(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년03월30일
(11) 등록번호 10-2516168
(24) 등록일자 2023년03월27일

(51) 국제특허분류(Int. Cl.)
G01N 33/68 (2006.01)
(52) CPC특허분류
G01N 33/6824 (2013.01)
G01N 33/6845 (2013.01)
(21) 출원번호 10-2022-7009243(분할)
(22) 출원일자(국제) 2017년05월02일
심사청구일자 2022년04월20일
(85) 번역문제출일자 2022년03월21일
(65) 공개번호 10-2022-0038549
(43) 공개일자 2022년03월28일
(62) 원출원 특허 10-2018-7034711
원출원일자(국제) 2017년05월02일
심사청구일자 2020년04월29일
(86) 국제출원번호 PCT/US2017/030702
(87) 국제공개번호 WO 2017/192633
국제공개일자 2017년11월09일
(30) 우선권주장
62/330,841 2016년05월02일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
WO2015164212 A1
WO2015166768 A1

(73) 특허권자
엔코디아, 인코포레이티드
미국 캘리포니아 92121 샌 디에고 스위트 120 로젤 스트리트 11045
(72) 발명자
치, 마크
미국 캘리포니아 92121 샌 디에고 스위트 비 플린트코트 애비뉴 11125
군데르손, 케빈
미국 캘리포니아 92121 샌 디에고 스위트 비 플린트코트 애비뉴 11125
웨이너, 마이클, 필립
미국 코네티컷 06437 길포드 레이튼 트레일 52
(74) 대리인
박장원

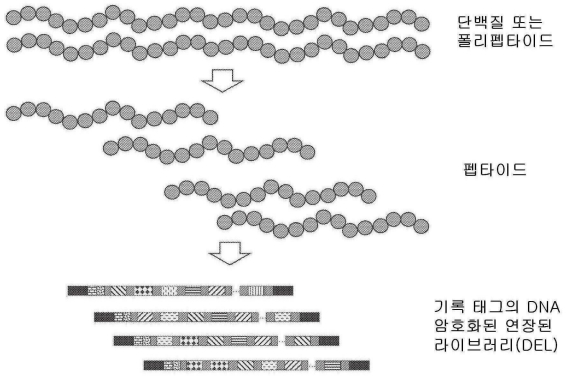
전체 청구항 수 : 총 7 항 심사관 : 차명훈

(54) 발명의 명칭 **암호화 핵산을 사용한 거대분자 분석**

(57) 요 약

핵산 암호화를 사용하여 펩타이드, 폴리펩타이드, 및 단백질을 포함하는, 거대 분자를 분석하는 방법이 개시되어 있다.

대 표 도



(52) CPC특허분류
G01N 2458/10 (2013.01)

(30) 우선권주장
62/339,071 2016년05월19일 미국(US)
62/376,886 2016년08월18일 미국(US)

명 세 서

청구범위

청구항 1

세포 집단을 포함하는 샘플로부터의 단일 세포의 펩타이드를 분석하는 방법으로서, 하기 단계를 포함하는 방법:

(a) 다수의 구획 내로 상기 세포 집단의 세포를 분배하는 단계로서, 여기서

i) 펩타이드를 포함하는 단일 세포 및 다수의 비드로부터의 하나 이상의 비드는 단일 구획 내에 포함되고;

ii) 상기 다수의 비드 중 각각의 비드는 그에 부착된 다수의 핵산 기록 태그를 포함하고, 여기서 상기 다수의 핵산 기록 태그 중 각각의 핵산 기록 태그는 각각의 구획에 특이적인 바코드 및 상기 펩타이드를 고정화하도록 구성된 기능성 모이어티를 포함하고;

(b) 상기 구획 내에서 세포를 용해시키고 상기 단일 구획 내에 포함된 하나 이상의 비드 상에서 단일 세포로부터 방출된 펩타이드의 고정화를 가능케 하여, 그에 의하여 단일 세포의 고정화된 바코드화된 펩타이드를 포함하는 하나 이상의 비드를 생성하는 단계로서, 여기서 상기 고정화된 바코드화된 펩타이드 각각은 상기 다수의 핵산 기록 태그 중의 핵산 기록 태그와 관련된 것이고;

(c) 다수의 구획으로부터의 상기 고정화된 바코드화된 펩타이드를 포함하는 비드를 방출하고 방출된 비드를 혼주시키는 (pooling) 단계; 및

(d) 상기 방출된 비드 상에서 상기 고정화된 바코드화된 펩타이드를 분석하는 단계로서, 여기서 분석은 다음 단계를 포함한다:

(i) 상기 고정화된 바코드화된 펩타이드를 상기 고정화된 바코드화된 펩타이드에 결합할 수 있는 다수의 결합제와 접촉시키는 단계;

(ii) 다수의 결합제로부터의 결합제의 고정화된 바코드화된 펩타이드와의 결합 후에, 상기 결합제에 관한 정보를 획득하여, 그에 의하여 상기 단일 세포의 상기 고정화된 바코드화된 펩타이드를 분석하는 단계.

청구항 2

제1항에 있어서,

구획 내로 세포를 분배하기 이전에, 상기 세포는 중합체-형성 소단위를 함유하는 소적 내에 캡슐화되고,

하나 이상의 비드 상에서 단일 세포로부터 방출된 상기 펩타이드의 고정화 이전에, 상기 펩타이드는 프로테아제로 소화되고,

상기 프로테아제 소화에 의하여 생성된 펩타이드는 상기 하나 이상의 비드 상에 고정화되는 것인, 방법.

청구항 3

제1항에 있어서,

하나 이상의 비드 상에서 단일 세포로부터 방출된 상기 펩타이드의 고정화 이전에, 상기 펩타이드는 프로테아제로 소화되고,

상기 프로테아제 소화에 의하여 생성된 펩타이드는 상기 하나 이상의 비드 상에 고정화되는 것인, 방법.

청구항 4

제1항에 있어서,

구획 내로 세포를 분배하기 이전에 또는 비드 상에서 상기 펩타이드의 고정화 이전에, 상기 펩타이드는, 상기 비드에 부착된 핵산 기록 태그의 기능성 모이어티와 반응할 수 있는 반응성 커플링 모이어티를 상기 펩타이드에 제공하도록 구성된 표지제와 접촉되는 것인, 방법.

청구항 5

◈청구항 5은(는) 설정등록료 납부시 포기되었습니다.◈

제4항에 있어서, 반응성 커플링 모이어티는 클릭 화학 반응을 통하여 핵산 기록 태그와 반응하는 것인, 방법.

청구항 6

◈청구항 6은(는) 설정등록료 납부시 포기되었습니다.◈

제4항에 있어서, 표지제와 접촉시키기 이전에, 상기 펩타이드는 변성되는 것인, 방법.

청구항 7

◈청구항 7은(는) 설정등록료 납부시 포기되었습니다.◈

제4항에 있어서,

하나 이상의 비드 상에서 단일 세포로부터 방출된 상기 펩타이드의 고정화 이전에, 상기 펩타이드는 프로테아제로 소화되고,

상기 프로테아제 소화에 의하여 생성된 펩타이드는 상기 하나 이상의 비드 상에 고정화되는 것인, 방법.

청구항 8

제1항 내지 제7항 중 어느 하나의 항에 있어서, 100개 또는 그 이상의 펩타이드가 동시에 분석되는 것인, 방법.

청구항 9

제1항 내지 제7항 중 어느 하나의 항에 있어서, 다수의 비드 중 각각의 비드 상에서, 핵산 기록 태그는 상기 비드의 표면상 또는 용적 내에서 서로 50 ㎚ 이상으로 떨어져 이격되는 것인, 방법.

청구항 10

제8항에 있어서, 방출된 비드 상에서 고정화된 바코드화된 펩타이드를 분석하는 단계는 하기 단계를 포함하는 것인, 방법:

(e) 상기 고정화된 바코드화된 펩타이드와 결합할 수 있는 다수의 결합제와 방출된 비드를 접촉시키는 단계로서, 여기서 각각의 결합제는 상기 결합제에 관한 확인 정보를 지닌 핵산 암호화 태그를 포함하고;

(f) 결합제의 고정화된 바코드화된 펩타이드와의 결합 후에, 상기 결합제에 관한 확인 정보를 지닌 핵산 암호화 태그의 정보를 상기 고정화된 바코드화된 펩타이드와 관련된 핵산 기록 태그로 이전시켜 상기 고정화된 바코드화된 펩타이드와 관련된 연장된 핵산 기록 태그를 생성시키는 단계로서, 여기서 이전은 프라이머 연장 또는 연결을 포함하는 것이고; 및

(g) 다수의 결합제로부터 정보의 이전 이후에 생성된 다수의 연장된 핵산 기록 태그를 분석하여, 그에 의하여 상기 단일 세포의 상기 고정화된 바코드화된 펩타이드를 분석하는 단계.

청구항 11

◈청구항 11은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서, 다수의 연장된 핵산 기록 태그는 핵산 시퀀싱 방법을 수행함으로써 병렬로 분석되는 것인, 방법.

청구항 12

◈청구항 12은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서, 핵산 암호화 태그의 정보를 핵산 기록 태그로 이전하는 것은 DNA 리가제에 의하여 또는 DNA 폴리머라제에 의하여 매개되는 것인, 방법.

청구항 13

◈청구항 13은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서, 다수의 결합제로부터의 하나 이상의 결합제는 고정화된 바코드화된 펩타이드의 N-말단 아미노산 (NTAA)에 결합할 수 있거나, 또는 작은 화학적 모이어티로 변형된 고정화된 바코드화된 펩타이드의 N-말단 아미노산 (NTAA)에 결합할 수 있는 것인, 방법.

청구항 14

◈청구항 14은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서, 단계 (e) 및 (f)는 단계 (g) 전에 1회 이상 반복되는 것인, 방법.

청구항 15

◈청구항 15은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서,

펩타이드를 고정화하도록 구성된 기능성 모이어티는 단백질 리가제의 펩타이드 기질 모티프를 포함하고,

상기 방법은 다수의 구획에 단백질 리가제를 첨가하여, 그에 의하여 펩타이드를 비드에 커플링시키는 단계를 더 포함하는 것인, 방법.

청구항 16

◈청구항 16은(는) 설정등록료 납부시 포기되었습니다.◈

제10항에 있어서, 방출된 비드 상에서 고정화된 바코드화된 펩타이드를 분석하는 것은 상기 고정화된 바코드화된 펩타이드의 적어도 부분 서열을 결정하는 것을 포함하는 것인, 방법.

발명의 설명

기 술 분 야

[0001] 서열 목록에 관한 기술

[0002] 본 출원과 관련된 서열 목록은 서류 사본 대신에 텍스트 문서양식으로 제공되며, 본원에서 명세서내에 참고로 포함된다. 서열 목록을 포함하는 텍스트 파일의 명칭은 760229_401WO_서열_LISTING.txt이다. 텍스트 파일은 38.7 KB이며, 2017년 5월 2일에 생성되었고, EFS-Web을 통해 전자적으로 제출된다.

[0003] 배경

[0004] 기술 분야

[0005] 본 개시내용은 일반적으로 펩타이드, 폴리펩타이드, 및 단백질을 포함하고, 분자 인식 사건을 암호화하는 핵산 및 바코드화(barcoding)를 사용하는, 거대분자(macromolecule)의 분석에 관한 것이다.

배 경 기 술

[0006] 관련 분야의 설명

[0007] 단백질은 세포 생물학 및 생리학에서 필수적인 역활을 담당하여, 많은 상이한 생물학적 기능을 수행하고 이를 촉진한다. 상이한 단백질 분자의 레퍼토리는 해독후 변형(post-translational modification: PTM)에 의해 유도된 추가의 다양성으로 인하여, 전사체(transcriptome)보다 강력하고, 훨씬 더 복잡하다. 또한, 세포내 단백질은 환경, 생리학적 상태, 및 질환 상태에 반응하여 현저하게 변화한다(발현 수준 및 변형 상태에서). 따라서, 단백질은 특히 게놈 정보와 관련하여 크게 탐험되지 않은 다량의 관련 정보를 포함한다. 일반적으로, 혁신이 게놈 분석과 관련하여 프로테오믹스 분석(proteomics analysis)에 있어서 지체되어 왔다. 유전체학 분야에서, 차세대 서열분석(next-generation sequencing: NGS)은 단일 기구 실행(single instrument run)에서 수십억개의 DNA 서열의 분석을 가능하게 함으로써 이러한 분야를 변환시켰지만, 단백질 분석 및 펩타이드 서열분석에서는, 처리량이 여전히 제한되어 있다.

[0008] 여전히 이러한 단백질 정보는 건강 및 질환에서 프로테오믹스 역학의 보다 나은 이해와 정밀한 의약이 가능하도

록 돕는 것을 직접적으로 필요로 한다. 따라서, 이러한 프로테오믹스 정보를 소형화하여 이의 수집을 고도로-병렬화하기 위한 "차-세대" 도구를 개발하는 것에 큰 관심이 있다.

[0009] 단백질의 고-병렬(high-parallel) 거대분자 특성화 및 인식은 몇가지 이유로 도전받고 있다. 친화성-기반 검정의 사용은 흔히 몇가지 주요 도전으로 인하여 어렵다. 하나의 유의적인 도전은 동종의 거대분자의 수집에 대해 친화성 제제의 수집의 판독물을 멀티플렉싱(multiplexing)화하는 것이고; 다른 도전은 친화성 제제와 오프-표적(off-target) 거대분자 사이의 교차-반응성을 최소화하는 것이며; 세번째 도전은 효율적인 고-처리량 판독물 플랫폼(high-throughput readout platform)을 개발하는 것이다. 이러한 문제점의 한가지 예는 하나의 목표가 샘플 속에서 대부분 또는 모든 단백질을 확인하고 정량화하는 것인 프로테오믹스에서 발생한다. 또한, 단일 분자 수준에서 단백질 상의 다양한 해독후 변형(PTM)을 특성화하는 것이 바람직하다. 현재, 이는 고-처리량 방식으로 달성할 방대한 업무이다.

[0010] 단백질 또는 펩타이드 거대분자의 분자 인식 및 특성화는 전형적으로 면역검정을 사용하여 수행된다. ELISA, 멀티플렉스(다수의x) ELISA(예컨대, 스포티드 항체 배열(spotted antibody array), 액체 입자 ELISA 배열), 디지탈 ELISA(예컨대, 콴테릭스(Quanterix), 싱귤렉스(Singulex)), 역상 단백질 배열(RPPA), 및 많은 다른 것들을 포함하는 많은 상이한 면역검정 포맷이 존재한다. 이들 상이한 면역검정 플랫폼 모두는 고 친화성 및 고-특이성(또는 선택성) 항체(결합제)의 개발, 샘플 및 분석물 수준 둘 다에서 멀티플렉스에 대한 제한된 능력, 제한된 민감성 및 역학적 범위, 및 교차-반응성 및 배경 시그널(signal)을 포함하는 유사한 도전을 마주하고 있다. 펩타이드 서열분석(에드만 분해(Edman degradation) 또는 질량 분광법)을 통한 직접적인 단백질 특성화와 같은 결합제에 구속받지 않는 시도는 유용한 대안적 시도를 제공한다. 그러나, 이들 시도들 중 어느 것도 매우 병렬이거나 고-처리량이 아니다.

[0011] 에드만 분해를 기반으로 하는 펩타이드 서열분석은 우선 1950년대에 페르 에드만(Pehr Edman)에 의해 처음 제안되었다; 즉 일련의 화학적 변형을 통한 펩타이드 상의 N-말단 아미노산의 단계적 분해 및 하부(downstream) HPLC 분석(후에 질량 분광법 분석으로 대체됨). 제1 단계에서, N-말단 아미노산은 온화한 염기성 조건(NMP/메탄올/H_2O) 하에서 페닐 이소티오시아네이트(PITC)로 변형되어 페닐티오카바모일(PTC) 유도체를 형성한다. 제2 단계에서, PTC-변형된 아미노 그룹을 산(무수 TFA)으로 처리하여 절단된 사이클릭 ATZ(2-아닐리노-5(4)-티오졸리논) 변형된 아미노산을 생성하여, 펩타이드 상에 새로운 N-말단을 남긴다. 절단된 사이클릭 ATZ-아미노산은 PTH-아미노산 유도체로 전환되어 역상 HPLC에 의해 분석된다. 이러한 과정은 펩타이드 서열을 포함하는 모든 또는 부분적인 수의 아미노산이 N-말단 끝으로부터 제거되어 확인될 때까지 반복적인 양식으로 계속된다. 일반적으로, 에드만 분해 펩타이드 서열분석은 느리며 하루당 단지 몇개의 펩타이드의 한정된 처리량을 갖는다.

[0012] 최근 10 내지 15년 동안, MALDI, 전자분무 질량 분광학(MS), 및 LC-MS/MS를 사용한 펩타이드 분석은 에드만 분해로 크게 대체되어 왔다. MS 계측(Riley et al., 2016, Cell Syst 2:142-143)에 있어서의 최근 진전에도 불구하고, MS는 고 장비 비용, 고급화 사용자에 대한 요건, 불량한 정량화 능력, 및 측정이 프로테옴의 역학적 범위에 이르도록 하는 제한된 능력을 포함하는 몇가지 단점을 여전히 겪고 있다. 예를 들면, 단백질은 상이한 수준의 효능에서 이온화되기 때문에, 샘플 사이의 절대적인 정량화 및 심지어 상대적인 정량화가 도전 중에 있다. 질량 태그(mass tag)의 실행은 상대적인 정량화를 개선시키는데 도움을 주었지만, 프로테옴의 표지화(labeling)를 필요로 한다. 역학적 범위는 추가로 복잡한데, 여기서 샘플 내의 단백질의 농도는 매우 큰 범위(혈장의 경우 10 차수 이상)에 걸쳐 변할 수 있다. MS는 전형적으로 보다 풍부한 종(species)을 단지 분석하여, 풍부하지 않은 단백질 챌린징(challenging)을 특성화한다. 최종적으로 샘플 처리량은 전형적으로 작동당 수천개의 펩타이드로 한정되며, 데이타 독립적인 분석(DIA)의 경우, 이러한 처리량은 실제 버튼-업 고-처리량 프로테옴 분석(bottoms-up high-throughput proteome analysis)의 경우 부적절하다. 또한, 각각의 샘플에 대해 기록된 수천개의 복잡한 MS 스펙트럼을 풀기위해(de-convolute) 유의적인 계산 요건이 존재한다.

발명의 내용

해결하려는 과제

[0013] 따라서, 서열분석 및/또는 분석에 대한 적용과 함께 거대분자 서열분석 및/또는 분석에 관한 개선된 기술, 및 이를 달성하기 위한 생성물, 방법 및 키트(kit)에 대해 당해 분야에서의 필요성이 남아있다. 고-병렬화되고, 정밀하며, 민감성이고 고-처리량인 프로테오믹스 기술에 대한 필요성이 존재한다. 본 개시내용은 이들 및 다른 필요성을 충족시킨다.

[0014] 본 발명의 이들 및 다른 국면은 다음의 상세한 설명을 참고로 명백해질 것이다. 이러한 목적으로, 다양한 참고 문헌이 본원에 제시되어 있으며, 이는 특정의 배경 정보, 과정, 화합물 및/또는 조성물을 보다 상세히 기술하며, 이들의 전문은 각각 본원에 참고로 포함된다.

과제의 해결 수단

[0015] 간단한 요약

[0016] 본 개시내용의 구현예는 일반적으로 고-병렬, 고 처리량의 디지탈 거대분자 분석, 특히 펩타이드 분석 방법에 관한 것이다.

[0017] 제1 구현예는:

[0018] (a) 거대분자 및, 고체 지지체에 결합된(joined) 관련 기록 태그를 제공하는 단계;

[0019] (b) 거대분자를 거대분자에 결합할 수 있는 제1 결합제와 접촉시키는 단계(여기서 제1 결합제는 제1 결합제에 관한 확인 정보를 지닌 제1 암호화 태그를 포함한다);

[0020] (c) 제1 암호화 태그의 정보를 기록 태그에 이전(transfer)시켜 제1 차수의 연장된 기록 태그(a first order extended recording tag)를 생성시키는 단계;

[0021] (d) 거대분자를 거대분자에 결합할 수 있는 제2 결합제와 접촉시키는 단계(여기서 제2 결합제는 제2 결합제에 관한 확인 정보를 지닌 제2 암호화 태그를 포함한다);

[0022] (e) 제2 암호화 태그의 정보를 제1 차수의 연장된 기록에 이전시켜 제2 차수의 연장된 기록 태그를 생성시키는 단계; 및

[0023] (f) 제2 차수 연장된 기록 태그를 분석하는 단계를 포함하는, 거대분자의 분석 방법이다.

[0024] 제2 구현예는 단계 (b) 및 (d)를 순차적인 순서로 수행하는, 제1 구현예의 방법이다.

[0025] 제3 구현예는 단계 (b) 및 (d)를 동시에 수행하는, 제1 구현예의 방법이다.

[0026] 제4 구현예는 단계 (e)와 (f) 사이에 다음의 단계를 추가로 포함하는, 제1 구현예의 방법이다:

[0027] (x) 제2 결합제를 거대분자에 결합할 수 있는 제3(또는 보다 높은 차수의) 결합제로 대체함으로써 단계 (d) 및 (e)를 1회 이상 반복하는 단계(여기서 제3(또는 보다 높은 차수의) 결합제는 제3(또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및

[0028] (y) 제3의(또는 보다 높은 차수의) 암호화 태그의 정보를 제2의(또는 보다 높은 차수의) 연장된 기록 태그로 이전시켜 제3의(또는 보다 높은 차수의) 연장된 기록 태그를 생성시키는 단계; 및

[0029] 여기서 제3의(또는 보다 높은 차수의) 연장된 기록 태그는 단계 (f)에서 분석된다.

[0030] 제5의 구현예는:

[0031] (a) 거대분자, 관련된 제1의 기록 태그 및 고체 지지체에 결합된 관련된 제2의 기록 태그를 제공하는 단계;

[0032] (b) 거대분자를 거대분자에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다);

[0033] (c) 제1의 암호화 태그의 정보를 제1의 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성시키는 단계;

[0034] (d) 거대분자를 거대분자에 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서, 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다);

[0035] (e) 제2의 암호화 태그의 정보를 제2의 기록 태그로 이전하여 제2의 연장된 기록 태그를 생성시키는 단계; 및

[0036] (f) 제1의 및 제2의 연장된 기록 태그를 분석하는 단계를 포함하는, 거대분자의 분석 방법이다.

[0037] 제6의 구현예는 단계 (b) 및 (d)를 순차적인 순서로 수행하는 제5 구현예의 방법이다.

[0038] 제7의 구현예는 단계 (b) 및 (d)를 동시에 수행하는 제5 구현예의 방법이다.

[0039] 제8 구현예는 단계 (a)가 고체 지지체에 결합된 관련된 제3의(또는 보다 높은 차수의) 기록 태그를 제공함을 추

가로 포함하는 제5 구현예의 방법이다.

[0040] 제9 구현예는 단계 (e)와 (f) 사이에 다음의 단계를 추가로 포함하는 제8 구현예의 방법이다:

[0041] (x) 제2의 결합제를 거대분자에 결합할 수 있는 제3의(또는 보다 높은 차수의) 결합제로 대체함으로써 단계 (d) 및 (e)를 1회 이상 반복하는 단계(여기서 제3의(또는 보다 높은 차수의) 결합제는 제3의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3의 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및

[0042] (y) 제3의 (또는 보다 높은 차수의) 암호화 태그의 정보를 제3의 (또는 보다 높은 차수의) 기록 태그로 이전시켜 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 생성시키는 단계;

[0043] 및 여기서 제1, 제2 및 제3의(또는 보다 높은 차수의) 연장된 기록 태그는 단계 (f)에서 분석된다.

[0044] 제10의 구현예는 제1의 암호화 태그, 제2의 암호화 태그, 및 어떠한 보다 높은 차수의 암호화 태그가 결합 주기 특이적인 스페이서 서열을 포함하는 제5 내지 제9의 구현예 중 어느 하나의 방법이다.

[0045] 제11의 구현예는:

[0046] (a) 펩타이드, 및 고체 지지체에 결합된 관련 기록 태그를 제공하는 단계;

[0047] (b) 펩타이드의 N-말단 아미노산 (NTAA)을 화학제(chemical agent)로 변형시키는 단계;

[0048] (c) 펩타이드를 변형된 NTAA에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다);

[0049] (d) 제1의 암호화 태그의 정보를 기록 태그로 이전시켜 연장된 기록 태그를 생성시키는 단계; 및

[0050] (e) 연장된 기록 태그를 분석하는 단계를 포함하는, 펩타이드를 분석하는 방법이다.

[0051] 제12의 구현예는 단계 (c)가 펩타이드를 제2의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제2의 (또는 보다 높은 차수의) 암호화 태그를 포함하는 제2의 (또는 보다 높은 차수의) 결합제와 접촉시킴을 추가로 포함하는, 제11의 구현예의 방법이며, 여기서 제2의 (또는 보다 높은 차수의) 결합제는 단계 (b)의 변형된 NTAA 이외의 변형된 NTAA에 결합할 수 있다.

[0052] 제13의 구현예는 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉이 펩타이드가 제1의 결합제와 접촉된 후 순차적인 순서로 일어나는 제12의 구현예의 방법이다.

[0053] 제14의 구현예는 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉이 펩타이드가 제1의 결합제와 접촉하는 것과 동시에 일어나는 제12의 구현예의 방법이다.

[0054] 제15의 구현예는 제11 내지 제14의 구현예 중 어느 하나의 방법이며, 여기서 화학제는 이소티오시아네이트 유도체, 2,4-디니트로벤젠설포닉 (DNBS), 4-설포닐-2-니트로플루오로벤젠(SNFB) 1-플루오로-2,4-디니트로벤젠, 단실 클로라이드, 7-메톡시쿠마린 아세트산, 티오아실화 시약, 티오아세틸화 시약, 또는 티오벤질화 시약이다.

[0055] 제16의 구현예는:

[0056] (a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계;

[0057] (b) 펩타이드의 N-말단 아미노산 (NTAA)을 화학제로 변형시켜 변형된 NTAA를 수득하는 단계;

[0058] (c) 펩타이드를 변형된 NTAA에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다);

[0059] (d) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성시키는 단계;

[0060] (e) 변형된 NTAA를 제거하여 새로운 NTAA를 노출시키는 단계;

[0061] (f) 펩타이드의 새로운 NTAA를 화학제로 변형시켜 새로이 변형된 NTAA를 수득하는 단계;

[0062] (g) 펩타이드를 새로이 변형된 NTAA에 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다);

[0063] (h) 제2의 암호화 태그의 정보를 제1의 연장된 기록 태그로 이전시켜 제2의 연장된 기록 태그를 생성하는 단계; 및

[0064] (i) 제2의 연장된 기록 태그를 분석하는 단계를 포함하는, 펩타이드의 분석 방법이다.

[0065] 제17의 구현예는:

[0066] (a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계;

[0067] (b) 펩타이드를 펩타이드의 N-말단 아미노산 (NTAA)에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다);

[0068] (c) 제1의 암호화 태그의 정보를 기록 태그로 이전시켜 연장된 기록 태그를 생성하는 단계; 및

[0069] (d) 연장된 기록 태그를 확인하는 단계를 포함하는, 펩타이드의 분석 방법이다.

[0070] 제18의 구현예는 제17의 구현예의 방법이며, 여기서 단계 (b)는 펩타이드를 제2의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제2의 (또는 보다 높은 차수의) 암호화 태그를 포함하는 제2의 (또는 보다 높은 차수의) 결합제와 접촉시킴을 추가로 포함하며, 여기서 제2의 (또는 보다 높은 차수의) 결합제는 펩타이드의 NTAA 이외의 다른 NTAA에 결합할 수 있다.

[0071] 제19 구현예는 제18의 구현예의 방법이며, 여기서 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉은 펩타이드가 제1의 결합제와 접촉한 후 순차적인 순서로 일어난다.

[0072] 제20의 구현예는 제18의 구현예의 방법이며, 여기서 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉은 펩타이드가 제1의 결합제와 접촉하는 것과 동시에 일어난다.

[0073] 제21의 구현예는:

[0074] (a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계;

[0075] (b) 펩타이드를 펩타이드의 N-말단 아미노산 (NTAA)에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다);

[0076] (c) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성시키는 단계;

[0077] (d) NTAA를 제거하여 펩타이드의 새로운 NTAA를 노출시키는 단계;

[0078] (e) 펩타이드를 새로운 NTAA에 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다);

[0079] (h) 제2의 암호화 태그의 정보를 제1의 연장된 기록 태그로 이전시켜 제2의 연장된 기록 태그를 생성하는 단계; 및

[0080] (i) 제2의 연장된 기록 태그를 분석하는 단계를 포함하는, 펩타이드의 분석 방법이다.

[0081] 제22의 구현예는 제1 내지 제10의 구현예 중의 어느 하나의 방법이며, 여기서 거대분자는 단백질, 폴리펩타이드 또는 펩타이드이다.

[0082] 제23의 구현예는 제1 내지 제10의 구현예 중 어느 하나의 방법이며, 여기서 거대분자는 펩타이드이다.

[0083] 제24의 구현예는 제11 내지 제23의 구현예 중 어느 하나의 방법이며, 여기서 펩타이드는 생물학적 샘플로부터의 단백질의 단편화(fragmenting)에 의해 수득된다.

[0084] 제25의 구현예는 제1 내지 제10의 구현예 중 어느 하나의 방법이며, 여기서 거대분자는 지질, 탄수화물, 또는 마크로사이클(macrocycle)이다.

[0085] 제26의 구현예는 제1 내지 제25의 구현예 중 어느 하나의 방법이며, 여기서 상기 기록 태그는 DNA 분자, 슈도-상보성 염기를 지닌 DNA, RNA 분자, BNA 분자, XNA 분자, LNA 분자, PNA 분자, γPNA 분자, 또는 이의 조합이다.

[0086] 제27의 구현예는 제1 내지 제26의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 보편적인 프라이밍 부위(universal priming site)를 포함한다.

[0087] 제28의 구현예는 제27의 구현예의 방법이며, 여기서 보편적인 프라이밍 부위는 증폭, 서열분석, 또는 둘 다를 위한 프라이밍 부위를 포함한다.

[0088] 제29의 구현예는 제1 내지 제28의 구현예중 어느 하나의 방법이며, 여기서 기록 태그는 유일한 분자 확인인자

(unique molucule identifier: UMI)를 포함한다.

[0089] 제30의 구현예는 제1 내지 제29의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 바코드를 포함한다.

[0090] 제31의 구현예는 제1 내지 제30의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 이의 3'-말단(terminus)에 스페이서(spacer)를 포함한다.

[0091] 제32의 구현예는 제1 내지 제31의 구현예 중 어느 하나의 방법이며, 여기서 거대분자 및 관련된 기록 태그는 고체 지지체에 공유결합으로 결합된다.

[0092] 제33의 구현예는 제1 내지 제32의 구현예 중 어느 하나의 방법이며, 여기서 고체 지지체는 비드, 다공성 비드, 다공성 매트릭스, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩(silicon wafer chip), 유동 통과 칩(flow through chip), 전자 장치를 변환하는 시그널을 포함하는 바이오칩, 미세역가 웰(microtitre well), ELISA 플레이트, 스피닝 인터페로메트리 디스크(spinning interferometry disc), 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구이다.

[0093] 제34의 구현예는 제33의 구현예의 방법이며, 여기서 고체 지지체는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드이다.

[0094] 제35의 구현예는 제1 내지 제34의 구현예 중 어느 하나의 방법이며, 여기서 다수의 거대분자 및 관련된 기록 태그는 고체 지지체에 결합된다.

[0095] 제36의 구현예는 제35의 구현예의 방법이며, 여기서 다수의 거대분자는 > 50 nm의 평균 거리에서 고체 지지체 상에 이격되어 있다.

[0096] 제37의 구현예는 제1 내지 제36의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 폴리펩타이드 또는 단백질이다.

[0097] 제38의 구현예는 제37의 구현예의 방법이며, 여기서 결합제는 변형된 아미노펩타이드, 변형된 아미노 아실 tRNA 신테타제(synthetase), 변형된 안티칼린, 또는 변형된 ClpS이다.

[0098] 제39의 구현예는 제1 내지 제38의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 거대분자에 선택적으로 결합할 수 있다.

[0099] 제40의 구현예는 제1 내지 제39의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 DNA 분자, RNA 분자, BNA 분자, XNA 분자, LNA 분자, PNA 분자, γPNA 분자, 또는 이의 조합이다.

[0100] 제41의 구현예는 제1 내지 제40의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 암호기 서열(encorder sequence)을 포함한다.

[0101] 제42의 구현예는 제1 내지 제41의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 스페이서, 결합 주기 특이적인 서열, 유일한 분자 확인인자, 보편적인 프라이밍 부위, 또는 이의 어떠한 조합을 포함한다.

[0102] 제43의 구현예는 제1 내지 제42의 구현예 중 어느 하나의 방법이며, 여기서 결합제 및 암호화 태그는 링커(linker)에 의해 결합된다.

[0103] 제44의 구현예는 제1 내지 제42의 구현예 중 어느 하나의 방법이며, 여기서 결합제 및 암호화 태그는 SpyTag/SpyCatcher 또는 SnoopTag/SnoopCatcher 펩타이드-단백질 쌍에 의해 결합된다.

[0104] 제45의 구현예는 제1 내지 제44의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그의 정보의 기록 태그로의 이전은 DNA 리가제에 의해 매개된다.

[0105] 제46의 구현예는 제1 내지 제44의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그의 정보의 기록 태그로의 이전은 DNA 폴리머라제에 의해 매개된다.

[0106] 제47의 구현예는 제1 내지 제44의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그의 정보의 기록 태그로의 이전은 화학적 연결에 의해 매개된다.

[0107] 제48의 구현예는 제1 내지 제47의 구현예 중 어느 하나의 방법이며, 여기서 연장된 기록 태그의 분석은 핵산 서열분석 방법을 포함한다.

[0108] 제49의 구현예는 제48의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 합성에 의한 서열분석, 연결에 의한

서열분석, 하이브리드화에 의한 서열분석, 폴로니(polony) 서열분석, 이온 반도체 서열분석, 또는 피로서열분석(pyrosequencing)이다.

[0109] 제50의 구현예는 제48의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 또는 진전된 현미경을 사용한 DNA의 직접적인 영상화(direct imaging)이다.

[0110] 제51의 구현예는 제1 내지 제50의 구현예 중 어느 하나의 방법이며, 여기서 연장된 기록 태그는 분석 전에 증폭된다.

[0111] 제52의 구현예는 제 1 내지 제51의 구현예 중 어느 하나의 방법이며, 여기서 연장된 기록 태그에 함유된 암호화 태그 정보의 순서는 거대분자에 대해 결합제가 결합하는 순서에 관한 정보를 제공한다.

[0112] 제53의 구현예는 제1 내지 제52의 구현예 중 어느 하나의 방법이며, 여기서 연장된 기록 태그 상에 함유된 암호화 태그 정보의 빈도는 결합제가 거대분자에 결합하는 빈도에 관한 정보를 제공한다.

[0113] 제54의 구현예는 제1 내지 제53의 구현예 중 어느 하나의 방법이며, 여기서 다수의 거대분자를 나타내는 다수의 연장된 기록 태그는 병렬로 분석된다.

[0114] 제55의 구현예는 제 54의 구현예의 방법이며, 여기서 다수의 거대분자를 나타내는 다수의 연장된 기록 태그는 멀티플렉스화된 검정으로 분석된다.

[0115] 제56의 구현예는 제1 내지 제55의 구현예 중 어느 하나의 방법이며, 여기서 다수의 연장된 기록 태그는 분석 전에 표적 농축 검정을 겪는다.

[0116] 제57의 구현예는 제1 내지 제56의 구현예 중 어느 하나의 방법이며, 여기서 다수의 연장된 기록 태그는 분석 전에 공제 검정(subtraction assay)을 겪는다.

[0117] 제58의 구현예는 제1 내지 제57의 구현예 중 어느 하나의 방법이며, 여기서 다수의 연장된 기록 태그는 분석 전에 매우 풍부한 종을 감소시키는 정상화 검정을 겪는다.

[0118] 제59의 구현예는 제1 내지 제58의 구현예 중 어느 하나의 방법이며, 여기서 NTAA는 변형된 아미노펩티다제, 변형된 아미노산 tRNA 신테타제, 온화한 에드만 분해, 에드마나제 효소(Edmanase enzyme), 또는 무수 TFA에 의해 제거된다.

[0119] 제60의 구현예는 제1 내지 제59의 구현예 중 어느 하나의 방법이며, 여기서 적어도 하나의 결합제는 말단 아미노산 잔기에 결합한다.

[0120] 제61의 구현예는 제1 내지 제60의 구현예 중 어느 하나의 방법이며, 여기서 적어도 하나의 결합제는 해독후 변형된 아미노산에 결합한다.

[0121] 제62의 구현예는:

[0122] (a) 샘플 내의 다수의 단백질 복합체(complex), 단백질, 또는 폴리펩타이드를 다수의 구획(compartment)내로 분배하는 단계(여기서 각각의 구획은 고체 지지체에 임의로 결합된 다수의 구획 태그를 포함하고, 여기서 다수의 구획 태그는 개개 구획내에서 동일하고 다른 구획의 구획 태그와는 상이하다);

[0123] (b) 다수의 단백질 복합체, 단백질, 및/또는 폴리펩타이드를 다수의 펩타이드로 단편화하는 단계;

[0124] (c) 다수의 펩타이드가 다수의 구획 태그와 다수의 구획 내에서 어닐링하거나 결합하도록 하기에 충분한 조건하에서 다수의 펩타이드를 다수의 구획 태그와 접촉시킴으로써, 다수의 구획 태그된 펩타이드를 생성하는 단계;

[0125] (d) 구획 태그된 펩타이드를 다수의 구획으로부터 수집하는 단계; 및

[0126] (e) 제1 내지 제21의 구현예 중 어느 하나 및 제26 내지 제61의 구현예 중 어느 하나의 방법에 따른 하나 이상의 구획 태그된 펩타이드를 분석하는 단계를 포함하는, 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드를 포함하는 샘플로부터 하나 이상의 펩타이드를 분석하는 방법이다.

[0127] 제63의 구현예는 제62의 구현예의 방법이며, 여기서 구획은 미세유동 소적(microfluidic droplet)이다.

[0128] 제64의 구현예는 제62의 구현예의 방법이며, 여기서 구획은 마이크로웰이다.

[0129] 제65의 구현예는 제62의 구현예의 방법이며, 여기서 구획은 표면 상에 분리된 영역이다.

[0130] 제66의 구현예는 제62 내지 제65의 구현예 중 어느 하나의 방법이며, 여기서 각각의 구획은 평균적으로 단일 세

포(single cell)를 포함한다.

[0131] 제67의 구현예는:

[0132] (a) 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드를 다수의 보편적인 DNA 태그로 표지하는 단계;

[0133] (b) 샘플내 다수의 표지된 단백질 복합체, 단백질, 또는 폴리펩타이드를 다수의 구획내로 분배하는 단계(여기서 각각의 구획은 다수의 구획 태그를 포함하고, 여기서 다수의 구획 태그는 개개의 구획 내에서 동일하며 다른 구획의 구획 태그와는 상이하다);

[0134] (c) 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드와 다수의 구획내 다수의 구획 태그의 어닐링 또는 결합을 허용하기에 충분한 조건하에서 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드를 다수의 구획 태그와 접촉시킴으로써, 다수의 구획 태그된 단백질 복합체, 단백질 또는 폴리펩타이드를 생성하는 단계;

[0135] (d) 구획 태그된 단백질 복합체, 단백질, 또는 폴리펩타이드를 다수의 구획으로부터 수집하는 단계;

[0136] (e) 구획 태그된 단백질 복합체, 단백질, 또는 폴리펩타이드를 구획 태그된 펩타이드내로 임의로 단편화하는 단계; 및

[0137] (f) 제1 내지 제21의 구현예 및 제26 내지 제61의 구현예 중 어느 하나의 방법에 따른 하나 이상의 구획 태그된 펩타이드를 분석하는 단계를 포함하는, 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드를 포함하는 샘플로부터 하나 이상의 펩타이드를 분석하는 방법이다.

[0138] 제68의 구현예는 제62 내지 제67의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그 정보는 프라이머 연장 또는 연결을 통해 펩타이드와 관련된 기록 태그로 이전된다.

[0139] 제69의 구현예는 제62 내지 제68의 구현예 중 어느 하나의 방법이며, 여기서 고체 지지체는 비드를 포함한다.

[0140] 제70의 구현예는 제69의 구현예의 방법이며, 여기서 비드는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드이다.

[0141] 제71의 구현예는 제62 내지 제70의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그는 단일 가닥(single stranded) 또는 이중 가닥 핵산 분자를 포함한다.

[0142] 제72의 구현예는 제62 내지 제71의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그는 바코드 및 임의로 UMI를 포함한다.

[0143] 제73의 구현예는 제71의 구현예의 방법이며, 여기서 고체 지지체는 비드이고구획 태그는 바코드를 포함하며, 또한 여기서 이에 결합된 다수의 구획 태그를 포함하는 비드는 분할-및-풀 합성(split-and-pool synthesis)에 의해 형성된다.

[0144] 제74의 구현예는 제71의 구현예의 방법이고, 여기서 고체 지지체는 비드이고 구획 태그는 바코드를 포함하며, 또한 여기서 이에 결합된 다수의 구획 태그를 포함하는 비드는 개개의 합성 또는 고정화에 의해 형성된다.

[0145] 제75의 구현예는 제62 내지 제74의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그는 기록 태그 내의 성분이고, 여기서 기록 태그는 임의로 스페이서, 유일한 분자 확인인자, 보편적인 프라이밍 부위, 또는 이의 어떠한 조합을 추가로 포함한다.

[0146] 제76의 구현예는 제62 내지 제75의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그는 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드 상의 내부 아미노산 또는 N-말단 아미노산과 반응할 수 있는 기능성 모이어티(moiety)를 추가로 포함한다.

[0147] 제77의 구현예는 제76의 구현예의 방법이고, 여기서 기능성 모이어티는 NHS 그룹이다.

[0148] 제78의 구현예는 제76의 구현예의 방법이고, 여기서 기능성 모이어티는 알데하이드 그룹이다.

[0149] 제79의 구현예는 제62 내지 제78의 구현예 중 어느 하나의 방법이며, 여기서 다수의 구획 태그는 구획 태그를 구획내로 프린팅, 스포팅(spotting), 잉크-젯팅(ink-jetting), 또는 이의 조합에 의해 형성된다.

[0150] 제80의 구현예는 제62 내지 제79의 구현예 중 어느 하나의 방법이며, 여기서 구획 태그는 펩타이드를 추가로 포함한다.

[0151] 제81의 구현예는 제80의 구현예의 방법이며, 여기서 구획 태그 펩타이드는 단백질 리가제 인식 서열을

포함한다.

[0152] 제82의 구현예는 제81의 구현예의 방법이며, 여기서 단백질 리가제는 부텔라제 I 또는 이의 동족체이다.

[0153] 제83의 구현예는 제62 내지 제82의 구현예 중 어느 하나의 방법이며, 여기서 다수의 폴리펩타이드는 프로테아제로 단편화된다.

[0154] 제84의 구현예는 제83의 구현예의 방법이며, 여기서 프로테아제는 메탈로프로테아제이다.

[0155] 제85의 구현예는 제84의 구현예의 방법이며, 여기서 메탈로프로테아제의 활성은 금속성 양이온의 광-활성화된 방출에 의해 조절된다.

[0156] 제86의 구현예는 다수의 폴리펩타이드를 다수의 구획내로 분배하기 전에 샘플로부터 하나 이상의 풍부한 단백질을 공제함을 추가로 포함하는 제62 내지 제85의 구현예 중 어느 하나의 방법이다.

[0157] 제87의 구현예는 다수의 펩타이드를 구획 태그와 결합시키기 전에 고체 지지체로부터 구획 태그를 방출시킴을 추가로 포함하는, 제62 내지 제86의 구현예 중 어느 하나의 방법이다.

[0158] 제88의 구현예는 단계 (d) 이후에, 구획 태그된 펩타이드를 기록 태그와 함께 고체 지지체에 결합시킴을 추가로 포함하는, 제62의 구현예의 방법이다.

[0159] 제89의 구현예는 구획 태그된 펩타이드 상의 구획 태그의 정보를 관련된 기록 태그로 이전시킴을 추가로 포함하는, 제88의 구현예의 방법이다.

[0160] 제90의 구현예는 단계 (e) 이전에 구획 태그된 펩타이드로부터 구획 태그를 제거함을 추가로 포함하는, 제89의 구현예의 방법이다.

[0161] 제91의 구현예는 이로부터의 분석된 펩타이드가 분석된 펩타이드의 구획 태그 서열을 기준으로 하여 유도되는 단일 세포의 동일성(identity)을 확인함을 추가로 포함하는, 제62 내지 제90의 구현예 중 어느 하나의 방법이다.

[0162] 제92의 구현예는 이로부터의 분석된 펩타이드가 분석된 펩타이드의 구획 태그 서열을 기준으로 하여 유도되는 단백질 또는 단백질 복합체의 동일성을 측정함을 추가로 포함하는, 제62 내지 제90의 구현예 중 어느 하나의 방법이다.

[0163] 제93의 구현예는:

[0164] (a) 다수의 거대분자 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계;

[0165] (b) 다수의 거대분자를 다수의 거대 분자에 결합할 수 있는 다수의 결합제와 접촉시키는 단계(여기서 각각의 결합제는 결합제에 관한 확인 정보를 지닌 암호화 태그를 포함한다);

[0166] (c) (i) 기록 태그와 관련된 거대분자의 정보를 거대 분자에 결합된 결합제의 암호화 태그로 이전시켜 연장된 암호화 태그를 생성하거나; (ii) 거대분자에 결합된 결합제의 거대분자 관련된 기록 태그 및 암호화 태그의 정보를 디-태그 작제물(di-tag construct)로 이전시키는 단계;

[0167] (d) 연장된 암호화 태그 또는 디-태그 작제물을 수집하는 단계;

[0168] (e) 임의로 하나 이상의 결합 주기를 위해 단계 (b) 내지 (d)를 반복하는 단계;

[0169] (f) 연장된 암호화 태그 또는 디-태그 작제물의 수집물을 분석하는 단계를 포함하는, 다수의 거대분자를 분석하는 방법이다.

[0170] 제94의 구현예는 제93의 구현예의 방법이며, 여기서 거대분자는 단백질이다.

[0171] 제95의 구현예는 제93의 구현예의 방법이며, 여기서 거대분자는 펩타이드이다.

[0172] 제96의 구현예는 제95의 구현예의 방법이며, 여기서 펩타이드는 생물학적 샘플로부터 단백질을 분획화하여 수득된다.

[0173] 제97의 구현예는 제93 내지 제96의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 DNA 분자, RNA 분자, PNA 분자, BNA 분자, XNA, 분자, LNA 분자, γPNA 분자, 또는 이의 조합이다.

[0174] 제98의 구현예는 제93 내지 제97의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 유일한 분자 확인인자

(UMI)를 포함한다.

[0175] 제99의 구현예는 제93 내지 제98의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 구획 태그를 포함한다.

[0176] 제100의 구현예는 제93 내지 제99의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 보편적인 프라이밍 부위를 포함한다.

[0177] 제101의 구현예는 제93 내지 제100의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 이의 3'-말단에 스페이서를 포함한다.

[0178] 제102의 구현예는 제93 내지 제101의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그의 3'-말단은 차단되어 폴리머라제에 의한 기록 태그의 연장을 방지하며 거대분자에 결합된 결합제의 거대분자 관련된 기록 태그 및 암호화 태그의 정보는 디-태그 작제물로 이전된다.

[0179] 제103의 구현예는 제93 내지 제102 중 어느 하나의 방법이며, 여기서 암호화 태그는 암호기 서열을 포함한다.

[0180] 제104의 구현예는 제93 내지 제103 중 어느 하나의 방법이며, 여기서 암호화 태그는 UMI를 포함한다.

[0181] 제105의 구현예는 제93 내지 제104의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 보편적인 프라이밍 부위를 포함한다.

[0182] 제106의 구현예는 제93 내지 제105의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 이의 3'-말단에서 스페이서를 포함한다.

[0183] 제107의 구현예는 제93 내지 제106의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 결합 주기 특이적인 서열을 포함한다.

[0184] 제108의 구현예는 제93 내지 제107의 구현예 중 어느 하나의 방법이며, 여기서 결합제 및 암호화 태그는 링커에 의해 결합된다.

[0185] 제109의 구현예는 제93 내지 제108의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그의 정보의 암호화 태그로의 이전은 프라이머 연장에 의해 달성된다.

[0186] 제110의 구현예는 제93 내지 제108의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그의 정보를 암호화 태그로 이전하는 것은 연결에 의해 달성된다.

[0187] 제111의 구현예는 제93 내지 제108의 구현예 중 어느 하나의 방법이며, 여기서 디-태그 작제물은 갭 충전(gap fill), 프라이머 연장, 또는 둘 다에 의해 발생된다.

[0188] 제112의 구현예는 제93 내지 제97, 제107, 제108, 및 제111의 구현예 중 어느 하나의 방법이며, 여기서 디-태그 분자는 기록 태그로부터 기원한 보편적인 프라이밍 부위, 기록 태그로부터 기원한 구획 태그, 기록 태그로부터 기원한 유일한 분자 확인인자, 기록 태그로부터 기원한 임의의 스페이서, 암호화 태그로부터 기원한 암호기 서열, 암호화 태그로부터 기원한 유일한 분자 확인인자, 암호화 태그로부터 기원한 임의의 스페이서, 및 암호화 태그로부터 기원한 보편적인 프라이밍 부위를 포함한다.

[0189] 제113의 구현예는 제93 내지 제112의 구현예 중 어느 하나의 방법이며, 여기서 거대분자 및 관련된 기록 태그는 고체 지지체에 공유결합으로 결합된다.

[0190] 제114의 구현예는 제113의 구현예의 방법이며, 여기서 고체 지지체는 비드, 다공성 비드, 다공성 매트릭스, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 통과 칩, 전자 장치를 변환하는 시그널을 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구이다.

[0191] 제115의 구현예는 제114의 구현예의 방법이며, 여기서 고체 지지체는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드이다.

[0192] 제116의 구현예는 제93 내지 제115의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 폴리펩타이드 또는 단백질이다.

[0193] 제117의 구현예는 제116의 구현예의 방법이며, 여기서 결합제는 변형된 아미노펩타이드, 변형된 아미노 아실 tRNA 신테타제, 변형된 안티칼린, 또는 변형된 ClpS이다.

[0194] 제118의 구현예는 제95 내지 제117의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 단일 아미노산 잔기, 디펩타이드, 트리펩타이드 또는 펩타이드의 해독후 변형(modification)에 결합한다.

[0195] 제119의 구현예는 제118의 구현예의 방법이며, 여기서 결합제는 N-말단 아미노산 잔기, C-말단 아미노산 잔기, 또는 내부 아미노산 잔기에 결합한다.

[0196] 제120의 구현예는 제118의 구현예의 방법이며, 여기서 결합제는 N-말단 펩타이드, C-말단 펩타이드, 또는 내부 펩타이드에 결합한다.

[0197] 제121의 구현예는 제119의 구현예의 방법이며, 여기서 결합제는 N-말단 아미노산 잔기에 결합하고 N-말단 아미노산 잔기는 각각의 결합 주기 후 절단된다.

[0198] 제122의 구현예는 제119의 구현예의 방법이며, 여기서 결합제는 C-말단 아미노산 잔기에 결합하고 C-말단 아미노산 잔기는 각각의 결합 주기 후에 절단된다.

[0199] 제123의 구현예는 제121의 구현예의 방법이며, 여기서 N-말단 아미노산 잔기는 에드만 분해를 통해 절단된다.

[0200] 제124의 구현예는 제93의 구현예의 방법이며, 여기서 결합제는 아미노산의 부위-특이적인 공유결합 표지 또는 해독 후 변형이다.

[0201] 제125의 구현예는 제93 내지 제124의 구현예 중 어느 하나의 방법이며, 여기서 단계 (b) 이후에, 거대분자 및 관련된 결합제를 포함하는 복합체는 고체 지지체로부터 해리되고 소적 또는 미세유동 소적의 유화액내로 분배된다.

[0202] 제126의 구현예는 제125의 구현예의 방법이며, 여기서 각각의 미세유동 소적은 평균적으로, 거대분자 및 결합제를 포함하는 하나의 복합체를 포함한다.

[0203] 제127의 구현예는 제125 또는 제126의 구현예의 방법이며, 여기서 기록 태그는 연장된 암호화 태그 또는 디-태그 작제물의 생성 전에 증폭된다.

[0204] 제128의 구현예는 제125 내지 제127 중 어느 하나의 구현예의 방법이며, 여기서 유화액 융합 PCR을 사용하여 기록 태그 정보를 암호화 태그로 이전시키거나 디-태그 작제물의 집단을 생성한다.

[0205] 제129의 구현예는 제93 내지 제128 중 어느 하나의 구현예의 방법이며, 여기서 연장된 암호화 태그 또는 디-태그 작제물의 수집물은 분석 전에 증폭된다.

[0206] 제130의 구현예는 제93 내지 제129 중 어느 하나의 구현예의 방법이며, 여기서 연장된 암호화 태그 또는 디-태그 작제물의 수집물의 분석은 핵산 서열분석 방법을 포함한다.

[0207] 제131의 구현예는 제130의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석(polony sequencing), 이온 반도체 서열분석, 또는 피로서열분석(pyrosequencing)이다.

[0208] 제132의 구현예는 제130의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 또는 진전된 현미경을 사용한 DNA의 직접적인 영상(imaging)이다.

[0209] 제133의 구현예는 제130의 구현예의 방법이며, 여기서 거대분자의 부분 조성물은 유일한 구획 태그 및 임의로 UMI를 사용하여 다수의 연장된 암호화 태그 또는 디-태그 작제물을 분석함으로써 측정된다.

[0210] 제134의 구현예는 제1 내지 제133의 구현예 중 어느 하나 방법이며, 여기서 분석 단계는 염기당 오차율(per base error rate)이 > 5%, > 10%, > 15%, > 20%, > 25%, 또는 > 30%인 서열분석 방법을 사용하여 수행된다.

[0211] 제135의 구현예는 제1 내지 제134의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그, 기록 태그, 또는 둘 다의 확인 성분은 오차 교정 코드를 포함한다.

[0212] 제136의 구현예는 제135의 구현예의 방법이며, 여기서 확인 성분은 암호기 서열, 바코드, UMI, 구획 태그, 주기 특이적인 서열, 또는 이의 어떠한 조합으로부터 선택된다.

[0213] 제137의 구현예는 제135 또는 제136의 구현예의 방법이며, 여기서 오차 교정 코드는 해밍 코드(Hamming code), 리 거리 코드(Lee distance code), 비대칭 리 거리 코드(symmetric Lee distance code), 리드-솔로몬 코드(Reed-Solomon code), 및 레벤슈타인-테넨골트스 코드(Levenshtein-Tenengolts code)로부터 선택된다.

[0214] 제138의 구현예는 제1 내지 제134의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그, 기록 태그, 또는 둘 다의 확인 성분은 유일한 전류 또는 이온 플럭스(ion flux) 또는 광학 신호를 생성할 수 있으며, 여기서 분석 단계는 확인 성분을 확인하기 위한 유일한 전류 또는 이온 흐름 또는 광학 신호의 검출을 포함한다.

[0215] 제139의 구현예는 제138의 구현예의 방법이며, 여기서 확인 성분은 암호기 서열, 바코드, UMI, 구획 태그, 주기 특이적인 서열, 또는 이의 어떠한 조합으로부터 선택된다.

[0216] 제140의 구현예는:

[0217] (a) 다수의 거대분자 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계;

[0218] (b) 다수의 거대분자를 동종의 거대분자에 결합할 수 있는 다수의 결합제와 접촉시키는 단계(여기서 각각의 결합제는 결합제에 관한 확인 정보를 지닌 암호화 태그를 포함한다);

[0219] (c) 제1의 결합제의 제1의 암호화 태그의 정보를 제1의 거대분자와 관련된 제1의 기록 태그로 이전시켜 제1의 순서의 연장된 기록 태그를 생성시키는 단계(여기서 제1의 결합제는 제1의 거대분자에 결합한다);

[0220] (d) 다수의 거대분자를 동종의 거대분자에 결합할 수 있는 다수의 결합제와 접촉시키는 단계;

[0221] (e) 제2의 결합제의 제2의 암호화 태그의 정보를 제1의 순서의 연장된 기록 태그로 이전시켜 제2의 순서의 연장된 기록 태그를 생성시키는 단계(여기서 제2의 결합제는 제1의 거대분자에 결합한다);

[0222] (f) "n"회 결합 주기를 위해 단계 (d) 내지 (e)를 임의로 반복하는 단계(여기서 제1의 거대분자에 결합하는 각각의 결합제의 각각의 암호화 태그의 정보는 앞서의 결합 주기로부터 생성된 연장된 기록 태그로 이전되어 제1의 거대분자를 나타내는 n번째 순서의 연장된 기록 태그를 생성한다);

[0223] (g) n번째 순서의 연장된 기록 태그를 분석하는 단계를 포함하는, 다수의 거대분자를 분석하는 방법이다.

[0224] 제141의 구현예는 제140의 구현예의 방법이며, 여기서 다수의 거대분자를 나타내는 다수의 n번째 순서의 연장된 기록 태그가 생성되어 분석된다.

[0225] 제142의 구현예는 제140 또는 제141의 구현예의 방법이며, 여기서 거대분자는 단백질이다.

[0226] 제143의 구현예는 제142의 구현예의 방법이며, 여기서 거대분자는 펩타이드이다.

[0227] 제144의 구현예는 제143의 구현예의 방법이며, 여기서 펩타이드는 생물학적 샘플로부터 단백질을 단편화함으로써 수득된다.

[0228] 제145의 구현예는 제140 내지 제144의 구현예 중 어느 하나의 방법이며, 여기서 다수의 거대분자는 다수의, 혼주된 샘플(pooled sample)로부터의 거대분자를 포함한다.

[0229] 제146의 구현예는 제140 내지 제145의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 DNA 분자, RNA 분자, PNA 분자, BNA 분자, XNA, 분자, LNA 분자, γPNA 분자, 또는 이의 조합이다.

[0230] 제147의 구현예는 제140 내지 제146의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 유일한 분자 확인 인자(UMI)를 포함한다.

[0231] 제148의 구현예는 제140 내지 제147의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 구획 태그를 포함한다.

[0232] 제149의 구현예는 제140 내지 제148의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 보편적인 프라이밍 부위를 포함한다.

[0233] 제150의 구현예는 제140 내지 제149의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그는 이의 3'-말단에 스페이서를 포함한다.

[0234] 제151의 구현예는 제140 내지 제150의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 암호기 서열을 포함한다.

[0235] 제152의 구현예는 제140 내지 제151의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 UMI를 포함한다.

[0236] 제153의 구현예는 제140 내지 제152의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 보편적인 프라이밍 부위를 포함한다.

[0237] 제154의 구현예는 제140 내지 제153의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 이의 3'-말단에 스페이서를 포함한다.

[0238] 제155의 구현예는 제140 내지 제154의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 결합 주기 특이적인 서열을 포함한다.

[0239] 제156의 구현예는 제140 내지 제155의 구현예 중 어느 하나의 방법이며, 여기서 암호화 태그는 유일한 분자 확인인자를 포함한다.

[0241] 제157의 구현예는 제140 내지 제156의 구현예 중 어느 하나의 방법이며, 여기서 결합제 및 암호화 태그는 링커에 의해 결합된다.

[0242] 제158의 구현예는 제140 내지 제157의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그의 정보의 암호화 태그로의 이전은 프라이머 연장에 의해 매개된다.

[0243] 제159의 구현예는 제140 내지 제158의 구현예 중 어느 하나의 방법이며, 여기서 기록 태그의 정보의 암호화 태그로의 이전은 연결에 의해 매개된다.

[0244] 제160의 구현예는 제140 내지 제159의 구현예 중 어느 하나의 방법이며, 여기서 다수의 거대분자, 관련된 기록 태그, 또는 둘 다는 고체 지지체에 공유결합으로 결합된다.

[0245] 제161의 구현예는 제140 내지 제160의 구현예 중 어느 하나의 방법이며, 여기서 고체 지지체는 비드, 다공성 비드, 다공성 매트릭스, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 통과 칩, 전자 장치를 변환하는 시그널을 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구이다.

[0246] 제162의 구현예는 제161의 구현예의 방법이며, 여기서 고체 지지체는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드이다.

[0247] 제163의 구현예는 제140 내지 제162의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 폴리펩타이드 또는 단백질이다.

[0249] 제164의 구현예는 제163의 구현예의 방법이며, 여기서 결합제는 변형된 아미노펩티다제, 변형된 아미노 아실 tRNA 신테타제, 변형된 안티칼린, 또는 이의 항체 또는 결합 단편이다.

[0250] 제165의 구현예는 제142 내지 제164의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 단일 아미노산 잔기, 디펩타이드, 트리펩타이드 또는, 펩타이드의 해독후 변형에 결합한다.

[0251] 제166의 구현예는 제165의 구현예의 방법이며, 여기서 결합제는 N-말단 아미노산 잔기, C-말단 아미노산 잔기, 또는 내부 아미노산 잔기에 결합한다.

[0252] 제167의 구현예는 제165의 구현예의 방법이며, 여기서 결합제는 N-말단 펩타이드, C-말단 펩타이드, 또는 내부 펩타이드에 결합한다.

[0253] 제168의 구현예는 제142 내지 제164의 구현예 중 어느 하나의 방법이며, 여기서 결합제는 변형된 N-말단 아미노산 잔기, 변형된 C-말단 아미노산 잔기, 또는 변형된 내부 아미노산 잔기의 화학 표지(chemical label)에 결합된다.

[0254] 제169의 구현예는 제166 또는 제168의 구현예의 방법이며, 여기서 결합제는 N-말단 아미노산 잔기 또는, 변형된 N-말단 아미노산 잔기의 화학 표지에 결합되고, N-말단 아미노산 잔기는 각각의 결합 주기 후 절단된다.

[0255] 제170의 구현예는 제166 또는 제168의 구현예의 방법이며, 여기서 결합제는 C-말단 아미노산 잔기 또는, 변형된 C-말단 아미노산 잔기의 화학 표지에 결합되고, C-말단 아미노산 잔기는 각각의 결합 주기 후 절단된다.

[0256] 제171의 구현예는 제169의 구현예의 방법이며, 여기서 N-말단 아미노산 잔기는 에드만 분해, 에드마나제, 변형된 아미노 펩티다제, 또는 변형된 아실펩타이드 하이드롤라제를 통해 절단된다.

[0257] 제172의 구현예는 제163의 구현예의 방법이며, 여기서 결합제는 아미노산의 부위-특이적인 공유결합 표지 또는 해독 후 변형이다.

[0259] 제173의 구현예는 제140 내지 제172의 구현예 중 어느 하나의 방법이며, 여기서 다수의 n번째 순서의 연장된 기

록 태그는 분석 전에 증폭된다.

[0260] 제174의 구현예는 제140 내지 제173의 구현예 중 어느 하나의 방법이며, 여기서 n번째 순서의 연장된 기록 태그를 분석하는 것은 핵산 서열분석 방법을 포함한다.

[0261] 제175의 구현예는 제174의 구현예의 방법이며, 여기서 다수의 거대분자를 나타내는 다수의 n번째 순서의 연장된 기록 태그가 병렬로 분석된다.

[0262] 제176의 구현예는 제174 또는 제175의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석, 이온 반도체 서열분석, 또는 피로서열분석이다.

[0263] 제177의 구현예는 제174 또는 제175의 구현예의 방법이며, 여기서 핵산 서열분석 방법은 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 또는 진전된 현미경을 사용한 DNA의 직접적인 영상화이다.

도면의 간단한 설명

[0264] 본 발명의 비-제한적인 구현예는 첨부되는 도면을 참고로 실시예의 방식으로 설명될 것이며, 이는 개략적이고, 확장되어야 하는 것으로 의도되지는 않는다. 설명의 목적을 위해서, 모든 성분이 모든 도에서 표지되지는 않고 나타낸 본 발명의 각각의 구현예의 모든 성분은 아니며, 여기서 설명은 당해 분야의 통상의 기술자가 본 발명을 이해하도록 하는데 필수적이지는 않다.

도 1a 내지 1b: 도 1a는 도에 나타낸 기능적 요소에 대한 핵심을 나타낸다. **도 1b**는 단백질 코드를 DNA 코드로 변환하는 일반적인 개요를 나타내며, 여기서 다수의 단백질 또는 폴리펩타이드는 다수의 펩타이드로 단편화되고, 이는 이후에 연장된 기록 태그의 라이브러리(library)로 전환된다. 연장된 기록 태그는 펩타이드 서열을 나타내는 DNA 암호화된 라이브러리를 구성한다. 라이브러리는 어떠한 차세대 서열분석(NGS) 플랫폼 상의 서열로 적절하게 변형된다.

도 **2a 내지 2d**는 단일 또는 다수의 기록 태그로 동시-국재화되거나 동시-표지된 고정된 단백질과 상호작용하는 암호화 태그를 포함하는 다수의 주기의 결합제(예컨대, 항체, 안티칼린, N-레코그닌 단백질(예컨대, TP-의존적인 Clp 프로테아제 어댑터 단백질(ClpS)), 아프타머 등 및 이의 변이체/동족체)을 사용하는, 본원에 개시된 방법에 따른 단백질 거대분자 분석의 예를 나타낸다. 기록 태그는 보편적인 프라이밍 부위, 바코드(예컨대, 분배 바코드, 구획 바코드, 분획 바코드), 임의의 유일한 분자 확인인자(UMI) 서열, 및 암호화 태그의 정보 이전에 사용된 스페이서 서열(Sp)로 구성된다. 스페이서 서열(Sp)은 모든 결합 주기에 걸쳐 일정할 수 있으며, 결합제 특이적이거나 결합 주기 수 특이적일 수 있다. 암호화 태그는 결합제에 대한 확인 정보를 제공하는 암호기 서열, 임의의 UMI, 및 기록 태그 상의 상보성 스페이서 서열에 하이브리드화하는 스페이서 서열로 구성되어, 암호화 태그 정보의 기록 태그로의 이전을 촉진한다(예컨대, 본원에서 폴리머라제 연장으로 지칭되는 프라이머 연장). 도 **2a**는 단백질에 대한 동종의 결합제의 사이클릭 결합을 통해 연장된 기록 태그를 생성하는 과정 및 결합제의 암호화 태그로부터 단백질의 기록 태그로의 상응하는 정보 이전을 나타낸다. 일련의 순차적인 결합 및 암호화 태그 정보 이전 단계 후, 결합제(예컨대, 항체 1(Ab1), 항체 2(Ab2), 항체 3(Ab3),, ... 항체 "n" (Ab n))에 대한 확인 정보를 제공하는 "n"번째 결합 주기로부터의 암호기 서열, 기록 태그로부터의 바코드/임의의 UMI 서열, 결합제의 암호화 태그로부터의 임의의 UMI 서열, 및 증폭 및 디지탈 차-세대 서열분석에 의한 분석을 촉진하는 라이브러리 작제물의 각각의 말단에서 플랭킹 보편적인 프라이밍 서열(flanking universal priminng sequence)을 포함하는 태그 정보를 암호화하는 결합제를 함유하는, 최종적인 연장된 기록 태그가 생산된다. 도 **2b**는 DNA 바코드화된 기록 태그를 지닌 단백질을 표지하기 위한 개략도의 예를 나타낸다. 상부 패널에서, N-하이드록시석신이미드(NHS)는 아민 반응성 커플링제이며, 디벤조사이클로옥틸(DBCO)은 고체 기질(substrate)의 표면에 대한 "클릭(click)" 커플링에 유용한 가닥화된 알킨이다. 이러한 개략도에서, 기록 태그는 NHS 모이어티를 통해 단백질의 라이신(K) 잔기(및 임의로 N-말단 아미노산)의 ε 아민에 커플링된다. 하단 패널에서, 헤테로이기능성 링커인, NHS-알킨은 라이신(K) 잔기의 ε 아민을 표지시켜 알킨 "클릭" 모이어티를 생성하는데 사용된다. 아지드-표지된 DNA 기록 태그는 이후에 표준 클릭 화학을 통해 이들 반응성 알킨 그룹에 용이하게 부착될 수 있다. 더욱이, DNA 기록 태그는 또한 역 iEDDA 반응을 통해 TCO-유도체화된 서열분석 기질에 하부(downstream) 커플링하기 위한 직각 메틸테트라진(mTet) 모이어티로 설계될 수 있다. 도 **2c**는 기록 태그를 사용한 단백질 분석 방법의 2가지 예를 나타낸다. 상단 패널에서, 단백질 거대분자는 포획제(capture agent)를 통해 고체 지지체에 고정되고 임의로 가교-결합된다. 단백질 또는 포획제는 기록 태그로 표지될 수 있다. 하단 패널에서, 관련된 기록 태그를 지닌 단백질은 고체 지지체에 직접 고정된다. 도 **2d**는 동종의 결합제를 암호화하는

DNA를 사용한 단순한 단백질 면역검정 및 수득되는 연장된 기록 태그의 서열분석에 대한 전체적인 작업흐름의 예를 나타낸다. 단백질은 기록 태그를 통해 바코드화되고(즉, 색인이 달리고) 사이클릭 결합 분석 이전에 혼주되어, 샘플 처리량을 크게 증가시키고 결합 시약을 절약할 수 있다. 이러한 시도는 효과적으로 디지탈적이고, 단순하며 역상 단백질 검정(RPPA)을 수행하기 위한 보다 확장가능한 시도이다.

도 3a 내지 3d는 펩타이드 서열을 나타내는 DNA 연장된 기록 태그의 작제에 의한 분해-기반의 펩타이드 서열분석 검정을 위한 과정을 나타낸다. 이는 N-말단 아미노산(NTAA) 결합, 펩타이드에 부착된 기록 태그에 대한 암호화 태그 정보 전달, NTAA 절단, 및 고체 지지체 상에서 모두, 사이클릭 방식으로 과정의 반복의 사이클릭 과정을 사용하는 에드만 분해-유사 시도를 통해 달성된다. 펩타이드의 N-말단 분해로부터의 연장된 기록 태그의 예시적인 작제의 개관이 제공된다: (A) 펩타이드의 N-말단 아미노산을 표지한다(예컨대, 페닐티오카바모일(PTC), 디니트로페닐(DNP), 설포닐 니트로페닐(SNP), 아세틸, 또는 구아니딜 모이어티 사용); (B) 결합제 및 표지된 NTAA에 결합된 관련된 암호화 태그를 나타낸다; (C) 고체 지지체(예컨대, 비드)에 결합되고 기록 태그(예컨대, 삼기능성 링커를 통해)와 관련된 펩타이드를 나타내며, 여기서 펩타이드의 NTAA에 대한 결합제의 결합시, 암호화 태그의 정보는 기록 태그(예컨대, 프라이머 연장을 통해)로 이전되어 연장된 기록 태그를 생성한다; (D) 표지된 NTAA를 화학적 또는 효소적 수단으로 절단하여 새로운 NTAA를 노출시킨다. 화살표로 나타낸 바와 같이, 주기를 "n"회 반복하여 최종적인 연장된 기록 태그를 생성한다. 최종의 연장된 기록 태그는 보편적인 프라이밍 부위에 의해 임의로 플랭킹(flanking)되어 하부 증폭 및 DNA 서열분석을 촉진한다. 전방(전방)의 보편적인 프라이밍 부위(예컨대, 일루미나(Illumina)의 P5-S1의 서열)는 원래의 기록 태그 설계의 부분일 수 있으며 역 보편적인 프라이밍 부위(예컨대, 일루미나의 P7-S2의 서열)를 기록 태그의 연장시 최종 단계로서 가할 수 있다. 이러한 최종 단계는 결합제와는 별도로 수행할 수 있다.

