및 소수성 상호작용).

[0087] "스캐폴드"는 특정한 특수 기하구조로 1종 이상의 다양성 노드를 디스플레이하는 화학적 모이어티를 의미한다. 다양성 노드는 전형적으로는 라이브러리 합성 동안 스캐폴드에 부착되지만, 일부 경우에서 1개의 다양성 노드는 라이브러리 합성 전에 스캐폴드에 부착될 수 있다 (예를 들어, 1종 이상의 빌딩 블록 및/또는 1종 이상의 태그의 추가). 일부 실시양태에서, 스캐폴드는, 그것이 라이브러리 합성 동안 직교형으로 탈보호될 수 있고 후속적으로 상이한 다양성 노드와 반응할 수 있도록 유도체화된다.

[0088] "소분자" 약물 또는 "소분자" 약물 후보는 분자량이 약 1,000 달톤 미만인 분자를 의미한다. 소분자는 (예를 들어, 화합물 라이브러리 또는 천연 공급원으로부터) 단리되거나 또는 공지된 화합물의 유도체화에 의해 수득된 유기 또는 무기물일 수 있다.

[0089] "실질적 동일성" 또는 "실질적으로 동일한"은 각각 참조 서열과 동일한 폴리펩티드 또는 폴리뉴클레오타이드 서열을 갖거나, 또는 2개의 서열을 최적으로 정렬했을 때 참조 서열 내의 상응하는 위치에서 각각 동일한 아미노산 잔기 또는 뉴클레오타이드를 명시된 백분율로 갖는 폴리펩티드 또는 폴리뉴클레오타이드 서열을 의미한다. 예를 들어, 참조 서열과 "실질적으로 동일한" 아미노산 서열은 참조 아미노산 서열과 적어도 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 동일성을 갖는다. 폴리펩티드의 경우에, 비교 서열의 길이는 일반적으로 적어도 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개의 인접 아미노산, 보다 바람직하게는 적어도 25, 50, 75, 90, 100, 150, 200, 250, 300 또는 350개의 인접 아미노산, 및 가장 바람직하게는 전장 아미노산 서열일 것이다. 핵산의 경우에, 비교 서열의 길이는 일반적으로 적어도 5개의 인접 뉴클레오타이드, 바람직하게는 적어도 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25개의 인접 뉴클레오타이드, 및 가장 바람직하게는 전장 뉴클레오타이드 서열일 것이다. 서열 동일성은 디폴트 설정된 서열 분석 소프트웨어 (예를 들어, 53705 위스콘신주 매디슨 유니버시티 애비뉴 1710 위스콘신 대학교 바이오테크놀로지 센터 제네틱스 컴퓨터 그룹의 서열 분석 소프트웨어 패키지)를 사용하여 측정될 수 있다. 이러한 소프트웨어는 다양한 치환, 결실 및 다른 변형에 상동성 정도를 할당함으로써 유사한 서열을 매칭시킬 수 있다.

[0090] "실질적으로"는 관심 특징 또는 특성의 전체 또는 거의-전체 범위 또는 정도를 나타내는 정성적 조건을 의미한다. 생물학적 및 화학적 현상은, 그것이 있을지라도, 드물게 완료되고/거나 완료까지 진행되거나 또는 절대적 결과를 달성하거나 회피한다는 것을 생물학 분야의 통상의 기술자는 이해할 것이다. 용어 "실질적으로"는 따라서 본원에서 많은 생물학적 및 화학적 현상에 내재된 완료의 잠재적인 결여를 포착하기 위해 사용된다.

[0091] "태그" 또는 "올리고뉴클레오타이드 태그"는 정보를 코딩하는 적어도 일부인 라이브러리의 올리고뉴클레오타이드 부분을 의미한다. 이러한 정보의 비제한적 예는 성분 (즉 각각 스캐폴드 태그 또는 빌딩 블록 태그에서와 같은 스캐폴드 또는 빌딩 블록), 라이브러리 내 헤드피스, 라이브러리의 아이덴티티 (즉 아이덴티티 태그에서와 같은), 라이브러리의 사용 (즉 사용 태그에서와 같은), 및/또는 라이브러리 구성원의 기원 (즉 기원 태그에서와 같은)의 추가 (예를 들어, 결합 반응에 의함)를 포함한다. 태그 세트는 임의로 동일하거나 거의-동일한 질량 태그로 구성될 수 있으며, 이에 의해 질량 분광측정법에 의한 라이브러리의 분석적 평가가 용이해질 수 있다.

[0092] "테일피스"는 모든 선행 태그의 추가 후에 복합체에 부착되고, 라이브러리의 아이덴티티, 라이브러리의 사용 및/또는 라이브러리 구성원의 기원을 코딩하는 라이브러리의 올리고뉴클레오타이드 부분을 의미한다.

[0093] "프라이머"는 올리고뉴클레오타이드 주형에 어닐링된 다음 폴리머라제에 의해 주형-의존성 방식으로 연장될 수 있는 올리고뉴클레오타이드를 의미한다.

[0094] 본 발명의 다른 특색 및 이점은 하기 상세한 설명 및 청구범위로부터 명백할 것이다.

### 도면의 간단한 설명

[0095] 도 1은 공유 부착된 코딩된 소분자의 합성을 위해 코딩 올리고뉴클레오타이드 태그 및 보호된 1급 아민의 둘 다의 화학적 라이게이션을 위한 부위를 제공하는 헤드피스 올리고뉴클레오타이드로서 사용된 이중-가닥 헤어핀 구조를 예시하는 영상이다.

도 2는 예시적인 라이게이션 반응의 진행을 예시하는 겔의 영상이다.

도 3은 예시적인 라이게이션 반응의 진행을 예시하는 2종의 LCMS 트레이스의 영상이다.

도 4a는 보호된 아민의 탈보호 반응을 예시하는 영상이다.

도 4b는 탈보호 반응의 진행을 예시하는 겔의 영상이다.

도 4c는 탈보호 반응의 진행을 예시하는 LCMS 트레이스의 영상이다.

도 5a는 HP006과 1-시아노이미다졸의 반응의 생성물의 질량 스펙트럼의 영상이다.

도 5b는 HP006과 1-시아노이미다졸의 반응을 예시하는 영상이다.

### 발명을 실시하기 위한 구체적인 내용

- [0096] 코딩된 화학적 엔티티
- [0097] 본 발명은 화학적 엔티티, 1종 이상의 태그, 및 제1 화학적 엔티티 및 1종 이상의 태그와 작동가능하게 회합된 헤드피스를 포함하는 코딩된 화학적 엔티티를 생성하는 방법을 특색으로 한다. 화학적 엔티티, 헤드피스, 태그, 연결 및 이관능성 스페이서가 하기에 추가로 기재된다.
- [0098] 화학적 엔티티
- [0099] 본 발명의 화학적 엔티티 또는 구성원 (예를 들어, 소분자 또는 펩티드)은 1종 이상의 빌딩 블록을 포함할 수 있고, 임의로 1종 이상의 스캐폴드를 포함할 수 있다.
- [0100] 스캐폴드 S는 단일 원자 또는 분자 스캐폴드일 수 있다. 예시적인 단일 원자 스캐폴드는 탄소 원자, 붕소 원자, 질소 원자 또는 인 원자 등을 포함한다. 예시적인 다원자 스캐폴드는 시클로알킬 기, 시클로알케닐 기, 헤테로시클로알킬 기, 헤테로시클로알케닐 기, 아릴 기 또는 헤테로아릴 기를 포함한다. 헤테로아릴 스캐폴드의 특정한 실시양태는 트리아진, 예컨대 1,3,5-트리아진, 1,2,3-트리아진 또는 1,2,4-트리아진; 피리미딘; 피라진; 피리다진; 푸란; 피롤; 피롤린; 피롤리딘; 옥사졸; 피라졸; 이속사졸; 피란; 피리딘; 인돌; 인다졸; 또는 퓨린을 포함한다.
- [0101] 스캐폴드 S는 임의의 유용한 방법에 의해 태그에 작동가능하게 연결될 수 있다. 한 예에서, S는 헤드피스에 직접적으로 연결된 트리아진이다. 이러한 예시적인 스캐폴드를 수득하기 위해, 트리클로로트리아진 (즉, 3개의 염소를 갖는 트리아진의 염소화 전구체)을 헤드피스의 친핵성 기와 반응시킨다. 이러한 방법을 사용하여 S는 치환에 이용가능한 염소가 있는 3개의 위치를 가지며, 여기서 2개의 위치는 이용가능한 다양성 노드이고, 1개의 위치는 헤드피스에 부착된다. 다음으로, 빌딩 블록  $A_n$ 을 스캐폴드의 다양성 노드에 추가하고, 빌딩 블록  $A_n$ 을 코딩하는 태그  $A_n$  ("태그  $A_n$ ")을 헤드피스에 라이게이션시키며, 여기서 이들 2 단계는 임의의 순서로 수행될 수 있다. 이어서, 빌딩 블록  $B_n$ 을 남은 다양성 노드에 추가하고, 빌딩 블록  $B_n$ 을 코딩하는 태그  $B_n$ 을 태그  $A_n$ 의 말단에 라이게이션시킨다. 또 다른 예에서, S는 태그에 작동가능하게 연결된 트리아진이며, 여기서 트리클로로트리아진은 태그의 PEG, 지방족 또는 방향족 링커의 친핵성 기 (예를 들어, 아미노 기)와 반응한다. 빌딩 블록 및 회합된 태그는 상기 기재된 바와 같이 추가될 수 있다.
- [0102] 또 다른 예에서, S는 빌딩 블록  $A_n$ 에 작동가능하게 연결된 트리아진이다. 이러한 스캐폴드를 수득하기 위해, 2개의 다양성 노드 (예를 들어, 친전자성 기 및 친핵성 기, 예컨대 Fmoc-아미노산)를 갖는 빌딩 블록  $A_n$ 을 링커의 친핵성 기 (예를 들어, 헤드피스에 부착된 PEG, 지방족 또는 방향족 링커의 말단 기)와 반응시킨다. 이어서, 트리클로로트리아진을 빌딩 블록  $A_n$ 의 친핵성 기와 반응시킨다. 이러한 방법을 사용하여, S의 모든 3개의 염소 위치는 빌딩 블록에 대한 다양성 노드로서 사용된다. 본원에 기재된 바와 같이, 추가의 빌딩 블록 및 태그를 추가할 수 있고, 추가의 스캐폴드  $S_n$ 을 추가할 수 있다.
- [0103] 예시적인 빌딩 블록  $A_n$ 은, 예를 들어 아미노산 (예를 들어, 알파-, 베타-, 감마-, 델타- 및 엡실론-아미노산, 뿐만 아니라 천연 및 비천연 아미노산의 유도체), 아민과의 화학적-반응성 반응물 (예를 들어, 아지드 또는 알킨 쉐), 또는 티올 반응물 또는 그의 조합을 포함한다. 빌딩 블록  $A_n$ 의 선택은, 예를 들어 링커에 사용되는 반응성 기의 성질, 스캐폴드 모이어티의 성질 및 화학적 합성에 사용되는 용매에 따라 달라진다.
- [0104] 예시적인 빌딩 블록  $B_n$  및  $C_n$ 는 화학적 엔티티의 임의의 유용한 구조 단위, 예컨대 임의로 치환된 방향족 기 (예를 들어, 임의로 치환된 페닐 또는 벤질), 임의로 치환된 헤테로시클릴 기 (예를 들어, 임의로 치환된 퀴놀리닐, 이소퀴놀리닐, 인돌릴, 이소인돌릴, 아자인돌릴, 벤즈이미다졸릴, 아자벤즈이미다졸릴, 벤즈이속사졸릴, 피리디닐, 피페리딜 또는 피롤리디닐), 임의로 치환된 알킬 기 (예를 들어, 임의로 치환된 선형 또는 분지

형  $C_{1-6}$  알킬 기 또는 임의로 치환된  $C_{1-6}$  아미노알킬 기), 또는 임의로 치환된 카르보시클릴 기 (예를 들어, 임의로 치환된 시클로프로필, 시클로헥실 또는 시클로헥세닐)를 포함한다. 특히 유용한 빌딩 블록  $B_n$  및  $C_n$ 은, 반응성 기이거나 또는 반응성 기를 형성하도록 화학적으로 변형될 수 있는 하나 또는 임의적인 치환기를 갖는 임의로 치환된 기 (예를 들어, 본원에 기재된 임의의 것)와 같은, 1종 이상의 반응성 기를 갖는 것을 포함한다. 예시적인 반응성 기는 1개 이상의 아민 ( $-NR_2$ , 여기서 각각의 R은 독립적으로 H 또는 임의로 치환된  $C_{1-6}$  알킬임), 히드록시, 알콕시 ( $-OR$ , 여기서 R은 임의로 치환된  $C_{1-6}$  알킬, 예컨대 메톡시임), 카르복시 ( $-COOH$ ), 아마이드 또는 화학적-반응성 치환기를 포함한다. 예를 들어, 태그  $B_n$  또는  $C_n$ 에 제한 부위를 도입할 수 있으며, 여기서 복합체는 PCR 및 상응하는 제한 효소 중 1종에 의한 제한 소화를 수행함으로써 확인될 수 있다.

[0105] 가역적 고정화를 위한 부위

[0106] 일부 실시양태에서, 코딩된 화학적 엔티티는 임의로 가역적 고정화를 위한 부위를 포함한다. 가역적 고정화는 코딩된 라이브러리의 분할-및-혼합 합성 동안 완충제-교환 및 시약/오염물 제거를 용이하게 하기 위해 사용될 수 있다. 예를 들어, 제1 화학적 엔티티에 빌딩 블록을 추가하기 위한 화학 반응 후에, 복합체는 가역적으로 고정될 수 있다. 이어서 과량의 시약 및 용매가 제거될 수 있고, 라이게이션 반응을 위한 시약 및 용매가 추가되고, 이어서 복합체는 지지체로부터 탈착될 수 있다. 이러한 방법은, 라이브러리 및 올리고뉴클레오타이드 태그를 구축하는데 사용되는 단계가 용액 중에서 수행되도록 하는 한편 또는 대안적으로 신생 라이브러리를 가역적으로 고정시키는 한편, 고체 지지 합성의 이익, 예컨대 정제 및/또는 후속 단계와 상용성이지 않은 용매 및 시약의 제거의 용이성을 통합한다.

[0107] 예시적인 가역적 고정화 전략은 이중 및 삼중-가닥을 포함한, 치환된 올리고뉴클레오타이드 (2'-변형, PNA, LNA 등)를 포함한 올리고뉴클레오타이드 혼성화; 올리고뉴클레오타이드-이온 교환 상호작용 (예를 들어 DEAE-셀룰로스에 의함); 소분자-소분자 상호작용 (예를 들어 아다만탄-시클로텍스트린); 가역적 화학 (예를 들어 디설피드 결합 형성); 가역적 광화학 (예를 들어 시아노비닐 우리딘 광-가교); 가역적 화학적 가교 (예를 들어 외인성 추가된 반응성 엔티티에 의함); 고정된 금속 친화성 크로마토그래피 (예를 들어, 고정된 Ni-NTA와 His<sub>6</sub>); 항체-에피토프 상호작용 (예를 들어 고정된 항-FLAG 항체 및 FLAG 펩티드); 단백질-단백질 상호작용; 단백질-소분자 상호작용 (예를 들어 고정된 스트렙타비딘과 이미노비오틴 또는 고정된 말토스-결합 단백질 및 말토스); 가역적 올리고뉴클레오타이드 라이게이션 (예를 들어 제한된 dsDNA의 라이게이션에 이은 제한); 및 소수성 상호작용 (예를 들어 불소 태그 및 소수성 표면)을 포함한다. 일부 실시양태에서, 가역적 고정화를 위한 부위는 본원에 기재된 임의의 가역적 고정화 전략의 결합 쌍의 1종의 구성원, 예를 들어, 핵산, 펩티드 또는 소분자를 포함한다.

[0108] 헤드피스

[0109] 코딩된 화학적 엔티티에서, 헤드피스는 각각의 화학적 엔티티를 그의 코딩 올리고뉴클레오타이드 태그에 작동가능하게 연결시킨다. 일반적으로, 헤드피스는 추가로 유도체화될 수 있는 적어도 2개의 관능기를 갖는 출발 올리고뉴클레오타이드이며, 여기서 제1 관능기는 제1 화학적 엔티티 (또는 그의 성분)을 헤드피스에 작동가능하게 연결시키고, 제2 관능기는 1종 이상의 태그를 헤드피스에 작동가능하게 연결시킨다. 이관능성 스페이서는 헤드피스와 화학적 엔티티 사이의 스페이싱 모이어티로서 임의로 사용될 수 있다.

[0110] 헤드피스의 관능기는 화학적 엔티티의 성분과의 공유 결합 및 태그와의 또 다른 공유 결합을 형성하는데 사용될 수 있다. 성분은 다양성 노드를 갖는 스캐폴드 또는 빌딩 블록과 같은 소분자의 임의의 일부일 수 있다. 대안적으로, 헤드피스는 화학적 엔티티의 성분과 공유 연결을 형성하는데 사용되는 관능기 (예를 들어, 히드록실, 아민, 카르복실, 술폰히드릴, 알킬닐, 아지도 또는 포스페이트 기)로 종결되는 스페이서 (예를 들어, 라이브러리에서 형성하고자 하는 소분자로부터 헤드피스를 분리시키는 스페이싱 모이어티)를 제공하기 위해 유도체화될 수 있다. 스페이서는 헤드피스의 5'-말단, 내부 위치 중 하나, 또는 3'-말단에 부착될 수 있다. 스페이서가 내부 위치 중 하나에 부착되는 경우에, 스페이서는 유도체화된 염기 (예를 들어, 우리딘의 C5 위치)에 작동가능하게 연결될 수 있거나, 또는 관련 기술분야에 공지된 표준 기술을 사용하여 올리고뉴클레오타이드 내에 내부적으로 배치될 수 있다. 예시적인 스페이서가 본원에 기재되어 있다.

[0111] 헤드피스는 임의의 유용한 구조를 가질 수 있다. 헤드피스는, 예를 들어 1 내지 100개 뉴클레오타이드 길이, 바람직하게는 5 내지 20개 뉴클레오타이드 길이, 및 가장 바람직하게는 5 내지 15개 뉴클레오타이드 길이일 수 있다. 헤드피스는 단일-가닥 또는 이중-가닥일 수 있고, 본원에 기재된 바와 같은 천연 또는 변형된 뉴클레오타이드로 이루어질 수 있다. 예를 들어, 화학적 모이어티는 헤드피스의 3'-말단 또는 5'-말단에 작동가능하게 연결될 수



있다. 특정한 실시양태에서, 헤드피스는 서열 내의 상보적 염기에 의해 형성된 헤어핀 구조를 포함한다. 예를 들어, 화학적 모이어티는 헤드피스의 내부 위치, 3'-말단 또는 5'-말단에 작동가능하게 연결될 수 있다.

[0112] 일반적으로, 헤드피스는 5'- 또는 3'-말단 상에 중합, 효소적 라이게이션 또는 화학 반응에 의해 올리고뉴클레오타이드 태그의 결합을 가능하게 하는 비-자기-상보적 서열을 포함한다. 헤드피스는 올리고뉴클레오타이드 태그의 라이게이션 및 임의적인 정제 및 인산화 단계를 가능하게 할 수 있다. 마지막 태그의 추가 후에, 추가의 어댑터 서열을 마지막 태그의 5'-말단에 추가할 수 있다. 예시적인 어댑터 서열은 프라이머-결합 서열 또는 표지 (예를 들어, 비오틴)를 갖는 서열을 포함한다. 많은 빌딩 블록 및 상응하는 태그가 사용되는 경우에 (예를 들어, 100개), 필요한 개수의 태그를 생성하기 위해 올리고뉴클레오타이드 합성 단계 동안 혼합-및-분할 전략을 사용할 수 있다. DNA 합성을 위한 이러한 혼합-및-분할 전략은 관련 기술분야에 공지되어 있다. 생성된 라이브러리 구성원은 관심 표적(들)에 대한 결합 엔티티의 선택 후에 PCR에 의해 증폭될 수 있다.

[0113] 헤드피스 또는 복합체는 1개 이상의 프라이머-결합 서열을 임의로 포함할 수 있다. 예를 들어, 헤드피스는 헤어핀의 루프 영역에 증폭을 위한 프라이머-결합 영역으로서 역할을 하는 서열을 가지며, 여기서 프라이머-결합 영역은 헤드피스 내의 서열보다 그의 상보적인 프라이머 (예를 들어, 플랭킹 식별자 영역을 포함할 수 있음)에 대해 더 높은 용융 온도를 갖는다. 다른 실시양태에서, 복합체는 1종 이상의 빌딩 블록을 코딩하는 1종 이상의 태그의 각 측면에 (예를 들어, PCR 반응을 가능하게 하는) 2개의 프라이머-결합 서열을 포함한다. 대안적으로, 헤드피스는 5'- 또는 3'-말단에 1개의 프라이머-결합 서열을 함유할 수 있다. 다른 실시양태에서, 헤드피스는 헤어핀이고 루프 영역은 프라이머-결합 부위를 형성하거나 또는 프라이머-결합 부위는 루프의 3' 측의 헤드피스에 대한 올리고뉴클레오타이드의 혼성화를 통해 도입된다. 헤드피스의 3'-말단에 상동성인 영역을 함유하고 그의 5'-말단에서 (예를 들어, PCR 반응을 가능하게 하는) 프라이머-결합 영역을 보유하는 프라이머 올리고뉴클레오타이드는 헤드피스에 혼성화될 수 있고, 빌딩 블록 또는 빌딩 블록의 추가를 코딩하는 태그를 함유할 수 있다. 프라이머 올리고뉴클레오타이드는 생물정보학 분석을 위해 포함된 추가의 정보, 예컨대 무작위화된 뉴클레오타이드의 영역, 예를 들어 2 내지 16개 뉴클레오타이드 길이를 함유할 수 있다.

[0114] 헤드피스는 헤어핀 구조를 임의로 포함할 수 있고, 이러한 구조는 임의의 유용한 방법에 의해 달성될 수 있다. 예를 들어, 헤드피스는 예컨대 왓슨-크릭 DNA 염기 쌍형성 (예를 들어, 아데닌-티민 및 구아닌-시토신)에 의해 및/또는 위블 염기 쌍형성 (예를 들어, 구아닌-우라실, 이노신-우라실, 이노신-아데닌 및 이노신-시토신)에 의해 분자간 염기 쌍형성 파트너를 형성하는 상보적 염기를 포함할 수 있다. 또 다른 예에서, 헤드피스는 비변형된 뉴클레오타이드와 비교 시 더 높은 친화도 듀플렉스 형성을 형성할 수 있는 변형 또는 치환된 뉴클레오타이드를 포함할 수 있으며, 이러한 변형 또는 치환된 뉴클레오타이드는 관련 기술분야에 공지되어 있다. 또 다른 예에서, 헤드피스는 헤어핀 구조를 형성하기 위해 1개 이상의 가교된 염기를 포함한다. 예를 들어, 단일 가닥 내의 염기 또는 상이한 이중 가닥 내의 염기는, 예를 들어 프소랄렌을 사용함으로써 가교될 수 있다.

[0115] 헤드피스 또는 복합체는 검출을 가능하게 하는 1종 이상의 표지를 임의로 포함할 수 있다. 예를 들어, 헤드피스, 1종 이상의 올리고뉴클레오타이드 태그 및/또는 1종 이상의 프라이머 서열은 동위원소, 방사성영상화제, 마커, 트래이서, 형광 표지 (예를 들어, 로다민 또는 플루오레세인), 화학발광 표지, 양자점 및 리포터 분자 (예를 들어, 비오틴 또는 his-태그)를 포함할 수 있다.

[0116] 다른 실시양태에서, 헤드피스 또는 태그는 반-, 감소된- 또는 비-수성 (예를 들어, 유기) 조건 하에서 용해도를 지지하도록 변형될 수 있다. 예를 들어, 그의 상보적 염기에 수소 결합하는 능력은 유의하게 방해하지 않으면서, T 또는 C 염기의 C5 위치를 지방족쇄로 변형시킴으로써 헤드피스 또는 태그의 뉴클레오타이드 염기를 보다 소수성이 되게 할 수 있다. 예시적인 변형 또는 치환된 뉴클레오타이드는 5'-디메톡시트리틸-5-(1-프로피닐)-2'-데옥시시티딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 5'-디메톡시트리틸-5-(1-프로피닐)-2'-데옥시우리딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 5'-디메톡시트리틸-5-플루오로-2'-데옥시우리딘, 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 및 5'-디메톡시트리틸-5-(피렌-1-일-에티닐)-2'-데옥시우리딘 또는 3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트이다.

[0117] 또한, 헤드피스 올리고뉴클레오타이드는 유기 용매 중 용해도를 촉진시키는 변형으로 산재될 수 있다. 예를 들어, 아조벤젠 포스포르아미다이트는 헤드피스 설계 내로 소수성 모이어티를 도입할 수 있다. 소수성 아미다이트의 헤드피스 내로의 이러한 삽입은 분자 내의 어디에서나 이루어질 수 있다. 그러나, 삽입은, 일단 선택이 완료되면 라이브러리 합성 또는 후속 PCR 동안, 또는 태그 디컨볼루션에 사용되는 경우에 마이크로어레이 분석 동안, 추가의 DNA 태그를 사용한 후속 태그부착을 방해할 수 없다. 본원에 기재된 헤드피스 설계에의 이러한

추가는 헤드피스를, 예를 들어 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% 또는 100% 유기 용매 중에서 용해되게 할 것이다. 따라서, 헤드피스 설계에의 소수성 잔기의 추가는 헤드피스가 올리고뉴클레오타이드 태그부착에 적격이 되게 하면서, 반- 또는 비-수성 (예를 들어, 유기) 조건에서의 개선된 용해도를 가능하게 한다. 추가로, 라이브러리에 후속 도입된 DNA 태그는 T 또는 C 염기의 C5 위치에서 또한 변형될 수 있으며, 이로써 이들은 또한 라이브러리 합성의 후속 단계를 위해 라이브러리를 보다 소수성있게 하고 유기 용매 중에서 보다 가용성이 되게 한다.

[0118] 특정한 실시양태에서, 헤드피스 및 제1 태그는 동일한 엔티티일 수 있으며, 즉 복수의 헤드피스-태그 엔티티는 공통 부분 (예를 들어, 프라이머-결합 영역)은 모두 공유하고 또 다른 부분 (예를 들어, 코딩 영역)은 모두 상이하게 구축될 수 있다. 이들은 "분할" 단계에서 사용될 수 있고, 코딩하는 이벤트가 발생한 이후에 풀링될 수 있다.

[0119] 특정한 실시양태에서, 헤드피스는 예를 들어, 예컨대 특정 라이브러리와 관련된 특정한 서열을 사용하는 것에 의해, 제1 분할(들) 단계를 코딩하는 서열 또는 라이브러리의 아이덴티티를 코딩하는 서열을 포함함으로써 정보를 코딩할 수 있다.

[0120] 올리고뉴클레오타이드 태그

[0121] 본원에 기재된 올리고뉴클레오타이드 태그 (예를 들어, 태그 또는 헤드피스의 부분 또는 테일피스의 부분)는 임의의 유용한 정보, 예컨대 분자, 화학적 엔티티의 부분, 성분 (예를 들어, 스캐폴드 또는 빌딩 블록)의 추가, 라이브러리의 헤드피스, 라이브러리의 아이덴티티, 1종 이상의 라이브러리 구성원의 사용 (예를 들어, 라이브러리의 분취물에서의 구성원의 사용), 및/또는 라이브러리 구성원의 기원 (예를 들어, 기원 서열의 사용에 의함)을 코딩하는데 사용될 수 있다.

[0122] 올리고뉴클레오타이드의 임의의 서열은 임의의 정보를 코딩하는데 사용될 수 있다. 따라서, 1개의 올리고뉴클레오타이드 서열은 2종 이상의 유형의 정보를 코딩하거나 또는 1종 이상의 유형의 정보를 또한 코딩하는 출발 올리고뉴클레오타이드를 제공하기 위한 것과 같은 1종 초과 목적을 제공할 수 있다. 예를 들어, 제1 태그는 제1 빌딩 블록의 추가, 뿐만 아니라 라이브러리의 확인을 코딩할 수 있다. 또 다른 예에서, 화학적 엔티티를 태그에 작동가능하게 연결시키는 출발 올리고뉴클레오타이드를 제공하기 위해 헤드피스가 사용될 수 있으며, 여기서 헤드피스는 라이브러리의 아이덴티티를 코딩하는 서열 (즉, 라이브러리-확인 서열)을 추가로 포함한다. 따라서, 본원에 기재된 임의의 정보는 분리된 올리고뉴클레오타이드 태그 내에 코딩될 수 있거나, 또는 동일한 올리고뉴클레오타이드 서열 (예를 들어, 올리고뉴클레오타이드 태그, 예컨대 태그 또는 헤드피스) 내에 조합 및 코딩될 수 있다.

[0123] 빌딩 블록 서열은 빌딩 블록의 아이덴티티 및/또는 빌딩 블록으로 수행되는 결합 반응의 유형을 코딩한다. 이러한 빌딩 블록 서열은 태그에 포함되어 있으며, 여기서 태그는 하기 기재된 1종 이상의 유형의 서열 (예를 들어, 라이브러리-확인 서열, 사용 서열 및/또는 기원 서열)을 임의로 포함할 수 있다.

[0124] 라이브러리-확인 서열은 특정한 라이브러리의 아이덴티티를 코딩한다. 2종 이상의 라이브러리가 혼합되게 하기 위해, 라이브러리 구성원은 예컨대 라이브러리-확인 태그 (즉, 라이브러리-확인 서열을 포함하는 올리고뉴클레오타이드)에, 라이게이션된 태그에, 헤드피스 서열의 일부에, 또는 테일피스 서열에 1종 이상의 라이브러리-확인 서열을 함유할 수 있다. 이들 라이브러리-확인 서열은 코딩 관계를 추론하는데 사용될 수 있으며, 여기서 태그의 서열은 번역되고, 화학적 (합성) 히스토리 정보와 상호관련된다. 따라서, 이들 라이브러리-확인 서열은 선택, 증폭, 정제, 서열분석 등을 위해 2종 이상의 라이브러리가 함께 혼합되게 한다.

[0125] 사용 서열은 라이브러리의 개별 분취물 내의 1종 이상의 라이브러리 구성원의 히스토리 (즉, 사용)를 코딩한다. 예를 들어, 분리된 분취물은 상이한 반응 조건, 빌딩 블록 및/또는 선택 단계로 처리될 수 있다. 특히, 이러한 서열을 사용하여 이러한 분취물을 확인하고 그의 히스토리 (사용)를 추론할 수 있으며, 이에 의해 선택, 증폭, 정제, 서열분석 등을 위해 샘플을 함께 혼합시키고자 하는 목적으로 히스토리 (사용)가 상이한 동일한 라이브러리의 분취물이 함께 혼합되게 한다 (예를 들어, 별개의 선택 실험). 이들 사용 서열은 헤드피스, 테일피스, 태그, 사용 태그 (즉, 사용 서열을 포함하는 올리고뉴클레오타이드), 또는 본원에 기재된 임의의 다른 태그 (예를 들어, 라이브러리-확인 태그 또는 기원 태그)에 포함될 수 있다.

[0126] 기원 서열은 라이브러리 구성원의 기원을 코딩하는 임의의 유용한 길이 (예를 들어, 약 6개 올리고뉴클레오타이드)의 축중성 (무작위, 확률적으로-생성된) 올리고뉴클레오타이드 서열이다. 이러한 서열은 모든 측면에서 그 외에는 동일한 라이브러리 구성원을 서열 정보에 의해 구별가능한 엔티티로 확률적으로 세분하는 역할을 하며, 이로써 고유한 전구 주형 (예를 들어, 선택된 라이브러리 구성원)으로부터 유래된 증폭 생성물

의 관찰은 동일한 전구 주형 (예를 들어, 선택된 라이브러리 구성원)으로부터 유래된 다중 증폭 생성물의 관찰과 구별될 수 있다. 예를 들어, 라이브러리 형성 후 및 선택 단계 전에, 각각의 라이브러리 구성원은 예컨대 기원 태그에서 상이한 기원 서열을 포함할 수 있다. 선택 후에, 선택된 라이브러리 구성원을 증폭시켜 증폭 생성물을 생성할 수 있고, (예를 들어 기원 태그 내에) 기원 서열을 포함하는 것으로 예상되는 라이브러리 구성원의 부분을 관찰할 수 있고, 각각의 다른 라이브러리 구성원 내의 기원 서열과 비교할 수 있다. 기원 서열은 축중성이기 때문에, 각각의 라이브러리 구성원의 각각의 증폭 생성물은 상이한 기원 서열을 가질 것이다. 그러나, 증폭 생성물에서 동일한 기원 서열의 관찰은 동일한 주형 분자로부터 유래된 다중 앰플리콘을 나타낼 수 있다. 증폭-후와 대조적으로, 증폭 전에 코딩 태그의 집단의 통계 및 집단통계를 결정하는 것이 바람직한 경우에, 기원 태그가 사용될 수 있다. 이들 기원 서열은 헤드피스, 테일피스, 태그, 기원 태그 (즉 기원 서열을 포함하는 올리고뉴클레오타이드), 또는 본원에 기재된 임의의 다른 태그 (예를 들어, 라이브러리-확인 태그 또는 사용 태그)에 포함될 수 있다.

[0127] 본원에 기재된 임의의 유형의 서열이 헤드피스에 포함될 수 있다. 예를 들어, 헤드피스는 빌딩 블록 서열, 라이브러리-확인 서열, 사용 서열 또는 기원 서열 중 1종 이상을 포함할 수 있다.

[0128] 본원에 기재된 임의의 이들 서열이 테일피스에 포함될 수 있다. 예를 들어, 테일피스는 라이브러리-확인 서열, 사용 서열 또는 기원 서열 중 1종 이상을 포함할 수 있다.