도 4a 내지 4b는 본원에 개시된 방법에 따른 예시적인 단백질 서열분석 작업흐름을 나타낸다. 도 4a는 화살표로 연결되는 박스로 나타낸 특수한 구현예를 지닌, 담회색 파선으로 윤곽을 나타낸 대안적 방식을 지닌 예시적인 작업흐름을 나타낸다. 작업흐름의 각각의 단계에 대한 대안적인 방식은 화살표로 박스 아래에 나타낸다. 도 4b는 정보 이전의 효능을 증진시키기 위한 사이클릭 결합 및 암호화 태그 정보 이전 단계를 수행하는데 있어서의 선택사항을 나타낸다. 분자당 다수의 기록 태그를 사용할 수 있다. 더욱이, 제공된 결합 현상을 위하여, 기록 태그로의 암호화 태그 정보의 이전은 수회 수행될 수 있거나, 대안적으로는, 표면 증폭 단계를 사용하여 연장된 기록 태그 라이브러리 등의 카피를 생성할 수 있다.

도 5의 A 내지 B는 거대분자(예컨대, 펩타이드)와 관련된 기록 태그로 결합제의 암호화 태그의 확인 정보를 이전시켜 연장된 기록 태그를 생성시키기 위한 프라이머 연장을 사용하는 연장된 기록 태그의 예시적인 작제의 개관을 나타낸다. 결합제에 관한 확인 정보를 지닌 유일한 암호기 서열을 포함하는 암호화 태그는 일반적인 스페이서 서열(Sp')에 의해 각각의 말단에서 임의로 플랭킹된다. 도 5의 A는 비드에 연결된 기록-태그 표지된 펩타이드의 NTAA에 대한 암호화 태그 결합을 포함하는 NTAA 결합제를 나타낸다. 기록 태그는 상보성 스페이서 서열(Sp)을 통해 암호화 태그로 어닐링되며 프라이머 연장 반응은 프라이밍 부위로서 스페이서(Sp)를 사용하여 기록 태그로 암호화 태그 정보의 이전을 매개한다. 암호화 태그는 결합제와 먼쪽의 말단에서 단일 가닥 스페이서(Sp')를 지닌 듀플렉스(duplex)로서 나타낸다. 이러한 구조는 기록 태그내 내부 부위에 대한 암호화 태그의 하이브리드화를 최소화시키며 암호화 태그의 단일 가닥 스페이서 오버행(overhang)(Sp')을 지닌 기록 태그의 말단 스페이서(sp) 서열의 하이브리드화를 선호한다. 더욱이, 연장된 기록 태그는 올리고뉴클레오타이드(암호기에 대해 상보성임, 스페이서 서열)로 예비-어닐링되어 내부 기록 태그 서열 요소에 대한 암호화 태그의 하이브리드화를 차단할 수 있다. 도 5b는 결합의 "n" 주기 후 생산된 최종의 연장된 기록 태그("***"는 연장된 기록 태그에 나타나지 않은 개입된 결합 주기(intervening binding cycle)를 나타낸다) 및 암호화 태그 정보의 이전 및 3'-말단에서 보편적인 프라이밍 부위의 첨가를 나타낸다.

도 6은 효소 연결을 통해 연장된 기록 태그로 이전된 암호화 태그 정보를 나타낸다. 2개의 상이한 거대분자를 이들 각각의 기록 태그와 함께 나타내며, 기록 태그 연장은 병렬로 진행된다. 연결은 이중 가닥 암호화 태그를 설계하여 스페이서 서열(Sp)이 기록 태그 상에 상보성 스페이서(Sp')와 어닐링하는 "점성 말단(sticky end)" 오버행을 가지도록 함으로써 촉진시킬 수 있다. 이중 가닥 암호화 태그의 상보성 쇄는 정보를 기록 태그로 이전한다. 연결을 사용하여 기록 태그를 연장시키는 경우, 연장 방향은 나타낸 바와 같은 5' 내지 3'일 수 있거나 임의로 3' 내지 5'일 수 있다.

도 7은 스페이서 서열을 연장된 기록 태그내로 삽입시키지 않고 기록 태그 또는 연장된 기록 태그의 3' 뉴클레오타이드를 암호화 태그(또는 이의 상보체)의 5' 뉴클레오타이드에 연결시키기 위해 화학적 연결을 통해 기록 태그로 암호화 태그 정보를 이전시키는 "스페이서가 없는(spacer-less)" 시도를 나타낸다. 연장된 기록 태그 및

암호화 태그의 배향은 또한 역전되어 기록 태그의 5' 말단이 암호화 태그(또는 상보체)의 3' 말단에 연결되도록 할 수 있다. 나타낸 실시예에서, 기록 태그("기록 헬퍼") 상의 상보성 "헬퍼" 올리고뉴클레오타이드 서열과 암호화 태그 사이의 하이브리드화를 사용하여 복합체를 안정화시켜 암호화 태그 상보성 가닥에 대한 기록 태그의 특이적인 화학적 연결이 가능하도록 할 수 있다. 수득되는 연장된 기록 태그는 스페이서 서열을 포함하지 않는다. 또한 DNA, PNA, 또는 유사한 핵산 중합체를 사용할 수 있는 화학적 연결(예컨대, 아지드 및 알킨 모이어티(삼중 선 기호로 나타냄)를 사용)의 "클릭 화학" 버전이 나타나 있다.

도 **8**의 **A** 내지 **B**는 N-말단 아미노산 분해 이전에 연장된 기록 태그내로 펩타이드의 해독 후 변형(PTM)의 기록의 예시적인 방법을 나타낸다. 도 **8**의 **A**: 결합제(예컨대, 포스포타이로신 항체에 대한 확인 정보를 지닌 암호화 태그를 포함하는 포스포타이로신 항체)에 관한 확인 정보를 지닌 암호화 태그를 포함하는 결합제는 펩타이드에 결합시킬 수 있다. 포스포타이로신이 나타낸 바와 같이, 기록 태그-표지된 펩타이드에 존재하는 경우, 포스포타이로신에 대한 포스포타이로신 항체의 결합시, 암호화 태그 및 기록 태그는 상보성 스페이서 서열을 통해 어닐링되어 암호화 태그 정보가 기록 태그로 이전됨으로써 연장된 기록 태그를 생성한다. 도 **8**의 **B**: 연장된 기록 태그는 주요한 아미노산 서열(예컨대, "aa_1", "aa_2", "aa_3", ..., "aa_N") 및 펩타이드의 해독-후 변형(예컨대, "PTM_1", "PTM_2") 둘 다에 대한 암호화 태그 정보를 포함할 수 있다.

도 **9**의 **A** 내지 **B**는 결합제를 거대분자에 결합시키고 결합제에 부착된 암호화 태그의 정보를 고체 지지체(예컨대, 비드)에 부착된 단일의 거대분자의 부위에 동시-국재화된 다수의 기록 태그 중에서 개개 기록 태그로 이전시킴으로써, 거대분자를 총괄적으로 나타내는 다수의 연장된 암호화 태그를 생성하는 다수 주기의 과정을 나타낸다. 이러한 도에서, 단지 예의 목적으로, 거대분자는 펩타이드이고 각각의 주기는 결합제를 N-말단 아미노산(NTAA)에 결합시키고, 암호화 태그 정보를 기록 태그에 이전시킨 후, NTAA를 제거하여 새로운 NTAA를 노출시킴에 의한 결합 사건을 기록함을 포함한다. 도 **9**의 **A**는 거대분자를 지닌 고체 지지체 상에 동시-국재화된 다수의 기록 태그(보편적인 전방 프라이밍 서열 및 UMI 포함)를 나타낸다. 개개의 기록 태그는 결합제의 암호화 태그 내 일반적인 스페이서 서열에 대해 상보성인 일반적인 스페이서 서열(Sp)을 지니며, 이는 암호화 태그 정보를 기록 태그로 이전시키기 위해 연장 반응물을 프라이밍하는데 사용될 수 있다. 도 **9**의 **B**는 결합의 각각의 연속적인 주기에 사용된 주기-특이적인 NTAA 결합제의 상이한 혼주물을 나타내며, 각각의 혼주물은 특이적인 스페이서 서열을 갖는다.

도 **10**의 **A** 내지 **C**는 결합제에 부착된 암호화 태그의 정보를 고체 지지체(예컨대, 비드)에 부착된 단일의 거대분자의 부위에 동시-국재화된 다수의 기록 태그 중에서 하나의 기록 태그로 이전시킴으로써, 거대분자를 총괄적으로 나타내는 다수의 연장된 암호화 태그를 생성하는 다수 주기를 포함하는 예시적인 방식을 나타낸다. 이러한 도에서, 단지 예의 목적으로, 거대분자는 펩타이드이고 각각의 주기는 NTAA에 대한 결합, 결합 사건의 기록에 이은, NTAA를 제거하여 새로운 NTAA를 노출시킴을 포함한다. 도 **10**의 **A**는 거대분자를 지닌 고체 지지체 상에 동시-국재화된 다수의 기록 태그(보편적인 전방 프라이밍 서열 및 UMI 포함), 바람직하게는 비드당 단일 분자를 나타낸다. 개개의 기록 태그는 상이한 "주기 특이적인" 서열(예컨대, C_1, C_2, C_3, ... C_n)를 지닌 이들의 3'-말단에서 상이한 스페이서 서열을 지닌다. 바람직하게는 각각의 비드에서 기록 태그는 동일한 UMI 서열을 공유한다. 결합의 제1의 주기(주기 1)에서, 다수의 NTAA 결합제는 거대분자와 접촉된다. 주기 1에 사용된 결합제는 기록 태그의 주기 1 C_1 스페이서 서열에 대해 상보성인 공통의 5'-스페이서 서열(C'1)을 지닌다. 주기 1에 사용된 결합제는 또한 주기 2 스페이서 C_2에 대해 상보성인 3'-스페이서 서열(C'_2)을 지닌다. 결합 주기 1 동안에, 제1의 NTAA 결합제는 거대분자의 유리된 N-말단에 결합하며, 제1의 암호화 태그의 정보는 상보성 C'_1 스페이서 서열에 하이브리드화된 C_1 서열로부터의 프라이머 연장을 통해 동종의 기록 태그로 이전된다. NTAA를 제거하여 새로운 NTAA를 노출시킨 후, 결합 주기 2는 주기 1 결합제의 3'-스페이서 서열과 동일한 주기 2 5'-스페이서 서열(C'_2) 및 공통의 주기 3 3'-스페이서 서열을 지닌 다수의 NTAA 결합제를 거대분자와 접촉시킨다. 제2의 NTAA 결합제는 거대분자의 NTAA에 결합하고, 제2의 암호화 태그의 정보는 상보성 C_2 및 C'_2 스페이서 서열로부터의 프라이머 연장을 통해 동종의 기록 태그로 이전된다. 이들 주기는 "n"회 결합 주기까지 반복되며, 여기서 마지막 연장된 기록 태그는 보편적인 역 프라이밍 서열로 캡핑(capping)되어, 단일의 거대분자와 함께 동시-국재화된 다수의 연장된 기록 태그를 형성하며, 여기서 각각의 연장된 기록 태그는 하나의 결합 주기로부터의 암호화 태그 정보를 지닌다. 각각의 연속적인 결합 주기에 사용된 결합제의 각각의 세트는 암호화 태그내에 주기 특이적인 스페이서 서열을 지니므로, 결합 주기 정보는 수득되는 연장된 기록 태그에서 결합제 정보와 관련될 수 있다. 도 **10**의 **B**는 각각의 연속적인 주기의 결합에 사용된 주기-특이적인 결합제의 상이한 혼주물을 나타내며, 각각의

혼주물은 주기 특이적인 스페이서 서열을 갖는다. 도 10의 C는 거대분자의 부위에서 동시-국재화된 연장된 기록 태그의 수집물이 주기 특이적인 스페이서 서열을 사용한 연장된 기록 태그의 PCR 조립을 기반으로 순차적인 순서로 조립되어, 거대분자의 정돈된 서열을 제공할 수 있는 방법을 나타낸다. 바람직한 방식에서, 각각의 연장된 기록 태그의 다수의 카피는 콘카테네이션(concatenation) 이전에 증폭을 통해 생성된다.

도 11의 A 및 B는 기록 태그로부터 암호화 태그 또는 디-태그 작제물로의 정보 이전을 나타낸다. 기록 결합 정보의 2가지 방법은 (A) 및 (B)에 나타낸다. 결합제는 본원에 기술된 바와 같은 어떠한 유형의 결합제일 수 있다; 항-포스포타이로신 결합제는 예시 목적으로만 나타낸다. 연장된 암호화 태그 또는 디-태그 작제를 위해, 암호화 태그로부터 기록 태그로의 결합 정보의 이전보다는, 정보가 기록 태그로부터 암호화 태그로 이전되어 연장된 암호화 태그(A)를 생성하거나, 정보가 기록 태그 및 암호화 태그 둘 다로부터 제3의 디-태그-형성 작제물(B)로 이전된다. 디-태그 및 연장된 암호화 태그는 기록 태그(바코드, 임의의 UMI 서열, 및 임의의 구획 태그(CT) 서열(나타내지 않음)을 포함함) 및 암호화 태그를 포함한다. 디-태그 및 연장된 암호화 태그는 기록 태그로부터 용출되어, 수집되고, 임의로 증폭되며 차세대 서열분석기에서 판독될 수 있다.

도 12의 A 내지 D는 결합 사건의 PNA 조합 바코드/UMI 기록 태그 및 디-태그 검출의 설계를 나타낸다. 도 12의 A에서 4개의 요소 PNA 단어 서열(A,'-B, B'-C, 및 C')의 화학적 연결을 통한 조합 PNA 바코드/UMI의 작제가 나열되어 있다. DNA 아암(arm)의 하이브리드화는 PNA 바코드/UMI의 조합적 조립을 위한 스페이서가 거의 없는 조합적 주형(spacer-less combinatorial template)을 생성하기 위해 포함된다. 화학적 연결을 사용하여 어닐링된 PNA "단어"를 함께 스티치(stitch)하는데 사용된다. 도 12의 B는 기록 태그의 PNA 정보를 DNA 중간체로 이전시키기 위한 방법을 나타낸다. DNA 중간체는 정보를 암호화 태그로 이전시킬 수 있다. 즉, 상보성 DNA 단어 서열을 PNA에 어닐링하고 화학적으로 연결할 수 있다(PNA 주형을 사용하는 리가제가 발견된 경우 임의로 효소적으로 연결함). 도 12의 C에서, DNA 중간체는 스페이서 서열, Sp를 통해 암호화 태그와 상호작용하도록 설계된다. 가닥-치환 프라이머 연장 단계는 연결된 DNA를 치환하고 DNA 중간체로부터의 기록 태그 정보를 암호화 태그로 이전시켜 연장된 암호화 태그를 생성한다. 터미네이터 뉴클레오타이드(terminator nucleotide)를 DNA 중간체의 말단에 포함시켜 프라이머 연장을 통한 암호화 태그 정보의 DNA 중간체로의 이전을 방지할 수 있다. 도 12의 D: 대안적으로, 정보를 암호화 태그로부터 DNA 중간체로 이전시켜 디-태그(di-tag) 작제물을 생성할 수 있다. 터미네이터 뉴클레오타이드는 암호화 태그의 말단 내로 혼입되어 DNA 중간체로부터의 기록 태그 정보의 암호화 태그로의 이전을 방지할 수 있다.

도 13의 A 내지 E는 구획 바코드화된 비드 상의 프로테옴 분배, 및 펩타이드 서열 조성을 나타내는 요소의 라이브러리를 생성하기 위한 유화액 융합 PCR을 통한 후속적인 디-태그 조립으로 나타낸다. 펩타이드의 아미노산 성분은 N-말단 서열분석을 통해 또는 대안적으로 아미노산 특이적인 화학적 표지 또는 암호화 태그와 관련된 결합제의 부착(공유결합성 또는 비-공유결합성)을 통해 후속적으로 특성화할 수 있다. 암호화 태그는 보편적인 프라이밍 서열, 및 아미노산 동일성에 대한 암호기 서열, 구획 태그, 및 아미노산 UMI로 구성된다. 정보 이전 후, 디태그(ditag)를 기록 태그 UMI를 통해 원래의 분자로 다시 맵핑(mapping)한다. 도 13의 A에서, 프로테옴은 바코드화된 비드를 지닌 소적으로 구획화된다. 관련된 기록 태그(구획 바코드 정보를 포함)를 지닌 펩타이드는 비드 표면에 부착된다. 소적 유화액은 분배된 펩타이드를 지닌 바코드화된 비드를 방출시킴으로써 깨뜨린다. 도 13의 B에서, 펩타이드 상의 특이적인 아미노산 잔기는 부위-특이적인 표지화 모이어티에 접합된 DNA 암호화 태그로 화학적으로 표지된다. DNA 암호화 태그는 아미노산 바코드 정보 및 임의로 아미노산 UMI를 포함한다. 도 13의 C: 표지된 펩타이드-기록 태그 복합체가 비드로부터 방출된다. 도 13의 D: 표지된 펩타이드-기록 태그 복합체는 구획당 평균적으로, 1개 미만의 펩타이드-기록 태그 복합체가 존재하도록 나노 또는 마이크로유화액으로 유화된다. 도 13의 E: 유화액 융합 PCR은 기록 태그 정보(예컨대, 구획 바코드)를 아미노산 잔기에 부착된 DNA 암호화 태그 모두로 이전시킨다.

도 14는 유화된 펩타이드 기록 태그 - 암호화 태그 복합체로부터 연장된 암호화 태그의 생성을 나타낸다. 도 13의 C로부터의 펩타이드 복합체는 PCR 시약과 함께 소적당 평균 1개의 펩타이드 복합체를 지닌 소적으로 동시-유화한다. 3개의-프라이머 융합 PCR 시도를 사용하여 펩타이드와 관련된 기록 태그를 증폭시키고, 증폭된 기록 태그를 다수의 결합제 암호화 태그 또는 공유결합으로 표지된 아미노산의 암호화 태그에 융합시키고, 암호화 태그를 프라이머 연장을 통해 연장시켜 기록 태그로부터의 펩타이드 UMI 및 구획 태그 정보를 암호화 태그로 이전시키고, 수득되는 연장된 암호화 태그를 증폭시킨다. 존재하는 각각의 아미노산 암호기 서열-UMI 암호화 태그에 대한 상이한 종과 함께 다수의 연장된 암호화 태그 종이 존재한다. 이러한 방식으로, 펩타이드내 아미노산의 동일성 및 수(count) 둘 다가 측정될 수 있다. U1 보편적인 프라이머 및 Sp 프라이머는 $U2_{tr}$ 보편적인 프라이머보다는 보다 높은 융점의 Tm을 가지도록 설계된다. 이는 2-단계 PCR로 가능하며 여기서 제1의 아주 적은 주기가

보다 높은 어닐링 온도에서 수행되어 기록 태그를 증폭시킨 후, 보다 낮은 Tm으로 진전됨으로써 PCR 동안 기록 태그 및 암호화 태그는 서로 프라이밍되어 연장된 암호화 태그를 생산하도록 하고, U1 및 $U2_{tr}$ 보편적인 프라이머를 사용하여 수득되는 연장된 암호화 태그 생성물의 증폭을 프라이밍한다. 특정의 구현예에서, $U2_{tr}$ 프라이머로부터의 조기성숙한 폴리머라제 연장은 광-불안정성 3' 차단 그룹을 사용함으로써 방지할 수 있다(Young et al., 2008, Chem. Commun. (Camb) 4:462-464). 기록 태그를 증폭시키는 제1의 라운드의 PCR, 및 암호화 태그 Sp_{tr}이 기록 태그의 증폭된 Sp' 서열에서 암호화 태그의 연장을 프라이밍하는 제2의 라운드의 융합 PCR 단계 후에, $U2_{tr}$의 3' 차단 그룹이 제거되며, 보다 높은 온도의 PCR이 U1 및 $U2_{tr}$ 프라이머를 사용한 연장된 암호화 태그의 증폭을 위해 개시된다.

도 15는 단백질의 맵핑가능성(mappability) 및 상화(phasing)를 촉진하는 프로테옴 분배 및 바코딩의 용도를 나타낸다. 펩타이드 서열분석에서, 단백질은 전형적으로 펩타이드로 분해된다. 이러한 과정에서 모 단백질 분자로부터 기원한 개개 펩타이드와 모 단백질 분자에 대한 이들의 관련성 사이에 관한 정보가 손실된다. 이러한 정보를 재구성하기 위하여, 개개 펩타이드 서열을 이들이 기원할 수 있는 단백질 서열의 수집물로 다시 맵핑한다. 이러한 세트내에서 유일한 매치(match)를 발견하는 업무는 짧고/짧거나 부분적인 펩타이드 서열의 사용, 및 수집물의 크기 및 복잡성(예컨대, 프로테오옴 서열 복잡성)이 증가하므로 보다 더 어렵게 된다. 프로테옴의 바코드화된(예컨대, 구획 태그된) 구획 또는 분할(분배)로의 분배, 및 구획 태그의 펩타이드로의 후속적인 분해, 및 펩타이드에 대한 구획 태그의 결합(joining)은 펩타이드 서열이 맵핑될 필요가 있는 "단백질" 공간을 감소시켜, 복잡한 단백질 샘플의 경우에 업무를 크게 단순화시킨다. 펩타이드로의 분해 전에 유일한 분자 확인인자(UMI)를 사용한 단백질의 표지화는 원래의 단백질 분자로 다시 펩타이드의 맵핑을 촉진하여 동일한 단백질 분자로부터 기원한 해독후 변형된(PTM) 변이체와 개개 프로테오폼(proteoform)의 확인 사이의 상 정보의 주석(annotation)을 허용한다. **도 15**의 **A**는 단백질을 분배 바코드를 포함하는 기록 태그로 표지하는 것을 포함하는 프로테옴 분배 및 기록-태그 표지된 펩타이드로의 후속적인 분획화의 예를 나타낸다. **도 15**의 **B:** 부분적인 펩타이드 서열 정보 또는 심지어 조성 정보를 위해, 이러한 맵핑은 매우 변질될 수 있다. 그러나, 부분적인 펩타이드 서열 또는 동일한 단백질로부터의 다수의 펩타이드로부터의 정보와 커플링된 조성 정보는 원래의 단백질 분자를 유일하게 확인하도록 한다.

도 16은 구획 태그된 비드 서열 설계의 예시적인 방식을 나타낸다. 구획 태그는 개개 구획을 확인하기 위한 X_{5-20}의 바코드 및 구획 태그가 결합된 펩타이드를 확인하기 위한 N_{5-10}의 유일한 분자 확인인자(UMI)를 포함하며, 여기서 X 및 N은 변성된 핵염기 또는 핵염기 단어를 나타낸다. 구획 태그는 단일 가닥(상단에 묘사)이거나 이중 가닥(하단에 묘사)일 수 있다. 임의로, 구획 태그는 목적한 펩타이드(좌측에 묘사)에 결합시키기 위한 단백질 리가제(예컨대, 부텔라제 I)에 대한 인식 서열을 지닌 펩타이드 서열을 포함하는 키메라 분자일 수 있다. 대안적으로, 화학적 모이어티는 목적한 펩타이드에 대한 커플링을 위해 구획 태그에 포함될 수 있다(예컨대, 우측 묘사에 나타낸 바와 같은 아지드).

도 17의 **A 및 B**는 다음을 나타낸다: **(A)** 다수의 펩타이드를 나타내는 다수의 연장된 기록 태그; 및 **(B)** 표준 하이브리드 포획 기술을 통한 표적 펩타이드 농축의 예시 방법. 예를 들면, 하이브리드 포획 농축은 펩타이드의 라이브러리를 나타내는 연장된 기록 태그의 라이브러리로부터 목적한 하나 이상의 펩타이드("표적 펩타이드")를 나타내는 연장된 기록 태그에 하이브리드화하는 하나 이상의 바이오티닐화된 "미끼(bait)" 올리고뉴클레오타이드를 사용할 수 있다. 미끼 올리고뉴클레오타이드:표적 연장된 기록 태그 하이브리드화 쌍은 하이브리드화 후 바이오틴 태그를 통해 용액으로부터 끌어당겨서 펩타이드 또는 목적한 펩타이드를 나타내는 연장된 기록 태그의 농축된 분획을 생성한다. 연장된 기록 태그의 분리("끌어당김")는 예를 들면, 스트렙타비딘-코팅된 자기 비드를 사용하여 달성할 수 있다. 바이오틴 모이어티는 비드 상에서 스트렙타비딘에 결합하며, 분리는 용액이 제거되거나 교환되는 동안 자석을 사용하여 비드를 국재화시킴으로써 달성된다. 바람직하지 않거나 과도하게 풍부한 펩타이드를 나타내는 연장된 기록 태그에 경쟁적으로 하이브리드화하는 비-바이오티닐화된 경쟁인자 농축된 올리고뉴클레오타이드는 하이브리드 포획 검정의 하이브리드화 단계에 임으로 포함시켜 농축된 표적 펩타이드의 양을 조절할 수 있다. 비-바이오티닐화된 경쟁인자 올리고뉴클레오타이드는 표적 펩타이드에 대한 하이브리드화에 대해 경쟁하지만, 하이브리드화 듀플렉스는 바이오틴 모이어티의 부재로 인하여 포획 단계 동안에 포획되지 않는다. 따라서, 농축된 연장된 기록 태그 분획은 경쟁인자 올리고뉴클레오타이드 대 바이오티닐화된 "미끼" 올리고뉴클레오타이드의 비를 큰 역학적 범위에 걸쳐 조정함으로써 조절할 수 있다. 이러한 단계는 샘플 내에서 단백질이 풍부한 역학적 범위 논쟁에 촛점을 맞추기 위해 중요할 것이다.

도 18의 **A** 및 **B**는 단일 세포 및 대량의 프로테옴을 개개 소적으로 분배하는 예시적인 방법을 나타내며, 각각의

소적은 펩타이드를 이들의 원래의 단백질 복합체, 또는 단일 세포로부터 기원하는 단백질과 관련시키기 위한 이에 부착된 다수의 구획 태그를 갖는 비드를 포함한다. 구획 태그는 바코드를 포함한다. 소적 형성 후 소적 성분의 조작: **(A)** 개개 소적으로 단일 세포 분배에 이은 세포 용해로 세포 프로테옴의 방출, 및 세포 프로테옴을 펩타이드로 분해하기 위한 단백질분해, 및 충분한 단백질분해 후 프로테아제의 불활성화; **(B)** 다수의 소적으로 다량의 프로테옴 분배(여기서 개개 소적은 단백질 복합체를 포함하며 이후 단백질 분해에 의해 단백질 복합체는 펩타이드로 분해되고, 충분한 단백질분해 후 프로테아제를 불활성화시킨다). 열-불안정성 금속-프로테아제를 사용하여 프로테아제를 활성화하기 위한 광-케이지된(photo-caged) 2가 양이온의 광-방출 후 캡슐화된 단백질을 펩타이드로 분해할 수 있다. 프로테아제는 열 불활성화된 후 충분히 단백질분해될 수 있거나, 2가 양이온이 킬레이트화될 수 있다. 소적은 펩타이드의 N- 또는 C-말단 아미노산에 연결될 수 있는 핵산 바코드(기록 태그로부터 분리됨)를 포함하는 하이브리드하되거나 방출가능한 구획 태그를 함유한다.

도 19의 **A** 및 **B**는 개개 소적으로의 단일 세포 및 다량의 프로테오믹스의 분배를 나타내며, 각각의 소적은 이에 부착된 구획 태그를 지닌 다수의 이기능성 기록 태그를 갖는 비드를 포함함으로써 펩타이드를 이들의 원래의 단백질 또는 단백질 복합체, 또는 단백질을 원래의 단일 세포에 관련시킨다. 소적 형성 후 소적 성분의 조작: **(A)** 개개 소적으로의 단일 세포 분배에 이은 세포 분해로 세포 프로테옴 방출, 및 단백질분해로 세포 프로테옴을 펩타이드로 분해, 및 충분한 단백질분해 후 프로테아제의 불활성화; **(B)** 다수의 소적으로 다량의 프로테옴 분배(여기서 개개 소적은 단백질 복합체를 포함하며 이후 단백질분해에 의해 단백질 복합체를 펩타이드로 분해하고, 충분한 단백질분해 후 프로테아제를 불활성화시킨다). 열 불안정성 금속-프로테아제를 사용하여 광-케이지된 2가 양이온(예컨대, Zn2+)의 광-방출 후 캡슐화된 단백질을 펩타이드로 분해할 수 있다.). 프로테아제는 충분한 단백질분해 후 열-불활성화시키거나 2가 양이온을 킬레이트화할 수 있다. 소적은 펩타이드의 N- 또는 C-말단 아미노산에 연결될 수 있는 핵산 바코드(기록 태그로부터 분리됨)를 포함하는 하이브리드화되거나 방출될 수 있는 구획 태그를 함유한다.

도 20의 A 내지 **L**은 펩타이드에 부착된 구획 바코드화된 기록 태그의 생성을 나타낸다. 구획 바코드화 기술(예컨대, 미세유동성 소적 속의 바코드화된 비드 등)을 사용하여 구획-특이적인 바코드를 특수한 구획내에 캡슐화된 분자 성분으로 이전할 수 있다. **(A)** 특수한 구현예에서, 단백질 분자는 변성되고, 라이신 잔기(K)의 ε-아미노 그룹은 활성화된 보편적인 DNA 태그 분자(5' 말단에서 NHS 모이어티와 함께 나타낸 보편적인 프라이밍 서열(U1)을 포함함)에 화학적으로 접합된다. 보편적인 DNA 태그를 폴리펩타이드에 접합시킨 후, 과도한 보편적인 DNA 태그를 제거한다. **(B)** 보편적인 DNA 태그된-폴리펩타이드를 비드에 결합된 핵산 분자에 하이브리드화하며, 여기서 개개의 비드에 결합된 핵산 분자는 구획 태그(바코드) 서열의 유일한 집단을 포함한다. 구획화는 샘플을 소적(쇄선의 타원형으로 나타냄)과 같은 상이한 물리적 구획으로 분리함으로써 일어날 수 있다. 대안적으로, 구획화는 추가의 물리적인 분리에 대한 필요성 없이, 예를 들면, 비드 상의 구획 DNA 태그에 대해 폴리펩타이드 상의 보편적인 DNA 태그의 어닐링을 통해서, 표지된 폴리펩타이드를 비드 표면에 고정화시켜 직접 달성할 수 있다. 단일의 폴리펩타이드 분자는 단일의 비드와 만 상호작용한다(예컨대, 단일의 폴리펩타이드는 다수의 비드로 연결되지 않는다). 그러나, 다수의 폴리펩타이드는 동일한 비드와 상호작용할 수 있다. 구획 바코드 서열(BC) 외에, 비드에 결합된 핵산 분자는 일반적인 Sp(스페이서) 서열, 유일한 분자 확인인자(UMI), 및 폴리펩타이드 DNA 태그, U1'에 대해 상보성인 서열로 구성될 수 있다. **(C)** 보편적인 DNA 태그된 폴리펩타이드를 비드에 결합된 구획 태그로 어닐링한 후, 구획 태그를 부착 링커의 절단을 통해 비드로부터 방출한다. **(D)** 어닐링된 U1 DNA 태그 프라이머는 주형으로서 비드로부터 기원하는 구획 태그 핵산 분자를 사용하여 폴리머라제-기반 프라이머 연장을 통해 연장된다. 프라이머 연장 단계는 (C)에 나타낸 바와 같이 비드로부터 구획 태그의 방출 후에, 또는, 임의로 구획 태그가 여전히 비드(나타내지 않음)에 부착되어 있는 동안 수행할 수 있다. 이는 비드 상의 구획 태그로부터의 바코드 서열을 폴리펩타이드 상의 U1 DNA-태그 서열에 효과적으로 쓴다. 이러한 새로운 서열은 기록 태그를 구성한다. 프라이머 연장 후, 프로테아제, 예컨대, Lys-C(라이신 잔기의 C-말단 측면에서 절단함), Glu-C(글루탐산 잔기의 C-말단 측면에서 및 보다 낮은 정도의 글루탐산 잔기로 절단함), 또는 프로테이나제 K와 같은 무작위 프로테아제를 사용하여 폴리펩타이드를 펩타이드 단편으로 절단한다. **(E)** 각각의 펩타이드 단편은 본원에 개시된 바와 같은 하부 펩타이드 서열분석을 위해 이의 C-말단 라이신에서 기록 태그를 구성하는 연장된 DNA 태그 서열로 표지된다. **(F)** 기록 태그된 펩타이드는 가닥화된 알킨 표지, DBCO를 통해 아지드 비드에 커플링된다. 아지드 비드는 또한 임의로 기록 태그에 대해 상보성인 포획 서열을 함유함으로써 DBCO-아지드 고정화의 효능을 촉진한다. 원래의 비드로부터 펩타이드를 제거하여 새로운 고체 지지체(예컨대, 비드)에 재-고정화하는 것은 펩타이드 사이의 최적의 분자간 공간을 허용함으로써 본원에 개시된 펩타이드 서열분석 방법을 촉진한다. **도 20**의 **G** 내지 **L**은 알킨-예비-표지된 폴리펩타이드(도 2b에 기술된 바와 같음)에 대한 DNA 태그의 클릭 화학적 접합을 사용하는 것을 제외하고는 도 20의 A 내지 F에 나타낸 바와 유사한 개념을 나타낸다. 아지

드 및 mTet 화학은 직각이어서 DNA 태그에 대한 클릭 접합 및 서열분석 기질에 대한 클릭 iEDDA 접합(mTet 및 TCO)을 허용한다.

도 21은 단일 세포에 대한 유동-포커싱 T-이음(junction) 및 비드를 사용한 구획 태그된(예컨대, 바코드) 구획화를 사용한 예시적인 방법을 나타낸다. 2개의 수성 흐름으로, 세포 분해 및 프로테아제 활성화(Zn^{2+} 혼합)가 소적 형성시 용이하게 개시될 수 있다.

도 22의 **A** 및 **B**는 예시적인 태그화 세부사항을 나타낸다. **(A)** 구획 태그 (DNA-펩타이드 키메라)를 부텔라제 I를 사용한 펩타이드 연결을 사용하여 펩타이드 상에 부착시킨다. **(B)** 구획 태그 정보를 펩타이드 서열분석의 시행 이전에 관련된 기록 태그로 이전시킨다. 임의로, 펩타이드 결합 N-말단을 아스파르트산 잔기로 선택적으로 절단하는 엔도펩타다제 AspN을 사용하여 기록 태그로의 정보 이전 후 구획 태그를 절단할 수 있다.

도 23의 **A** 내지 **C**: 조직 슬라이스의 공간 프로테오믹스-기반 분석의 배열-기반 바코드. **(A)** 공간적으로-암호화된 DNA 바코드(BC_{ij}로 나타낸 특징적인 바코드)의 배열을 조직 슬라이스(FFPE 또는 동결됨)와 조합한다. 일 구현예에서, 조직 슬라이스를 고정시키고 투과성이 되도록 한다. 바람직한 구현예에서, 배열 특징적인 크기는 세포 크기보다 더 작다(사람 세포의 경우 ~10 μm). **(B)** 배열-장착된 조직 슬라이스를 시약으로 처리하여 역 가교-결합시키고(예컨대, 시트라콘산 무수물을 사용한 항원 회수 프로토콜(Namimatsu, Ghazizadeh et al. 2005)), 이후 여기의 단백질을 부위-반응성 DNA 표지로 표지하며, 이는 모든 단백질 분자를 DNA 기록 태그(예컨대, 항원 회수 후 유리된 라이신 표지화)로 효과적으로 표지한다. 표지 및 세척 후, 배열 결합된 DNA 바코드 서열을 절단하고 장착된 조직 슬라이스내로 확산되어 내부의 단백질에 부착된 DNA 기록 태그에 대해 하이브리드화하도록 한다. **(C)** 배열-장착된 조직을 이제 폴리머라제 연장에 적용시켜 하이브리드화된 바코드의 정보를 단백질을 표지하는 DNA 기록 태그로 이전시킨다. 바코드 정보를 이전한 후, 배열-장착된 조직을 슬라이드로부터 긁어내고, 임의로 프로테아제로 분해하고, 단백질 또는 펩타이드를 용액내로 추출한다.

도 24a 및 **24b**는 비드에 고정되고 암호화 태그에 부착된 결합제에 의해 분석된 2개의 상이한 예시적인 DNA 표적 거대분자(AB 및 CD)를 나타낸다. 이러한 모델 시스템은 결합제로부터 인접한 리포팅 태그로의 암호화 태그 이전의 단일 분자 거동을 나타내기 위해 제공된다. 바람직한 구현예에서, 암호화 태그는 프라이머 연장을 통해 연장된 재암호화 태그내로 혼입된다. **도 24a**는 AB 거대분자와 A-특이적인 결합제("A'", AB 거대분자의 "A" 성분에 대해 상보성인 올리고뉴클레오타이드 서열)의 상호작용 및 프라이머 연장을 통한 기록 태그, 및 B-특이적인 결합제("B'", AB 거대분자의 "B" 성분에 대해 상보성인 올리고뉴클레오타이드 서열)와 관련된 암호화 태그의 정보의 이전 및 관련된 암호화 태그의 정보의 프라이머 연장을 통한 재암호화 태그로의 이전을 나타낸다. 암호화 태그 A 및 B는 상이한 서열의 것이며 본 예시에서 용이한 확인을 위해, 또한 길이가 상이하다. 상이한 길이는 겔 전기영동에 의한 암호화 태그 이전(transfer)의 분석을 촉진시키지만, 차세대 서열분석에 의한 분석에는 필요하지 않다. A' 및 B' 결합제의 결합은 단일 결합 주기에 대한 대안적인 가능성으로서 나타나 있다. 제2의 주기를 가하는 경우, 연장된 기록 태그는 또한 연장될 수 있다. A' 또는 B' 결합제가 제1의 및 제2의 주기에 첨가되는 것에 따라, 연장된 기록 태그는 AA, AB, BA, 및 BB 형태의 암호화 태그 정보를 함유할 수 있다. 따라서, 연장된 기록 태그는 결합 사건의 순서 및 결합제의 동일성에 관한 정보를 함유한다. 유사하게, 도 **24b**는 CD 거대분자와 C-특이적인 결합제("C'''", CD 거대분자의 "C" 성분에 대해 상보성인 올리고뉴클레오타이드 서열)의 상호작용 및 프라이머 연장을 통한 관련된 암호화 태그의 정보의 기록 태그로의 이전, 및 D-특이적인 결합제("D'", CD 거대분자의 "D" 성분에 대해 상보성인 올리고뉴클레오타이드 서열) 및 프라이머 연장을 통한 관련된 암호화 태그의 정보의 기록 태그로의 이전을 나타낸다. 암호화 태그 C 및 D는 서열이 상이하며, 이를 나타내는데 있어 확인을 용이하게 하기 위해 또한 길이가 상이하다. 상이한 길이는 겔 전기영동에 의해 암호화 태그 이전의 분석을 촉진하지만, 차세대 서열분석에 의한 분석에는 필요하지 않다. C' 및 D' 결합제의 결합은 단일 결합 주기에 대한 대안적인 가능성으로서 나타낸다. 제2의 주기를 가하는 경우, 연장된 기록 태그는 추가로 연장될 수 있다. C' 또는 D' 결합제가 제1의 및 제2의 주기에 가해지느냐에 따라서, 연장된 기록 태그는 CC, CD, DC, 및 DD 형태의 암호화 태그 정보를 함유할 수 있다. 암호화 태그는 임의로 UMI를 포함할 수 있다. 암호화 태그 내 UMI의 혼입은 추가의 정보가 결합 사건에 대해 기록되도록 하며; 이는 결합 사건이 개개의 결합제의 수준에서 구별되도록 한다. 이는 개개의 결합제가 하나 이상의 결합 사건에 관여할 수 있는 경우 유용할 수 있다(예컨대, 이의 결합 친화성은 이것이 풀려서 하나 이상의 사건에 관여하기에 충분히 흔하게 재-결합할 수 있도록 한다). 이는 또한 오류-수정에 유용할 수 있다. 예를 들면, 일부 환경 하에서, 암호화 태그는 동일한 결합 주기에서 정보를 기록 태그에 2회 이상 이전시킬 수 있다. UMI의 사용은 이들이 단일 결합 사건에 모두 연결된 정보 이전 사건을 반복되는 경향이 있음을 나타낼 수 있다.

도 25는 비드에 고정되고 암호화 태그에 부착된 결합제에 의해 검정된 예시적인 DNA 표적 거대분자(AB)를 나타낸다. A-특이적인 결합제 ("A'", AB 거대분자의 A 성분에 대해 상보성인 올리고뉴클레오타이드)는 AB 거대분자와 상호작용하며 관련된 암호화 태그의 정보는 연결에 의해 기록 태그로 이전된다. B-특이적인 결합제("B'", AB 거대분자의 B 성분에 대해 상보성인 올리고뉴클레오타이드)는 AB 거대분자와 상호작용하며 관련된 암호화 태그의 정보는 연결에 의해 기록 태그로 이전된다. 암호화 태그 A 및 B는 길이가 상이하며 이를 나타내는데 있어서 확인을 용이하게 하기 위해 또한 길이가 상이하다. 상이한 길이는 겔 전기영동에 의한 암호화 태그 이전의 분석을 촉진하지만 차세대 서열분석에 의한 분석에는 필요하지 않다.

도 26의 **A** 및 **B**는 프라이머 연장을 통한 결합/암호화 태그 이전을 위한 예시적인 DNA-펩타이드 거대분자를 나타낸다. **도 26**의 **A**는 비드 상에 고정된 예시적인 올리고뉴클레오타이드-펩타이드 표적 거대분자("A" 올리고뉴클레오타이드-cMyc 펩타이드)를 나타낸다. cMyc-특이적인 결합제 (예컨대 항체)는 거대분자의 cMyc 펩타이드 부위에 결합하며 관련된 암호화 태그의 정보는 기록 태그로 이전된다. cMyc 암호화 태그의 정보의 기록 태그로의 이전은 겔 전기영동에 의해 분석될 수 있다. **도 26**의 **B**는 비드 상에 고정된 예시적인 올리고뉴클레오타이드-펩타이드 표적 거대분자 ("C" 올리고뉴클레오타이드-헤마글루티닌(HA) 펩타이드)를 나타낸다. HA-특이적인 결합제 (예컨대, 항체)는 거대분자의 HA 펩타이드 부위와 상호작용하며 관련된 암호화 태그의 정보는 기록 태그로 이전된다. 암호화 태그의 정보의 기록 태그로의 이전은 겔 전기영동에 의해 분석될 수 있다. cMyc 항체-암호화 태그 및 HA 항체-암호화 태그의 결합은 단일 결합 주기에 대한 대안적인 가능성으로서 나타난다. 제2의 결합 주기를 수행하는 경우, 연장된 기록 태그는 추가로 연장될 수 있다. cMyc 항체-암호화 태그 또는 HA 항체-암호화 태그가 제1 및 제2의 결합 주기에서 가해지는 것에 따라, 연장된 기록 태그는 cMyc-HA, HA-cMyc, cMyc-cMyc, 및 HA-HA 형태의 암호화 태그 정보를 함유할 수 있다. 나타나지는 않았지만, 추가의 결합제를 또한 도입하여 거대분자의 A 및 C 올리고뉴클레오타이드 성분의 검출을 가능하도록 할 수 있다. 따라서, 상이한 유형의 골격을 포함하는 하이브리드 거대분자를 기록 태그로의 정보의 이전 및 연장된 기록 태그의 판독을 통해 분석할 수 있으며, 이는 결합 사건의 순서 및 또한 결합제의 동일성을 함유한다.

도 27a 내지 **27d. 오류-정정 바코드의 생성**. **(A)** 65개의 오류-정정 바코드(서열 번호: 1 내지 65)의 소세트를 R 소프트웨어 패키지 'DNABardodes'(https://bioconductor.riken.jp/packages/3.3/bioc/manuals/DNABarcodes/man/DNABarcodes.pdf)로부터 기원한 77개의 바코드의 세트로부터 커맨드 매개변수(command parameter)[create.dnabardodes(n=15, dist=10)]를 사용하여 선택하였다. 이러한 알고리즘은 4개의 치환의 거리에 대해 치환 오류를 교정할 수 있고 9개의 치환 오류를 검출할 수 있는 15-머 "햄밍(Hamming)" 바코드를 생성한다. 65개 바코드의 소세트는 다양한 나노공극 전류 수준(나노공극-기반 서열분석의 경우)을 나타내지 않거나 이러한 세트의 다른 구성원과 너무 관련된 바코드를 필터링함으로써 생성하였다. **(B)** 공극을 통과하는 15-머(mer) 바코드에 대한 예측된 나노포어 전류 수준의 플롯. 예측된 전류는 각각의 15-머 바코드 단어를 5-머 단어를 오버랩핑한 11개의 복합 세트로 분할하고, 5-머 R9 나노공극 전류 수준 검색 표(nanopore current level look-up table)(주형_median68pA.5mers.model (https://github.com/jts/nanopolish/tree/master/etc/r9-models)를 사용하여 계산함으로써 바코드가 한번에 1개의 염기로 나노공극을 통과함에 따라 상응하는 전류 수준을 예측하였다. (B)로부터 인식될 수 있는 바와 같이, 65개 바코드의 이러한 세트는 이의 구성원 각각에 대한 유일한 전류 신호를 나타낸다. **(C)** 나노공극 서열분석용의 모델 연장된 기록 태그로서 PCR 생성물의 생성은 DTR 및 DTR 프라이머의 오버랩핑 세트를 사용하여 나타낸다. PCR 앰플리콘을 이후에 연결하여 연관되어 연장된 기록 태그 모델을 형성한다. **(D)** 도 27c에 나타낸 바와 같이 생성된 예시적인 "연장된 기록 태그" 모델(판독물 길이 734개 염기)의 나노공극 서열분석 판독물. MinIon R9.4 판독물은 품질 점수가 7.2(불량한 판독 품질)이다. 그러나, 바코드 서열은 불량한 품질 판독물(Qscore = 7.2)을 사용하는 경우에도 *lalign*을 사용하여 용이하게 확인할 수 있다. 15-머 스페이서 요소는 밑줄쳐져 있다. 바코드는 BC 또는 BC' 설계로 나타낸, 전방 또는 역방 배향으로 정렬될 수 있다.

도 28의 **A** 내지 **D. 기록 태그를 지닌 단백질의 분석물-특이적인 표지화. (A)** 이의 천연 구조에서 목적한 단백질 분석물을 표적화하는 결합제는 DNA 기록 태그에서 상보성인 분석물-특이적인 바코드(BC_A)에 하이브리드화하는 상보성 분석물-특이적인 바코드(BC_A')를 포함한다. 대안적으로, DNA 기록 태그는 절단가능한 링커를 통해 결합제에 부착될 수 있으며 DNA 기록 태그는 단백질에 직접 "클릭"되고 후속적으로 결합제로부터(절단가능한 링커를 통해) 절단된다. DNA 기록 태그는 목적한 단백질에 커플링하기 위한 반응성 커플링 모이어티(예를 들면, 클릭 화학 시약(예컨대, 아지드, mTet, 등), 및 다른 기능성 성분(예컨대, 보편적인 프라이밍 서열 (P1), 샘플 바코드 (BC), 분석물 특이적인 바코드 (BC_A), 및 스페이서 서열 (Sp))을 포함한다. 샘플 바코드(BC)를 또한 사용하여 단백질을 표지하고 상이한 샘플로부터 이를 구별할 수 있다. DNA 기록 태그는 또한 기질 표면에 후속적인 커

플링을 위한 직교 커플링 모이어티(예컨대, mTet)를 포함할 수 있다. 목적한 단백질에 기록 태그를 클릭 화학 커플링하기 위해, 단백질은 DNA 기록 태그(예컨대, 단백질 상의 알킨 모이어티는 DNA 기록 태그 상의 아지드 모이어티에 대해 동종이다)에서 클릭 화학 커플링 모이어티에 대해 동종인 클릭 화학 커플링 모이어티로 예비-표지된다. DNA 기록 태그를 클릭 화학 커플링용 커플링 모이어티로 표지하기 위한 시약의 예는 라이신 표지를 위한 알킨-NHS 시약, 광친화성 표지를 위한 알킨-벤조페논 시약 등을 포함한다. **(B)** 결합제가 인접한 표적 단백질에 결합한 후, 기록 태그(예컨대, 아지드) 상에서 반응성 커플링 모이어티는 인접한 단백질에서 동종의 클릭 화학 커플링 모이어티(삼중선 기호로 나타냄)에 공유결합으로 부착한다. **(C)** 표적 단백질 분석물이 기록 태그로 표지된 후, 부착된 결합제는 우라실-특이적인 절제 시약(예컨대, USER™)을 사용한 우라실(U)의 분해에 의해 제거된다. **(D)** DNA 기록 태그 표지된 표적 단백질 분석물은 클릭 화학(알킨-아지드 결합 쌍, 메틸 테트라진(mTET)-*트랜스*-사이클로옥텐(TCO) 결합 쌍 등)과 같은 적합한 생물접합 화학 반응을 사용하여 기질 표면에 고정시킨다. 특정 구현예에서, 전체의 표적 단백질-기록 태그 표지화 검정은 결합제의 혼주물 및 기록 태그의 혼주물을 사용하여 많은 상이한 표적 단백질을 포함하는 단일 튜브 속에서 수행한다. 샘플 바코드(BC_S)를 포함하는 기록 태그를 지닌 샘플 내에서 단백질 분석물의 표적화된 표지화 후, 다수의 단백질 분석물 샘플을 (D)에서 고정화 단계 전에 혼주시킬 수 있다. 따라서, 특정의 구현예에서, 수백개의 샘플에 걸친 수천개 이하의 단백질 분석물을 표지하고 단일 튜브 차세대 단백질 검정(NGPA)에서 고정시킬 수 있으므로, 비싼 친화성 시약(예컨대, 항체)을 크게 절약할 수 있다.

도 29의 **A** 내지 **E. 폴리펩타이드에 대한 DNA 기록 태그의 접합. (A)** 변성된 폴리펩타이드를 알킨-NHS 에스테르(아세틸렌-PEG-NHS 에스테르) 시약 또는 알킨-벤조페논과 같은 이기능성 클릭 화학 시약으로 표지하여 알킨-표지된(삼중선 기호) 폴리펩타이드를 생성한다. 알킨은 또한 가닥화된 알킨, 예를 들면, 디벤조사이클로옥틸(DBCO) 등을 포함하는 사이클로옥틴일 수 있다. **(B)** 알킨-표지된 폴리펩타이드에 화학적으로 커플링된 DNA 기록 태그 설계의 예가 나타나 있다. 기록 태그는 보편적인 프라이밍 서열 (P1), 바코드 (BC), 및 스페이서 서열(Sp)을 포함한다. 기록 태그는 지질 표면에 커플링하기 위한 mTet 모이어티 및 표지된 폴리펩타이드의 알킨 모이어티와 커플링하기 위한 아지드 모이어티로 표지된다. **(C)** 변성된, 알킨-표지된 단백질 또는 폴리펩타이드는 알킨 및 아지드 모이어티를 통해 기록 태그로 표지된다. 임의로, 기록 태그-표지된 폴리펩타이드는 예컨대, 구획 비드에 부착된 상보성 서열에 대한 어닐링 및 프라이머 연장(또한 폴리머라제 연장으로 언급됨)을 통해, 또는 도 20의 H 내지 J에 나타낸 바와 같이, 구획 바코드로 추가로 표지될 수 있다. **(D)** 기록 태그-표지된 폴리펩타이드의 프로테아제 분해는 기록 태그-표지된 펩타이드의 집단을 생성한다. 일부 구현예에서, 일부 펩타이드는 어떠한 기록 태그로도 표지되지 않을 것이다. 다른 구현예에서, 일부 펩타이드는 부착된 하나 이상의 기록 태그를 가질 수 있다. **(E)** 기록 태그-표지된 펩타이드는 TCO 그룹으로 기능화된 기질 표면과 펩타이드에 부착된 기록 태그의 mTet 모이어티 사이의 역 전자 요구 디엘-앨더(inverse electron demand Diels-Alder: iEDDA) 클릭 화학 반응을 사용하여 기질 표면 위에 고정시킨다. 특정 구현예에서, 정화(clean-up) 단계는 나타낸 상이한 단계 사이에서 사용할 수 있다. 직교 클릭 화학(예컨대, 아지드-알킨 및 mTet-TCO)의 사용은 기록 태그를 지닌 폴리펩타이드의 클릭 화학 표지화, 및 기질 표면에 기록 태그-표지된 폴리펩타이드의 클릭 화학 고정화 둘 다를 허용한다(참고: McKay et al., 2014, Chem. Biol. 21:1075-1101, 이의 전문이 참고로 포함됨).

도 30의 **A** 내지 **E. 폴리펩타이드의 초기 DNA 태그 표지화 후 기록 태그 내로의 쓰기 샘플 바코드(Writing sample barcode). (A)** 변성된 폴리펩타이드를 알킨-NHS 시약 또는 알킨-벤조페논과 같은 이기능성 클릭 화학 시약으로 표지하여 알킨-표지된 폴리펩타이드를 생성한다. **(B)** 폴리펩타이드의 알킨(또는 대안적인 클릭 화학 모이어티) 표지화 후, 보편적인 프라이밍 서열 (P1)을 포함하고 아지드 모이어티 및 mTet 모이어티로 표지된 DNA 태그를 아지드-알킨 상호작용을 통해 폴리펩타이드에 커플링한다. 다른 클릭 화학 상호작용을 사용할 수 있음이 이해된다. **(C)** 샘플 바코드 정보(BC_S') 및 다른 기록 태그 기능성 성분(예컨대, 보편적인 프라이밍 서열 (P1'), 스페이서 서열 (Sp'))을 포함하는 기록 태그 DNA 작제물을 상보성의 보편적인 프라이밍 서열(P1-P1')을 통해 DNA 태그-표지된 폴리펩타이드에 어닐링한다. 기록 태그 정보를 폴리머라제 연장에 의해 DNA 태그로 이전시킨다. **(D)** 기록 태그-표지된 폴리펩타이드의 프로테아제 분해는 기록 태그-표지된 펩타이드의 집단을 생성한다. **(E)** 기록 태그-표지된 펩타이드는 TCO 그룹으로 기능화된 표면과 펩타이드에 부착된 기록 태그의 mTet 모이어티 사이의 역 전자 요구 디엘스-앨더(iEDDA) 클릭 화학 반응을 사용하여 기질 표면 위에 고정시킨다. 특정 구현예에서, 정화 단계는 나타낸 상이한 단계 사이에서 사용할 수 있다. 직교 클릭 화학(예컨대, 아지드-알킨 및 mTet-TCO)의 사용은 기록 태그를 지닌 폴리펩타이드의 클릭 화학 표지, 및 기질 표면에 기록 태그-표지된 폴리펩타이드의 클릭 화학 고정화 둘 다를 허용한다(참고: McKay et al., 2014, Chem. Biol. 21:1075-1101, 이의 전문은 참고로 포함된다).

도 31의 A 내지 **E. 폴리펩타이드의 바코드화를 위한 비드 구획화.** **(A)** 폴리펩타이드를 헤테로이기능성 클릭 화학 시약이 들어있는 용액 속에서 표준 생물접합 또는 광친화성 표지화 기술을 사용하여 표지한다. 가능한 표지화 부위는 라이신 잔기의 ε-아민(예컨대, 나타낸 바와 같은 NHS-알킨을 지닌) 또는 펩타이드의 탄소 골격(예컨대, 벤조페논-알킨을 지님)을 포함한다. **(B)** 보편적인 프라이밍 서열(P1)을 포함하는 아지드-표지된 DNA 태그를 표지된 폴리펩타이드의 알킨 모이어티에 커플링한다. **(C)** DNA 태그-표지된 폴리펩타이드를 DNA 기록 태그 표지된 비드에 상보성 DNA 서열(P1 및 P1')을 통해 어닐링한다. 비드 상의 DNA 기록 태그는 스페이서 서열(Sp'), 구획 바코드 서열(BC_P'), 임의의 유일한 분자 확인인자(UMI), 및 보편적인 서열(P1')을 포함한다. DNA 기록 태그 정보를 폴리펩타이드 위의 DNA 태그로 폴리머라제 연장(대안적으로, 연결을 사용할 수 있었다)을 통해 이전시켰다. 정보 이전 후, 수득되는 폴리펩타이드는 구획 바코드를 포함하는 몇가지 기능적 요소를 함유하는 다수의 기록 태그를 포함한다. **(D)** 기록 태그-표지된 폴리펩타이드의 프로테아제 분해는 기록 태그-표지된 펩타이드의 집단을 생성한다. 기록 태그-표지된 펩타이드는 비드로부터 해리되며, **(E)** 서열분석 기질(예컨대, 나타낸 바와 같이 mTet와 TCO 모이어티 사이의 iEDDA 클릭 화학 사용) 위에 재-고정화된다.

도 32a 내지 **32h. 차세대 단백질 검정(NGPA)을 위한 작업흐름의 예.** 단백질 샘플을 몇가지 기능성 단위, 예컨대, 보편적인 프라이밍 서열(P1), 바코드 서열(BC), 임의의 UMI 서열, 및 스페이서 서열 (Sp)(결합제 암호화 태그를 사용한 정보 이전을 가능하도록 함)로 구성된 DNA 기록 태그로 표지한다. **(A)** 표지된 단백질을 기질(예컨대, 비드,다공성 비드 또는 다공성 매트릭스)에 고정시킨다(수동적으로 또는 공유결합으로). **(B)** 기질을 단백질로 차단하고, 임의로 스페이서 서열에 대해 상보성인 경쟁인자 올리고뉴클레오타이드(Sp')를 가하여 분석물 기록 태그 서열의 비-특이적인 상호작용을 최소화한다. **(C)** 분석물-특이적인 항체(관련된 암호화 태그를 지님)를 기질-결합된 단백질과 함께 항온처리한다. 암호화 태그는 후속적인 우라실 특이적인 절단을 위한 우라실 염기를 포함할 수 있다. **(D)** 항체 결합 후, 과도한 경쟁인자 올리고뉴클레오타이드(Sp')를, 첨가한 경우 세척제거한다. 암호화 태그를 상보성 스페이서 서열을 통해 기록 태그에 일시적으로 어닐링하고, 암호화 태그 정보를 프라이머 연장 반응에서 기록 태그로 이전시켜 연장된 기록 태그를 생성한다. 고정된 단백질이 변성된 경우, 결합된 항체 및 어닐링된 암호화 태그를 0.1N NaOH와 같은 알칼리성 세척 조건 하에서 제거할 수 있다. 고정된 단백질이 천연 구조인 경우, 온화한 조건이 결합된 항체 및 암호화 태그를 제거하기 위해 요구될 수 있다. 온화한 항체 제거 조건의 예는 패널 E 내지 H에 요약되어 있다. **(E)** 암호화 태그로부터 기록 태그로 정보 이전 후, 암호화 태그를 우라실-특이적인 절개 시약(예컨대, USER™) 효소 혼합물을 사용하여 이의 우라실 부위에 닉(nick)을 생성한다(절단한다). **(F)** 결합된 항체를 단백질로부터 고-염, 저/고 pH 세척액을 사용하여 제거한다. 항체에 부착되어 남아있는 트렁케이트된(truncated) DNA 암호화 태그는 짧고 또한 신속하에 용출제거된다. 보다 긴 DNA 암호화 태그 단편은 기록 태그에 어닐링되어 남아있거나 남아있지 않는다. **(G)** 제2의 결합 주기는 단계 (B) 내지 (D)에서와 같이 개시하며, 제2의 프라이머 연장 단계는 제2 항체로부터의 암호화 태그 정보를 프라이머 연장을 통해 연장된 기록 태그로 이전한다. **(H)** 2개의 결합 주기의 결과는 기록 태그에 부착된 제1의 항체 및 제2의 항체로부터의 결합 정보의 콘카테네이트(concatenate)이다.

도 33의 A 내지 **D. 다수의 결합제 및 효소적으로 매개된 서열 정보 이전을 사용한 단일-단계 차세대 단백질 검정(NGPA).** 고정된 단백질 분자를 사용한 NGPA 검정을 2개의 동종의 결합제(예컨대, 항체)로 동시 결합시켰다. 다수의 동종의 항체 결합 사건 후, 합해진 프라이머 연장 및 DNA 닉생성(nicking) 단계를 사용하여 결합된 항체의 암호화 태그로부터의 정보를 기록 태그로 이전시킨다. 암호화 태그에서 탈자 기호(^)는 이중 가닥 DNA 닉생성 엔도뉴클레아제 부위를 나타낸다. **(A)** 나타낸 실시예에서, 단백질의 에피토프 1(Epi#1)에 결합된 항체의 암호화 태그는 상보성 스페이서 서열의 하이브리드화 후 프라이머 연장 단계에서 암호화 태그 정보(예컨대, 암호기 서열)를 기록 태그로 이전한다. **(B)** 일단 연장된 기록 태그와 암호화 태그 사이의 이중 가닥 DNA 듀플렉스가 형성되면, 37℃에서 활성인, Nt.BsmAI와 같은 이중-가닥 DNA의 1개 가닥만을 절단하는 닉생성(nicking) 엔도뉴클레아제를 사용하여 암호화 태그를 절단한다. 닉생성 단계 후, 트렁케이트된 암호화 태그-결합제 및 연장된 기록 태그로부터 형성된 듀클렉스는 열역학적으로 불안정하며 해리된다. 보다 긴 암호화 태그 단편은 기록 태그에 어닐링되어 남거나 남지 않을 수 있다. **(C)** 이는 단백질의 에피토프 #2(Epi#2)에 결합된 항체로부터의 암호화 태그가 상보성 스페이서 서열을 통해 연장된 기록 태그에 어닐링하도록 하며 Epi#2 항체의 암호화 태그로부터의 정보를 프라이머 연장을 통해 연장된 기록 태그로 이전함으로써 연장된 기록 태그가 추가로 연장되도록 한다. **(D)** 다시 한번, 이중 가닥 DNA 듀플렉스가 Epi#2 항체의 연장된 기록 태그와 암호화 태그 사이에 형성된 후에, 암호화 태그는 Nb.BssSI와 같은 닉생성 엔도뉴클레아제에 의해 닉생성된다. 특정의 구현예에서, 프라이머 연장(또한 폴리머라제 연장으로 지칭됨) 동안 비-가닥 치환 폴리머라제의 사용이 바람직하다. 비-가닥 치환 폴리머라제는 하나 이상의 단일 염기에 의해 기록 태그에 어닐링되어 남아있는 절단된 암호화 태그의 연장을

방지한다. (A) 내지 (D)의 과정은 인접하여 결합된 결합제의 암호화 태그 모두가 하이브리드화, 연장된 기록 태그로의 정보 이전, 및 닉생성 단계에 의해 "소비"될 때까지 자체적으로 반복할 수 있다. 암호화 태그는 제공된 분석물(예컨대, 동종의 단백질)에 대해 특이적인 모든 결합제(예컨대, 항체)에 대해 동일한 암호기 서열을 포함할 수 있거나, 에피토프-특이적인 암호기 서열을 포함할 수 있거나, 상이한 분자 사건 사이를 구별하기 위한 유일한 분자 확인인자(UMI)를 포함할 수 있다.

도 34의 **A** 내지 **C: 기질 표면에서 반응성 모이어티의 적정을 사용한 기록 태그-펩타이드 고정화의 조절된 밀도. (A)** 기질 표면의 펩타이드 밀도는 기질 표면의 기능성 커플링 모이어티를 조절함으로서 적정할 수 있다. 이는 기질의 표면을 적절한 비의 활성 커플링 분자로 유도체화하여 커플링 분자를 "더미(dummy)"화함으로써 달성할 수 있다. 나타낸 실시예에서, NHS-PEG-TCO 시약(활성의 커플링 분자)은 TCO가 있는 아민 표면을 유도체화하기 위한 정의된 비로 NHS-mPEG(더미 분자)와 조합됨이 실시예에 나타나 있다. 기능화된 PEG는 300 내지 40,000 이상에 이르는 다양한 분자량으로 된다. **(B)** 이기능성 5' 아민 DNA 기록 태그(mTet는 다른 기능성 모이어티이다)는 석신이미딜 4-(N-말레이미도메틸)사이클로헥산-1(SMCC) 이기능성 가교결합제를 사용하여 펩타이드의 N-말단 Cys 잔기에 커플링된다. 기록 태그 상의 내부 mTet-dT 그룹은 아지드-dT 그룹으로부터 m테트라진-아지드를 사용하여 생성된다. **(C)** 기록 태그 표지된 펩타이드는 mTet 및 TCO의 iEDDA 클릭 화학 반응을 사용하여 (A)로부터 활성화된 기질 표면으로 고정된다. mTet-TCO iEDDA 커플링 반응은 매우 신속하고, 효율적이며, 안정하다(mTet-TCO가 Tet-TCO보다 더 안정하다).

도 35의 **A** 내지 **C. 차세대 단백질 서열분석 (NGPS) 결합 주기 - 특이적인 암호화 태그. (A)** 주기-특이적인 N-말단 아미노산 (NTAA) 결합제 암호화 태그를 사용한 NGPS 검정의 설계. NTAA 결합제(예컨대, N-말단 DNP-표지된 타이로신에 대해 특이적인 항체)를 보편적인 프라이밍 서열(P1), 바코드(BC) 및 스페이서 서열(Sp)을 포함하는 기록 태그와 관련된 펩타이드의 DNP-표지된 NTAA에 결합시킨다. 결합제가 펩타이드의 동종의 NTAA에 결합하면, NTAA 결합제와 관련된 암호화 태그는 기록 태그와 근접하게 되며 상보성 스페이서 서열을 통해 기록 태그로 어닐링된다. 암호화 태그 정보는 프라이머 연장을 통해 기록 태그로 이전된다. 암호화 태그가 나타내는 결합 주기의 트랙을 유지하기 위해, 암호화 태그는 주기-특이적인 바코드를 포함할 수 있다. 특정의 구현예에서, 분석물에 결합하는 결합제의 암호화 태그는 주기 수와는 독립적으로 동일한 암호기 바코드를 가지며, 이는 유일한 결합 주기-특이적인 바코드와 조합된다. 다른 구현예에서, 분석물에 대한 결합제용 암호화 태그는 조합된 분석물-결합 주기 정보에 대한 유일한 암호기 바코드를 포함한다. 어느 시도에서도, 일반적인 스페이서 서열을 각각의 결합 주기에서 결합제의 암호화 태그용으로 사용할 수 있다. **(B)** 당해 실시예에서, 각각의 결합 주기로부터의 결합제는 결합 주기를 확인하기 위한 짧은 결합 주기-특이적인 바코드를 가지며, 이는 결합제를 확인하는 암호기 바코드와 함께, 특수한 결합제-결합 주기 조합을 확인하는 유일한 조합 바코드를 제공한다. **(C)** 결합 주기의 완료 후, 연장된 기록 태그는 캡핑 주기 단계를 사용하는 증폭가능한 라이브러리로 전환될 수 있으며, 여기서 예를 들면, 보편적인 프라이밍 서열 P2 및 스페이서 서열 Sp'에 연결된 보편적인 프라이밍 서열 P1'을 포함하는 캡(cap)은 상보성 P1 및 P1' 서열을 통해 연장된 기록 태그로 초기에 어닐링되어 캡이 연장된 기록 태그에 근접하도록 한다. 연장된 기록 태그 및 캡에서 상보성 Sp 및 Sp' 서열은 어닐링되며 프라이머 연장은 제2의 보편적인 프라이머 서열(P2)을 연장된 기록 태그에 가한다.

도 36a 내지 **도 36e. 암호화 태그로부터 기록 태그로의 정보 이전을 입증하기 위한 DNA 기반 모델 시스템.**

예시적인 결합 및 분자내 쓰기는 올리고뉴클레오타이드 모델 시스템에 의해 입증되었다. 암호화 태그내에서 표적화제 A' 및 B'를 설계하여 기록 태그내 표적 결합 영역 A 및 B에 하이브리드화시켰다. 기록 태그(RT) 혼합물을 2개의 재암호화 태그, saRT_Abc_v2(A 표적) 및 saRT_Bbc_V2(B 표적)를 동등한 농도에서 혼주하여 제조하였다. 기록 태그는 이들의 5' 말단에서 바이오티닐화되며 유일한 표적 결합 영역, 보편적인 전방 프라이머 서열, 유일한 DNA 바코드, 및 8개 염기의 일반적인 스페이서 서열(Sp)을 함유한다. 암호화 태그는 8개 염기의 일반적인 스페이서 서열(Sp')에 의해 플랭킹된 유일한 암호기 바코드를 함유하며, 이들 중 하나는 폴리에틸렌 글리콜 링커를 통해 A 또는 B 표적제에 공유결합으로 연결된다. **(A)** 바이오티닐화된 기록 태그 올리고뉴클레오타이드(saRT_Abc_v2 and saRT_Bbc_V2)를 바이오티닐화된 더미-T10 올리고뉴클레오타이드와 함께 스트렙타비딘 비드에 고정시켰다. 기록 태그를 A 또는 B 포획 서열(동종의 결합제 - A' 및 B' 각각으로 인식됨), 및 상응하는 바코드(rtA_BC 및 rtB_BC)로 설계하여 결합 표적을 확인하였다. 이러한 모델 시스템에서 모든 바코드를 65개의 15-머 바코드(서열 번호: 1 내지 65)의 세트로부터 선택하였다. 일부 경우에, 15-머 바코드를 합하여 용이한 겔 분석을 위해 보다 긴 바코드를 구성하였다. 특히, rtA_BC = BC_1 + BC_2; rtB_BC = BC_3. 기록 태그의 A 및 B 서열에 대해 동종인 결합제에 대한 2개의 암호화 태그, 즉 CT_A'-bc(암호기 바코드 = BC_5) 및 CT_B'-bc(암호기 바코드 = BC_5+BC_6)를 또한 합성하였다. 암호화 태그 서열(단일 가닥 Sp' 서열이 남아 있음)의 일부에 대해 상

보성인 차단 올리고(DupCT_A'BC 및 DupCT_AB'BC)를 비드-고정된 기록 태그에 대해 암호화 태그를 어닐링하기 전에 암호화 태그에 임의로 재-어닐링하였다. 가닥 치환 폴리머라제는 폴리머라제 연장 동안 차단 올리고를 제거한다. 바코드 키(삽입물)는 기록 태그 및 암호화 태그에서 기능성 바코드에 대한 15-머 바코드의 지정을 나타낸다. **(B)** 기록 태그 바코드 설계 및 암호화 태그 암호기 바코드 설계는 기록 태그와 암호화 태그 사이의 "분자내" 대 "분가간" 상호작용의 용이한 겔 분석을 제공한다. 이러한 설계에서, 바람직하지 않은 "분가간" 상호작용(A 기록 태그와 B' 암호화 태그, 및 B 기록 태그와 A' 암호화 태그)은 바람직한 "분자간"(A 기록 태그와 A' 암호화 태그; B 기록 태그와 B' 암호화 태그) 상호작용 생성물보다 더 길거나 더 짧은 시들은(wither) 겔 생성물을 생성한다. 프라이머 연장 단계는 A' 및 B' 암호화 태그 바코드(ctA'_BC, ctB'_BC)를 역 상보체 바코드(ctA_BC 및 ctB_BC)로 변화시킨다. **(C)** 프라이머 연장 검정은 암호화 태그로부터 기록 태그로의 정보 이전, 및 PCR 분석을 위한 어닐링된 EndCap 올리고 상의 프라이머 연장을 통한 어댑터 서열의 첨가를 입증하였다. **(D)** 더미-T20 올리고의 사용을 통한 기록 태그의 표면 밀도의 적정을 통한 "분자내" 정보의 이전의 최적화. 바이오티닐화된 기록 태그 올리고를 1:0, 1:10, 1;10000으로 모든 하향된 방식의 다양한 비에서 바이오티닐화된 더미-T20 올리고와 혼합하였다. 감소된 기록 태그 밀도($1:10^3$ 및 $1:10^4$)에서, "분자내" 상호작용은 "분자간" 상호작용보다 우세하다. **(F)** DNA 모델 시스템의 단순한 연장으로서, 나노-태그$_{15}$ 펩타이드-스트렙타비딘 결합 쌍을 포함하는 단순한 단백질 결합 시스템이 나열되어 있지만(K_D ~4 nM)(Perbandt et al., 2007, Proteins 67:1147-1153), 어떠한 수의 펩타이드-결합제 모델 시스템도 사용할 수 있다. 나노-태그$_{15}$ 펩타이드 서열은 (fM)DVEAWLGARVPLVET(서열 번호: 131)(fM = 포르밀-Met)이다. 나노-태그$_{15}$ 펩타이드는 또한 짧은, 굴곡성 링커 펩타이드(GGGGS) 및 DNA 기록 태그에 커플링시키기 위한 시스테인 잔기를 포함한다. 다른 예의 펩타이드 태그-동종의 결합제 쌍은: 칼모둘린 결합 펩타이드(CBP)-칼모둘린(K_D ~2 pM)(Mukherjee et al., 2015, J. Mol. Biol. 427: 2707-2725), 아밀로이드-베타(Aβ16-27) 펩타이드-US7/Lcn2 안티칼린(0.2 nM) (Rauth et al., 2016, Biochem. J. 473: 1563-1578), PA 태그/NZ-1 항체(K_D ~ 400 pM), FLAG-M2 Ab(28 nM), HA-4B2 Ab(1.6 nM), 및 Myc-9E10 Ab(2.2 nM)(Fujii et al., 2014, Protein Expr. Purif. 95:240-247)를 포함한다. **(E)** 프라이머 연장을 통한 결합제의 암호화 태그로부터 기록 태그로의 분자내 정보 이전의 시험으로서, 상보성 DNA 서열 "A"에 결합하는 올리고뉴클레오타이드 "결합제"를 시험 및 개발에 사용할 수 있다. 이러한 하이브리드화 사건은 필수적으로 fM 친화성보다 크다. 스트렙타비딘은 나노-태그$_{15}$ 펩타이드 에피토프에 대한 시험 결합제로서 사용할 수 있다. 펩타이드 태그-결합제 상호작용은 고 친화성이지만, 산성 및/또는 고 염 세척액으로 용이하게 파괴할 수 있다(Perbandt et al., supra).

도 37의 A 및 B. UMI-표지된 N 또는 C 말단으로부터의 정보를 폴리펩타이드의 바디(body)를 표지하는 DNA 태그로 이전시키기 위한 나노- 또는 마이크로-유화액 PCR의 용도. (A) 폴리펩타이드를 이의 N- 또는 C-말단에서 유일한 분자 확인인자(UMI)를 포함하는 핵산 분자로 표지한다. UMI는 후속되는 PCR을 프라이밍하기 위해 사용된 서열에 의해 플랭킹될 수 있다. 폴리펩타이드는 이후에 UMI를 플랭킹하는 프라이밍 서열에 대해 상보성인 서열을 포함하는 별도의 DNA 태그로 내부 부위에서 "바디 표지(body labeled)"된다. **(B)** 수득되는 표지된 폴리펩타이드를 증폭시키고 유화액 PCR(ePCR)(대안적으로, 유화액 시험관내 전사-RT-PCR(IVT-RT-PCR) 반응 또는 다른 적합한 증폭 반응이 사용될 수 있다) 반응시켜 N- 또는 C-말단 UMI를 증폭시킨다. 마이크로유화액 또는 나노유화액은 평균 소적 직경이 50내지 1000 nm 이고 평균적으로 소적당 1개 미만으로 적은 폴리펩타이드가 존재하도록 형성된다. PCR 전 및 후 소적 성분의 스냅샷(snapshot)은 각각 좌측 패널 및 우측 패널에 나타낸다. UMI 앰플리콘은 상보성 프라이밍 서열을 통해 내부 폴리펩타이드 바디 DNA 태그에 하이브리드화되며 UMI 정보는 프라이머 연장을 통해 앰플리콘으로부터 내부 폴리펩타이드 바디 DNA 태그로 이전된다.

도 38. 단일 세포 프로테오믹스. 세포를 캡슐화하고 중합체-형성 소단위(예컨대, 아크릴아미드)를 함유하는 소적 속에서 분해한다. 중합체-형성 소단위를 중합시키고(예컨대, 아크릴아미드), 단백질을 중합체 매트릭스에 가교-결합시킨다. 유화액 소적을 파괴하고 투과성 중합체 매트릭스에 부착된 단일 세포 단백질 분해물을 함유하는 중합된 겔 비드를 방출시킨다. 단백질을 용해물 및 캡슐화 완충액 속에 우레아와 같은 변성체를 포함시킴으로써 변성 상태로 또는 이들의 천연 구조로 중합체 매트릭스에 가교-결합시킨다. 구획 바코드 및 다른 기록 태그 성분(예컨대, 보편적인 프라이밍 서열(P1), 스페이서 서열(Sp), 임의의 유일한 분자 확인인자(UMI))를 함유하는 기록 태그를, 바코드화된 비드를 사용한 유화, 또는 조합 색인화를 포함하는, 당해 분야에 공지되고 본원에 개시된 다수의 방법을 사용하여 단백질에 부착시킨다. 단일 세포 단백질을 함유하는 중합된 겔 비드를 또한 기록 태그의 첨가 후 프로테이나제 분해물에 적용시켜 펩타이드 서열분석에 적합한 기록 태그 표지된 펩타이드를 생성한다. 특정의 구현예에서, 중합체 매트릭스를 트리스(2-카복시에틸)포스핀(TCEP) 또는 디티오트레이톨(DTT)과

같은 환원제에 대한 노출시 이황화물 가교-결합된 중합체와 같은 적절한 첨가제 속에 용해되도록 설계할 수 있다.

도 39의 A 내지 **E. 이기능성 N-말단 아미노산(NTAA) 개질인자 및 키메라 절단 시약을 사용한 아미노산 절단 반응의 향상.** **(A)** 및 **(B)** 고체-상 기질에 부착된 펩타이드를 바이오틴-페닐 이소티오시아네이트(PITC)와 같은 이기능성 NTAA 변형인자(modifier)로 변형시킨다. **(C)** 저 친화성 에드마나제(> μM Kd)를 스트렙타비딘-에드마나제 키메라 단백질을 사용하여 바이오틴-PITC 표지된 NTAA로 보충한다. **(D)** 에드마나제 절단 효능은 바이오틴-스트렙타비딘 상호작용의 결과로서 효과적인 국소 농도에 있어서의 증가로 인하여 크게 개선된다. **(E)** 절단된 바이오틴-PITC 표지된 NTAA 및 관련된 스트렙타비딘-에드마나제 키메라 단백질은 절단 후 확산 제거된다. 다수의 기타 생물접합 보충 전략을 또한 사용할 수 있다. 아지드 변형된 PITC는 시판되며(4-아지도페닐 이소티오시아네이트, Sigma), 아지드-PITC의 알킨-바이오틴과의 클릭 화학 반응을 통해 바이오틴-PITC와 같은 PITC의 다른 생물접합체로의 다수의 단순한 변환을 허용한다.

도 40의 A 내지 **I: 단백질 분해물(겔 비드 속에 캡슐화될 수 있다)로부터 C-말단 기록 태그-표지된 펩타이드의 생성.** **(A)** 변성된 폴리펩타이드를 산 무수물과 반응시켜 라이신 잔기를 표지한다. 일 구현예에서, 알킨(mTet)-치환된 시트라콘상 무수물 + 프로피온산 무수물의 혼합물을 사용하여 라이신을 mTet로 표지한다.(줄무니 직사각형으로 나타냄). **(B)** 결과는 프로피온산 그룹으로 차단된 라이신의 분획(폴리펩타이드 쇄 위에 사각형으로 나타냄)을 지닌 알킨(mTet)-표지된 폴리펩타이드이다. 알킨(mTet) 모이어티는 클릭-화학 기반의 DNA 표지화에 유용하다. **(C)** DNA 태그(채워진 사각형으로 나타냄)를 알킨 또는 mTet 모이어티 각각에 대해 아지드 또는 *트랜스*-사이클로옥텐(TCO) 표지를 사용하여 클릭 화학으로 부착시킨다. **(D)** 바코드 및 스페이서(Sp) 서열 및 보편적인 프라이밍 서열과 같은 기능성 요소들을 도 31에 나타낸 바와 같은 프라이머 연장 단계를 사용하여 DNA 태그에 첨부함으로써 기록 태그-표지된 폴리펩타이드를 생산한다. 바코드는 샘플 바코드, 분배 바코드, 구획 바코드, 공간 배치 바코드, 등, 또는 이의 어떠한 조합일 수 있다. **(E)** 수득되는 기록 태그-표지된 폴리펩타이드는 프로테아제를 사용하거나 화학적으로 기록 태그-표지된 펩타이드로 단편화한다. **(F)** 설명을 위해 2개의 태그로 표지된 펩타이드 단편을 나타낸다. **(G)** 기록 태그 내에 보편적인 프라이밍 서열에 대해 상보성인 보편적인 프라이밍 서열을 포함하는 DNA 태그를 펩타이드의 C-말단에 연결한다. C-말단 DNA 태그는 또한 펩타이드를 표면에 접합시키기 위한 모이어티를 포함한다. **(H)** C-말단 DNA 태그 상의 상보성 보편적인 프라이밍 서열 및 확률적으로 선택된 기록 태그를 어닐링한다. 분자내 프라이머 연장 반응을 사용하여 기록 태그로부터의 정보를 C-말단 DNA 태그로 이전한다. **(I)** 펩타이드 상의 내부 기록 태그를 말레산 무수물을 통해 라이신 잔기에 커플링시키며, 이러한 커플링은 산성 pH에서 가역성이다. 내부 기록 태그를 산성 pH에서 펩타이드의 라이신 잔기로부터 절단하여, C-말단 기록 태그를 남긴다. 새로이 노출된 라이신 잔기를 프로피온산 무수물과 같은 가수분해할 수 없는 무수물로 임의 차단할 수 있다.

도 41. NGPS 검정의 바람직한 구현예에 대한 작업흐름.

도 42의 A 내지 **D. NGPS 서열분석 검정의 예시적인 단계.** 기록 태그-표지되고, 표면 결합된 펩타이드 상에서 N-말단 아미노산(NTAA) 아세틸화 또는 아미드화 단계가 NTAA 결합제가 아세틸화된 NTAA 또는 천연 NTAA에 결합하도록 가공되었는지의 여부에 따라, NTAA 결합제에 의한 결합 전 또는 후에 일어날 수 있다. 첫번째의 경우에, **(A)** 펩타이드는 NTAA에서 아세트산 무수물을 사용한 화학적 수단에 의해 또는 N-말단 아세틸트랜스퍼라제(NAT)를 사용하여 효소적으로 NTAA에서 초기에 아세틸화된다. **(B)** NTAA는 가공된 안티칼린, 아미노아실 tRNA 신테타제(aaRS), ClpS 등과 같은 NTAA 결합제에 의해 인식된다. DNA 암호화 태그는 결합제에 부착되며 특수한 NTAA 결합제를 확인하는 바코드 암호기 서열을 포함한다. **(C)** NTAA 결합제에 의한 아세틸화된 NTAA의 결합 후, DNA 암호화 태그는 상보성 서열을 통해 기록 태그로 일시적으로 어닐링되며 암호화 태그 정보는 폴리머라제 연장을 통해 기록 태그로 이전된다. 대안적인 구현예에서, 기록 태그 정보는 폴리머라제 연장을 통해 암호화 태그로 이전된다. **(D)** 아세틸화된 NTAA를 아세틸화된 펩타이드로부터 말단 아세틸화된 아미노산의 가수분해를 촉매하는 가공된 아실펩타이드 하이드롤라제(APH)로부터 절단한다. 아세틸화된 NTAA의 절단 후, 주기는 새로이 노출된 NTAA의 아세틸화로 시작하여 자체적으로 반복한다. N-말단 아세틸화는 예시적인 방식의 NTAA 변형/절단으로서 사용되지만, 구아닐 모이어티와 같은 다른 N-말단 모이어티도 절단 화학에서 수반된 변화로 치환할 수 있다. 구아니딘화가 사용되는 경우, 구아닐화된 NTAA는 0.5 내지 2% NaOH 용액(참고: Hamada, 2016, 이의 전문은 본원에 참고로 포함됨)을 사용하여 온화한 조건 하에서 절단될 수 있다. APH는 차단된 펩타이드로부터 Nα-아세틸화된 아미노산의 제거를 촉매할 수 있으며 이는 프롤릴 올리고펩티다제(POP) 계열(clan SC, 계열 S9)에 속한다. 이는 진핵, 세균 및 고세균 세포에서 N-말단적으로 아세틸화된 단백질의 중요한 조절인자이다.

도 43의 A 및 B. **예시적인 기록 태그 - 암호화 태그 설계 특징.** **(A)** 예시적인 기록 태그 관련된 단백질(또는

펩타이드) 및 관련된 암호화 태그를 지닌 결합된 결합제(예컨대, 안티칼린)의 구조. 티미딘(T) 염기를 암호화 태그 위의 스페이서 (Sp')와 바코드(BC') 서열 사이에 삽입하여 프라이머 연장 반응에서 확률적인 비-주형화된 3' 말단 아데노신(A) 첨가를 수용한다. **(B)** DNA 암호화 태그를 SpyCatcher-SpyTag 단백질-펩타이드 상호작용을 통해 결합제 (예컨대, 안티칼린)에 부착한다.

도 44의 **A** 내지 **E. 기록 태그 (A) 및 (B)에 대한 절단제의 하이브리드화를 사용한 NTAA 절단 반응의 향상.** 고체 상 기질(예컨대, 비드)에 부착된 기록 태그-표지된 펩타이드를 NTAA(Mod)에서 예컨대, PITC, DNP, SNP, 아세틸 개질화제, 구아니딜화 등을 사용하여 변형시키거나 표지한다. **(C)** 절단 효소(예컨대, 아실펩타이드 하이드롤라제(APH), 아미노 펩티다제(AP), 에드마나제 등)을 기록 태그 위에서 보편적인 프라이밍 서열에 대해 상보성인 보편적인 프라이밍 서열을 포함하는 DNA 태그에 부착한다. 절단 효소를 절단 효소의 DNA 태그 및 기록 태그 위에서 상보성인 보편적인 프라이밍 서열의 하이브리드화를 통해 변형된 NTAA에 보충한다. **(D)** 이러한 하이브리드화 단계는 NTAA에 대한 절단 효소의 효과적인 친화성을 크게 개선시킨다. **(E)** 절단된 NTAA는 확산 제거되며 관련된 절단 효소는 하이브리드화된 DNA 태그에 의해 제거될 수 있다.

도 45. 펩타이드 리가제 + 프로테아제 + 디아미노펩티다제를 사용한 사이클릭 분해 펩타이드 서열분석. 부텔라제 I은 TEV-부텔라제 I 펩타이드 기질(TENLYFQNHV, 서열 번호: 132)을 질의 펩타이드(query peptide)의 NTAA에 연결한다. 부텔라제는 펩타이드 기질의 C-말단에서 NHV 모티프(motif)를 필요로 한다. 연결 후, 담배 에치 바이러스(Tobacco Etch Virus: TEV) 프로테아제를 사용하여 글루타민(Q) 잔기 이후 키메라 펩타이드 기질을 절단하여, 질의 펩타이드의 N-말단에 부착된 아스파라긴(N) 잔기를 갖는 키메라 펩타이드를 남긴다. N-말단으로부터 2개의 아미노산 잔기를 절단하는 이다미노펩티다제(DAP) 또는 디펩티딜-펩티다제는 질의 펩타이드에서 아스파라긴 잔기(N) 및 원래의 NTAA를 효과적으로 제거하는 2개의 아미노산에 의해 N-첨가된 질의 펩타이드를 단축한다. 새로이 노출된 NTAA는 본원에 제공된 결합제를 사용하여 판독하며, 이후 전체 주기를 서열분석된 "n개 펩타이드에 대해 "n" 회 반복한다. 스트렙타비딘-DAP 금속효소 키메라 단백질의 사용 및 바이오틴 모이어티를 N-말단 아스파라긴 잔기에 테더링(tethering)하는 것은 DAP 가공성을 제어하도록 할 수 있다.

상세한 설명

본원에 구체적으로 정의되지 않은 용어는 개시내용 및 본문의 측면에서 당해 분야의 숙련가에게 제공될 수 있는 의미를 제공할 수 있다. 명세서에 사용된 바와 같이, 그러나, 달리 명시하지 않는 한, 용어는 나타낸 의미를 갖는다.

I. 서론

본 개시내용은 부분적으로 단백질 및 펩타이드 특성화 및 서열 분석에 직접 적용되는, 고-병렬, 고 처리량의 디지탈 거대분자 특성화 및 정량화의 방법을 제공한다(참고, 도 1b, 도 2a). 본원에 기술된 방법은 핵산 분자 또는 서열분석가능한 중합체 형태의 확인 정보를 지닌 암호화 태그를 포함하는 결합제를 사용하며, 여기서 결합제는 목적한 거대분자와 상호작용한다. 각각의 주기가 고체 지지체에 위에 고정된, 다수의 거대분자를 노출시키고, 바람직하게는 혼주된 샘플을 나타내는 과정을 포함하는, 다수의 연속된 결합 주기를 수행한다. 각각의 결합 주기 동안에, 거대 분자에 결합하는 각각의 결합체의 동일성, 및 임의로 결합 주기 수는 결합제 암호화 태그로부터의 정보를 거대 분자와 동시-국재화된 기록 태그에 이전시킴으로써 기록된다. 대안적인 구현예에서, 관련된 거대분자에 대한 정보를 확인함을 포함하는 기록 태그로부터의 정보는 결합된 결합제의 암호화 태그(예컨대, 연장된 암호화 태그를 형성하기 위해) 또는 제3의 "디-태그" 작제물에 이전될 수 있다. 결합 사건의 다수의 주기는 거대 분자와 동시-국재화된 기록 태그 상의 역사적인 결합 정보를 확립함으로써, 제공된 거대분자에 대한 일시적인 결합 역사를 나타내는 동시-선형 순서로 다수의 암호화 태그를 포함하는 연장된 기록 태그를 생산한다. 또한, 주기-특이적인 암호화 태그를 사용하여 각각의 주기로부터 정보를 추적함으로써, 주기가 몇 가지 이유로 생략되는 경우, 연장된 기록 태그가 후속 주기에서 정보를 지속적으로 수집하고 잃어버린 정보를 갖는 주기를 확인할 수 있다.

대안적으로, 암호화 태그로부터의 정보를 기록 태그로의 쓰기 또는 이전하는 대신, 관련된 거대분자에 대한 정보를 연장된 암호화 태그를 형성하는 암호화 태그에 대해서 또는 제3의 디-태그 작제물에 대해서 확인함을 포함하여 정보를 기록 태그로부터 이전시킬 수 있다. 수득되는 연장된 암호화 태그 또는 디-태그는 후속적인 서열분석을 위한 각각의 결합 주기 후 수집할 수 있다. 바코드(예컨대, 분배 태그, 구획 태그, 샘플 태그, 분획 태그, UMI, 또는 이의 어떠한 조합)를 포함하는 기록 태그 상의 정보를 확인하는 것을 사용하여 원래의 거대분자에 대해 역으로 연장된 암호화 태그 또는 디-태그 서열 판독물을 맵핑할 수 있다. 이러한 방식으로 거대분자의 결합 역사의 핵산 암호화된 라이브러리 묘사가 생성된다. 이러한 핵산 암호화된 라이브러리는 증폭시키고, 매우

고-처리량의 차세대 디지탈 서열분석 방법을 사용하여 분석함으로써 작동당 수만 내지 수천만개의 분자가 분석되도록 할 수 있다. 결합 정보의 핵산 암호화된 라이브러리의 생성은 이것이 하이브리드화를 사용하도록 하는 DNA-기반 기술에 의한 농축, 삭감, 및 정상화를 가능하도록 한다는 점에서 다른 방식으로 유용하다. 이러한 DNA-기반 방법은 단백질 라이브러리와 같은, 다른 유형의 거대분자 라이브러리의 직접적인 조작에 이용가능한 것보다 용이하게 및 신속하게 확장가능하며 주문가능하고, 보다 비용-효율적이다. 따라서, 결합 정보의 핵산 암호화된 라이브러리는 하나 이상의 기술에 의한 서열분석 전에 가공되어 서열의 묘사를 농축시키고/시키거나 감하고/감하거나 표준화할 수 있다. 이는 목적한 최대 정보가 이의 개개 구성원이 많은 차수의 크기에 걸쳐 초기에 변할 수 있는 매우 큰 라이브러리로부터 훨씬 더 효과적으로, 신속하게 및 비용-효율적이 되도록 할 수 있다. 중요하게는, 라이브러리 묘사를 조작하기 위한 이러한 핵산-기반 기술은 통상의 방법보다 더 직각이어서 이들과 함께 사용될 수 있다. 예를 들면, 알부민과 같은, 일반적이고, 매우 풍부한 단백질을 단백질-기반 방법을 사용하여 공제할 수 있으며, 이는 대부분의 그러나 전부가 아닌 바람직하지 않은 단백질을 제거할 수 있다. 후속적으로, 연장된 기록 태그 라이브러리의 알부민-특이적인 구성원이 추가로 공제됨으로써, 보다 완전한 전체적인 공제가 달성될 수 있다.

일 국면에서, 본 개시내용은 에드만-유사 분해 시도를 사용하는 펩타이드 서열분석을 위한 고-병렬화된 시도를 제공하여, DNA 기록 태그-표지된 펩타이드의 큰 수집물(예컨대, 수백만 내지 수십억개)로부터의 서열분석을 허용한다. 이러한 기록 태그 표지된 펩타이드는 단백질 샘플의 단백질분해성 소화 또는 제한된 가수분해로부터 기원하며, 기록 태그 표지된 펩타이드는 기질에서 적절한 분자간 이격(spacing)으로 서열분석 기질(예컨대, 다공성 비드)에 무작위로 고정된다. 페닐티오카바모닐(PTC), 디니트로페놀(DNP), 설포닐 니트로페놀(SNP), 단실, 7-메톡시 쿠마린, 아세틸, 또는 구아니딜과 같은, NTAA 절단 반응을 촉매하거나 보충하는 작은 화학적 모이어티를 사용한 펩타이드의 N-말단 아미노산(NTAA) 잔기의 변형은 에드만-유사 분해 공정의 사이클릭 제어를 허용한다. 변형된 화학적 모이어티는 또한 동종의 NTAA 결합제에 대한 향성된 결합 친화성을 제공할 수 있다. 각각의 고정된 펩타이드의 변형된 NTAA는 암호화 태그를 포함하는 동종의 NTAA 결합제를 결합시키고, 암호화 태그로부터의 암호화 태그 정보(예컨대, 결합제에 대한 확인 정보를 제공하는 암호기 서열)를 펩타이드의 기록 태그로 이전시킴(예컨대, 프라이머 연장 또는 연결)으로써 확인된다. 후속적으로, 변형된 NTAA는 화학적 방법 또는 효소적 수단에 의해 제거된다. 특정의 구현예에서, 효소(예컨대, 에드마나제)는 가공되어 변형된 NTAA의 제거를 촉매한다. 다른 구현예에서, 아미노펩티다제 또는 아실 펩타이드 하이드롤라제와 같은, 천연적으로 존재하는 엑소펩티다제를 가공하여 적합한 화학적 변형의 존재하에서 말단 아미노산만을 절단할 수 있다.

II. 정의

다음의 설명에서, 특정의 구체적인 세부사항을 다양한 구현예의 완전한 이해를 제공하기 위해 제시한다. 그러나, 당해 분야의 숙련가는 본 발명의 화합물이 이들 세부사항없이도 제조되어 사용될 수 있음을 이해할 것이다. 다른 예에서, 잘-공지된 구조는 나타나 있지 않고 구현예의 불필요한 애매한 설명을 피하기 위해 상세히 기술된다. 내용이 달리 필요로하지 않는 한, 이후에 기술되는 명세서 및 청구범위 전체에서, 단어 "포함하다" 및 이의 변형, 예를 들면, "포함하며", 및 "포함하는"은 개방된, 포괄적인 의미, 즉, "포함하나, 이에 한정되지 않는"으로 해석되어야 한다. 또한, 용어 "포함하는"(및 "포함하다(comprise, comprises)" 또는 "갖는" 또는 "포괄하는"과 같은 관련 용어)은 다른 특정 구현예, 예를 들면, 본원에 기술된 물질의 특정 조성, 조성물, 방법, 또는 과정 등의 구현예가 기술된 특징"으로 이루어지거나" "필수적으로 이루어질" 수 있음을 배제하는 것으로 의도되지 않는다. 본원에 제공된 제목은 단지 편의를 위한 것이며, 청구된 구현예의 영역 또는 의미를 해석하는 것이 아니다.

본 명세서 전체에서 "하나의 구현예" 또는 "일 구현예"에 대한 참고는 구현예와 관련하여 기술된 특수한 특징, 구조 또는 특성이 적어도 하나의 구현예에 포함됨을 의미한다. 따라서, 본 명세서 전체에서 다양한 위치에 어구 "일 구현예에서" 또는 구현예에서"의 출현은 필수적으로 동일한 구현예 전부를 지칭하는 것은 아니다. 또한, 특수한 특징, 구조, 또는 특성은 하나 이상의 구현예에서 어떠한 적합한 방식으로 조합될 수 있다.

본원에 사용된 바와 같은, 단수 형("a," "an" 및 "the")은 내용에서 달리 명확하게 나타내지 않는 한 복수 참고를 포함한다. 따라서, 예를 들어 "하나의 펩타이드"에 대한 참고는 하나 이상의 펩타이드, 또는 펩타이드의 혼합물을 포함한다. 또한, 및 구체적으로 기술하거나 내용으로부터 명확하지 않은 한, 본원에 사용된 바와 같은, 용어 "또는"은 포괄적인 것으로 이해되며 "또는"과 "및" 둘 다를 포함하는 것으로 이해된다.

본원에 사용된 바와 같은, 용어 "거대분자"는 보다 작은 소단위로 구성된 큰 분자를 포함한다. 거대분자의 예는 펩타이드, 폴리펩타이드, 단백질, 핵산, 탄수화물, 지질, 마크로사이클(macrocycle)을 포함하나, 이에 한정되지

않는다. 거대분자는 또한 서로 공유결합으로 연결된, 2개 이상의 유형의 거대분자의 조합(예컨대, 핵산에 연결된 펩타이드)으로 구성된 키메라 거대분자를 포함한다. 거대분자는 또한 2개 이상의 거대 분자의 비-공유결합성 복합체로 구성된 "거대분자 조립체"를 포함할 수 있다. 거대분자 조립체는 동일한 유형의 거대분자(예컨대, 단백질-단백질) 또는 2개 이상의 상이한 유형의 거대분자(예컨대, 단백질-DNA)로 구성될 수 있다.

본원에 사용된 바와 같은, 용어 "펩타이드"는 펩타이드, 폴리펩타이드 및 단백질을 포함하며, 펩타이드 결합에 의해 결합된 2개 이상의 아미노산의 쇄를 포함하는 분자를 지칭한다. 일반적인 용어에서, 20 내지 30개 이상의 아미노산을 갖는 펩타이드는 일반적으로 폴리펩타이드로 지칭되며, 50개 이상의 아미노산을 갖는 펩타이드는 단백질로서 일반적으로 지칭된다. 펩타이드의 아미노산은 대부분 전형적으로 L-아미노산이지만, 또한 D-아미노산, 변형된 아미노산, 아미노산 동족체, 아미노산 모사체(mimetics), 또는 이의 어떠한 조합일 수 있다. 펩타이드는 천연적으로 존재하거나, 합성적으로 생산되거나, 재조합적으로 발현될 수 있다. 펩타이드는 또한 아미노산 쇄를 변형시키는 추가의 그룹, 예를 들면, 해독후 변형을 통해 첨가된 기능성 그룹을 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "아미노산"은 펩타이드의 단량체성 소단위로서 제공되는, 각각의 아미노산에 대해 특이적인 아미노 그룹, 카복실산 그룹, 및 측쇄를 포함하는 유기 화합물을 지칭한다. 아미노산은 20개의 표준, 천연적으로 존재하거나 기본형의 아미노산 및 비-표준 아미노산을 포함한다. 표준의, 천연적으로 존재하는 아미노산은 알라닌(A 또는 Ala), 시스테인(C 또는 Cys), 아스파르트산(D 또는 Asp), 글루탐산(E 또는 Glu), 페닐알라닌(F 또는 Phe), 글리신(G 또는 Gly), 히스티딘(H 또는 His), 이소루이신(I 또는 Ile), 라이신(K 또는 Lys), 루이신(L 또는 Leu), 메티오닌(M 또는 Met), 아스파라긴(N 또는 Asn), 프롤린(P 또는 Pro), 글루타민(Q 또는 Gln), 아르기닌(R 또는 Arg), 세린(S 또는 Ser), 트레오닌(T 또는 Thr), 발린(V 또는 Val), 트립토판(W 또는 Trp), 및 타이로신(Y 또는 Tyr)을 포함한다. 아미노산은 L-아미노산 또는 D-아미노산일 수 있다. 비-표준 아미노산은 변형된 아미노산, 아미노산 동족체, 아미노산 모사체, 비-표준 단백질원성 아미노산, 또는 천연적으로 존재하거나 화학적으로 합성된 비-단백질원성 아미노산일 수 있다. 비-표준 아미노산의 예는 셀레노시스테인, 피롤라이신, N-포르밀메티오닌, β-아미노산, 호모-아미노산, 프롤린 및 피루브산 유도체, 3-치환된 알라닌 유도체, 글리신 유도체, 환-치환된 페닐알라닌 및 타이로신 유도체, 선형 코어 아미노산, N-메틸 아미노산을 포함하나, 이에 한정되지 않는다.

본원에 사용된 바와 같은, 용어 "해독후 변형"은 리보소옴에 의한 이의 해독이 완료된 후 펩타이드 상에서 일어나는 변형을 지칭한다. 해독후 변형은 공유결합성 변형 또는 효소적 변형일 수 있다. 해독후 변형의 예는 아실화, 아세틸화, 알킬화(메틸화 포함), 바이오티닐화, 부티릴화, 카바밀화, 카보닐화, 탈아미드화, 탈이민화, 디프타미드 형성, 이황화물 브릿지 형성, 엘리미닐화(eliminylation), 플라빈 부착, 포르밀화, 감마-카복실화, 글루타밀화, 글리실화, 글리코실화, 글리피화(glypiation), 헴(heme) C 부착, 하이드록실화, 하이푸신 형성(hypusine formation), 요오드화, 이소프레닐화, 지질화, 리포일화, 말로닐화, 메틸화, 미리스토일화, 산화, 팔미토일화, 페길화, 포스포판테테이닐화, 인산화, 프레닐화, 프로피오닐화, 레티닐리덴 스키프 염기 형성(retinylidene Schiff base formation), S-글루타티오닐화, S-니트로실화, S-설페닐화, 셀렌화, 석시닐화, 설핀화, 우비퀴틴화(ubiquitination), 및 C-말단 아미드화를 포함하나, 이에 한정되지 않는다. 해독후 변형은 펩타이드의 아미노 말단 및/또는 카복실 말단의 변형을 포함한다. 말단 아미노 그룹의 변형은 데스-아미노, N-저급 알킬, N-디-저급 알킬, 및 N-아실 변형을 포함하나, 이에 한정되지 않는다. 말단 카복시 그룹의 변형은 아미드, 저급 알킬 아미드, 디알킬 아미드, 및 저급 알킬 에스테르 변형(예컨대, 여기서 저급 알킬은 C_1-C_4 알킬이다)을 포함하나, 이에 한정되지 않는다. 해독후 변형은 또한 아미노 및 카복시 말단 사이에 속하는 아미노산의, 상술한 것과 같은 그러나 이에 한정되지 않는 변형을 포함한다. 용어 해독 후 변형은 또한 하나 이상의 검출가능한 표지를 포함하는 펩타이드 변형을 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "결합제"는 거대분자 또는 거대 분자의 성분 또는 특징과 결합하거나, 연합하거나, 단일화하거나, 인식하거나, 또는 조합하는 핵산 분자, 펩타이드, 폴리펩타이드, 단백질, 탄수화물, 또는 거대 분자 또는 거대분자의 성분 또는 특징에 결합하거나, 연합하거나, 단일화하거나, 인식하거나 또는 조합하는 소 분자를 지칭한다. 결합제는 거대분자 또는 거대분자의 성분 또는 특징과 공유결합성 연합체 또는 비-공유결합성 연합체를 형성할 수 있다. 결합제는 또한 핵산 분자-펩타이드 키메라 결합제 또는 탄수화물-펩타이드 키메라 결합제와 같은 2개 이상의 유형의 분자로 구성된 키메라 결합제일 수 있다. 결합제는 천연적으로 존재하거나, 합성적으로 생산되거나, 재조합적으로 발현된 분자일 수 있다. 결합제는 거대 분자의 단일 단량체 또는 소단위(예컨대, 펩타이드의 단일 아미노산)에 결합하거나 거대분자(예컨대, 디-펩타이드, 트리-펩타이드, 또는 보다 긴 펩타이드의 보다 높은 차수의 펩타이드, 폴리펩타이드, 또는 단백질 분자)의 다수의 연결된 소단위에 결합할 수 있다. 결합제는 선형 분자 또는 3차원 구조(또는 구조로 지칭됨)를 가진 분자에 결합할 수

있다. 예를 들면, 항체 결합제는 선형 펩타이드, 폴리펩타이드, 또는 단백질에 결합하거나, 구조 펩타이드, 폴리펩타이드, 또는 단백질에 결합할 수 있다. 결합제는 펩타이드, 폴리펩타이드, 또는 단백질 분자의 N-말단 펩타이드, C-말단 펩타이드, 또는 중간에 있는(intervening) 펩타이드에 결합할 수 있다. 결합제는 펩타이드 분자의 N-말단 아미노산, C-말단 아미노산, 또는 중간에 있는 아미노산에 결합할 수 있다. 결합제는 바람직하게는 변형되지 않거나 표지되지 않은 아미노산에 걸쳐 화학적으로 변형되거나 표지된 아미노산에 결합할 수 있다. 예를 들면, 결합제는 바람직하게는 상기 모이어티를 지니지 않은 아미노산에 걸쳐서 아세틸 모이어티, 구아닐 모이어티, 단실 모이어티, PTC 모이어티, DNP 모이어티, SNP 모이어티 등으로 변형된 아미노산에 결합할 수 있다. 결합제는 펩타이드 분자의 해독 후 변형에 결합할 수 있다. 결합제는 거대 분자의 성분 또는 특징에 대해 선택적인 결합을 나타낼 수 있다(예컨대, 결합제는 20개의 가능한 천연 아미노산 잔기 중 하나에 선택적으로 결합할 수 있으며 다른 19개의 천연 아미노산 잔기에 매우 낮은 친화성으로 결합하거나 또는 결코 결합하지 않을 수 있다). 결합제는 선택적인 결합을 거의 나타내지 않을 수 있으며, 여기서 결합제는 거대 분자의 다수의 성분 또는 특징과 결합할 수 있다(예컨대, 결합제는 2개 이상의 상이한 아미노산 잔기에 대해 유사한 친화성으로 결합할 수 있다). 결합제는 암호화 태그를 포함하며, 이는 링커에 의해 결합제에 결합된다.

본원에 사용된 바와 같은, 용어 "링커"는 하나 이상의 뉴클레오타이드, 뉴클레오타이드 유사체, 아미노산, 펩타이드, 폴리펩타이드, 또는 2개의 분자를 결합시키는데 사용된 비-뉴클레오타이드 화학적 모이어티를 지칭한다. 링커는 결합제와 암호화 태그, 기록 태그와 거대분자(예컨대, 펩타이드), 거대분자와 고체 지지체, 기록 태그와 고체 지지체 등을 결합시키는데 사용될 수 있다. 특정의 구현예에서, 링커는 효소 반응 또는 화학 반응(예컨대, 클릭 화학)을 통해 2개의 분자를 결합시킨다.

본원에 사용된 바와 같은, 용어 "프로테오믹스"는 세포, 조직 및 체액 내에서 프로테오믹스의 정량적 분석, 및 세포내 및 조직내에서 프로테오믹스의 상응하는 공간적 분포를 지칭한다. 또한, 프로테오믹스 연구는 생물학 및 정의된 생물학적 또는 화학적 자극의 함수로서 시간내 지속적으로 변화하는, 프로테옴의 역학적 상태를 포함한다.

본원에 사용된 바와 같은, 용어 "비-동종의 결합제"는 상응하는 거대분자, 특징, 성분, 또는 소단위에 대해 고친화성으로 결합하는 "동종의 결합제"와 비교하여 특수한 결합 주기에서 조사하는(interrogate) 거대분자, 특징, 성분, 또는 소단위에 대해 결합하지 않거나 낮은 친화성으로 결합하는 결합제를 지칭한다. 예를 들면, 펩타이드 분자의 타이로신 잔기가 결합 반응에서 조사되는 경우, 비-동종의 결합제는 타이로신 잔기에 대해 낮은 친화성으로 결합하거나 전혀 결합하지 않음으로써, 비-동종의 결합제가 동종의 결합제로부터의 암호화 태그 정보를 기록 태그로 이전하기에 적합한 조건 하에서 암호화 태그 정보를 기록 태그로 효과적으로 이전시키지 않는 것이다. 대안적으로, 펩타이드 분자의 타이로신 잔기가 결합 반응에서 조사되는 경우, 비-동종의 결합제는 타이로신 잔기에 대해 저 친화성으로 결합하거나 전혀 결합하지 않음으로써, 기록 태그 정보가 연장된 기록 태그보다는 연장된 암호화 태그를 포함하는 구현예에 대해 적합한 조건 하에서 암호화 태그를 효율적으로 이전하지 않는다.

유리 아미노 그룹을 갖는 펩타이드 쇄의 한쪽 말단에서 말단 아미노산은 본원에서 "N-말단 아미노산"(NTAA)으로 언급된다. 유리 카복실 그룹을 갖는 쇄의 다른 말단에서 말단 아미노산은 본원에서 "C-말단 아미노산"(CTAA)으로 지칭된다. 펩타이드를 구성하는 아미노산는 순차적으로, 길이내 "*n*"번 아미노산인 펩타이드를 사용하여 순차적으로 번호매김될 수 있다. 본원에 사용된 바와 같은, NTAA는 *n*번째 아미노산(또한 본원에서 "*n*번 NTAA"로 지칭됨)으로 고려된다. 이러한 명명법을 사용하면, 다음 아미노산은 *n-1*번 아미노산이고, 이후 *n-2*번 아미노산이며, N-말단 끝으로부터 C-말단 끝까지 펩타이드의 길이가 이와 같이 정해진다. 특정의 구현예에서, NTAA, CTAA, 또는 둘 다는 화학적 모이어티로 변형되거나 표지된다.

본원에 사용된 바와 같은, 용어 "바코드"는 거대분자 (예컨대, 단백질, 폴리펩타이드, 펩타이드), 결합제, 결합 주기로부터의 결합제의 세트, 샘플 거대분자, 샘플의 세트, 구획(예컨대, 소적, 비드, 또는 분리된 위치) 내 거대분자, 구획의 세트내 거대분자, 거대분자의 분획, 거대분자 분획의 세트, 공간적 영역 또는 공간적 영역의 세트, 거대분자의 라이브러리, 또는 결합제의 라이브러리에 대한 유일한 확인인자 태그 또는 원래의 정보를 제공하는, 염기가 약 2 내지 약 30개(예컨대, 2, 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 또는 30개 염기)인 핵산 분자를 지칭한다. 바코드는 인공 서열 또는 천연적으로 존재하는 서열일 수 있다. 특정의 구현예에서, 바코드의 집단내 각각의 바코드는 상이하다. 다른 구현예에서, 바코드의 집단내 바코드 중 일부는 상이한데, 예컨대, 바코드의 집단 속의 바코드의 적어도 약 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 또는 99%는 상이하다. 바코드의 집단은 무작위로 생성되거나 비-무작위로 생성될 수 있다. 특정의 구현예에서, 바코드의 집

단은 오류 수정 바코드이다. 바코드를 사용하여 멀티플렉스화된 서열분석 데이타를 계산적으로 디컨벌루트(deconvolute)하고 개개의 거대분자, 샘플, 라이브러리 등으로부터 기원한 서열 판독물을 확인할 수 있다. 바코드를 또한 향상된 맵핑을 위해 작은 구획으로 분포된 거대분자의 수집의 디컬벌루션(deconvolution)에 사용할 수 있다. 예를 들면, 펩타이드를 프로테옴으로 역으로 맵핑하는 대신에, 펩타이드를 이의 원래의 단백질 분자 또는 단백질 복합체로 역으로 맵핑할 수 있다.

"샘플 태그"로도 지칭되는 "샘플 바코드"는 이러한 샘플로부터 거대분자가 유도되는 것을 확인한다.

"공간 바코드"는 거대 분자가 기원하는 2-D 또는 3-D 조직 단면의 영역을 지칭한다. 공간 바코드는 분자 단면에서 분자 병리학에 사용될 수 있다. 공간 바코드는 조직 단면(들)로부터의 다수의 샘플 또는 라이브러리의 멀티플렉스 서열분석을 허용한다.

본원에 사용된 바와 같은, 용어 "암호화 태그" 2 내지 100 사이의 어떠한 정수를 포함하고, 이들 사이에, 이의 관련된 결합제에 대한 확인 정보를 포함하는 약 2개의 염기 내지 약 100개의 염기의 핵산 분자를 지칭한다. "암호화 태그"는 또한 "서열분석가능한 중합체"로부터 제조될 수 있다(참고: 예컨대, Niu et al., 2013, Nat. Chem. 5:282-292; Roy et al., 2015, Nat. Commun. 6:7237; Lutz, 2015, Mzcromolecules 48:4759-4767; 이들 각각은 이의 전문이 본원에 참고로 포함된다). 암호화 태그는 암호기 서열을 포함하며, 이는 한 면에서 하나의 스페이서로 임의로 플랭킹되고 다른 면에서 스페이서로 플랭킹된다. 암호화 태그는 또한 임의의 UMI 및/또는 임의로 결합 주기-특이적인 바코드로 구성될 수 있다. 암호화 태그는 단일 가닥 또는 이중 가닥일 수 있다. 이중 가닥의 암호화 태그는 평활 말단, 오버행 말단(overhanging end), 또는 둘 다일 수 있다. 암호화 태그는 결합제에 직접 부착된 암호화 태그, 결합제에 직접 부착된 암호화 태그에 하이브리드화하는 상보성 서열(예컨대, 이중 가닥 암호화 태그의 경우), 또는 연장된 기록 태그 속에 존재하는 암호화 태그 정보를 지칭할 수 있다. 특정 구현예에서, 암호화 태그는 결합 주기 특이적인 스페이서 또는 바코드, 유일한 분자 확인인자, 보편적인 프라이밍 부위, 또는 이의 어떠한 조합을 추가로 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "암호기 서열" 또는 "암호기 바코드"는 이의 관련된 결합제에 대한 확인 정보를 제공하는 길이가 약 2개의 염기 내지 약 30개의 염기(예컨대, 2, 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 또는 30개의 염기)인 핵산 분자를 지칭한다. 특정 구현예에서, 암호기 서열은 이의 관련 결합제 및 결합제가 사용된 결합 주기에 대한 확인 정보를 제공한다. 다른 구현예에서, 암호기 서열은 암호화 태그내에서 별도의 결합 주기-특이적인 바코드와 조합된다. 대안적으로, 암호기 서열은 2개 이상의 상이한 결합제의 세트의 구성원에 속하는 것으로서 이의 관련된 결합제를 확인할 수 있다. 일부 구현예에서, 이러한 확인 수준은 분석 목적에 충분하다. 예를 들면, 아미노산에 결합하는 결합제를 포함하는 일부 구현예에서, 특수한 위치에서 아미노산 잔기를 명확하게 정의하기 보다는, 펩타이드가 특수한 위치에서 2개의 아미노산 중 하나를 포함하는 것을 아는 것이 충분할 수 있다. 다른 예에서, 일반적인 암호기 서열을 폴리클로날 항체에 사용하며, 이는 단백질 표적의 하나 이상의 에피토프를 인식하고 다양한 특이성을 갖는 항체의 혼합물을 포함한다. 다른 구현예에서, 암호기 서열이 가능한 결합제의 세트를 확인하는 경우, 순차적인 탈암호화 시도를 사용하여 각각의 결합제의 유일한 확인을 생산할 수 있다. 이는 반복된 결합 주기에서 제공된 결합제에 대한 암호기 서열을 변화시켜 달성할 수 있다(참고: Gunderson et al., 2004, Genome Res. 14:870-7). 다른 주기로부터 암호화 정보와 조합되는 경우, 각각의 결합 주기로부터 암호화 태그 정보를 부분적으로 확인하는 것은 결합제에 대한 유일한 확인인자를 생산하는데, 예컨대, 개개 암호화 태그(또는 암호기 서열)보다는 암호화 태그의 특수한 조합은 결합제에 대한 유일한 확인 정보를 제공한다. 바람직하게는, 결합제의 라이브러리내 암호기 서열은 동일하거나 유사한 수의 염기를 지닌다.

본원에 사용된 바와 같은 용어 "결합 주기 특이적인 태그", "결합 주기 특이적인 바코드" 또는 "결합 주기 특이적인 서열"은 특수한 결합 주기내에서 사용된 결합제의 라이브러리를 확인하는데 사용된 유일한 서열을 지칭한다. 결합 주기 특이적인 태그는 약 2개 염기 내지 약 8개 염기(예컨대, 2, 3, 4, 5, 6, 7, 또는 8개 염기) 길이를 포함할 수 있다. 결합 주기 특이적인 태그는 스페이서 서열의 부분, 암호기 서열의 부분, UMI의 부분, 또는 암호화 태그내 별도의 성분으로서, 결합제의 암호화 태그 내에 포함될 수 있다.

본원에 사용된 바와 같은, 용어 "스페이서"(Sp)는 기록 태그 또는 암호화 태그의 말단에 존재하는 약 1개 염기 내지 약 20개 염기(예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 염기)의 길이의 핵산 분자를 지칭한다. 특정 구현예에서, 스페이서 서열은 한쪽 말단 또는 양쪽 말단에서 암호기 서열을 플랭크한다. 거대분자에 대한 결합제의 결합 후, 이들의 관련된 암호화 태그와 기록 태그 각각에서 상보성 스페이서 서열 사이의 어닐링은 프라이머 연장 반응 또는 연결을 통한 결합 정보의 기록 태그, 암호화

태그, 또는 디-태그 작제물로의 이전을 허용한다. Sp'는 Sp에 상보성인 스페이서 서열을 지칭한다. 바람직하게는 결합제의 라이브러리내 스페이서 서열은 동일한 수의 염기를 지닌다. 일반적인(공유되거나 동일한) 스페이서를 결합제의 라이브러리에 사용할 수 있다. 스페이서 서열은 특수한 결합 주기에 사용된 결합제를 추적하기 위하여 "주기 특이적인" 서열을 가질 수 있다. 스페이서 서열(Sp)은 모든 결합 주기에 걸쳐 일정할 수 있으며, 특수한 부류의 결합주기에 대해 특이적이거나, 결합 주기 수 특이적일 수 있다. 거대분자 부류-특이적인 스페이서는 부류-특이적인 스페이서를 통해 후속적인 결합 주기내 동일한 부류의 거대분자를 인식하는 다른 결합제의 암호화 태그로 완성된 결합/연장 주기로부터의 연장된 기록 태그내에 존재하는 동종의 결합제의 암호화 태그 정보의 어닐링을 허용한다. 정확한 동종의 쌍의 순차적인 결합만이 상호작용하는 스페이서 요소 및 효과적인 프라이머 연장을 생성한다. 스페이서 서열은 기록 태그 내의 상보성 스페이서 서열에 어닐링하기에 충분한 수의 염기를 포함함으로써 프라이머 연장(또는 폴리머라제 연장으로 지칭됨) 반응을 개시할 수 있거나, 연결 반응을 위한 "스플린트(splint)"를 제공하거나, "점성 말단(sticky end)" 연결 반응을 매개할 수 있다. 스페이서 서열은 암호화 태그내에 암호기 서열보다 더 적은 수의 염기를 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "기록 태그"는 이것이 관련된 거대분자에 대한 확인 정보를 포함하는 핵산 분자 또는 서열분석가능한 중합체 분자를 지칭한다(참고: 예컨대, Niu et al., 2013, Nat. Chem. 5:282-292; Roy et al., 2015, Nat. Commun. 6:7237; Lutz, 2015, Macromolecules 48:4759-4767; 이들 각각은 이의 전문이 참고로 포함됨). 특정 구현예에서, 결합제가 거대분자에 결합한 후, 결합제에 연결된 암호화 태그로부터의 정보를 결합제가 거대분자에 결합하는 동안 거대분자와 관련된 기록 태그로 이전시킬 수 있다. 다른 구현예에서, 결합제가 거대분자에 결합한 후, 거대분자와 관련된 기록 태그로부터의 정보를 결합제가 거대 분자에 결합하는 동안 결합제에 연결된 암호화 태그로 이전시킬 수 있다. 재암호화 태그는 거대분자에 직접 연결되거나, 다기능성 링커를 통해 거대분자에 연결되거나, 고체 지지체에서 이의 근접성(또는 동시-국재화)로 인하여 거대분자와 관련될 수 있다. 기록 태그는 연결이 암호화 태그 정보를 기록 태그에 이전시키거나, 또는 역으로 사용된 방법과 혼용성인 한, 이의 5' 말단 또는 3' 말단을 통해 또는 내부 부위에서 연결될 수 있다. 기록 태그는 다른 기능성 성분, 예컨대, 보편적인 프라이밍 부위, 유일한 분자 확인인자, 바코드(예컨대, 샘플 바코드, 분획 바코드, 공간 바코드, 구획 태그 등), 암호화 태그의 스페이서 서열에 대해 상보성인 스페이서 서열, 또는 이의 어떠한 조합에 대해 상보성인 스페이서 서열을 추가로 포함할 수 있다. 기록 태그의 스페이서 서열은 바람직하게는 폴리머라제 연장을 사용하여 암호화 태그 정보를 기록 태그로 이전시키는 구현예에서 기록 태그의 3'-말단에 존재한다.

본원에 사용된 바와 같은, 용어 "폴리머라제 연장"으로 또한 언급되는, "프라이머 연장"은 핵산 폴리머라제(예컨대, DNA 폴리머라제)에 의해 촉매된 반응을 지칭하며, 이에 의해 상보성 쇄에 어닐링하는 핵산 분자(예컨대, 올리고뉴클레오타이드 프라이머, 스페이서 서열)는 주형으로서 상보성 쇄를 사용하여 폴리머라제에 의해 연장된다.

본원에 사용된 바와 같은, 용어 "유일한 분자 확인인자" 또는 "UMI"는 각각의 거대분자(예컨대, 펩타이드)에 대한 유일한 확인인자 태그 또는 UMI가 연결된 결합제를 제공하는 약 3 내지 약 40개 염기의 길이(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, 또는 40개 염기의 길이)의 핵산 분자를 지칭한다. 거대분자 UMI를 사용하여 다수의 연장된 기록 태그로부터 서열분석 데이타를 계산적으로 디컨벌루트하여 개개 거대분자로부터 기원한 연장된 기록 태그를 확인할 수 있다. 결합제 UMI를 사용하여 특수한 거대분자에 결합하는 각각의 개개 결합제를 확인할 수 있다. 예를 들면, UMI를 사용하여 특수한 펩타이드 분자에 대해 발생하는 단일 아미노산에 대해 특이적인 결합제에 대한 다수의 개개 결합 사건을 확인할 수 있다. UMI 및 바코드가 둘 다 결합제 또는 거대분자의 내용에서 지칭되는 경우, 바코드는 개개 결합제 또는 거대분자(예컨대, 샘플 바코드, 구획 바코드, 결합 주기 바코드)에 대한 UMI 이외의 확인 정보를 지칭한다.

본원에 사용된 바와 같은, 용어 "보편적인 프라이밍 부위" 또는 "보편적인 프라이머" 또는 "보편적인 프라이밍 서열"은 라이브러리 증폭 및/또는 서열분석 반응에 사용될 수 있는 핵산 분자를 지칭한다. 보편적인 프라이밍 부위는 PCR 증폭을 위한 프라이밍 부위(프라이머 서열), 일부 차세대 서열분석 플랫폼에서 브릿지 증폭을 가능하도록 하는 유동 셀(flow cell) 표면에서 상보성 올리고뉴클레오타이드에 어닐링하는 유동 셀 어댑터 서열, 서열분석 프라이밍 부위, 또는 이의 조합을 포함하나, 이에 한정되지 않는다. 보편적인 프라이밍 부위는 차세대 디지탈 서열분석과 함께 일반적으로 사용된 것들을 포함하는, 다른 유형의 증폭에 사용될 수 있다. 예를 들면, 연장된 기록 태그 분자는 서열분석 주형으로서 사용될 수 있는 DNA 나노볼을 형성하기 위한 롤링 서클 증폭(rolling circle amplification)을 위해 사용된 환형화되고 보편적인 프라이밍 부위일 수 있다(Drmanac et

al., 2009, Science 327:78-81). 대안적으로, 기록 태그 분자는 보편적인 프라이밍 부위로부터의 폴리머라제 연장에 의해 직접 환형화되거나 서열분석될 수 있다(Korlach et al., 2008, Proc. Natl. Acad. Sci. 105:1176-1181). "보편적인 프라이밍 부위" 또는 "보편적인 프라이머"와 함께 사용되는 경우 용어 "전방(forward)"은 또한 "5'" 또는 "센스(sense)"로서 지칭될 수 있다. "보편적인 프라이밍 부위" 또는 "보편적인 프라이머"와 함께 사용되는 경우 용어 "역방(reverse)"은 또한 "3'" 또는 "안티센스"로 지칭될 수 있다.

본원에 사용된 바와 같은, 용어 "연장된 기록 태그"는 이에 대해 적어도 하나의 결합제의 암호화 태그의 정보(또는 이의 상보성 서열)가 결합제의 결합 후 거대분자로 이전된 기록 태그를 지칭한다. 암호화 태그의 정보는 기록 태그에 직접적으로(예컨대, 연결) 또는 간접적으로(예컨대, 프라이머 연장) 이전될 수 있다. 암호화 태그의 정보는 효소적으로 또는 화학적으로 기록 태그에 이전될 수 있다. 연장된 기록 태그는 1, 2, 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, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200개 이상의 암호화 태그의 결합제 정보를 포함할 수 있다. 연장된 기록 태그의 염기 서열은 이들의 암호화 태그에 의해 확인된 결합제의 일시적인 및 순차적인 결합 순서를 반영할 수 있으며, 암호화 태그에 의해 확인된 결합제의 결합의 부분적 순차적인 순서를 반영할수 있거나, 암호화 태그에 의해 확인된 결합제의 어떠한 결합 순서도 반영하지 않을 수 있다. 특정 구현예에서, 연장된 기록 태그 속에 존재하는 암호화 태그 정보는 분석되는 거대분자 서열과 적어도 25%, 30%, 35% , 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98%, 99%, 또는 100%의 동일성을 나타낸다. 연장된 기록 태그가 100% 동일성으로 분석되는 거대분자 서열을 나타내지 않는 특정 구현예에서, 오류는 결합제에 의한 오프-표적 결합, 또는 "손실된" 결합 주기(예컨대, 결합제가 결합 주기 동안 거대 분자에 결합하지 못하기 때문에, 실패한 프라이머 연장 반응으로 인하여), 또는 둘 다에 기인할 수 있다.

본원에 사용된 바와 같은, 용어 "연장된 암호화 태그"는 이에 대해 적어도 하나의 기록 태그(또는 이의 상보성 서열)의 정보가 결합제의 결합 후 이전된 암호하 태그, 이에 대해 암호화 태그가 결합된 암호화 태그, 기록 태그가 관련된 거대분자를 지칭한다. 기록 태그의 정보는 암호화 태그에 직접적으로(예컨대, 연결), 또는 간접적으로(예컨대, 프라이머 연장) 이전될 수 있다. 기록 태그의 정보는 효소적으로 또는 화학적으로 이전될 수 있다. 특정 구현예에서, 연장된 암호화 태그는 하나의 결합 사건을 반영하는, 하나의 기록 태그의 정보를 포함한다. 본원에 사용된 바와 같은, 용어 "디-태그" 또는 "디-태그 작제물" 또는 "디-태그 분자"는 이에 대해 적어도 하나의 기록 태그(또는 이의 상보성 서열) 및 적어도 하나의 암호화 태그(또는 이의 상보성 서열)의 정보가 결합제의 결합 후 이전된 핵산 분자, 이에 대해 암호화 태그가 결합된 핵산 분자, 이에 대해 기록 태그가 관련된 거대분자를 지칭한다(참고: 도 11b). 기록 태그의 정보 및 암호화 태그는 디-태그에 간접적으로(예컨대, 프라이머 연장) 이전될 수 있다. 기록 태그의 정보는 효소적으로 또는 화학적으로 이전될 수 있다. 특정 구현예에서, 디-태그는 기록 태그의 UMI, 기록 태그dml 구획 태그, 기록 태그의 보편적인 프라이밍 부위, 암호화 태그의 UMI, 암호화 태그의 암호기 서열, 결합 주기 특이적인 바코드, 암호화 태그의 보편적인 프라이밍 부위, 또는 이의 어떠한 조합을 포함한다.

본원에 사용된 바와 같은, 용어 "고체 지지체", "고체 표면" 또는 "고체 기질" 또는 "기질"은 이에 대해 거대분자(예컨대, 펩타이드)가 공유결합 및 비-공유결합성 상호작용, 또는 이의 어떠한 조합에 의해 직접 또는 간접적으로 관련될 수 있는 다공성 및 비-다공성 물질을 포함하는 어떠한 고체 물질도 지칭한다. 고체 지지체는 2차원(예컨대, 평편 표면) 또는 3차원(예컨대, 겔 매트릭스 또는 비드)일 수 있다. 고체 지지체는 비드, 마이크로비드, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 웨이퍼 칩, 유통 통과 칩, 유동 셀, 시그널 신호전달 전자장치를 포함하는 바이오칩, 채널, 미세역가 웰, ELISA 플레이트, 스피닝 간섭계 디스크(spinning interferometry disc), 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 중합체 매트릭스, 나노입자, 또는 미세구를 포함하나, 이에 한정되지 않는 어떠한 지지체 표면일 수 있다. 고체 지지체용 물질은 아크릴아미드, 아가로즈, 셀룰로즈, 니트로셀룰로즈, 유리, 금, 석영, 폴리스티렌, 폴리에틸렌 비닐 아세테이트, 폴리프로필렌, 폴리메타크릴레이트, 폴리에틸렌, 폴리에틸렌 옥사이드, 폴리실리케이트, 폴리카보네이트, 테플론, 플루오로카본, 나일론, 규소 고무, 다가무수물, 폴리글리콜산, 폴리악트산, 폴리오르토에스테르, 기능화된 실란, 폴리프로필푸메레이트, 콜라겐, 글리코스아미노글리칸, 폴리아미노산, 덱스트란, 또는 이의 어떠한 조합을 포함하나 이에 한정되지 않는다. 고체 지지체는 또한 박 필름, 막, 병, 접시, 섬유, 제직 섬유, 튜브, 입자, 미세구, 미세입자, 또는 이의 어떠한 조합과 같은 형상화된 중합체를 포함한다. 예를 들면, 고체 표면이 비드인 경우, 비드는 세라믹 비드, 폴리스티렌 비드, 중합체 비드, 메틸스티렌 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 제어된 공극 비드를 포함할 수 있으나, 이에 한정되지 않는다. 비드는 구형 또는 불규칙한 형태일 수 있다. 비드의 크기는 예컨대, 100 nm, 내지 밀리미터,

예컨대, 1 ㎜의 범위일 수 있다. 특정 구현예에서, 비드는 약 0.2 마이크론 내지 약 200 마이크론, 또는 약 0.5 마이크론 내지 약 5 마이크론의 범위일 수 있다. 일부 구현예에서, 비드는 직경이 약 1, 1.5, 2, 2.5, 2.8, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 15, 또는 20 μm일 수 있다. 특정 구현예에서, "비드" 고체 지지체는 개개 비드 또는 다수의 비드를 지칭할 수 있다.

본원에 사용된 바와 같은, 용어 "핵산 분자" 또는 "폴리뉴클레오타이드"는 3'-5' 포스포디에스테르 결합에 의해 연결된 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드, 및 폴리뉴클레오타이드 유사체를 함유하는 단일 가닥 또는 이중 가닥 폴리뉴클레오타이드를 지칭한다. 핵산 분자는 DNA, RNA, 및 cDNA를 포함하나, 이에 한정되지 않는다. 폴리뉴클레오타이드 유사체는 천연 폴리뉴클레오타이드에서 발견된 표준 포스포디에스테르 연결이외의 골격 및, 임으로 리보스 또는 데옥시리보스 이외의 변형된 당 모이어티 또는 모이어티들을 지닐 수 있다. 폴리뉴클레오타이드 유사체는 표준 폴리뉴클레오타이드 염기에 대해 왓슨-크릭 염기 쌍화(Watson-Crick base pairing)에 의해 수소 결합할 수 있는 염기를 함유하며, 여기서 유사체 골격은 올리고뉴클레오타이드 유사체 분자와 표준 폴리뉴클레오타이드내 염기 사이에 서열-특이적인 양식에서 이러한 수소 결합을 허용하는 방식으로 염기를 나타낸다. 폴리뉴클레오타이드 유사체의 예는 크세노(xeno) 핵산(XNA), 브릿지된 핵산(BNA), 글리콜 핵산(GNA), 펩타이드 핵산(PNA), gPNA, 모르폴리노 폴리뉴클레오타이드, 록킹된(locked) 핵산(LNA), 트레오즈 핵산(TNA), 2'-O-메틸 폴리뉴클레오타이드, 2'-O-알킬 리보실 치환된 폴리뉴클레오타이드, 포스포로티오에이트 폴리뉴클레오타이드, 및 보로노포스페이트 폴리뉴클레오타이드를 포함하나, 이에 한정되지 않는다. 폴리뉴클레오타이드 유사체는 예를 들면, 7-데아자 푸린 유사체, 8-할로푸린 유사체, 5-할로피리미딘 유사체를 포함하는 푸린 또는 피리미딘 유사체, 또는 하이포크산틴, 니트로아졸, 이소카보스티릴 유사체, 아졸 카복스아미드, 및 방향족 트리아졸 유사체를 포함하는 어떠한 염기와도 쌍을 이룰 수 있는 염기 유사체, 또는 친화성 결합을 위한 바이오틴 모이어티와 같은 추가의 기능성을 지닌 염기 유사체를 포함하나, 이에 한정되지 않는다.

본원에 사용된 바와 같은, "핵산 서열분석"은 핵산 분자 또는 핵산 분자의 샘플내 뉴클레오타이드의 순서의 측정을 의미한다.

본원에 사용된 바와 같은, "차 세대 서열분석"은 수백만 내지 수십억의 분자를 병렬로 서열분석하도록 하는 고-처리량 서열분석 방법을 지칭한다. 차 세대 서열분석 방법의 예는 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석(polony sequencing), 이온 반도체 서열분석 및 피로서열분석을 포함하나, 이에 한정되지 않는다. 프라이머를 고체 기질에 및 상보성 서열을 핵산 분자에 부착함으로써, 핵산 분자는 프라이머를 경유해 고체 기질로 하이브리드화될 수 있으며 이후 다수의 카피가 폴리머라제를 사용함으로써 고체 기질 위의 별개의 영역내에 생성되어 증폭될 수 있다(이들 그룹화는 때때로 폴리머라제 콜로니 또는 콜로니들로 지칭된다). 결과적으로, 서열분석 과정 동안, 특수한 위치에서의 뉴클레오타이드는 수회(예컨대, 수백 또는 수천회) 서열분석될 수 있으며 - 이러한 포함 깊이(depth of coverage)를 "깊은 서열분석(deep sequencing)"으로 지칭한다. 고 처리량 핵산 서열분석 기술의 예는 제공자에 의해 검토되는 바와 같이, 병렬 비드 배열, 합성에 의한 서열분석, 연결에 의한 서열분석, 모세관 전기영동, "바이오칩", 미세배열, 병렬 마이크로칩, 및 단일-분자 배열과 같은 양식을 포함하는, Illumina, BGI, Qiagen, Thermo-Fisher, 및 Roche 에 의해 제공된 플랫폼을 포함한다(*Science* 311:1544-1546, 2006).

본원에 사용된 바와 같은, "단일 분자 서열분석" 또는 "3 세대 서열분석"은 차-세대 서열분석 방법을 지칭하며, 여기서 단일 분자 서열분석 장치로부터의 판독물은 DNA의 단일 분자의 서열분석에 의해 생성된다. 상화된 시도(phased approach)로 서열분석하기 위해 많은 DNA 분자를 병렬로 클로닝하기 위해 증폭에 의존하는 차 세대 서열분석 방법과는 달리, 단일 분자 서열분석은 DNA의 단일 분자를 조사(interrogate)하므로 증폭 또는 동기화(synchronization)를 필요로 하지 않는다. 단일 분자 서열분석은 각각의 염기 혼입 후 서열분석 반응을 중지하는 것을 필요로 하는 방법('세척-및-스캔(wash-and-scan)' 주기) 및 판독 단계 사이에서 중단할 필요가 없는 방법을 포함한다. 단일 분자 서열분석 방법의 예는 단일 분자 실시간 서열분석(Pacific Biosciences), 나노공극-기반 서열분석(Oxford Nanopore), 듀플렉스 차단된 나노공극 서열분석, 및 발달된 현미경을 사용한 DNA의 직접적인 영상화를 포함한다.

본원에 사용된 바와 같은, 거대분자를 "분석하는 것"은 거대분자의 성분 모두 또는 일부를 정량화하거나, 특성화하거나, 구별하거나, 또는 이의 조합을 의미한다. 예를 들면, 펩타이드, 폴리펩타이드, 또는 단백질을 분석하는 것은 펩타이드의 아미노산 서열(연속 또는 비-연속) 모두 또는 일부를 측정하는 것을 포함한다. 거대분자를 분석하는 것은 또한 거대 분자의 성분의 부분적인 확인을 포함한다. 예를 들면, 거대분자 단백질 서열내 아미노산의 부분적인 확인은 가능한 아미노산의 소세트에 속하는 단백질 속의 아미노산을 확인할 수 있다. 분석은 전형적으로 *n*개의 NTAA로 개시하며, 이후 펩타이드의 다음 아미노산(즉, *n-1*, *n-2*, *n-3* 등)으로 진행된다. 이는 *n*

번째 NTAA의 절단에 의해 달성되며, 이에 의해 펩타이드의 *n-1*번 아미노산은 N-말단 아미노산(본원에서 "*n-1*번 NTAA"으로 지칭됨)으로 전환된다. 펩타이드를 분석하는 것은 또한 펩타이드 위에서 해독 후 변형의 존재 및 빈도를 포함할 수 있으며, 이는 펩타이드 상에서 해독 후 변형의 순차적인 순서에 관한 정보를 포함하거나, 포함하지 않을 수 있다. 펩타이드를 분석하는 것은 또한 상이한 유형의 분석을 조합하는 것, 예를 들면 에피토프 정보, 아미노산 서열 정보, 해독 후 변형 정보, 또는 이의 어떠한 조합을 수득하는 것을 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "구획"은 거대분자의 샘플로부터 거대분자의 소세트를 분리하거나 단리하는 물리적 영역 또는 용적을 지칭한다. 예를 들면, 구획은 개개 세포를 다른 세포로부터, 또는 샘플의 프로테옴의 소세트를 샘플의 프로테옴의 나머지로부터 분리할 수 있다. 구획은 수성 구획(예컨대, 미세유동 소적), 고체 구획(예컨대, 플레이트, 튜브, 바이알, 겔 비드 상의 피코역가 웰 또는 미세역가 웰), 또는 표면 상의 분리된 영역일 수 있다. 구획은 이에 대해 거대 분자가 고정될 수 있는 하나 이상의 비드를 포함할 수 있다.

본원에 사용된 바와 같은, 용어 "구획 태그" 또는 "구획 바코드"는 하나 이상의 구획(예컨대, 미세유동 소적) 내에 구성성분(예컨대, 단일 세포의 프로테옴)에 대한 확인 정보를 포함하는 약 4개 염기 내지 약 100개 염기(4개 염기, 100개 염기, 및 이들 사이의 어떠한 정수)의 단일 또는 이중 가닥 핵산 분자를 지칭한다. 구획 바코드는 샘플 속의 거대분자의 소세트, 예컨대, 동일한 물리적 구획 또는 다수(예컨대, 수백만 내지 수십억)의 구획의 그룹으로 분리된, 샘플내 거대분자, 예컨대, 단백질 샘플의 소세트를 확인한다. 따라서, 구획 태그를 사용하여 구성성분이 함께 혼주된 후에서 조차, 상이한 구획 태그를 갖는 다른 구획에서 이들로부터 동일한 구획을 가진 하나 이상의 구획으로부터 기원한 구성성분을 구별할 수 있다. 각각의 구획 또는 2개 이상의 구획의 그룹 내에서 단백질 및/또는 펩타이드를 유일한 구획 태그로 표지함으로서, 개개의 구획 또는 구획의 그룹내에 동일한 단백질, 단백질 복합체, 또는 세포로부터 기원한 펩타이드를 확인할 수 있다. 구획 태그는 바코드를 포함하며, 이는 하나 또는 둘 다의 측면에서 스페이서 서열, 및 임의의 보편적인 프라이머에 의해 임의로 플랭킹된다. 스페이서 서열은 기록 태그의 스페이서 서열에 대해 상보성이어서, 구획 태그 정보를 기록 태그로 이전할 수 있다. 구획 태그는 또한 특히 구획 태그가 본원에 기술된 하부 펩타이드 분석 방법(downstream peptide analysis method)에 사용될 기록 태그를 포함하는 구현예의 경우, 보편적인 프라이밍 부위, 유일한 분자 확인인자(이에 부착된 펩타이드에 대한 확인 정보를 제공하기 위한), 또는 둘 다를 포함할 수 있다. 구획 태그는 펩타이드를 커플링시키기 위한 기능성 모이어티(예컨대, 알데하이드, NHS, mTet, 알킨 등)을 포함할 수 있다. 대안적으로, 구획 태그는 목적한 펩타이드로 구획 태그를 연결시키는 단백질 리가제에 대한 인식 서열을 포함하는 펩타이드를 포함할 수 있다. 구획은 단일 구획 태그, 임의의 UMI 서열에 대한 다수의 동일한 구획 태그 세이브(save), 또는 2개 이상의 상이한 구획 태그를 포함할 수 있다. 특정 구현예에서 각각의 구획은 유일한 구획 태그(1 대 1 맵핑)를 포함한다. 다른 구현예에서, 보다 큰 집단의 구획으로부터의 다수의 구획은 동일한 구획 태그(다수 대 1 맵핑(many-to-one mapping))을 포함한다. 구획 태그는 구획(예컨대, 비드)내에서 결합되거나 구획 자체의 표면(예컨대, 피코역가 웰의 표면)에 결합된 고체 지지체일 수 있다. 대안적으로, 구획 태그는 구획내 용액 속에 존재하지 않을 수 있다.