[0129] 본원에 기재된 임의의 태그는 고정된 서열을 갖는 5'- 또는 3'-말단에서의 또는 그에 근접한 연결기를 포함할 수 있다. 연결기는 반응성 기 (예를 들어, 화학적-반응성 기 또는 광-반응성 기)를 제공함으로써 또는 연결을 가능하게 하는 작용제 (예를 들어, 연결기(들) 또는 가교 올리고뉴클레오타이드 내의 삽입성 모이어티 또는 가역적 반응성 기의 작용제)를 위한 부위를 제공함으로써 연결 (예를 들어, 화학적 연결)의 형성을 용이하게 한다. 각각의 5'-연결기는 동일하거나 상이할 수 있고, 각각의 3'-연결기는 동일하거나 상이할 수 있다. 1종 초과 태그를 갖는 예시적인 비제한적 복합체에서, 각각의 태그는 5'-연결기 및 3'-연결기를 포함할 수 있으며, 여기서 각각의 5'-연결기는 동일한 서열을 갖고, 각각의 3'-연결기는 동일한 서열을 갖는다 (예를 들어, 여기서 5'-연결기의 서열은 3'-연결기의 서열과 동일하거나 상이할 수 있음). 연결기는 1개 이상의 연결을 위해 사용될 수 있는 서열을 제공한다. 릴레이 프라이머의 결합 또는 가교 올리고뉴클레오타이드의 혼성화를 가능하게 하기 위해, 연결기는 연결 (예를 들어, 폴리머라제가 감소된 판독 또는 전위 능력을 갖게 하는 연결, 예컨대 화학적 연결)을 가능하게 하는 1종 이상의 관능기를 포함할 수 있다.

[0130] 이들 서열은 올리고뉴클레오타이드에 대해 본원에 기재된 임의의 변형, 예컨대 유기 용매 중의 용해도를 촉진시키거나 (예를 들어, 예컨대 헤드피스에 대해 본원에 기재된 임의의 것), 천연 포스포디에스테르 연결의 유사체 (예를 들어, 포스포로티오에이트 유사체)를 제공하거나, 또는 1종 이상의 비-천연 올리고뉴클레오타이드 (예를 들어, 2'-치환된 뉴클레오타이드, 예컨대 2'-O-메틸화 뉴클레오타이드 및 2'-플루오로 뉴클레오타이드, 또는 본원에 기재된 임의의 것)를 제공하는 1종 이상의 변형을 포함할 수 있다.

[0131] 이들 서열은 올리고뉴클레오타이드에 대해 본원에 기재된 임의의 특징을 포함할 수 있다. 예를 들어, 이들 서열은 20개 미만의 뉴클레오타이드인 태그 (예를 들어, 본원에 기재된 바와 같음)에 포함될 수 있다. 다른 예에서, 이들 서열 중 1종 이상을 포함하는 태그는 거의 동일한 질량을 갖거나 (예를 들어, 각각의 태그는 특정 가변성을 코딩하는 특정 세트 내의 태그 사이의 평균 질량에서 약 +/-10%인 질량을 가짐); 프라이머-결합 (예를 들어, 불변) 영역이 결여되어 있거나; 불변 영역이 결여되어 있거나; 또는 감소된 길이 (예를 들어 30개 뉴클레오타이드 미만, 25개 뉴클레오타이드 미만, 20개 뉴클레오타이드 미만, 19개 뉴클레오타이드 미만, 18개 뉴클레오타이드 미만, 17개 뉴클레오타이드 미만, 16개 뉴클레오타이드 미만, 15개 뉴클레오타이드 미만, 14개 뉴클레오타이드 미만, 13개 뉴클레오타이드 미만, 12개 뉴클레오타이드 미만, 11개 뉴클레오타이드 미만, 10개 뉴클레오타이드 미만, 9개 뉴클레오타이드 미만, 8개 뉴클레오타이드 미만 또는 7개 뉴클레오타이드 미만의 길이)의 불변 영역을 갖는다.

[0132] 이러한 길이의 라이브러리 및 올리고뉴클레오타이드를 서열분석하는 전략은, 판독 충실도 또는 서열분석 심도를 각각 증가시키기 위해 연속 또는 연쇄 전략을 임의로 포함할 수 있다. 특히, 프라이머-결합 영역이 결여되어 있는 코딩된 라이브러리의 선택은 SELEX에 대한 문헌에 기재된 바 있고, 예컨대 본원에 참조로 포함되는 문헌 [Jarosch et al., Nucleic Acids Res. 34: e86 (2006)]에 기재되어 있다. 예를 들어, 라이브러리 구성원은 복합체의 5'-말단에 제1 어댑터 서열 및 복합체의 3'-말단에 제2 어댑터 서열을 포함하도록 (예를 들어, 선택 단계 후에) 변형될 수 있으며, 여기서 제1 서열은 제2 서열에 실질적으로 상보적이고, 듀플렉스의 형성을 발생시킨다. 수율을 추가로 개선시키기 위해, 2개의 고정된 현수 뉴클레오타이드 (예를 들어, CC)가 5'-말단에 추가된다.



- [0133] 연결
- [0134] 본 발명의 연결은 정보를 코딩하는 올리고뉴클레오타이드 사이 (예를 들어, 예컨대 헤드피스와 태그 사이, 2개의 태그 사이 또는 태그와 테일피스 사이)에 존재한다. 예시적인 연결은 포스포디에스테르, 포스포네이트 및 포스포로티오에이트를 포함한다. 일부 실시양태에서, 폴리머라제는 1개 이상의 연결을 통해 감소된 판독 또는 전위 능력을 갖는다. 특정 실시양태에서, 화학적 연결은 화학적-반응성 기 예컨대 모노포스페이트 및/또는 히드록실기, 광-반응성 기, 삽입성 모이어티, 가교 올리고뉴클레오타이드 또는 가역적 공-반응성 기 중 1종 이상을 포함한다.
- [0135] 연결은 폴리머라제가 그러한 연결을 통해 감소된 판독 또는 전위 능력을 갖는지 여부를 결정하기 위해 시험될 수 있다. 이러한 능력은 임의의 유용한 방법, 예컨대 액체 크로마토그래피-질량 분광측정법, RT-PCR 분석, 서열 집단통계 및/또는 PCR 분석에 의해 시험될 수 있다.
- [0136] 일부 실시양태에서, 화학적 라이게이션은 연결을 제공하기 위해 1개 이상의 화학적-반응성 쌍, 예컨대 모노포스페이트 및 히드록실의 사용을 포함한다. 본원에 기재된 바와 같이, 판독가능한 연결은 화학적 라이게이션에 의해, 예를 들어, 시아노이미다졸 및 2가 금속 공급원 (예를 들어,  $ZnCl_2$ )의 존재 하에 5'- 또는 3'-말단 상의 모노포스페이트, 모노포스포티오에이트 또는 모노포스포네이트와 5'- 또는 3'-말단 상의 히드록실 기의 반응에 의해 합성될 수 있다.
- [0137] 다른 예시적인 화학적-반응성 쌍은 휘스겐 1,3-쌍극자 고리화첨가 반응을 통해 트리아졸을 형성하는 임의로 치환된 알킬닐 기 및 임의로 치환된 아지도 기를 포함하는 쌍; 딜스-알더 반응을 통해 시클로알케닐을 형성하는 4  $\pi$ -전자계를 갖는 임의로 치환된 디엔 (예를 들어, 임의로 치환된 1,3-불포화 화합물, 예컨대 임의로 치환된 1,3-부타디엔, 1-메톡시-3-트리메틸실릴옥시-1,3-부타디엔, 시클로펜타디엔, 시클로헥사디엔 또는 푸란) 및 2  $\pi$ -전자계를 갖는 임의로 치환된 친디엔체 또는 임의로 치환된 친헤테로디엔체 (예를 들어, 임의로 치환된 알케닐 기 또는 임의로 치환된 알킬닐 기)를 포함하는 쌍; 개환 반응을 통해 헤테로알킬을 형성하는 친핵체 (예를 들어, 임의로 치환된 아민 또는 임의로 치환된 티올)과 변형 헤테로시클릴 친전자체 (예를 들어, 임의로 치환된 에폭사이드, 아지리딘, 아지리디늄 이온 또는 에피술포늄 이온)를 포함하는 쌍; 예컨대 5'-아이오도 dT를 함유하는 올리고뉴클레오타이드와 3'-포스포로티오에이트 올리고뉴클레오타이드의 스플린트 라이게이션에서와 같은 포스포로티오에이트 기와 아이오도 기를 포함하는 쌍; 상업적으로 입수가능한 3'-글리세틸-변형된 올리고뉴클레오타이드를 산화시키는 것에 의해 임의로 수득될 수 있는 3'-알데히드-변형된 올리고뉴클레오타이드와 5'-아미노 올리고뉴클레오타이드 (즉 환원성 아미노화 반응에서) 또는 5'-히드라지도 올리고뉴클레오타이드의 반응에서와 같은, 임의로 치환된 아미노 기와 알데히드 기 또는 케톤 기를 포함하는 쌍; 임의로 치환된 아미노 기 및 카르복실산 기 또는 티올 기의 쌍 (예를 들어, 숙신이미딜 트랜스-4-(말레이미딜메틸)시클로헥산-1-카르복실레이트 (SMCC) 또는 1-에틸-3-(3-디메틸아미노프로필)카르보디이미드 (EDAC)의 사용의 존재 또는 부재 하); 임의로 치환된 히드라진 및 알데히드 또는 케톤 기의 쌍; 임의로 치환된 히드록실아민 및 알데히드 또는 케톤 기의 쌍; 또는 친핵체 및 임의로 치환된 알킬 할라이드의 쌍이다.
- [0138] 백금 착물, 알킬화제 또는 푸란-변형된 뉴클레오타이드는 가닥간 또는 가닥내 연결을 형성하기 위한 화학적-반응성 기로서 또한 사용될 수 있다. 이러한 작용제는 2개의 올리고뉴클레오타이드 사이에서 사용될 수 있고, 가교 올리고뉴클레오타이드 내에 임의로 존재할 수 있다.
- [0139] 예시적인 비제한적 백금 착물은 시스플라틴 (예를 들어, GG 가닥내 연결을 형성하는 시스-디아민디클로로백금(II)), 트랜스플라틴 (예를 들어, GXG 가닥간 연결을 형성하는 트랜스-디아민디클로로백금(II), 여기서 X는 임의의 뉴클레오타이드일 수 있음), 카르보플라틴, 피콜라틴 (ZD0473), 오르마플라틴, 또는 예를 들어 GC, CG, AG 또는 GG 연결을 형성하는 옥살리플라틴을 포함한다. 임의의 이들 연결은 가닥간 또는 가닥내 연결일 수 있다.
- [0140] 예시적인 비제한적 알킬화제는 질소 머스타드 (예를 들어, GG 연결을 형성하는 메클로레타민), 클로람부실, 멜팔란, 시클로포스파미드, 시클로포스파미드의 전구약물 형태 (예를 들어, 4-히드로퍼옥시시클로포스파미드 및 이포스파미드), 1,3-비스(2-클로로에틸)-1-니트로소우레아 (BCNU, 카르무스틴), 아지리딘 (예를 들어, GG 또는 AG 연결을 형성하는 미토마이신 C, 트리에틸렌멜라민 또는 트리에틸렌티오포스포르아미드 (티오-테파)), 헥사메틸멜라민, 알킬 술포네이트 (예를 들어, GG 연결을 형성하는 부술판) 또는 니트로소우레아 (예를 들어, GG 또는 CG 연결을 형성하는 2-클로로에틸니트로소우레아, 예컨대 카르무스틴 (BCNU), 클로로조토신, 로무스틴 (CCNU) 및 세무스틴 (메틸-CCNU))을 포함한다. 임의의 이들 연결은 가닥간 또는 가닥내 연결일 수 있다.
- [0141] 푸란-변형된 뉴클레오타이드가 또한 연결을 형성하는데 사용될 수 있다. 계내 산화 시 (예를 들어, N-브로모숙신

이미드 (NBS)에 의함), 푸란 모이어티는 상보적 염기와 반응하여 가닥간 연결을 형성하는 반응성 옥소-에날 유도체를 형성한다. 일부 실시양태에서, 푸란-변형된 뉴클레오티드는 상보적 A 또는 C 뉴클레오티드와 함께 연결을 형성한다. 예시적인 비제한적 푸란-변형된 뉴클레오티드는 임의의 2'-(푸란-2-일)프로파노일아미노-변형된 뉴클레오티드; 또는 2-(푸란-2-일)에틸 글리콜 핵산의 비-시클릭, 변형된 뉴클레오티드를 포함한다.

[0142] 광-반응성 기가 또한 반응성 기로서 사용될 수 있다. 예시적인 비제한적 광-반응성 기는 삽입성 모이어티, 프소랄렌 유도체 (예를 들어, 프소랄렌, HMT-프소랄렌 또는 8-메톡시프소랄렌), 임의로 치환된 시아노비닐카르바졸 기, 임의로 치환된 비닐카르바졸 기, 임의로 치환된 시아노비닐 기, 임의로 치환된 아크릴아미드 기, 임의로 치환된 다이아지린 기, 임의로 치환된 벤조페논 (예를 들어, 4-벤조일벤조산 또는 벤조페논 이소티오시아네이트의 숙신이미딜 에스테르), 임의로 치환된 5-(카르복시)비닐-우리딘 기 (예를 들어, 5-(카르복시)비닐-2'-데옥시우리딘) 또는 임의로 치환된 아지드 기 (예를 들어, 아릴 아지드 또는 할로젠화 아릴 아지드, 예컨대 4-아지도-2,3,5,6-테트라플루오로벤조산 (ATFB)의 숙신이미딜 에스테르)를 포함한다.

[0143] 삽입성 모이어티가 또한 반응성 기로서 사용될 수 있다. 예시적인 비제한적 삽입성 모이어티는 프소랄렌 유도체, 알칼로이드 유도체 (예를 들어, 베르베린, 팔마틴, 코랄린, 상귀나린 (예를 들어, 이미늄 또는 그의 알칸올 아민 형태) 또는 아리스톨로락탐-β-D-글루코시드), 에티뎀 양이온 (예를 들어, 브로민화 에티뎀), 아크리딘 유도체 (예를 들어, 프로플라빈, 아크리플라빈 또는 암사크린), 안트라시클린 유도체 (예를 들어, 독소루비신, 에피루비신, 다우노루비신 (다우노마이신), 이다루비신 및 아클라루비신) 또는 탈리도미드를 포함한다.

[0144] 가교 올리고뉴클레오티드의 경우에, 가닥간 또는 가닥내 연결을 형성하기 위해 임의의 유용한 반응성 기 (예를 들어, 본원에 기재된 것)가 또한 사용될 수 있다. 예시적인 반응성 기는 화학적-반응성 기, 광-반응성 기, 삽입성 모이어티, 및 가역적 공-반응성 기를 포함한다. 가교 올리고뉴클레오티드와 함께 사용하기 위한 가교제는, 비제한적으로, 알킬화제 (예를 들어, 본원에 기재된 바와 같음), 시스플라틴 (시스-디아민디클로로백금(II)), 트랜스-디아민디클로로백금(II), 프소랄렌, HMT-프소랄렌, 8-메톡시프소랄렌, 푸란-변형된 뉴클레오티드, 2-플루오로-데옥시이노신 (2-F-dI), 5-브로모-데옥시시토신 (5-Br-dC), 5-브로모-데옥시우리딘 (5-Br-dU), 5-아이오도-데옥시시토신 (5-I-dC), 5-아이오도-데옥시우리딘 (5-I-dU), 숙신이미딜 트랜스-4-(말레이미딜메틸)시클로핵산-1-카르복실레이트, SMCC, EDAC 또는 숙신이미딜 아세틸티오아세테이트 (SATA)를 포함한다.

[0145] 올리고뉴클레오티드는, 다양한 티올 반응성 기 예컨대 말레이미드, 할로젠, 및 아이오도아세트아미드와 반응할 수 있고 따라서 2개의 올리고뉴클레오티드를 가교시키는데 사용할 수 있는 티올 모이어티를 함유하도록 또한 변형될 수 있다. 티올 기는 올리고뉴클레오티드의 5'- 또는 3'- 말단에 연결될 수 있다.

[0146] 듀플렉스 올리고뉴클레오티드 사이에서 피리미딘 (예를 들어, 티미딘) 위치에서의 가닥간 가교를 위해, 삽입성, 광-반응성 모이어티 프소랄렌이 선택될 수 있다. 프소랄렌은 듀플렉스 내로 삽입되고, 자외선 (약 254 nm) 조사 시 피리미딘과, 우세하게는 5'-TpA 부위에서 공유적 가닥간 가교를 형성한다. 프소랄렌 모이어티는 변형된 올리고뉴클레오티드에 공유 부착될 수 있다 (예를 들어, 알칸 쇠, 예컨대 C<sub>1-10</sub> 알킬 또는 폴리에틸렌 글리콜 기, 예컨대 -(CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>-에 의함, 여기서 n은 1 내지 50의 정수임). 예시적인 프소랄렌 유도체가 또한 사용될 수 있으며, 여기서 비제한적 유도체는 4'-(히드록시에톡시메틸)-4,5',8-트리메틸프소랄렌 (HMT-프소랄렌) 및 8-메톡시프소랄렌을 포함한다.

[0147] 연결을 도입하기 위해 가교 올리고뉴클레오티드의 다양한 부분이 변형될 수 있다. 예를 들어, 2개의 인접한 올리고뉴클레오티드를 연결시키는데 올리고뉴클레오티드 내의 말단 포스포로티오에이트가 또한 사용될 수 있다. 올리고뉴클레오티드 내의 가교 변형으로서 할로젠화 우라실/시토신이 또한 사용될 수 있다. 예를 들어, 2-플루오로-데옥시이노신 (2-F-dI) 변형된 올리고뉴클레오티드는 디설파이드-함유 디아민 또는 티오프로필아민과 반응하여 디설파이드 연결을 형성할 수 있다.

[0148] 하기 기재된 바와 같이, 가역적 공-반응성 기는 시아노비닐카르바졸 기, 시아노비닐 기, 아크릴아미드 기, 티올 기 또는 술포닐에틸 티오에테르로부터 선택되는 것을 포함한다. 상보적 가닥 내의 피리미딘 염기 (예를 들어, 시토신, 티민 및 우라실, 뿐만 아니라 그의 변형된 염기)에의 가교를 위해 임의로 치환된 시아노비닐카르바졸 (CNV) 기가 또한 올리고뉴클레오티드에서 사용될 수 있다. CNV 기는 366 nm에서의 조사 시 인접한 피리미딘 염기와의 [2+2] 고리화첨가를 촉진시켜, 가닥간 가교를 발생시킨다. 312 nm에서의 조사는 가교를 역전시키고, 따라서 올리고뉴클레오티드 가닥의 가역적 가교 방법을 제공한다. 비제한적 CNV 기는 3-시아노비닐카르바졸이며, 이는 카르복시비닐카르바졸 뉴클레오티드로서 (예를 들어, 3-카르복시비닐카르바졸-1'-β-데옥시리보시드-5'-트리포스페이트로서) 포함될 수 있다.

- [0149] CNV 기는 반응성 시아노 기를 또 다른 반응성 기로 대체하여 임의로 치환된 비닐카르바졸 기를 제공하기 위해 변형될 수 있다. 비닐카르바졸 기를 위한 예시적인 비제한적 반응성 기는  $-CONR_{N1}R_{N2}$ 의 아미드 기를 포함하며, 여기서 각각의  $R_{N1}$  및  $R_{N2}$ 는 동일하거나 상이할 수 있고, 독립적으로 H 및  $C_{1-6}$  알킬, 예를 들어,  $-CONH_2$ ; 카르복실 기  $-CO_2H$ ; 또는  $C_{2-7}$  알콕시카르보닐 기 (예를 들어, 메톡시카르보닐)이다. 또한, 반응성 기는 비닐 기의 알파 또는 베타 탄소 상에 위치할 수 있다. 예시적인 비닐카르바졸 기는 본원에 기재된 바와 같은 시아노비닐카르바졸 기; 아미도비닐카르바졸 기 (예를 들어, 아미도비닐카르바졸 뉴클레오타이드, 예컨대 3'-아미도비닐카르바졸-1'- $\beta$ -데옥시리보시드-5'-트리포스페이트); 카르복시비닐카르바졸 기 (예를 들어, 카르복시비닐카르바졸 뉴클레오타이드, 예컨대 3'-카르복시비닐카르바졸-1'- $\beta$ -데옥시리보시드-5'-트리포스페이트); 및  $C_{2-7}$  알콕시카르보닐 비닐카르바졸 기 (예를 들어, 알콕시카르보닐비닐카르바졸 뉴클레오타이드, 예컨대 3'-메톡시카르보닐비닐카르바졸-1'- $\beta$ -데옥시리보시드-5'-트리포스페이트)를 포함한다. 추가의 임의로 치환된 비닐카르바졸 기 및 이러한 기를 갖는 뉴클레오타이드가 미국 특허 번호 7,972,792 및 문헌 [Yoshimura and Fujimoto, Org. Lett. 10:3227-3230 (2008)]의 화학식에서 제공되며, 이들 둘 다는 그 전문이 본원에 참조로 포함된다.
- [0150] 다른 가역적 반응성 기는 디술피드를 형성하는 티올 기 및 또 다른 티올 기, 뿐만 아니라 술포닐에틸 티오에테르를 형성하는 티올 기 및 비닐 술포ن 기를 포함한다. 티올-티올 기는 비스-((N-아이오도아세틸)피페라지닐)술포노다민과의 반응에 의해 형성되는 연결을 임의로 포함할 수 있다. 다른 가역적 반응성 기 (예를 들어, 예컨대 일부 광-반응성 기)는 임의로 치환된 벤조페논 기를 포함한다. 비제한적 예는 벤조페논 우라실 (BPU)이며, 이는 BPU-함유 올리고뉴클레오타이드 듀플렉스의 가닥간 가교의 부위- 및 서열-선택적 형성에 사용될 수 있다. 이러한 가교는 가열 시 역전되어, 2개의 올리고뉴클레오타이드 가닥의 가역적 가교 방법을 제공할 수 있다.
- [0151] 다른 실시양태에서, 화학적 라이게이션은 예를 들어 선택-후 PCR 분석 및 서열분석을 위해 포스포디에스테르 결합의 유사체를 도입하는 것을 포함한다. 포스포디에스테르의 예시적인 유사체는 포스포로티오에이트 연결 (예를 들어, 포스포로티오에이트 기 및 이탈기, 예컨대 아이오도 기의 사용에 의해 도입된 것과 같음), 포스포르아미드 연결, 또는 포스포로티오에이트 연결 (예를 들어, 포스포로티오에이트 기 및 이탈기, 예컨대 아이오도 기의 사용에 의해 도입된 것과 같음)을 포함한다.
- [0152] 본원에 기재된 임의의 기의 경우에 (예를 들어, 화학적-반응성 기, 광-반응성 기, 삽입성 모이어티, 가교 올리고뉴클레오타이드 또는 가역적 공-반응성 기), 기는 올리고뉴클레오타이드의 말단에 또는 그에 근접하여 또는 5'-와 3'-말단 사이에 혼입될 수 있다. 또한, 1개 이상의 기가 각각의 올리고뉴클레오타이드 내에 존재할 수 있다. 반응성 기의 쌍이 요구되는 경우에, 올리고뉴클레오타이드는 기의 쌍 사이의 반응을 용이하게 하도록 설계될 수 있다. 피리미딘 염기와 공-반응하는 시아노카르바졸 기의 비제한적 예에서, 제1 올리고뉴클레오타이드는 5'-말단에 또는 그에 근접하여 시아노카르바졸 기를 포함하도록 설계될 수 있다. 이러한 예에서, 제2 올리고뉴클레오타이드는 제1 올리고뉴클레오타이드와 상보적이라도, 그리고 제1 및 제2 올리고뉴클레오타이드가 혼성화되는 경우에 시아노비닐카르바졸 기와 정렬되는 위치에서 공-반응성 피리미딘 염기를 포함하도록 설계될 수 있다. 본원의 임의의 기, 및 1개 이상의 기를 갖는 임의의 올리고뉴클레오타이드는, 기 사이의 반응이 1개 이상의 연결의 형성을 용이하게 하도록 설계될 수 있다.
- [0153] 이관능성 스페이서
- [0154] 헤드피스와 화학적 엔티티 사이의 이관능성 스페이서는 적절한 스페이싱 모이어티를 제공하기 위해 및/또는 유기 용매 중에서의 헤드피스의 용해도를 증가시키기 위해 달라질 수 있다. 헤드피스와 소분자 라이브러리를 커플링시킬 수 있는 매우 다양한 스페이서가 상업적으로 입수가능하다. 스페이서는 전형적으로 선형 또는 분지형쇄로 이루어지고,  $C_{1-10}$  알킬, 1 내지 10개의 원자의 헥테로알킬,  $C_{2-10}$  알케닐,  $C_{2-10}$  알키닐,  $C_{5-10}$  아릴, 3 내지 20개 원자의 시클릭 또는 폴리시클릭계, 포스포디에스테르, 펩티드, 올리고사카라이드, 올리고뉴클레오타이드, 올리고머, 중합체 또는 폴리 알킬 글리콜 (예를 들어, 폴리 에틸렌 글리콜, 예컨대  $-(CH_2CH_2O)_nCH_2CH_2-$ , 여기서 n은 1 내지 50의 정수임) 또는 그의 조합을 포함할 수 있다.
- [0155] 이관능성 스페이서는 라이브러리의 헤드피스와 화학적 엔티티 사이에 적절한 스페이싱 모이어티를 제공할 수 있다. 특정 실시양태에서, 이관능성 스페이서는 3개의 부분을 포함한다. 부분 1은 DNA와 공유 결합을 형성하는 반응성 기, 예컨대 예를 들어, DNA 상의 아미노 기 (예를 들어, 아미노-변형된 dT)와 반응하도록 N-히드록시 수신이미드 (NHS) 에스테르에 의해 바람직하게 활성화된 카르복실산, (표준 올리고뉴클레오타이드 화학에 의해 달성되는) 단일-가닥 헤드피스의 5' 또는 3'-말단을 변형시키는 아미다이트, 화학적-반응성 쌍 (예를 들어, Cu(I) 촉매의 존재 하의 아지도-알킨 고리화첨가, 또는 본원에 기재된 임의의 것), 또는 티올 반응성 기일 수 있다.

부분 2는 또한 화학적 엔티티, 빌딩 블록  $A_n$  또는 스캐폴드와 공유 결합을 형성하는 반응성 기일 수 있다. 이러한 반응성 기는 예를 들어 아민, 티올, 아지드 또는 알킨일 수 있다. 부분 3은 부분 1과 2 사이에 도입된, 가변 길이의 화학적 불활성 스페이싱 모이어티일 수 있다. 이러한 스페이싱 모이어티는 에틸렌 글리콜 단위의 쇠 (예를 들어, 상이한 길이의 PEG), 알칸, 알켄, 폴리엔 쇠 또는 펩티드 쇠일 수 있다. 스페이서는 유기 용매 중 헤드피스의 용해도를 개선시키기 위한 소수성 모이어티 (예컨대 예를 들어, 벤젠 고리), 뿐만 아니라 라이브러리 검출 목적으로 사용되는 형광 모이어티 (예를 들어 플루오레세인 또는 Cy-3)를 갖는 분지 또는 삽입물을 함유할 수 있다. 헤드피스 설계 내의 소수성 잔기는 유기 용매 중 라이브러리 합성을 용이하게 하기 위해 스페이서 설계와 함께 달라질 수 있다. 예를 들어, 헤드피스 및 스페이서 조합은 옥탄올:물 계수 ( $P_{oct}$ )가 예를 들어 1.0 내지 2.5인 적절한 잔기를 갖도록 설계된다.

[0156] 스페이서는 주어진 소분자 라이브러리 설계에 대해 실험적으로 선택될 수 있고, 이로써 라이브러리는 유기 용매, 예를 들어 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% 또는 100% 유기 용매 중에서 합성될 수 있다. 스페이서는 유기 용매 중에 헤드피스를 가용화시키는 적절한 쇠 길이를 선택하기 위해 라이브러리 합성 전에 모델 반응을 사용하여 다르게 할 수 있다. 예시적인 스페이서는 증가된 알킬 쇠 길이, 증가된 폴리 에틸렌 글리콜 단위, (헤드피스 상의 포스페이트 음전하를 중화시키는) 양전하를 갖는 분지형 중 또는 증가된 소수성 양 (예를 들어, 벤젠 고리 구조의 추가)을 갖는 것을 포함한다.

[0157] 상업적으로 입수가 가능한 스페이서의 예는 아미노-카르복실산 스페이서, 예컨대 펩티드인 것 (예를 들어, Z-Gly-Gly-Gly-Osu (N-알파-벤질옥시카르보닐-(글리신)<sub>3</sub>-N-숙신이미드 에스테르) 또는 Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (N-알파-벤질옥시카르보닐-(글리신)<sub>6</sub>-N-숙신이미드 에스테르, 서열식별번호: 1)), PEG (예를 들어, Fmoc-아미노 PEG2000-NHS 또는 아미노-PEG (12-24)-NHS), 또는 알칸 산 쇠 (예를 들어, Boc-ε-아미노카프로산-Osu); 화학적-반응성 쌍 스페이서, 예컨대 펩티드 모이어티 (예를 들어, 아지도호모알라닌-Gly-Gly-Gly-Osu (서열식별번호: 2) 또는 프로파르길글리신-Gly-Gly-Gly-Osu (서열식별번호: 3)), PEG (예를 들어, 아지도-PEG-NHS), 또는 알칸 산 쇠 모이어티 (예를 들어, 5-아지도펜탄산, (S)-2-(아지도메틸)-1-Boc-피롤리딘, 4-아지도아닐린 또는 4-아지도-부탄-1-산 N-히드록시숙신이미드 에스테르)와 함께 조합된, 본원에 기재된 화학적-반응성 쌍; 티올-반응성 스페이서, 예컨대 PEG인 것 (예를 들어, SM(PEG)<sub>n</sub> NHS-PEG-말레이미드), 알칸 쇠 (예를 들어, 3-(피리딘-2-일디술팜닐)-프로피온산-Osu 또는 술포숙신이미드 6-(3'-[2-피리딜디티오]-프로피온아미도)헥사노에이트)); 및 올리고뉴클레오타이드 합성을 위한 아미다이트, 예컨대 아미노 개질제 (예를 들어, 6-(트리플루오로아세틸아미노)-헥실-(2-시아노에틸)-(N,N-디이소프로필)-포스포르아미다이트), 티올 개질제 (예를 들어, S-트리틸-6-메르캅토헥실-1-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트 또는 화학적-반응성 쌍 개질제 (예를 들어, 6-헥실-1-일-(2-시아노에틸)-(N,N-디이소프로필)-포스포르아미다이트, 3-디메톡시트리틸옥시-2-(3-(3-프로파르길옥시프로판아미도)프로판아미도)프로필-1-O-숙시노일, 장쇄 알킬아미노 CPG 또는 4-아지도-부탄-1-산 N-히드록시숙신이미드 에스테르))를 포함한다. 추가의 스페이서가 관련 기술분야에 공지되어 있고, 라이브러리 합성 동안 사용될 수 있는 것은 5'-O-디메톡시트리틸-1',2'-디데옥시리보스-3'-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 9-O-디메톡시트리틸-트리에틸렌 글리콜, 1-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 3-(4,4'-디메톡시트리틸옥시)프로필-1-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트; 및 18-O-디메톡시트리틸 헥사에틸렌글리콜, 1-[(2-시아노에틸)-(N,N-디이소프로필)]-포스포르아미다이트를 포함하나, 이에 제한되지는 않는다. 본원의 임의의 스페이서는 탠덤으로 서로 상이한 조합으로 추가되어 상이한 목적하는 길이의 스페이서를 생성할 수 있다.

[0158] 스페이서는 또한 분지형일 수 있으며, 여기서 분지형 스페이서는 관련 기술분야에 널리 공지되어 있고, 예는 대칭 또는 비대칭 더블러 또는 대칭 트레블러로 이루어질 수 있다. 예를 들어, 문헌 [Newcome et al., Dendritic Molecules: Concepts, Synthesis, Perspectives, VCH Publishers (1996); Boussif et al., Proc. Natl. Acad. Sci. USA 92:7297-7301 (1995); 및 Jansen et al., Science 266:1226 (1994)]을 참조한다.

[0159] 효소적 라이게이션 및 화학적 라이게이션 기술

[0160] 태그를 헤드피스에 추가하여 복합체를 생성하는데 다양한 라이게이션 기술이 사용될 수 있다. 따라서, 본원에 기재된 임의의 결합 단계는 임의의 유용한 라이게이션 기술, 예컨대 효소적 라이게이션 및/또는 화학적 라이게이션을 포함할 수 있다. 이들 결합 단계는 1종 이상의 태그의 헤드피스 또는 복합체로의 추가를 포함할 수 있다. 특정한 실시양태에서, 임의의 올리고뉴클레오타이드에 대해 사용되는 라이게이션 기술은 전사 및/또는 역전사되어 라이브러리의 디코딩을 가능하게 하거나 또는 1종 이상의 DNA 또는 RNA 폴리머라제에 의한 주형-의존성



중합을 가능하게 할 수 있는 생성된 생성물을 제공한다.

[0161] 일반적으로, 효소적 라이게이션은 전사 및/또는 역전사될 수 있는 천연 포스포디에스테르 결합을 갖는 올리고뉴클레오타이드를 생성한다. 효소 라이게이션의 예시적인 방법이 본원에 제공되고, 1종 이상의 RNA 또는 DNA 리가제, 예컨대 T4 RNA 리가제 1 또는 2, T4 DNA 리가제, 서크리가제(CircLigase)<sup>TM</sup> ssDNA 리가제, 서크리가제<sup>TM</sup> II ssDNA 리가제 및 써모파지(ThermoPhage)<sup>TM</sup> ssDNA 리가제 (프로카자임 리미티드(Prokaryote Ltd.), 아이슬랜드 레이카비크)의 사용을 포함한다.