본원에 사용된 바와 같은, 용어 "분배"는 샘플내 거대분자의 집단으로부터 거대분자의 소집단에 대한 유일한 바코드의 무작위 지정을 지칭한다. 특정 구현예에서, 분배는 거대분자를 구획내로 분배시킴으로써 달성할 수 있다. 분배는 단일 구획내 거대분자 또는 구획의 집단으로부터의 다수의 구획내에 거대분자로 구성될 수 있다.

본원에 사용된 바와 같은, "분배 태그" 또는 "분배 바코드"는 분배에 대한 확인 정보를 포함하는 약 4개 염기 내지 약 100개 염기(4개 염기, 100개 염기, 및 이들 사이의 어떠한 정수 포함)의 단일 또는 이중 가닥 핵산 분자를 지칭한다. 특정 구현예에서, 거대분자에 대한 분배 태그는 거대분자를 동일한 바코드로 표지된 구획(들)내로 분배하는 것으로부터 발생하는 동일한 구획 태그를 지칭한다.

본원에 사용된 바와 같은, 용어 "분획"은 크기, 소수성, 등전점, 친화성 등에 의한 분획화와 같은, 물리적 또는 화학적 분리 방법을 사용하여 샘플 또는 기관의 나머지로부터 분류된 샘플내 거대분자(예컨대, 단백질)의 소집단을 지칭한다. 분리 방법은 HPLC 분리, 겔 분리, 친화성 분리, 세포 분획화, 세포 세포기관 분획화, 조직 분획화 등을 포함한다. 유체 유동, 자성(magnetism), 전류, 질량, 밀도 등과 같은 물리적 특성을 또한 분리에 사용할 수 있다.

본원에 사용된 바와 같은, 용어 "분획 바코드"는 분획내에 거대분자에 대한 확인 정보를 포함하는 약 4개 염기 내지 약 100개 염기(4개 염기, 100개 염기, 및 이들 사이의 어떠한 정수 포함)의 단일 또는 이중 가닥 핵산 분자를 지칭한다.

III. 거대분자의 분석 방법

본원에 기술된 방법은 거대분자 분석을 위한 고-병렬화된 시도(highly-parallelized approach)를 제공한다. 고도로 멀티플렉스화된 거대분자 결합 검정은 차 세대 서열분석에 의한 판독을 위해 핵산 분자 라이브러리로 전환된다. 본원에 제공된 방법은 단백질 또는 펩타이드 서열분석에 특히 유용하다.

바람직한 구현예에서, 단백질 샘플은 바코드(예컨대, 샘플 바코드, 구획 바코드) 및 임의의 유일한 분자 확인인자를 포함하는 적어도 하나의 핵산 기록 태그를 지닌 단일 분자 수준에서 표지된다. 단백질 샘플은 단백질분해성 소화를 겪어 기록 태그 표지된 펩타이드의 집단(예컨대, 수백만 내지 수십억)을 생산한다. 이러한 기록 태그 표지된 펩타이드는 혼주되어 고체 지지체(예컨대, 다공성 비드)에 무작위로 고정된다. 혼주되고, 고정된, 기록 태그 표지된 펩타이드는 다수의, 연속된 결합 주기에 적용되며, 각각의 결합 주기는 관련된 결합제를 확인하는 암호기 서열을 포함하는 암호화 태그로 표지된 다수의 결합제(예컨대, 천연적으로 존재하는 아미노산 20개 모두에 대한 결합제)에 대한 노출을 포함한다. 각각의 결합 주기 동안, 펩타이드에 대한 결합제의 결합에 관한 정보는 결합제의 암호화 태그 정보를 기록 태그로 이전함으로써(또는 기록 태그 정보를 암호화 태그로 이전하거나 기록 태그 정보 및 암호화 태그 정보 둘 다를 별도의 디-태그 작제물로 이전함) 포획된다. 결합 주기가 완료되면, 분석된 펩타이드의 결합 역사를 나타내는 연장된 기록 태그(또는 연장된 암호화 태그 또는 디-태그 작제물)의 라이브러리가 생성되며, 이는 매우 높은-처리량의 차 세대 디지탈 서열분석 방법을 사용하여 분석할 수 있다. 기록 태그내에서의 핵산 바코드의 사용은 거대한 양의 펩타이드 서열 분석의 데콘벌루션을 허용하는데, 예컨대, 어떠한 샘플, 세포, 프로테옴의 소세트, 또는 단백질, 펩타이드 서열이 기원하는지를 확인하도록 한다.

일 국면에서, (a) 고체 지지체에 결합된 거대분자 및 관련된 또는 동시-국재화된 기록 태그를 제공하는 단계; (b) 거대분자를 거대분자에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서, 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다); (c) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 제1의 순서의 연장된 기록 태그를 생성시키는 단계; (d) 거대분자를 거대분자에 결합할 수 있는 제2의 결합제와 잡촉시키는 단계(여기서, 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다); (e) 제2의 암호화 태그의 정보를 제1의 순서의 연장된 기록 태그로 이전시켜 제2의 순서의 연장된 기록 태그를 생성시키는 단계; 및 (f) 제2의 순서의 연장된 태그를 분석하는 단계를 포함하여, 거대분자를 분석하는 방법이 제공된다(참고:, 예컨대, 도 2a 내지 2d).

특정 구현예에서, 접촉 단계 (b) 및 (d)는 순차적인 순서로 수행되는데, 에컨대, 제1의 결합제 및 제2의 결합제는 별도의 결합 주기 반응에서 거대분자와 접촉된다. 다른 구현예에서, 접촉 단계 (b) 및 (d)는 동시에, 예컨대, 제1의 결합제, 제2의 결합제, 및 임의로 추가의 결합제를 포함하는 단일의 결합 주기 반응으로 수행된다. 바람직한 구현예에서, 접촉 단계 (b) 및 (d)는 각각 거대분자를 다수의 결합제와 접촉시킴을 포함한다.

특정 구현예에서, 방법은 또한 단계 (e)와 단계 (f) 사이에 다음 단계를 포함한다: (x) 제2의 결합제를 거대분자에 결합할 수 있는 제3의 (또는 보다 높은 차수의) 결합제로 치환함으로써 단계 (d) 및 (e)를 1회 이상 반복하는 단계(여기서 제3의 (또는 보다 높은 차수의) 결합제는 제3의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3의 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및 (y) 제3의 (또는 보다 높은 차수의) 암호화 태그의 정보를 제2의 (또는 보다 높은 차수의) 연장된 기록 태그로 이전하여 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 생성하는 단계; 및 (z) 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 분석하는 단계.

제3의 (또는 보다 높은 차수의) 결합제는 제1의 결합제 및 제2의 결합제로부터의 별도의 결합 주기에서 거대분자와 접촉될 수 있다. 대안적으로, 제3의 (또는 보다 높은 차수의) 결합제는 제1의 결합제, 및 제2의 결합제와의 단일 결합 주기 반응에서 거대분자와 접촉될 수 있다.

제2의 국면에서, (a) 거대분자, 고체 지지체에 결합된 관련된 제1의 기록 태그 및 관련된 제2의 기록 태그를 제공하는 단계; (b) 거대분자를 거대분자에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다); (c) 제1의 암호화 태그의 정보를 제1의 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성시키는 단계; (d) 거대분자를 거대분자에 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다); (e) 제2의 암호화 태그의 정보를 제2의 기록 태그로 이전시켜 제2의 연장된 기록 태그를 생성하는 단계; 및 (f) 제1 및 제2의 연장된 기록 태그를 분석하는 단계를 포함하는, 거대분자를 분석하는 방법이 제공된다.

특정 구현예에서, 접촉 단계 (b) 및 (d)는 순차적인 순서로 수행되는데, 예컨대, 제1의 결합제 및 제2의 결합제는 별도의 결합 주기 반응에서 거대분자와 접촉된다. 다른 구현예에서, 접촉 단계 (b) 및 (d)는 동시에,

예컨대, 제1의 결합제, 제2의 결합제, 및 임의로 추가의 결합제를 포함하는 단일 결합 주기 반응으로 수행된다.

특정 구현예에서, 단계 (a)는 고체 지지체에 결합된 관련된 제3의 (또는 보다 높은 차수의) 기록 태그를 제공함을 추가로 포함한다. 추가의 구현에에서, 방법은 또한 단계 (e)와 단계 (f) 사이에, 다음의 단계를 포함한다: (x) 제2의 결합제를 거대분자에 결합할 수 있는 제3의 (또는 보다 높은 차수의) 결합제와 치환함으로써 단계 (d) 및 (e)를 1회 이상 반복하는 단계(여기서 제3의 (또는 보다 높은 차수의) 결합제는 제3의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3의 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및 (y) 제3의 (또는 보다 높은 차수의) 암호화 태그의 정보를 제3의 (또는 보다 높은 차수의) 기록 태그로 이전하여 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 생성시키는 단계; 및 (z) 제1, 제2 및 제3의(또는 보다 높은 차수의) 연장된 기록 태그를 분석하는 단계를 포함한다.

제3의 (또는 보다 높은 차수의) 결합제는 제1의 결합제 및 제2의 결합제로부터 별도의 결합 주기 반응으로 거대분자와 접촉할 수 있다. 대안적으로, 제3의 (또는 보다 높은 차수의) 결합제는 제1의 결합제, 및 제2의 결합제와의 단일 결합 주기 반응으로 거대 분자와 접촉될 수 있다.

특정 구현예에서, 제1의 암호화 태그, 제2의 암호화 태그, 및 어떠한 보다 고 차수의 암호화 태그 각각은 결합 주기 특이적인 서열을 갖는다.

제3의 국면에서, (a) 펩타이드 및 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계; (b) 펩타이드의 N-말단 아미노산(NTAA)을 화학적 모이어티로 변형시켜 변형된 NTAA를 생산하는 단계; (c) 펩타이드를 변형된 NTAA에 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다); (d) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 연장된 기록 태그를 생성하는 단계; 및 (e) 연장된 기록 태그를 분석하는 단계를 포함하여, 펩타이드를 분석하는 방법이 제공된다(참고: 예컨대, 도 3).

특정 구현예에서, 단계 (c)는 또한 제2의 (또는 보다 높은 차수의) 암호화 태그를 제2의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보와 접촉시킴을 포함하여 펩타이드를 제2의 (또는 보다 높은 차수의) 결합제와 잡촉시킴을 포함하며, 여기서 제2의 (또는 보다 높은 차수의) 결합제는 단계 (b)의 변형된 NTAA이외에 변형된 NTAA에 결합할 수 있다. 추가의 구현예에서, 펩타이드를 제2의 (또는 보다 높은 차수의) 결합제와 접촉시키는 것은 펩타이드가 제1의 결합제와 접촉된 후 순차적으로 일어나는데, 예컨대, 제1의 결합제 및 제2의 (또는 보다 높은 차수의) 결합제는 별도의 결합 주기 반응에서 펩타이드와 접촉된다. 다른 구현예에서, 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉은, 예컨대, 제1의 결합제 및 제2의 (또는 보다 높은 차수의) 결합제를 포함하는 단일의 결합 주기 반응에서와 같이, 펩타이드가 제1의 결합제와 접촉하는 것과 동시에 일어난다.

특정 구현예에서, 화학적 모이어티는 화학적 반응 또는 효소적 반응을 통해 NTAA에 첨가된다.

특정 구현예에서, NTAA를 변형시키는데 사용된 화학적 모이어티는 페닐티오카바모일(PTC), 디니트로페놀(DNP) 모이어티; 설포닐옥시니트로페닐(SNP) 모이어티, 단실 모이어티; 7-메톡시 쿠마린 모이어티; 티오아실 모이어티; 티오아세틸 모이어티; 아세틸 모이어티; 구아니드닐(guanidnyl) 모이어티; 또는 티오벤질 모이어티이다.

화학적 모이어티는 화학제를 사용하여 NTAA에 첨가할 수 있다. 특정 구현예에서, NTAA를 PTC 모이어티로 변형시키기 위한 화학제는 이소티오시아네이트 또는 이의 유도체이거나, NTAA를 DNP 모이어티로 변형시키기 위한 화학제는 2,4-디니트로벤젠설폰산(DNBS) 또는 1-플루오로-2,4-디니트로벤젠 (DNFB)과 같은 아릴 할라이드이거나; NTAA를 설포닐옥시니트로페닐(SNP) 모이어티로 변형시키기 위한 화학제는 4-설포닐-2-니트로플루오로벤젠(SNFB)이거나; NTAA를 단실 그룹으로 변형시키기 위한 화학제는 단실 클로라이드와 같은 설포닐 클로라이드이거나; NTAA를 7-메톡시 쿠마린 모이어티로 변형시키기 위한 화학제는 7-메톡시쿠마린 아세트산(MCA)이거나; NTAA를 티오아실 모이어티로 변형시키기 위한 화학제는 티오아실화 시약이거나; NTAA를 티오아세틸 모이어티로 변형시키기 위한 화학제는 티오아세틸화 시약이거나; NTAA를 아세틸 모이어티로 변형시키기 위한 화학제는 아세틸화 시약(예컨대, 아세트산 무수물)이거나; NTAA를 구아니드닐(아미디닐) 모이어티로 변형시키기 위한 화학제는 구아니디닐화 시약이거나, NTAA를 티오벤질 모이어티로 변형시키기 위한 화학제는 티오벤질화시약이다.

본 개시내용의 제4의 국면은:

(a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계; (b) 펩타이드의 N-말단 아미노산(NTAA)을 화학적 모이어티로 변형시켜 변형된 NTAA를 생산하는 단계; (c) 펩타이드를 변형된 NTAA에 결합할 수

있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다); (d) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성시키는 단계; (e) 변형된 NTAA를 제거하여 새로운 NTAA를 노출시키는 단계; (f) 펩타이드의 새로운 NTAA를 화학적 모이어티로 변형시켜 새로이 변형된 NTAA를 생산하는 단계; (g) 펩타이드를 새로이 변형된 NTAA에 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다); (h) 제2의 암호화 태그의 정보를 제1의 연장된 기록 태그로 이전시켜 제2의 연장된 기록 태그를 생성시키는 단계; 및 (i) 제2의 연장된 기록 태그를 분석하는 단계를 포함하여, 펩타이드를 분석하는 방법이 제공된다.

특정 구현예에서, 접촉 단계 (c) 및 (g)는 순차적인 순서로 수행되는데, 예컨대, 제1의 결합제 및 제2의 결합제는 별도의 결합 주기 반응에서 펩타이드와 접촉된다.

특정 구현예에서, 방법은 단계 (h)와 단계 (i) 사이에 다음의 단계를 추가로 포함한다: (x) 제2의 결합제를 변형된 NTAA에 결합할 수 있는 제3의 (또는 보다 높은 차수의) 결합제로 치환함으로써 단계 (e), (f), 및 (g)를 1회 이상 반복하는 단계(여기서 제3의 (또는 보다 높은 차수의) 결합제는 제3의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3의 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및 (y) 제3의 (또는 보다 높은 차수의) 암호화 태그의 정보를 제2의 (또는 보다 높은 차수의) 연장된 기록 태그로 이전시켜 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 생성하는 단계; 및 (z) 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 분석하는 단계.

특정 구현예에서, 화학적 모이어티는 화학적 반응 또는 효소적 반응을 통해 NTAA에 첨가된다.

특정 구현예에서, 화학적 모이어티는 페닐티오바카모일(PTC), 디니트로페놀(DNP) 모이어티; 설포닐옥시니트로페닐(SNP) 모이어티, 단실 모이어티; 7-메톡시 쿠마린 모이어티; 티오아실 모이어티; 티오아세틸 모이어티; 아세틸 모이어티; 구아닐 모이어티; 또는 티오벤질 모이어티이다.

화학적 모이어티는 화학제를 사용하에 NTAA에 첨가할 수 있다. 특정 구현예에서, NTAA를 PTC 모이어티로 변형시키기 위한 화학제는 페닐 이소티오시아네이트 또는 이의 유도체이거나; NTAA를 DNP 모이어티로 변형시키기 위한 화학제는 2,4-디니트로벤젠설폰산(DNBS) 또는 1-플루오로-2,4-디니트로벤젠(DNFB)과 같은 아실 할라이드이거나; NTAA를 설포닐옥시니트로페닐(SNP) 모이어티로 변형시키기 위한 화학제는 4-설포닐-2-니트로플루오로벤젠(SNFB)이거나; NTAA를 단실 그룹으로 변형시키기 위한 화학제는 단실 클로라이드와 같은 설포닐 클로라이드이거나; NTAA를 7-메톡시 쿠마린 모이어티로 변형시키기 위한 화학제는 7-메톡시쿠마린 아세트산(MCA)이거나; NTAA를 티오아실 모이어티로 변형시키기 위한 화학제는 티오아실화 시약이거나; NTAA를 티오아세틸 모이어티로 변형시키기 위한 화학제는 티오아세틸화 시약이거나; NTAA를 아세틸 모이어티로 변형시키기 위한 화학제는 아세틸화제(예컨대, 아세트산 무수물)이거나; NTAA를 구아닐 모이어티로 변형시키기 위한 화학제는 구아니디닐화 시약이거나, NTAA를 티오벤질 모이어티로 변형시키기 위한 화학제는 티오벤질화 시약이다.

제5의 국면에서, (a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계; (b) 펩타이드를 펩타이드의 N-말단 아미노산(NTAA)과 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한 확인 정보를 지닌 제1의 암호화 태그를 포함한다); (c) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 연장된 기록 태그를 생성시키는 단계; 및 (d) 연장된 기록 태그를 분석하는 단계를 포함하는, 펩타이드를 분석하는 방법이 제공된다.

특정 구현예에서, 단계 (b)는 펩타이드를 제2의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제2의 (또는 보다 높은 차수의) 암호화 태그를 포함하는 제2의 (또는 보다 높은 차수의) 결합제와 접촉시킴을 추가로 포함하며, 여기서 제2의 (또는 보다 높은 차수의) 결합제는 펩타이드의 NTAA 이외의 NTAA에 결합할 수 있다. 추가의 구현예에서, 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉은 펩타이드가 제1의 결합제와 접촉한 후 순차적인 순서로 일어나는데, 예컨대, 제1의 결합제 및 제2의 (또는 보다 높은 차수의) 결합제는 펩타이드와 별도의 결합 주기 반응으로 접촉한다. 다른 구현예에서, 펩타이드와 제2의 (또는 보다 높은 차수의) 결합제의 접촉은 펩타이드가 제1의 결합제와 접촉한 후 동시에, 예컨대, 제1의 결합제 및 제2의 (또는 보다 높은 차수의) 결합제를 포함하는 단일의 결합 주기 반응으로 일어난다.

제6의 국면에서,

(a) 펩타이드 및, 고체 지지체에 결합된 관련된 기록 태그를 제공하는 단계; (b) 펩타이드를 펩타이드의 N-말단 아미노산(NTAA)과 결합할 수 있는 제1의 결합제와 접촉시키는 단계(여기서 제1의 결합제는 제1의 결합제에 관한

확인 정보를 지닌 제1의 암호화 태그를 포함한다); (c) 제1의 암호화 태그의 정보를 기록 태그로 이전하여 제1의 연장된 기록 태그를 생성하는 단계; (d) NTAA를 제거하여 펩타이드의 새로운 NTAA를 노출시키는 단계; (e) 펩타이드를 새로운 NTAA와 결합할 수 있는 제2의 결합제와 접촉시키는 단계(여기서 제2의 결합제는 제2의 결합제에 관한 확인 정보를 지닌 제2의 암호화 태그를 포함한다); (f) 제2의 암호화 태그의 정보를 제1의 연장된 기록 태그로 이전시켜 제2의 연장된 기록 태그를 생성하는 단계; 및 (g) 제2의 연장된 기록 태그를 분석하는 단계를 포함하여, 펩타이드를 분석하는 방법이 제공된다.

특정 구현예에서, 방법은 단계 (f) 및 단계 (g) 사이에 다음의 단계를 포함한다; (x) 제2의 결합제를 거대분자에 결합할 수 있는 제3의 (또는 보다 높은 차수의) 결합제로 치환함으로써 단계 (d), (e), 및 (f)를 1회 이상 반복하는 단계(여기서 제3의 (또는 보다 높은 차수의) 결합제는 제3의 (또는 보다 높은 차수의) 결합제에 관한 확인 정보를 지닌 제3의 (또는 보다 높은 차수의) 암호화 태그를 포함한다); 및 (y) 제3의 (또는 보다 높은 차수의) 암호화 태그의 정보를 제2의 (또는 보다 높은 차수의) 연장된 기록 태그로 이전시켜 제3의 (또는 보다 높은 차수의) 연장된 기록 태그를 생성하는 단계(여기서 제3의 (또는 보다 높은 차수의) 연장된 기록 태그는 단계 (g)에서 분석된다).

특정 구현예에서, 접촉 단계 (b) 및 (e)는 순차적인 순서로 수행되는데, 예컨대, 제1의 결합제 및 제2의 결합제는 별도의 결합 주기 반응으로 펩타이드와 접촉된다.

본원에 제공된 어떠한 구현예에서, 방법은 다수의 거대분자를 병렬로 분석함을 포함한다. 바람직한 구현예에서, 방법은 다수의 펩타이드를 병렬로 분석함을 포함한다.

본원에 제공된 어떠한 구현예에서, 거대분자 (또는 펩타이드)를 결합제와 접촉시키는 단계는 거대분자 (또는 펩타이드)를 다수의 결합제와 접촉시킴을 포함한다.

본원에 제공된 어떠한 구현예에서, 거대분자는 단백질, 폴리펩타이드, 또는 펩타이드일 수 있다. 추가의 구현예에서, 펩타이드는 생물학적 샘플로부터 단백질 또는 폴리펩타이드를 단편화하여 수득될 수 있다.

본원에 제공된 어떠한 구현예에서, 거대분자는 탄수화물, 지질, 핵산, 또는 마크로사이클이거나 이를 포함할 수 있다.

본원에 제공된 어떠한 구현예에서, 기록 태그는 DNA 분자, 변형된 염기를 지닌 DNA 분자, RNA 분자, BNA, 분자, XNA 분자, LNA 분자, PNA 분자, γPNA 분자 (Dragulescu-Andrasi et al., 2006, J. Am. Chem. Soc. 128:10258-10267), GNA 분자, 또는 이의 어떠한 조합일 수 있다.

본원에 제공된 어떠한 구현예에서, 기록 태그는 보편적인 프라이밍 부위를 포함할 수 있다. 추가의 구현예에서, 보편적인 프라이밍 부위는 증폭, 연결, 서열분석, 또는 이의 조합을 위한 프라이밍 부위를 포함한다.

본원에 제공된 어떠한 구현예에서, 기록 태그는 유일한 분자 확인인자, 구획 태그, 분배 바코드, 샘플 바코드, 분획 바코드, 스페이서 서열, 또는 이의 어떠한 조합을 포함할 수 있다.

본원에 제공된 어떠한 구현예에서, 암호화 태그는 유일한 분자 확인인자 (UMI), 암호기 서열, 결합 주기 특이적인 서열, 스페이서 서열, 또는 이의 어떠한 조합을 포함할 수 있다.

본원에 제공된 어떠한 구현예에서, 암호화 태그내 결합 주기 특이적인 서열은 결합 주기-특이적인 스페이서 서열일 수 있다.

특정 구현예에서, 결합 주기 특이적인 서열은 암호기 서열로부터 별도의 바코드로서 암호화된다. 다른 구현예에서, 암호기 서열 및 결합 주기 특이적인 서열은 결합제에 대해 및 각각의 결합 주기에 대해 유일한 단일 바코드로 설정된다.

특정 구현예에서, 스페이서 서열은 다수의 결합 주기로부터의 결합제 중에서 공유되는 일반적인 결합 주기 서열을 포함한다. 다른 구현예에서, 스페이서 서열은 동일한 결합 죽로부터의 결합제 중에서 공유된 유일한 결합 주기 서열을 포함한다.

본원에 제공된 어떠한 구현예에서, 기록 태그는 바코드를 포함할 수 있다.

본원에 제공된 어떠한 구현예에서, 거대분자 및 관련된 기록 태그(들)은 고체 지지체에 공유 결합으로 결합될 수 있다.

본원에 제공된 어떠한 구현예에서, 고체 지지체는 비드, 다공성 비드, 다공성 매트릭스, 확장가능한 겔 비드 또

는 매트릭스, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 통과 칩, 시그널 변환 전자를 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구일 수 있다..

본원에 제공된 어떠한 구현예에서, 고체 지지체는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드일 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 거대분자 및 관련된 기록 태그는 고체 지지체에 결합될 수 있다. 추가의 구현예에서, 다수의 거대분자는 ≥ 50 nm, ≥ 100 nm, 또는 ≥ 200 nm의 평균 거리로 고체 지지체 위에서 이격된다.

본원에 제공된 어떠한 구현예에서, 결합제는 폴리펩타이드 또는 단백질일 수 있다. 추가의 구현예에서, 결합제는 변형된 또는 변이체 아미노펩티다제, 변형된 또는 변이체 아미노 아실 tRNA 신테타제, 변형된 또는 변이체 안티칼린, 또는 변형된 또는 변이체 ClpS이다.

본원에 제공된 어떠한 구현예에서, 결합제는 거대분자에 선택적으로 결합할 수 있다.

본원에 제공된 어떠한 구현예에서, 암호화 태그는 DNA 분자, 변형된 염기를 지닌 DNA 분자, RNA 분자, BNA 분자, XNA 분자, LNA 분자, GNA 분자, PNA 분자, gPNA 분자, 또는 이의 조합일 수 있다.

본원에 제공된 어떠한 구현예에서, 결합제 및 암호화 태그는 링커에 의해 결합될 수 있다.

본원에 제공된 어떠한 구현예에서, 결합제 및 암호화 태그는 SpyTag/SpyCatcher 또는 SnoopTag/SnoopCatcher 펩타이드-단백질 쌍(Zakeri, et al., 2012, Proc Natl Acad Sci U S A 109(12): E690-697; Veggiani et al., 2016, Proc. Natl. Acad. Sci. USA 113:1202-1207, 이들 각각은 이의 전문이 참고로 포함된다)일 수 있다.

본원에 제공된 어떠한 구현예에서, 암호화 태그의 정보의 기록 태그로의 이전은 DNA 리가제에 의해 매개된다. 대안적으로, 암호화 태그의 정보의 기록 태그로의 이전은 DNA 폴리머라제 또는 화학적 연결에 의해 매개된다.

본원에 제공된 어떠한 구현예에서, 연장된 기록 태그의 분석은 핵산 서열분석을 포함한다. 추가의 구현예에서, 핵산 서열분석은 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석, 이온 반도체 서열분석, 또는 피로서열분석이다. 다른 구현예에서, 핵산 서열분석은 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 나노갭 터널링 서열분석(nanogap tunneling sequencing) 또는 발달된 현미경을 사용한 DNA의 직접적인 영상화이다.

본원에 제공된 어떠한 구현예에서, 연장된 기록 태그는 분석 전에 증폭될 수 있다.

본원에 제공된 어떠한 구현예에서, 연장된 기록 태그에 함유된 암호화 태그 정보의 순서는 거대분자에 대해 결합제가 결합하는 순서에 관한 정보 및, 따라서 결합제에 의해 검출된 분석물의 서열을 제공한다.

본원에 제공된 어떠한 구현예에서, 연장된 기록 태그 상에 함유된 특수한 암호화 태그 정보(예컨대, 암호기 서열)의 빈도는 특수한 결합제가 거대분자에 결합하는 빈도, 및 따라서 결합제에 의해 검출된 거대분자내 분석물의 빈도에 관한 정보를 제공한다.

본원에 개시된 어떠한 구현예에서, 다수의 거대분자(예컨대, 단백질) 샘플(여기서 각각의 샘플 내 거대분자의 집단은 샘플 특이적인 바코드를 포함하는 기록 태그로 표지된다)은 혼주될 수 있다. 이러한 거대분자 샘플의 혼주물은 단일-반응 튜브 내에서 결합 주기에 적용될 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 거대분자를 나타내는 다수의 연장된 기록 태그는 병렬로 분석될 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 거대분자를 나타내는 다수의 연장된 기록 태그는 멀티플렉스화된 검정으로 분석될 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 연장된 기록 태그는 분석 전에 표적 농축 검정을 겪을 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 연장된 기록 태그는 분석 전에 공제 검정(subtraction assay)을 겪을 수 있다.

본원에 제공된 어떠한 구현예에서, 다수의 연장된 기록 태그는 분석 전에 매우 풍부한 종을 감소시키는 정상화 검정을 겪을 수 있다.

본원에 제공된 어떠한 구현예에서, NTAA는 변형된 아미노펩티다제, 변형된 아미노산 tRNA 신테타제, 온화한 에드만 분해, 에드마나제 효소, 또는 무수 TFA에 의해 제거될 수 있다.

본원에 제공된 어떠한 구현예에서, 적어도 하나의 결합제는 말단 아미노산 잔기에 결합될 수 있다. 특정의 구현예에서, 말단 아미노산 잔기는 N-말단 아미노산 또는 C-말단 아미노산이다.

본원에 제공된 어떠한 구현예에서, 적어도 하나의 결합제는 해독후 변형된 아미노산에 결합될 수 있다.

전술한 구현예의 특징은 다음의 단락에서 추가로 상세히 제공된다.

IV. 거대분자

일 국면에서, 본 개시내용은 거대분자의 분석에 관한 것이다. 거대분자는 보다 작은 소단위로 구성된 큰 분자이다. 특정 구현예에서, 거대분자는 단백질, 단백질 복합체, 폴리펩타이드, 펩타이드, 핵산 분자, 탄수화물, 지질, 마크로사이클, 또는 키메라 거대분자이다.

본원에 개시된 방법에 따라 분석된 거대분자(예컨대, 단백질, 폴리펩타이드, 펩타이드)는, 세포(원발성 세포 및 배양된 세포주 둘 다), 세포 분해물 또는 추출물, 엑소조(exosome), 조직 및 조직 추출물을 포함하는 세포 기관 또는 소낭; 생검; 태아 물질(fecal matter); 사실상 어떠한 유기체의 체액(예를 들면, 혈액, 전혈, 혈청, 혈장, 뇨, 림프액, 담즙액, 뇌척수액, 간질액, 수성 또는 유리액(vitreous homor), 초유, 가래, 양수, 타액, 항문 및 질 분비액, 땀 및 정액, 누출액, 삼출물(예컨대, 종기 또는 감염 또는 염증의 어떠한 다른 부위로부터 수득된 유화액) 또는 관절(정상의 관절 또는 류마티스 관절염, 골관절염, 통풍 또는 패혈성 관절염과 같은 질환이 있는 관절)과 같은 생물학적 샘플과, 바람직한, 미생물군집-함유 샘플을 포함하는 포유동물-기원한 샘플, 및 특히 바람직한 미생물군집-함유 샘플을 포함하는 사람-기원한 샘플; 환경 샘플(예를 들면, 공기, 농업, 물 및 토양 샘플); 미생물 생물막 및/또는 군집으로부터 기원한 샘플, 및 또한 미생물 포자를 포함하는 미생물 샘플; 세포외 유화액, 세포 배양물로부터의 세포외 상층액, 세균내 봉입체(inclusion body), 미토콘드리아 구획을 포함하는 세포 구획, 및 세포 주변세포질을 포함하는 연구 샘플을 포함하나 이에 한정되지 않는 적합한 공급원 또는 샘플로부터 수득될 수 있다.

특정 구현예에서, 거대분자는 단백질, 단백질 복합체, 폴리펩타이드, 또는 펩타이드이다. 펩타이드, 폴리펩타이드, 또는 단백질의 아미노산 서열 정보 및 해독 후 변형은 차 세대 서열분석 방법을 통해 분석될 수 있는 핵산 암호화된 라이브러리로 형질도입된다. 펩타이드는 L-아미노산, D-아미노산, 또는 둘 다를 포함할 수 있다. 펩타이드, 폴리펩타이드, 단백질, 또는 단백질 복합체는 표준의, 천연적으로 존재하는 아미노산, 변형된 아미노산(예컨대, 해독 후 변형), 아미노산 유사체, 아미노산 모사체, 또는 이의 어떠한 조합을 포함할 수 있다. 일부 구현예에서, 펩타이드, 폴리펩타이드, 또는 단백질은 천연적으로 존재하거나, 합성적으로 생산되거나, 재조합적으로 발현된다. 상술한 펩타이드 구현예 중 어느 것에서도, 펩타이드, 폴리펩타이드, 단백질, 또는 단백질 복합체는 해독 후 변형을 추가로 포함할 수 있다.

표준의, 천연적으로 존재하는 아미노산은 알라닌(A 또는 Ala), 시스테인(C 또는 Cys), 아스파르트산(D 또는 Asp), 글루탐산(E 또는 Glu), 페닐알라닌(F 또는 Phe), 글리신(G 또는 Gly), 히스티딘(H 또는 His), 이소루이신(I 또는 Ile), 라이신(K 또는 Lys), 루이신(L 또는 Leu), 메티오닌(M 또는 Met), 아스파라긴(N 또는 Asn), 프롤린(P 또는 Pro), 글루타민(Q 또는 Gln), 아르기닌(R 또는 Arg), 세린(S 또는 Ser), 트레오닌(T 또는 Thr), 발린(V 또는 Val), 트립토판(W 또는 Trp), 및 타이로신(Y 또는 Tyr)을 포함한다. 비-표준 아미노산은 셀레노시스테인, 피롤라이신, N-포르밀메티오닌, β-아미노산, 호모-아미노산, 프롤린 및 피루브산 유도체, 3-치환된 알라닌 유도체, 글리신 유도체, 환-치환된 페닐알라닌 및 타이로신 유도체, 선형 코어 아미노산, 및 N-메틸 아미노산을 포함한다.

펩타이드, 폴리펩타이드, 또는 단백질의 해독후 변형(PTM)은 공유결합성 변형 또는 효소적 변형일 수 있다. 해독후 변형의 예는 아실화, 아세틸화, 알킬화(메틸화 포함), 바이오티닐화, 부티릴화, 카바밀화, 카보닐화, 탈아미드화, 탈이민화, 디프타미드 형성, 이황화물 브릿지 형성, 엘리미닐화, 플라빈 부착, 포르밀화, 감마-카복실화, 글루타밀화, 글리실화, 글리코실화(예컨대, N-연결된, O-연결된, c-연결된, 포스포글리코실화), 글리피화(glypiation), 헴(heme) C 부착, 하이드록실화, 하이푸신 형성, 요오드화, 이소프레닐화, 지질화, 리포일화, 말로닐화, 메틸화, 미리스토일화, 산화, 팔미토일화, 페길화, 포스포판테테이닐화, 인산화, 프레닐화, 프로피오닐화, 레티닐리덴 스키프 염기 형성, S-글루타티오닐화, S-니트로실화, S-설페닐화, 셀렌화, 석시닐화, 설핀화, 우비퀴틴화, 및 C-말단 아미드화를 포함하나, 이에 한정되지 않는다. 해독후 변형은 펩타이드, 폴리펩타이드, 또는 단백질의 아미노 말단 및/또는 카복실 말단의 변형을 포함한다. 말단 아미노 그룹의 변형은 데스-아미노,

N-저급 알킬, N-디-저급 알킬, 및 N-아실 변형을 포함하나, 이에 한정되지 않는다. 말단 카복시 그룹의 변형은 아미드, 저급 알킬 아미드, 디알킬 아미드, 및 저급 알킬 에스테르 변형(예컨대, 여기서 저급 알킬은 C_1-C_4 알킬이다)을 포함하나, 이에 한정되지 않는다. 해독후 변형은 또한 아미노와 카복시 말단 사이에 속하는 아미노산의, 상술한 것과 같은 그러나 이에 한정되지 않는 변형을 포함한다. 해독 후 변형은 세포내 단백질의 "생물학", 예컨대, 이의 활성, 구조, 안정성, 또는 국재화를 조절할 수 있다. 포스포릴화는 가장 일반적인 해독후 변형이며 특히 세포 시그널링에 있어서, 단백질 조절에 중요한 역활을 한다(Prabakaran et al., 2012, Wiley Interdiscip Rev Syst Biol Med 4: 565-583). 글리코실화와 같은, 단백질에 당의 첨가는 단백질 폴딩(folding)을 촉진하고, 안정성을 개선시키며, 조절 기능을 변형시키는 것으로 밝혀졌다. 단백질에 대한 지질의 부착은 세포 막에 대한 표적화를 가능하도록 한다. 해독후 변형은 또한 하나 이상의 검출가능한 표지를 포함하기 위한 펩타이드, 폴리펩타이드, 또는 단백질 변형을 포함할 수 있다.

특정 구현예에서, 펩타이드, 폴리펩타이드, 또는 단백질은 단편화될 수 있다. 예를 들면, 단편화된 펩타이드는 생물학적 샘플과 같은 단백질을 단편화함으로써 수득할 수 있다. 펩타이드, 폴리펩타이드, 또는 단백질은 프로테아제 또는 엔도펩티다제에 의한 단편화를 포함하는, 당해 분야에 공지된 어떠한 수단에 의해서도 단편화될 수 있다. 일부 구현예에서, 펩타이드, 폴리펩타이드, 또는 단백질의 단편화는 특이적인 프로테아제 또는 엔도펩티다제의 사용에 의해 표적화된다. 특이적인 프로테아제 또는 엔도펩티다제는 특이적인 컨센서스(consensus) 서열(예컨대, ENLYFQ\S 컨센서스 서열에 대해 특이적인 TEV 프로테아제)에서 절단한다. 다른 구현예에서, 펩타이드, 폴리펩타이드, 또는 단백질의 단편화는 비-표적화되거나 비-특이적인 프로테아제 또는 엔도펩티다제이 사용에 의해 부작위화된다. 비-특이적인 프로테아제는 컨센서스 서열보다는 특이적인 아미노산 잔기에 결합하여 분해할 수 있다(예컨대, 프로테이나제 K는 비-특이적인 세린 프로테아제이다). 프로테이나제 및 엔도펩티다제는 당해 분야에 잘 공지되어 있으며 단백질 또는 폴리펩타이드를 보다 작은 펩타이드 단편으로 절단하는데 사용될 수 있는 이의 예는 프로테이나제 K, 트립신, 키모트립신, 펩신, 서몰라이신, 트롬빈, 인자 Xa, 푸린, 엔도펩티다제, 파파인, 펩신, 서브틸리신, 엘라스타제, 엔테로키나제, Genenase™ I, 엔도프로테이나제 LysC, 엔도프로테이나제 AspN, 엔도프로테이나제 GluC, 등을 포함한다(Granvogl et al., 2007, Anal Bioanal Chem 389: 991-1002). 특정 구현예에서, 펩타이드, 폴리펩타이드, 또는 단백질은 프로테이나제 K에 의해, 또는 임의로 신속한 활성화를 가능하도록 하는 프로테이나제 K의 열분해성 버젼으로 단편화된다. 프로테이나제 K는 우레아 및 SDS와 같은 변성 시약 속에서 상당히 안정하여, 완전히 변성된 단백질의 분해를 가능하도록 한다. 펩타이드로의 단백질 및 폴리펩타이드 단편화는 DNA 태그 또는 DNA 기록 태그의 부착 전 또는 후에 수행될 수 있다.

화학적 시약을 또한 사용하여 단백질을 펩타이드 단편으로 소화시킬 수 있다. 화학적 시약은 특이적인 아미노산 잔기에서 절단할 수 있다(예컨대, 시아노겐 브로마이드는 메티오닌 잔기의 C-말단에서 펩타이드 결합을 가수분해한다). 폴리펩타이드 또는 단백질을 보다 작은 펩타이드로 단편화하기 위한 화학적 시약은 시아노겐 브로마이드(CNBr), 하이드록실아민, 하이드라진, 포름산, BNPS-스카톨[2-(2-니트로페닐설페닐)-3-메틸인돌], 요오도소벤조산, ·NTCB +Ni(2-니트로-5-티오시아노벤조산) 등을 포함한다.

특정 구현예에서, 효소적 또는 화학적 절단 후, 수득되는 펩타이드 단편은 대략적으로 동일한 바람직한 길이, 예컨대, 약 10개 아미노산 내지 약 70개 아미노산, 약 10개 아미노산 내지 약 60개 아미노산, 약 10개 아미노산 내지 약 50개 아미노산, 약 10개 내지 약 40개 아미노산, 약 10개 내지 약 30개 아미노산, 약 20개 아미노산 내지 약 70개 아미노산, 약 20개 아미노산 내지 약 60개 아미노산, 약 20개 아미노산 내지 약 50개 아미노산, 약 20 내지 약 40개 아미노산, 약 20개 내지 약 30개 아미노산, 약 30개 아미노산 내지 약 70개 아미노산, 약 30개 아미노산 내지 약 60개 아미노산, 약 30개 아미노산 내지 약 50개 아미노산, 또는 약 30개 아미노산 내지 약 40개 아미노산이다. 절단 반응은 단백질 또는 폴리펩타이드 샘플을 프로테이나제 또는 엔도펩티다제 절단 부위를 함유하는 펩타이드 서열을 포함하는 짧은 시험 FRET(형광성 공명 에너지 전달) 펩타이드로 스파이킹(spiking)함으로써, 바람직하게는 실시간으로 모니터링할 수 있다. 완전한 FRET 펩타이드에서, 형광성 그룹 및 퀀처(quencher) 그룹은 절단 부위, 및 퀀처와 형광성을 저하시키기 위한 형광단 납 사이의 형광서 공명 에너지 전달을 함유하는 펩타이드 서열의 한쪽 말단에 부착된다. 프로테아제 또는 엔도펩티다제에 의한 시험 펩타이드의 절단 시, 퀀처 및 형광단은 분리되어 형광성에 있어서 큰 증가를 제공한다. 절단 반응은 특정의 형광성 강도가 달성되는 경우 중단되어 재생가능한 절단 종점이 달성되도록 한다.

거대분자의 샘플(예컨대, 펩타이드, 폴리펩타이드, 또는 단백질)은 고체 지지체에 부착되기 전에 단백질 분획화 방법을 겪을 수 있으며, 여기서 단백질 또는 펩타이드는 세포 국재화, 분자량, 소수성, 또는 등전점, 또는 단백질 농축 방법과 같은 하나 이상의 특성에 의해 분리된다. 대안적으로, 또는 추가로, 단백질 농축 방법을 사용하여 특이적인 단백질 또는 펩타이드를 선택하거나(참고: 예컨대, Whiteaker et al., 2007,al. Biochem. 362:44-

54, 이의 전문은 참고로 포함됨) 특수한 해독후 변형을 위해 선택할 수 있다(참고: 예컨대, Huang et al., 2014. J. Chromatogr. A 1372:1-17, 이의 전문은 참고로 포함됨). 대안적으로, 면역글로불린, 또는 IgG와 같은 면역글로불린(Ig) 동형과 같은 단백질의 특수한 부류 또는 부류들은 분석을 위해 친화성 농축되거나 선택될 수 있다. 면역글로불린 분자의 경우, 서열의 분석 및 친화성 결합에 포함된 초가변성 서열의 풍부성 또는 빈도는, 이들이 질환 진행에 대한 반응시 변하고 건강, 면역, 및/도는 질환 표현형과 관련되므로, 특히 중요하다. 과도하게 풍부한 단백질은 또한 표준 면역친화성 방법을 사용하여 샘플로부터 감할 수 있다. 풍부한 단백질의 고갈은 혈장 샘플의 경우 유용할 수 있으며, 여기서 단백질 구성성분의 80% 이상은 알부민 및 면역글로불린이다. PROTIA 및 PROT20(Sigma-Aldrich)과 같은 몇가지 시판 제품이 과도하게 풍부한 단백질의 혈장 샘플의 고갈에 이용가능하다.

특정 구현예에서, 거대분자는 단백질 또는 폴리펩타이드로 구성된다. 일 구현예에서, 단백질 또는 폴리펩타이드는 표준 아민 커플링 화학을 통해 DNA 기록 태그로 표지된다(참고: 예컨대, 도 2b, 2c, 28, 29, 31, 40). ε-아미노 그룹(예컨대, 라이신 잔기의) 및 N-말단 아미노 그룹은 반응의 pH에 의존하여, 아민-반응성 커플링제로 표지시키기에 특히 민감하다(Mendoza and Vachet 2009). 특수한 구현예에서(참고: 예컨대, 도 2b 및 도 29), 기록 태그는 반응성 모이어티(예컨대, 고체 표면, 다기능성 링커, 또는 거대분자에 대한 접합을 위해), 링커, 보편적인 프라이밍 서열, 바코드(예컨대, 구획 태그, 분배 바코드, 샘플 바코드, 분획 바코드, 또는 이의 어떠한 조합), 임의의 UMI, 및 암호화 태그로 및/또는 암호화 태그로부터 정보 이전을 촉진시키기 위한 스페이서(Sp) 서열로 구성된다. 다른 구현예에서, 단백질은 보편적인 DNA 태그로 우선 표지될 수 있으며, 바코드-Sp 서열(샘플, 구획, 슬라이드 상의 물리적 위치 등을 나타냄)이 효소적 또는 화학적 커플링 단계를 통해 후에 단백질에 부착된다(참고: 예컨대, 도 20, 30, 31, 40). 보편적인 DNA 태그는 단백질 또는 폴리펩타이드 거대분자를 표지하는데 사용된 뉴클레오타이드의 짧은 서열을 포함하며 바코드(예컨대, 구획 태그, 기록 태그 등)의 부착점으로서 사용될 수 있다. 예를 들면, 기록 태그는 이의 말단에서 보편적인 DNA 태그에 대해 상보성인 서열을 포함할 수 있다. 특정 구현예에서, 보편적인 DNA 태그는 보편적인 프라이밍 서열이다. 표지된 단백질 상의 보편적인 DNA 태그의 기록 태그내(예컨대, 비드에 결합된) 상보성 서열에 하이브리드화시, 어닐링된 보편적인 DNA 태그는 프라이머 연장을 통해 연장되어, 기록 태그 정보를 DNA 태그된 단백질에 이전시킬 수 있다. 특수한 구현예에서, 단백질은 펩타이드로 프로테이나제 분해 전에 보편적인 DNA 태그로 표지된다. 소화물로부터 표지된 펩타이드 상의 보편적인 DNA 태그는 이후 정보성이고 효과적인 기록 태그로 전환될 수 있다.

특정 구현예에서, 단백질 거대분자는 친화성 포획 시약(및 임의로 공유결합으로 가교결합된)에 의해 고체 지지체에 고정될 수 있으며, 여기서 기록 태그는 친화성 포획 시약과 직접 관련되거나, 대안적으로, 단백질은 기록 태그를 지닌 고체 지지체에 직접 고정될 수 있다(참고: 예컨대, 도 2c).

V. 고체 지지체

본 개시내용의 거대분자는 고체 지지체의 표면(또한 "기질 표면"으로 지칭됨)에 결합된다. 고체 지지체는 비드, 마이크로비드, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 슈소 웨이퍼 칩, 유동 셀, 유통 통과 칩, 시그널 신호전달 전자 장치를 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 간섭계 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구를 포함하나, 이에 한정되지 않는, 어떠한 다공성 또는 비-다공성 지지체 표면일 수 있다. 고체 지지체용 물질은 아크릴아미드, 아가로즈, 셀룰로즈, 니트로셀룰로즈, 유리, 금, 석영, 폴리스티렌, 폴리에틸렌 비닐 아세테이트, 폴리프로필렌, 폴리메타크릴레이트, 폴리에틸렌, 폴리에틸렌 옥사이드, 폴리실리케이트, 폴리카보네이트, 테플론, 플루오로카본, 나일론, 규소 고무, 다가무수물, 폴리글리콜산, 폴리악트산, 폴리오르토에스테르, 기능화된 실란, 폴리프로필푸머레이트, 콜라겐, 글리코스아미노글리칸, 폴리아미노산, 또는 이의 어떠한 조합을 포함하나 이에 한정되지 않는다. 고체 지지체는 또한 박 필름, 막, 병, 접시, 섬유, 직조 섬유, 튜브, 입자, 비드, 미세입자, 또는 이의 어떠한 조합과 같은 형상화된 중합체를 포함한다. 예를 들면, 고체 표면이 비드인 경우, 비드는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 제어된 공극 비드를 포함할 수 있으나, 이에 한정되지 않는다.

특정의 구현예에서, 고체 지지체는 유동 셀이다. 유동 셀 구조는 상이한 차 세대 서열분석 플랫폼에서 변할 수 있다. 예를 들면, 일루미나 유동 셀은 현미경 슬라이드와 유사한 평면의 임의로 투명한 표면이며, 이는 이의 표면에 결합된 올리고뉴클레오타이드 앵커(anchor)의 론(lawn)을 함유한다. 주형 DNA는 유동 셀 표면에서 올리고뉴클레오타이드에 대해 상보성인 말단에 연결된 어댑터를 포함한다. 채택된(adapted) 단일 가닥 DNA는 유동 셀에 결합되어 서열분석 전에 고체-상 "브릿지" PCR에 의해 증폭된다. 454 유동 셀(454 Life Sciences)은 "피코역가" 플레이트, ~1백만 6천 75-피코리터 웰을 지닌 섬유광 슬라이드(fiber optic slide)를 지지한다. 전단된

(sheared) 주형 DNA의 각각의 개개 분자는 별도의 비드 상에 포획되고, 각각의 비드는 오일 유화액 내에서 수성 PCR 반응 혼합물의 개별 소적 내에서 구획화된다. 주형은 PCR에 의해 비드 표면에서 클론적으로 증폭되며, 이후 주형-로딩된 비드는 이상적으로 웰당 1 또는 더 적은 비드를 사용하여, 서열분석 반응을 위한 피코역가 플레이트의 웰로 분배된다. 454 시스템과 같은 Applied Biosystems로부터의 SOLiD(지지된 올리고뉴클레오타이드 연결 및 검출) 장치는 유화액 PCR에 의해 주형 분자를 증폭시킨다. 증폭된 주형을 함유하지 않는 비드를 추려내는 단계 후, 비드-결합된 주형을 유동 셀 위에 침착시킨다. 유동 셀은 또한 TWIST™ DNA 합성 컬럼(Glen Research)과 같은 단순한 필터 프릿(filter frit)일 수 있다.

특정 구현예에서, 고체 지지체는 개개의 비드 또는 다수의 비드를 지칭할 수 있는 비드이다. 일부 구현예에서, 비드는 하부 분석(예컨대, SOLiD 또는 454)에 사용될 선택된 차 세대 서열분석 플랫폼과 혼용성이다. 일부 구현예에서, 고체 지지체는 아가로즈 비드, 상자성 비드, 폴리스티렌 비드, 중합체 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 유리 비드, 또는 제어된 공극 비드이다. 추가의 구현예에서, 비드는 결합 기능성(예컨대, 아민 그룹, 바이오틴 표지된 거대분자, 항체에 결합하기 위한 스트렙타비딘과 같은 친화성 리간드)으로 코팅되어 거대분자에 대한 결합을 촉진할 수 있다.

단백질, 폴리펩타이드, 또는 펩타이드는 공유결합성 및 비-공유결합성 상호작용, 또는 이의 어떠한 조합을 포함하는, 당해 분야에 공지된 어떠한 수단에 의해서 간접적으로 또는 직접적으로 고체 지지체에 결합될 수 있다(참고: 예컨대, Chan et al., 2007, PLoS One 2:e1164; Cazalis et al., Bioconj. Chem. 15:1005-1009; Soellner et al., 2003, J. Am. Chem. Soc. 125:11790-11791; Sun et al., 2006, Bioconjug. Chem. 17-52-57; Decreau et al., 2007, J. Org. Chem. 72:2794-2802; Camarero et al., 2004, J. Am. Chem. Soc. 126:14730-14731; Girish et al., 2005, Bioorg. Med. Chem. Lett. 15:2447-2451; Kalia et al., 2007, Bioconjug. Chem. 18:1064-1069; Watzke et al., 2006, Angew Chem. Int. Ed. Engl. 45:1408-1412; Parthasarathy et al., 2007, Bioconjugate Chem. 18:469-476; 및 Bioconjugate Techniques, G. T. Hermanson, Academic Press (2013), 및 이들의 전문은 각각 본원에 참고로 포함된다). 예를 들면, 펩타이드는 연결 반응에 의해 고체 지지체에 결합될 수 있다. 대안적으로, 고체 지지체는 펩타이드를 고체 지지체에 직접 또는 간접적으로 결합시키는 것을 촉진하는 제제 또는 코팅을 포함할 수 있다. 단백질, 핵산, 탄수화물 및 소 분자를 포함하는 어떠한 적합한 분자 또는 물질도 이러한 목적을 위해 사용할 수 있다. 예를 들면, 일 구현예에서 제제는 친화성 분자이다. 다른 예에서, 제제는 아지드 그룹이고, 당해 그룹은 다른 분자내에서 알키닐 그룹과 반응하여 고체 지지체와 다른 분자 사이의 연합 또는 결합을 촉진할 수 있다.

단백질, 폴리펩타이드, 또는 펩타이드는 "클릭 화학"으로 지칭된 방법을 사용하여 고체 지지체에 결합시킬 수 있다. 이러한 목적을 위해 신속하고 실질적으로 비가역성인 어떠한 반응도 단백질, 폴리펩타이드, 또는 펩타이드를 고체 지지체에 부착시키는데 사용할 수 있다. 예시적인 반응은 트리아졸을 형성시키는 아지드와 알킨의 구리 촉매된 반응(Huisgen 1,3-이극성 고리첨가반응), 변형률(strain)-촉진된 아지드 알킨 고리첨가반응(SPAAC), 디엔과 디엔필(디엘스-앨더(Diels-Alder)) 반응, 변형률-촉진된 알킨-니트론 고리첨가반응, 쇄 알켄과 아지드, 테트라진 또는 테트라졸의 반응, 알킨과 아지드[3+2] 고리첨가반응, 알켄과 테트라진 역 전자 요구 디엘스-앨더(inverse electron demand Diels-Alder: IEDDA) 반응(예컨대, m-테트라진(mTet) 과 트랜스-사이클로옥텐(TCO)), 알켄 및 테트라졸 광반응, 아지드와 포스핀의 스타우딩거 연결(Staudinger ligation), 및 친핵성 원자에 친핵성 공격에 의한 이탈 그룹의 치환과 같은 다양한 치환 반응(Horisawa 2014, Knall, Hollauf et al. 2014)을 포함한다. 예시적인 치환 반응은 아민과: 활성화된 에스테르; N-하이드록시석신이미드 에스테르; 이소시아네이트; 이소티오스시아네이트 등의 반응을 포함한다.

일부 구현예에서 거대분자 및 고체 지지체는 2개의 상보성 반응 그룹, 예를 들면, 앞서의 "클릭" 반응 중 하나의 생성물인 기능성 그룹의 반응에 의해 형성될 수 있는 기능성 그룹에 의해 결합된다. 다양한 구현예에서, 기능성 그룹은 알데하이드, 옥심, 하이드라존, 하이드라지드, 알킨,아민, 아지드, 사이클라지드, 아실할라이드, 니트릴, 니트론, 설프하이드릴, 니트릴, 니트론, 설프하이드릴 디설파이드, 설포닐 할라이드, 이소티오시아네이트, 이미도에스테르, 활성화된 에스테르(예컨대, N-하이드록시석신이미드 에스테르, 펜티노산 STP 에스테르), 케톤, α,β-불포화된 카보닐, 알켄, 말레이미드, α-할로이미드, 에폭사이드, 아지리딘, 테트라진, 테트라졸, 포스핀, 바이오틴 또는 티이란 기능 그룹과 상보성 반응 그룹의 반응에 의해 형성된다. 예시적인 반응은 아민(예컨대, 1급 아민)과 N-하이드록시석신이미드 에스테르 또는 이소티오시아네이트의 반응이다.

여전히 다른 구현예에서, 기능성 그룹은, 알켄, 에스테르, 아미드, 티오에스테르, 디설파이드, 카보사이클릭, 헤테로사이클릭 또는 헤테로아릴 그룹을 포함한다. 추가의 구현예에서, 기능성 그룹은 알켄, 에스테르, 아미드,

티오에스테르, 티오우레아, 디설파이드, 카보사이클릭, 헤테로사이클릭 또는 헤테로아릴 그룹을 포함한다. 다른 구현예에서, 기능성 그룹은 아미드 또는 티오우레아를 포함한다. 일부 보다 특이적인 구현예에서, 기능성 그룹은 트리아졸릴 기능성 그룹, 아미드, 또는 티오우레아 기능성 그룹이다.

바람직한 구현예에서, iEDDA 클릭 화학은 신속하고 낮은 투입 농도에서 고 수율을 전달하므로, 거대분자(예컨대, 단백질, 폴리펩타이드, 펩타이드)를 고체 지지체에 고정시키는데 사용된다. 다른 바람직한 구현에에서, m-테트라진은 개선된 결합 안정성을 가지므로, 테트라진 이외에 m-테트라진을 iEDDA 클릭 화학 반응에 사용한다.

바람직한 구현예에서, 기질 표면은 TCO로 기능성화되며, 기록 태그-표지된 단백질, 폴리펩타이드, 펩타이드는 부착된 m-테트라진 모이어티를 통해 TCO 코팅된 기질 표면에 고정된다(도 34).

단백질, 폴리펩타이드, 또는 펩타이드는 이의 C-말단, N-말단, 또는 내부 아미노산에 의해, 예를 들면, 아민, 카복실, 또는 설프하이드릴 그룹을 통해 고체 지지체의 표면에 고정될 수 있다. 아민 그룹에 커플링하는데 사용된 표준 활성화된 지지체는 CNBr-활성화된, NHS-활성화된, 알데하이드-활성화된, 아즐락톤-활성화된, 및 CDI-활성화된 지지체를 포함한다. 카복실 커플링에 사용된 표준 활성화된 지지체는 아민 지지체에 대한 카보디이미드-활성화된 카복실 모이어티 커플링을 포함한다. 시스테인 커플링은 말레이미드, 요오도아세틸, 및 피리딜 디설파이드 활성화된 지지체를 사용할 수 있다. 펩타이드 카복시 말단 고정화의 대안적 방식은 이들을 절단하지 않고 이들의 C-말단에서 라이신 또는 아르기닌 잔기를 함유하는 펩타이드에 결합하는 트립신의 촉매적으로 불활성인 유도체인, 안하이드로트립신을 사용한다.

특정 구현예에서, 단백질, 폴리펩타이드, 또는 펩타이드는 단백질, 폴리펩타이드, 또는 펩타이드의 라이신 그룹에 대해 고체 표면 결합된 링커의 공유결합성 부착을 통해 고체 지지체에 고정된다.

기록 태그는 고체 지지체에 고정화 전 또는 후에 단백질, 폴리펩타이드, 또는 펩타이드에 부착될 수 있다. 예를 들면, 단백질, 폴리펩타이드, 또는 펩타이드는 기록 태그로 우선 표지될 수 있으며 이후 커플링을 위한 2개의 기능성 모이어티를 포함하는 기록 태그를 통해 고체 표면에 고정될 수 있다(참고: 도 28). 기록 태그의 하나의 기능성 모이어티는 단백질에 커플링하며, 다른 기능성 모이어티는 기록 태그-표지된 단백질을 고체 지지체에 고정시킨다.

대안적으로, 단백질, 폴리펩타이드, 또는 펩타이드는 단백질, 폴리펩타이드 또는 펩타이드를 기록 태그로 표지하기 전에 고체 지지체에 고정된다. 예를 들면, 단백질은 우선 클릭 화학 모이어티와 같은 반응성 그룹으로 유도체화될 수 있다. 활성화된 단백질 분자는 이후 적합한 고체 지지체에 부착된 후 상보성 클릭 화학 모이어티를 사용하여 기록 태그로 표지된다. 예로서, 알킨 및 mTet 모이어티로 유도체화된 단백질은 아지드 및 TCO로 유도체화된 비드에 고정될 수 있으며 아지드 및 TCO로 표지된 기록 태그에 부착된다.

거대분자(예컨대, 단백질, 폴리펩타이드, 또는 펩타이드)를 고체 지지체에 부착시키기 위한 본원에 제공된 방법을 또한 사용하여 기록 태그를 고체 지지체에 부착시키거나 기록 태그를 거대분자(예컨대, 단백질 폴리펩타이드, 또는 펩타이드)에 부착시킬 수 있다.

특정 구현예에서, 고체 지지체의 표면은 패시베이팅(passivating)(차단)되어 결합제에 대한 비-특이적인 흡수를 최소화한다. "패시베이팅된" 표면은 결합제의 비-특이적인 결합을 최소화하기 위해 물질의 외부 층으로 처리된 표면을 지칭한다. 표면을 패시베이팅하는 방법은 폴리에틸렌 글리콜(PEG)(Pan et al., 2015, Phys. Biol. 12:045006), 폴리실록산(예컨대, Pluronic F-127), 스타 중합체(star polymer)(예컨대, star PEG)(Groll et al., 2010, Methods Enzymol. 472:1-18), 소수성 디클로로디메틸실란(DDS) + 자가-조립된 트윈-20(Hua et al., 2014, Nat. Methods 11:1233-1236), 및 다이아몬드-유사 탄소(DLC), DLC + PEG(Stavis et al., 2011, Proc. Natl. Acad. Sci. USA 108:983-988)를 지닌 패시베이팅 표면을 포함하는, 형광성 단일 분자 분석 문헌으로부터의 표준 방법을 포함한다. 공유결합성 표면 변형 이외에, 트윈-20, 용액 중 폴리실록산(Pluronic series), 폴리비닐 알코올, (PVA), 및 BSA 및 카제인과 같은 단백질과 같은 계면활성제를 포함하는, 다수의 패시베이팅 제제를 사용할 수 있다. 대안적으로, 단백질, 폴리펩타이드, 또는 펩타이드의 밀도는 단백질, 폴리펩타이드 또는 펩타이드를 고체 지지체에 고정시키는 경우 경쟁인자 또는 "더미" 반응성 분자를 스파이킹함으로써 고체 기질의 용적내에서 또는 표면에서 적정할 수 있다(참고: 도 36a).

다수의 거대분자가 동일한 고체 지지체에 고정되는 특정 구현예에서, 예컨대, 결합제가 제1의 거대분자에 결합하고 이의 암호화 태그 정보가 제1의 거대분자와 관련된 기록 태그보다는 이웃하는 거대분자와 관련된 기록 태그로 이전되는 경우, 거대분자는 적절하게 떨어져서 가교-결합 또는 분자간 사건의 발생을 감소시키거나 방지할

수 있다. 고체 지지체 상에서 이격된 거대분자(예컨대, 단백질, 폴리펩타이드, 또는 펩타이드 이격)를 제어하기 위해, 기능성 커플링 그룹(예컨대, TCO)의 밀도를 기질 표면에서 적정할 수 있다(참고: 도 34). 일부 구현예에서, 다수의 거대분자는 고체 지지체의 표면 위에서 또는 용적(예컨대, 다공성 지지체) 내에서 약 50 nm 내지 약 500 nm, 또는 약 50 nm 내지 약 400 nm, 또는 약 50 nm 내지 약 300 nm, 또는 약 50 nm 내지 약 200 nm, 또는 약 50 nm 내지 약 100 nm의 거리로 이격될 수 있다. 일부 구현예에서, 다수의 거대분자는 고체 지지체의 표면에서 적어도 50 nm, 적어도 60 nm, 적어도 70 nm, 적어도 80 nm, 적어도 90 nm, 적어도 100 nm, 적어도 150 nm, 적어도 200 nm, 적어도 250 nm, 적어도 300 nm, 적어도 350 nm, 적어도 400 nm, 적어도 450 nm, 또는 적어도 500 nm의 평균 거리로 이격된다. 일부 구현예에서, 다수의 거대분자는 고체 지지체의 표면에서 적어도 50 nm의 평균 거리로 이격된다. 일부 구현예에서, 거대분자는 고체 지지체의 표면 또는 용적 내에서, 경험적으로, 분자간 내지 분자내 사건의 상대적인 빈도가 <1:10; <1:100; <1:1,000; 또는 <1:10,000이 되도록 이격된다. 적합한 간격 빈도는 기능성 검정(참고: 실시예 23)을 사용하여 경험적으로 측정할 수 있으며, 희석에 의해 및/또는 기질 표면에서 부착 부위에 대해 경쟁하는 "더미" 스페이서 분자를 스파이킹함으로써 달성할 수 있다.

예를 들면, 도 34에 나타낸 바와 같이, PEG-5000(MW ~ 5000)을 사용하여 기질 표면(예컨대, 비드 표면)에서 펩타이드 사이에 격자 사이 공간(interstitial space)을 차단한다. 또한, 펩타이드를 PEG-5000 분자에 또한 부착된 기능성 모이어티에 커플링한다. 바람직한 구현예에서, 이는 NHS-PEG-5000-TCO + NHS-PEG-5000-메틸의 혼합물을 아민-유도체화된 비드에 커플링함으로써 달성된다(참고: 도 34). 2개의 PEG(TCO 대 메틸) 사이의 화학량론적 비를 적정하여 기질 표면 위의 기능성 커플링 모이어티(TCO 그룹)의 적절한 밀도를 생성한다; 메틸-PEG는 커플링에 대해 불활성이다. TCO 그룹들 사이의 유효 간격은 표면에서 TCO 그룹의 밀도를 측정함으로써 계산할 수 있다. 특정 구현예에서, 고체 지지체 상의 커플링 모이어티(예컨대, TCO) 사이의 평균 간격은 적어도 50 nm, 적어도 100 nm, 적어도 250 nm, 또는 적어도 500 nm이다. 비드의 PEG5000-TCO/메틸 유도체화 후, 표면에서 과도한 NH_2 그룹은 반응성 무수물(예컨대 아세트산 또는 석신산 무수물)로 퀀칭(quenching)된다.

VI. 기록 태그

적어도 하나의 기록 태그는 거대분자와 관련되거나 이와 직접적으로 또는 간접적으로 동시-국재화되며 고체 지지체에 결합된다(참고: 예컨대, 도 5). 기록 태그는 DNA, RNA, PNA, γPNA, GNA, BNA, XNA, TNA, 폴리뉴클레오타이드 유사체, 또는 이이 조합을 포함할 수 있다. 기록 태그는 단일 가닥, 또는 부분적으로 또는 완전히 이중 가닥일 수 있다. 기록 태그는 평활 말단 또는 오버행 말단을 가질 수 있다. 특정 구현예에서, 거대분자에 대한 결합제의 결합시, 결합제의 암호화 태그의 확인 정보는 기록 태그로 이전되어 연장된 기록 태그를 생성한다. 연장된 기록 태그에 대한 추가의 연장은 후속적인 결합 주기에서 이루어질 수 있다.

기록 태그는 공유결합성 및 비-공유결합성 상호작용, 또는 이의 어떠한 조합을 포함하는, 당해 분야에 공지된 어떠한 수단에 의해서도 고체 지지체에, 직접 또는 간접적으로(예컨대, 링커를 통해) 결합될 수 있다. 예를 들면, 기록 태그는 연결 반응에 의해 고체 지지체에 결합될 수 있다. 대안적으로, 고체 지지체는 기록 태그를 고체 지지체에 직접 또는 간접적으로 결합시키는 것을 촉진하는 제제 또는 코팅을 포함할 수 있다. 핵산 분자를 고체 지지체(예컨대, 비드)에 고정화시키는 전략는 미국 특허 제5,900,481호; Steinberg et al. (2004, Biopolymers 73:597-605); Lund et al., 1988 (Nucleic Acids Res. 16: 10861-10880); 및 Steinberg et al. (2004, Biopolymers 73:597-605)에 기술되어 있으며, 이들 각각은 이의 전문이 참고로 본원에 포함된다.

특정 구현예에서, 거대분자(예컨대, 펩타이드) 및 관련된 기록 태그의 동시-국재화는 거대분자 및 기록 태그를 고체 지지체 표면에 직접 부착된 이기능성 링커에 접합시킴으로써 달성된다. Steinberg et al. (2004, Biopolymers 73:597-605). 추가의 구현예에서, 삼기능성 모이어티를 사용하여 고체 지지체(예컨대, 비드)를 유도체화하고, 수득되는 이기능성 모이어티를 거대분자 및 기록 태그 둘 다에 커플링시킨다.

거대분자 및 고체 지지체의 부착을 위해 기술된 것과 같은 방법 및 시약(예컨대, 클릭 화학 시약 및 광친화성 표지화 시약)을 또한 기록 태그의 부착에 사용할 수 있다.

특수한 구현예에서, 단일의 기록 태그는 거대분자(예컨대, 펩타이드)에, 바람직하게는 탈-차단된 N- 또는 C-말단 아미노산에 대한 부착을 통해 부착시킬 수 있다. 다른 구현예에서, 다수의 기록 태그는 거대분자(예컨대, 단백질, 폴리펩타이드, 또는 펩타이드), 바람직하게는 라이신 잔기 또는 펩타이드 골격에 부착된다. 일부 구현예에서, 다수의 기록 태그로 표지된 거대분자(예컨대, 단백질 또는 폴리펩타이드)는 보다 작은 펩타이드로 단편화되거나 소화되며, 각각의 펩타이드는 평균적으로 하나의 기록 태그로 표지된다.

특정 구현예에서, 기록 태그는 임의의, 유일한 분자 확인인자(UMI)를 포함하며, 이는 UMI가 관련된 각각의 거대

분자(예컨대, 단백질, 폴리펩타이드, 펩타이드)에 대한 유일한 확인인자 태그를 제공한다. UMI는 길이가 약 3 내지 약 40개 염기, 약 3 내지 약 30개 염기, 약 3 내지 약 20개 염기, 또는 약 3 내지 약 10개 염기, 또는 약 3 내지 약 8개 염기일 수 있다. 일부 구현예에서, UMI는 길이가 약 3개 염기, 4개 염기, 5개 염기, 6개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개 염기, 15개 염기, 16개 염기, 17개 염기, 18개 염기, 19개 염기, 20개 염기, 25개 염기, 30개 염기, 35개 염기, 또는 40개 염기이다. UMI를 사용하여 다수의 연장된 기록 태그로부터의 서열분석 데이타를 탈-컨벌루트시켜 개개의 거대분자로부터 서열 판독물을 확인할 수 있다. 일부 구현예에서, 거대분자의 라이브러리 내에서, 각각의 거대분자는 단일의 기록 태그와 관련되며, 각각의 기록 태그는 유일한 UMI를 포함한다. 다른 구현예에서, 기록 태그의 다수의 카피는 단일의 거대분자와 관련되며, 기록 태그의 각각의 카피는 동일한 UMI를 포함한다. 일부 구현예에서, UMI는 결합제의 암호화 태그내에서 스페이서 또는 암호기 서열과는 상이한 염기 서열을 가짐으로서 서열 분석 동안 이들 성분을 구별하기에 용이하다.

특정 구현예에서, 기록 태그는 바코드, 예컨대, 존재하는 경우 UMI 이외의 다른 것을 포함한다. 바코드는 길이가 약 3 내지 약 30개 염기, 약 3 내지 약 25개 염기, 약 3 내지 약 20개 염기, 약 3 내지 약 10개 염기, 약 3 내지 약 10개 염기, 약 3 내지 약 8개 염기인 핵산 분자이다. 일부 구현예에서, 바코드는 약 3개 염기, 4개 염기, 5개 염기, 6개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개 염기, 15개 염기, 20개 염기, 25개 염기, 또는 30개 염기이다. 일 구현예에서, 바코드는 다수의 샘플 또는 라이브러리의 멀티플렉싱 서열분석을 허용한다. 바코드는 이로부터 거대분자(예컨대, 펩타이드)가 기원하는 분배, 분획, 구획, 샘플, 공간 위치, 또는 라이브러리를 확인하는데 사용될 수 있다. 바코드를 사용하여 멀티플렉스화된 서열 데이타를 디컨벌루트하고 개개의 샘플 또는 라이브러리로부터 서열 판독물을 확인한다. 예를 들면, 바코드화된 비드는 예컨대, 프로테옴을 분배시키려는 목적으로, 샘플의 유화액 및 분배를 포함하는 방법에 유용하다.

바코드는 구획 태그를 나타낼 수 있으며, 여기서 고체 지지체 상의 소적, 미세웰, 물리적 영역과 같은 구획이 유일한 바코드로 지정된다. 구획과 특이적인 바코드의 연합은 구획 속에서 단일의 바코드화된 비드의 캡슐화에 의해, 예컨대, 바코드화된 소적을 구획에 직접 침지시키거나 첨가함에 의해, 바코드 시약을 구획에 직접 인쇄하거나 투입함에 의해 등과 같은 어떠한 다수의 방법으로도 달성할 수 있다. 구획내 바코드 시약을 사용하여 구획-특이적인 바코드를 구획내 이의 거대분자 또는 단편에 가할 수 있다. 구획내로의 단백질 분배에 응용시, 바코드를 사용하여 구획내 이들의 원래의 단백질 분자로 다시 분석된 펩타이드를 맵핑할 수 있다. 이는 단백질 확인을 크게 용이하도록 한다. 구획 바코드를 또한 사용하여 단백질 복합체를 확인할 수 있다.

다른 구현예에서, 구획의 집단의 소세트에 존재하는 다수의 구획을 소세트를 나타내는 유일한 바코드에 지정할 수 있다.

대안적으로, 바코드는 샘플 확인 바코드일 수 있다. 샘플 바코드는 단일 반응 용기 속에서 샘플의 세트의 멀티플렉스화된 분석에 유용하거나 단일의 고체 기질 또는 고체 기질의 수집물(예컨대, 평면 슬라이드, 단일 튜브 또는 용기 등에 함유된 비드의 집단)에 고정된다. 많은 상이한 샘플로부터의 거대분자는 샘플-특이적인 바코드를 지닌 기록 태그로 표지될 수 있으며, 이후, 모든 샘플은 고체 지지체에 대한 고정화, 사이클릭 결합, 및 기록 태그 분석 전에 함께 혼주된다. 대안적으로, 샘플은 DNA-암호화된 라이브러리, 및 DNA-암호화딘 라이브러리의 PCR 증폭 동안 부착된 샘플 바코드의 생성 후까지 별도로 유지될 수 있으며, 이후 서열분석 전에 함께 혼합된다. 이러한 시도는 상이한 풍부성 부류의 분석물(예컨대, 단백질)을 검정하는 경우 유용할 수 있다. 예를 들면, 샘플은 분할되어 바코드화될 수 있으며, 하나의 부위는 결합제를 사용하여 낮은 풍부성 분석물로 가공되고, 다른 부위는 결합제를 사용하여 보다 더 큰 풍부성 분석물로 가공된다. 특수한 구현예에서, 이러한 시도는 단백질 분석물의 표준 발현 수준의 "스위트 스폿(sweet spot) 내에 있는 특수한 단백질 분석물 검정의 역학적 범위를 조절하는 것을 돕는다.

특정 구현예에서, 다수의 상이한 샘플로부터의 펩타이드, 폴리펩타이드, 또는 단백질은 샘플-특이적인 바코드를 함유하는 기록 태그로 표지된다. 다중-샘플 바코드화된 펩타이드, 폴리펩타이드, 또는 단백질은 사이클릭 결합 반응 전에 함께 혼합될 수 있다. 이러한 방식으로, 디지탈 역상 단백질 배열(RPPA)에 대한 고도로-멀티플렉스화된 대안이 효과적으로 생성된다(Guo, Liu et al. 2012, Assadi, Lamerz et al. 2013, Akbani, Becker et al. 2014, Creighton and Huang 2015). 디지탈 RPPA-유사 검정의 생성은 해독 조사, 생물마커 입증, 약물 발견, 임상, 및 정밀 의약에서 다수의 응용을 갖는다.

특정 구현예에서, 기록 태그는 보편적인 프라이밍 부위, 예컨대, 전방 또는 5' 보편적인 프라이밍 부위를 포함

한다. 보편적인 프라이밍 부위는 라이브러리 증폭 반응을 프라이밍하고/하거나 서열분석하기 위해 사용될 수 있는 핵산 서열이다. 보편적인 프라이밍 부위는 PCR 증폭에 대한 프라이밍 부위, 유동 셀 표면(예컨대, 일루미나 차세대 서열분석)에서 상보성 올리고뉴클레오타이드에 어닐링하는 유동 셀 어댑터 서열, 서열분석 프라이밍 부위, 또는 이의 조합을 포함할 수 있으나, 이에 한정되지 않는다. 보편적인 프라이밍 부위는 약 10개 염기 내지 약 60개 염기일 수 있다. 일부 구현예에서, 보편적인 프라이밍 부위는 일루미나 P5 프라이머(5'-AATGATACGGCGACCACCGA-3' - 서열 번호: 133) 또는 일루미나 P7 프라이머(5'-CAAGCAGAAGACGGCATACGAGAT - 3' - 서열 번호: 134)를 포함한다.

특정 구현예에서, 기록 태그는 이의 말단, 예컨대, 3' 말단에 스페이서를 포함한다. 본원에 사용된 바와 같은, 기록 태그의 내용에서 스페이서 서열에 대한 참고는 이의 동종의 결합제와 관련된 스페이서 서열, 또는 이의 동종의 결합제와 관련된 스페이서 서열에 대해 상보성인 스페이서 서열을 포함한다. 기록 태그에서 말단, 예컨대, 3', 스페이서는 제1의 결합 주기 동안(예컨대, 프라이머 연장 또는 점성 말단 연결을 위한 상보성 스페이서 서열의 어닐링을 통해) 이의 암호화 태그로부터 기록 태그로 동종의 결합제의 확인 정보의 이전을 허용한다.

일 구현예에서, 스페이서 서열은 길이가 약 1-20개 염기, 길이가 약 2-12개 염기, 또는 길이가 5-10개 염기이다. 스페이서의 길이는 암호화 태그 정보를 기록 태그로 이전시키기 위한 프라이머 연장 반응의 온도 및 반응 조건과 같은 인자에 의존할 수 있다.

바람직한 구현예에서, 기록내 스페이서 서열은 기록 태그내 다른 영역에 대해 최소의 상보성을 갖도록 설계되며, 유사하게 암호화 태그 내 스페이서 서열은 암호화 태그내 다른 영역에 대해 최소의 상보성을 가질 수 있다. 다시 말해서, 기록 태그 및 암호화 태그의 스페이서 서열은 기록 태그 또는 암호화 태그 속에 존재하는 유일한 분자 확인인자, 바코드(예컨대, 구획, 분배, 샘플, 공간 위치), 보편적인 프라이머 서열, 암호기 서열, 주기 특이적인 서열 등과 같은 성분에 대해 최소의 서열 상보성을 가질 수 있다.

결합제 스페이서에 대해 기술된 바와 같이, 일부 구현예에서, 거대분자의 라이브러리와 관련된 기록 태그는 일반적인 스페이서 서열을 공유한다. 다른 구현예에서, 거대분자의 라이브러리와 관련된 기록 태그는 이들의 종종의 결합제의 결합 주기 특이적인 스페이서 서열에 대해 상보성인 결합 주기 특이적인 스페이서 서열을 가지며, 이는 비-콘카테네이트화된 연장된 기록 태그를 사용하는 경우 유용할 수 있다(참고: 도 10).

연장된 기록 태그의 수집은 이러한 사실 후에 콘카테네이트될 수 있다(참고: 예컨대, 도 10). 결합 주기가 완료된 후, 비드 고체 지지체(각각의 비드는 평균 비드당 1개 또는 1개 이하의 거대분자를 포함하고, 각각의 거대분자는 거대 분자의 부위에서 동시-국재화된 연장된 기록 태그의 수집물을 갖는다)는 유화액 속에 둔다. 유화액은 각각의 소적이, 평균적으로 최대 1개 비드에 의해 점유되도록 형성된다. 임의의 조립 PCR 반응은 유화액 속에서 수행되어 바드 상의 거대분자와 동시-국재화된 연장된 기록 태그를 증폭시키고 이들을 별도의 연장된 기록 태그 상의 상이한 주기 특이적인 서열 사이에서 프라이밍에 의해 동일-선형의 순서로 조립한다(Xiong, Peng et al. 2008). 이후 유화액은 깨져서 조립된 연장된 기록 태그가 서열분석된다.

다른 구현예에서, DNA 기록 태그는 보편적인 프라이밍 서열(U1), 하나 이상의 바코드 서열(BCs), 및 제1의 결합 주기에 대해 특이적인 스페이서 서열(Sp1)로 구성된다. 제1의 결합 주기에서, 결합제는 Sp1 상보성 스페이서, 암호기 바코드, 및 임의의 주기 바코드, 및 제2의 스페이서 요소(Sp2)로 구성된 DNA 암호화 태그를 사용한다. 적어도 2개의 상이한 스페이서 요소를 사용하는 유용성은 제1의 결합 주기가 잠재적으로 수개의 DNA 기록 태그 중 하나를 선택하고 단일의 DNA 기록 태그가 연장되어 연장된 DNA 기록 태그의 말단에서 새로운 Sp2 스페이서 요소를 생성한다는 것이다. 제2 및 후속적인 결합 주기에서, 결합제는 Sp1' 보다는 Sp2' 스페이서 만을 함유한다. 이러한 방식으로, 제1의 주기로부터 단지 단일의 연장된 기록 태그가 후속적인 주기에서 연장된다. 다른 구현예에서, 제2의 및 후속적인 주기는 결합제 특이적인 스페이서를 사용할 수 있다.

일부 구현예에서, 기록 태그는 5' 내지 3' 방향으로: 보편적인 전방(또는 5') 프라이밍 서열, UMI, 및 스페이서 서열을 포함한다. 일부 구현예에서, 기록 태그는 5' 내지 3' 방향으로: 보편적인 전방(또는 5') 프라이밍 서열, 임의의 UMI, 바코드 (예컨대, 샘플 바코드, 분배 바코드, 구획 바코드, 공간 바코드, 또는 이의 어떠한 조합), 및 스페이서 서열을 포함한다. 일부 다른 구현예에서, 기록 태그는 5' 내지 3' 방향으로: 보편적인 전방(또는 5') 프라이밍 서열, 바코드(예컨대, 샘플 바코드, 분배 바코드, 구획 바코드, 공간 바코드, 또는 이의 어떠한 조합), 임의의 UMI, 및 스페이서 서열을 포함한다.

조합 시도를 사용하여 변형된 DNA 및 PNA로부터 UMI를 생성할 수 있다. 하나의 예에서, UMI는 짧은 단어 서열(4-15머)의 세트와 함께 "화학적 연결"에 의해 작제될 수 있으며, 상기 서열은 서로에 대해 직각이 되도록 작제

될 수 있다(Spiropulos and Heemstra 2012). DNA 주형을 사용하여 "단어" 중합체의 화학적 연결을 지시한다. DNA 주형은 용액 속에서 소-성분을 함께 단순히 혼합함으로써 조합 주형 구조의 조립이 가능한 하이브리드화 아암(hybridizing arm)으로 작제된다(참고: 도 12c). 특정 구현예에서, 이러한 설계에 "스페이서" 서열은 존재하지 않는다. 단어 공간의 크기는 10개의 단어 내지 10,000개 이상의 단어로 변할 수 있다. 특정 구현예에서, 단어는 이들이 서로 상이하여 교차 하이브리드화하지 않고, 여전히 비교적 균일한 하이브리드화 조건을 지니도록 선택된다. 일 구현예에서, 단어의 길이는 소세트내에서 약 1000개의 단어를 사용하여 10개 염기의 순서로 이루어질 것이다(이는 총 10-머 단어 공간 ~ 4^{10} = 1 백만 단어의 단지 0.1%이다). 이러한 단어의 세트(소세트내 1000개)는 함께 콘카테네이트되어 복잡성 = 1000^{n} 파워(power)인 최종의 조합 UMI를 생성할 수 있다. 함께 콘카테네이트된 4개 단어의 경우, 이는 10^{12}개의 상이한 요소의 UMI 다양성을 생성한다. 이들 UMI 서열은 단일 분자 수준에서 거대분자(펩타이드, 단백질 등)에 첨부된다. 일 구현예에서, UMI의 다양성은 UMI가 부착된 거대분자의 분자의 수를 초과한다. 이러한 방식으로, UMI는 목적한 거대분자를 유일하게 확인한다. 조합 단어 UMI의 사용은, 단일 염기 분해(resolution)가 다수 염기의 길이의 단어를 판독하는 것을 요구하지 않기 때문에, 고 오류율 서열(예컨대, 나노공극 서열분석기, 나노갭 터널링 서열분석 등)에서 판독을 용이하게 한다. 조합 단어를 또한 사용하여 구획 태그, 분배 바코드, 공간 바코드, 샘플 바코드, 암호기 서열, 주기 특이적인 서열, 및 바코드와 같은 기록 태그 또는 암호화 태그의 다른 동일성-정보 성분을 생성할 수 있다. 오류-내성 단어(코드)를 지닌 나노공극 서열분석 및 DNA 암호화 정보는 당해 분야에 알려져 있다(참고: 예컨대, Kiah et al., 2015, *Codes for DNA sequence profiles.* IEEE International Symposium on Information Theory (ISIT); Gabrys et al., 2015, *symmetric Lee distance codes for DNA-based storage.* IEEE Symposium on Information Theory (ISIT); Laure et al., 2016, *coding in 2D: Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes*. Angew. Chem. Int. Ed. doi:10.1002/anie.201605279; Yazdi et al., 2015, IEEE Transactions on Molecular, Biological and Multi-Scale Communications 1:230-248; 및 Yazdi et al., 2015, Sci Rep 5:14138, 이들 각각은 이의 전문이 참고로 포함된다). 따라서, 특정 구현예에서 본원에 기술된 구현예 중의 어느 하나에서 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 작제물은 오류 수정 코드인 확인 성분(예컨대, UMI, 암호기 서열, 바코드, 구획 태그, 주기 특이적인 서열 등)으로 구성된다. 일부 구현예에서, 오류 수정 코드는: 해밍 코드(Hamming code), 리 거리 코드(Lee distance code), 비대칭 리 거리 코드(asymmetric Lee distance code), 리드-솔로몬 코드(Reed-Solomon code), 및 레벤슈타인-테넨골트스 코드(Levenshtein-Tenengolts code)로부터 선택된다. 나노공극 서열분석의 경우 전류 또는 이온 플럭스 프로파일(ionic flux profile) 및 비대칭 염기 콜링 오류(asymmetric base calling error)는 사용된 나노공극 및 생화학의 유형에 대해 고유하며, 이러한 정보를 사용하여 전술한 오류 수정 시도를 사용하여 보다 풍부한 DNA 코드를 설계할 수 있다. 강력한 DNA 나노공극 서열분석 바코드를 사용하기 위한 대안으로서, DNA 염기 콜링을 전적으로 피하면서 바코드 서열의 전류 또는 이온 플럭스 신호를 직접 사용할 수 있으며(이의 전문이 참고로 포함된 미국 특허 제7,060,507호), 라스즐로(Laszlo) 등에 의해 기술된 바와 같이(2014, Nat. Biotechnol. 32:829-833, 이의 전문은 참고로 포함된다) 예측된 전류/플럭스 신호로 역 맵핑함으로써 바코드 서열을 즉시 확인할 수 있다. 이러한 논문에서, 라스즐로 등은 나노공극을 통해 상이한 단어 스트링(string)이 통과하는 경우, 생물학적 나노공극, MspA에 의해 생성된 전류 신호(current signature), 및 수득되는 전류 신호를 보편적인 서열로부터의 가능한 전류 신호의 인 실리코(*in silico*) 예측에 대해 역으로 맵핑함으로써 DNA 쇄를 맵핑하고 확인하는 능력을 기술하고 있다(2014, Nat. Biotechnol. 32:829-833). 유사한 개념을 DNA 코드 및 나노갭 터널링 전류-기반 DNA 서열분석에 의해 생성된 전기 시그널에 적용할 수 있다(Ohshiro et al., 2012, Sci Rep 2: 501).

따라서, 특정 구현예에서, 암호화 태그, 기록 태그, 또는 둘 다의 성분을 확인하는 것은 유일한 전류 또는 이온 플럭스 또는 광학 신호를 생성할 수 있으며, 여기서 본원에 제공된 방법들 중 어느 것의 분석 단계는 확인 성분을 확인하기 위한 유일한 전류 또는 이온 플럭스 또는 광학 신호를 포함한다. 일부 구현예에서, 확인 성분은 암호기 서열, 바코드, UMI, 구획 태그, 주기 특이적인 서열, 또는 이의 어떠한 조합으로부터 선택된다.

특정 구현예에서, 샘플내 모든 또는 실질적인 양의 거대분자(예컨대, 단백질, 폴리펩타이드, 또는 펩타이드)(예컨대, 적어도 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%)는 기록 태그로 표지된다. 거대분자의 표지화는 거대분자를 고체 지지체에 고정화하기 전 또는 후에 일어날 수 있다.

다른 구현예에서, 샘플내 거대분자(예컨대, 단백질, 폴리펩타이드, 또는 펩타이드)는 기록 태그로 표지된다. 특수한 구현예에서, 샘플로부터의 거대분자의 소세트는 기록 태그를 사용하여 표적화된(분석물 특이적인) 표지화를 겪는다. 단백질의 표적화된 기록 태그 표지화는 기록 태그내에서 짧은 표적-특이적인 DNA 포획 프로브, 예컨

대, 분석물-특이적인 바코드에 연결된 표적 단백질-특이적인 결합제(예컨대, 항체, 아프타머 등)을 사용하여 달성할 수 있으며, 상기 결합제는 기록 태그내 상보성의 표적-특이적인 미끼 서열, 예컨대, 분석물-특이적인 바코드에 어닐링된다(참고: 도 28a). 기록 태그는 표적 단백질 상에 존재하는 동종의 반응성 모이어티에 대한 반응성 모이어티(예컨대, 클릭 화학 표지화, 광친화성 표지화)를 포함한다. 예를 들면, 기록 태그는 알킨-유도체화된 단백질과 상호작용시키기 위한 아지드 모이어티를 포함할 수 있거나, 기록 태그는 천연 단백질 등과 상호작용시키기 위한 벤조페논을 포함할 수 있다(참고: 도 28의 A 및 B). 표적 단백질 특이적인 결합제에 의한 표적 단백질의 결합시, 기록 태그 및 표적 단백질은 이들의 상응하는 반응성 모이어티를 통해 커플링된다(참고: 도 28의 B 및 C). 표적 단백질이 기록 태그로 표지된 후, 표적-단백질 특이적인 결합제는 표적-단백질 특이적인 결합제에 연결된 DNA 포획 프로브의 소화에 의해 제거될 수 있다. 예를 들면, DNA 포획 프로브는 우라실 염기를 함유하도록 설계될 수 있으며, 이는 이후에 우라실-특이적인 절개 시약(예컨대, USER™)을 사용한 소화에 대해 표적화되고, 표적-단백질 특이적인 결합제는 표적 단백질로부터 해리될 수 있다.

하나의 예에서, 표적 단백질의 세트에 대해 특이적인 항체를 상보성 미끼 서열(예컨대, 도 28에서 분석물 바코드 BC_A')을 사용하여 설계한 기록 태그로 하이브리드화된 DNA 포획 프로브(예컨대, 도 28에서 분석물 바코드 BC_A)로 표지할 수 있다. 단백질의 샘플-특이적인 표지화는 샘플-특이적인 바코드를 포함하는 기록 태그 상에서 상보성 미끼 서열로 하이브리드화하는 DNA-포획 프로브 표지된 항체를 사용하여 달성할 수 있다.

다른 예에서, 표적 단백질-특이적인 아프타머를 샘플내 단백질의 소세트의 표적화된 기록 태그 표지화에 사용한다. 표적 특이적인-아프타머는 기록 태그내 상보성 미끼와 어닐링하는 DNA 포획 프로브에 연결된다. 기록 태그는 상응하는 반응성 모이어티를 가진 표적 단백질에 커플링하기 위한 반응성의 화학적 또는 광-반응성 화학적 프로브(예컨대, 벤조페논(BP))를 포함한다. 아프타머는 이의 표적 단백질 분자에 결합하여, 기록 태그가 표적 단백질에 매우 근접하도록 하여, 표적 단백질에 대한 기록 태그의 커플링을 야기한다.

작은 분자 단백질 친화성 리간드에 부착된 광-반응성 화학적 프로브를 사용한 광친화성(PA) 단백질 표지화는 이미 기술되어 있다(Park, Koh et al. 2016). 대표적인 광-반응성 화학적 프로브는 앞서 기술된 바와 같이(Smith and Collins 2015) 조사 파장 하에서 활성화된, 벤조페논(반응성 이라디칼, 365 nm), 페닐디아지린(반응성 탄소, 365 nm), 및 페닐아지드(반응성 니트렌 유리 라디칼)을 기반으로 한 프로브를 포함한다. 바람직한 구현예에서, 단백질 샘플내 표적 단백질은 리(Li) 등이 개시한 방법을 사용하여 샘플 바코드를 포함하는 기록 태그로 표지되며, 여기서 벤조페논 표지된 기록 태그내 미끼 서열은 동종의 결합제(예컨대, 핵산 아프타머(참고: 도 28)에 부착된 DNA 포획 프로브에 하이브리드화된다(Li, Liu et al. 2013). 광친화성 표지된 단백질 표적의 경우, 광친화성 모이어티는 표적 단백질보다는 항체를 자가-표지할 수 있으므로, 표적 단백질-특이적인 결합제로서 DNA/RNA 아프타머의 사용이 항체에 걸쳐 바람직하다. 대조적으로, 광친화성 표지화는 단백질보다는 핵산에 대해 덜 효율적이어서, 아프타머를 DNA-지시된 화학적 또는 강-표지화에 대해 보다 더 우수한 비히클로 만든다. 광-친화성 표지화와 유사하게, 로젠(Rosen) 등에 의해 기술된 것(Rosen, Kodal et al. 2014, Kodal, Rosen et al. 2016)과 유사한 방식으로 아프타머 결합 부위에 근접하게 반응성 라이신(또는 다른 모이어티)의 DNA-지시된 화학적 표지화를 사용할 수 있다.

전술한 구현예에서, 하이브리드화 외에 다른 유형의 연결을 사용하여 표적 특이적인 결합제 및 기록 태그를 연결할 수 있다(참고: 도 28의 A). 예를 들면, 2개의 모이어티를 포획된 표적 단백질(또는 다른 거대분자)이 도 28의 B에 나타낸 바와 같이 기록 태그에 공유결합으로 연결되면 결합제를 절단하여 방출하도록 설계된 링커를 사용하여, 공유결합으로 연결시킬 수 있다. 적합한 링커는 3' 말단과 같은 기록 태그의 다양한 위치에, 또는 기록 태그의 5' 말단에 부착된 링커내에서 부착될 수 있다.

VII. 결합제 및 암호화 태그

본원에 기술된 방법은 거대분자에 결합할 수 있는 결합제를 사용한다. 결합제는 거대분자의 성분 또는 특징에 결합할 수 있는 어떠한 분자(예컨대, 펩타이드, 폴리펩타이드, 단백질, 핵산, 탄수화물, 소 분자 등)일 수 있다. 결합제는 천연적으로 존재하거나, 합성적으로 생산되거나, 재조합적으로 발현된 분자일 수 있다. 결합제는 거대 분자(예컨대, 펩타이드의 단일 아미노산)의 단일 단량체 또는 소단위에 결합할 수 있거나 거대분자(예컨대, 디펩타이드, 트리펩타이드, 또는 보다 긴 펩타이드 분자의 보다 높은 차수의 펩타이드)의 다중 연결된 소단위에 결합할 수 있다.

특정 구현예에서, 결합제는 공유결합으로 결합하도록 설계될 수 있다. 공유결합성 결합은 조건적이도록 설계되거나 정확한 모이어티에 결합시 선호된다. 예를 들면, NTAA 및 이의 동종의 NTAA-특이적인 결합제는 각각 반응

성 그룹으로 변형되어 일단 NTAA-특이적인 결합제가 동종의 NTAA에 결합되면, 커플링 반응이 수행되어 2개 사이에 공유결합성 연결을 생성하도록 수행된다. 동종의 반응성 그룹을 결여한 다른 위치에 대한 결합제의 비-특이적인 결합은 공유결합성 부착을 생성하지 않을 수 있다. 결합제와 이의 표적 사이의 공유결합성 결합은 비-특이적으로 결합된 결합제를 제거하기 위해 사용될 보다 엄격한(stringent) 세척을 허용하므로, 검정의 특이성을 증가시킨다.

특정 구현예에서, 결합제는 선택적인 결합제일 수 있다. 본원에 사용된 바와 같은, 선택적인 결합은 상이한 리간드(예컨대, 아미노산 또는 아미노산의 부류)에 대한 결합에 비해 특이적인 리간드(예컨대, 아미노산 또는 아미노산의 부류)에 우선적으로 결합하는 결합제의 능력을 지칭한다. 선택성은 일반적으로 결합제를 사용하여 복합체내 하나의 리간드의 다른 리간드로의 치환 반응하기 위한 평형 상수로서 지칭된다. 전형적으로, 이러한 선택성은 리간드의 공간 기하학 및/또는 수소 결합에 의해 또는 반데르 바알스력(Van der Waals force)(비-공유결합성 상호작용)에 의해 또는 결합제에 대한 가역성 또는 비-가역성 공유결합성 부착에 의해서와 같이, 리간드가 결합제에 결합하는 방식 및 정도와 관련된다. 선택성은 상대적일 수 있으며, 절대적인 것과 대치되는 것이고, 리간드 농도를 포함하는 상이한 인자들이 이에 영향을 미칠 수 있음이 이해될 수 있다. 따라서, 하나의 예에서, 결합제는 20개의 표준 아미노산 중 하나에 선택적으로 결합한다. 비-선택적인 결합의 예에서, 결합제는 20개의 표준 아미노산 중 2개 이상에 결합할 수 있다.

본원에 개시된 방법의 실시에서, 거대분자의 특징 또는 성분에 선택적으로 결합하는 결합제의 능력은 거대분자와 관련된 기록 태그에 이의 암호화 태그 정보의 이전, 암호화 태그에 기록 태그 정보의 이전, 또는 디-태그 분자로 암호화 태그 정보 및 기록 태그 정보의 이전을 허용하기에 충분해야만 할 필요가 있다. 따라서, 선택성은 거대분자가 노출되는 다른 결합제에 대해서만 상대적일 필요가 있다. 또한, 결합제의 선택성은 특이적인 아미노산에 대해 절대적일 필요가 없을 뿐 아니라, 비극성 또는 비-극성 측면 쇄를 지니거나 전기적으로(양성으로 또는 음성으로) 하전된 측쇄를 지니거나, 방향족 측쇄를 지니거나, 일부 특이적인 부류 또는 크기의 측쇄를 지닌 아미노산 등과 같은 아미노산의 부류에 대해 선택적일 수 있다.

특수한 구현예에서, 결합제는 목적한 거대분자에 대해 높은 친화성 및 높은 선택성을 갖는다. 특히, 낮은 오프-율(off-rate)의 고 결합 친화성은 암호화 태그와 기록 태그 사이의 정보 이전에 효율적이다. 특정 구현예에서, 결합제는 Kd가 < 10 nM, <5 nM, < 1 nM, < 0.5 nM, 또는 < 0.1 nM이다. 특수한 구현예에서, 결합제는 거대분자에 이의 Kd의 >10X, >100X, 또는 >1000X로 첨가되어 결합을 구동시킴으로서 완료한다. 단일 단백질 분자에 대한 항체의 결합 동력학의 상세한 논의는 Chang et al. (Chang, Rissin et al. 2012)에 기술되어 있다.

펩타이드의 소 N-말단 아미노산(NTAAs)에 대한 결합제의 친화성을 증가시키기 위하여, NTAA를 디니트로페놀(DNP)과 같은 "면역원성" 합텐(hapten)으로 변형시킬 수 있다. 이는 상거 시약(Sanger's reagent), 디니트로플루오로벤젠 (DNFB)을 사용한 사이클릭 서열분석 시도에서 시행될 수 있으며, 이는 DNP 그룹을 NTAA의 아민 그룹에 부착시킨다. 시판되는 항-DNP 항체는 낮은 nM 범위(~8 nM, LO-DNP-2)의 친화성을 가지며(Bilgicer, Thomas et al. 2009); 따라서, 이는 고-친화성 NTAA 결합제를 DNP(DNFB를 통해)로 변형된 다수의 NTAA에 대해 가공하고 특수한 NTAA에 대해 양호한 결합 선택성을 동시에 달성하는 것이 가능할 수 있는 이유이다. 다른 실시예에서, NTAA는 4-설포닐-2-니트로플루오로벤젠(SNFB)을 사용하여 설포닐 니트로페놀(SNP)로 변형시킬 수 있다. 유사한 친화성 향상은 또한 아실 그룹 또는 아미디닐(구아니디닐) 그룹과 같은 대안적 NTAA 변형제로 달성할 수 있다.

특정 구현예에서, 결합제는 NTAA, CTAA, 개재하는 아미노산, 디펩타이드 (2개의 아미노산의 서열), 트리펩타이드(3개의 아미노산의 서열), 또는 펩타이드 분자의 보다 높은 차수의 펩타이드에 결합할 수 있다. 일부 구현예에서, 결합제의 라이브러리내 각각의 결합제는 특수한 아미노산, 예를 들면 20개의 표준의 천연적으로 존재하는 아미노산 중 하나에 선택적으로 결합한다. 표준의, 천연적으로-존재하는 아미노산은 알라닌(A 또는 Ala), 시스테인(C 또는 Cys), 아스파르트산(D 또는 Asp), 글루탐산(E 또는 Glu), 페닐알라닌(F 또는 Phe), 글리신(G 또는 Gly), 히스티딘(H 또는 His), 이소루이신(I 또는 Ile), 라이신(K 또는 Lys), 루이신(L 또는 Leu), 메티오닌(M 또는 Met), 아스파라긴(N 또는 Asn), 프롤린(P 또는 Pro), 글루타민(Q 또는 Gln), 아르기닌(R 또는 Arg), 세린(S 또는 Ser), 트레오닌(T 또는 Thr), 발린(V 또는 Val), 트립토판(W 또는 Trp), 및 타이로신(Y 또는 Tyr)을 포함한다.

특정 구현예에서, 결합제는 아미노산의 해독 후 변형에 결합할 수 있다. 일부 구현예에서, 펩타이드는 하나 이상의 해독 후 변형을 포함하며, 이는 동일하거나 상이할 수 있다. 펩타이드의 NTAA, CTAA, 개재하는 아미노산, 또는 이의 조합은 해독 후 변형될 수 있다. 아미노산에 대한 해독 후 변형은 아실화, 아세틸화, 알킬화(메틸화 포함), 바이오티닐화, 부티릴화, 카바밀화, 카보닐화, 탈아미드화, 탈이민화, 디프타미드 형성, 이황화물 브릿

지 형성, 엘리미닐화, 플라빈 부착, 포르밀화, 감마-카복실화, 글루타밀화, 글리실화, 글리코실화, 글리피화, 헴 C 부착, 하이드록실화, 하이푸신 형성, 요오드화, 이소프레닐화, 지질화, 리포일화, 말로닐화, 메틸화, 미리스토일화, 산화, 팔미토일화, 페길화, 포스포판테테이닐화, 인산화, 프레닐화, 프로피오닐화, 레티닐리덴 스키프 염기 형성, S-글루타티오닐화, S-니트로실화, S-설페닐화, 셀렌화, 석시닐화, 설핀화, 우비퀴틴화, 및 C-말단 아미드화를 포함한다(참고: 또한 Seo and Lee, 2004, J. Biochem. Mol. Biol. 37:35-44).

특정 구현예에서, 렉틴은 단백질, 폴리펩타이드, 또는 펩타이드의 글리코실화 상태를 검출하기 위한 결합제로서 사용된다. 렉틴은 유리 탄수화물 또는 당단백질의 글리칸 에피토프를 선택적으로 인식할 수 있는 탄수화물-결합 단백질이다 다양한 글리코실화 상태를 인식하는 렉틴의 목록(예컨대, 코어-푸코즈, 시알산, N-아세틸-D-락토스아민, 만노즈, N-아세틸-글루코사민)은 다음을 포함한다: A,AA,AL,BA,CA,CG,CL,OL,SA, BanLec, BC2L-A, BC2LCN, BPA, BPL, 칼세파(Calsepa), CGL2, CNL, Con, ConA, DBA, 디스코이딘(Discoidin), DSA, ECA, EEL, F17AG, Gal1, Gal1-S, Gal2, Gal3, Gal3C-S, Gal7-S, Gal9, GNA, GRFT, GS-I, GS-II, GSL-I, GSL-II, HHL, HIHA, HPA, I, II, 자칼린(Jacalin), LBA, LCA, LEA, LEL, 렌틸(Lentil), 롯스(Lotus), LSL-N, LTL, MAA, MAH, MAL_I, 말렉틴(Malectin), MOA, MPA, MPL, NPA, 오리사타(Orysata), PA-IIL, PA-IL, PALa, PHA-E, PHA-L, PHA-P, PHAE, PHAL, PNA, PPL, PSA, PSL1a, PTL, PTL-I, PWM, RCA120, RS-Fuc, SAMB, SBA, SJA, SNA, SNA-I, SNA-II, SSA, STL, TJA-I, TJA-II, TxLCI, UDA, UEA-I, UEA-II, VFA, VVA, WFA, WGA (참고: Zhang et al., 2016, MABS 8:524-535).

특정 구현예에서, 결합제는 변형되거나 표지된 NTAA에 결합될 수 있다. 변형되거나 표지된 NTAA는 PITC, 1-플루오로-2,4-디니트로벤젠(상거 시약, DNFB), 단실 클로라이드 (DNS-Cl, 또는 1-디메틸아미노프탈렌-5-설포닐 클로라이드), 4-설포닐-2-니트로플루오로벤젠(SNFB), 아세틸화 시약, 구아니딘화 시약, 티오아실화 시약, 티오아세틸화 시약, 또는 티오벤질화시약로 표지된 것일 수 있다.

특정 구현예에서, 결합제는 아프타머(예컨대, 펩타이드 아프타머, DNA 아프타머, 또는 RNA 아프타머), 항체, 안티칼린, ATP-의존성 Clp 프로테아제 어댑터 단백질(ClpS), 항체 결합 단편, 항체 모사체, 펩타이드, 펩티도미메틱(peptidomimetic), 단백질, 또는 폴리뉴클레오타이드(예컨대, DNA, RNA, 펩타이드 핵산 (PNA), γPNA, 브릿지된 핵산(BNA), 크세노 핵산(XNA), 글리세롤 핵산(GNA), 또는 트레오즈 핵산(TNA), 또는 이의 변이체)일 수 있다.

본원에 사용된 바와 같은, 용어 항체 및 항체들은 광의적 의미로 사용되어 예를 들면 면역글로불린 A, 면역글로불린 G, 면역글로불린 D, 면역글로불린 E, 및 면역글로불린 M에 한정되지 않는 완전한 항체 분자뿐만 아니라, 적어도 하나의 에피토프에 면역-특이적으로 결합하는 항체 분자의 어떠한 면역반응성 성분(들)을 포함한다. 항체는 천연적으로 존재하거나, 합성적으로 생산되거나, 재조합적으로 발현된다. 항체는 융합 단백질일 수 있다. 항체는 항체 모사체일 수 있다. 항체의 예는 Fab 단편, Fab' 단편, $F(ab')_2$ 단편, 단일 쇄 항체 단편(scFv), 미니항체, 디아보디(diabody), 가교결합된 항체 단편, Affibody™, 나노보디(nanobody), 단일 도메인 항체, DVD-Ig 분자, 알파보디(alphabody), 애피머(affimer), 아피틴(affitin), 사이클로타이드, 분자 등을 포함하나, 이에 한정되지 않는다. 항체 가공 또는 단백질 가공 기술을 사용하여 유도된 면역반응성 생성물은 또한 용어 항체의 의미내에 표현하여 존재한다. 관련 프로토콜을 포함하는, 항체 및/또는 단백질 가공의 상세한 설명은 다른 곳 중에서도, J. Maynard and G. Georgiou, 2000,n. Rev. Biomed. Eng. 2:339-76; Antibody Engineering, R. Kontermann and S. Dubel, eds., Springer Lab Manual, Springer Verlag (2001); U.S. Patent No. 5,831,012; and S. Paul, Antibody Engineering Protocols, Humana Press (1995)에서 찾을 수 있다.

항체를 사용하여, 펩타이드를 특이적이로 인식하는 핵산 및 펩타이드 아프타머를 공지된 방법을 사용하여 생산할 수 있다. 더 낮은 결합 친화성을 갖는 아프타머가 경우에 따라 선택될 수 있지만, 아프타머는 전형적으로 매우 높은 친화성을 사용하여 매우 특이적인, 구조-의존적 방식으로 표적 분자에 결합한다. 아프타머는 메틸 또는 하이드록실 그룹의 존재 또는 부재와 같은 매우 작은 구조적 차이를 기반으로 표적 사이를 구별하는 것으로 밝혀졌으며 특정의 아프타머는 D-와 L-거울상이성체 사이를 구별한다. 약물, 금속 이온, 및 유기 염료, 스트렙타비딘, VEGF, 및 바이러스 단백질을 포함하나 이에 한정되지 않는 펩타이드, 바이오틴, 및 단백질을 포함하는 소분자 표적에 결합하는 아프타머가 수득되어 왔다. 아프타머는 유리 표면 및 미세구에 부착되는 경우 바이오티닐화, 플루오레세인 표지 후 기능성 활성을 보유하는 것으로 밝혀졌다(참고: Jayasena, 1999, Clin Chem 45:1628-50; Kusser2000, J. Biotechnol. 74: 27-39; Colas, 2000, Curr Opin Chem Biol 4:54-9). 아르기닌 및 AMP에 특이적으로 결합하는 아프타머가 또한 기술되어 왔다(참고: Patel and Suri, 2000, J. Biotech. 74:39-60). 특이적인 아미노산에 결합하는 올리고뉴클레오타이드 아프타머는 골드(Gold) 등(1995, Ann. Rev.

Biochem. 64:763-97)에 기술되어 왔다. 아미노산에 결합하는 RNA 아프타머 또한 기술되어 왔다(Ames and Breaker, 2011, RNA Biol. 8; 82-89; Mannironi et al., 2000, RNA 6:520-27; Famulok, 1994, J. Am. Chem. Soc. 116:1698-1706).

결합제는 유전적 가공에 의해 아미노산 서열내에 하나 이상의 돌연변이를 도입하여 거대분자(예컨대, NTAA, CTAA, 또는 해독 후 변형된 아미노산 또는 펩타이드)의 특이적인 성분 또는 특징에 결합하는 가공된 단백질을 생산함으로써 천연적으로 존재하거나 합성적으로-생산한 단백질을 변형시켜 제조할 수 있다. 예를 들면, 엑소펩티다제(예컨대, 아미노펩티다제, 카복시펩티다제), 엑소프로테아제, 돌연변이된 엑소프로테아제, 돌연변이된 안티칼린, 돌연변이된 ClpSs, 항체, 또는 tRNA 신테타제를 변형시켜 특수한 NTAA에 선택적으로 결합하는 결합제를 생성할 수 있다. 다른 예에서, 카복시펩티다제는 변형되어 특수한 CTAA에 선택적으로 결합하는 결합제를 생성할 수 있다. 결합제는 또한 변형된 NTAA 또는 변형된 CTAA, 예를 들면, 해독 후 변형(예컨대, 인산화된(phosphorylated) NTAA 또는 인산화된 CTAA)를 갖는 것 또는 표지(예컨대, PTC, 1-플루오로-2,4-디니트로벤젠(상거 시약, DNFB 사용), 단실 클로라이드(DNS-Cl, 또는 1-디메틸아미노나프탈렌-5-설포닐 클로라이드 사용)로, 또는 티오아실화 시약, 티오아세틸화 시약, 아세틸화 시약, 아미드화(구아니딘화) 시약, 또는 티오벤질화시약)을 사용하여 변형된 것에 특이적으로 결합하도록 설계되고, 변형되어, 이용될 수 있다. 단백질의 직접적인 발달을 위한 전략은 당해 분야에 공지되어 있으며(예컨대, Yuan et al., 2005, Microbiol. Mol. Biol. Rev. 69:373-392에 의해 고찰됨), 상 디스플레이, 리보소옴 디스플레이, mRNA 디스플레이, CIS 디스플레이, CAD 디스플레이, 유화액, 세포 표면 디스플레이 방법, 효소 표면 디스플레이, 세균 표면 디스플레이 등을 포함한다.

일부 구현예에서, 변형된 NTAA에 특이적으로 결합하는 결합제를 사용할 수 있다. 예를 들면, NTAA를 페닐이소티오시아네이트(PITC)와 반응시켜 페닐티오카바모일-NTAA 유도체를 형성할 수 있다. 이러한 방식으로, 결합제를 페닐티오카바모일 모이어티의 페닐 그룹 및 NTAA의 알카-탄소 R 그룹 둘 다에 선택적으로 결합하도록 양식화할 수 있다. 이러한 방식으로 PITC의 사용은 하기 논의된 바와 같은 에드만 분해에 의한 NTAA의 후속적인 절단을 허용한다. 다른 구현예에서, NTAA는 상거 시약(DNFB)과 반응하여 DNP-표지된 NTAA를 생성한다(참고: 도 3). 임의로, DNFB는 1-에틸-3-메틸이미다졸륨 비스[(트리플루오로메틸)설포닐]이미드([emim][Tf2N])와 같은 이온성 액체와 함께 사용되며, 여기서 DNFB는 매우 가용성이다. 이러한 방식으로, 결합제를 가공하여 NTAA 상의 DNP 및 R 그룹의 조합에 선택적으로 결합시킬 수 있다. DNP 모이어티의 첨가는 결합제와 NTAA의 상호작용을 위한 보다 큰 "핸들(handle)"을 제공하며, 보다 큰 친화성 상호작용을 유도해야 한다. 여전히 다른 구현예에서, 결합제는 펩타이드의 아미노펩티다제 분해의 사이클릭 제어를 제공하는 DNP-표지된 NTAA를 인식하도록 가공된 아미노펩티다제일 수 있다. 일단 DNP-표지된 NTAA가 절단되면, DNFB 유도체화의 다른 주기를 수행하여 새로이 노출된 NTAA를 결합시키고 절단한다. 바람직한 특수한 구현예에서, 아미노펩티다제는 아연에 의해 활성화된 아미노펩티다제와 같은 단량체성 금속성(metallo)-프로테아제이다(Calcagno and Klein 2016). 다른 예에서, 결합제는 설포닐 니트로페놀(SNP)로 예컨대, 4-설포닐-2-니트로플루오로벤젠(SNFB)을 사용하여 변형시킨 NTAA에 선택적으로 결합할 수 있다. 여전히 다른 구현예에서, 결합제는 아세틸화되거나 아미드화된 NTAA에 선택적으로 결합할 수 있다.

NTAA를 변형시키는데 사용될 수 있는 시약은 트리플루오로에틸 이소티오시아네이트, 알릴 이소티오시아네이트, 및 디메틸아미노아조벤젠 이소티오시아네이트를 포함한다.

결합제는 변형된 NTAA에 대한 고 친화성, 변형된 NTAA에 대한 고 특이성, 또는 둘 다를 위해 가공될 수 있다. 일부 구현예에서, 결합제는 파아지 디스플레이를 사용하여 촉망되는 친화성 스캐폴드(scaffold)를 통해 개발할 수 있다.

표지된(바이오티닐화된) NTAA의 개개 또는 소 그룹에 결합하여 절단하는 가공된 아미노펩티다제 돌연변이가 기술되어 있다(참고: PCT 공보 제WO2010/065322호, 이의 전문은 참고로 포함된다). 아미노펩티다제는 단백질 또는 펩타이드의 N-말단으로부터 아미노산을 절단하는 효소이다. 천연의 아미노펩티다제는 매우 제한된 특이성을 가지며, 전진된 방식으로 N-말단 아미노산을 유전적으로 절단하여, 다른 것 이후 하나의 아미노산을 절단한다(Kishor et al., 2015, Anal. Biochem. 488:6-8). 그러나, 잔기 특이적인 아미노펩티다제가 확인되었다(Eriquez et al., J. Clin. Microbiol. 1980, 12:667-71; Wilce et al., 1998, Proc. Natl. Acad. Sci. USA 95:3472-3477; Liao et al., 2004, Prot. Sci. 13:1802-10). 아미노펩티다제는 특이적인 모이어티(예컨대, PTC, DNP, SNP 등)로 표지된 표준 아미노산을 나타내는 20개의 상이한 NTAA에 특이적으로 결합하도록 가공될 수 있다. 펩타이드의 N-말단의 단계적 분해의 제어는 표지의 존재하에서만 활성인(예컨대, 결합 활성 또는 촉매 활성) 가공된 아미노펩티다제를 사용함에 의해 달성된다. 다른 예에서, 하브라녹(Havranak) 등(미국 특허 공보 제2014/0273004호)은 특이적인 NTAA 결합제로서 아미노아실 tRNA 신테타제(aaRS)를 가공하는 것을 기술하였다. aaRS의 아미노산 결합 포켓(pocket)은 동종의 아미노산을 결합시키는 고유의 능력을 가지지만, 일반적으로 불량

한 결합 친화성 및 특이성을 나타낸다. 더욱이, 이들 천연의 아미노산 결합제는 N-말단 표지를 인식하지 않는다. aaRS 스캐폴드의 지시된 발달을 사용하여 N-말단 표지와 관련하여 N-말단 아미노산을 인식한 보다 높은 친화성, 보다 높은 특이성 결합제를 생성할 수 있다.

다른 예에서, 고-선택성의 ClpS는 또한 문헌에 기술되어 왔다. 에밀리(Emili) 등은 파아지 디스플레이를 통해 이. 콜라이(*E. coli*) ClpS 단백질의 지시된 발달로 아스파르트산, 아르기닌, 트립토판, 및 루이신 잔기의 경우 NTAA에 특이적으로 결합하는 능력을 지닌 4개의 상이한 변이체가 생성됨을 기술하였다(이의 전문이 참고로 포함된 미국 특허 제9,566,335호).

특수한 구현예에서, 안티칼린은 표지된 NTAA에 대해 고 친화성 및 고 특이성 둘다를 위해 가공된다(예컨대 DNP, SNP, 아세틸화된 등). 안티칼린 스캐폴드의 특정의 다양성은 이들의 베타 바렐 구조에 의해, 단일 아미노산에 결합하기에 적합한 형상을 갖는다. N-말단 아미노산(변형이 있거나 없는)은 잠재적으로 맞춰져서 이러한 "베타 배럴(beta barrel)" 버켓 속에서 인식된다. 가공된 신규 결합 활성을 지닌 고 친화성 안티칼린이 기술되어 왔다(참고: Skerra, 2008, FEBS J. 275: 2677-2683). 예를 들면, 플루오레세인 및 디곡시게닌에 대해 고 친화성 결합(저 nM)을 지닌 안티칼린이 가공되었다(Gebauer and Skerra 2012). 새로운 결합 기능을 위한 대안의 스캐폴드의 가공 또한 반타(Banta) 등(2013, Annu. Rev. Biomed. Eng. 15:93-113)에 의해 검토되어 왔다.

제공된 일가 결합제의 기능성 친화성(항원항체결합력)은 일가 결합제의 이가 또는 보다 높은 차수의 다량체를 사용함에 의해 적어도 일 차수의 규모까지 증가될 수 있다(Vauquelin and Charlton 2013). 항원항체결합력은 다수의, 동시, 비-공유결합성 결합 상호작용의 축적된 강도를 지칭한다. 개개의 결합 상호작용은 용이하게 해리될 수 있다. 그러나, 다수의 결합 상호작용이 동시에 존재하면, 단일 결합 상호작용의 일시적인 해리는 결합 단백질이 확산 제거되도록 허용하지 않으며 결합 상호작용은 회복되는 경향이 있다. 결합제의 항원항체결합력을 증가시키는 대안적 방법은 결합제에 부착된 암호화 태그 및 거대 분자와 관련된 기록 태그내 상보성 서열을 포함하는 것이다.

일부 구현예에서, 변형된 C-말단 아미노산(CTAA)에 선택적으로 결합하는 결합제를 이용할 수 있다. 카복시펩티다제는 유리 카복실 그룹을 함유하는 말단 아미노산을 절단하는 프로테아제이다. 다수의 카복시펩티다제는 아미노산 선호도를 나타내는데, 예컨대, 카복시펩티다제 B는 아르기닌 및 라이신과 같은 염기성 아미노산에서 우선적으로 절단한다. 카복시펩티다제는 변형되어 특수한 아미노산에 선택적으로 결합하는 결합제를 생성할 수 있다. 일부 구현예에서, 카복시펩티다제는 가공되어 변형 모이어티 및 CTAA의 알파-탄소 R 그룹 둘 다에 선택적으로 결합할 수 있다. 따라서, 가공된 카복시펩티다제는 C-말단 표지의 맥락에서 표준 아미노산을 나타내는 20개의 상이한 CTAA를 특이적으로 인식할 수 있다. 펩타이드의 C-말단으로부터 단계식 변형의 제어는 표지의 존재하에서만 활성(예컨대, 결합 활성 또는 촉매 활성)인 가공된 카복시펩티다제를 사용함으로써 달성된다. 하나의 예에서, CTAA는 파라-니트로아닐리드 또는 7-아미노-4-메틸코우마리닐 그룹에 의해 변형될 수 있다.

본원에 기술된 방법에서 사용하기 위해 결합제를 생성하도록 가공될 수 있는 다른 잠재적인 스캐폴드는 다음을 포함한다: 안티칼린, 아미노산 tRNA 신테타제(aaRS), ClpS, Affilin®, Adnectin™, T 세포 수용체, 아연 핑거 단백질(zinc finger protein), 티오레독신, GST A1-1, DARPin, 아피머, 아피틴, 알파보디, 아비머, 쿠니츠 도메인 펩타이드(Kunitz domain peptide), 모노바디(monobody), 단일 도메인 항체, EETI-II, HPSTI, 인트라바디(intrabody), 리포칼린, PHD-핑거, V(NAR) LDTI, 에비바디(evibody), Ig(NAR), 노틴(knottin), 막시바디(maxibody), 네오카르지노스타틴, pVIII, 텐다미스타트, VLR, 단백질 A 스캐폴드, MTI-II, 에코틴, GCN4, Im9, 쿠니츠 도메인, 마이크로바디, PBP, 트랜스-바디, 테트라넥틴, WW 도메인, CBM4-2, DX-88, GFP, iMab, Ldl 수용체 도메인 A, Min-23, PDZ-도메인, 조류 판크레아틱 폴리펩타이드, 카비브로톡신(charybdotoxin)/10Fn3, 도메인 항체(Dab), a2p8 안키린 반복체, 곤충 방어 A 펩타이드, 설계된 AR 단백질, C-형 렉틴 도메인, 스타필로코쿠스 뉴클레아제, Src 상동성 도메인 3(SH3), 또는 Src 상동성 도메인 2(SH2).

결합제를 가공하여 고온 및 온화한-변성 조건(예컨대, 우레아, 구아니디늄 티오시아네이트, 이온성 용액 등)을 견디도록 할 수 있다. 변성제의 사용은 α-나선 구조, β-헤어핀(hairpin), β-가닥, 및 결합제가 선형 펩타이드 에피토프에 결합하는 것을 방해할 수 있는 다른 이러한 구조와 같은 표면 결합된 펩타이드내에서 2차 구조를 감소시킬 수 있다. 일 구현예에서, 1-에틸-3-메틸이미다졸륨 아세테이트([EMIM]+[ACE])와 같은 이온성 액체를 사용하여 결합 주기 동안에 펩타이드 2차 구조를 감소시킨다(Lesch, Heuer et al. 2015).

기술된 어떠한 결합제도 또한 결합제에 관한 확인 정보를 함유하는 암호화 태그를 포함한다. 암호화 태그는 이의 관련된 결합제에 대해 유일한 확인 정보를 제공하는 약 3개의 염기 내지 약 100개의 염기의 핵산 분자이다. 암호화 태그는 약 3 내지 약 90개 염기, 약 3 내지 약 80개 염기, 약 3 내지 약 70개 염기, 약 3 내지 약 60개

염기, 약 3개 염기 내지 약 50개 염기, 약 3개 염기 내지 약 40개 염기, 약 3개 염기 내지 약 30개 염기, 약 3개 염기 내지 약 20개 염기, 약 3개 염기 내지 약 10개 염기, 또는 약 3개 염기 내지 약 8개 염기를 포함할 수 있다. 일부 구현예에서, 암호화 태그는 길이가 약 3개 염기, 4개 염기, 5개 염기, 6개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개 염기, 15개 염기, 16개 염기, 17개 염기, 18개 염기, 19개 염기, 20개 염기, 25개 염기, 30개 염기, 35개 염기, 40개 염기, 55개 염기, 60개 염기, 65개 염기, 70개 염기, 75개 염기, 80개 염기, 85개 염기, 90개 염기, 95개 염기, 또는 100개 염기이다. 암호화 태그는 DNA, RNA, 폴리뉴클레오타이드 유사체, 또는 이의 조합으로 구성될 수 있다. 폴리뉴클레오타이드 유사체는 PNA, γPNA, BNA, GNA, TNA, LNA, 모르폴리노 폴리뉴클레오타이드, 2'-O-Methyl 폴리뉴클레오타이드, 알킬 리보실 치환된 폴리뉴클레오타이드, 포스포로티오에이트 폴리뉴클레오타이드, 및 7-데아자 푸린 유사체를 포함한다.

암호화 태그는 관련된 결합제에 관한 확인 정보를 제공하는 암호기 서열을 포함한다. 암호기 서열은 길이가 약 3개 염기 내지 약 30개 염기, 약 3개 염기 내지 약 20개 염기, 약 3개 염기 내지 약 10개 염기, 또는 약 3개 염기 내지 약 8개 염기이다. 일부 구현예에서, 암호기 서열은 길이가 약 3개 염기, 4개 염기, 5개 염기, 6개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개 염기, 15개 염기, 20개 염기, 25개 염기, 또는 30개 염기이다. 암호기 서열의 길이는 생성될 수 있는 유일한 암호기 서열을 측정한다. 보다 짧은 암호화 서열은 보다 작은 수의 유일한 암호화 서열을 생성하며, 이는 적은 수의 결합제를 사용하는 경우 유용할 수 있다. 보다 긴 암호기 서열이 거대분자의 집단을 분석하는 경우 바람직할 수 있다. 예를 들면, 5개 염기의 암호기 서열은 5'-NNNNN-3'(서열 번호: 135)을 가질 수 있으며, 여기서 N은 어떠한 천연적으로 존재하는 뉴클레오타이드, 또는 유사체일 수 있다. 4개의 천연적으로 존재하는 뉴클레오타이드 A, T, C, 및 G를 사용하여, 길이가 5개 염기인 유일한 암호기 서열의 총 수는 1,024개이다. 일부 구현예에서, 유일한 암호기 서열의 총 수는 예를 들면, 모든 염기가 동일하거나, 적어도 3개의 연속된 염기가 동일하거나, 또는 둘 다인 암호기 서열을 배제시킴으로써 감소될 수 있다. 특이적인 구현예에서, ≥ 50 유일한 암호기 서열의 세트가 결합제 라이브러리에 사용된다.

일부 구현예에서, 암호화 태그 또는 기록 태그의 확인 성분, 예컨대, 암호기 서열, 바코드, UMI, 구획 태그, 분배 바코드, 샘플 바코드, 공간 영역 바코드, 주기 특이적인 서열 또는 이의 어떠한 조합은 해밍 거리(Hamming distance), 리 거리(Lee distance), 비대칭 리 거리(asymmetric Lee distance), 리드-솔로몬(Reed-Solomon), 레벤슈타인-테넨골트스(Levenshtein-Tenengolts), 또는 오류-수정을 위한 유사한 방법에 적용된다. 해밍 거리는 동일한 길이의 2개의 스트링 사이에 상이한 다수의 위치를 지칭한다. 이는 하나의 스트링을 다른 것으로 변화시키는데 요구되는 치환의 최소 수를 측정한다. 해밍 거리를 사용하여 거리가 충분히 떨어져 있는 암호기 서열을 선택함으로서 오류를 수정할 수 있다. 따라서, 암호기 서열이 5개 염기인 예에서, 이용가능한 암호기 서열의 수는 256개의 유일한 암호기 서열로 감소한다($1 \rightarrow 4^4$ 암호기 서열 = 256개 암호기 서열의 해밍 거리). 다른 구현예에서, 암호기 서열, 바코드, UMI, 구획 태그, 주기 특이적인 서열, 또는 이의 어떠한 조합을 설계하여 주기 해독 과정(cyclic decoding process)에 의해 판독을 용이하게 한다(Gunderson, 2004, Genome Res. 14:870-7). 다른 구현예에서, 단일 염기 분해(resolution)를 필요로 하기 보다는, 다수의 염기(길이가 ~5 내지 20개 염기)의 단어가 판독될 필요가 있으므로, 암호기 서열, 바코드, UMI, 구획 태그, 분배 바코드, 공간 바코드, 샘플 바코드, 주기 특이적인 서열, 또는 이의 어떠한 조합을 설계하여 낮은 정확도의 나노공극 서열분석에 의해 판독한다. 본 개시내용의 방법에서 사용될 수 있는 15-머, 오류-수정 해밍 바코드의 소세트는 서열 번호: 1 내지 65에 제시되어 있으며 이들의 상응하는 역 상보성 서열은 서열 번호: 66 내지 130에 제시되어 있다.

일부 구현예에서, 결합제의 라이브러리내 각각의 유일한 결합제는 유일한 암호기 서열을 갖는다. 예를 들면, 20개의 유일한 암호기 서열을 20개의 표준 아미노산에 결합하는 20개의 결합제의 라이브러리에 사용할 수 있다. 추가의 암호화 태그 서열을 사용하여 변형된 아미노산(예컨대, 해독 후 변형된 아미노산)을 확인할 수 있다. 다른 예에서, 30개의 유일한 암호기 서열을 20개의 표준 아미노산 및 10개의 독 후 변형된 아미노산(예컨대, 인산화된 아미노산, 아세틸화된 아미노산, 메틸화된 아미노산)에 결합하는 30개의 결합제의 라이브러리에 사용할 수 있다. 다른 구현예에서, 2개 이상의 상이한 결합제는 동일한 암호기 서열을 공유할 수 있다. 예를 들면, 상이한 표준 아미노산에 각각 결합하는 2개의 결합제는 동일한 암호기 서열을 공유할 수 있다.

특정 구현예에서, 암호화 태그는 하나의 말단 또는 양쪽 말단에 스페이서 서열을 추가로 포함한다. 스페이서 서열은 약 1개 염기 내지 약 20개 염기, 약 1개 염기 내지 약 10개 염기, 약 5개 염기 내지 약 9개 염기, 또는 약 4개 염기 내지 약 8개 염기이다. 일부 구현예에서, 스페이서는 길이가 약 1개 염기, 2개 염기, 3개 염기, 4개 염기, 5개 염기, 6개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개

염기, 15개 염기 또는 20개 염기이다. 일부 구현예에서, 암호화 태그내 스페이서는 암호기 서열보다 짧은데, 예컨대, 암호기 서열보다 적어도 1개 염기, 2,개 염기, 3개 염기, 4개 염기, 5개 염기, 6,개 염기, 7개 염기, 8개 염기, 9개 염기, 10개 염기, 11개 염기, 12개 염기, 13개 염기, 14개 염기, 15개 염기, 20개 염기, 또는 25개 염기가 더 짧다. 다른 구현예에서, 암호화 태그내 스페이서는 암호기 서열과 동일한 길이이다. 특정 구현예에서, 스페이서는 결합제 특이적이어서 앞서의 결합 주기로부터의 스페이서 만이 현재의 결합 주기내 적절한 결합제로부터의 스페이서와 상호작용한다. 예는 항체 둘 다가 거대분자에 순차적으로 결합하는 경우 정보 이전만을 허용하는 스페이서 서열을 함유하는 동종의 항체의 쌍일 수 있다. 스페이서 서열은 프라이머 연장 반응용 프라이머 어닐링 부위, 또는 연결 반응에서 분할 또는 점성 말단으로서 사용될 수 있다. 암호화 태그 상의 5' 스페이서(참고: 도 5의 A, "*Sp'")는 기록 태그 상의 3' 스페이서에 대해 슈도 상보성 염기를 임의로 함유하여 T_m을 증가시킬 수 있다(Lehoud et al., 2008, Nucleic Acids Res. 36:3409-3419).

일부 구현예에서, 결합제의 수집물 내의 암호화 태그는 검정에 사용된 일반적인 스페이서 서열을 공유한다(예컨대, 다수의 결합 주기에서 사용된 전체 라이브러리는 이들의 암호화 태그내에서 일반적인 스페이서를 소유한다). 다른 구현예에서, 암호화 태그는 특수한 결합 주기를 확인하는 결합 주기 태그로 구성된다. 다른 구현예에서, 결합제의 라이브러리 내 암호화 태그는 결합 주기 특이적인 스페이서 서열을 갖는다. 일부 구현예에서, 암호화 태그는 하나의 결합 주기 특이적인 스페이서 서열을 포함한다. 예를 들면, 제1의 결합 주기에 사용된 결합제에 대한 암호화 태그는 "주기 1" 특이적인 스페이서 서열을 포함하고, 제2의 결합 주기에 사용된 결합제에 대한 암호화 태그는 "주기 2" 특이적인 스페이서 서열을 포함하며, 이렇게 "n"회 결합 주기까지 지속된다. 추가의 구현예에서, 제1의 결합 주기에 사용된 결합제에 대한 암호화 태그는 "주기 1" 특이적인 스페이서 서열 및 "주기 2" 특이적인 스페이서 서열을 포함하고, 제2의 결합 주기에 사용된 결합제에 대한 암호화 태그는 "주기 2" 특이적인 스페이서 서열 및 "주기 3" 특이적인 스페이서 서열을 포함하며, 이렇게 "n"회 결합 주기까지 지속된다. 이러한 구현예는 결합 주기가 완료된 후 비-콘카테네이트된 연장된 기록 태그의 후속적인 PCR 조립체에 유용하다(참고: 도 10). 일부 구현예에서, 스페이서 서열은 기록 태그 또는 연장된 기록 태그내에서 상보성 스페이서 서열에 어닐링하여 프라이머 연장 반응 또는 점성 말단 연결 반응을 개시하는 충분한 수의 염기를 포함한다.

주기 특이적인 스페이서 서열을 또한 사용하여 기록 태그의 집단이 거대분자와 관련된 경우 암호화 태그의 정보를 단일 기록 태그 상에 콘카테네이트시하는데 사용할 수 있다. 제1의 결합 주기는 암호화 태그로부터의 정보를 무작위-선택된 기록 태그로 이전하며, 후속적인 결합 주기는 주기 의존적 스페이서 서열을 사용하여 연장된 기록 태그 만을 프라이밍할 수 있다. 보다 구체적으로, 제1의 결합 주기에 사용된 결합제에 대한 보다 특이적인 암호화 태그는 "주기 1" 특이적인 스페이서 서열 및 "주기 2" 특이적인 스페이서 서열을 포함하며, 제2의 결합 주기에 사용된 결합제에 대한 암호화 태그는 "주기 2" 특이적인 스페이서 서열 및 "주기 3" 특이적인 스페이서 서열을 포함하고, 이렇게 "n"회 결합 주기까지 지속된다. 제1의 결합 주기로부터의 결합제의 암호화 태그는 상보성 주기 1 특이적인 스페이서 서열을 통해 기록 태그에 어닐링할 수 있다. 암호화 태그 정보가 기록 태그로 이전되면, 주기 2 특이적인 스페이서 서열은 결합 주기 1의 말기에 연장된 기록 태그의 3' 말단에 위치한다. 제2의 결합 주기로부터의 결합제의 암호화 태그는 상보성 주기 2 특이적인 스페이서 서열을 통해 연장된 기록 태그에 어닐링할 수 있다. 암호화 태그 정보가 연장된 기록 태그로 이전되면, 주기 3 특이적인 스페이서 서열은 결합 주기 2의 말기에 연장된 기록 태그의 3' 말단에 위치하며, 이렇게 "n"회 결합 주기까지 지속된다. 이러한 구현예는 다수의 결합 주기 중에서 특수한 결합 주기내 결합 정보의 이전이 앞서의 결합 주기를 경험한(연장된) 기록 태그에서만 일어날 것이라는 것을 제공한다. 그러나, 때때로, 결합제는 동종의 거대분자에 결합하지 못한다. "뒤쫓는(chase)" 단계로서 각각의 결합 주기 후 결합 주기 특이적인 스페이서를 포함하는 올리고뉴클레오타이드를 사용하여 결합 주기의 사건이 실패하는 경우에도 동시 발생한 결합 주기를 유지시킬 수 있다. 예를 들면, 동종의 결합제가 결합 주기 1 동안에 거대 분자에 결합하는데 실패한 경우, 주기 1 특이적인 스페이서, 주기 2 특이적인 스페이서, 및 "널(null)" 암호기 서열 둘 다를 포함하는 올리고뉴클레오타이드를 사용하여 결합 주기 1 후 뒤쫓는 단계를 가한다. "널" 암호기 서열은 암호기 서열 또는, 바람직하게는 "널" 결합 주기를 양성적으로 확인하는 특이적인 바코드의 부재일 수 있다. "널" 올리고뉴클레오타이드는 주기 1 특이적인 스페이서을 통해 기록 태그에 어닐링할 수 있으며, 주기 2 특이적인 스페이서는 기록 태그로 이전된다. 따라서, 결합 주기 2로부터의 결합제는 실패된 결합 주기 1 사건에도 불구하고 주기 2 특이적인 스페이서를 통해 연장된 기록 태그에 어닐링할 수 있다. "널" 올리고뉴클레오타이드는 연장된 기록 태그내에서 실패된 결합 사건으로서 결합 주기 1을 표시한다.

바람직한 구현예에서, 결합 주기-특이적인 암호기 서열은 암호화 태그에 사용된다. 결합 주기-특이적인 암호기

서열은 완전하게 유일한 분석물(예컨대, NTAA)-결합 주기 암호기 바코드의 사용을 통해 또는 주기-특이적인 바코드에 결합된 분석물(예컨대, NTAA) 암호기 서열의 조합 사용을 통해 달성될 수 있다(참고: 도 35). 조합 시도를 사용하는 장점은 보다 적은 수의 총 바코드를 설계할 필요가 있다는 것이다. 20개의 분석물 결합제는 10회 주기에 거쳐 사용되며, 단지 20개의 분석물 암호기 서열 바코드 및 10개의 결합 주기 특이적인 바코드를 설계할 필요가 있다. 대조적으로, 결합 주기가 결합제 암호기 서열에 포매된(embedded) 경우, 총 200개의 독립된 암호기 바코드를 설계할 필요가 있다. 암호기 서열 내에 직접 포매된 결합 주기 정보의 장점은 나노공극 판독물에서 오류-수정 바코드를 사용하는 경우 암호화 태그의 전체 길이가 최소화될 수 있다는 것이다. 오류-내성 바코드의 사용은 서열분석 플랫폼 및 보다 오류-경향성이 있는 시도를 사용하여 매우 정밀한 바코드 확인을 허용하지만, 신속한 속도의 분석, 보다 낮은 비용, 및/또는 보다 휴대가능한 계측과 같은 다른 장점도 갖는다. 하나의 이러한 예는 나노공극-계 서열분석 판독물이다.

일부 구현예에서, 암호화 태그는 결합제에 근접한 제2의(3') 스페이서 서열내에 절단가능한 또는 닉생성가능한(nickable) DNA 쇄를 포함한다(참고, 도 32). 예를 들면, 3' 스페이서는 우라실-특이적인 절개 시약(USER)으로 닉생성(nicking)할 수 있는 하나 이상의 우라실 염기를 가질 수 있다. USER은 우라실의 위치에 단일의 뉴클레오타이드 갭을 형성한다. 다른 예에서, 3' 스페이서는 듀플렉스의 하나의 쇄만을 가수분해하는 닉생성 엔도뉴클레아제에 대한 인식 서열을 포함할 수 있다. 바람직하게는, 3' 스페이서 서열을 절단하거나 닉생성하는데 사용된 효소는 하나의 DNA 쇄(암호화 태그의 3' 스페이서)에서만 작용하여 (연장된) 기록 태그에 속하는 듀플렉스내 다른 쇄가 완벽하게 남아있도록 한다. 이는 프라이머 연장이 일어나서 후속적인 결합 주기에 이용가능한 연장된 기록 태그 상에 단일 가닥 DNA 스페이서 서열을 남긴 후 (연장된) 기록 태그로부터 결합제의 비-변성 제거를 허용하므로, 이들 구현예는 이들의 천연 구조에서 단백질을 분석하는 검정에 특히 유용하다.

암호화 태그는 또한 팔린드로믹(PALINDROMIC) 서열을 함유하도록 설계될 수 있다. 팔린드로믹 서열의 암호화 태그내로의 도입은 암호화 태그 정보가 이전되면서 초기의, 성장하는 연장된 기록 태그가 자체적으로 폴딩되도록 한다. 연장된 기록 태그는 보다 복잡한 구조로 폴딩되어, 바람직하지 않은 분자간 결합 및 프라이머 연장 사건을 효과적으로 감소시킨다.