[0162] 화학적 라이게이션이 또한 전사 또는 역전사될 수 있는 올리고뉴클레오타이드를 생성하는데 사용될 수 있거나, 또는 달리 주형-의존성 폴리머라제를 위한 주형으로서 사용될 수 있다. 전사 또는 역전사될 수 있는 올리고뉴클레오타이드를 제공하는 화학적 라이게이션 기술의 효능은 시험될 필요가 있을 수 있다. 이러한 효능은 임의의 유용한 방법, 예컨대 액체 크로마토그래피-질량 분광측정법, RT-PCR 분석, PCR 분석, 전기영동, 및/또는 서열분석에 의해 시험될 수 있다. 특정한 실시양태에서, 화학적 라이게이션은 전사 또는 역전사될 수 있는 스페이싱 모이어티를 제공하기 위해 1종 이상의 화학적-반응성 쌍의 사용을 포함한다. 본 발명의 방법의 예가 도 1에 약술되어 있고, 여기서 이중-가닥 헤어핀 구조는 공유 부착된 코딩 소분자의 합성을 위해 코딩 올리고뉴클레오타이드 태그 및 보호된 1급 아민의 둘 다의 화학적 라이게이션을 위한 부위를 제공하는 이관능성 헤드피스 올리고뉴클레오타이드로서 사용된다. 헤드피스는 3'- 및 5'-포스페이트 기 둘 다를 보유하고, 그의 각각은 시아노이미다졸 및 2가 금속 이온 예컨대  $Zn^{2+}$ 를 사용하여 상응하는 상보적 비인산화 올리고뉴클레오타이드에 라이게이션될 수 있다. 동일한 구조물은 T4 DNA 리가제를 사용한 효소적 라이게이션에 의하면 단지 반-라이게이션될 수 있고, 이는 이러한 효소가 도 1에 제시된 바와 같이 단지 5'-포스페이트의 3'-히드록실 올리고뉴클레오타이드에의 라이게이션만 지지하고, 3'-포스페이트의 5'-히드록실 올리고뉴클레오타이드에의 라이게이션은 지지하지 않기 때문이다. 비보호된 1급 아민은 시아노이미다졸과 반응하여 구아닌딘 부가물을 생성하는 것이 관찰되었지만, 아민의 Fmoc 보호는 이의 발생을 방지할 수 있고, 보호된 아민은 화학적 라이게이션 반응 조건 하에 탈보호되지 않는다. Fmoc는 피페리딘에 의해 용이하게 제거된다.

[0163] 효소적 라이게이션 또는 화학적 라이게이션을 촉진시키기 위한 반응 조건

[0164] 본원에 기재된 방법은 헤드피스와 태그 사이 또는 2개의 태그 사이의 효소적 또는 화학적 라이게이션을 촉진시키는 1종 이상의 반응 조건을 포함할 수 있다. 이들 반응 조건은 본원에 기재된 바와 같은, 태그내 변형된 뉴클레오타이드를 사용하는 것; 상이한 길이를 갖는 공여자 태그 및 수용자 태그를 사용하고, 태그의 농도를 달리하는 것; 상이한 유형의 리가제, 뿐만 아니라 그의 조합 (예를 들어, 서크리가제<sup>TM</sup> DNA 리가제 및/또는 T4 RNA 리가제)을 사용하고, 그의 농도를 달리하는 것; 상이한 분자량을 갖는 폴리 에틸렌 글리콜 (PEG)을 사용하고, 그의 농도를 달리하는 것; 비-PEG 집중화제 (예를 들어, 베타인 또는 소 혈청 알부민)의 사용; 라이게이션 온도 및 지속시간을 달리하는 것; ATP,  $Co(NH_3)_6Cl_3$  및 효모 무기 피로포스페이트를 포함한, 다양한 작용제의 농도를 달리하는 것; 효소적으로 또는 화학적으로 인산화된 올리고뉴클레오타이드 태그를 사용하는 것; 3'-보호된 태그를 사용하는 것; 및 프리아데닐화 태그를 사용하는 것을 포함한다. 이들 반응 조건은 화학적 라이게이션을 또한 포함한다.

[0165] 헤드피스 및/또는 태그는 1종 이상의 변형 또는 치환된 뉴클레오타이드를 포함할 수 있다. 바람직한 실시양태에서, 헤드피스 및/또는 태그는 효소적 라이게이션을 촉진시키는 1종 이상의 변형 또는 치환된 뉴클레오타이드, 예컨대 2'-O-메틸 뉴클레오타이드 (예를 들어, 2'-O-메틸 구아닌 또는 2'-O-메틸 우라실), 2'-플루오로 뉴클레오타이드, 또는 라이게이션에 대해 기질로서 사용되는 임의의 다른 변형된 뉴클레오타이드를 포함한다. 대안적으로, 헤드피스 및/또는 태그는 화학적 라이게이션을 지지하는 1종 이상의 화학적 반응성 기 (예를 들어, 임의로 치환된 알킬닐 기 및 임의로 치환된 아지도 기)를 포함하도록 변형된다. 임의로, 태그 올리고뉴클레오타이드는 둘 다의 말단에서 화학적 반응성 기로 관능화되고, 임의로 이들 말단 중 하나는 보호되어, 기는 독립적으로 다루어질 수 있고 부반응은 감소될 수 있다 (예를 들어, 중합 부반응 감소).

[0166] 본원에 기재된 바와 같이, 포스포디에스테르, 포스포네이트 또는 포스포로티오에이트 연결을 생성하는 화학적 라이게이션은 시아노이미다졸 및 2가 금속 이온 예컨대  $Zn^{2+}$ 의 존재 하에서의 5'- 또는 3'-포스페이트, 포스포네이트 또는 포스포로티오에이트와 5'- 또는 3'-히드록실 기의 반응에 의해 수행될 수 있다.

[0167] 효소적 라이게이션은 1종 이상의 리가제를 포함할 수 있다. 예시적인 리가제는 서크리가제<sup>TM</sup> ssDNA 리가제 (에피센터 바이오테크놀로지스(EPICENTRE Biotechnologies) 위스콘신주 매디슨), 서크리가제<sup>TM</sup> II ssDNA 리가제 (또한 에피센터 바이오테크놀로지스로부터의 것), 써모파지<sup>TM</sup> ssDNA 리가제 (프로카자임 리미티드, 아이슬란드

레이카비크), T4 RNA 리가제 및 T4 DNA 리가제를 포함한다. 바람직한 실시양태에서, 라이게이션은 RNA 리가제, 또는 RNA 리가제 및 DNA 리가제의 조합을 사용하는 것을 포함한다. 라이게이션은 1종 이상의 리가제와 함께 조합하여, 1종 이상의 가용성 다가 양이온, 예컨대  $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ 를 추가로 포함할 수 있다.

[0168] 라이게이션 단계 전 또는 후에, 복합체 또는 코딩된 화학적 엔티티는 정제될 수 있다. 일부 실시양태에서, 복합체 또는 코딩된 화학적 엔티티는 교차-반응을 발생시킬 수 있고 코딩 프로세스에 "노이즈"를 도입할 수 있는 미반응 헤드피스 또는 태그를 제거하기 위해 정제될 수 있다. 일부 실시양태에서, 복합체 또는 코딩된 화학적 엔티티는 리가제의 라이게이션 활성을 억제 또는 저하시킬 수 있는 임의의 시약 또는 미반응 출발 물질을 제거하기 위해 정제될 수 있다. 예를 들어, 오르토포스페이트는 저하된 라이게이션 활성을 발생시킬 수 있다. 특정 실시양태에서, 화학적 또는 라이게이션 단계에 도입되는 엔티티는 후속 화학적 또는 라이게이션 단계가 가능하도록 제거될 필요가 있을 수 있다. 복합체 또는 코딩된 화학적 엔티티를 정제하는 방법은 본원에 기재되어 있다. 복합체의 정제는 복합체의 가역적 고정화에 이어 후속 단계 전 정제 및 방출에 의해 수행될 수 있다.

[0169] 효소적 및 화학적 라이게이션은 300 달톤 초과 (예를 들어, 600 달톤, 3,000 달톤, 4,000 달톤, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 15,000, 20,000, 25,000, 30,000, 35,000, 40,000, 또는 45,000 달톤 초과)의 평균 분자량을 갖는 폴리 에틸렌 글리콜을 포함할 수 있다. 특정한 실시양태에서, 폴리 에틸렌 글리콜은 약 3,000 달톤 내지 9,000 달톤 (예를 들어, 3,000 달톤 내지 8,000 달톤, 3,000 달톤 내지 7,000 달톤, 3,000 달톤 내지 6,000 달톤 및 3,000 달톤 내지 5,000 달톤)의 평균 분자량을 갖는다. 바람직한 실시양태에서, 폴리 에틸렌 글리콜은 약 3,000 달톤 내지 약 6,000 달톤 (예를 들어, 3,300 달톤 내지 4,500 달톤, 3,300 달톤 내지 5,000 달톤, 3,300 달톤 내지 5,500 달톤, 3,300 달톤 내지 6,000 달톤, 3,500 달톤 내지 4,500 달톤, 3,500 달톤 내지 5,000 달톤, 3,500 달톤 내지 5,500 달톤 및 3,500 달톤 내지 6,000 달톤, 예컨대 4,600 달톤)의 평균 분자량을 갖는다. 폴리 에틸렌 글리콜은 임의의 유용한 양으로, 예컨대 약 25% (w/v) 내지 약 35% (w/v), 예컨대 30% (w/v)로 존재할 수 있다.

[0170] 복합체의 뉴클레오타이드 서열을 결정하는 방법

[0171] 본 발명은 복합체의 뉴클레오타이드 서열을 결정하는 방법을 특색으로 하며, 이로써 어셈블리된 태그 서열의 서열과 화학적 엔티티의 구조 단위 (또는 빌딩 블록) 사이의 코딩 관계가 확립될 수 있다. 특히, 화학적 엔티티의 아이덴티티 및/또는 히스토리는 올리고뉴클레오타이드의 염기 서열로부터 추론될 수 있다. 이러한 방법을 사용하여 다양한 화학적 엔티티 또는 구성원 (예를 들어, 소분자 또는 펩티드)을 포함하는 라이브러리를 특정한 태그 서열로 다룰 수 있다.

[0172] 본원에 기재된 임의의 연결은 가역적 또는 비가역적일 수 있다. 가역적 연결은 광-반응성 연결 (예를 들어, 시아노비닐카르바졸 기 및 티미딘) 및 산화환원 연결을 포함한다. 추가의 연결이 본원에 기재되어 있다.

[0173] 대안적 실시양태에서, "판독불가능한" 연결은 판독가능한 또는 적어도 전위가능한 연결을 생성하기 위해 효소적으로 복구될 수 있다. 효소적 복구 프로세스는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있고, 피리미딘 (예를 들어, 티미딘) 이량체 복구 메카니즘 (예를 들어, 포토리아제 또는 글리코실라제 (예를 들어, T4 피리미딘 이량체 글리코실라제 (PDG))를 사용하는 것), 염기 절제 복구 메카니즘 (예를 들어, 글리코실라제, 아퓨린/아피리미딘 (AP) 엔도뉴클레아제, Flap 엔도뉴클레아제 또는 폴리 ADP 리보스 폴리머라제 (예를 들어, 인간 아퓨린/아피리미딘 (AP) 엔도뉴클레아제, APE 1; 엔도뉴클레아제 III (Nth) 단백질; 엔도뉴클레아제 IV; 엔도뉴클레아제 V; 포름아미도피리미딘 [fapy]-DNA 글리코실라제 (Fpg); 인간 8-옥소구아닌 글리코실라제 1 ( $\alpha$  이소형) (hOGG1); 인간 엔도뉴클레아제 VIII-유사 1 (hNEIL1); 우라실-DNA 글리코실라제 (UDG); 인간 단일-가닥 선택적 일관능성 우라실 DNA 글리코실라제 (SMUG1); 및 인간 알킬아데닌 DNA 글리코실라제 (hAAG))를 사용하는 것, 이는 복구를 위해 1종 이상의 엔도뉴클레아제, DNA 또는 RNA 폴리머라제 및/또는 리가제와 임의로 조합될 수 있음), 메틸화 복구 메카니즘 (예를 들어, 메틸 구아닌 메틸트랜스퍼라제를 사용하는 것), AP 복구 메카니즘 (예를 들어, 아퓨린/아피리미딘 (AP) 엔도뉴클레아제 (예를 들어, APE 1; 엔도뉴클레아제 III; 엔도뉴클레아제 IV; 엔도뉴클레아제 V; Fpg; hOGG1; 및 hNEIL1)를 사용하는 것, 이는 복구를 위해 1종 이상의 엔도뉴클레아제, DNA 또는 RNA 폴리머라제 및/또는 리가제와 임의로 조합될 수 있음), 뉴클레오타이드 절제 복구 메카니즘 (예를 들어, 절제 복구 교차-상보적 단백질 또는 절제 뉴클레아제를 사용하는 것, 이는 복구를 위해 1종 이상의 엔도뉴클레아제, DNA 또는 RNA 폴리머라제 및/또는 리가제와 임의로 조합될 수 있음), 및 미스매치 복구 메카니즘 (예를 들어, 엔도뉴클레아제 (예를 들어, T7 엔도뉴클레아제 I; MutS, MutH, 및/또는 MutL)를 사용하는 것, 이는 복구를 위해 1종 이상의 엑소뉴클레아제, 엔도뉴클레아제, 헬리카제, DNA 또는 RNA 폴리머라제 및/또는 리가제와 임의로 조합될 수 있음)을 포함하나 이에 제한되지는 않는다. 이들 복구 메카니즘 유형을 용이하게 제공

하기 위해 상업적 효소 혼합물, 예를 들어 Taq DNA 리가제, 엔도뉴클레아제 IV, Bst DNA 폴리머라제, Fpg, 우라실-DNA 글리코실라제 (UDG), T4 PDG (T4 엔도뉴클레아제 V) 및 엔도뉴클레아제 VIII을 포함하는 PreCR® 복구 믹스 (뉴잉글랜드 바이오랩스 인크.(New England Biolabs Inc.), 매사추세츠주 입스위치)가 이용가능하다.

[0174] 코딩된 라이브러리를 태그부착하는 방법

[0175] 본 발명은 올리고뉴클레오타이드 태그를 화학적 엔티티와 작동가능하게 회합시키는 방법을 특색으로 하며, 이로써 태그 서열과 화학적 엔티티의 구조 단위 (또는 빌딩 블록) 사이의 코딩 관계가 확립될 수 있다. 특히, 화학적 엔티티의 아이덴티티 및/또는 히스토리는 올리고뉴클레오타이드의 염기 서열로부터 추론될 수 있다. 이러한 방법을 사용하여 다양한 화학적 엔티티 또는 구성원 (예를 들어, 소분자 또는 펩티드)을 포함하는 라이브러리를 특정한 태그 서열로 코딩할 수 있다.

[0176] 일반적으로, 이들 방법은 화학적으로 구현될 수 있는 적어도 1개의 관능기 및 올리고뉴클레오타이드 태그가 결합 (또는 라이게이션)될 수 있는 적어도 1개의 관능기를 갖는 헤드피스의 사용을 포함한다. 결합은 임의의 유용한 수단, 예컨대 효소적 결합 (예를 들어, RNA 리가제 및/또는 DNA 리가제 중 1종 이상에 의한 라이게이션)에 의해 또는 화학적 결합에 의해 (예를 들어, 2개의 관능기, 예컨대 친핵체 및 이탈기 사이의 치환 반응에 의해) 이루어질 수 있다.

[0177] 라이브러리 내에 다수의 화학적 엔티티를 생성하기 위해, 헤드피스를 함유하는 용액을 다중 분취물로 나누고, 이를 다수의 물리적으로 분리된 구획, 예컨대 멀티웰 플레이트의 웰에 배치할 수 있다. 일반적으로, 이는 "분할" 단계이다. 각각의 구획 또는 웰 내에서, 각각의 분취물 내의 올리고뉴클레오타이드 태그를 사용하여 연속적인 화학 반응 및 라이게이션 단계를 수행한다. 화학 반응 조건과 태그의 서열 사이의 관계가 연관되었다. 반응 및 라이게이션 단계는 임의의 순서로 수행될 수 있다. 이어서, 반응되고 라이게이션된 분취물을 조합하거나 또는 "폴딩하고", 이 지점에서 임의로 정제를 수행할 수 있다. 정제는 복합체의 가역적 고정화, 용매 및 임의의 시약/오염물의 제거, 이어서 후속 단계 전 복합체의 방출에 의해 수행될 수 있다. 이들 분할 및 폴딩 단계를 임의로 반복할 수 있다.

[0178] 다음으로, 본원에 기재된 바와 같이, 특정한 특징 또는 기능에 대해 라이브러리를 시험하고/거나 선택할 수 있다. 예를 들어, 태그부착된 화학적 엔티티의 혼합물을, 제1 집단은 특정한 생물학적 표적에 결합하는 구성원에 대해 풍부화되고 제2 집단은 덜 풍부화된 것인, 적어도 2개의 집단으로 분리할 수 있다 (예를 들어, 음성 선택 또는 양성 선택에 의한). 이어서, 제1 집단을 (예를 들어, 칼럼 상에서 용리시켜 관심 표적을 제공하거나 또는 분취물을 관심 표적과 함께 인큐베이션시키는 것에 의해) 선택적으로 포획할 수 있고, 임의로, 예컨대 임의적인 세척, 정제, 음성 선택, 양성 선택 또는 분리 단계에 의해 추가로 분석하거나 시험할 수 있다.