일부 구현예에서, 암호화 태그는 동일한 분석물을 인식하는 결합제로 이미 연장된 기록 태그 상에서만 프라이밍 연장할 수 있는 분석물-특이적인 스페이서를 포함한다. 연장된 기록 태그는 분석물-특이적인 스페이서 및 암호기 서열을 포함하는 암호화 태그를 사용하는 일련의 결합 사건으로부터 제조될 수 있다. 일 구현예에서, 제1의 결합 사건은 다음 결합 주기에서 사용하기 위한 일반적인 3' 스페이서 프라이머 서열 및 5' 말단에서의 분석물-특이적인 스페이서 서열로 구성된 암호화 태그를 지닌 결합제를 사용하며; 후속적인 결합 주기는 이후 암호화된 분석물-특이적인 3' 스페이서 서열을 지닌 결합제를 사용한다. 이러한 설계는 정확한 일련의 동종의 결합 사건으로부터 유일하게 생성되는 증폭가능한 라이브러리 요소를 생성한다. 오프-표적 및 교차-반응성 결합 상호작용은 증폭가능하지 않는 연장된 기록 태그를 가져올 것이다. 하나의 예에서, 특수한 거대분자 분석물에 대한 동종의 결합제의 쌍을 2개의 결합 주기에서 사용하여 분석물을 확인한다. 제1의 동종의 결합제는 기록 태그의 특수한 스페이서 서열 상에 프라이밍 연장을 위한 일반적인 스페이서 3' 서열, 및 다음 결합 주기에서 사용할, 5' 말단에서의 암호화된 분석물-특이적인 스페이서로 구성된 암호화 태그를 함유한다. 일치된 동종의 결합제 쌍의 경우, 제2의 결합제의 3' 분석물-특이적인 스페이서는 제1의 결합제의 5' 분석물-특이적인 스페이서와 일치한다. 이러한 방식으로, 결합제의 동종의 쌍의 정확한 결합 만이 증폭가능한 연장된 기록 태그를 생성할 것이다. 교차-반응성 결합제는 기록 태그 상에서 프라임 연장할 수 없을 것이며, 증폭가능한 연장된 기록 태그 생성물로 생성되지 않는다. 이러한 시도는 본원에 개시된 방법의 특이성을 크게 향상시킨다. 동일한 원리를 3중 결합제 세트에 적용할 수 있으며, 여기서, 3회 주기의 결합을 사용한다. 제1의 결합 주기에서, 기록 태그 상의 일반적인 3' Sp 서열은 결합제 암호화 태그의 일반적인 스페이서와 상호작용한다. 프라이머 연장은 특이적인 5' 스페이서를 포함하는 암호화 태그 정보를 기록 태그로 이전시킨다. 후속적인 결합 주기는 결합제의 암호화 태그 상에 분석물 특이적인 스페이서를 사용한다.

특정 구현예에서, 암호화 태그는 또한 암호화 태그가 연결된 결합제에 대한 유일한 분자 확인인자를 추가로 포함할 수 있다. 결합제에 대한 UMI는 서열분석 판독물에 대한 연장된 암호화 태그 또는 디-태그 분자를 사용하는 구현예에서 유용할 수 있으며, 이는 암호기 서열과 함께 거대분자에 대한 결합제 및 유일한 결합 사건의 수의 확인과 관련된 정보를 제공한다.

다른 구현예에서, 암호화 태그는 무작위 서열(N개의 세트, 여기서 N= A, C, G, T로부터 무작위 선택, 단어의 세트로부터 무작위 선택이다)을 포함한다. 일련의 "n"회의 결합 주기 및 암호화 태그 정보의 (연장된) 기록 태그로의 이전 후, 최종의 연장된 기록 태그 생성물은 일련의 무작위 서열로 구성될 것이며, 이는 최종의 연장된 기

록 태그에 대한 "복합체" 유일한 분자 확인인자(UMI)를 총괄적으로 형성한다. 예를 들어, 각각의 암호화 태그가 (NN) 서열(4*4=16개의 가능한 서열)을 함유하는 경우, 10회의 서열분석 주기 후, 10개의 분포된 2-머의 조합 세트가 형성되어 연장된 기록 태그 생성물에 대해 16^{10} ~ 10^{12}개의 가능한 복합체 UMI 서열의 총 다양성이 생성된다. 펩타이드 서열분석 실험이 ~10^{9}개의 분자를 사용하는 것을 고려할 때, 이러한 다양성은 서열분석 실험을 위한 UMI의 효과적인 세트를 사용하기에 훨씬 더 충분하다. 증가된 다양성은 암호화 태그내 보다 긴 무작위 영역(NNN, NNNN 등)을 단순히 사용함으로써 달성할 수 있다.

암호화 태그는 3' 스페이서 서열의 3' 말단에서 포함된 터미네이터 뉴클레오타이드를 포함한다. 결합제가 거대분자에 결합하고 이들의 상응하는 암호화 태그 및 기록 태그가 상보성 스페이서 서열을 통해 어닐링된 후, 프라이머 연장이 암호화 태그로부터의 정보를 기록 태그로 이전하거나, 또는 기록 태그로부터의 정보를 기록 태그로 이전하는 것이 가능하다. 암호화 태그의 3' 말단에서 터미네이터 뉴클레오타이드의 첨가는 기록 태그 정보의 암호화 태그로의 이전을 방지한다. 연장된 암호화 태그의 생성을 포함하는 본원에 기술된 구현예의 경우, 기록 태그의 3' 말단에 터미네이터 뉴클레오타이드를 포함시켜 암호화 태그 정보의 기록 태그로의 이전을 방지하는것이 바람직할 수 있다.

암호화 태그는 단일 가닥 분자, 이중 가닥 분자, 또는 부분적으로 이중 가닥일 수 있다. 암호화 태그는 평활 말단, 오버행 말단, 또는 각각 하나를 포함할 수 있다. 일부 구현예에서, 암호화 태그는 부분적으로 이중 가닥이며, 이는 성장하는연장된 기록 태그내에서 내부 암호기 및 스페이서 서열로 암호화 태그의 어닐링을 방지한다.

암호화 태그는 공유결합성 및 비-공유결합성 상호작용을 포함하는, 당해 분야에 공지된 어떠한 수단에 의해서도, 결합제에 직접 또는 간접적으로 결합될 수 있다. 일부 구현예에서, 암호화 태그는 효소적으로 또는 화학적으로 결합제에 결합될 수 있다. 일부 구현예에서, 암호화 태그는 연결을 통해 결합제에 결합될 수 있다. 다른 구현예에서, 암호화 태그는 친화성 결합 쌍(예컨대, 바이오틴 및 스트렙타비딘)을 통해 결합제에 결합된다.

일부 구현예에서, 결합제는 SpyCatcher-SpyTag 상호작용을 통해 암호화 태그에 결합된다(참고: 도 43b). SpyTag 펩타이드는 자발적인 이소펩타이드 연결을 통해 SpyCatcher 단백질에 대해 비가역성 공유결합성 결합을 형성함으로써 강요 및 가혹한 조건(force and harsh condition)을 견디는 펩타이드 상호작용을 생성하는 유전적으로 암호화된 방식을 부여한다(Zakeri et al., 2012, Proc. Natl. Acad. Sci. 109:E690-697; Li et al., 2014, J. Mol. Biol. 426:309-317). 결합제는 SpyCatcher 단백질을 포함하는 융합 단백질로서 발현될 수 있다. 일부 구현예에서, SpyCatcher 단백질은 결합제의 N-말단 또는 C-말단에 첨부된다. SpyTag 펩타이드는 표준 접합 화학을 사용하여 암호화 태그에 커플링시킬 수 있다(Bioconjugate Techniques, G. T. Hermanson,cademic Press (2013)).

다른 구현예에서, 결합제는 SnoopTag-SnoopCatcher 펩타이드-단백질 상호작용을 통해서 암호화 태그에 결합된다. SnoopTag 펩타이드는 SnoopCatcher 단백질과 이소펩타이드 결합을 형성한다(Veggiani et al., Proc. Natl. Acad. Sci. USA, 2016, 113:1202-1207). 결합제는 SnoopCatcher 단백질을 포함하는 융합 단백질로서 발현될 수 있다. 일부 구현예에서, SnoopCatcher 단백질은 결합제의 N-말단 또는 C-말단에 첨부된다. SnoopTag 펩타이드는 표준 접합 화학을 사용하여 암호화 태그에 커플링시킬 수 있다.

여전히 다른 구현예에서, 결합제는 HaloTag® 단백질 융합 태그 및 이의 화학적 리간드를 통해 암호화 태그에 결합된다. HaloTag는 합성 리간드(HaloTag 리간드)에 공유결합으로 결합하도록 설계된 변형된 할로알칸 데할로게나제이다(Los et al., 2008, ACS Chem. Biol. 3:373-382). 합성 리간드는 다양한 유용한 분자에 부착된 클로로알칸 링커를 포함한다. Halotag와 매우 특이적인 사이클로알칸 링커 사이의 공유결합 형태는 생리학적 조건 하에서 신속하게 일어나고 필수적으로 비가역성이다.

특정 구현예에서, 거대분자는 또한 비-동종의 결합제와 접촉된다. 본원에 사용된 바와 같은, 비-동종의 결합제는 고려되는 특수한 거대분자 이외의 상이한 거대분자 특징 또는 성분에 대해 선택적인 결합제를 지칭한다. 예를 들면, *n*개의 NTAA가 페닐알라닌인 경우, 펩타이드는 페닐알라닌, 타이로신, 및 아스파라긴 각각에 대해 선택적인 3개의 결합제와 접촉하며, 페닐알라닌에 대해 선택적인 결합제는 *n*번째 NTAA(즉, 페닐알라닌)에 대해 선택적으로 결합할 수 있는 제1 결합제일 수 있는 반면, 다른 2개의 결합제는 이러한 펩타이드에 대해 비-동종의 결합제일 수 있다(이들이 페닐알라닌 이외의 NTAA에 대해 선택성이므로). 그러나, 타이로신 및 아스파라긴 결합제는 샘플내 다른 펩타이드에 대해 동종의 결합제일 수 있다. *n*번 NTAA(페닐알라닌)가 이후에 펩타이드로부터 절단됨으로써, 펩타이드의 *n-1*번 아미노산은 *n-1*번 NTAA(예컨대, 타이로신)으로 전환되었고, 펩타이드는 이후에

동일한 3개의 결합제와 접촉되었으며, 타이로신에 대해 선택적인 결합제는 *n-1*번 NTAA(즉, 타이로신)에 대해 선택적으로 결합할 수 있는 제2의 결합제일 수 있지만 다른 2개의 결합제는 비-동종의 결합제일 수 있다(이들이 타이로신 이외의 NTAA에 대해 선택적이므로).

따라서, 제제가 결합제 또는 비-동종의 결합제인지의 여부는 결합에 이용가능한 특수한 거대분자 특징 또는 성분의 특성에 의존할 것이다. 또한, 다수의 거대분자가 멀티플렉스화된 반응에서 분석되는 경우, 하나의 거대분자에 대한 결합제는 다른 것에 대한 비-동종의 결합제일 수 있으며, 역으로도 가능하다. 따라서, 결합제에 관한 다음의 설명은 본원에 기술된 어떠한 유형의 결합제(즉, 동종의 및 비-동종의 결합제 둘 다)에도 적용가능함이 이해될 수 있다.

VIII. 암호화 태그 정보의 기록 태그로의 사이클릭 이전

본원에 기술된 방법에서, 결합제가 거대분자에 결합시, 이의 연결된 암호화 태그의 확인 정보는 거대분자와 관련된 기록 태그로 이전됨으로써 "연장된 기록 태그"를 생성한다. 연장된 기록 태그는 수행된 각각의 결합 주기를 나타내는 결합제의 암호화 태그로부터의 정보를 포함할 수 있다. 그러나, 예컨대, 결합제는 거대분자에 결합하는데 실패하므로, 암호화 태그가 손실되거나, 손상되거나 결함이 있거나, 프라이머 연장 반응이 실패하므로, 연장된 기록 태그는 또한 "손실된" 결합 주기를 경험할 수 있다. 결합 사건이 일어난 경우에도, 예컨대, 암호화 태그가 손상되거나 결손되었기 때문에, 오류가 프라이머 연장 반응에 도입되었기 때문에, 암호화 태그로부터의 정보의 기록 태그로의 이전은 불완전하거나 100% 미만의 정확도일 수 있다. 따라서, 연장된 기록 태그는 이의 관련된 거대분자에서 발생하는 결합 사건의 100%, 또는 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 65%, 55%, 50%, 45%, 40%, 35%, 30% 이하를 나타낼 수 있다. 더욱이, 연장된 기록 태그에 나타난 암호화 태그 정보는 상응하는 암호화 태그에 대해 적어도 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 또는 100% 동일성을 가질 수 있다.

특정 구현예에서, 연장된 기록 태그는 다수의, 연속적인 결합 사건을 나타내는 다수의 암호화 태그로부터의 정보를 포함할 수 있다. 이들 구현예에서, 단일의, 콘카테네이트된 연장된 기록 태그는 단일의 거대분자를 대표할 수 있다(참고: 도 2a). 본원에 지칭된 바와 같이, 암호화 태그 정보의 기록 태그로의 이전은 또한 다수의, 연속적인 결합 사건을 포함하는 방법에서 일어날 수 있는 연장된 기록 태그로의 이전을 포함한다.

특정 구현예에서, 결합 사건 정보는 사이클릭 양식으로 암호화 태그로부터 기록 태그로 이전된다(참고: 도 2a 및 2c). 교차-반응성 결합 사건은 적어도 2개의 상이한 암호화 태그, 2회 이상의 독립적인 결합 사건의 확인, 동일한 부류의 결합제(특수한 단백질에 대해 동종인)에 대한 맵핑을 요구함으로써 서열분석 후 정보적으로 여과될 수 있다. 임의의 샘플 또는 구획 바코드는 기록 태그, 및 또한 임의의 UMI 서열에 포함될 수 있다. 암호화 태그는 또한 암호기 및 스페이서 서열과 함께 임의의 UMI 서열을 함유할 수 있다. 보편적인 프라이밍 서열(U1 및 U2)는 또한 증폭 및 NGS 서열분석을 위한 연장된 기록 태그 속에 포함될 수 있다(참고: 도 2a).

특이적인 결합제와 관련된 암호화 태그 정보는 다양한 방법을 사용하여 기록 태그로 이전될 수 있다. 특정 구현예에서, 암호화 태그의 정보는 프라이머 연장을 통해 기록 태그로 이전된다(Chan, McGregor et al. 2015). 기록 태그 또는 연장된 기록 태그의 3'-말단 상의 스페이서 서열은 암호화 태그의 3' 말단의 상보성 스페이서 서열과 어닐링하며 폴리머라제(예컨대, 가닥-치환 폴리머라제)는 주형으로서 어닐링된 암호화 태그를 사용하여 기록 태그 서열을 연장한다(참고: 도 5 내지 7). 일부 구현예에서, 암호화 태그 암호기 서열 및 5' 스페이서에 대해 상보성인 올리고뉴클레오타이드를 암호화 태그에 예비-어닐링시켜 연장된 기록 태그 내에 존재하는 내부 암호기 및 스페이서 서열에 대한 암호화 태그의 하이브리드화를 방지한다. 단일 가닥으로 남은, 암호화 태그 상의 3' 말단 스페이서는 바람직하게는 기록 태그 상의 말단 3' 스페이서에 결합한다. 다른 구현예에서, 초기의 기록 태그는 단일 가닥 결합 단백질로 코팅하여 암호화 태그가 내부 부위로 어닐링하는 것을 방지할 수 있다. 대안적으로, 초기의 기록 태그를 또한 RecA(또는 uvsX와 같은 방출된 동족체)로 코팅하여 3' 말단이 완전히 이중 가닥인 암호화 태그로 침입하는 것을 촉진할 수 있다(Bell et al., 2012, Nature 491:274-278). 이러한 구조는 암호화 태그가 내부 기록 태그 요소와 상호작용하는 것을 방지하며, 여전히 연장된 기록 태그의 RecA 코팅된 3' 테일에 의한 가닥 침입에 민감하다(Bell, et al., 2015, Elife 4: e08646). 단일 가닥 결합 단백질의 존재는 가닥 치환 반응을 촉진시킬 수 있다.

바람직한 구현예에서, 프라이머 연장에 사용되는 DNA 폴리머라제는 가닥-치환 활성을 지니며 3'-5 엔소뉴클레아제 활성이 제한되어 있거나 고갈되어 있다. 이러한 폴리머라제 중 많은 예 중의 수개는 클레노우 엑소(Klenow exo)-(DNA Pol 1의 클레노우 단편), T4 DNA 폴리머라제 엑소-, T7 DNA 폴리머라제 엑소(Sequenase 2.0), Pfu 엑소-, 벤트(Vent) 엑소-, 딥 벤트 엑소(Deep Vent exo)-, Bst DNA 폴리머라제 큰 단편 엑소-, Bca Pol, 9° N

Pol, 및 Phi29 Pol 엑소-를 포함한다. 바람직한 구현예에서, DNA 폴리머라제는 실온 내지 45℃ 이하에서 활성이다. 다른 구현예에서, 호열성 폴리머라제의 "따뜻한 출발" 버젼을 사용하여 폴리머라제가 활성화되어 약 40℃ 내지 50℃에서 사용되도록 한다. 예시적인 따뜻한 출발 폴리머라제는 Bst 2.0 Warm Start DNA 폴리머라제(New England Biolabs)이다.

가닥-치환 복제에 유용한 첨가제는 세균, 바이러스, 또는 이. 콜라이의 SSB 단백질과 같은 진핵 기원의 다수의 단일-가닥 DNA 결합 단백질(SSB 단백질), 파아지 T4 유전자 32 생성물, 파아지 T7 유전자 2.5 단백질, 파아지 Pf3 SSB, 복제 단백질 A RPA32 및 RPA14 소단위(Wold, 1997); 아데노바이러스 DNA-결합 단백질, 헤르페스 단성 단백질 ICP8, BMRF1 폴리머라제 악세서리 소 단위, 헤르페스 바이러스 UL29 SSB-유사 단백질과 같은 다른 DNA 결합 단백질 중 어느 것; 파아지 T7 헬리카제/프리마제, 파아지 T4 유전자 41 헬리카제, 이. 콜라이 Rep 헬리카제. 이. 콜라이 recBCD 헬리카제, recA, 이. 콜라이 및 진핵세포 토포이소머라제(Champoux, 2001)와 같은, DNA 복제에 관여하는 것으로 알려진 다수의 복제 복합체 단백질 중 어느 것을 포함한다.

재암호화 태그의 말단 스페이서 서열이 연장 자가-연장을 프라이밍하는 경우와 같은, 잘못된(Mis)-프라이밍 또는 자가(self)-프라이밍 사건은 단일의 표준 결합 단백질(T4 유전자 32, 이. 콜라이 SSB 등), DMSO(1-10%), 포름아미드(1-10%), BSA(10-100 ug/ml), TMACl(1-5 mM), 황산암모늄(10-50 mM), 베타인(1-3M), 글리세롤(5-40%), 또는 에틸렌 글리콜(5-40%)을 프라이머 연장 반응에 혼입시킴으로써 최소화시킬 수 있다.

대부분의 A형 폴리머라제는 클레노우 엑소-T7 DNA 폴리머라제 엑소- (Sequenase 2.0)와 같이, 3' 엑소뉴클레아제 활성(내인성 또는 가공된 제거)이 고갈되어 있고, Taq 폴리머라제는 뉴클레오타이드, 바람직하게는 아데노신 염기(서열의 내용에 따라 보다 적은 정도의 G 염기로)의 듀플렉스 증폭 생성물의 3' 평활 말단에 대한 비-주형화된 첨가를 촉매한다. Taq 폴리머라제의 경우, 3' 피리미딘(C>T)은 비-주형화된 아데노신 첨가를 최소화하지만, 3' 푸린 뉴클레오타이드 (G>A)는 비-주형화된 아데노신 첨가를 선호한다. 프라이머 연장을 위해 Taq 폴리머라제를 사용하는 구현예에서, 결합제로부터 먼 스페이서 서열과 인접한 바코드 서열(예컨대, 암호기 서열 또는 주기 특이적인 서열) 사이의 암호화 태그내 티미딘 염기의 치환은 기록 태그의 스페이서 서열의 3' 말단에서 비-주형화된 아데노신 뉴클레오타이드의 산발적인 혼입을 수용한다(도 43a). 이러한 방식으로, 연장된 기록 태그(비-주형화된 아데노신 염기가 있거나 없는)는 암호화 태그에 어닐링되어 프라이머 연장을 겪는다.

대안적으로, 비-주형화된 염기의 첨가는 돌연변이체 폴리머라제(중온성 또는 호열성)를 사용함으로써 감소시킬 수 있으며, 여기서 비-주형화된 말단 트랜스퍼라제 활성은 특히 O-나선 영역에서 하나 이상의 전 돌연변이에 의해 감소되었다(참고: 미국 특허 제7,501,237호)(Yang,statke et al. 2002). P3' 엑소뉴클레아제 결핍성이고 가닥-치환능을 가진 Pfu exo-는 또한 비-주형화된 말단 트랜스퍼라제 활성을 가지지 않는다.

다른 구현예에서, 최적의 폴리머라제 연장 완충액은 pH 6 내지 9에서 트리스-아세테이트, 트리스-HCl, HEPES 등과 같은 40-120 mM의 완충제로 구성된다.

연장된 기록 태그의 내부 영역을 지닌 연장된 기록 태그의 말단 스페이서 서열의 자가-어닐링에 의해 개시된 자가-프라이밍/잘못된-프라이밍 사건은 기록/연장된 기록 태그내에서 슈도-상보성 염기를 포함시킴으로써 최소화될 수 있다(Lahoud, Timoshchuk et al. 2008),(Hoshika, Chen et al. 2010). 슈도-상보성 염기는 화학적 변형의 존재로 인하여 각각 서로 듀플렉스의 형성을 위한 유의적으로 감소된 하이브리드화 친화성을 나타낸다. 그러나, 많은 슈도-상보성 변형된 염기는 천연의 DNA 또는 RNA 서열과 강력한 염기 쌍을 형성할 수 있다. 특정 구현예에서, 암호화 태그 스페이서 서열은 다수의 A 및 T 염기로 구성되며, 시판되는 슈도-상보성 염기 2-아미노아데닌 및 2-티오티민은 포스포르아미디트 올리고뉴클레오타이드 합성을 사용하여 기록 태그내로 도입시킨다. 추가의 슈도상보성 염기는 반응에 슈도-상보성 뉴클레오타이드를 첨가함으로써 프라이머 연장 동안 연장된 기록 태그내로 혼입될 수 있다.(Gamper, Arar et al. 2006).

용액 속의 암호화 태그 표지된 결합제와 면역화된 단백질의 기록 태그의 비-특이적인 상호작용을 최소화하기 위하여, 기록 태그 스페이서 서열에 대해 상보성인 경쟁인자(또한 차단제로 지칭됨) 올리고뉴클레오타이드를 결합 반응에 가하여 비-특이적인 상호작용을 최소화시킨다(도 32의 A-D). 올리고뉴클레오타이드의 차단은 비교적 짧다. 과도한 경쟁인자 올리고뉴클레오타이드가 프라이머 연장 전에 결합 반응으로부터 세척제거되며, 이는 특히 약하게 상승된 온도(예컨대, 30 내지 50℃)에 노출되는 경우, 어닐링된 경쟁인자 올리고뉴클레오타이드를 기록 태그로부터 효과적으로 해리한다. 올리고뉴클레오타이드의 차단은 이의 3' 말단에서 터미네이터 뉴클레오타이드를 포함함으로써 프라이머 연장을 방지할 수 있다.

특정 구현예에서, 암호화 태그 상의 상보성 스페이서 서열에 대한 기록 태그 상의 스페이서 서열의 어닐링은 프

라이머 연장 반응 조건 하에서 준안정성이다(즉, 어닐링 Tm은 반응 온도와 유사하다). 이는 암호화 태그의 스페이서 서열이 기록 태그의 스페이서 서열로 어닐링된 어떠한 차단 올리고뉴클레오타이드도 치환시키도록 한다.

특이적인 결합제와 관련된 암호화 태그 정보는 또한 연결을 통해 기록 태그로 이전시킬 수 있다(참고: 예컨대, 도 6 및 7). 연결은 평활 말단 연결 또는 점성 말단 연결일 수 있다. 연결은 효소적 연결 반응일 수 있다. 리가제의 예는 T4 DNA 리가제, T7 DNA 리가제, T3 DNA 리가제, Taq DNA 리가제, 이. 콜라이 DNA 리가제, 9° N DNA 리가제, Electroligase®을 포함하나, 이에 한정되지 않는다. 대안적으로, 연결은 화학적 연결 반응이다(참고: 도 7). 예시에서, 스페이서가 없는 연결은 암호화 태그 상에 아암(arm)이 있는 "기록 헬퍼" 서열의 하이브리드화를 사용하여 달성된다. 어닐링된 상보체 서열은 표준 화학적 연결 또는 "클릭 화학"을 사용하여 화학적으로 연결시킨다(Gunderson, Huang et al. 1998, Peng, Li et al. 2010, El-Sagheer, Cheong et al. 2011, El-Sagheer, Sanzone et al. 2011, Sharma, Kent et al. 2012, Roloff 및 Seitz 2013, Litovchick, Clark et al. 2014, Roloff, Ficht et al. 2014).

다른 구현예에서, PNA의 이전은 발표된 기술을 사용하여 화학적 연결로 달성할 수 있다. PNA의 구조는 이것이 5' N-말단 아민 그룹 및 반응하지 않은 3' C-말단 아미드를 갖도록 하는 것이다. PNA의 화학적 연결은 말단이 화학적으로 활성이 되도록 변형되는 것을 요구한다. 이는 전형적으로 5' N-말단을 시스테이닐 모이어티로 유도체화하고 3' C-말단을 티오에스테르 모이어티로 유도체화함으로써 이루어진다. 이러한 변형된 PNA는 표준의 천연 화학적 연결 조건을 사용하여 용이하게 커플링된다(Roloff et al., 2013, Bioorgan. Med. Chem. 21:3458-3464).

일부 구현예에서, 암호화 태그 정보는 토포이소머라제를 사용하여 이전할 수 있다. 토포이소머라제는 기록 태그 상의 토포-하전된 3' 포스페이트를 암호화 태그의 5' 말단, 또는 이의 상보체에 연결하기 위해 사용될 수 있다(Shuman et al., 1994, J. Biol. Chem. 269:32678-32684).

본원에 기술된 바와 같이, 결합제는 해독후 변형된 아미노산에 결합될 수 있다. 따라서, 펩타이드 거대분자를 포함하는 특정 구현예에서, 연장된 기록 태그는 아미노산 서열 및 해독 후 변형에 관한 암호화 태그 정보를 포함한다. 일부 구현예에서, 내부 해독후 변형된 아미노산(예컨대, 인산화, 글리코실화, 석시닐화, 우비퀴틴화, S-니트로실화, 메틸화, N-아세틸화, 지질화 등)의 검출은 말단 아미노산(예컨대, NTAA 또는 CTAA)의 검출 및 절단 전에 달성된다. 하나의 예에서, 펩타이드는 PTM 변형을 위해 결합제와 접촉되며, 관련된 암호화 태그 정보는 상술한 바와 같이 기록 태그로 이전한다(참고: 도 8의 A). 일단 아미노산 변형에 관한 암호화 태그 정보의 검출 및 이전이 완성되면, PTM 개질 그룹을 N-말단 또는 C-말단 분해 방법을 사용하여 주요 아미노산 서열에 대한 암호화 태그 정보의 검출 및 이전 전에 제거할 수 있다. 따라서, 수득되는 연장된 기록 태그는 주요 아미노산 서열 정보를 따라, 순차적인 순서가 아니지만, 펩타이드 서열내 해독 후 변형의 존재를 나타낸다(참고: 도 8의 B).

일부 구현예에서, 내부 해독후 변형된 아미노산의 검출은 주요 아미노산 서열의 검출과 동시에 일어날 수 있다. 하나의 예에서, NTAA(또는 CTAA)는 단독으로 또는 결합제의 라이브러리(예컨대, 20개의 표준 아미노산 및 선택된 해독후 변형된 아미노산에 대한 결합제로 구성된 라이브러리)의 일부로서, 해독후 변형된 아미노산에 대해 특이적인 결합제와 접촉된다. 말단 아미노산 절단 및 결합제(또는 결합제의 라이브러리)와의 접촉의 연속적인 주기가 이어진다. 따라서, 수득되는 연장된 기록 태그는 주요 아미노산 서열과 관련하여 해독 후 변형의 존재 및 순서를 나타낸다.

특정 구현예에서, 기록 태그의 앙상블(ensemble)을 거대분자당 사용하여 암호화 태그 정보 이전의 전체적인 강인성(robustness) 및 효능을 증진시킬 수 있다(참고: 예를 들면, 도 9). 단일의 기록 태그보다는 제공된 거대분자와 관련된 기록 태그의 앙상블의 사용은 기록 태그에 대한 암호화 태그의 잠재적으로 보다 높은 커플링 수율 및 라이브러리의 보다 높은 전체적인 수율로 인하여 라이브러리 구조의 효능을 개선시킨다. 단일의 콘카테네이트된 연장된 기록 태그의 수율은 콘카테네이션의 단계별 수율에 직접 의존하는 반면, 암호화 태그 정보를 수용할 수 있는 다수의 기록 태그의 사용은 콘카테네이션의 기하급수적 손실을 겪지 않는다.

이러한 구현예의 예는 도 9 및 10에 나타나 있다. 도 9의 A 및 10의 A의 예에서, 다수의 기록 태그는 단일의 거대분자(단일 비드에 대한 단일 거대 분자의 공간적 동시-국재화 또는 국한으로 인해)와 관련된다. 결합제는 주기적인 양식으로 고체 지지체에 노출되며 이들의 상응하는 암호화 태그는 정보를 각각의 주기에서 동시-국재화된 다수의 기록 태그 중 하나로 이전한다. 도 9의 A에 나타낸 실시예에서, 결합 주기 정보는 암호화 태그 상에 존재하는 스페이서로 암호화된다. 각각의 결합 주기의 경우, 결합제의 세트는 설계된 주기-특이적인 스페이서 서열로 표시된다(도 9의 A 및 9의 B). 예를 들면, NTAA 결합제의 경우에, 동일한 아미노산 잔기에 대한 결합제

는 상이한 암호화 태그로 표지되거나 스페이서 서열내에 주기-특이적인 정보를 포함함으로써 결합제 동일성 및 주기 수 둘 다를 나타낸다.

도 9의 A에 나타낸 바와 같이, 결합의 제1의 주기(주기 1)에서, 다수의 NTAA 결합제는 거대분자와 접촉된다. 주기 1에 사용된 결합제는 기록 태그의 스페이서 서열에 대해 상보성인 일반적인 스페이서 서열을 지닌다. 주기 1에 사용된 결합제는 또한 주기 1 특이적인 서열을 포함하는 3'-스페이서 서열을 지닌다. 결합 주기 1 동안에, 제1의 NTAA 결합제는 거대분자의 유리 말단에 결합하며, 제1의 암호화 태그 및 기록 태그내 일반적인 스페이서 서열의 상보성 서열은 어닐링되고 제1의 암호화 태그의 정보는 일반적인 스페이서 서열로부터의 프라이머 연장을 통해 동종의 기록 태그로 이전된다. NTAA를 제거하여 새로운 NTAA를 노출시킨 후, 결합 주기 2는 기록 태그의 스페이서 서열에 대해 상보성인 일반적인 스페이서 서열을 지닌 다수의 NTAA 결합제와 접촉된다. 주기 2에 사용된 결합제는 또한 주기 2 특이적인 서열을 포함하는 3'-스페이서 서열을 지닌다. 제2의 NTAA 결합제는 거대분자의 NTAA에 결합하며, 제2의 암호화 태그의 정보는 프라이머 연장을 통해 기록 태그로 이전된다. 이러한 주기는 "n"회 결합 주기까지 반복하여, 단일의 거대분자와 동시-국재화된 다수의 연장된 기록 태그를 생성하며, 여기서 각각의 연장된 기록 태그는 하나의 결합 주기로부터 암호화 태그 정보를 지닌다. 각각의 연속된 결합 주기에 사용된 결합제의 각각의 세트가 암호화 태그내에 주기 특이적인 스페이서 서열을 지니므로, 결합 주기 정보는 수득되는 연장된 기록 태그내 결합제 정보와 관련될 수 있다.

대안적인 구현예에서, 다수의 기록 태그는 도 9의 A에서와 같이 고체 지지체(예컨대, 비드) 상의 단일의 거대분자와 관련되지만, 이 경우 특수한 결합 주기에 사용된 결합제는 현재의 결합 주기에 대한 주기-특이적인 스페이서 및 다음 결합 주기에 대한 주기 특이적인 스페이서에 의해 플랭킹된다(도 10의 A 및 10의 B). 이러한 설계의 이유는 최종의 조립 PCR 단계(도 10의 C)를 지지함으로써 연장된 기록 태그의 집단을 단일의 동시-선형의, 연장된 기록 태그로 전환하는 것이다. 단일의, 동시-선형의 연장된 기록 태그의 라이브러리는 서열분석 전에 농축, 공제 및/또는 표준화 방법에 적용시킬 수 있다. 제1의 결합 주기(주기 1)에서, 제1의 결합제의 결합 시, 주기 1 특이적인 스페이서(C'1)를 포함하는 암호화 태그의 정보는 이의 말단에서 상보성 주기 1 특이적인 스페이서(C1)를 포함하는 기록 태그로 이전된다. 제2의 결합 주기(주기 2)에서, 제2의 결합제의 결합시, 주기 2 특이적인 스페이서(C'2)를 포함하는 암호화 태그의 정보는 이의 말단에서 상보성 주기 2 특이적인 스페이서(C2)를 포함하는 상이한 기록 태그로 이전된다. 이러한 과정은 n번째 결합 주기까지 지속된다. 일부 구현예에서, 연장된 기록 태그내 n번째 암호화 태그는 보편적인 역 프라이밍 서열로 캡핑(capping)되는데, 예컨대, 보편적인 역 프라이밍 서열은 n번째 암호화 태그 설계의 일부로서 포함될 수 있거나 보편적인 역 프라이밍 서열은 테일된 프라이머를 사용하는 증폭 반응과 같은 n번째 결합 주기 후에 반응에 가해질 수 있다. 일부 구현예에서, 각각의 결합 주기에서 거대분자는 이들의 상응하는 결합제 및 결합 주기 정보에 관한 확인 정보를 포함하는 암호화 태그에 결합된 결합제의 수집에 노출된다(도 9 및 도 10). 특수한 구현예에서, *n*번째 결합 주기의 완료 후, 연장된 기록 태그로 코팅된 연장 기질은 오일 유화액 속에 두어 평균적으로 대략 1개 이하의 비드/소적이 존재하도록 한다. 조립 PCR을 이후에 사용하여 비드로부터 연장된 기록 태그를 증폭시키고 다수의 별도의 기록 태그를 별도의 연장된 기록 태그내에서 주기 특이적인 스페이서 서열을 통한 프라이밍에 의해 공선형 순서로 조립된다(도 10의 C)(Xiong et al., 2008, FEMS Microbiol. Rev.32:522-540). 대안적으로, 결합제의 암호화 태그를 지닌 주기-특이적인 스페이서를 사용하는 대신에, 주기 특이적인 스페이서를 각각의 결합 주기 동안 또는 후에 연장된 기록 태그에 별도로 가할 수 있다. 단일의 거대분자 대 단일의 콘카테네이트화된 연장된 기록 태그를 총괄적으로 나타내는, 연장된 기록 태그의 집단을 사용하는 한가지 장점은 보다 높은 농도의 기록 태그가 암호화 태그 정보의 이전의 효능을 증가시킬 수 있다는 것이다. 또한, 결합 주기를 수회 반복하여 동종의 결합 사건의 완료를 보증할 수 있다. 또한, 연장된 기록 태그의 표면 증폭은 정보 이전의 풍부성을 제공할 수 있다(참고: 도 4b). 암호화 태그 정보가 항상 이전되지 않는 경우, 대부분의 경우에 이는 암호화 태그 정보의 불완전한 수집을 사용하여 단백질과 같은 매우 높은 정보 함량을 갖는 거대분자를 확인할 수 있다. 심지어 짧은 펩타이드를 매우 큰 수의 가능한 단백질 서열로 구현할 수 있다. 예를 들면, 10-머 펩타이드는 20^{10}개의 가능한 서열을 갖는다. 따라서, 결실 및/또는 모호성(ambiguity)을 함유할 수 있는 부분적이거나 불완전한 서열을 여전히 유일하게 여전히 맵핑할 수 있다.

이들의 천연 구조의 단백질이 질의되는 일부 구현예에서, 사이클릭 결합 검정은 결합제와 근접한 스페이서내에서 절단가능하거나 닉생성가능한 DNA 쇄로 구성된 결합제를 지닌 암호화 태그로 수행된다(도 32). 예를 들면, 결합제에 근접한 스페이서는 우라실-특이적인 절개 시약(USER)에 의해 닉생성될 수 있는 하나 이상의 우라실 염기를 지닐 수 있다. 다른 실시예에서, 결합제에 근접한 스페이서는 듀플렉스의 하나의 가닥 만을 가수분해하는 닉생성 엔도뉴클레아제에 대한 인식 서열을 포함할 수 있다. 이러한 설계는 연장된 기록 태그로부터의 결합제의

제거를 허용하며 후속적인 면역검정 주기를 위한 유리된 단일 가닥 DNA 스페이서 요소를 생성한다. 바람직한 구현예에서, 우라실 염기는 암호화 태그내로 포함되어 프라이머 연장 단계 후 결합제의 효소적 USER 제거를 허용한다(도 32e 내지 32f). 우라실의 USER 절개 후, 결합제 및 트렁케이트된 암호화 태그는 고 염(4M NaCl, 25% 포름아미드) 및 약한 열을 포함하는 다양한 약한 조건 하에 제거하여 단백질-결합제 상호작용을 파괴할 수 있다. 기록 태그 상에 어닐링되어 남아있는 다른 트렁케이트된 암호화 태그 DNA 조각(stub)(도 32f)은 약간의 승온에서 용이하게 해리된다.

결합제에 근접한 스페이서 요소 내 절단가능하거나 닉생성가능한 DNA 가닥으로 구성된 암호화 태그는 또한 다수의 결합된 결합제로부터 암호화 태그 정보의 이전을 위한 단일의 균질한 검정을 허용한다(참고: 도 33). 바람직한 구현예에서, 결합제의 근접한 암호화 태그는 닉생성 엔도뉴클레아제 서열 모티프를 포함하며, 이는 dsDNA와 관련하여 정의된 서열 모티프에서 닉생성 엔도뉴클레아제에 의해 인식되어 닉생성된다. 다수의 결합제의 결합 후, 조합된 폴리머라제 연장(가닥-치환 활성의 결여) + 닉생성 엔도뉴클레아제 시약 혼합물을 사용하여 암호화 태그의 근접한 기록 태그 또는 연장된 기록 태그로의 반복된 이전을 생성한다. 각각의 이전 단계 후, 수득되는 연장된 기록 태그-암호화 태그 듀플렉스를 닉생성 엔도뉴클레아제로 닉형성시켜 결합제에 부착된 트렁케이트된 스페이서를 방출하고 연장된 기록 태그 3' 스페이서 서열을 노출시키며, 이는 추가의 근접한 결합된 결합제의 암호화 태그에 어닐링시킬 수 있다(도 33의 B 내지 D). 암호화 태그 스페이서 서열내 닉생성 모티프의 치환을 설계하여 준안정한 하이브리드를 생성할 수 있으며, 이는 비-절단된 암호화 태그 스페이서 서열로 용이하게 교환될 수 있다. 이러한 방식으로, 2개 이상의 결합제를 동일한 단백질 분자에 동시 결합시키는 경우, 기록 태그 상으로 다양하게 결합된 결합제로부터의 암호화 태그 정보의 콘카테네이션을 통한 결합 정보가 어떠한 사이클릭 시약 교환없이 단일의 반응 혼합물 속에서 발생한다(도 33의 C 및 D). 이러한 구현예는 특히 단백질 상의 다가 에피토프에 대한 폴리클로날 항체(또는 모노클로날 항체의 혼합된 집단)을 사용하는 차 세대 단백질 검정(NGPA)에서 특히 유용하다.

변성된 단백질, 폴리펩타이드, 및 펩타이드의 분석을 포함하는 구현예의 경우, 결합된 결합제 및 어닐링된 암호화 태그는 고 변성 조건(예컨대, 0.1 내지 0.2 N NaOH, 6M 우레아, 2.4 M 구아니디늄 이소티오시아네이트, 95% 포름아미드 등)을 사용함으로써 프라이머 연장 후 제거할 수 있다.

IX. 암호화 태그 또는 디-태그 작제물로의 기록 태그 정보의 사이클릭 이전

다른 국면에서, 거대분자에 대한 결합제의 결합 후 암호화 태그로부터 기록 태그로의 정보를 적기보다는, 정보를 임의의 UMI 서열(예컨대 특수한 펩타이드 또는 단백질 분자의 확인) 및 적어도 하나의 바코드(예컨대, 구획 태그, 분배 바코드, 샘플 바코드, 공간 위치 바코드 등)를 포함하는 기록 태그로부터 암호화 태그로 이전시킴으로서 연장된 암호화 태그를 생성시킬 수 있다(참고: 도 11의 A). 특정 구현예에서, 결합제 및 관련된 연장된 암호화 태그를 각각의 결합 주기 후 및, 임의로 에드만 분해화학 단계 전에 수집한다. 특정 구현예에서, 암호화 태그는 결합 주기 특이적인 태그를 포함한다. 사이클릭 에드만 분해시 NTAA의 검출과 같은 모든 결합 주기의 완료 후, 연장된 암호화 태그의 완전한 수집물을 증폭시키고 서열분석하며, 펩타이드상의 정보를 UMI(펩타이드 동질성), 암호기 서열 (NTAA 결합제), 구획 태그(단일 세포 또는 프로테오믹스의 소집단), 결합 주기 특이적인 서열(주기 수), 또는 이의 어떠한 조합 사이의 해리로부터 측정할 수 있다. 동일한 구획 태그/UMI 서열을 지닌 라이브러리 요소를 동일한 세포, 프로테옴의 소세트, 분자 등으로 다시 맵핑하고, 펩타이드 서열을 재구성할 수 있다. 이러한 구현예는 기록 태그가 에드만 분해 과정 동안에 지속적으로 너무 많이 손상되는 경우에 유용할 수 있다.

(a) 거대분자 및, 고체 지지체에 결합된 관련 기록 태그를 제공하는 단계; (b) 다수의 거대분자를 다수의 거대분자에 결합할 수 있는 다수의 결합제와 접촉시키는 단계(여기서 각각의 결합제는 결합제에 관한 확인 정보와 함께 암호화 태그를 포함한다); (c) (i) 거대분자 관련된 기록 태그의 정보를 거대분자에 결합된 결합제의 암호화 태그에 이전시켜 연장된 암호화 태그를 생성하거나(참고: 도 11의 A); (ii) 거대분자 관련된 기록 태그 및, 거대분자에 결합된 결합제의 암호화 태그의 정보를 디-태그 작제물로 이전시키는 단계(참고: 도 11의 B); (d) 연장된 암호화 태그 또는 디-태그 작제물을 수집하는 단계; (e) 임의로 단계 (b) 내지 (d)를 하나 이상의 결합 주기 동안 반복하는 단계; (f) 연장된 암호화 태그 또는 디-태그 작제물의 수집물을 분석하는 단계를 포함하는, 다수의 거대분자를 분석하는 방법이 본원에 제공된다.

특정 구현예에서, 기록 태그로부터 암호화 태그로의 정보 이전은 프라이머 연장 단계를 사용하여 달성할 수 있으며, 여기서 기록 태그의 3' 말단은 임의로 차단되어 기록 태그의 프라이머 연장을 방지한다(참고: 예컨대, 도 11의 A). 수득되는 연장된 암호화 태그 및 관련된 결합제는 각각의 결합 사건 및 정보 이전의 완료 후 수집할

수 있다. 도 11의 B에 나타낸 예에서, 기록 태그는 보편적인 프라이밍 부위 (U2'), 바코드(예컨대, 구획 태그 "CT"), 임의의 UMI 서열, 및 일반적인 스페이서 서열(Sp1)로 구성된다. 특정 구현예에서, 바코드는 개개의 구획을 나타내는 구획 태그이며, UMI는 질의되는 특수한 단백질 또는 펩타이드 분자에 대해 다시 서열 판독물을 맵핑할 수 있다. 도 11의 B의 예에 나타낸 바와 같이, 암호화 태그는 일반적인 스페이서 서열(Sp2'), 결합제 암호기 서열, 및 보편적인 프라이밍 부위(U3)로 구성된다. 암호화 태그-표지된 결합제의 도입 이전에, 기록 태그의 U2' 보편적인 프라이밍 부위에 대해 상보성이고 보편적인 프라이밍 서열 U1 및 주기 특이적인 태그를 포함하는 올리고뉴클레오타이드(U2)를 기록 태그 U2'에 어닐링한다. 또한, 어댑터 서열, Sp1'-Sp2를 기록 태그 Sp1에 어닐링한다. 이러한 어댑터 서열은 또한 암호화 태그의 Sp2' 서열과 상호작용하여, 기록 태그 및 암호화 태그를 서로 근접하도록 할 수 있다. 갭-충전된 연장 연결 검정은 결합 사건 전 또는 후에 수행된다. 갭 충전이 결합 주기 전에 수행되는 경우, 결합-후 주기 프라이머 연장 단계를 사용하여 디-태그 형성을 완료한다. 다수의 결합 주기에 걸친 디-태그의 수집 후, 디-태그의 수집을 서열분석하고 UMI 서열을 통해 원래의 펩타이드 분자에 다시 맵핑한다. 효능을 최대화하기 위해, UMI 서열의 다양성은 UMI에 의해 태그된 다수의 단일 분자의 다양성을 초과하여야만 함이 이해된다.

특정 구현예에서, 거대분자는 단백질 또는 펩타이드이다. 펩타이드는 생물학적 샘플로부터 단백질을 단편화하여 수득할 수 있다.

기록 태그는 DNA 분자, RNA 분자, PNA 분자, BNA 분자, XNA 분자, LNA 분자, γPNA 분자, 또는 이의 조합일 수 있다. 기록 태그는 이것이 관련된 거대분자(예컨대, 펩타이드)를 확인하는 UNI를 포함한다. 특정 구현예에서, 기록 태그는 구획 태그를 추가로 포함한다. 기록 태그는 또한 보편적인 프라이밍 부위를 포함할 수 있으며, 이는 하부 증폭에 사용될 수 있다. 특정 구현예에서, 기록 태그는 이의 3' 말단에서 스페이서를 포함한다. 스페이서는 암호화 태그내 스페이서에 대해 상보성일 수 있다. 기록 태그의 3'-말단은 차단되어(예컨대, 광-분해성 3' 차단 그룹) 폴리머라제에 의한 기록 태그의 연장을 방지하여, 거대분자 관련된 기록 태그의 정보의 암호화 태그로의 이전 또는 거대분자 관련된 기록 태그 및 암호화 태그의 정보의 디-태그 작제물로의 이전을 촉진할 수 있다.

암호화 태그는 암호화제가 연결된 결합제를 확인하는 암호기 서열을 포함한다. 특정 구현예에서, 암호화 태그는 암호화 태그가 연결된 각각의 결합제에 대한 유일한 분자 확인인자(UMI)를 추가로 포함한다. 암호화 태그는 보편적인 프라이밍 부위를 포함할 수 있으며, 이는 하부 증폭에 사용될 수 있다. 암호화 태그는 이의 3'-말단에 스페이서를 포함할 수 있다. 스페이서는 기록 태그내 스페이서에 대해 상보성일 수 있며 프라이머 연장 반응을 개시하여 기록 태그 정보를 암호화 태그로 이전시킬 수 있다. 암호화 태그는 또한 이로부터 연장된 암호화 태그 또는 디-태그가 기원한 결합 주기를 확인하기 위한, 결합 주기 특이적인 서열을 추가로 포함할 수 있다.

기록 태그의 정보의 암호화 태그로의 이전은 프라이머 연장 또는 연결에 의해 영향받을 수 있다. 기록 태그 및 암호화 태그의 정보의 디-태그 작제물로의 이전은 갭 충전 반응, 프라이머 연장 반응, 또는 둘 다에 의해 생성될 수 있다.

디-태그 분자는 연장된 기록 태그의 것과 유사한 기능성 성분을 포함할 수 있다. 디-태그 분자는 기록 태그로부터 기원한 보편적인 프라이밍 부위, 기록 태그로부터 기원한 바코드(예컨대, 구획 태그), 기록 태그로부터 기원한 임의의 유일한 분자 확인인자(UMI), 기록 태그로부터 기원한 임의의 스페이서, 암호화 태그로부터 기원한 암호기 서열, 암호화 태그로부터 기원한 임의의 유일한 분자 확인인자, 결합 주기 특이적인 서열, 암호화 태그로부터 기원한 임의의 스페이서, 및 암호화 태그로부터 기원한 보편적인 프라이밍 부위를 포함할 수 있다.

특정 구현예에서, 기록 태그는 바코드 암호화 단어의 조합적 콘카테네이션을 사용하여 생성시킬 수 있다. 조합적 암호화 단어의 사용은 이에 의해 어닐링 및 화학적 연결을 사용하여 PNA 기록 태그로부터의 정보를 암호화 태그 또는 디-태그 작제물로 이전시킬 수 있는 방법을 제공한다(참고: 예컨대, 도 12의 A 내지 D). 본원에 개시된 펩타이드를 분석하는 방법이 에드만 분해를 통한 말단 아미노산의 분해를 포함하는 특정 구현예에서, 이는 PNA와 같이, 에드만 분해의 가혹한 조건에 대해 내성인 기록 태그를 사용하는 것이 바람직할 수 있다. 에드만 분해 프로토콜에서 하나의 가혹한 단계는 N-말단 아미노산을 절단하기 위한 무수 TFA 처리이다. 이러한 단계는 전형적으로 DNA를 파괴할 것이다. DNA와는 대조적으로, PNA는 산 가수분해에 대해 매우 내성이다. PNA를 사용한 도전은 정보 이전의 효소적 방법이 보다 어려워진다는 것인데, 즉, 화학적 연결을 통한 정보 이전이 바람직한 방식이다. 도 11의 B에서, 기록 태그 및 암호화 태그 정보는 효소적 갭-충전 연장 연결 단계를 사용하여 기록되지만, 이는 폴리머라제가 PNA를 사용하는 것으로 개발되지 않는 한, 현재 PNA 주형을 사용하여 실현가능하지 않다. PNA 기록 태그로부터 바코드 및 UMI의 암호화 태그로의 쓰기는 용이하게 증폭되지 않는 생성물인, 화학적

연결의 요건으로 인하여 문제가 된다. 화학적 연결 방법은 문헌에서 집중적으로 기술되어 왔다(Gunderson et al. 1998, Genome Res. 8:1142-1153; Peng et al., 2010, Eur. J. Org. Chem. 4194-4197; El-Sagheer et al., 2011, Org. Biomol. Chem. 9:232-235; El-Sagheer et al., 2011, Proc. Natl. Acad. Sci. USA 108:11338-11343; Litovchick et al., 2014,rtif. DNA PNA XNA 5: e27896; Roloff et al., 2014, Methods Mol. Biol. 1050:131-141).

조합적 PNA 바코드 및 UMI 서열을 생성하기 위하여, n-머 라이브러리로부터 PNA 단어의 세트를 조합적으로 연결할 수 있다. 각각의 PNA 단어가 1,000개 단어의 공간(space)으로부터 기원하는 경우, 4개의 조합 서열은 $1,000^4 = 10^{12}$개 코드의 암호화 종을 생성한다. 이러한 방식으로, 4,000개의 상이한 DNA 주형 서열의 세트로부터 시작하여, 10^{12}개 초과의 PNA 코드를 생성할 수 있다(도 12의 A). 보다 작거나 큰 암호화 공간이 다수의 콘카테네이트화된 단어를 조절하거나, 다수의 기본적인 단어를 조절하여 생성할 수 있다. 따라서, PNA 기록 태그에 하이브리드화된 DNA 서열을 사용한 정보 이전은 DNA 단어 조립 하이브리드화 및 화학적 연결을 사용하여 완료할 수 있다(참고: 도 12의 B). PNA 주형 상에서 DNA 단어의 조립 및 DNA 단어의 화학적 연결 후, 수득되는 중간체를 사용하여 암호화 태그로/로부터 정보를 이전시킬 수 있다(참고: 도 12의 C 및 도 12의 D).

특정 구현예에서, 거대분자 및 관련된 기록 태그는 고체 지지체에 공유결합으로 결합된다. 고체 지지체는 비드, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 통과 칩, 시그널 변환 전자를 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구일 수 있다. 고체 지지체는 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드일 수 있다.

특정 구현예에서, 결합제는 단백질 또는 폴리펩타이드이다. 일부 구현예에서, 결합제는 변형된 또는 변이체 아미노펩티다제, 변형된 또는 변이체 아미노 아실 tRNA 신테타제, 변형된 또는 변이체 안티칼린, 변형된 또는 변이체 ClpS, 또는 변형된 또는 변이체 항체 또는 이의 결합 단편이다. 특정 구현예에서, 결합제는 단일 아미노산 잔기, 디-펩타이드, 트리-펩타이드, 또는 펩타이드의 해독 후 변형에 결합한다. 일부 구현예에서, 결합제는 N-말단 아미노산 잔기, C-말단 아미노산 잔기, 또는 내부 아미노산 잔기에 결합한다. 일부 구현예에서, 결합제는 N-말단 펩타이드, C-말단 펩타이드, 또는 내부 펩타이드에 결합한다. 일부 구현예에서, 결합제는 펩타이드의 해독 후 변형의 아미노산의 부위-특이적인 공유결합성 표지이다.

특정 구현예에서, 단계 (b)에서 다수의 거대분자와 다수의 결합제의 접촉 후, 거대분자 및 관련된 결합제를 포함하는 복합체는 고체 지지체로부터 해리되어 소적 또는 미세유동 소적의 유화액으로 분배된다. 일부 구현예에서, 각각의 미세유동 소적은 거대분자 및 결합제를 포함하는 최대 하나의 복합체를 포함한다.

특정 구현예에서, 기록 태그는 연장된 암호화 태그 또는 디-태그 작제물을 생성하기 전에 증폭된다. 거대분자 및 관련된 결합제를 포함하는 복합체가 소적 또는 미세유동 소적으로 구배되어 소적당 최대 하나의 복합체가 존재하도록 하는 구현예에서, 기록 태그의 증폭은 정보를 암호화 태그 또는 디-태그 작제물로 이전하기 위한 주형으로서 추가의 기록 태그를 제공한다(참고: 도 13 및 도 14). 유화액 융합 PCR을 사용하여 기록 태그 정보를 암호화 태그로 이전하거나 디-태그 작제물의 집단을 생성할 수 있다.

생성된 연장된 암호화 태그 또는 디-태그 작제물의 수집물은 분석 전에 증폭될 수 있다. 연장된 암호화 태그 또는 디-태그 작제물의 분석물의 분석은 핵산 서열분석 방법, 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석, 이온 반도체 서열분석, 또는 피로서열분석이다. 핵산 서열분석 방법은 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 또는 진전된 현미경을 사용한 DNA의 직접적인 영상화일 수 있다.

PITC, 생거 제제(DNFB), SNFB, 아세틸화 시약, 아미드화(구아니딘화) 시약 등과 같은 N-말단 아민을 화학적으로 표지하는 에드만 분해 및 방법은 또한 표준 핵산에서 내부 아미노산 및 엑소사이클릭 아민 또는 아데닌, 구아닌, 및 사이토신과 같은 PNA 염기를 변형시킬 수 있다. 특정의 구현예에서, 펩타이드의 라이신 잔기의 ε-아민은 산 무수물, 구아니딘화제, 또는 유사한 차단 시약을 사용하여 서열분석 전에 차단한다. DNA 염기의 엑소사이클릭 아민은 펩타이드의 훨씬 덜 반응성인 주요 N-말단 아민이지만, 내부 아미노산 및 DNA 염기에서 엑소사이클릭 아민에 대한 비-표적 활성을 감소시키는 N-말단 아민에 대한 아민 반응성 제제의 반응성을 제어하는 것은 서열분석 검정에서 중요하다. 변형 반응의 선택성은 pH, 용매(수성 대 유기, 비양성자성, 극성 비양자성, 이온성 액체 등), 염기 및 촉매, 보조-촉매, 온도, 및 시간과 같은 반응 조건을 조절함으로써 조정할 수 있다.

또한, DNA 염기에서 사이클릭 아민의 반응성은 DNA가 ssDNA 또는 dsDNA 형태인지에 의해 조정된다. 변형을 최소화하기 위해, NTAA 화학적 변형 전에, 기록 태그를 상보성 DNA 프로브: P1', {샘플 BCs}', {Sp-BC}' 등으로 가수분해할 수 있다. 다른 구현예에서, 보호된 엑소사이클릭 아민을 가진 핵산의 사용을 또한 사용할 수 있다(Ohkubo, Kasuya et al. 2008). 여전히 다른 구현예에서, SNFB와 같은 "거의 반응성이 아닌" 아민 표지 화합물은 DNA상에서 내부 아미노산 및 엑소사이클릭 아민의 오프-표적 표지를 이주한다(Carty and Hirs 1968). SNFB는 파라설포닐 그룹에 파라 니트로 그룹을 제거하는 보다 많은 전자가 존재하여, DFB보다 SNFB에 의한 거의 활성이 아닌 불소 치환을 가져온다는 사실로 인하여 DNFB보다 거의 반응성이 아니다.

NTAA α-아민 변형을 최적화하고 오프-표적 아미노산 변형 또는 DNA 변형을 최소화하기 위한 커플링 조건 및 커플링 시약의 적정은 화학 및 반응 조건(농도, 온도, 시간, pH, 용매 유형 등)의 조심스러운 선택을 통해 가능하다. 예를 들면, DNFB는 물 속에서에 대해 아세토니트릴과 같은 비양성자성 용매 속에서 2급 아민과 보다 용이하게 반응하는 것으로 알려져 있다. 엑소사이클릭 아민의 약한 변형은 여전히 상보성 프로브가 서열과 하이브리드화하도록 할 수 있지만 폴리머라제-기반 프라이머 연장을 파괴하는 경향이 있을 수 있다. 또한 수소 결합을 여전히 허용하면서 엑소사이클릭 아민을 보호하는 것이 가능하다. 이는 최근 공보에 기술되었는데 여기서 보호된 염기는 여전히 목적한 표적에 하이브리드화할 수 있다(Ohkubo, Kasuya et al. 2008). 일 구현예에서, 가공된 폴리머라제를 사용하여 DNA 암호화 태그 주형내에서 기록 태그의 연장 동안 보호된 염기를 지닌 뉴클레오타이드를 혼입할 수 있다. 다른 구현에에서, 가공된 폴리머라제를 사용하여 PNA 기록 태그 주형 상의 암호화 태그의 연장 동안 기록 태그 PNA 주형(보호된 염기의 존재 또는 부재하에서)상에 뉴클레오타이드를 혼입시킨다. 다른 구현예에서, 정보는 외인성 올리고뉴클레오타이드를 PNA 기록 태그로 어닐링함으로써 기록 태그를 암호화 태그로 이전시킬 수 있다. 하이브리드화의 특이성은 n-머 단어의 조립을 기반으로 하여, 서열 공간내에서 명백한 UMI를 선택함으로써 촉진시킬 수 있다(Gerry, Witowski et al. 1999).

한편 에드만-유사 N-말단 펩타이드 분해 서열분석을 사용하여 펩타이드의 선형 아미노산 서열을 측정할 수 있으며, 대안적인 구현예를 사용하여 연장된 기록 태그, 연장된 암호화 태그, 및 디-태그를 사용하는 방법으로 펩타이드의 부분적인 조성 분석을 수행할 수 있다. 결합제 또는 화학적 표지를 사용하여 펩타이드 상의 N-말단 및 내부 아미노산 또는 아미노산 변형 둘 다를 확인할 수 있다. 화학적 제제는 부위-특이적인 방식으로 아미노산(예컨대, 표지)를 공유결합으로 변형시킬 수 있다(Sletten and Bertozzi 2009, Basle, Joubert et al. 2010)(Spicer and Davis 2014). 암호화 태그는 단일 아미노산을 표적화하는 화학적 표지제(chemical lebeling agent)에 부착시켜 암호화 및 부위-특이적인 표지된 아미노산의 후속적인 확인을 촉진시킬 수 있다(참고: 도 13).

펩타이드 조성적 분석은 펩타이드의 사이클릭 분해를 필요로 하지 않으므로, 태그를 함유하는 DNA를 가혹한 에드만 화학에 노출시키는 쟁점을 피한다. 사이클릭 결합 방식에서, 연장된 암호화 태그 또는 디-태그를 사용하여 조성 정보(아미노산 또는 디펩타이드/트리펩타이드 정보), PTM 정보, 및 주요 아미노산 서열을 제공할 수 있다. 일 구현예에서, 이러한 조성 정보는 본원에 기재된 연장된 암호화 태그 또는 디-태그 시도를 사용하여 판독할 수 있다. UMI 및 구획 태그 정보와 조합되는 경우, 연장된 암호화 태그 또는 디-태그의 수집은 펩타이드에 있어서 조합 정보 및 이들의 원래의 구획 단백질 또는 단백질을 제공한다. 동일한 구획 태그(및 표면적으로 기원하는 단백질 분자)에 대해 역으로 맵핑하는 연장된 암호화 태그 또는 디-태그의 수집은 부분적인 조성 정보를 지닌 펩타이드를 맵핑하는 강력한 도구이다. 전체 프로테옴에 대해 역으로 맵핑하는 대신에, 구획 태그된 펩타이드의 수집은 단백질 분자의 제한된 소세트에 대해 역으로 맵핑하여, 맵핑의 유일성을 크게 증가시킨다.

본원에 사용된 결합제는 단일 아미노산, 디펩타이드, 트리펩타이드, 또는 심지어 보다 긴 펩타이드 서열 모티프를 인식할 수 있다. 테쓸러(Tessler)(2011, Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single Molecule Detection. Ph.D., Washington University in St. Louis)는 비교적 선택성 이펩타이드 항체가 하전된 디펩타이드 에피토프의 소세트에 대해 생성될 수 있음을 입증하였다(Tessler 2011). 대안적 단백질 스캐폴드(예컨대, aaRS, 안티칼린, ClpSs 등) 및 아프타머에 대한 지시된 발달의 적용을 사용하여 디펩타이드/트리펩타이드 결합제의 세트를 확장할 수 있다. 단일 단백질 분자에 대해 역 맵핑으로 커플링된 디펩타이드/트리펩타이드 조합 분석으로부터의 정보는 각각의 단백질 분자를 유일하게 확인하고 정량화하는데 충분할 수 있다. 최대, 총 400개의 가능한 디펩타이드 조합이 존재한다. 그러나, 가장 빈번하고 가장 항원성(하전된, 친수성, 소수성)인 디펩타이드의 소세트는, 이에 대해 결합제를 생성시키기에 충분하다. 이러한 수는 40 내지 100개의 상이한 결합제의 세트를 구성할 수 있다. 40개의 상이한 결합제의 세트의 경우, 평균 10-머 펩타이드는 적어도 하나의 결합제에 의해 결합되는 약 80%의 변화를 갖는다. 이러한 정보를 동일한 단백질 분자로부터 유래하는 모든 펩타이드와 조합하는 것은 단백질 분자의 확인을 가능하도록 할 수 있

다. 펩타이드 및 이의 유래하는 단백질에 관한 모든 이러한 정보를 합하여 보다 정확하고 정밀한 단백질 서열 특성화를 제공할 수 있다.

최근에 부분적인 펩타이드 서열 정보를 사용하는 디지탈 단백질 특성화 검정이 제안되었다(Swaminathan et al., 2015, PLoS Comput. Biol. 11:e1004080)(Yao, Docter et al. 2015). 즉, 이러한 시도는 시스테인, 라이신, 아르기닌, 타이로신, 아스파르테이트/글루탐산과 같은 표준 화학을 사용하여 용이하게 표지된 아미노산의 형광성 표지를 사용한다(Basle, Joubert et al. 2010). 부분적인 펩타이드 서열 정보를 사용한 챌린지는 프로테옴으로의 역 맵핑이 확인된 유일한 단백질이 없는, 1 대 다수의 해리이다. 이러한 1 대 다수의 맵핑 문제는 전체 프로테옴 공간을 펩타이드가 역 맵핑되는 단백질 분자의 제한된 소세트로 감소시킴으로써 해결할 수 있다. 필수적으로, 단일의 부분 펩타이드 서열은 100 또는 1000개의 상이한 단백질 서열로 역 맵핑될 수 있지만, 수개의 펩타이드(예를 들면, 단일의 단백질 분자의 소화물로부터 기원한 10개의 펩타이드)의 세트는 모두 구획내 단백질 분자의 소세트내에 함유된 단일의 단백질 분자로 역 맵핑한 후, 단백질 분자의 동일성을 유추하는 것이 보다 용이하다. 예를 들면, 동일한 분자로부터 기원하는 모든 펩타이드에 대한 펩타이드 프로테옴 맵의 교차는 가능한 단백질 동일성의 세트를 크게 제한한다(참고: 도 15).

특히, 부분적인 펩타이드 서열 또는 조성물의 맵핑화능은 구획의 태그 및 UNI를 획기적으로 사용함으로써 유의적으로 향상된다. 즉, 프로테옴은 초기에 바코드화된 구획으로 분배되며, 여기서 구획 바코드는 또한 UMI 서열에 부착된다. 구획 바코드는 구획에 대해 유일한 서열이며, UMI는 구획내 각각의 바코드화된 분자에 대해 유일한 서열이다(참고: 도 16). 일 구현예에서, 이러한 분배는 이의 전문이 참고로 포함된 PCT 공보 제 WO2016/061517호에 개시된 것과 유사한 방법을 사용하여, DNA 태그 표지된 폴리펩타이드를 비드에 부착된 DNA 구획 바코드에 대한 하이브리드화를 통해 비드의 표면과 직접 상호작용시킴으로써 달성된다(참고: 도 31). 프라이머 연장 단계는 비드-연결된 구획 바코드로부터의 정보를 폴리펩타이드 상의 DNA 태그로 이전한다(도 20). 다른 구현예에서, 이러한 분배는 바코드화된 비드 및 단백질 분자를 함유하는 UMI를 유화액의 소적내로 동시-캡슐화함으로써 달성된다. 또한, 소적은 임의로 단백질을 펩타이드로 소화시키는 프로테아제를 함유한다. 다수의 프로테아제를 사용하여 리포터 태그된 폴리펩타이드를 소화시킬 수 있다(Switzar, Giera et al. 2013). 부텔라제 I와 같은 효소-리가제와 프로테아제의 동시-캡슐화는 페길화와 같은, 효소에 대한 변형을 불러서, 프로테아제 소화에 대해 내성이 되도록 한다(Frokjaer and Otzen 2005, Kang, Wang et al. 2010). 소화 후, 펩타이드는 바코드-UMI 태그에 연결된다. 바람직한 구현예에서, 바코드-UMI 태그는 비드에 보유되어 하부 생화학적 조작을 촉진한다(참고: 도 13).

펩타이드에 대한 바코드-UMI 연결 후, 유화액을 깨뜨려 비드를 수거한다. 바코드화된 펩타이드는 이들의 주요 아미노산 서열, 또는 이들의 아미노산 조성에 의해 특성화될 수 있다. 펩타이드에 대한 정보의 유형 둘 다를 사용하여 이를 다시 프로테오믹스의 소세트에 역 맵핑할 수 있다. 일반적으로, 서열 정보는 조성 정보보다는 프로테오믹스의 훨씬 더 작은 소세트에 역 맵핑한다. 그럼에도 불구하고, 다수의 펩타이드(서열 또는 조성)로부터의 정보를 동일한 구획 바코드와 합함으로써, 단백질 또는 펩타이드가 기원하는 단백질을 유일하게 확인할 수 있다. 이러한 방식으로, 전체 프로테오믹스를 특성화하고 정량할 수 있다. 펩타이드에 있어서 주요 서열 정보는 펩타이드 서열을 나타내는 DNA 암호화된 라이브러리(DEL)의 연장된 기록 태그 생성과 함께 펩타이드 서열분석 반응을 수행함으로써 유도시킬 수 있다. 바람직한 구현예에서, 기록 태그는 구획 바코드 및 UMI 서열로 구성된다. 이러한 정보는 암호화 태그로부터 이전된 주요 또는 PTM 아미노산 정보와 함께 사용하여 최종의 맵핑된 펩타이드 정보를 생성한다.

펩타이드 서열 정보에 대한 대안은 구획 바코드 및 UMI에 연결된 펩타이드 아미노산 또는 디펩타이드/트리펩타이드 조합 정보를 생성하는 것이다. 이는 UMI-바코드화된 펩타이드를 지닌 비드를 아미노산 표지화 단계에 적용시킴으로써 달성되며, 여기서 각각의 펩타이드 상의 선택된 아미노산(내부)은 아미노산 코드 정보 및 다른 아미노산 UMI(AA UMI)를 포함하는 DNA 태그로 부위-특이적으로 표지된다(참고: 도 13). 화학적 표지로 가장 처리가능한 아미노산(AA)는 라이신, 아르기닌, 시스테인, 타이로신, 트립토판, 및 아스파르테이트/글루타메이트이지만, 이는 또한 다른 AA에 대한 표지 개략도를 개발하기에 용이할 수 있다(Mendoza and Vachet, 2009). 제공된 펩타이드는 동일한 유형의 몇가지 AA을 함유할 수 있다. 동일한 유형의 다수의 아미노산의 존재는 부착된 AA UMI 표지로 인해 구별될 수 있다. 각각의 표지화 분자는 아미노산의 계수를 가능하도록 하는 DNA 태그내에서 상이한 UMI를 갖는다. 화학적 표지화에 대한 대안은 결합제로 AA를 "표지"하는 것이다. 예를 들면, AA 코드 정보 및 AA UMI를 포함하는 암호화 태그를 사용하여 표지된 타이로신-특이적인 항체를 사용하여 펩타이드의 타이로신 모두를 표시할 수 있었다. 이러한 시도를 사용하는 주의사항(caveat)은 큰 거대한 항체를 사용하여 직면하는 입체 장애이며, 이상적으로는 보다 작은 scFv, 안티칼린, 또는 ClpS 변이체가

이러한 목적으로 사용될 수 있다.

일 구현예에서, AA를 태그화한 후, 정보는 펩타이드 복합체를 구획화함으로써 단일의 펩타이드가 소적당 함유되도록 하고 유화액 융합 PCR을 수행하여 구획화된 펩타이드의 아미노산 조성을 특성화하는 연장된 암호화 태그 또는 디-태그의 세트를 작제함으로써 펩타이드 상에 결합되거나 공유결합으로 커플링된 결합제와 관련된 기록 태그와 다수의 암호화 태그 사이에 이전된다. 디-태그를 서열분석한 후, 동일한 바코드를 지닌펩타이드 상의 정보를 단일의 단백질 분자로 역 맵핑할 수 있다.