[0179] 마지막으로, 선택된 집단 내의 1종 이상의 구성원 (또는 화학적 엔티티)의 화학적 히스토리는 작동가능하게 연결된 올리고뉴클레오타이드의 서열에 의해 결정될 수 있다. 서열과 코딩된 라이브러리 구성원 화학적 히스토리를 상호관련시켜볼 때, 이러한 방법은 선택된 특징 (예를 들어, 표적 단백질에 결합하고 이에 의해 치료 효과를 도출하는 증가된 경향)을 갖는 라이브러리의 개별 구성원을 확인할 수 있다. 이어서 추가의 시험 및 최적화를 위해, 그의 회합된 올리고뉴클레오타이드 태그의 존재 또는 부재 하의 확인된 라이브러리 구성원을 합성함으로써 후보 치료 화합물을 제조할 수 있다.

[0180] 본원에 기재된 방법은 라이브러리를 다양화하거나 또는 라이브러리의 구성원을 조사하기 위해 임의의 개수의 임의적인 단계를 포함할 수 있다. 본원에 기재된 임의의 태그부착 방법을 위해, 연속적인 "n"개의 태그는 추가의 "n"개의 라이게이션, 분리 및/또는 인산화 단계에 의해 추가될 수 있다. 예시적인 임의적인 단계는 1종 이상의 제한 엔도뉴클레아제를 사용한 라이브러리 구성원-회합된 코딩 올리고뉴클레오타이드의 제한; 예를 들어, 임의의 복구 효소, 예컨대 본원에 기재된 것에 의한 회합된 코딩 올리고뉴클레오타이드의 복구; 예를 들어 증폭 및 서열 분석을 위한 프라이밍 서열을 제공하거나 또는 서열의 고정화를 위한 표지, 예컨대 비오틴을 제공하는 1종 이상의 어댑터 서열과 같은 1종 이상의 어댑터 서열의, 라이브러리 구성원-회합된 코딩 올리고뉴클레오타이드의 말단 중 하나 또는 둘 다에의 라이게이션; 역전사효소, 전사효소 또는 또 다른 주형-의존성 폴리머라제를 사용한 복합체 내의 어셈블리된 태그의 역전사 또는 전사에 이어 임의로 역전사; 예를 들어 PCR을 사용한 복합체 내의 어셈블리된 태그의 증폭; 예를 들어 박테리아 형질전환의 사용, 에멀전 형성, 희석, 표면 포획 기술 등에 의한 복합체 내의 어셈블리된 태그의 1종 이상의 집단의 클론 단리물 생성; 예를 들어 뉴클레오타이드의 주형-의존성 중합을 위한 주형으로서 클론 단리물을 사용함으로써 복합체 내의 어셈블리된 태그의 1종 이상의 집단의 클론 단리물의 증폭; 및 예를 들어 가역적 종결인자 화학에 의해 형광 표지된 뉴클레오타이드와 함께 주형-의존성 중합을 위한 주형으로서 클론 단리물을 사용하는 것에 의한 복합체 내의 어셈블리된 태그의 1종 이상의 집단의 클론 단



리물의 서열 결정을 포함한다. 올리고뉴클레오타이드 태그를 증폭시키고 서열분석하는 추가의 방법은 본원에 기재되어 있다.

[0181] 이들 방법을 사용하여, 예를 들어 선택 단계에서 특정한 특징 또는 기능을 갖는 임의의 수의 화학적 엔티티를 확인 및 발견할 수 있다. 라이브러리 내의 목적하는 기능을 갖는 구성원 또는 관련 구성원 중 적어도 1종을 동시에 풍부화시키면서, 라이브러리를 적어도 2개의 부분으로 분할하기 위한 기준으로서 목적하는 특징 또는 기능이 사용될 수 있다. 특정한 실시양태에서, 방법은 관심 치료 단백질에 결합하거나 그를 불활성화시키는 소형 약물-유사 라이브러리 구성원을 확인하는 것을 포함한다. 또 다른 실시양태에서, 규정된 화학적 조건 하에서의 선택된 빌딩 블록의 반응이 1종 이상의 분자가 특정한 단백질에 대한 치료제로서의 유용성을 가질 수 있는 복수의 분자의 조합 (또는 분자의 라이브러리)을 생성하도록, 화학 반응의 순서를 설계하고 빌딩 블록 세트를 선택한다. 예를 들어, 화학 반응 및 빌딩 블록은 통상적으로 키나제 억제제에 존재하는 구조적 기를 갖는 라이브러리를 생성하도록 선택된다. 임의의 이들 경우에서, 올리고뉴클레오타이드 태그는 라이브러리 구성원의 화학적 히스토리를 코딩하고, 각각의 경우에 화학적 가능성의 집합은 임의의 특정한 태그 조합에 의해 나타내어질 수 있다.

[0182] 한 실시양태에서, 라이브러리의 적어도 1종의 구성원이 표적에 결합하는데 적합한 조건 하에서 화학적 엔티티의 라이브러리 또는 그의 부분을 생물학적 표적과 접촉시킨 후, 표적에 결합하지 않은 라이브러리 구성원을 제거하고, 표적과 회합된 1종 이상의 올리고뉴클레오타이드 태그를 분석한다. 이러한 방법은 관련 기술분야에 공지된 방법에 의해 태그를 증폭시키는 것을 임의로 포함할 수 있다. 예시적인 생물학적 표적은 효소 (예를 들어, 키나제, 포스파타제, 메틸라제, 데메틸라제, 프로테아제 및 DNA 복구 효소), 단백질:단백질 상호작용에 수반되는 단백질 (예를 들어, 수용체에 대한 리간드), 수용체 표적 (예를 들어, GPCR 및 RTK), 이온 채널, 박테리아, 바이러스, 기생충, DNA, RNA, 프리온 및 탄수화물을 포함한다.

[0183] 또 다른 실시양태에서, 표적에 결합한 화학적 엔티티를 증폭에 적용시키지 않지만 직접 분석한다. 예시적인 분석 방법은 소실 공명 광자 결정 분석을 포함한 마이크로어레이 분석; (예를 들어, his-태그를 사용하는 것에 의한) 태그를 디킨블루션하기 위한 비드-기반 방법; 무-표지 광자 결정 바이오센서 분석 (예를 들어, 매사추세츠 주 워번 소재의 SRU 바이오시스템즈 인크.(SRU Biosystems, Inc.)로부터의 바인드(BIND)® 판독기); 또는 (예를 들어 태그의 라이브러리에 존재하는 서열에 상보적인 고정된 올리고뉴클레오타이드의 어레이를 사용하는 것에 의한) 혼성화-기반 접근법을 포함한다.

[0184] 또한, 화학적-반응성 쌍 (또는 관능기)은 고체상 올리고뉴클레오타이드 합성 스킴에 용이하게 포함될 수 있고, 올리고뉴클레오타이드의 효율적인 화학적 라이게이션을 지지할 것이다. 또한, 생성된 라이게이션된 올리고뉴클레오타이드는 1종 이상의 폴리머라제를 사용한 주형-의존성 중합을 위한 주형으로서 작용할 수 있다. 따라서, 코딩된 라이브러리를 태그부착하기 위한 본원에 기재된 임의의 결합 단계는 효소적 라이게이션 및/또는 화학적 라이게이션 기술 중 하나 이상을 포함하도록 변형될 수 있다. 예시적인 라이게이션 기술은 효소 라이게이션, 예컨대 1종 이상의 RNA 리가제 및/또는 DNA 리가제의 사용; 및 화학적 라이게이션, 예컨대 화학적-반응성 쌍 (예를 들어, 임의로 치환된 알킬 및 아지도 관능기를 포함하는 쌍)의 사용을 포함한다.

[0185] 또한, 1종 이상의 라이브러리는 분할-및-혼합 단계에서 조합될 수 있다. 2종 이상의 라이브러리의 혼합을 가능하게 하기 위해, 라이브러리 구성원은 본원에 기재된 바와 같이, 예컨대 라이브러리-확인 태그 내에, 라이게이션된 태그 내에 또는 헤드피스 서열의 일부로서 1종 이상의 라이브러리-확인 서열을 함유할 수 있다.

[0186] 라이브러리 내에 화학적 엔티티를 코딩하는 방법

[0187] 본 발명의 방법은 올리고뉴클레오타이드 태그에 의해 코딩되는 다양한 개수의 화학적 엔티티를 갖는 라이브러리를 합성하는데 사용될 수 있다. 빌딩 블록 및 코딩 DNA 태그의 예는 미국 특허 출원 공개 번호 2007/0224607에서 발견되며, 그의 빌딩 블록 및 태그는 본원에 참조로 포함된다.

[0188] 각각의 화학적 엔티티는 1종 이상의 빌딩 블록 및 임의로 스캐폴드로부터 형성된다. 스캐폴드는 특정한 기하구조로 1종 이상의 다양성 노드를 제공하는 역할을 한다 (예를 들어, 트리아진은 헤테로아릴 고리 또는 선형 기하구조 주위에 공간적으로 배열된 3개의 노드를 제공함).

[0189] 빌딩 블록 및 그의 코딩 태그는 직접적으로 또는 (예를 들어, 스페이서를 통해) 간접적으로 헤드피스에 추가되어 복합체를 형성할 수 있다. 헤드피스가 스페이서를 포함하는 경우에, 빌딩 블록 또는 스캐폴드는 스페이서의 말단에 추가된다. 스페이서가 부재하는 경우에, 빌딩 블록은 헤드피스에 직접적으로 추가될 수 있거나, 또는 빌딩 블록 그 자체가 헤드피스의 관능기와 반응하는 스페이서를 포함할 수 있다. 예시적인 스페이서 및 헤드피

스가 본원에 기재되어 있다.

[0190] 스캐폴드는 임의의 유용한 방식으로 추가될 수 있다. 예를 들어, 스캐폴드는 스페이서 또는 헤드피스의 말단에 추가될 수 있고, 연속적인 빌딩 블록은 스캐폴드의 이용가능한 다양성 노드에 추가될 수 있다. 또 다른 예에서, 빌딩 블록  $A_n$  은 먼저 스페이서 또는 헤드피스에 추가된 다음, 스캐폴드 S의 다양성 노드는 빌딩 블록  $A_n$  내의 관능기와 반응한다. 특정한 스캐폴드를 코딩하는 올리고뉴클레오타이드 태그는 헤드피스 또는 복합체에 임의로 추가될 수 있다. 예를 들어,  $S_n$ 은  $n$ 개의 반응 용기 내의 복합체에 추가되고 (여기서  $n$ 은 1 초과와 정수임), 태그  $S_n$  (즉, 태그  $S_1, S_2, \dots, S_{n-1}, S_n$ )은 복합체의 관능기에 결합된다.

[0191] 빌딩 블록은 다중 합성 단계에 첨가될 수 있다. 예를 들어, 부착된 스페이서를 임의로 갖는 헤드피스의 분취물을  $n$ 개의 반응 용기로 분리한다 (여기서  $n$ 은 2 이상의 정수임). 제1 단계에서, 빌딩 블록  $A_n$ 을 각각의  $n$ 개의 반응 용기에 첨가하고 (즉, 빌딩 블록  $A_1, A_2, \dots, A_{n-1}, A_n$ 을 반응 용기 1, 2,  $\dots$ ,  $n-1, n$ 에 첨가함), 여기서  $n$ 은 정수이며 각각의 빌딩 블록  $A_n$ 은 고유하다. 제2 단계에서, 스캐폴드 S를 각각의 반응 용기에 첨가하여  $A_n$ -S 복합체를 형성한다. 임의로, 스캐폴드  $S_n$ 을 각각의 반응 용기에 첨가하여  $A_n$ - $S_n$  복합체를 형성할 수 있으며, 여기서  $n$ 은 2 초과와 정수이고, 각각의 스캐폴드  $S_n$ 은 고유할 수 있다. 제3 단계에서, 빌딩 블록  $B_n$ 을  $A_n$ -S 복합체를 함유하는 각각의  $n$ 개의 반응 용기에 첨가하고 (즉, 빌딩 블록  $B_1, B_2, \dots, B_{n-1}, B_n$ 을  $A_1$ -S,  $A_2$ -S,  $\dots$ ,  $A_{n-1}$ -S,  $A_n$ -S 복합체를 함유하는 반응 용기 1, 2,  $\dots$ ,  $n-1, n$ 에 첨가함), 여기서 각각의 빌딩 블록  $B_n$ 은 고유하다. 추가 단계에서, 빌딩 블록  $C_n$ 을  $B_n$ - $A_n$ -S 복합체를 함유하는 각각의  $n$ 개의 반응 용기에 첨가할 수 있으며 (즉, 빌딩 블록  $C_1, C_2, \dots, C_{n-1}, C_n$ 을  $B_1$ - $A_1$ -S,  $\dots$ ,  $B_n$ - $A_n$ -S 복합체를 함유하는 반응 용기 1, 2,  $\dots$ ,  $n-1, n$ 에 첨가함), 여기서 각각의 빌딩 블록  $C_n$ 은 고유하다. 생성된 라이브러리는  $n^3$ 개의 태그를 갖는  $n^3$ 개의 복합체를 가질 것이다. 이러한 방식으로, 추가의 합성 단계를 사용하여 추가의 빌딩 블록에 결합시킴으로써 라이브러리를 추가로 다양화할 수 있다.

[0192] 라이브러리를 형성한 후, 생성된 복합체를 임의로 정제하고, 중합 또는 예를 들어 테일피스에 대한 라이게이션 반응에 적용시킬 수 있다. 이러한 일반적인 전략은 추가의 다양성 노드 및 빌딩 블록 (예를 들어, D, E, F 등)을 포함하도록 확대될 수 있다. 예를 들어, 제1 다양성 노드는 빌딩 블록 및/또는 S와 반응하고, 올리고뉴클레오타이드 태그에 의해 코딩된다. 이어서, 추가의 빌딩 블록은 생성된 복합체와 반응하고, 후속 다양성 노드는 추가의 빌딩 블록에 의해 유도체화되며, 이는 중합 또는 라이게이션 반응에 사용되는 프라이머에 의해 코딩된다.

[0193] 코딩된 라이브러리를 형성하기 위해, 각각의 합성 단계 후 또는 전에 올리고뉴클레오타이드 태그를 복합체에 첨가한다. 예를 들어, 각각의 반응 용기에의 빌딩 블록  $A_n$ 의 첨가 전 또는 후에 태그  $A_n$ 을 헤드피스의 관능기에 결합시킨다 (즉, 태그  $A_1, A_2, \dots, A_{n-1}, A_n$ 을 헤드피스를 함유하는 반응 용기 1, 2,  $\dots$ ,  $n-1, n$ 에 첨가함). 각각의 태그  $A_n$ 은 각각의 고유한 빌딩 블록  $A_n$ 과 상호관련된 특유의 서열을 가지며, 태그  $A_n$ 의 서열을 결정하는 것은 빌딩 블록  $A_n$ 의 화학 구조를 제공한다. 이러한 방식으로, 추가의 태그를 사용하여 추가의 빌딩 블록 또는 추가의 스캐폴드를 코딩한다.

[0194] 추가로, 복합체에 첨가된 마지막 태그는 프라이머-결합 서열을 포함할 수 있거나, 또는 (예컨대, 라이게이션에 의한) 프라이머-결합 서열의 결합을 가능하게 하는 관능기를 제공할 수 있다. 프라이머-결합 서열은 복합체의 올리고뉴클레오타이드 태그를 증폭시키고/거나 서열분석하는데 사용될 수 있다. 예시적인 증폭 및 서열분석 방법은 폴리머라제 연쇄 반응 (PCR), 선형 쇄 증폭 (LCR), 롤링 써클 증폭 (RCA) 또는 관련 기술분야에 공지된 핵산 서열을 증폭 또는 결정하는 임의의 다른 방법을 포함한다.

[0195] 이들 방법을 사용하여, 다수의 코딩된 화학적 엔티티를 갖는 대형 라이브러리를 형성할 수 있다. 예를 들어, 헤드피스를 스페이서 및 1,000개의 상이한 변이체를 포함하는 (즉,  $n = 1,000$ ) 빌딩 블록  $A_n$ 과 반응시킨다. 각각의 빌딩 블록  $A_n$ 에 대해, DNA 태그  $A_n$ 은 헤드피스에 라이게이션되거나 또는 프라이머 연장된다. 이들 반응은 1,000-웰 플레이트 또는 10 x 100 웰 플레이트에서 수행될 수 있다. 모든 반응을 풀링하고, 임의로 정제하고, 제2 플레이트 세트에 분할할 수 있다. 다음으로, 1,000개의 상이한 변이체를 또한 포함하는 빌딩 블록  $B_n$ 을 사

용하여 동일한 절차를 수행할 수 있다. DNA 태그  $B_n$ 을  $A_n$ -헤드피스 복합체에 라이게이션시킬 수 있고, 모든 반응을 풀링할 수 있다. 생성된 라이브러리는 1,000,000개의 상이한 태그 조합에 의해 태그부착된  $1,000 \times 1,000$  개의  $A_n \times B_n$  조합 (즉, 1,000,000개의 화합물)을 포함한다. 동일한 접근법을 확장시켜 빌딩 블록  $C_n$ ,  $D_n$ ,  $E_n$  등을 추가시킬 수 있다. 이어서, 생성된 라이브러리를 사용하여 표적에 결합하는 화합물을 확인할 수 있다. DNA 태그의 PCR 및 서열분석에 의해 라이브러리에 결합하는 화학적 엔티티의 구조를 임의로 평가함으로써 풍부화된 화합물을 확인할 수 있다.