특수한 구현예에서, 태그된 펩타이드 복합체는 작은 미니-구획(예컨대, 마이크로-유화액)내로 분배된 비드로부터 해리되어(참고: 도 13) 평균적으로 단일의 표지된/결합된 결합제 펩타이드 복합체가 제공된 구획에 잔류하도록 한다. 특수한 구현예에서, 이러한 구획화는 마이크로-유화액 소적의 생성을 통해 달성된다(Shim, Ranasinghe et al. 2013, Shembekar, Chaipan et al. 2016). 펩타이드 복합체 외에, PCR 시약은 또한 3개의 프라이머(U1, Sp, 및 $U2_{tr}$)와 함께 소적 내에서 동시-캡슐화된다. 소적 형성 후, 아주 적은 주기의 유화액 PCR이 U1 및 Sp 어닐링보다 더 높은 어닐링 온도에서 수행되어(~5 내지 10회 주기) 기록 태그 생성물을 증폭시킨다(참고: 도 13). PCR의 이러한 초기 5 내지 10회 주기 후, 어닐링 온도는 아미노산 코드 태그의 $U2_{tr}$ 및 Sp_{tr}이 증폭에 관여하도록 강하되며, 다른 ~10회 라운드가 수행된다. 3개의-프라이머 유화액 PCR은 펩타이드 및 이의 아미노산 조성의 디-태그 라이브러리 표시를 생성하는 모든 AA 코드 태그를 지닌 펩타이드 UMI-바코드와 효과적으로 조합한다. 3개의 프라이머 PCR를 수행하는 양식 및 태그의 콘카테네이션을 또한 사용할 수 있다. 다른 구현예는 광-탈차단(photo-deblocking)에 의해 활성화된 3' 차단된 U2 프라이머의 사용, 또는 불안정한 차단된 3' 폴리뉴클레오타이드의 3' 탈차단을 개시하는 오일 가용성 환원제의 첨가이다. 유화액 PCR 후, 다른 라운드의 PCR을 일반적인 프라이머에서 수행하여 NGS 서열분석을 위한 라이브러리 요소를 포맷화할 수 있다.

이러한 방식으로, 라이브러리 요소의 상이한 서열 성분을 계수 및 분류 목적을 위해 사용할 수 있다. 제공된 펩타이드(구획 바코드-UMI 조합에 의해 확인된)의 경우, 많은 라이브러리 요소가 존재하며, 각각은 확인 AA 코드 태그 및 AA UMI를 지닌다(참고: 도 13). AA 코드 및 관련된 UMI를 사용하여 제공된 펩타이드 내에서 제공된 아미노산 유형의 발생을 계수한다. 따라서, 펩타이드(아마도 GluC, LysC, 또는 Endo AsnN 소화물)은 공간 순서화와 관계없이 이의 아미노산 조성(예컨대, 2개의 Cys, 1개의 Lys, 1개의 Arg, 2개의 Tyr 등)에 의해 특성화된다. 그럼에도 불구하고 이는 펩타이드를 프로테옴의 소세트로 맵핑하기에 충분한 신호를 제공하며, 동일한 단백질 분자로부터 기원한 다른 펩타이드와 함께 사용되는 경우, 단백질을 유일하게 확인하여 정량화한다.

X. 말단 아미노산(TAA) 표지화 방법

특정 구현예에서, 펩타이드의 말단 아미노산(예컨대, NTAA 또는 CTAA)은 본원에 기술된 방법에서 펩타이드를 결합제와 접촉시키기 전에 변형시키거나 표지한다.

일부 구현예에서, NTAA는 페닐이소티오시아네이트(PITC)와 반응하여 페닐티오카바모일(PTC)-NTAA 유도체를 생성한다. 에드만 분해는 전형적으로 페닐 이소티오시아네이트(PITC)를 사용하여 N-말단을 표지한다. PITC는 본원에 개시된 방법에 매우 적합한 2개의 특성을 갖는다: (1) PITC는 N-말단 아민 그룹을 고 효율로 표지하고; (2) 수득되는 PTC 유도체화된 NTAA는 산 처리시 자가-이성체화를 겪어서, 나머지 펩타이드로부터 아미노산의 절단을 야기한다.

NTAA를 표지하는데 사용될 수 있는 다른 시약은 다음을 포함한다: 4-설포페닐 이소티오시아네이트, 3-피리딜 이소티오시아네이트(PYITC), 2-피페리디닐에틸 이소티오시아네이트(PEITC), 3-(4-모르폴리노)프로필 이소티오시아네이트(MPITC), 3-(디에틸아미노)프로필 이소티오시아네이트(DEPTIC)(Wang et al., 2009,al Chem 81: 1893-1900), (1-플루오로-2,4-디니트로벤젠(생거 시약, DNFB), 단실 클로라이드 (DNS-Cl, 또는 1-디메틸아미노나프탈렌-5-설포닐 클로라이드), 4-설포닐-2-니트로플루오로벤젠(SNFB), 아세틸화 시약, 아미드화(구아니딘화) 시약, 2-카복시-4,6-디니트로클로로벤젠, 7-메톡시쿠마린 아세트산, 티오아실화 시약, 티오아세틸화 시약, 및 티오벤질화시약. NTAA가 표지화에 대해 차단되는 경우, N-아세틸 블록을 아실 펩타이드 하이드롤라제(APH)로 제거하는 것과 같은 말단을 차단시키지 않는 다수의 시도가 존재한다(Farries, Harris et al., 1991, Eur. J. Biochem. 196:679-685). 펩타이드의 N-말단을 차단시키지 않는 방법은 당해 분야에 공지되어 있다(참고: 예컨대, Krishna et al., 1991,al. Biochem. 199:45-50; Leone et al., 2011, Curr. Protoc. Protein Sci., Chapter 11:Unit11.7; Fowler et al., 2001, Curr. Protoc. Protein Sci., Chapter 11: Unit 11.7, 이들 각각은 이의 전문이 본원에 참고로 포함된다).

단실 클로라이드는 펩타이드의 유리 아민 그룹과 반응하여 NTAA의 단실 유도체를 생성한다. DNFB 및 SNFB는 펩

타이드의 α-아민 그룹과 반응하여 각각 DNP-NTAA, 및 SNP-NTAA를 생산한다. 또한, DNFB 및 SNFB 둘 다는 라이신 잔기의 ε-아민과 반응한다. DNFB는 또한 타이로신 및 히스티딘 아미노산 잔기와 반응한다. SNFB는 DNFB보다 아민 그룹에 대해 더 우수한 선택성을 가지며 NTAA 변형에 바람직하다(Carty and Hirs 1968). 특정 구현예에서, 라이신 ε-아민은 펩타이드로의 폴리펩타이드 프로테아제 분해 전에 유기 무수물로 예비-차단된다.

다른 유용한 NTAA 개질제는 공지된 효소, 즉, N-말단 아세틸화된 아미노산을 절단하여 단일 아미노산에 의해 펩타이드를 효과적으로 단축시키는 아실 펩타이드 하이드롤라제(APH)가 아세틸화된 NTAA를 제거하기 위해 존재하므로 아세틸 그룹이다{Chang, 2015 #373;Friedmann, 2013 #374}. NTAA는 아세트산 무수물로 화학적으로 아세틸화될 수 있거나 N-말단 아세틸트랜스퍼라제(NAT)로 효소적으로 아세틸화될 수 있다{Chang, 2015 #373; Friedmann, 2013 #374}. 아미드화된 NTAA의 입증된 절단 화학이 문헌에 공지되어 있으므로, 여전히 다른 유용한 NTAA 개질제는 아미디닐(구아니디닐) 모이어티이며, 즉, N-말단 아미드화된 펩타이드를 0.5 내지 2%의 NaOH와 함께 온화하게 항온처리하면 N-말단 아미노산이 절단된다{Hamada, 2016 #383}. 이는 온화한 에드만-유사 화학적 N-말단 분해 펩타이드 서열분석 과정을 제공한다. 더욱이, 특정의 아미드화(구아니딘화) 시약 및 하부 NaOH 절단은 DNA 암호화와 매우 혼용성이다.

NTAA에서 DNP/SNP, 아세틸, 또는 아미디닐(구아니디닐) 그룹의 존재는 가공된 결합제와의 상호작용을 위한 보다 우수한 취급을 제공한다. 다수의 시판되는 DNP 항체는 낮은 nM 친화성으로 존재한다. NTAA를 표지화하는 다른 방법은 트리플리가제(Liebscher et al., 2014,gew Chem Int Ed Engl 53:3024-3028) 및 아미노 아실 트랜스퍼라제(Wagner, et al., 2011, J Am Chem Soc 133:15139-15147)를 사용한 표지화를 포함한다.

이온성 액체의 존재하에서, 이소티오시아네이트는 1차 아민에 대해 향상된 반응성을 가진 것으로 밝혀졌다. 이온성 액체는 유기 화학 반응에서 탁월한 용매(및 촉매로서 제공한다)이며 이소티오시아네이트와 아민의 반응을 향상시켜 티오우레아를 형성할 수 있다. 예는 페닐 이소티오시아네이트(PITC)에 의한 방향족 및 지방족 아민의 신속하고 효율적인 표지화를 위한 이온성 액체 1-부틸-3-메틸-이미다졸륨 테트라플루오로보레이트[Bmim][BF4]의 사용이다(Le, Chen et al. 2005). 에드만 분해는 PITC와 같은 이소티오시아네이트와, 펩타이드의 아민 N-말단의 반응을 포함한다. 따라서, 일 구현예에서 이온성 액체를 사용하여 온화한 표지화 및 분해 조건을 제공함으로써 에드만 분해 공정의 효율을 개선시킨다. 예를 들면, 25℃에서 10분 동안 이온성 액체[Bmim][BF4] 속에서 5%(vol./vol.)의 PITC의 사용은 55℃에서 60분 동안 피리딘, 에탄올 및 ddH2O (1:1:1 vol./vol./vol.)를 함유하는 용액 속에서 5%(vol./vol.) PITC를 사용하는 표준 에드만 PITC 유도체화 조건보다 더 효율적이다(Wang, Fang et al. 2009). 바람직한 구현예에서, 내부 라이신, 타이로신, 히스티딘, 및 시스테인 아미노산은 펩타이드로의 단편화 전에 폴리펩타이드내에서 차단된다. 이러한 방식으로, NTAA의 펩타이드 α-아민 그룹 만이 펩타이드 서열분석 반응 동안 변형되기 쉽다. 이는 특히 DNFB(생거 시약) 및 단실 클로라이드를 사용하는 경우 관련된다.

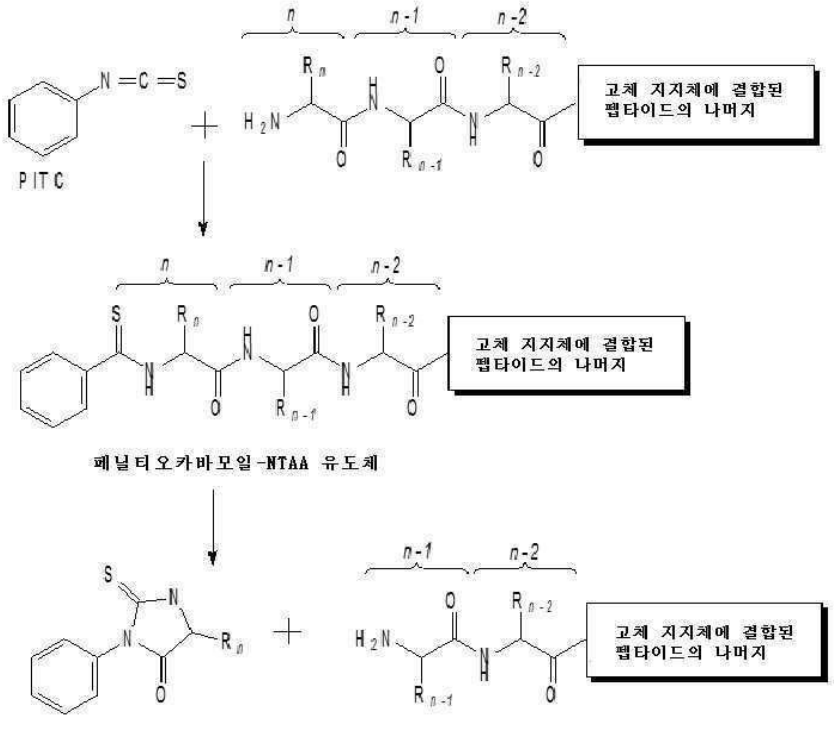
특정 구현예에서, NTAA는 NTAA 표지화 단계(특히 단백질의 원래의 N-말단) 전에 차단된다. 그럴 경우, N-아세틸 블록을 아실 펩타이드 하이드롤라제(APH)로 제거하는 것과 같이, N-말단을 차단하지 않는 다수의 시도가 존재한다(Farries, Harris et al. 1991). 펩타이드의 N-말단을 차단하지 않는 다수의 다른 방법이 당해 분야에 공지되어 있다(참고: 예컨대, Krishna et al., 1991,al. Biochem. 199:45-50; Leone et al., 2011, Curr. Protoc. Proein Sci., Chapter 11:Unit11.7; Fowler et al., 2001, Curr. Protoc. Protein Sci., Chapter 11: Unit 11.7, 이들 각각은 전문이 본원에 참고로 포함된다).

CTAA는 헤르만슨(Hermanson)(Hermanson 2013)에 의해 기술된 바와 같이 다수의 상이한 카복실-반응성 시약으로 변형될 수 있다. 다른 예에서, CTAA는 혼합된 무수물과 이소티오시아네이트로 변형되어 티오하이드라토인을 생성한다((Liu and Liang 2001) 및 미국 특허 제5,049,507호). 티오하이단토인 변형된 펩타이드는 염기 속에서 승온에서 절단되어 끝에서 두번째의 CTAA를 노출시켜, C-말단 기반의 펩타이드 분해 서열분석 시도를 효과적으로 생성한다(Liu and Liang 2001). CTAA에 대해 이룰 수 있는 다른 변형은 파라-니트로아닐리드 그룹의 첨가 및 7-아미노-4-메틸코우마리닐 그룹의 첨가를 포함한다.

XI. 말단 아미노산 절단 방법

펩타이드를 분석하는 것과 관련된 특정 구현예에서, 결합제에 의한 말단 아미노산(N-말단 또는 C-말단)의 결합 및 암호화 태그 정보의 기록 태그로의 이전, 기록 태그 정보의 암호화 태그로의 이전, 기록 태그 정보 및 암호화 태그 정보의 디-태그 작제물로의 이전 후, 말단 아미노산이 제거되거나 펩타이드로부터 절단되어 새로운 말단 아미노산을 노출시킨다. 일부 구현예에서, 말단 아미노산은 NTAA이다. 다른 구현예에서, 말단 아미노산은 CTAA이다.

말단 아미노산의 절단은 화학적 절단 및 효소적 절단을 포함하는, 어떠한 수의 공지된 기술로도 달성할 수 있다. 화학적 절단의 예는 에드만 분해이다. 펩타이드의 에드만 분해 동안에 *n*번째 NTAA가 페닐 이소티오시아네이트(PITC)와 온화한 알칼리성 조건 하에서 반응하여 페닐티오카바모일-NTAA 유도체를 형성한다. 다음에, 산성 조건 하에서, 페닐티오카바모일-NTAA 유도체는 절단되어 유리된 티아졸리논 유도체를 생성하며, 이에 의해 펩타이드의 *n-1*번 아미노산은 N-말단 아미노산(*n-1*번 NTAA)로 전환된다. 이러한 과정의 단계는 하기 나열되어 있다:



상기 기술된 바와 같은, 전형적인 에드만 분해는 장시간 항온처리 동안 엄격한 고온 화학 조건(예컨대, 무수 TFA)의 사용을 필요로 한다. 이러한 조건은 일반적으로 거대분자의 핵산 암호화와 혼용성이 아니다.

화학적 에드만 분해를 핵산 암호화-친화성 시도로 전환시키기 위해, 엄격한 화학적 단계를 온화한 화학적 분해 또는 효율적인 효소 단계로 대체한다. 일 구현예에서, 화학적 에드만 분해는 원래 기술된 것보다 온화한 조건으로 사용할 수 있다. 에드만 분해를 위한 몇가지 온화한 분해 조건이 무수 TFA를 아세토니트릴 중 트리에틸아민 아세테이트로 대체시킴을 포함하여, 문헌에 기술되어 있다(참고: 예컨대, Barrett, 1985, Tetrahedron Lett. 26:4375-4378, 이의 전문은 참고로 포함된다). NTAA의 절단은 또한 티오아세틸화 분해를 사용하여 달성할 수 있으며, 이는 에드만 분해와 비교하여 보다 온화한 절단 조건을 사용한다(참고: 미국 특허 제4,863,870호).

다른 구현예에서, 무수 TFA를 사용한 절단은 PITC-유도체화된 N-말단 아미노산을 온화한 조건하에서 자를수 있는 펩타이드의 카보닐 그룹 상의 티오우레아 황 원자의 친핵성 부착을 통해 PITC-유도체화된 N-말단 아미노산의 제거를 촉매하는 가공된 효소인, "에드마나제"를 사용하여 대체될 수 있다(참고: 미국 특허 공보 제US2014/0273004호, 이의 전문은 참고로 포함됨). 에드마나제는 트리파노소마 크루지(*Trypanosoma cruzi*)로부터의 시스테인 프로테아제인, 크루자인을 변형시켜 제조하였다(Borgo, 2014). C25G 돌연변이는 촉매적 시스테인 잔기를 제거하지만 3개의 돌연변이(G65S, 138C, L160Y)를 선택하여 에드만 시약(PITC)의 페닐 모이어티를 사용한 입체적 조화(steric fit)를 생성하였다.

NTAA의 효소적 절단은 또한 아미노펩티다제에 의해 달성될 수 있다. 아미노펩티다제는 천연적으로, 단량체성 및 다량체성 효소로 존재하며, 금속 또는 ATP-의존성일 수 있다. 천연의 아미노펩티다제는 매우 제한된 특이성을 가지므로 일반적으로 진행적인 방식으로 N-말단 아미노산을 절단하여, 다른 것 이후 하나의 아미노산을 절단한다. 여기에 기술된 방법의 경우, 아미노펩티다제를 가공하여 N-말단 표지로 변형시키는 경우에만 NTAA에 대한 특이적인 결합 또는 촉매 활성을 지니도록 가공할 수 있다. 예를 들면, 아미노펩티다제를 가공하여 이것이 DNP/SNP, PTC, 단실 클로라이드, 아세틸, 아미디닐 등과 같은 그룹으로 변형되는 경우 이것이 N-말단 아미노산만을 절단하도록 가공될 수 있다. 이러한 방식으로, 아미노펩티다제는 N-말단으로부터 처음으로 단일 아미노산을 절단하며, 분해 주기의 제어를 허용한다. 일부 구현예에서, 변형된 아미노펩티다제는 아미노산 잔기 동일성

에 대해 비-선택성이지만 N-말단 표지에 대해서는 선택성이다. 다른 구현예에서, 변형된 아미노펩티다제는 아미노산 잔기 동일성 및 N-말단 표지 둘 다에 대해 선택성이다. 효소적 NTAA 분해의 특이성을 개질시키는 모델의 예는 보르고(Borgo) 및 하브라넥(Havranek)에 의해 묘사되어 있는데, 여기서 구조-기능 보도된 설계를 통해, 메티오닌 아미노펩티다제는 루이신 아미노펩티다제로 전환된다(Borgo and Havranek 2014). 유사한 시도를 DNP/SNP-변형된 NTAA와 같은 변형된 NTAA를 사용하여 취할 수 있으며, 여기서 아미노펩티다제가 가공되어(구조-기능 기반의 설계 및 지시된 발전 둘 다를 사용) DNP/SNP 그룹이 존재하는 N-말단 아미노산 만을 절단한다. 표지된(바이오티닐화된) NTAA의 개개 또는 소 그룹에 결합하여 이를 절단하는 가공된 아미노펩티다제 돌연변이체는 기술되어 있다(참고: PCT 공보 제WO2010/065322호).

특정 구현예에서, 콤팩트(compact) 단량체성 금속효소 아미노펩티다제를 가공하여 DNP-표지된 NTAA를 인식하고 절단한다. 단량체성 금속-아미노펩티다제의 사용은 2가지 중요한 장점을 갖는다: 1) 콤팩트 단량체성 단백질은 파아지 디스플레이를 사용하여 나타내고 스크리닝하기 훨씬 용이하다; 2) 금속-아미노펩티다제는 이의 활성이 적절한 금속 양이온의 첨가 또는 제거에 의해 의도대로 작동/비작동될 수 있다는 점에서 유일한 장점을 갖는다. 예시적인 아미노펩티다제는 스트렙토마이세스 아종(Streptomyces sp.) KK506(SKAP)(Yoo,hn et al. 2010), 스트렙토마이세스 그리세우스(Streptomyces griseus)(SGAP), 비브리오 프로테올라이티쿠스(Vibrio proteolyticus)(VPAP)와 같은 M28 계열의 아미노펩티다제를 포함한다(Spungin and Blumberg 1989, Ben-Meir, Spungin et al. 1993). 이들 효소는 안정하고, 강력하며, 실온 및 pH 8.0에서 활성이고, 펩타이드 분석을 위해 바람직한 온화한 조건과 혼용성이다.

다른 구현예에서, 사이클릭 절단은 아미노펩티다제를 N-말단 아미노산 표지의 존재하에서만 활성이 되도록 가공함으로써 획득된다. 더욱이, 아미노펩티다제는 비-특이적이 되도록 가공함으로써, 이것이 다른 것보다 하나의 특수한 아미노산을 선택적으로 인식하는 것이 아니라, 표지된 N-말단을 인식하도록 할 수 있다. 바람직한 구현예에서, 금속펩티다제 단량체성 아미노펩티다제(예컨대, 비브로(*Vibro*) 루이신 아미노펩티다제)(Hernandez-Moreno, Villasenor et al. 2014)를 가공하여 변형된 NTAA(예컨대, PTC, DNP, SNP, 아세틸화된, 아실화된, 등)만을 절단한다.

여전히 다른 구현예에서, 사이클릭 절단은 가공된 아실펩티다제 하이드롤라제(APH)를 사용하여 달성하여 아세틸화된 NTAA를 절단한다. APH는 차단된 펩타이드로부터 Nα-아세틸화된 아미노산의 제거를 촉매할 수 있는 세린 프로테아제이며, 진핵 세포, 세균 세포 및 고세균 세포에서 N-말단적으로 아세틸화된 단백질의 중요한 조절인자이다. 특정 구현예에서, APH는 이량체성이고 단지 엑소펩티다제 활성을 갖는다(Gogliettino, Balestrieri et al. 2012, Gogliettino, Riccio et al. 2014). 가공된 APH는 보다 높은 친화성을 가지며 내인성 또는 야생형 APH보다 거의 선택성이 아니다.

여전히 다른 구현예에서, NTAA의 아미드화(구아니디닐화)를 사용하여 NaOH를 사용한 표지된 NTAA의 온화한 절단을 가능하도록 한다(Hamada, 2016, 이의 전문은 본원에 참고로 포함됨). 다음을 포함하는 다수의 아미드화(구아니디닐화) 시약이 당해 분야에 공지되어 있다: S-메틸이소티오우레아, 3,5-디메틸피라졸-1-카복스아미딘, S-에틸티오우로늄 브로마이드, S-에틸티오우로늄 클로라이드, O-메틸이소우레아, O-메틸이소우로늄 설페이트, O-메틸이소우레아 하이드로겐 설페이트, 2-메틸-1-니트로이소우레아, 아미노이미노메탄설폰산, 시안아미드, 시아노구아니드, 디시아노디아미드, 3,5-디메틸-1-구아닐피라졸 니트레이트 및 3,5-디메틸 피라졸, N,N'-비스(오르토-클로로-Cbz)-S-메틸이소티오우레아 및 N,N'-비스(오르토-브로모-Cbz)-S-메틸이소티오우레아(Katritzky, 2005, 이의 전문은 참고로 포함됨).

NTAA 표지화, 결합, 및 분해 작업흐름의 예는 다음과 같다(참고: 도 41 및 42): 단백질분해성 소화물로부터 기록 태그 표지된 펩타이드(예컨대, 5천만 내지 10억)를 단일의 분자 서열분석 기질(예컨대, 다공성 비드) 상에서 적절한 분자간 간격으로 무작위로 고정시킨다. 사이클릭 방식으로 각각의 펩타이드의 N-말단 아미노산(NTAA)을 작은 화학적 모이어티(예컨대, DNP, SNP, 아세틸)로 변형시켜 NTAA 분해 공정의 사이클릭 제어를 제공하고 동종의 결합제에 의해 결합 친화성을 향상시킨다. 각각의 고정된 펩타이드의 변형된 N-말단 아미노산(예컨대, DNP-NTAA, SNP-NTAA, 아세틸-NTAA)을 동종의 NTAA 결합제로 결합시키고, 결합된 NTAA 결합과 관련된 암호화 태그로부터의 정보를 고정된 펩타이드와 관련된 기록 태그에 이전시킨다. NTAA 인식, 결합, 및 암호화 태그 정보의 기록 태그로의 이전 후, 표지된 NTAA를 표지의 존재하에서만 NTAA 절단할 수 있는, 가공된 아미노펩티다제(예컨대, DNP-NTAA 또는 SNP-NTAA의 경우) 또는 가공된 APH(예컨대, 아세틸-NTAA의 경우)에 대해 노출시켜 제거한다. 다른 NTAA 표지(예컨대, PITC)를 또한 적합하게 가공된 아미노펩티다제와 함께 사용할 수 있다. 특수한 구현예에서, 단일의 가공된 아미노펩티다제 또는 APH는 N-말단 아미노산 표지를 지닌 모든 가능한 NTAA(해독 후 변형 변이체 포함)를 공통적으로 절단한다. 다른 특수한 구현예에서, 2, 3, 4개 이상의 가공된 아미노펩티다제

또는 APH를 사용하여 표지된 NTAA의 레퍼토리를 절단한다.

DNP 또는 SNP 표지된 NTAA에 대해 활성을 지닌 아미노펩티다제를 아포(apo)-효소 상(금속 보조인자의 부재하에서 불활성임)에서 강력한-결합 선택을 조합하는 스크린에 이어서 벤질페니실린에 대한 금속-베타-락타마제 효소를 가공하는데 있어서 폰사드(Ponsard) 등이 기술한 시도와 같은(Ponsard, Galleni et al. 2001, Fernandez-Gacio, Uguen et al. 2003) 기능성 효소 선택 단계에 의해 선택한다. 이러한 2-단계 선택은 Zn^{2+} 이온의 첨가에 의해 활성화된 금속-AP를 사용함을 포함한다. 고정된 펩타이드 기질에 대한 강한 결합 선택 후에, Zn^{2+}를 도입하고 DNP 또는 SNP로 표지된 NTAA를 가수분해할 수 있는 촉매적으로 활성인 파아지는 결합된 파아지의 상층액내로의 방출을 이끈다. 반복된 선택 라운드를 수행하여 DNP 또는 SNP 표지된 NTAA 절단을 위한 활성 AP를 농축시킨다.

본원에 제공된 어떠한 구현예에서, NTAA 절단 시약을 NTAA에 보충하는 것은 키메라 절단 효소 및 키메라 NTAA 변형제를 통해 향상시킬 수 있으며, 여기서 키메라 절단 효소 및 키메라 NTAA 변형제 각각은 서로 강력하게 결합반응할 수 있는 모이어티(예컨대, 바이오틴-스트렙타비딘)를 포함한다(참고: 도 39). 예를 들면, NTAA는 바이오틴-PITC로 변형시킬 수 있으며, 키메라 절단 효소(스트렙타비딘-에드마나제)는 스트렙타비딘-바이오틴 상호작용을 통해 변형된 NTAA로 보충되어, 절단 효소의 친화성 및 효능을 증진시킨다. 변형된 NTAA는 절단되고 관련된 절단 효소와 함께 펩타이드로부터 확산 제거된다. 키메라 에드마나제의 예에서, 이러한 시도는 μM 내지 준-피코몰의 친화성 K_D를 효과적으로 증가시킨다. 유사한 절단 향상이 또한 기록 태그와 상호작용하는 절단제에서 DNA 태그를 사용하여 테써링(tethering)함으로써 실현될 수 있다(도 44).

NTAA 절단에 대한 대안으로서, 디펩티딜 아미노 펩티다제(DAP)를 사용하여 펩타이드로부터 적어도 2개의 N-말단 아미노산을 절단할 수 있다. 특정 구현예에서, 단일의 NTAA는 절단될 수 있다(참고: 도 45): 도 45는 N-말단 분해에 대한 시도를 나타내며, 여기서 부텔라제 I 펩타이드 기질의 N-말단 연결은 TEV 엔도펩티다제 기질을 펩타이드의 N-말단에 부착시킨다. 부착 후, TEV 엔도펩티다제는 질의의 펩타이드(서열분석을 겪는 펩타이드)로부터 새로이 연결된 펩타이드를 절단하여 NTAA에 부착된 단일의 아스파라긴(N)을 남긴다. N-말단으로부터 2개의 아미노산을 절단하는, DAP와의 항온처리는 원래의 NTAA의 전체적인 제거를 야기한다. 이러한 전체 과정은 N-말단 분해 과정에서 주기화될 수 있다.

CTAA 결합제에 관한 구현예의 경우, 펩타이드로부터 CTAA를 절단하는 방법은 당해 분야에 또한 공지되어 있다. 예를 들면, 미국 특허 제6,046,053호는 펩타이드 또는 단백질을 알킬 산 무수물과 반응시켜 카복시-말단을 옥사졸론으로 전환시켜, 산 및 알코올과 에스테르의 반응에 의해 C-말단 아미노산을 유리시킴을 개시하고 있다. CTAA의 효소적 절단은 또한 카복시펩티다제에 의해 달성될 수 있다. 수개의 카복시펩티다제는 아미노산 선호도를 나타내는데, 예컨대, 카복시펩티다제 B는 아르기닌 및 라이신과 같은 염기성 아미노산을 우선적으로 절단한다. 위에서 기술한 바와 같이, 카복시펩티다제는 또한 아미노펩티다제와 같은 양식으로 변형되어 C-말단 표지를 갖는 CTAA에 특이적으로 결합하는 카복시펩티다제를 가공할 수 있다. 이러한 방식으로, 카복시펩티다제는 C-말단으로부터 한번에 단일 아미노산 만을 절단하여, 분해 주기가 제어되도록 한다. 일부 구현예에서, 변형된 카복시펩티다제는 아미노산 잔기 동일성에 대해 비-선택성이지만 C-말단 표지에 대해서는 선택성이다. 다른 구현예에서, 변형된 카복시펩티다제는 아미노산 잔기 동일성 및 C-말단 표지 둘 다에 대해 선택적이다.

XII. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 프로세싱 및 분석

연장된 기록 태그, 연장된 암호화 태그, 및 목적한 거대분자(들)을 나타내는 디-태그 라이브러리를 프로세싱하고 다양한 핵산 서열분석 방법을 사용하여 분석할 수 있다. 서열분석 방법의 예는 쇄 말단 서열분석(생거 서열분석); 합성에 의한 서열분석, 연결에 의한 서열분석, 하이브리드화에 의한 서열분석, 폴로니 서열분석, 이온 반도체 서열분석, 또는 피로서열분석과 같은 차 세대 서열분석 방법; 및 단일 분자 실시간 서열분석, 나노공극-기반 서열분석, 듀플렉스 차단된 서열분석, 및 진전된 현미경을 사용한 DNA의 직접적인 영상화와 같은, 차세대 서열분석 방법을 포함하나, 이에 제한되지 않는다.

연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 다양한 방식으로 증폭시킬 수 있다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 예컨대, PCR 또는 유화액 PCR을 통해 대수적 증폭을 겪을 수 있다. 유화액 PCR은 보다 균일한 증폭을 생산하는 것으로 알려져 있다(Hori, Fukano et al. 2007). 대안적으로, 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 예컨대, T7 RNA 폴리머라제를 사용한 주형 DNA의 시험관내(*in vitro*) 전사를 통한 선형 증폭을 겪을 수 있다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 이에 함유된 보편적인 전방 프라이밍 부위 및 보편적인 역 프

라이밍 부위와 혼화성인 프라이머를 사용하여 증폭시킬 수 있다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 또한 테일화된 프라이머를 사용하여 증폭시켜 서열을 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 5'-말단, 3'-말단 또는 말단들 둘 다에 가할 수 있다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 말단에 가할 수 있는 서열은 라이브러리 특이적인 지수 서열을 포함함으로써 단일의 서열분석 실행, 어댑터 서열, 판독물 프라이머 서열, 또는 서열분석 플랫폼을 위해 상용성인 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리를 제조하기 위한 어떠한 다른 서열내에서 다수의 라이브러리의 멀티플렉싱을 허용한다. 차세대 서열분석을 위한 제조시 라이브러리 증폭의 예는 다음과 같다: 20 μl의 PCR 반응 용적을 ~1 mg의 비드(~ 10 ng), 200 uM dNTP, 1 μM의 각각의 전방 및 역방 증폭 프라이머, 0.5 μl(1U)의 Phusion Hot Start 효소(New England Biolabs)로부터 용출된 연장된 기록 태그 라이브러리를 사용하여 설정하고 다음의 사이클링 조건에 적용한다: 98℃에서 30초에 이은 20 주기의 98℃에서 10초, 60℃에서 30초, 72℃에서 30초에 이어, 72℃에서 7분에 이어, 4℃에서 유지.

특정 구현예에서, 증폭 전, 동안 또는 후에, 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 표적 농축을 겪을 수 있다. 표적 농축을 사용하여 서열분석 전에 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리로부터 목적한 거대분자를 나타내는 연장된 기록 태그를 선택적으로 포획하거나 증폭시킬 수 있다. 단백질 서열을 위한 표적 농축은 표적 단백질에 대해 고도로-특이적인 결합제를 생산하는데 있어서 고 비용 및 곤란성으로 인하여 도전 중에 있다. 항체는 매우 비-특이적이고 수천개의 단백질에 걸쳐 생산을 확대시키기가 힘들다. 본 개시내용의 방법은 단백질 코드를 DNA 라이브러리에 이용가능한 광범위한 표적화된 DNA 농축 전략을 사용할 수 있도록 할 수 있는 핵산 코드로 전환시킴으로써 이러한 문제를 극복한다. 목적한 펩타이드는 이들의 상응하는 연장된 기록 태그를 농축시킴으로써 샘플 속에 농축시킬 수 있다. 표적화된 농축 방법은 당해 분야에 공지되어 있으며 하이브리드 포획 검정, TruSeq custom Amplicon(Illumina), 패드록(padlock) 프로브(또한 분자 전환 프로브로서 지칭됨) 등과 같은 PCR-기반 검정을 포함한다(참고: Mamanova et al., 2010, Nature Methods 7: 111-118; Bodi et al., J. Biomol. Tech. 2013, 24:73-86; Ballester et al., 2016, Expert Review of Molecular Diagnostics 357-372; Mertes et al., 2011, Brief Funct. Genomics 10:374-386; Nilsson et al., 1994, Science 265:2085-8; 이들 각각은 이들의 전문이 참고로 본원에 포함된다).

일 구현예에서, 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 하이브리드 포획-기반한 검정을 통해 농축된다(참고: 예컨대, 도 17의 A 및 도 17의 B). 하이브리드-포획 기반한 검정에서, 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 표적-특이적인 올리고뉴클레오타이드 또는 친화성 태그(예컨대, 바이오틴)로 표지된 "미끼 올리고뉴클레오타이드"에 하이브리드화된다. 표적-특이적인 올리고뉴클레오타이드에 하이브리드화된 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그는 친화성 리간드(예컨대, 스트렙타비딘 코팅된 비드)를 사용하여 이들의 친화성 태그를 통해 "떼어지고(pull down)", 배경(비-특이적인) 연장된 기록 태그는 세척 제거된다(참고: 예컨대, 도 17). 농축된 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그는 이후 양성 농축을 위해 수득된다(예컨대, 비드로부터 용출된다).

"반응계내(in situ)" 올리고뉴클레오타이드 합성 및 올리고뉴클레오타이드 혼주물의 후속적인 증폭에 의해 합성된 미끼 올리고뉴클레오타이드의 경우, 경쟁하는 미끼는 제공된 올리고뉴클레오타이드 배열내에서 보편적인 프라이머의 수개 세트를 사용함으로써 혼주물내로 가공할 수 있다. 각각의 유형의 보편적인 프라이머의 경우, 바이오티닐화된 프라이머 대 비-바이오티닐화된 프라이머의 비는 농축 비를 제어한다. 몇가지 프라이머 유형의 사용은 몇가지 농축 비가 최종의 올리고뉴클레오타이드 미끼 혼주물로 설계되도록 할 수 있다.

미끼 올리고뉴클레오타이드는 목적한 거대분자를 나타내는 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그에 대해 상보성이도록 설계할 수 있다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 내에서 스페이서 서열에 대한 미끼 올리고뉴클레오타이드의 상보성 정도는 0% 내지 100%, 및 이들 사이의 어떠한 정수일 수 있다. 이러한 매개변수는 약간의 농축 실험에 의해 용이하게 최적화할 수 있다. 일부 구현예에서, 암호기 서열에 대한 스페이서 길이는 암호화 태그 설계에서 최소화되거나 스페이서는 이들이 미끼 서열에 대한 하이브리드화에 대해 이용불가능하도록 설계된다. 하나의 시도는 보조인자의 존재하에서 2차 구조를 형성하는 스페이서를 사용하는 것이다. 이러한 2차 구조의 예는 G-쿼드러플렉스(quadruplex)이며, 이는 서로의 상단에 적층된(stacked)된 2개 이상의 구아닌 쿼테트(quartet)에 의해 형성된 구조이다(Bochman, Paeschke et al. 2012). 구아닌 쿼테트는 후그스텐 수소 결합(Hoogsteen hydrogen bonding)을 통해 연합된 4개의 구아닌 염기에 의해 형성된 사각형 평면 구조이다. G-쿼드러플렉스 구조는 양이온, 예컨대, K+ 이온 대 Li+ 이온의 존재하에서 안정화된다.

사용된 미끼 올리고뉴클레오타이드의 수를 최소화하기 위하여, 각각의 단백질로부터 비교적 유일한 펩타이드의 세트를 생물정보적으로 확인할 수 있으며 목적한 펩타이드의 상응하는 연장된 기록 태그 라이브러리 표시에 대해 상보성인 이들 미끼 올리고뉴클레오타이드 만을 하이브리드 포획 검정에 사용한다. 순차적인 라운드 또는 농축을 또한 동일하거나 상이한 미끼 세트를 사용하여 수행할 수 있다.

이의 단편(예컨대, 펩타이드)을 나타내는 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리내 전체 길이의 거대분자(예컨대, 단백질 또는 폴리펩타이드)를 농축시키기 위해, "타일된(tiled)" 미끼 올리고뉴클레오타이드를 단백질의 전체 핵산 표시를 따라 설계할 수 있다.

다른 구현예에서, 프라이머 연장 및 연결-기반 매개된 증폭 농축(AmpliSeq, PCR, TruSeq TSCA 등)을 사용하여 거대분자의 소세트를 나타내는 라이브러리 요소의 농축된 분획을 선택하고 모듈화할 수 있다. 경쟁하는 올리고를 또한 사용하여 프라이머 연장, 연결, 또는 증폭의 정도를 조율할 수 있다. 가장 단순한 시행에서, 이는 보편적인 프라이머 테일 및 5' 보편적인 프라이머 테일을 결여한 경쟁하는 프라이머를 포함하는 표적 특이적인 프라이머의 혼합물을 가짐으로써 달성할 수 있다. 초기 프라이머 연장 후, 5' 보편적인 프라이머 서열을 지닌 프라이머 만을 증폭할 수 있다. 보편적인 프라이머 서열을 지니거나 지니지 않은 프라이머는 증폭된 표적의 분획을 제어한다. 다른 구현예에서, 하이브리드화되지만 연장되지 않는 프라이머의 혼입을 사용하여 프라이머 연장, 연결, 또는 증폭을 겪는 라이브러리 요소의 분획을 조절할 수 있다.

표적화된 농축 방법을 또한 음성 선택 방식으로 사용하여 서열분석 전에 라이브러리로부터 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그를 선택적으로 제거할 수 있다. 따라서, 바이오티닐화된 미끼 올리고뉴클레오타이드 및 스트렙타비딘 코팅된 비드를 사용한 상술한 실시예에서, 상층액은 서열분석을 위해 보유되지만 비드에 결합된 미끼-올리고뉴클레오타이드:연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 하이브리드는 분석되지 않는다. 제거될 수 있는 바람직하지 않은 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 예는 예컨대, 단백질, 알부민, 면역글로불린 등에 대한 풍부한 거대분자 종에 걸쳐 나타나는 것들이다.

표적에 하이브리드화하지만 바이오틴 모이어티를 결여하는 경쟁인자 올리고뉴클레오타이드 미끼를 또한 하이브리드 포획 단계에서 사용하여 농축된 어떠한 특수한 유전자자리(locus)의 분획을 조절할 수 있다. 경쟁인자 올리고뉴클레오타이드 미끼는 표적에 대한 하이브리드화에 대해 농축 동안 당겨진 표적의 분획을 효과적으로 조절하는 표준 바이오티닐화된 미끼와 경쟁한다(도 17). 10 차수의 역학적 범위의 단백질 발현은 특히 알부민과 같은 매우 풍부한 종에 대해, 이러한 경쟁적 억제 시도를 사용하여 수개의 차수까지 압축할 수 있다. 따라서, 표준 하이브리드 포획에 대한 제공된 유전자자리에 대해 포획된 라이브러리 요소의 분획을 100%로부터 0% 농축까지 조절할 수 있다.

또한, 라이브러리 표준화 기술을 사용하여 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 라이브러리로부터 매우 풍부한 종을 제거할 수 있다. 이러한 시도는 트립신, LysC, GluC 등과 같은 부위-특이적인 프로테아제 분해에 의해 생성된 펩타이드로부터 기원하는 정의된 길이의 라이브러리에 대해 최고로 작업한다. 하나의 예에서, 표준화는 이중-가닥 라이브러리를 변성시키고 라이브러리 요소를 재-어닐링되도록 함으로써 달성할 수 있다. 풍부한 라이브러리 요소는 이중 분자 하이브리드화 동역학(bimolecular hybridization kinetics)의 2차 속도 상수(second-order rate constant)로 인하여 거의 풍부하지 않은 요소보다 더 신속하게 재-어닐링된다(Bochman, Paeschke et al. 2012). ssDNA 라이브러리 요소는 하이드록시아파타이트 컬럼(hydroxyapatite column) 상에서 크로마토그래피와 같은, 당해 분야에 공지된 방법(VanderNoot, et al., 2012, Biotechniques 53:373-380) 또는 라이브러리를 dsDNA 라이브러리를 파괴하는 캄챠카 크랩(Kamchatka crab)으로부터의 듀플렉스-특이적인 뉴클레아제(DSN)의 처리(Shagin et al., 2002, Genome Res. 12:1935-42)를 사용하여 풍부한 dsDNA 라이브러리 요소로부터 분리할 수 있다

고체 지지체에 대한 및/또는 수득되는 연장된 기록 태그 라이브러리의 부착 전에 거대 분자의 분획화, 농축, 및 공제 방법의 어떠한 조합도 서열분석 판독물을 절약하고 낮은 풍부성 종의 측정을 증진시킬 수 있다.

일부 구현예에서, 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그의 라이브러리는 연결 또는 말단-상보성 PCR에 의해 콘카테네이트화하여 각각 다수의 상이한 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그를 포함하는 긴 DNA 분자를 생성할 수 있다(Du et al., 2003, BioTechniques 35:66-72; Muecke et al., 2008, Structure 16:837-841; 미국 특허 제5,834,252호, 이들 각각은 이의 전문이 참고로 포함된다). 이러한 구현예는 나노공극 서열분석에 바람직하며, 여기서 DNA의 긴 가닥은 나노공극 서열분석 장치에 의해 분석된다.

일부 구현예에서, 직접적인 단일 분자 분석은 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그(참고: 예컨

대, Harris et al., 2008, Science 320:106-109)에서 수행된다. 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그는 유동 셀 또는 유동 셀 표면(임의로 패턴화된 마이크로셀) 위에 로딩하기 위해 혼용성인 비드와 같은, 고체 지지체 상에서 직접 분석될 수 있으며, 여기서 유동 셀 또는 비드는 단일 분자 서열기(sequencer) 또는 단일 분자 탈암호화 장치로 통합될 수 있다. 단일 분자 탈암호화의 경우, 탈암호화 올리고뉴클레오타이드의 수회 라운드의 혼주된 형광성-표지된 하이브리드화(Gunderson et al., 2004, Genome Res. 14:970-7)를 사용하여 연장된 기록 태그내 암호화 태그의 동일성 및 순서 둘 다를 확인할 수 있다. 암호화 태그의 결합 순서를 데컨볼루트하기 위해, 결합제는 상술한 바와 같이 주기-특이적인 암호화 태그로 표지할 수 있다(참고: 또한 Gunderson et al., 2004, Genome Res. 14:970-7). 주기-특이적인 암호화 태그는 단일의 거대분자를 나타내는 단일의, 콘카테네이트화된 연장된 기록 태그, 또는 단일의 거대분자를 나타내는 연장된 기록 태그의 수집 둘 다를 위해 작업할 것이다.

연장된 리포터 태그, 연장된 암호화 태그, 또는 디-태그 라이브러리의 서열분석에 이어서, 수득되는 서열은 이들의 UMI에 의해 붕괴될 수 있으며, 이후 이들의 상응하는 거대분자(예컨대, 펩타이드, 단백질, 단백질 복합체)와 관련되어 세포내에서 전체 거대분자 유형(예컨대, 펩타이드, 폴리펩타이드, 단백질 거대분자를 위한 프로테오믹스)에 대해 정렬될 수 있다. 수득되는 서열은 또한 이들의 구획에 의해 붕괴될 수 있으며 이들의 상응하는 구획 프로테옴과 관련될 수 있으며, 이는, 특수한 구현예에서, 단일 또는 매우 제한된 수의 단백질 분자를 함유한다. 단백질 확인 및 정량화 둘 다는 이러한 디지탈 펩타이드 정보로부터 용이하게 유도시킬 수 있다.

일부 구현예에서, 암호화 태그 서열은 특수한 서열분석 분석 플랫폼에 대해 최적화할 수 있다. 특수한 구현예에서, 서열분석 플랫폼은 나노공극 서열분석이다. 일부 구현예에서, 서열분석 플랫폼은 염기당 오류율이 > 5%, > 10%, >15%, > 20%, > 25%, 또는 > 30%이다. 예를 들면, 연장된 기록 태그가 나노공극 서열분석 장치를 사용하여 분석되어야 하는 경우, 바코드 서열(예컨대, 암호기 서열)을 설계하여 나노공극의 통과시 최적으로 전기적 구별 가능하도록 할 수 있다. 본원에 기술된 방법에 따른 펩타이드 서열분석은, 나노공극 서열분석에 대한 단일 염기 정확도가 오히려 여전히 낮지만(75%-85%), "암호기 서열"의 측정이 훨씬 보다 정밀하여야 한다(>99%)는 것을 고려할 때, 나노공극 서열분석에 매우 적합할 수 있다. 더욱이, 듀플렉스 차단된 나노공극 서열분석(DI)으로 불리는 기술은 분자 모터(molecular motor)에 대한 요구없이 나노공극 가닥 서열분석과 함께 사용되어, 시스템 설계를 크게 단순화시킬 수 있다(Derrington, Butler et al. 2010). DI 나노공극 서열분석을 통한 연장된 기록 태그의 판독물은 상보성 올리고뉴클레오타이드로 어닐링될 콘카테네이트화된 연장된 기록 라이브러리에서 스페이서 요소를 필요로 한다. 본원에 사용된 올리고뉴클레오타이드는 LNA, 또는 다른 변형된 핵산 또는 유사체를 포함함으로써 수득되는 듀플렉스의 효과적인 Tm을 상승시킬 수 있다. 이들 듀플렉스 스페이서 영역으로 장식된 단일-가닥의 연장된 기록 태그가 공극을 통과함에 따라, 이중 가닥 영역은 듀플렉스 영역에 인접하여 약 3개의 염기의 전류 판독을 가능하도록 하는 제한 구역에서 일시적으로 지연되기 시작할 것이다. DI 나노공극 서열분석에 대한 특수한 구현예에서, 암호기 서열은 스페이서 요소에 인접한 3개의 염기가 최대의 전기적으로 구별가능한 나노공극 시그널을 생성하도록 하는 방식으로 설계된다(Derrington et al., 2010, Proc. Natl. Acad. Sci. USA 107:16060-5). 모터가 없는 DI 서열분석에 대한 대안으로서, 스페이서 요소는 G-쿼테트와 같은 2차 구조를 채택하도록 설계될 수 있으며, 이는 이것이 인접한 암호기 서열의 판독을 가능하도록 하는 나노공극을 통과하면서 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그를 일시적으로 지연시킬 것이다(Shim, Tan et al. 2009, Zhang, Zhang et al. 2016). 엔진정지(stall)를 지나 진행된 후, 다음의 스페이서는 다시 일시적인 정지를 생성함으로써, 다음 암호기 서열의 판독 등을 가능하게 할 것이다.

본원에 개시된 방법은 다수의 거대분자(예컨대, 펩타이드)의 검출, 정량화 및/또는 서열분석을 포함하는 분석에 동시에(멀티플렉싱) 사용할 수 있다. 본원에 사용된 바와 같은 멀티플렉싱은 동일한 검정에서 다수의 거대분자의 분석을 지칭한다. 다수의 거대분자는 동일한 샘플 또는 상이한 샘플로부터 기원할 수 있다. 다수의 거대분자는 동일한 대상체 또는 상이한 대상체로부터 기원할 수 있다. 분석된 다수의 거대분자는 상이한 샘플로부터 기원한 상이한 거대분자(예컨대, 펩타이드), 또는 동일한 거대분자(예컨대, 펩타이드)일 수 있다. 다수의 거대분자는 2개 이상의 거대분자, 5개 이상의 거대분자, 10개 이상의 거대분자, 50개 이상의 거대분자, 100개 이상의 거대분자, 500개 이상의 거대분자, 1000개 이상의 거대분자, 5,000개 이상의 거대분자, 10,000개 이상의 거대분자, 50,000개 이상의 거대분자, 100,000개 이상의 거대분자, 500,000개 이상의 거대분자, 또는 1,000,000개 이상의 거대분자를 포함한다.

샘플 멀티플렉싱은 기록 태그 표지된 거대분자 샘플의 선행(upfront) 바코딩에 의해 달성될 수 있다. 각각의 바코드는 상이한 샘플을 나타내며, 샘플은 사이클릭 결합 검정 또는 서열 분석 전에 혼주될 수 있다. 이러한 방식으로, 많은 바코드-표지된 샘플이 단일 튜브 속에서 동시에 프로세싱될 수 있다. 이러한 시도는 역상 단백질 배

열(RPPA) 상에서 수행된 면역검정에 있어 유의적인 개선이다(Akbani, Becker et al. 2014, Creighton and Huang 2015, Nishizuka and Mills 2016). 이러한 방식으로, 본 개시내용은 필수적으로 단순한 작업흐름으로 RPPA 검정에 대한 고도의 디지탈 샘플 및 분석물 멀티플렉스된 대안을 제공한다.

XIII. NTAA 인식, 기록 태그 연장, 및 NTAA 절단의 사이클릭 라운드를 통한 거대분자 특성화

특정 구현예에서, 본 개시내용에서 제공된 거대분자를 분석하는 방법은 다수의 결합 주기를 포함하며, 여기서 거대분자는 다수의 결합제와 접촉하고, 결합제의 연속적인 결합은 핵산 기반의 암호화 태그 형태의 역사적인 결합 정보를 거대분자와 관련된 적어도 하나의 기록 태그로 이전한다. 이러한 방식으로 다수의 결합 사건에 대한 정보를 함유하는 역사적인 기록이 핵산 양식으로 생성된다.

N-말단 분해 기반 시도를 사용한 펩타이드 거대분자를 분석하는 방법에 관한 구현예에서(참고: 도 3, 도 4, 도 41, 및 도 42), 제1의 결합제의 *n*번 아미노산의 펩타이드의 *n*번 NTAA에 대한 접촉, 제1의 결합제의 암호화 태그 정보의 펩타이드와 관련된 기록 태그로의 이전, 이에 의한 제1의 순서의 연장된 기록 태그의 형성 후, *n*번 NTAA가 본원에 기술된 바와 같이 절단된다. *n*번 NTAA의 절단은 펩타이드의 *n-1*번 아미노산을 N-말단 아미노산으로 전환시키며, 이는 본원에서 *n-1*번 NTAA로 지칭된다. 본원에 기술된 바와 같이, *n*번 NTAA는 임의로 모이어티(예컨대, PTC, DNP, SNP, 아세틸, 아미디닐 등)으로 표지될 수 있으며, 이는 특히 NTAA의 표지된 형태에 결합하도록 가공된 절단 효소와 접합시키는데 유용하다. *n*번 NTAA가 표지된 경우, 이후 *n-1*번 NTAA가 동일한 모이어티로 표지된다. 제2의 결합제는 펩타이드와 접촉하여 *n-1*번 NTAA에 결합하며, 제2의 결합제의 암호화 태그 정보는 제1의 순서의 연장된 기록 태그로 이전함으로써 제2의 순서의 연장된 기록 태그(예컨대, 펩타이드를 나타내는 콘카테네이트화된 *n*번 순서의 연장된 기록 태그를 생성하기 위해), 또는 상이한 기록 태그(예컨대, 총괄적으로 펩타이드를 나타내는, 다수의 연장된 기록 태그를 생성하기 위해)를 생성한다. *n-1*번 NTAA의 절단은 펩타이드의 *n-2*번 아미노산을 N-말단 아미노산으로 전환시키며, 이는 본원에서 *n-2*번 NTAA로 지칭된다. 추가의 결합, 이전, 절단, 및 임의로 NTAA 표지화는 상술한 바와 같이 *n*번 아미노산까지 일어나서 *n*번 순서의 연장된 기록 태그 또는 *n*개의 연장된 기록 태그를 생성할 수 있으며, 이는 총괄적으로 펩타이드를 나타낸다. 본원에 사용된 바와 같은, 결합제, 암호화 태그, 또는 연장된 기록 태그를 지칭하는데 사용되는 경우 *n* "순서"는 *n*번 결합 주기(여기서 결합제 및 이의 관련된 암호화 태그가 사용된다) 또는 *n*번 결합 주기(여기서 연장된 기록 태그가 생성된다)를 지칭한다.

일부 구현예에서, 거대 분자, 및 임의로 어떠한 추가의 결합제(예컨대, 3번째 결합제, 4번째 결합제, 5번째 결합제 등)에 대한 제1의 결합제 및 제2의 결합제의 접촉은 동시에 수행된다. 예를 들면, 제1의 결합제 및 제2의 결합제, 및 임의로 어떠한 추가의 순서의 결합제도, 함께 혼주되어, 예를 들면 결합제의 라이브러리를 형성할 수 있다. 다른 실시예에서, 제1의 결합제 및 제2의 결합제, 및 임의로 어떠한 추가의 순서의 결합제도, 함께 혼주되기 보다는 거대분자에 동시에 가해진다. 일 구현예에서, 결합제의 라이브러리는 20개의 표준의, 천연적으로 존재하는 아미노산에 선택적으로 결합하는 적어도 20개의 결합제를 포함한다.

다른 구현예에서, 제1의 결합제 및 제2의 결합제, 및 임의로 어떠한 추가의 순서의 결합제를 별도의 결합 주기로 거대분자와 각각 접촉시키고, 순차적인 순서로 가한다. 특정 구현예에서, 병렬 시도는 시간을 절약하고 결합제가 경쟁 상태이므로 다수의 결합제의 동시 사용이 바람직하며, 이는 동종의 결합제에 의해 결합된 부위에 대해 비-동종의 결합제에 의한 비-특이적인 결합을 감소시킨다.

본원에 기술된 방법에 의해 생성된 최종의 연장된 기록 태그의 길이는 각각의 결합 주기로부터의 암호화 태그가 동일한 연장된 기록 태그 또는 다수의 연장된 기록 태그로 이전되는지의 여부에 상관없이, 암호화 태그(예컨대, 암호기 서열 및 스페이서)의 길이, 기록 태그(예컨대, 유일한 분자 확인인자, 스페이서, 보편적인 프라이밍 부위, 바코드)의 길이, 수행된 결합 주기의 수를 포함하는 다수의 인자에 의존한다. 펩타이드를 나타내고 에드만 분해 유사 절단 방법에 의해 생산된 콘카테네이트화된 연장된 기록 태그에 대한 예에서, 암호화 태그가 5개 염기의 스페이서에 의해 각각의 측면에서 플랭킹되는 5개 염기의 암호기 서열을 갖는 경우, 펩타이드의 결합제 역사를 나타내는, 최종의 연장된 기록 태그 상의 암호화 태그 정보는 10개의 염기 x 에드만 분해 주기의 수이다. 20-주기 작동의 경우, 연장된 기록은 적어도 200개 염기(초기 기록 태그 서열을 포함하지 않음)이다. 이러한 길이는 표준 차 세대 서열분석 장치와 혼용성이다.

최종의 결합 주기 및 최종의 결합제의 암호화 태그 정보의 연장된 기록 태그로의 이전 후, 기록기 태그는 연결, 프라이머 연장 또는 당해 분야에 공지된 다른 방법을 통해 보편적인 역 프라이밍 부위에 의해 캡핑할 수 있다. 일부 구현예에서, 기록 태그 내 보편적인 전방 프라이밍 부위는 최종의 연장된 기록 태그에 첨부된 보편적인 역 프라이밍 부위와 혼용성이다. 일부 구현예에서, 보편적인 역 프라이밍 부위는 일루미나 P7 프라이머(5'-

CAAGCAGAAGACGGCATACGAGAT - 3' - 서열 번호: 134) 또는 일루미나 P5 프라이머(5'-AATGATACGGCGACCACCGA-3' - 서열번호: 133)이다. 센스 또는 안티센스 P7은 기록 태그의 가닥 센스에 따라 첨부될 수 있다. 연장된 기록 태그 라이브러리는 고체 지지체(예컨대, 비드)로부터 직접 절단하거나 증폭시켜 전통적인 차 세대 서열분석 검정 및 프로토콜에 사용할 수 있다.

일부 구현예에서, 프라이머 연장 반응을 단일 가닥 연장된 기록 태그의 라이브러리에서 수행하여 이의 상보성 가닥을 카피한다.

NGPS 펩타이드 서열분석 검정은 주기적인 진행으로 수개의 화학적 및 효소적 단계를 포함한다. NGPS 서열분석이 단일 분자라는 사실은 과정에 대해 몇가지 주요 장점을 부여한다. 단일 분자 검정의 첫번째 주요 장점은 다양한 주기 화학적/효소적 단계에서 비효율성에 대한 강인성이다. 이는 암호화 태그 서열에 존재하는 주기-특이적인 바코드의 사용을 통해 가능하다.

주기-특이적인 암호화 태그를 사용하여, 본 발명자들은 각각의 주기로부터의 정보를 추적한다. 이는 단일 분자 서열분석 시도이므로, 서열분석 과정에서 각각의 결합/이전 주기에서 심지어 70% 효능이 맵핑가능한 서열 정보를 생성하는데 보다 더 충분하다. 예로서, 10개-염기 펩타이드 서열 "CPVQLWVDST"(서열 번호: 169)은 본 발명자들의 서열 플랫폼(여기서 X는 특정 아미노산이고; 아미노산의 존재는 주기 수 추적(tracking)에 의해 부여된다)에서 "CPXQXWXDXT"(서열 번호: 170)로 판독될 수 있다. 이러한 부분적인 아미노산 서열 판독은 이를 BLASTP를 사용하여 사람 p53 단백질로 유일하게 역 맵핑하는데 보다 더 충분하다. 따라서, 본 발명자들의 과정 어느 것도 강인하게 되는데 완벽하지 않다. 더욱이, 주기-특이적인 바코드가 본 발명자의 분배 개념과 조합하는 경우, 본 발명자들은 원래의 단백질 분자(구획 바코드를 통해)에 대해 맵핑하는 것을 알고 있으므로 단백질의 절대적인 확인은 10개 위치에서 확인된 단지 소수의 아미노산으로 달성될 수 있다.

XIV. 분획화, 구획화, 및 제한된 결합력 수지를 통한 단백질 표준화.

프로테오믹스 분석을 사용한 주요 챌린지 중 하나는 샘플내에서 단백질 풍부성에 있어서 큰 역학적 범위에 집중하는 것이다. 단백질은 혈장내에서 10 차수 초과의 역학적 범위(심지어 "상위 20개" 고갈된 혈장)에 이른다. 특정 구현예에서, 샘플로부터 특정 단백질 종(예컨대, 매우 풍부한 단백질)의 삭감은 분석 전에 수행된다. 이는 예를 들면, 상위 20개 혈장 단백질을 고갈하는, Sigma의 PROT20 면역-고갈 키트와 같은 상업적으로 이용가능한 단백질 고갈 시약을 사용하여 달성할 수 있다. 또한, 심지어 추가로 조절가능한 3 내지 4 차수까지 역학적 범위로 크게 감소시킨 시도를 하는 것이 유용할 수 있다. 특정 구현예에서, 단백질 샘플 역학적 범위는 전기영동 및 액체 크로마토그래피(Zhou, Ning et al. 2012), 또는 제한된 용량의 단백질 결합 비드/수지(예컨대, 하이드록실화된 실리카 입자)(McCormick 1989)로 로딩된 구획(예컨대, 소적)으로 분배시키고 결합된 단백질을 용출시킴으로써 조절할 수 있다. 각각의 구획화된 분획에서 과도한 단백질은 세척제거한다.

전기영동 방법의 예는 모세관 전기영동(CE), 모세관 등전점 포커싱(capillary isoelectric focusing: CIEF), 모세관 등속전기영동(CITP), 유리 유동 전기영동, 겔-용출된 액체 분획 포착 전기영동(gel-eluted liquid fraction entrapment electrophoresis: GELFrEE)을 포함한다. 액체 크로마토그래피 단백질 분리 방법의 예는 역상(RP), 이온 교환(IE), 크기 배제(SE), 소수성 상호작용 등을 포함한다. 구획 분배의 예는 유화액, 소적, 마이크로웰, 평평한 기판 상의 물리적으로 분리된 영역 등을 포함한다. 예시적인 단백질 결합 비드/수지는 페놀 그룹 또는 하이드록실 그룹(예컨대, Agilent Technologies로부터의 StrataClean 수지, LabTech로부터의 RapidClean 등)으로 유도체화된 실리카 나노입자를 포함한다. 비드/수지의 결합능을 제한함으로서, 제공된 분획에서 용출하는 매우 풍부한 단백질이 비드에 단지 부분적으로 결합하고 과도한 단백질이 제거될 것이다.

XV. 단일 세포 또는 분자 소샘플링(subsampling)의 프로테옴의 분배

다른 국면에서, 본 개시내용은 바코딩 및 분배 기술을 사용하여 샘플 속에서 단백질의 대량의 병렬 분석 방법을 제공한다. 단백질 분석에 대한 현재의 시도는 펩타이드 서열분석에 적합한 보다 짧은 펩타이드 분자로의 단백질 거대분자의 단편화를 포함한다. 따라서, 이러한 시도를 사용하여 수득된 정보는 단편화 단계에 의해 제한되며, 예컨대, 해독 후 변형, 각각의 샘플에서 발생하는 단백질-단백질 상호작용, 샘플 속에 존재하는 단백질 집단의 조성, 또는 특수한 세포 또는 세포의 집단으로부터와 같은, 단백질 거대분자의 기원을 포함하는, 단백질의 긴 범위의 연속성 정보를 배제한다. 단백질 분자내 해독 후 변형의 긴 범위의 정보(예컨대, 프로테오폼(proteoform) 특성화)는 생물학의 보다 완전한 그림, 및 어떠한 단백질에 속하는 어떠한 펩타이드가 근본적인 단백질 서열에 대한 펩타이드 서열의 보다 강력한 맵핑을 제공하는지에 대한 긴 범위의 정보를 제공한다(참고: 도 15의 A). 이는 펩타이드 서열분석 기술이 단지 5개 아미노산 유형으로부터의 정보와 같은 불완전한 아미노

산 서열 정보 만을 제공하는 경우에 특히 관련된다. 동일한 단백질 분자로부터 기원하는 다수의 펩타이드로부터의 정보와 조합된, 본원에 개시된 분배 방법을 사용함으로써, 단백질 분자의 동일성(예컨대, 프로테오폼)를 보다 정밀하게 평가할 수 있다. 구획 태그와 동일한 구획(들)로부터 기원한 단백질 및 펩타이드의 연합은 분자 및 세포 정보의 재구축을 촉진한다. 전형적인 프로테옴 분석에서, 세포는 용해되고 단백질은 짧은 펩타이드로 분해되어 단백질이 어떠한 세포 또는 세포형으로부터 기원하는지, 및 어떠한 펩타이드가 어떠한 단백질 또는 단백질 복합체로부터 기원하는지에 대한 전반적인 정보를 붕괴시킨다. 이러한 전반적인 정보는 세포 및 조직내 생물학 및 생화학을 이해하는데 있어 중요하다.

분배는 샘플내 거대분자의 집단으로부터 거대분자의 소집단으로 유일한 바코드의 무작위 지정을 지칭한다. 분배는 거대분자를 구획내에 분포시킴으로써 달성할 수 있다. 분배는 단일 구획내 거대분자 또는 구획의 집단으로부터의 다수의 구획내 거대분자로 구성될 수 있다.

다수(예컨대, 수백만 내지 수십억)의 구획으로부터 동일한 물리적 구획 또는 구획의 그룹 내로 또는 상으로 분리된 거대분자의 세트 또는 단백질 샘플의 소 세트는 유일한 구획 태그에 의해 확인된다. 따라서, 구획 태그를 사용하여 심지어 구성성분이 함께 혼주된 후에도, 상이한 구획 태그를 갖는 다른 구획(또는 구획의 그룹)에서의 것들로부터의 동일한 구획 태그를 갖는 하나 이상의 구획으로부터 기원한 구성성분을 구별할 수 있다.

본 개시내용은 복합체 프로테옴 샘플(예컨대, 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드) 또는 복합체 세포 샘플을 다수의 구획으로 분배함으로써 단백질 분석을 향상시키는 방법을 제공하며, 여기서 각각의 구획은 개개의 구획내에서 동일한(임의의 UMI 서열이 절약됨) 및 다른 구획의 구획 태그로부터 상이한 다수의 구획 태그를 포함한다(참고: 도 18 내지 20). 구획은 다수의 구획 태그가 이에 결합된 고체 지지체(예컨대, 비드)를 임의로 포함한다. 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드는 다수의 펩타이드로 단편화되며, 이는 이후에 다수의 구획 내에서 다수의 구획 태그를 지닌 다수의 펩타이드의 어닐링 또는 결합을 허용하기에 충분한 조건 하에서 다수의 구획 태그와 접촉함으로써 다수의 구획 태그된 펩타이드를 생성한다. 대안적으로, 다수의 단백질 복합체, 단백질, 또는 폴리펩타이드는 다수의 단백질 복합체, 단백질 또는 폴리펩타이드와 다수의 구획내 다수의 구획 태그의 어닐링 또는 결합을 허용하기에 충분한 조건 하에서 다수의 구획 태그에 결합함으로써, 다수의 구획 태그된 단백질 복합체, 단백질, 폴리펩타이드를 생성한다. 구획 태그된 단백질 복합체, 단백질, 또는 폴리펩타이드는 이후에 다수의 구획으로부터 수집되어 다수의 구획 태그된 펩타이드로 단편화된다. 하나 이상의 구획 태그된 펩타이드는 본원에 기술된 방법들 중 어느 것에 따라 분석된다.

특정 구현예에서, 구획 태그 정보는 프라이머 연장(도 5) 또는 연결(도 6)을 통해 거대분자(예컨대, 펩타이드)와 관련된 기록 태그로 이전된다.

일부 구현예에서, 구획 태그는 구획 내에서 용액 속에 유리된다. 다른 구현예에서, 구획 태그는 구획의 표면(예컨대, 마이크로역가 또는 피코역가 플레이트의 웰 바닥) 또는 비드 또는 구획내 비드에 직접 결합된다.

구획은 수성 구획(예컨대, 미세유동 소적) 또는 고체 구획일 수 있다. 고체 구획은 예를 들면, 나노입자, 미세구, 미세역가 또는 피코역가 웰 또는 배열내 분리된 영역, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 셀, 유동 통과 칩, 시그널 변환 전자 장치를 포함하는 바이오칩, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면을 포함한다. 특정 구현예에서, 각각의 | 구획은 평균적으로, 단일 세포를 함유한다.

고체 지지체는 비드, 마이크로비드, 배열, 유리 표면, 규소 표면, 플라스틱 표면, 필터, 막, 나일론, 규소 웨이퍼 칩, 유동 셀, 유동 통과 칩, 시그널 변환 전자 장치를 포함하는 바이오칩, 미세역가 웰, ELISA 플레이트, 스피닝 인터페로메트리 디스크, 니트로셀룰로즈 막, 니트로셀룰로즈-기반 중합체 표면, 나노입자, 또는 미세구를 포함하나, 이에 한정되지 않은 어떠한 지지체 표면일 수 있다. 고체 지지체용 물질은 아크릴아미드, 아가로즈, 셀룰로즈, 니트로셀룰로즈, 유리, 금, 석영, 폴리스티렌, 폴리에틸렌 비닐 아세테이트, 폴리프로필렌, 폴리메타크릴레이트, 폴리에틸렌, 폴리에틸렌 옥사이드, 폴리실리케이트, 폴리카보네이트, 테플론, 플루오로카본, 나일론, 규소 고무, 다가무수물, 폴리글리콜산, 폴리악트산, 폴리오르토에스테르, 기능화된 실란, 폴리프로필푸머레이트, 콜라겐, 글리코사미노글리칸, 폴리아미노산, 또는 이의 어떠한 조합을 포함하나, 이에 한정되지 않는다. 특정의 구현예에서, 고체 지지체는 비드, 예를 들면, 폴리스티렌 비드, 중합체 비드, 아가로즈 비드, 아크릴아미드 비드, 고체 코어 비드, 다공성 비드, 상자성 비드, 유리 비드, 또는 조절된 공극 비드이다.

샘플을 구획 태그된 비드를 지닌 구획으로 분배하는 다양한 방법은 셈베커(Shembekar) 등의 문헌(Shembekar, Chaipan et al. 2016)에서 고찰된다. 하나의 예에서, 프로테옴은 본원에 개시된 방법을 사용하여 기록될 단백질

분자 및 단백질 복합체 상의 전반적인 정보를 가능하도록 하는 유화액을 통해 소적으로 분배된다(참고: 예를 들면, 도 18 및 도 19). 특정 구현예에서, 프로테옴은 구획 태그된 비드, 활성화가능한 프로테아제(열, 광 등을 통해 직접적으로 또는 간접적으로), 및 프로테아제-내성이 되도록 가공된 펩타이드 리가제(예컨대, 변형된 라이신, 페길화 등)과 함께 구획(예컨대, 소적)내에서 분배된다. 특정 구현예에서, 프로테옴은 변성체로 처리함으로써 단백질 또는 폴리펩타이드의 구성 성분을 평가할 수 있다. 단백질의 천연 상태에 관한 정보가 요구되는 경우, 상호작용하는 단백질 복합체를 이들로부터 기원한 펩타이드의 후속적인 분석을 위해 구획내로 분배할 수 있다.