[0196] 각각의 빌딩 블록의 추가 이후 태그부착을 피하기 위해 또는 풀링 (또는 혼합)을 피하기 위해 이러한 방법은 변형될 수 있다. 예를 들어, 방법은 빌딩 블록  $A_n$ 을  $n$ 개의 반응 용기 (여기서,  $n$ 은 1 초과인 정수임)에 첨가하고 동일한 빌딩 블록  $B_1$ 을 각각의 반응 웰에 첨가하는 것에 의해 변형될 수 있다. 여기서,  $B_1$ 은 각각의 화학적 엔티티에 대해 동일하고, 따라서 이러한 빌딩 블록을 코딩하는 올리고뉴클레오타이드 태그는 필요하지 않다. 빌딩 블록을 첨가한 후, 복합체를 풀링시킬 수 있거나 또는 풀링시키지 않을 수 있다. 예를 들어, 마지막 빌딩 블록 첨가 단계 후에 라이브러리를 풀링시키지 않고, 풀을 개별적으로 스크리닝하여 표적에 결합하는 화합물(들)을 확인한다. 합성 후 모든 반응의 풀링을 피하기 위해, 예를 들어 결합 검정, 예를 들어 예를 들어 ELISA, SPR, ITC,  $T_m$  시프트, SEC 등을 사용하여 고처리량 포맷으로 (예를 들어, 384 웰 플레이트 및 1,536 웰 플레이트) 센서 표면 상에서 결합을 모니터링할 수 있다. 예를 들어, 빌딩 블록  $A_n$ 은 DNA 태그  $A_n$ 으로 코딩될 수 있고, 빌딩 블록  $B_n$ 은 웰 플레이트 내의 그 위치의 위치에 의해 코딩될 수 있다. 이어서, 후보 화합물은 결합 검정 (예를 들어, ELISA, SPR, ITC,  $T_m$  시프트, SEC 등)을 사용하고, 서열분석, 마이크로어레이 분석 및/또는 제한 소화 분석에 의해  $A_n$  태그를 분석함으로써 확인할 수 있다. 이러한 분석은 목적하는 분자를 생성하는 빌딩 블록  $A_n$  및  $B_n$ 의 조합의 확인을 가능하게 한다.

[0197] 증폭 방법은 유증수 에멀전을 형성하여 복수의 수성 마이크로반응기를 생성하는 것을 임의로 포함할 수 있다. 반응 조건 (예를 들어, 복합체의 농도 및 마이크로반응기의 크기)은 평균적으로 화합물 라이브러리의 적어도 1종의 구성원을 갖는 마이크로반응기를 제공하도록 조정될 수 있다. 각각의 마이크로반응기는 또한 표적, 복합체 또는 복합체의 부분 (예를 들어, 1종 이상의 태그)에 결합 및/또는 표적에 결합할 수 있는 단일 비드, 및 핵산 증폭을 수행하기 위한 1종 이상의 필요 시약을 갖는 증폭 반응 용액을 함유할 수 있다. 마이크로반응기에서 태그를 증폭시킨 후, 태그의 증폭된 카피를 마이크로반응기에서 비드에 결합시킬 것이고, 임의의 유용한 방법에 의해 코딩된 비드를 확인할 수 있다.

[0198] 일단 제1 라이브러리로부터 관심 표적에 결합하는 빌딩 블록을 확인하고 나면, 반복적인 방식으로 제2 라이브러리를 제조할 수 있다. 예를 들어, 본원에 기재된 바와 같이, 1 또는 2개의 추가의 다양성 노드를 추가할 수 있고, 제2 라이브러리를 생성하고 샘플링한다. 이러한 프로세스는 목적하는 분자 및 제약 특성을 갖는 분자를 생성하는데 필요한 횟수만큼 많이 반복할 수 있다.

[0199] 다양한 라이게이션 기술을 사용하여 스캐폴드, 빌딩 블록, 스페이서, 연결 및 태그를 추가할 수 있다. 따라서, 본원에 기재된 임의의 결합 단계는 임의의 유용한 라이게이션 기술 또는 기술들을 포함할 수 있다. 예시적인 라이게이션 기술은 본원에 기재된 바와 같은 효소적 라이게이션, 예컨대 1종 이상의 RNA 리가제 및/또는 DNA 리가제의 사용; 및 본원에 기재된 바와 같은 화학적 라이게이션, 예컨대 화학적-반응성 쌍의 사용을 포함한다.

[0200] 실시예

[0201] 실시예 1. 화학적 라이게이션을 위한 성분의 제조 (이중-가닥 헤드피스 및 이중-가닥 태그)

[0202] 그의 5' 말단에서 화학적으로 인산화된 헤드피스 HP006 서열식별번호: 1 - (p)CCTGTGTTZTTCACGGCCT (여기서 Z는 C6-아미노 dT 변형을 의미함)를 바이오서치 인크.(Biosearch Inc.)에서 획득하였다. 이어서 하기 절차를 사용하여 Fmoc-NH-PEG4-CH<sub>2</sub>CH<sub>2</sub>COOH (캠 펩 인크(Chem Pep Inc))를 사용하여 DMT-MM 아실화에 의해 HP006을 변형시켰다.

[0203] 50 당량의 Fmoc-NH-PEG4-CH<sub>2</sub>CH<sub>2</sub>COOH (캠 펩 인크)를 DMA (디메틸 아세트아미드, 아크로스(Acros)) 중에 용해시키고, 0.5 M 보레이트 완충제 pH 9.5 중에 용해된 1 당량의 HP006에, 새롭게 물 중에 용해된 50 당량의 DMT-MM (4-(4,6-디메톡시-1,3,5-트리아진-2-일)-4-메틸모르폴리늄 클로라이드, 아크로스)과 함께 첨가하였다. 반응이 2-4시간 동안 진행되도록 한 다음, 50 당량의 Fmoc-NH-PEG4-CH<sub>2</sub>CH<sub>2</sub>COOH 및 50 당량의 DMT-MM의 제2 첨가를 행하고, 이어서 반응이 밤새 진행되도록 하였다. 반응의 완료를 LCMS에 의해 모니터링하였다.

- [0204] 생성물을 에탄올 침전시키고, 3,000 MW 컷-오프 원심분리 스핀 필터 (밀리포어(Millipore))를 사용하여 크기 배제 스핀 여과에 의해 탈염시켰다. 생성물의 LCMS는 MW를 6,803.3 (계산치 6,802.5)으로서 확인시켜주었다.
- [0205] 올리고뉴클레오타이드 TagZA1+ $\Delta$ C<sub>50H</sub>: 서열식별번호: 2 - 5' CATCAAGACCCAGAAAG-3', TagZB\_CNIm\_bot30H; 서열식별번호: 3 - 5'-(p)TCTGGGTCTTGATGGCTATCC-3' (5' 말단에서 화학적으로 인산화됨), PrA\_CNIm\_bot5P; 서열식별번호: 4 - 5'-(p)TGGCTGAGG-3' (5' 말단에서 화학적으로 인산화됨) 및 PrA\_top\_extraC<sub>3P</sub>; 서열식별번호: 5 - 5'-(p)CAGCCAGGATAGC(p)-3' (5' 및 3' 말단 둘 다에서 화학적으로 인산화됨)를 IDT DNA에서 획득하였다.
- [0206] 이어서 올리고 tagZA1+ $\Delta$ C 및 TagZB\_CNIm\_bot30H를 물 중 2 mM 최종 농도로 용해시키고, 등몰비로 조합하여 이중 가닥 TagZA의 1 mM 용액을 제조하였다.
- [0207] 또한 올리고 PrA\_CNIm\_bot5P 및 PrA\_top\_extraC<sub>3P</sub>를 물 중 2 mM 최종 농도로 용해시키고, 등몰비로 조합하여 이중 가닥 "CNIm-PrA"의 1 mM 용액을 제조하였다.
- [0208] 이어서 T4 DNA 리가제 및 표준 라이게이션 프로토콜을 사용하여 Fmoc-아미노-PEG4-HP006을 1 당량의 이중 가닥 CNIm-PrA에 효소적으로 라이게이션시켰다. 생성된 올리고, (Fmoc-아미노-PEG4-HP013)를 에탄올 침전시키고, 일러스트라 NAP-5 칼럼 (지이 헬스케어 라이프 사이언스(GE Healthcare Life Science))을 사용하여 탈염시켰다. LCMS는 MW 13,772 (계산치 13,770.7)를 확인시켜주었다.
- [0209] 실시예 2: 이중-가닥 헤드피스의 이중-가닥 태그에 대한 화학적 라이게이션
- [0210] Fmoc-아미노-PEG4-HP013 및 이중-가닥 TagZA 올리고뉴클레오타이드를 800 mM NaCl 및 8 mM ZnCl<sub>2</sub>를 함유하는 80 mM MES 완충제 pH 6.0 중 0.33 mM의 최종 농도로 용해시켰다. 1-시아노이미다졸을 새롭게 DMF 중 1 M로 용해시키고, 1-시아노이미다졸의 150mM의 최종 농도까지 12-시간에 걸쳐 반응물에 1-2 첨가를 행하였다. 이어서 반응물을 4°C에서 밤새 인큐베이션하였다.
- [0211] 완료된 반응물을 변성 겔 전기영동 뿐만 아니라 LCMS에 의해 분석하였다. 이어서 샘플을 15% 변성 분석 TBE-8M 우레아 겔 상에서 분해시키고, 형광 염료를 함유하는 TLC 플레이트 상에서 UV 새도잉에 의해 가시화하였다 (254 nm). LCMS는 MW 25,417.3 (계산치 25,415.3)을 갖는 이중-가닥 라이게이션된 생성물의 ~70% 전환율의 형성을 확인시켜주었다. MW 20,254.7 및 18,935.4를 갖는 추가의 생성물이 관찰되었고, 이는 (반-라이게이션된) 상부 또는 하부 가닥 라이게이션 생성물에 상응하였다.
- [0212] 15% TBE-8M 우레아 변성 겔에 의한 화학적 라이게이션 생성물의 분석적 겔 전기영동을 도 2에 제시한다:
- [0213] 1- 출발 물질- Fmoc-아미노-PEG4-HP013
- [0214] 2- dsTag ZA, tagZA1 $\Delta$ C<sub>50H</sub> 및 TagZA1+CNIm\_bot30H의 등몰 혼합물
- [0215] 3, 4, 5- 시아노이미다졸 라이게이션 반응
- [0216] 6- 효소적 라이게이션 대조군 (T4 DNA 리가제)은 단지 하부 가닥, 3' OH와 5'포스페이트 사이의 접합만을 라이게이션하고; 3'포스페이트와 5' OH 사이의 접합은 이러한 효소에 의해 라이게이션되지 않는다.
- [0217] 화학적 라이게이션 생성물의 LCMS를 도 3에 제시한다. (각각의 패널에서 - 상부 UV (260nm) LC 트레이스, 중간 - TIC, 하부- 질량 스펙트럼)
- [0218] A.- 출발 물질: 이중 가닥 TagZA (MW 5,182 및 6,500.2 Da) 뿐만 아니라 Fmoc-아미노-PEG4-HP013 (13,772)의 혼합물.
- [0219] B-화학적 라이게이션 반응의 생성물: 이중 라이게이션: MW 25,417.3 (계산치 25,415.3). 반-라이게이션 (상부 또는 하부 가닥) 생성물: MW 20,254.7 및 18,935.4.
- [0220] 실시예 3. 화학적 라이게이션 반응 생성물의 Fmoc 탈보호.
- [0221] 1-시아노이미다졸 라이게이션 반응의 생성물을 에탄올 침전시키고, 물 중 용해시키고, 10% 피페리딘 중 실온에서 2시간 동안 인큐베이션하여 탈보호시켰다. 이러한 탈보호 단계 후에, 물질을 15% TBE-8M 우레아 겔 상에서 정제하였다. 정제된 샘플에 대해 수행된 LC-MS는 탈보호된 아미노-PEG4-HP013-TagZA (MW 25,192.4, 계산치 25,193.2) 뿐만 아니라 2종의 반-라이게이션된 탈보호된 생성물 (MW 18,738.6 및 20,029.3)의 존재를 확인시켜주었다.
- [0222] LC 트레이스의 통합은 전장 생성물의 상대 수율을 64%로 제공하였고, 한편 반-라이게이션된 생성물은 각각 대략



18%로 제공하였다. 가닥당 라이게이션의 효율은 83%로 추정될 수 있다.

- [0223] 피페리딘에 의한 아미노 탈보호의 개략도를 도 4a에 제시한다. 라이게이션 반응 생성물의 겔 정제: 15% TBE-우레아 겔, UV 새도잉을 도 4b에 제시한다. 정제된 물질의 LCMS 분석을 도 4c에 제시한다. 전장 라이게이션 생성물 MW 25,192.4 Da, 반-라이게이션된 생성물 MW 18,738.6 및 20,029.3 Da
- [0224] 실시예 4. Fmoc에 의한 아미노-기 보호를 위한 필요성의 설명.
- [0225] 상기 기재된 바와 같이, 루프 내 T에서 아미노-C6 링커를 특색으로 하는 HP006을 1-시아노이미다졸을 함유하는 반응 혼합물 중에서 4℃에서 12시간 동안 인큐베이션하였다. 인큐베이션 후에, HP006을 에탄올-침전시키고, 10% 피페리딘 중 실온에서 2시간 동안 인큐베이션하고, 다시 에탄올 침전시켰다.
- [0226] 이러한 물질의 LCMS 분석은 혼합물 중 2종의 생성물, MW 6,333.4 Da HP006 및 MW 6,426.4 반응 생성물 (30-40% 전환)이 존재한다는 것을 입증하였다. 94 Da의 추가는 HP006의 N-이미다졸 구아니딘 유도체의 형성에 상응한다. 아미노 기의 Fmoc 보호는 이러한 원치 않는 반응을 완전히 제거한다.
- [0227] HP006과 1-시아노이미다졸의 반응의 생성물의 디컨볼루션된 질량 스펙트럼을 도 5a에 제시한다. MW 6,333.4 Da 은 비변형된 HP006에 상응하고, MW 6,426.4는 HP006의 N-이미다졸 구아니딘 유도체에 상응한다.
- [0228] HP006의 N-이미다졸 구아니딘 유도체의 생성의 개략도를 도 5b에 제시한다.
- [0229] 실시예 5. 대안적 2가 금속 이온에 의한 화학적 라이게이션
- [0230] 시아노이미다졸-매개 화학적 라이게이션을 상기 기재된 바와 같이 대안적 2가 금속 8 mM의 치환에 의해 수행하였다. CoCl<sub>2</sub> (30% 전장 생성물, 70% 반-라이게이션된 생성물), MnCl<sub>2</sub> (75% 전장 생성물, 25% 반-라이게이션된 생성물 및 ZnCl<sub>2</sub> (60% 전장 생성물, 30% 반-라이게이션된 생성물)에 의해 유의한 라이게이션 수율이 관찰되었다. 납, 마그네슘, 주석 및 구리의 가용성 2가 염은 어떠한 유의한 라이게이션도 생성하지 않았다.
- [0231] 실시예 6. 대안적 플랭킹 뉴클레오타이드에 의한 화학적 라이게이션
- [0232] 하기 화학적으로 인산화된 올리고뉴클레오타이드를 IDT DNA로부터 획득하였다.

상부 가닥, 쌍 1:

PrA\_top: SEQ ID NO: 6 - 5'-(p)CAGCCAGGATAG-3';

Tag\_ZA1+ : 5'-(p)CCATCAAGACCCAGAAAG-3';

상부 가닥, 쌍 2:

PrA\_top\_extraC\_3P: 5'-(p)CAGCCAGGATAGCp-3';

tagZA1\_deltaC\_5OH: 5'-CATCAAGACCCAGAAAG-3'

(중첩 서열, 볼드체)

하부 가닥, 쌍 A:

PrA\_CNIm\_bot5P: 5'-pTGGCTGAGG-3';

TagZB\_CNIm\_bot3OH: 5'-pTCTGGGTCTTGATGGCTATCC-3'

하부 가닥, 쌍 B:

PrA\_CNIm\_bot5OH: 5'-TGGCTGAGG-3':

TagZB\_CNIm\_bot3P: 5'-pTCTGGGTCTTGATGGCTATCCp-3'

[0233]

- [0234] 표 2에 제시된 바와 같은 올리고뉴클레오타이드의 4종 조합을 1-시아노이미다졸 라이게이션의 효율에 대해 시험하였다. 하부 가닥은 6- 및 7-뉴클레오타이드 둘 다의 중첩으로 (80% 초과) 및 플랭킹 뉴클레오타이드의 둘 다의 시험된 조합 (C 대 C 및 C 대 T)에서 일관되게 높은 라이게이션 수율을 입증한 한편, 상부 가닥 라이게이션은 플랭킹 뉴클레오타이드의 아이덴티티에 분명하게 의존적이었고, 예를 들어 C 대 G의 라이게이션은 비효율적인 반면에 C 대 C 접합은 높은 수율로 라이게이션되었다.

[0235] 표 2. 라이게이션 접합 설계의 개요 및 화학적 라이게이션의 수율

반응	중첩 길이 (nt)	라이게이션 접합부 (상부 가닥)	하부 가닥 라이게이션 접합부	상대 라이게이션 전환 (상부 가닥)
1-A	6	C-3'+5'pG	C-3'+5'pT	20%
1-B	6	C-3'+5'pG	Cp-3'+5'-T	25%
2-A	7	Cp-3'+5'-C	C-3'+5'pT	90%
2-B	7	Cp-3'+5'-C	Cp-3'+5'-T	95%

[0236]

[0237]

다른 실시양태

[0238]

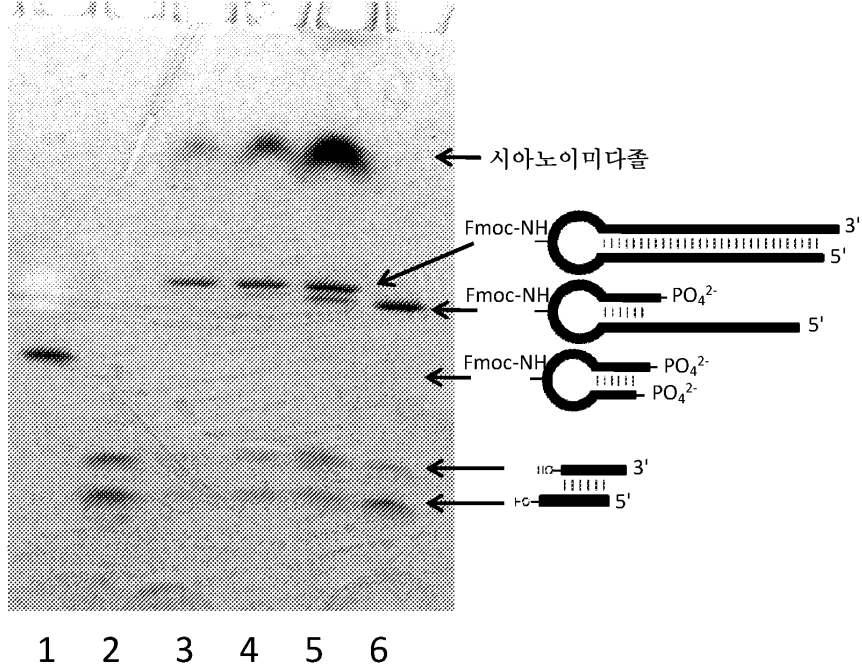
본 발명에 기재된 방법 및 시스템의 다양한 변형 및 변경이 본 발명의 범주 및 취지를 벗어나지 않으면서 관련 기술분야의 통상의 기술자에게 명백할 것이다. 본 발명이 구체적인 바람직한 실시양태와 관련하여 기재되었지만, 청구되는 바와 같은 발명은 이러한 구체적인 실시양태로 과도하게 제한되어서는 안된다는 것을 이해해야 한다. 실제로, 의학 분야, 약리학 분야 또는 관련 분야의 통상의 기술자에게 분명한 본 발명에 기재된 수행 방식의 다양한 변형은 본 발명의 범주에 속하는 것으로 의도된다.



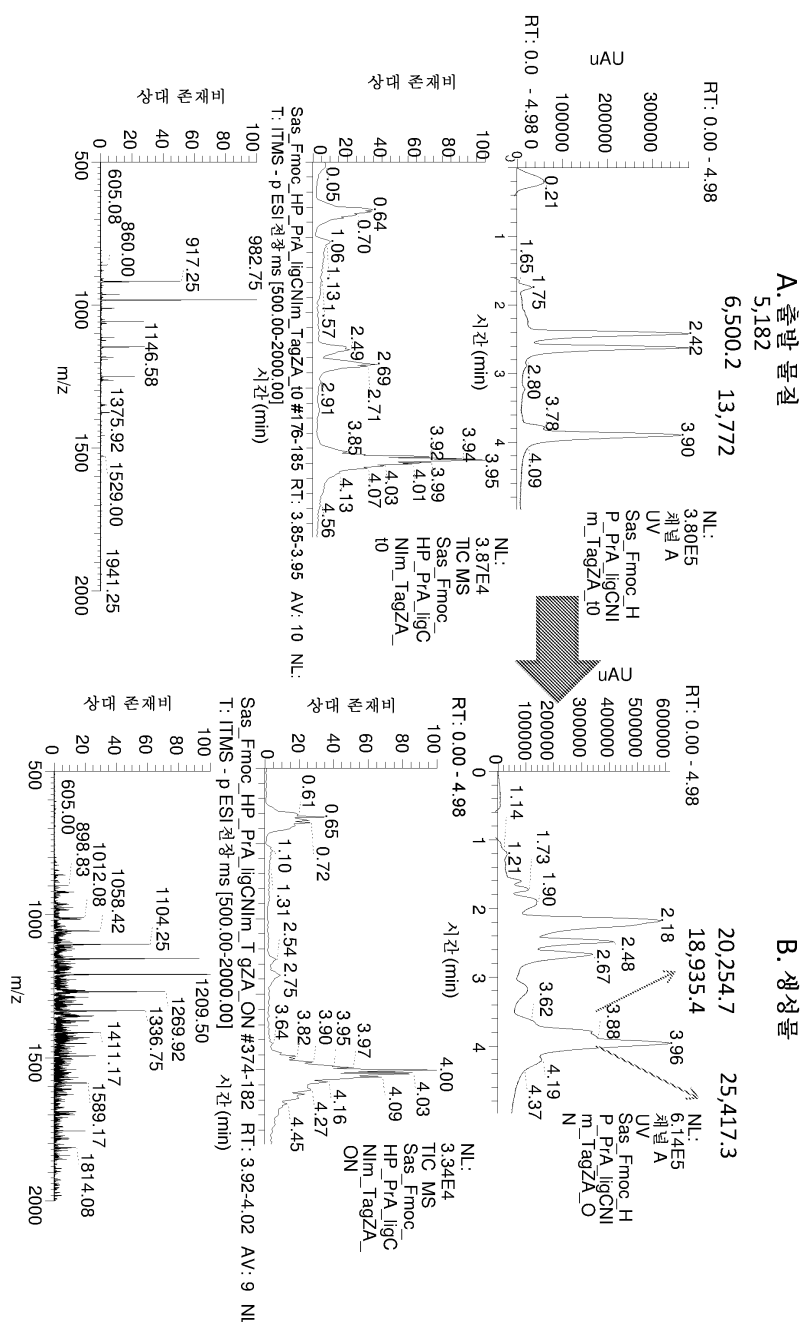


도면2

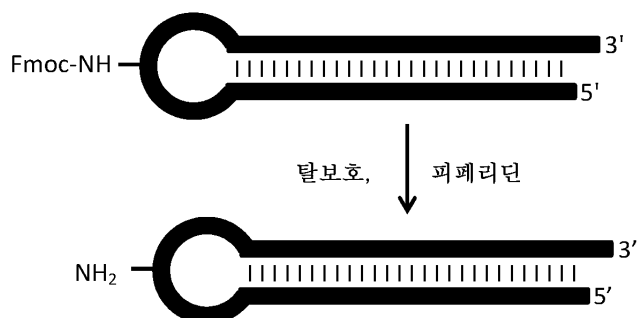
헤드피스	+		+	+	+	+
태그		+	+	+	+	+
시아노이미다졸			+	++	+++	
T4 DNA 리가제						+



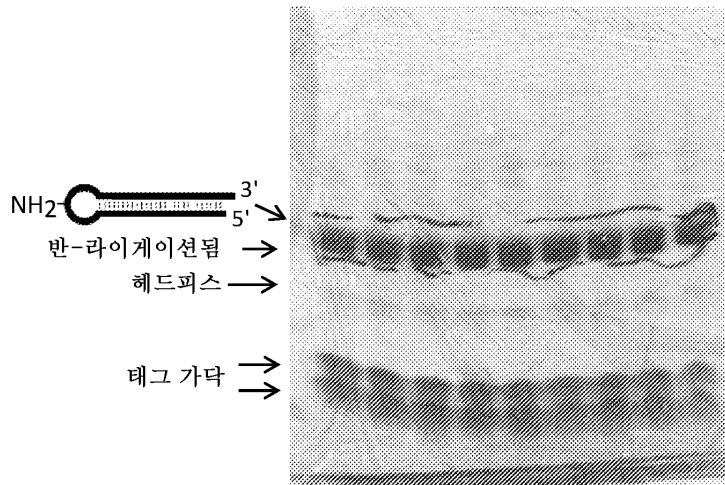
도면3



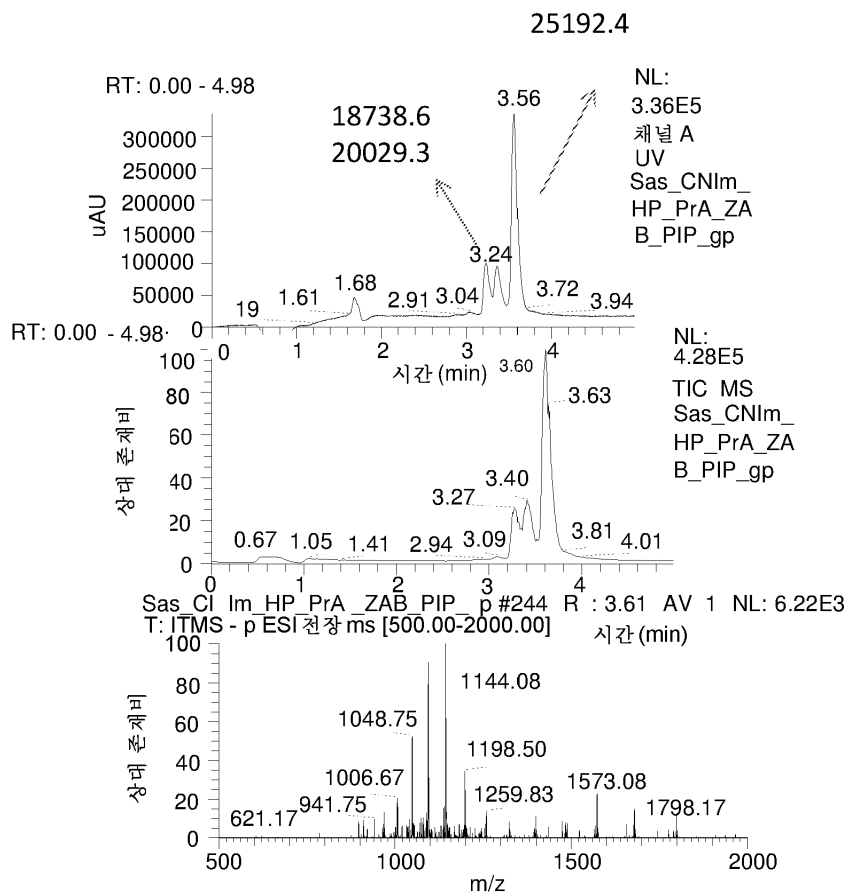
도면4a



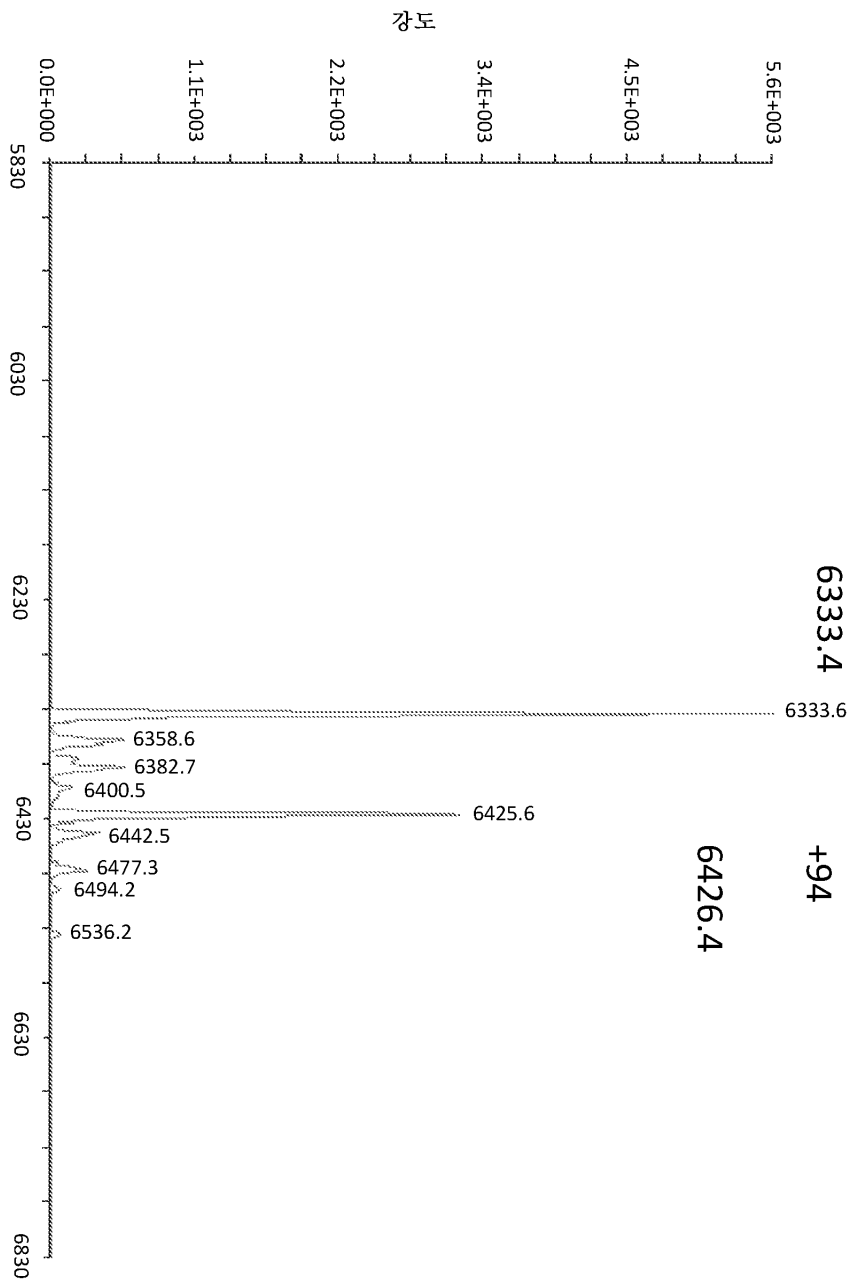
도면4b



도면4c

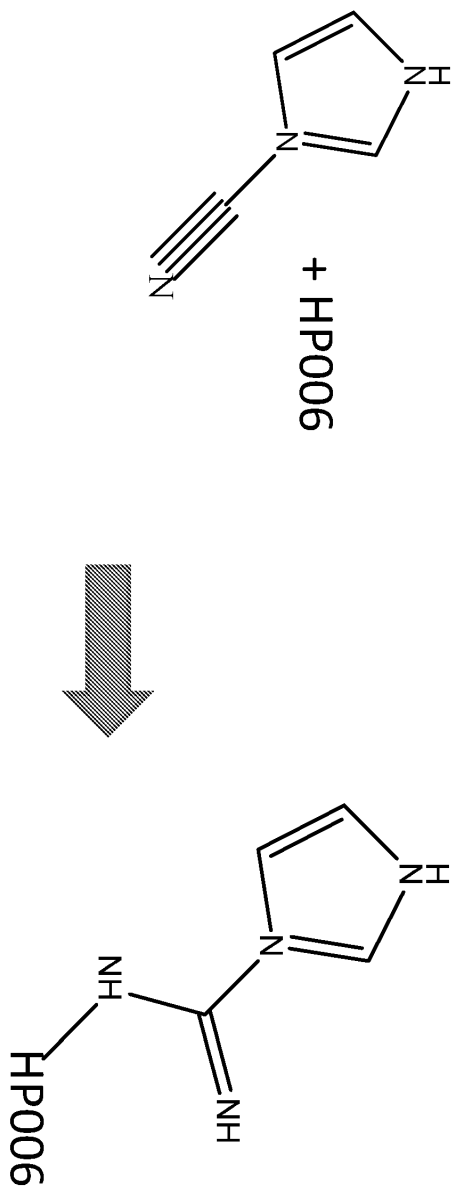


도면5a





도면5b



서열 목록

SEQUENCE LISTING

<110> X-CHEM, INC.

<120> METHODS FOR TAGGING DNA-ENCODED LIBRARIES

<130> 50719-033W02

<140> PCT/US 15/067667

<141> 2015-12-28

<150> US 62/098,037

<151> 2014-12-30

<160> 6  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Construct  
 <220><221> misc\_feature  
 <222> (9)..(9)  
 <223> n is T(C6-amino)  
 <400> 1  
 cctgtgttnt tcacggcct 19  
  
 <210> 2  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Construct  
 <400> 2  
 catcaagacc cagaaag 17  
  
 <210> 3  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Construct  
 <400> 3  
 tctgggtctt gatggctatc c 21  
  
 <210> 4  
 <211> 9  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Construct  
 <400> 4  
 tggctgagg 9

<210> 5	
<211> 13	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Construct	
<400> 5	
cagccaggat agc	13
<210> 6	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Construct	
<400> 6	
cagccaggat ag	12