구획 태그는 하나 또는 양쪽 측면에서 스페이서 또는 보편적인 프라이머 서열에 의해 임의로 플랭킹된 바코드를 포함한다. 프라이머 서열은 기록 태그의 3' 서열에 대해 상보성이므로, 구획 태그 정보를 프라이머 연장 반응을 통해 기록 태그로 이전할 수 있다(참고: 도 22의 A 및 B). 바코드는 고체 지지체 또는 구획에 부착된 단일 가닥 핵산 분자 또는 고체 지지체 또는 구획에 하이브리드화된 이의 상보성 서열, 또는 둘 다의 가닥으로 구성될 수 있다(참고: 예컨대, 도 16). 구획 태그는 펩타이드에 대한 커플링을 위해, 예를 들면 스페이서에 부착된 기능성 모이어티를 포함할 수 있다. 하나의 예에서, 기능성 모이어티(예컨대, 알데하이드)는 다수의 펩타이드에서 N-말단 아미노산 잔기와 반응할 수 있는 것이다. 다른 예에서, 기능성 모이어티는 다수의 펩타이드에서 내부 아미노산 잔기(예컨대, 라이신 또는 "클릭" 반응성 모이어티로 표지된 라이신)와 반응할 수 있다. 다른 구현예에서, 기능성 모이어티는 단순히 DNA 태그-표지된 단백질에 하이브리드화할 수 있는 상보성 DNA 서열일 수 있다. 대안적으로, 구획 태그는 단백질 리가제(예컨대, 부텔라제 I 또는 이의 동족체)에 대한 인식 서열을 포함함으로써 목적한 펩타이드에 대한 구획 태그의 연결을 허용하는 펩타이드를 추가로 포함하는 키메라 분자일 수 있다(참고: 도 22의 A). 구획 태그는 보다 큰 핵산 분자내의 성분일 수 있으며, 이는 임의로 이에 결합된 펩타이드, 스페이서 서열, 보편적인 프라이밍 부위, 또는 이의 어떠한 조합 상의 확인 정보를 제공하기 위한 유일한 분자 확인인자를 추가로 포함할 수 있다. 이러한 UMI 서열은 일반적으로 구획내 구획 태그의 집단 중에서 상이하다. 특정 구현예에서, 구획 태그는 기록 태그내 성분이어서, 개개의 구획 정보를 제공하는데 사용된 동일한 태그가 이에 부착된 펩타이드에 대한 개개 펩타이드 정보를 기록하는데 사용될 수 있다.

특정 구현예에서, 구획 태그는 구획 태그를 구획내로 인쇄하고, 스포팅(spotting) 하고, 잉크-젯팅(ink-jetting)함으로써 형성될 수 있다. 특정 구현예에서, 다수의 구획 태그된 비드가 형성되며, 여기서 하나의 바코드 유형이 문헌(Klein et al., 2015, Cell 161:1187-1201; Macosko et al., 2015, Cell 161:1202-1214; and Fan et al., 2015, Science 347:1258367)에 기술된 바와 같이 분할-및-혼주(split-and-pool) 올리고뉴클레오타이드 연결을 통해, 비드당 존재한다. 구획 태그된 비드는 또한 개개 합성 또는 고정화에 의해 형성될 수 있다. 특정 구현예에서, 구획 태그된 비드는 이기능성 기록 태그를 추가로 포함할 수 있으며, 여기서 하나의 부위는 기록 태그를 포함하는 구획 태그를 포함하고, 다른 부위는 소화된 펩타이드가 커플링될 수 있는 기능성 모이어티를 포함한다(도 19 및 도 20).

특정 구현예에서, 다수의 구획내 다수의 단백질 또는 폴리펩타이드는 프로테아제를 사용하여 다수의 펩타이드로 단편화시킨다. 프로테아제는 금속프로테아제일 수 있다. 특정 구현예에서, 금속프로테아제의 활성은 금속 양이온의 광-활성화된 방출에 의해 조절된다. 사용될 수 엔도펩티다제의 예는: 트립신, 키모트립신, 엘라스타제, 서몰라이신, 펩신, 클로스트리판, 글루타밀 엔도펩티다제(GluC), 엔도펩티다제 ArgC, 펩티딜-asp 메탈로-엔도펩티다제(AspN), 엔도펩티다제 LysC 및 엔도펩티다제 LysN을 포함한다. 이들의 활성화 방식은 완충액 및 2가 양이온 요건에 따라 변한다. 임의로, 단백질 또는 폴리펩타이드를 펩타이드 단편으로 충분히 소화한 후, 프로테아제는 불활성화된다(예컨대, 열, 플루오로-오일 또는 규소 오일 가용성 억제제, 예를 들면, 2가 양이온 킬레이트제).

구획 태그로 바코딩되는 펩타이드의 특정 구현예에서, 단백질 분자(임의로, 변성된 폴리펩타이드)는 DNA 태그를 단백질의 라이신 그룹의 ε-아민 모이어티에 대한 접합에 의해 또는 간접적으로 알킨과 같은 반응성 클릭 모이어티로 예비-표지된 단백질/폴리펩타이드에 대한 클릭 화학 부착을 통해 간접적으로 DNA 태그로 표지된다(도 2b 및 도 20의 A). DNA 태그-표지된 폴리펩타이드는 이후 구획 태그(예컨대, 소적내에 함유된 비드에 결합된 DNA 바코드)로 분배되며(참고: 도 20의 B), 여기서 구획 태그는 각각의 구획을 확인하는 바코드를 함유한다. 일 구현예에서, 단일의 단백질/폴리펩타이드 분자는 비드와 관련된 DNA 바코드의 단일 종으로 동시-캡슐화된다(참고: 도 20의 B). 다른 구현예에서, 구획은 DNA보다는 단백질에 적용되는 것을 제외하고는, PCR 공보 제 WO2016/061517호(이의 전문은 참고로 포함됨)에 기술된 것과 유사한 부착된 구획(비드) 태그를 지닌 비드의 표면을 구성할 수 있다. 구획 태그는 바코드(BC) 서열, 보편적인 프라이밍 부위(U1'), UMI 서열, 및 스페이서 서열(Sp)을 포함할 수 있다. 일 구현예에서, 분배와 동시에 또는 분배 후에, 구획 태그는 비드로부터 절단되어 폴리펩타이드에 부착된 DNA 태그에, 예를 들면 각각 DNA 태그 및 구획 태그 상의 상보성 U1 및 U1' 서열을 통해

하이브리드화된다. 비드 상의 분배를 위해, DNA 태그-표지된 단백질은 비드 표면 상의 구획 태그에 직접 하이브리드화될 수 있다(참고: 도 20의 C). 이러한 하이브리드화 단계 후, 하이브리드화된 DNA 태그를 지닌 폴리펩타이드를 구획(예컨대, "크랙된" 유화액, 또는 비드로부터 절단된 구획 태그)으로부터 추출하고, 폴리머라제-기반된 프라이머 연장 단계를 사용하여 바코드 및 UMI 정보를 폴리펩타이드 상의 DNA 태그에 써서 구획 바코드화된 기록 태그를 수득한다(참고: 도 20의 D). LysC 프로테아제 소화를 사용하여 폴리펩타이드를 보편적인 프라이밍 서열, 구획 태그, 및 UMI를 함유하는 기록 태그를 지닌 이들의 C-말단 라이신에서 표지된 구성성분 펩타이드로 절단할 수 있다(참고: 도 20의 E). 일 구현예에서, LysC 프로테아제는 DNA-태그된 라이신 잔기에 내성이 되도록 가공된다. 수득되는 기록 태그 표지된 펩타이드는 적절한 밀도에서 고체 지지체(예컨대, 비드) 상에 고정되어 기록 태그된 펩타이드 상의 분가간 상호작용을 최소화시킨다(참고: 도 20의 E 및 20의 F).

구획 태그에 대한 펩타이드의 부착(또는 이의 역)은 고정된 구획 태그에 대해, 또는 이의 상보성 서열(이중 가닥인 경우)에 대해 직접 이루어질 수 있다. 대안적으로, 구획 태그는 고체 지지체 또는 구획의 표면, 및 구획내에서 결합된 펩타이드 및 용액상 구획 태그로부터 탈착될 수 있다. 일 구현예에서, 구획 태그(예컨대, 올리고뉴클레오타이드의 말단) 상의 기능성 모이어티는 쉬프 염기(Schiff base)를 통해 펩타이드의 아민 N-말단에 직접 커플링된 알데하이드이다(참고: 도 16). 다른 구현예에서, 구획 태그는 단백질 리가제에 대한 펩타이드 모티프(n-X...XXCGSHV-c)를 포함하는 핵산-펩타이드 키메라 분자로서 작제된다. 핵산-펩타이드 구획 태그 작제물은 부텔라제 I 또는 이의 동족체와 같은 펩타이드 리가제를 사용하는 소화된 펩타이드에 접합된다. 부텔라제 I, 및 다른 아스파라기닐 엔도펩티다제(AEP) 동족체를 사용하여 올리고뉴클레오타이드-펩타이드 구획 태그 작제물의 C-말단을 소화된 펩타이드의 N-말단에 연결할 수 있다(Nguyen, Wang et al. 2014, Nguyen, Cao et al. 2015). 이러한 반응은 신속하고 매우 효율적이다. 수득되는 구획 태그된 펩타이드는 본원에 기술된 바와 같이 핵산 펩타이드 분석을 위해 고체 지지체에 후속적으로 고정시킬 수 있다.

특정 구현예에서, 고체 지지체 또는 구획의 표면에 결합된 구획 태그는 구획 태그를 다수의 단편화된 펩타이드와 결합시키기 전에 방출된다(참고: 도 18). 일부 구현예에서, 다수의 구획으로부터 구획 태그된 펩타이드의 수집 후, 구획 태그된 펩타이드는 기록 태그와 함께 결합된 고체 지지체이다. 이후에 구획 태그 정보는 구획 태그된 펩타이드 상의 구획 태그로부터 관련된 기록 태그(예컨대, 기록 탭(tab) 및 구획 태그내 상보성 스페이서 서열로부터 프라이밍된 프라이머 연장 반응을 통해)로 이전될 수 있다. 일부 구현예에서, 구획 태그는 이후 본원에 기술된 방법에 따라 펩타이드 분석 전에 구획 태그된 펩타이드로부터 제거된다. 추가의 구현예에서, 다수의 단백질을 소화하는데 초기 사용된 서열 특이적인 프로테아제(예컨대, Endo AspN)를 또한 사용하여 구획 태그 정보를 관련된 기록 태그로 이전한 후 펩타이드의 N 말단으로부터 구획 태그를 제거한다(참고: 도 22의 B).

구획-기반한 분배를 위한 시도는 T-연결 및 유동 포커싱을 사용하는 미세유동 장치를 통한 소적 형성, 교반을 사용한 유화액 생성 또는 작은 공극을 지닌 막(예컨대, 트랙 에치 막(track etch membrane)) 등을 통한 압출을 포함한다(참고: 도 21). 구획화를 사용한 챌린지는 구획의 내부에 집중하는 것이다. 특정 구현예에서, 유체 성분을 교환하는 것이 도전이므로, 이는 구획내에서 일련의 상이한 생화학적 단계를 수행하는 것이 어려울 수 있다. 앞서 기술된 바와 같이, 시약을 유화액의 플루오로-오일에 첨가함으로써, pH, 킬레이팅제, 환원제와 같은 소적 내부의 한정된 특징을 변형시킬 수 있다. 그러나, 수성 및 유기 상 둘 다에서 용해도를 갖는 화합물의 수는 한정되어 있다. 한가지 시도는 구획내 반응을 필수적으로 목적한 분자에 대한 바코드의 이전으로 제한하는 것이다.

단백질/펩타이드를 구획 태그(바코드)로 구성된 기록 태그로 표지한 후, 단백질/펩타이드를 적합한 밀도로 고체-지지체 상에 고정시켜 결합된 펩타이드 또는 단백질 분자에 부착된 상응하는 기록 태그/태그들에 대한 결합된 동종의 결합체의 암호화 태그로부터의 정보의 분자내 이전을 돕는다. 분자간 정보 이전은 고체-지지체의 표면 위에 분자내 공간을 조절함으로써 최소화된다.

특정 구현예에서, 구획 태그는 구획의 집단내 각각의 구획에 대해 유일할 필요가 없다. 구획의 집단내 구획의 소세트(2개, 3개, 4개 이상)는 동일한 구획 태그를 공유할 수 있다. 예를 들면, 각각의 구획은 샘플의 소집단을 포획하기 위해 작동하는 비드 표면의 집단으로 구성될 수 있다(많은 분자는 비드당 포획된다). 더욱이, 비드는 포획된 거대분자에 부착될 수 있는 구획 바코드를 포함한다. 각각의 비드는 단일의 구획 바코드 서열을 가지지만, 이러한 구획 바코드는 구획(동일한 바코드에 맵핑하는 많은 비드)내에서 다른 비드상에서 복제될 수 있다. 물리적 구획과 구획 바코드 사이에 다수-대-일 맵핑(many-to-one mapping)일 수 있으며(요구되지는 않는다), 더욱이 구획내 거대분자 사이에 다수-대-일 맵핑이 존재할 수 있다(요구되지는 않는다). 분배 바코드는 샘플내 거대분자의 집단으로부터 거대분자의 소샘플링에 대한 유일한 바코드의 지정으로 정의된다. 이러한 분배 바코드는 동일한 바코드로 표지된 구획내 거대분자의 분배로부터 발생된 동일한 구획 바코드로 구성될 수 있다. 물리적

구획의 사용은 원래의 샘플을 효과적으로 소샘플링하여 분배 바코드의 지정을 제공한다. 예를 들면, 10,000개의 상이한 구획 바코드로 표지된 비드의 세트가 제공된다. 또한, 제공된 검정에서 고려할 때, 1백만개의 비드의 집단이 검정에 사용된다. 평균적으로, 구획 바코드 당 100개의 비드가 존재한다(푸아송 분포(Poisson distribution)). 비드가 평균 천만개의 거대분자의 집합체를 포획하는 것으로 추가로 가정한다. 평균적으로 비드당 10개의 거대분자가 존재하며, 구획 바코드당 100개의 구획이 존재하므로, 분배 바코드(100개의 명백한 물리적 구획에 대해 100개의 구획으로 구성된다)당 효과적으로 1000개의 거대분자가 존재한다.

다른 구현예에서, 폴리펩타이드의 단일의 분자 분배 및 분배 바코드화는 N 또는 C 말단 또는 둘 다에서 폴리펩타이드를 (화학적으로 또는 효소적으로) 증폭가능한 DNA UMI 태그(예컨대, 기록 태그)로 표지함으로써 달성된다(참고: 도 37). DNA 태그는 도 2b에 나타낸 바와 같이 라이신과 같은 반응성 아미노산에 대한 비-특이적인 광-표지화 또는 특이적인 화학적 부착을 통해 폴리펩타이드(내부 아미노산)의 바디(body)에 부착된다. 펩타이드의 말단에 부착된 기록 태그로부터의 정보는 효소 유화액 PCR(Williams, Peisajovich et al. 2006, Schutze, Rubelt et al. 2011) 또는 유화액 시험관내 전사/역전사(IVT/RT) 단계를 통해 DNA 태그로 이전된다. 바람직한 구현예에서, 나노유화액을 사용하여 평균적으로, 크기가 50 nm 내지 1000 nm인 유화액 소적 당 1개보다 적은 폴리펩타이드가 존재하도록 한다(Nishikawa, Sunami et al. 2012, Gupta, Eral et al. 2016). 또한, PCR의 성분은 프라이머, dNTP, Mg2+, 폴리머라제, 및 PCR 완충액을 포함하는 수성 유화액 혼합물 속에 포함된다. IVT/RT가 사용되는 경우, 기록 태그는 T7/SP6 RNA 폴리머라제 프로모터 서열로 설계되어 폴리펩타이드의 바디에 부착된 DNA 태그에 하이브리드화하는 전사체를 생성한다(Ryckelynck, Baudrey et al. 2015). 리버스 트랜스크립타제(RT)는 하이브리드화된 RNA 분자로부터의 정보를 DNA 태그로 카피한다. 이러한 방식으로, 유화액 PCR 또는 IVT/RT를 사용하여 말단 기록 태그로부터의 정보를 폴리펩타이드의 바디에 부착된 다수의 DNA 태그로 효과적으로 이전시킬 수 있다.

비드내 겔화를 통한 세포 성분의 캡슐화는 단일 세포 분석에 대해 유용한 시도이다(Tamminen and Virta 2015, Spencer, Tamminen et al. 2016). 단일 세포 소적의 바코드화는 단일 세포로부터의 모든 성분을 동일한 확인인자로 표지되도록 한다(Klein, Mazutis et al. 2015, Gunderson, Steemers et al. 2016, Zilionis, Nainys et al. 2017). 구획 바코드화는 소적 결합(Raindance)에 의해, 바코드된 비드의 소적내로의 도입에 의해(10X Genomics), 또는 군더슨(Gunderson) 등의 문헌(Gunderson, Steemers et al. 2016) 및 PCT 공보 제 WO2016/130704호(이의 전문은 참고로 포함된다)에 기술된 바와 같이 분할-혼주 조합 바코딩을 사용하여 캡슐화 및 겔화 후 소적의 성분의 조합적 바코딩에 의해 유일한 바코드의 각각의 소적내로의 직접적인 혼입을 포함하는 다수의 방식으로 달성할 수 있다. 유사한 조합 표지화 개략도를 또한 에디(Adey) 등의 문헌(Vitak, Torkenczy et al. 2017)에 기술된 바와 같이 핵에 적용할 수 있다.

상기 소적 바코딩 시도는 DNA 분석에 사용되었으나 단백질 분석에서는 사용되지 않았다. 상기 소적 바코딩 플랫폼을 단백질을 사용하여 작업하기 위해 채택하는 것은 몇가지 혁신적인 단계를 필요로 한다. 첫번째는 바코드가 주로 DNA 서열로 구성되고, 이러한 DNA 서열 정보가 단백질 분석물에 제공될 필요가 있다는 것이다. DNA 분석물의 경우에, 이는 DNA 정보를 DNA 분석물로 이전하기에 비교적 간단하다. 대조적으로, DNA 정보를 단백질에 이전시키는 것은 특히, 단백질이 변성되어 하부 분석을 위해 펩타이드로 소화된 경우에, 보다 도전이 된다. 이는 각각의 펩타이드가 구획 바코드로 표지되는 것을 요구한다. 도전은 세포가 소적내로 캡슐화되면, 단백질을 변성시키기가 어려우며, 프로테아제는 수득되는 폴리펩타이드를 소화하고, 동시에 펩타이드를 DNA 바코드로 표지한다는 것이다. 수성 완충액으로 될 수 있는, 소적을 형성하는 중합체내에서 세포의 캡슐화 및 다공성 비드내로 이들의 중합화(겔화)는 소적 속의 세포와는 달리 다수의 상이한 반응 단계를 수행하는 비히클을 제공한다(Tamminen and Virta 2015, Spencer, Tamminen et al. 2016)(Gunderson, Steemers et al. 2016). 바람직하게는, 캡슐화된 단백질은 겔 매트릭스로 가교결합되어 겔 비드로부터 이들의 후속적인 확산을 방지한다. 이러한 겔 비드 양식은 겔 내에 포획된 단백질이 화학적으로 또는 효소적으로 변형되어 DNA 태그로 표지되고, 프로테아제 소화되며, 다수의 다른 개입에 적용되도록 한다. 도 38은 겔 매트릭스 속에서 단일 세포의 예시적인 캡슐화 및 용해를 나타낸다.

XVI. 조직 및 단일 세포 공간 프로테오믹스

바코드의 다른 용도는 공간적으로 분포된 DNA 바코드 서열의 배열을 표면 상의 조직에 공간 분절화(segmentation)하는 것이다. 조직 단백질이 배열 표면에 놓인 세포 조직내 단백질의 공간 위치를 반영하는 바코드를 포함하는 DNA 기록 태그로 표지되는 경우, 조직 슬라이스내 단백질 분해물의 공간 분포는, 스타흘(Stahl) 등의 문헌(2016, Science 353(6294):78-82) 및 크로세토(Crosetto) 등의 문헌(Corsetto, Bienko et al., 2015)에 기술된 바와 같이 공간 전사체학을 위해 수행되지만, 서열 분석 후 후에 재구축될 수 있다. 공간 바코드의

부착은 배열로부터 배열-결합된 바코드를 방출하고 이들을 조직 단면으로 확산시키거나, 대안적으로, 조직 단면 내 단백질을 DNA 기록 태그로 표지한 후, 단백질을 프로테아제로 소화시켜 배열 위에서 공간 바코드로 확산하여 가수분해할 수 있는 표지된 펩타이드를 방출할 수 있다. 바코드 정보는 이후 펩타이드에 부착된 기록 태그로 이전(효소적으로 또는 화학적으로)된다.

조직내 단백질의 공간 바코드화는 공간적으로 암호화된 DNA 배열 상에 DNA 기록 태그로 화학적으로 표지된, 고정된/투과된 조직 슬라이스를 위치시킴으로써 달성될 수 있으며, 여기서 배열 상의 각각의 특징은 공간적으로 확인가능한 바코드를 갖는다(참고: 도 23). 배열 바코드를 DNA 태그로 부착시키기 위하여, DNA 태그 표지된 펩타이드를 방출시키며, 이는 조직 슬라이스에서 가까운 근접한 배열 특징으로 확산하여 하이브리드화할 수 있다. 배열 바코드 정보는 화학적/효소적 연결 또는 폴리머라제 연장을 사용하여 DNA 태그로 이전시킬 수 있다. 대안적으로, 표지된 펩타이드가 배열 표면에 확산되도록 하기 보다는, 배열 상의 바코드 서열을 절단하여 조직 슬라이스 위의 근접한 부위내로 확산시켜 여기서 DNA 태그-표지된 단백질에 하이브리드화하도록 할 수 있다. 다시 한번, 바코드화 정보는 화학적/효소적 연결 또는 폴리머라제 연장에 의해 이전될 수 있다. 이러한 두번째 경우에, 프로테아제 분해는 바코드 정보의 이전 후 수행될 수 있다. 이러한 시도의 결과는 기록 태그-표지된 단백질 또는 펩타이드의 수집이며, 여기서 기록 태그는 원래의 조직 내에서 단백질/펩타이드 위치의 2-D 공간 정보를 지닌 바코드를 포함한다. 더욱이, 해독 후 변형의 공간 분포가 특성화될 수 있다. 이러한 시도는 민감하고 고도로-멀티플렉스화된 반응계내 디지탈 면역조직화학 검정을 제공하며, 훨씬 정밀한 진단 및 예후로 이끄는 현대적인 분자 병리학의 기초를 형성할 수 있다.

다른 구현예에서, 공간적 바코드화를 세포내에서 사용하여 세포 기관 및 세포 구획내 단백질 구성성분/PTM을 확인할 수 있다(Christoforou et al., 2016, Nat. Commun. 7:8992, 이의 전문은 참고로 포함됨). 다수의 시도를 사용하여 세포내 공간 바코드를 제공할 수 있으며, 이는 인접한 단백질에 부착될 수 있다. 일 구현예에서, 세포 또는 조직은 구성성분 세포기관내로 분획화된 소-세포, 및 바코드화된 상이한 단백질 세포기관 분획일 수 있다. 공간 세포 표지화의 다른 방법은 이의 전문이 참고로 포함된 문헌(Marx, 2015, Nat Methods 12:815-819)에 기술되어 있으며, 유사한 시도가 본원에서 사용될 수 있다.

발명을 실시하기 위한 구체적인 내용

[0265] 다음의 실시예는 설명 목적을 위해 제공되며, 제한하기 위한 것이 아니다.

[0266] 실시예

[0267] 실시예 1: 프로테이나제 K를 사용한 단백질 샘플의 소화

[0268] 펩타이드의 라이브러리를 트립신, 프로테이나제 K 등과 같은 프로테아제를 사용한 소화에 의해 단백질 샘플로부터 제조한다. 트립신은 바람직하게는 라이신 및 아르기닌과 같은 양성으로 하전된 아미노산의 C-말단 측면에서 절단하지만, 프로테이나제 K는 단백질을 따라 비-특이적으로 절단한다. 이와 같이, 프로테이나제 K 소화는 충분한 단백질분해를 제공하여 짧은 펩타이드(~ 30개 아미노산)를 생성하지만, 샘플을 과도하게 소화하지 않는, 바람직한 효소-대-폴리펩타이드 비를 사용한 조심스러운 적정을 필요로 한다. 일반적으로, 기능성 활성의 적정은 제공된 프로테이나제 K를 다량 사용하여 수행할 필요가 있다. 본 실시예에서, 단백질 샘플은 프로테이나제 K로 1시간 동안 37℃에서 1X PBS/1 mM EDTA/0.5 mM $CaCl_2$/0.5% SDS (pH 8.0) 중 1:10 내지 1:100 (w/w)의 효소:단백질 비로 소화된다. 항온처리 후, PMSF를 5 mM의 최종 농도로 가하여 추가의 소화를 억제한다.

[0269] 프로테이나제 K의 비활성(specific activity)은 "화학적 기질" 벤조일 아르기닌-p-니트로아닐린을 프로테이나제 K와 항온처리하고 ~410nm에서 흡수하는 황색의 p-니트로아닐린 생성물의 발달을 측정함으로써 측정할 수 있다. 효소 활성은 단위로 측정되며, 여기서 1 단위는 1 μmole의 생산된 p-니트로아닐리드/분과 동일하고, 특이적인 활성은 효소 활성의 단위/mg의 총 단백질로 측정된다. 비활성은 이후 효소 활성을 용액 속의 단백질의 총량으로 나누어 계산한다.

[0270] 실시예 2: 비드 프로테아제 소화 및 표지화에 있어서 SP3를 사용한 샘플 제조

[0271] 단백질을 추출하고 휴그스(Hughes) 등의 문헌(2014, Mol Syst Biol 10:757)에 기술된 바와 같이 SP3 샘플 제조 프로토콜을 사용하여 변성시켰다. 추출 후, 단백질 혼합물(및 비드)를 0.02% SDS가 보충된 1 mM EDTA가 들어있는 50 mM 붕산염 완충액(pH 8.0) 속에서 37℃에서 1시간 동안 용해시킨다. 단백질 용해 후, 이황화물 결합은 DTT를 5 mM의 최종 농도로 가하고, 샘플을 50℃에서 10분 동안 항온처리함으로써 환원시킨다. 시스테인은 요오도아세트아미드를 10 mM의 최종 농도로 가하여 알킬화하고 암실에서 실온으로 20분 동안 항온처리한다. 반응물

을 50 mM 붕산염 완충액으로 2배 희석시키고, Glu-C 또는 Lys-C를 1:50 (w/w)의 최종 프로테이나제:단백질의 비로 가한다. 샘플을 37℃에서(하룻밤(o/n), ~16 시간 동안) 항온처리하여 소화를 완료한다. 휴그스 등(상기 참고)에 의해 기술된 바와 같은 샘플 소화 후, 100% 아세토니트릴을 95% 아세토니트릴의 최종 농도로 가함으로써 펩타이드를 비드에 결합시키고 8분 항온처리시 아세토니트릴로 세척하였다. 세척 후, 펩타이드를 5분의 피펫 혼합 단계에 의해 10 μl의 2% DMSO 속에서 비드를 용출 제거하였다.

[0272] 실시예 3: 펩타이드에 대한 기록 태그의 커플링

[0273] DNA 기록 태그를 여러 방식으로 펩타이드에 커플링시킨다(참고: Aslam et al., 1998, Bioconjugation: Protein coupling Techniques for the Biomedical Sciences, Macmillan Reference LTD; Hermanson GT, 1996, Bioconjugate Techniques,cademic Press Inc., 1996). 하나의 시도에서, 올리고뉴클레오타이드 기록 태그를 카브디이미드 화학을 사용하여 펩타이드의 C-말단에 커플링하는 5' 아민, 및 클릭 화학을 사용하여 아지드 비드에 커플링하는, 내부 가닥화된 알킨, DBCO-dT(Glen Research, VA)로 작제한다. 기록 태그를 용액 속에서 큰 몰 초과량의 기록 태그를 사용하여 펩타이드에 커플링함으로써 카보디이미드 커플링을 완료시키고, 펩타이드-펩타이드 커플링을 제한한다. 대안적으로, 올리고뉴클레오타이드를 5' 가닥화된 알킨(DBCO-dT)으로 작제하고, 아지드-유도체화된 펩타이드(펩타이드의 C-말단에 대한 아지드-PEG-아민 및 카보디이미드 커플링을 통해)에 커플링하고, 알데하이드-반응성 HyNic 하이드라딘 비드에 커플링한다. 기록 태그 올리고뉴클레오타이드는 이러한 목적을 위해 내부 알데하이드 포르밀인돌(Trilink) 그룹으로 용이하게 표지시킬 수 있다. 대안적으로, C-말단 아민에 대한 커플링보다는, 기록 태그를 대신 내부 라이신 잔기(바람직하게는 Lys-C 소화 후, 또는 대안적으로 Glu-C 소화 후)에 커플링시킬 수 있다. 하나의 시도에서, 이는 라이신 아민을 NHS-아지드(또는 NHS-PEG-아지드) 그룹으로 활성화시킨 후 5' 아민-표지된 기록 태그에 커플링하여 달성할 수 있다. 다른 시도에서, 5' 아민-표지된 기록 태그를 DSS와 같은, 과도한 NHS 단독-이작용성 가교결합제와 반응시켜, 5' NHS 활성화된 기록 태그를 생성할 수 있다. 이러한 5' NHS 활성화된 기록 태그는 펩타이드의 라이신 잔기의 ε-아미노 그룹에 직접 커플링할 수 있다.

[0274] 실시예 4: 펩타이드 상의 아미노산의 부위-특이적인 표지화

[0275] 활성화된 DNA 태그로 직접 변형될 수 있거나(헤테로이기능성 아민노산 부위-특이적인 시약을 사용한 활성화 사용) DNA 태그 상에 동종의 클릭 모이어티를 부착시키기 위해 후에 사용되는 클릭 모이어티를 사용하여 아미노산을 부위-특이적으로 표지하는 클릭 화학 헤테로이기능성 시약을 통해 간접적으로 변형시킬 수 있는 단백질 또는 펩타이드상의 아미노산의 5개의 상이한 예(Lundblad 2014). 전형적인 단백질 투입물은 0.1% RapiGest™ SF 계면활성제, 및 5 mM TCEP를 함유하는 50 μl의 적절한 수성 완충액 속에 1 μg 단백질을 포함한다. RapiGest™ SD는 표지화 또는 소화를 개선시키기 위해 단배질을 폴리펩타이드로 변성시키기 위한 산 분해가능한 표면활성제로서 유용하다. 다음의 아미노산 표지 전략을 사용할 수 있다: 말레이미드 화학을 사용한 시스테인 --- 200 μM 설포-SMCC-활성화된 DNA 태그를 사용하여 100 mM MES 완충액 (pH 6.5) + 1% TX-100 속에서 시스테인을 1시간 동안 부위-특이적으로 표지하고; NHS 화학을 사용한 라이신 --- 200 μM DSS 또는 BS^{3-}활성화된 DNA 태그를 사용하여 붕산염 완충액 (50 mM, pH 8.5) + 1% TX-100 속에서 용액 상 단백질 또는 비드-결합된 펩타이드 위에서 라이신을 1시간 동안 실온에서 부위-특이적으로 표지하고; 타이로신을 4-페닐-3H-1,2,4-트리아졸린-3,5(4H)-디온(PTAD)로 변형시키거나 디아조늄 화학 --- 디아조늄 화학의 경우, DNA 태그를 EDC 및 4-카복시벤젠 디아조늄 테트라플루오로보레이트(Aikon International, 중국)으로 활성화시킨다. 타이로신을 사용한 디아조 연결을 단백질 또는 비드-결합된 펩타이드를 200 μM 디아조늄-유도체화된 DNA 태그와 함께 붕산염 완충액(50 mM, pH 8.5) + 1% TX-100 속에서 1시간 동안 빙상에서 항온처리하여 생성시킨다(Nguyen, Cao et al. 2015). 아스파르테이트/글루타메이트를 EDC 화학을 사용하여 변형시킨다 --- 아민-표지된 DNA 태그를 비드-결합된 펩타이드 및 100 mM EDC/50 mM 이미다졸과 함께 pH 6.5 MES 속에서 1시간 동안 실온에서 항온처리한다(Basle et al., 2010, Chem. Biol. 17:213-227). 표지한 후, 과도한 활성화된 DNA 태그를 C4 수지 ZipTips(Millipore)로부터 단백질 결합 용출을 사용하여 제거한다. 용출된 단백질을 50 μl이하의 1X PBS 완충액에 둔다.

[0276] 실시예 5: 가닥화된 알킨 기록 태그-표지된 펩타이드의 아지드-활성화된 비드로의 고정화

[0277] 아지드-유도체화된 Dynabeads® M-270 비드는 시판되는 아민 Dynabeads® M-270을 아지드 PEG NHS 에스테르 헤테로이기능성 링커(JenKem Technology, 텍사스) 와 반응시켜 생성한다. 또한, 아지드의 표면 밀도를 메톡시 또는 하이드록시 PEG NHS 에스테르 속에서 적절한 비로 혼합하여 적정할 수 있다. 제공된 펩타이드 샘플의 경우, 1 내지 2 mg의 아지드-유도체화된 Dynabeads® M-270 비드 (~1.3 x 10^8 비드)를 100 μl의 붕산염 완충액(50 mM

붕산나트륨, pH 8.5) 속에 희석시키고, 1 ng 기록 태그-펩타이드를 가하고 1시간 동안 23 내지 37℃에서 항온처리한다. 200 μl의 붕산염 완충액으로 3회 세척한다.

[0278] 실시예 6: 포르밀인돌 반응성 HyNic 비드의 생성

[0279] 아민 비드의 HyNic 유도체화는 포르밀인돌 반응성 비드를 생성한다. 20 mg의 Dynabeads® M-270 아민 비드(2.8 μm)를 200 ul의 붕산염 완충액 속에 현탁시킨다. 약하게 초음파처리한 후, 1 내지 2 mg의 설포-S-HyNic(석신이미딜 6-하이드라지노니코티네이트 아세톤 하이드라존, SANH)(제품 번호 S-1002, Solulink, 샌 디에고)를 가하고 반응 혼합물을 1시간 동안 실온에서 진탕한다. 이후에, 비드를 붕산염 완충액으로 2회, 및 시트트산 완충액(200 mM 시트르산나트륨)으로 1회 세척한다. 비드를 스트르산 완충액 속에서 10 mg/ml의 최종 농도로 현탁시킨다.

[0280] 실시예 7: 활성화된 비드에 대한 기록 태그 포르밀인돌-표지된 펩타이드의 고정화

[0281] 1 내지 2 mg의 분취량의 HyNic 활성화된 Dynabeads® M-270 비드 (~1.3 x 10^8 비드)를 50 mM 아닐린이 보충된 100 μl의 시트르산염 완충액 속에 희석시키고, ~1 ng 기록 태그 펩타이드 접합체를 가하고 1시간 동안 37℃에서 항온처리한다. 비드를 200 μl의 스트르산염 완충액으로 3회 세척하고, 100 μl의 붕산염 완충액 속에 재-현탁시킨다.

[0282] 실시예 8: 올리고뉴클레오타이드 모델 시스템 - 사이클릭 양식으로 암호화 태그의 확인 정보의 기록 태그로의 이전에 의한 결합제 역사의 기록

[0283] 핵산 암호화 태그 및 기록 태그의 경우, 정보는 결합된 결합제 상의 암호화 태그로부터 근접한 기록 태그로 연결 또는 표준 핵산 효소학을 사용한 프라이머 연장에 의해 이전시킬 수 있다. 이는 결합제 표적을 나타내는 5' 부위, 및 기록 태그를 나타내는 3' 부위를 지닌 올리고뉴클레오타이드로 이루어진 단순한 모델 시스템을 사용하여 입증할 수 있다. 올리고뉴클레오타이드는 dT-알킨 변형(DBCO-dT, Glen Research)을 통하여 클릭 화학을 사용하여 내부 부위에서 고정화시킬 수 있다. 도 24a에 나타낸 실시예에서, 고정화된 올리고뉴클레오타이드(AB 표적)는 2개의 표적 결합 영역인, 표지된 A 및 B를 함유하며, 여기에 동종의 올리고뉴클레오타이드 "결합제"인, A 올리고 및 B 올리고가 결합될 수 있다. A 올리고 및 B 올리고뉴클레오타이드를 일반적인 스페이서(Sp)를 통해 상호작용하여 초기 프라이머 연장(또는 연결)을 개시하는 암호화 태그(서열 및 길이에 있어서 상이함)에 연결시킨다. Sp의 길이는 짧게(예컨대, 6-9개 염기) 유지하여 결합제 결합 동안에 비-특이적인 상호작용을 최소화한다. 이러한 특수 실시예에서, 암호화 태그의 길이를 설계하여 겔 분석에 의해 "B" 올리고 결합 사건(20개의 염기 암호기 서열)으로부터 "A" 올리고 결합 사건(10개의 암호기 서열)을 용이하게 구별한다.

[0284] PAGE 겔 상의 단순한 분석은 A 또는 B 암호화 태그 이전의 효능의 측정을 가능하도록 하며, 실험 매개변수의 용이한 최적화를 가능하도록 한다. C 및 D가 A 및 B와 상호작용하지 않는 상이한 하이브리드화 서열인 것을 제외하고는, AB 표적 서열 이외에, 유사한 올리고뉴클레오타이드 CD 표적 서열을 사용한다(참고: 도 24b). 또한, C 및 D는 서열 및 길이가 상이하며, 각각 30개 염기 DNA 코드 및 40개 염기 DNA 코드를 포함하는, 암호화 태그를 함유한다. 제2의 표적 서열, CD의 목적은 AB와 CD 표적 분자 사이의 교차 상호작용을 평가하는 것이다. 특이적인 하이브리드화를 고려할 때, CD 표적에 대한 연장된 기록 태그는 분자간 교차가 AB 표적에 결합된 올리고에 연결된 A 또는 B 암호화 태그 사이에서 일어나지 않는 한 A 또는 B 암호화 태그 정보를 함유하지 않는다. 유사하게 AB 표적에 대한 연장된 기록 태그는 C 또는 D 암호화 태그 정보를 함유하지 않는다. AB 및 CD 표적이 물리적으로 매우 근접한(즉, < 50 nm) 상황에서, 혼선(cross talk)이 존재할 가능성이 있다. 따라서, 표면 상의 표적 거대분자를 적절히 이격시키는 것이 중요하다.

[0285] 이러한 올리고뉴클레오타이드 모델 시스템은 결합제 역사의 기록능을 완전히 특성화할 수 있도록 한다. 도 25는 프라이머 연장보다는 연결을 통한 정보 이전을 나타낸다. 겔 상에서 초기 최적화한 후, 다양한 결합 및 검정 프로토콜을 수행하고 서열분석으로 평가한다. 유일한 분자 확인인자(UMI) 서열을 계수 목적으로 사용하며, 단일의 거대분자로부터 기원하는 판독물의 확인을 가능하도록 하고 원래의 샘플내 전체적인 총 거대분자 복잡성의 척도를 제공한다. 예시적인 역사적 결합 프로토콜은 다음을 포함한다: A-B-C-B-A,-B-A-A-B-A,-B-C-D-A-C 등. 수득되는 최종의 생성물은 다음과 같이 판독되어야 한다: 각각 UMI-Sp-A-Sp-B-Sp-B-Sp-A-Sp + UMI-Sp-C-Sp; UMI-Sp-A-Sp-B-Sp-A-Sp-A-Sp-B-Sp-A; UMI-A-Sp-B-Sp-A + UMI-Sp-C-Sp-D-Sp-C-Sp. 이러한 분석의 결과는 추가의 최적화를 허용한다.

[0286] 실시예 9: 올리고뉴클레오타이드-펩타이드 모델 시스템 - 사이클릭 양식으로 암호화 태그의 확인 정보의 기록 태그로의 이전에 의한 결합제 역사의 기록

[0287] 올리고뉴클레오타이드 모델 시스템을 입증한 후, 펩타이드 모델 시스템은 펩타이드 에피토프 태그를 예시적인 표적 올리고뉴클레오타이드 서열의 5' 말단에 접합시켜 올리고뉴클레오타이드 시스템으로부터 작제한다(도 26의 A 및 B). 예시적인 펩타이드 에피토프 태그는 다음을 포함한다: FLAG (DYKDDDDK)(서열 번호: 171), V5 (GKPIPNPLLGLDST)(서열 번호: 172), c-Myc (EQKLISEEDL)(서열 번호: 173), HA (YPYDVPDYA)(서열 번호: 174), V5 (GKPIPNPLLGLDST)(서열 번호: 175), StrepTag II (NWSHPQFEK)(서열 번호: 176) 등. 임의의 Cys- Ser-Gly 링커는 펩타이드 에피토프 태그의 올리고뉴클레오타이드로의 커플링을 위해 포함시킬 수 있다. 실시예 7의 AB 올리고뉴클레오타이드 주형을 A_올리고뉴클레오타이드-cMyc 펩타이드 작제물로 교체하고, 실시예 7의 CD 올리고뉴클레오타이드 주형을 C_올리고뉴클레오타이드-HA 펩타이드 작제물로 교체한다(참고: 도 26). A_올리고뉴클레오타이드-cMyc 펩타이드 작제물은 또한 CSG 링커 및 N-말단 포스포타이로신을 함유한다. 유사하게, 동종의 펩타이드 결합제, cMyc 항체 및 HA 항체를 B 올리고뉴클레오타이드 암호화 태그, 및 D 올리고뉴클레오타이드 암호화 태그로 각각 태그화한다. 포스포타이로신 특이적인 항체는 별도의 "E" 암호화 태그로 태그화한다. 이러한 방식으로, 펩타이드 모델 시스템은 올리고뉴클레오타이드 시스템과 병렬하여, 올리고 결합 및 항체 결합 둘 다가 이러한 모델 시스템에서 시험된다.

[0288] 항-c-myc 항체(2G8D5, 마우스 모노클로날, GenScript), 항-HA 항체(5E11D8, 마우스 모노클로날, GenScript), strep-Tag II 항체(5A9F9, 마우스 모노클로날, GenScript), 또는 항-FLAG 항체(5AE85, 마우스 모노클로날, GenScript)를 사용한 고정된 DNA-펩타이드 작제물의 항체 염색을 1X PBST(PBS + 0.1% 트윈 20) 속에서 0.1 - 1 μg/ml을 사용하여 수행한다. 항온처리는 전형적으로 실온에서 30분 동안 수행한다. 1X PBST 중 1% PVP을 사용한 표준 예비-차단, 및 염색 후 세척을 또한 수행한다. 항체 탈-염색은 고 염(1M NaCl), 및 저 pH(글리신, pH 2.5) 또는 고 pH(트리에틸아민, pH 11.5)로 세척함으로써 효과적으로 달성한다.

[0289] 표적 올리고뉴클레오타이드는 아지드 비드에 대한 부착을 위해 내부 알킨 표지를 함유하며, 5' 말단은 윌리암(Williams) 등의 문헌(2010, Curr Protoc Nucleic Acid Chem. Chapter 4:Unit 4.41)에 기술된 바와 같이 펩타이드의 C-말단 시스테인에 대한 SMCC-매개된 부착을 위해 아미노 그룹을 함유한다. 대안적으로, 표준 카보디이미드 커플링을 올리고뉴클레오타이드 및 펩타이드의 접합 반응에 사용한다(Lu et al., 2010, Bioconjug. Chem. 21:187-202). 이 경우, 과도한 올리고를 사용하여 카보디이미드 반응 및 최소화된 펩타이드-펩타이드 커플링을 유도한다. 접합 후, 최종 생성물을 절제 및 PAGE 겔로부터의 용출에 의해 정제한다.

[0290] 실시예 10: 기록 태그에 대한 DNA/PNA 암호화 태그 상보체의 연결을 통한 암호화 태그 이전

[0291] 암호화 태그를 기록 태그에 대한 연결을 통해 직접 또는 간접적으로 이전시켜 연장된 기록 태그를 생성한다. 하나의 시행에서, 암호화 태그의 어닐링된 상보체를 기록 태그에 연결한다(도 25). 이러한 암호화 태그 상보체는 핵산(DNA 또는 RNA), 펩타이드 핵산(PNA), 또는 성장하는 기록 태그에 연결될 수 있는 일부 다른 암호화 분자일 수 있다. 연결은 표준 ATP-의존성 및 NADH-의존성 리가제를 사용하는 DNA 및 RNA의 경우에 효소적일 수 있거나, 연결은 DNA/RNA 및 특히 펩타이드 핵산, PNA 둘 다의 경우에 화학적-매개될 수 있다.

[0292] DNA의 효소적 연결을 위해, 어닐링된 암호화 태그는 기록 태그의 3' 하이드록실에 연결시키기 위한 5' 포스페이트를 필요로 한다. 예시적인 효소적 연결 조건은 다음과 같다(Gunderson, Huang et al. 1998): 표준 T4 DNA 연결 반응은 다음을 포함한다: 50 mM 트리스-HCl(pH 7.8), 10 mM MgCl2, 10 mM DTT, 1 mM ATP, 50 μg/ml BSA, 100 mM NaCl, 0.1% TX-100 및 2.0 U/μl T4 DNA 리가제(New England Biolabs). 이. 콜라이(*E. coli*) DNA 리가제 반응은 40 mM 트리스-HCl(pH 8.0), 10 mM $MgCl_2$, 5 mM DTT, 0.5 mM NADH, 50 μg/ml BSA, 0.1% TX-100, 및 0.025 U/μl 이. 콜라이 DNA 리가제(Amersham)를 포함한다. 태그 DNA 연결 반응은 20 mM 트리스-HCl(pH 7.6), 25 mM 아세트산칼륨, 10 mM 아세트산마그네슘, 10 mM DTT, 1 mM NADH, 50 μg/ml BSA, 0.1% 트리톤 X-100, 10% PEG, 100 mM NaCl, 및 1.0 U/μl Taq DNA 리가제(New England Biolabs)를 포함한다. T4 및 이. 콜라이 DNA 리가제 반응은 실온에서 1시간 동안 수행하며 태크(Taq) DNA 리가제 반응은 40℃에서 1시간 동안 수행한다.

[0293] DNA/PNA의 주형의 화학적 연결의 몇가지 방법을 DNA/PNA 암호화 태그 이전에 사용할 수 있다. 이는 표준 화학적 연결 및 클릭 화학 시도를 포함한다. 주형 DNA 연결을 위한 예시적인 화학적 연결 조건은 다음과 같다(Gunderson, Huang et al. 1998): 주형 3' 포스페이트 리포터 태그의 5' 포스페이트 암호화 태그로의 연결은 1시간 동안 실온에서 다음으로 이루어진 반응 속에서 일어난다: 50 mM 2-[N- 모르폴리노]에탄설폰산(MES)(KOH를 사용하여 pH 6.0), 10 mM $MgCl_2$, 0.001% SDS, 새로이 제조된 200 mM EDC, 50 mM 이미다졸(HCl을 사용하여 pH 6.0) 또는 50 mM HOBt(HCl을 사용하여 pH 6.0) 및 3.0-4.0 M TMACl(Sigma).

[0294] PNA의 주형-의존성 연결의 예시적인 조건은 NH_2-PNA-CHO 중합체(예컨대, 암호화 태그 상보체 및 연장된 기록기

태그)를 포함하며 브루드노(Brudno) 등의 문헌(Brudno, Birnbaum et al. 2010)에 기술되어 있다. PNA는 5' 아민 등가물 및 3' 알데하이드 등가물을 가지며, 여기서 화학적 연결은 2개의 모이어티를 커플링하여 쉬프 염기(Schiff base)를 생성하고 이는 후속적으로 나트륨 시아노보로하이드라이드로 환원된다. 이러한 커플링을 위한 전형적인 반응 조건은 다음과 같다: 100 mM TAPS (pH 8.5), 80 mM NaCl, 및 80 mM 나트륨 시아노보로하이드라이드를 실온에서 60분 동안. 5' 아미노 말단 1,2-아미노티올 변형 및 3' C-말단 티오에스테르 변형을 함유하는 기능화된 PNA를 사용한 천연의 화학적 연결을 위한 예시적인 반응 조건은 롤로프(Roloff) 등의 문헌(2014, Methods Mol. Biol. 1050:131-141)에 기술되어 있다. 다른 N- 및 C- 말단 PNA 모이어티를 또한 연결에 사용할 수 있다. 다른 실시예는 클릭 화학을 사용한 PNA의 화학적 연결을 포함한다. 펭(Peng) 등의 시도(2010, European J. Org. Chem. 2010: 4194-4197)를 사용하여, PNA를 5' 아지드 및 3' 알킨으로 유도체화하고 클릭 화학을 사용하여 연결할 수 있다. "클릭" 화학적 연결을 위한 예시적인 반응 조건은 다음과 같다: 10 mM 인산칼륨 완충액, 100 mM KCl, 5 mM THPTA(트리스-하이드록시프로필 트리아졸 아민), 0.5 mM $CuSO_4$, 및 2.5 mM Na-아스코르베이트를 함유하는 100 μl의 반응 혼합물 속에 주형화된 PNA-PNA을 지닌 1 내지 2 mg의 비드. 화학적 연결 반응물을 실온에서 1시간 동안 항온처리한다. PNA 연결의 다른 예시적인 반응은 사쿠라이(Sakurai) 등의 문헌(Sakurai, Snyder et al. 2005)에 기술되어 있다.

[0295] 실시예 11: DNA로 PNA의 해독

[0296] PNA는 PNA 주형으로 어닐링된 DNA 올리고뉴클레오타이드의 클릭 화학-매개된 중합을 사용하여 DNA로 해독한다. DNA 올리고는 반응성 5' 아지드 및 3' 알킨을 함유함으로써 DNA 폴리머라제에 의해 복제될 수 있는 뉴클레오타이드-간 트리아졸 연결을 생성한다(El-Sagheer et al., 2011, Proc. Natl. Acad. Sci. USA 108:11338-11343). PNA 중 모든 가능한 암호화 태그에 대해 상보성인 완전한 세트의 DNA 올리고(1X 하이브리드화 완충액 중 10 nM: 10 mM Na-붕산염(pH 8.5), 0.2 M NaCl)을 고체-상 결합된 PNA 분자와 함께 30분 동안 항온처리(23 내지 50℃)한다. 어닐링 후, 고체-상 결합된 PNA-DNA 작제물을 나트륨 아스코르베이트 완충액(10 mM 아스코르브산나트륨, 200 mM NaCl)으로 1회 세척한다. '클릭 화학' 반응 조건은 다음과 같다: 비드 상의 PNA-DNA를 새로운 아스코르브산나트륨 완충액 속에서 항온처리하고 10 mM THPTA + 2 mM $CuSO_4$의 혼합물과 1:1로 합하고 1시간 동안 실온에서 항온처리한다. 비드를 이후에 하이브리드화 완충액으로 1회 및 PCR 완충액으로 2회 세척한다. 화학적 연결 후, 수득되는 연결된 DNA 생성물을 엘-새그히어(El-Sagheer) 등의 문헌(2011, Proc. Natl. Acad. Sci. USA 108:11338-11343)에 기술된 바와 같은 조건 하에서 PCR로 증폭한다.

[0297] 실시예 12: 핵산 기록 및 암호화 태그와 혼화성인 온화한 N-말단 에드만 분해

[0298] N-말단 에드만 분해와 DNA 암호화 사이의 상용성은 이러한 시도가 펩타이드 서열분석에 작동하도록 한다. 무수 TFA를 사용하는 N-말단 에드만 분해에 대한 표준 조건은 DNA를 파괴한다. 그러나, 이러한 효과는 보다 온화한 절단 조건을 개발하고 변형된 DNA를 보다 큰 산 내성으로 개발함으로써 경감시킨다. N-말단 에드만 분해를 위한 보다 온화한 조건은 페닐티오카바모일(PTC)-펩타이드의 절단 최적화 및 절단 조건 하에서 DNA/PNA 암호화된 라이브러리의 측정된 안정성의 절단 최적화의 조합을 사용하여 개발한다. 또한, 천연의 DNA를 저 pH에서 탈퓨린화를 감소시키는 7-데아자 푸린, 및 탈피리미딘화를 감소시키는 5' 메틸 변형된 사이토신과 같은 염기 변형을 사용함으로써 산 가수분해에 대해 안정화시킬 수 있다(Schneider and Chait, 1995, Nucleic Acids Res. 23:1570-1575). T가 풍부한 암호화 태그는 또한 티민이 산 분획화에 대해 가장 안정한 염기임을 고려할 때 유용할 수 있다. 온화한 N-말단 에드만 분해에 대한 조건은 무수 TFA 절단을 바렛(Barrett) 등의 문헌(1985, Tetrahedron Lett. 26:4375-4378, 이의 전문은 참고로 포함됨)에 기술된 바와 같이 60℃에서 아세토니트릴 중 트리에틸아민 아세테이트를 사용하여 온화한 10분 염기 절단으로 대체한다. 이러한 온화한 조건은 대부분의 유형의 DNA 기록 및 암호화 태그와 혼용성이다. 대안으로서, PNA는 완전하게 산-안정성이므로 암호화 태그에 사용된다(Ray and Norden, 2000, FASEB J. 14:1041-1060).

[0299] DNA 암호화 태그/기록 태그를 사용하여 NTAA 결합제를 확인하고 온화한 N-말단 에드만 분해를 수행하기 위한 혼용성은 다음 검정을 사용하여 입증된다. 항-포스포타이로신 및 항-cMyc 항체 둘 다를 사용하여 모델 펩타이드를 판독한다. C-Myc 및 N-말단 포스포타이로신 검출, 암호화 태그 기록, 및 단일의 에드만 분해 단계를 사용한 N-말단 포스포타이로신의 제거. 이러한 단계 후, 펩타이드를 항-포스포타이로신 및 항-cMyc 항체로 다시 염색한다. N-말단 분해에 대한 기록 태그의 안정성을 qPCR로 분석한다. 포스포타이로신의 효과적인 제거는 서열분석, qPCR, 또는 겔 전기영동에 의해 분석된 최종 기록 태그 서열내 E-올리고뉴클레오타이드 암호화 태그 정보의 부재에 의해 나타난다.

[0300] 실시예 13: 구획 태그된 비드의 제조.

[0301] 구획 태그된 비드의 제조를 위해, 바코드를 분할-및-혼주물 합성 시도를 사용하여, 포스포르아미디트 합성을 사용하여 또는 분할-및-혼주물 연결을 통해 비드에 고정된 올리고뉴클레오타이드내로 혼입시킨다. 구획 태그는 유일한 분자 확인인자(UMI)를 추가로 포함함으로써 구획 태그가 결합된 각각의 펩타이드 또는 단백질 분자를 유일하게 표지할 수 있다. 예시적인 구획 태그 서열은 다음과 같다: 5'-NH_2-GCGCAATCAG-XXXXXXXXXXXX-NNNNN-TGCAAGGAT-3'(서열 번호: 177). XXXXXXXXXXXX(서열 번호: 178) 바코드 서열은 비드 합성시 분할-혼주에 의해 생성된 비드당 핵염기 서열의 고정된 집단이며, 여기서 고정된 서열은 비드마다 상이하다. NNNNN(서열 번호: 179) 서열을 비드내에서 무작위 처리하여 후속적으로 이에 대해 결합된 펩타이드에 대한 유일한 분자 확인인자(UMI)로서 제공한다. 바코드 서열을 마코스코(Macosko) 등의 문헌(2015, Cell 161:1202-1214, 이의 전문은 참고로 포함됨)에 기술된 바와 같이 분할-및-혼주 시도를 사용하여 비드 상에서 합성할 수 있다. UMI 서열은 변성 염기 혼합물(각각의 커플링 단계에서 존재하는 모든 4개의 포스포르아미디트 염기의 혼합물)을 사용하여 올리고뉴클레오타이드를 합성함으로써 생성시킬 수 있다. 5'-NH_2를 석시닐 4-(N-말레이미도메틸)사이클로헥산-1-카복실레이트(SMCC)로 활성화시키고 N-말단으로부터 C-말단까지의 서열 "CGGSSGSNHV"(서열 번호: 180)을 지닌 부텔라제 I 펩타이드 기질을 함유하는 시스테인을 윌리암(Williams) 등의 문헌(2010, Curr Protoc Nucleic Acid Chem. Chapter 4:Unit 4.41)에 기술된 변형된 프로토콜을 사용하여 SMCC 활성화된 구획 태그된 비드에 커플링시킨다. 즉, 200 μl의 자기 비드(10 mg/ml)를 1.5 ml의 에펜도르프 튜브에 둔다. 1 ml의 커플링 완충액(5 mM EDTA, 0.01% 트윈 20, pH 7.4이 들어있는 100 mM KH_2PO_4 완충액, pH 7.2)을 튜브에 가하고 약하게 와동시킨다. 새로이 제조된 40 μl의 설포-SMCC(DMSO 중 50 mg/ml, ThermoFisher)를 자기 비드에 가하고 혼합한다. 반응물을 1시간 동안 실온에서 회전 혼합기 속에서 항온처리한다. 항온처리 후, 비드를 자석 위에서 상층액으로부터 분리하고, 500 μl의 커플링 완충액으로 3회 세척한다. 비드를 400 μl의 커플링 완충액 속에 재-현탁시킨다. 1 mL의 CGGSSGSNHV(서열 번호: 180) 펩타이드를 자기 비드에 가한다(TCEP-환원(5 mM) 후 커플링 완충액 중 1 mg/mL 및 빙냉 아세톤 침전). 반응물을 실온에서 회전 혼합기 상에서 2시간 동안 항온처리한다. 반응물을 커플링 완충액으로 1회 세척한다. 400 μl의 퀀칭(quenching) 완충액(10 mg/mL 머캅토석신산, pH 7.4이 들어있는 100 mM KH_2PO_4 완충액, pH 7.2)을 반응 혼합물에 가하고 2시간 동안 회전 혼합기 속에서 항온처리한다. 반응 혼합물을 커플링 완충액으로 3회 세척한다. 수득되는 비드를 저장 완충액(0.02% NaN_3, 0.01% 트윈 20, pH 7.4이 들어있는 10 mM KH_2PO_4 완충액, pH 7.2)속에 재-현탁시키고 4℃에서 저장한다.

[0302] 실시예 14: 캡슐화된 비드 및 단백질의 생성

[0303] 구획 태그된 비드 및 단백질을 엔도프로테이나제 AspN(Endo AspN)와 같은 아연 금속-엔도펩티다제, 임의의 광-케이지된 Zn 킬레이터(예컨대, ZincCleav I), 및 가공된 열-내성 부텔라제 I 동족체(Bandara, Kennedy et al. 2009, Bandara, Walsh et al. 2011, Cao, Nguyen et al. 2015)와 합한다. 실시예 12로부터의 구획 태그된 비드를 단백질과 혼합하고 T-이음 미세유동 또는 유동 포커싱 장치(flow focusing device)를 통해 유화시켰다(참고: 도 21). 2개의 수성 유동 구조에서, 하나의 유동물 중 단백질 및 Zn^{2+}을 다른 유동물로부터 금속-엔도펩티다제와 합하여 소적 형성 즉시 소화를 개시할 수 있다. 하나의 유동 구조에서, 모든 시약을 예비혼합하고 함께 유화시킨다. 이는 UV 광에 대한 노출을 통해 소적 형성 후 단백질 소화를 개시하기 위해 임의의 광-케이지된 Zn 킬레이터(예컨대, ZincCleav I)의 사용을 필요로 한다. 농도 및 유동 조건을 평균적으로, 소적당 1개 미만의 비드가 존재하도록 조정한다. 최적화된 실험에서, 10^8개의 펨토-소적을 비드를 함유하는 소적의 약 10%의 점유율로 제조할 수 있다(Shim et al., 2013,CS Nano 7:5955-5964). 하나의 유동 시도에서, 소적을 형성시킨 후, 프로테아제는 유화액을 UV-365 nm 광에 노출시킴으로써 광-케이지된 Zn^{2+}를 방출시켜 활성화하여, 엔도 AspN 프로테아제를 활성화시킨다. 유화액을 1시간 동안 37℃에서 항온처리하여 단백질을 펩타이드로 분해한다. 분해 후, Endo AspN을 유화액을 80℃에서 15분 동안 가열함으로써 불활성화시킨다. 2개의 유동 제형에서, Zn^{2+}을 2개의 유동물을 소적내로 조합시키는 동안 도입한다. 이 경우에, 엔도 AspN은 광-활성화된 Zn^{2+} 케이징 분자를 사용하여 불활성화시킬 수 있으며, 여기서 킬레이터는 UV 광에 노출시, 또는 양친매성 Zn^{2+} 킬레이트화제를 2-알킬말론산 또는 EDTA-MO와 같은 오일 상에 가함으로써 활성화시킨다. 양친매성 EDTA 분자의 예는 다음을 포함한다: EDTA-MO, EDTA-BO, EDTA-BP, DPTA-MO, DPTA-BO, DPTA-BP 등(Ojha, Singh et al. 2010, Moghaddam, de Campo et al. 2012). 다른 양식을 또한 사용하여 양친매성 산 또는 염기를 유화액 오일에 첨가함을 통해 소적의 pH를 변화시킴을 포함하여 소적 내부내에서 반응을 조절할 수 있다. 예를 들면, 소적 pH를 물/오일 가용성 아세트산을 사용

하여 강하시킬 수 있다. 아세트산의 플루오로-유화액으로의 첨가는 아세트산 분자의 양친매성 특성으로 인하여 소적 구획내에서 pH의 환원을 이끈다(Mashaghi and van Oijen, 2015, Sci Rep 5:11837). 유사하게, 염기, 프로필 아민의 첨가는 소적 내부를 알칼리성화한다. 유사한 시도를 오일/물 수용성 산화환원 시약(resox reagent), 환원제, 킬레이트화제 및 촉매와 같은 양친매성 분자의 다른 유형에 대해 사용할 수 있다.

[0304] 구획화된 단백질을 펩타이드로 분해한 후, 펩타이드를 비드 상에서 구획 태그(올리고뉴클레오타이드 펩타이드 바코드 키메라)에 부틸라제 I 또는 화학적 연결(예컨대, 알데하이드-아미노 등)을 사용하여 연결한다(참고: 도 16 및 도 22의 A). 임의의 시도에서, 올리고-티오뎁시펩타이드 "화학적 기질"을 사용하여 비가역성 부텔라제 I 연결을 제조한다(Nguyen, Cao et al. 2015). 연결 후, 유화액을 "크랙킹(cracking)"하고, 대량으로 수집된 고정된 구획 태그된 펩타이드 작제물을 지닌 비드, 또는 구획 태그된 펩타이드를 비드로부터 절단하고, 대량으로 수집한다. 비드 고정된 구획 태그된 펩타이드가 기록 태그를 포함하는 경우, 이들 비드는 본원에 기술된 핵산 암호화 기반 펩타이드 분석 방법에서 직접 사용할 수 있다. 대조적으로, 구획 태그된 펩타이드를 비드 기질로부터 절단하는 경우, 구획 태그된 펩타이드는 구획 태그된 펩타이드의 C-말단에 접합시킴으로써 기록 태그와 관련시키고, 본원에 기술된 바와 같이 암호화 태그된 결합제와 서열분석 분석을 위해 고체 지지체 상에 고정시킨다. 기록 태그와 구획 태그된 펩타이드의 연합은 3기능성 링커 분자를 사용하여 달성할 수 있다. 관련된 기록 태그를 지닌 구획 태그된 펩타이드를 사이클릭 서열분석 분석을 위해 고체 지지체에 고정시킨 후, 구획 정보를 프라이머 연장 또는 연결을 사용하여 관련된 기록 태그로 이전시킨다(참고: 도 22의 B). 구획 태그 정보를 기록 태그로 이전한 후, 구획 태그를 원래의 펩타이드 분해에 사용된 동일한 효소를 사용하여 펩타이드로부터 절단할 수 있다(참고: 도 22의 B). 이는 펩타이드의 원래의 N-말단 말단을 회복시킴으로써 본원에 기술된 바와 같은 N-말단 분해 펩타이드 서열분석 방법이 가능하도록 한다.

[0305] 실시예 15: 3 프라이머 융합 유화액 PCR을 통해 아미노산-특이적인 암호화 태그로 공유결합적으로 변형된 펩타이드의 기록 태그의 연합에 의한 디-태그 생성

[0306] 구획 태그 및 분자 UMI로 구성된 기록 태그를 지닌 펩타이드를 암호화 태그 부위-특이적인 화학적 표지로 화학적으로 변형시킨다. 암호화 태그는 또한 변형된 펩타이드내 제공된 유형의 다수의 아미노산을 계수할 수 있도록 하는 UMI를 함유한다. 타이슨(Tyson) 및 아모르(Armor)(Tyson and Armour 2012)로부터의 변형된 프로토콜을 사용하여, 유화액 PCR을 1x PHUSION™ GC 반응 완충액(Thermo Fisher Scientific), 200 μM의 각각의 dNTP(New England Biolabs), 1 μM 프라이머 U1, 1 μM 프라이머 U2tr, 25 nM 프라이머 Sp, 14단위의 PHUSION™ 고 충실성(fidelity) DNA 폴리머라제(Thermo Fisher Scientific)를 함유하는, 100μl의 총 수성 용적 속에서 제조한다. 10 μl의 수성 상을 5 내지 10초마다 터너(Turner) 및 후를레스(Hurles)가 이미 기술한 바와 같이(2009, Nat. Protoc. 4:1771-1783) 1000 rpm에서 총 5분 동안 교반하면서 2 ml의 냉-바이알(cryo-vial) 중 열은 광 오일(Sigma)) 속에 용해된 200 μl의 오일 상(4.5% vol./vol.) Span 80, 0.4% vol./vol. 트윈 80 및 0.05% 트리톤 X-100에 가한다. 수득되는 유화액의 평균 소적 크기는 약 5 마이크론이었다. T-이음 및 유동 포커싱의 사용과 같은, 다른 유화액 생성 방법을 또한 사용할 수 있다(Brouzes, Medkova et al. 2009). 유화액 생성 후, 100 μl의 수성/오일 혼합물을 0.5 ml의 PCR 튜브로 이전시키고 제1의 라운드의 증폭을 다음 조건에서 수행하였다: 98℃에서 30초 동안; 40주기의 98℃에서 10초, 70℃에서 30초 및 7℃에서 30초에 이어서; 72℃에서 5분 동안 연장. 제2의 라운드의 증폭 반응은 다음 조건에서 수행한다: 98℃에서 30초; 40 주기의 98℃에서 10초, 55℃에서 30초 및 72℃에서 30초에 이어서; 4℃에서 유지. 유화액을 최종 주기의 PCR 후 가능한 한 신속하게 200 μl의 헥산(Sigma)을 PCR 튜브에 직접 가하여 파괴하고, 20초 동안 와동시키고, 13,000 g에서 3분 동안 원심분리한다.

[0307] 실시예 16: 서열분석 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 작제물

[0308] 기록 태그 또는 암호화 태그의 스페이서 (Sp) 또는 보편적인 프라이밍 부위를 서열의 바디 중 3개의 염기만(예컨대, A, C, 및 T), 및 서열의 5' 말단에서 4번째 염기(예컨대, G)를 사용하여 설계할 수 있다. 합성에 의한 서열분석(SBS)을 위해, 이는 표준 다크(dark)(표지되지 않고 종결되지 않은) 뉴클레오타이드(dATP, dGTP, 및 dTTP) 및 단일의 ffC 염료-표지된 가역성 터미네이터(예컨대, 완전한 기능성 사이토신 트리포스페이트)의 혼합물을 사용하여 스페이서 서열을 따라 신속한 다크 염기 혼입을 가능하도록 한다. 이러한 방식으로, 관련된 암호기 서열, 유일한 분자 확인인자(들), 구획 태그, 연장된 기록기 태그의 결합 주기 서열, 연장된 암호화 태그, 또는 디-태그 만을 SBS 서열분석하고, 관련없는 스페이서 또는 보편적인 프라이밍 서열을 "거를(skipped over)"수 있다. 스페이서에 대한 염기 및 서열의 5' 말단에서 4번째 염기의 확인은 변할 수 있으며 상기 확인은 설명 목적으로만 제공된다.

[0309] 실시예 17: 단백질 용해물의 제조

[0310] 다양한 샘플 유형으로부터 단백질 용해물을 제조하기 위한 당해 분야에 공지된 광범위한 프로토콜이 존재한다. 프로토콜 상의 많은 변형은 세포 유형 및 용해물 속의 추출된 단백질이 비-변성된 또는 변성된 상태로 분석되어야 하는지의 여부에 의존한다. NGPA 검정의 경우, 천연의 구조 또는 변성된 단백질을 고체 기질에 고정시킬 수 있다(참고: 도 32). 더욱이, 천연 단백질이 고정 후, 기질의 표면에 고정된 단백질을 변성시킬 수 있다. 변성된 단백질을 사용하는 장점은 2배이다. 우선, 많은 항체 시약은 선형 에피토프(예컨대, 웨스턴 블롯 항체(Western Blot Abs))에 결합하며, 변성된 단백질은 선형 에피토프에 대해 보다 우수한 접근을 제공한다. 둘째로, NGPA 검정 작업흐름은 고정된 단백질이 이미 변성되므로 어닐링된 암호화 태그가 알칼리성(예컨대, 0.1 NaOH) 스트리핑 조건(stripping condition)을 사용하여 연장된 기록 태그로부터 스트리핑될 수 있으므로, 변성된 단백질을 사용하는 경우 단순화된다. 이는 어닐링된 암호화 태그의 효소적 제거에 이은 결합 사건 및 정보 이전을 필요로 하는, 이들의 천연 구조의 단백질을 포함하는 검정을 사용하는 어닐링된 암호화 태그의 제거와는 대조적이다.

[0311] 비-변성된 단백질 용해 완충액의 예는 다음을 포함한다: 50 mm HEPES(pH 7.4), 150 mM NaCl, 1% 트리톤 X-100, 1.5 mM MgCl2, 10% 글리콜; 및 M-PER 포유동물 단백질 추출 시약(Thermo-Fisher)과 같은 시판 완충액으로 이루어진 RPPA 완충액. 변성 용해 완충액은 50 mm HEPES (pH 8.), 1% SDS를 포함한다. 우레아(1M-3M) 또는 구아니딘 HCl(1-8M)의 첨가를 또한 단백질 샘플을 변성시키는데 사용할 수 있다. 용해 완충액의 상기 성분 외에, 프로테아제 및 포스파타제 억제제가 일반적으로 포함된다. 프로테아제 억제제의 예 및 전형적인 농도는 아프트로티틴(2 μg/ml), 류펩틴(5-10 μg/ml), 벤즈아미딘(15 μg/ml), 펩스타틴 A(1 μg/ml), PMSF(1 mM), EDTA(5 mM), 및 EGTA(1 mM)를 포함한다. 포스파타제 억제제의 예는 Na 피로포스페이트(10 mM), 불화나트륨(5-100 mM) 및 나트륨 오르토바나데이트(1mM)를 포함한다. 추가의 첨가제는 단백질 샘플로부터 DNA를 제거하기 위한 DNAaseI 및 이황화물 결합을 환원시키기 위한 DTT와 같은 환원제를 포함할 수 있다.

[0312] 조직 배양 세포로부터 제조된 비-변성 단백질 용해물 프로토콜의 예는 다음과 같다: 부착 세포를 트립신 처리(PBS 중 0.05% 트립신-EDTA)하고, 원심분리(5분 동안 200g)하여 수집하고, 빙냉 PBS 속에서 2회 세척한다. 프로테아제/포스파타제 억제제 및 첨가제(예컨대, EDTA가 없는 완전한 억제제(Roche) 및 PhosStop (Roche)가 보충된 빙-냉 M-PER 포유동물 추출 시약(10^7개 세포/100 mm의 디쉬 또는 150 cm^2의 플라스크 당 ~1 mL)을 가한다. 수득되는 세포 현탁액을 회전하는 진탕기 상에서 4℃에서 20분 동안 항온처리한 후, 4℃에서 ~12,000 rpm(세포 유형에 의존)에서 20분 동안 원심분리하여 단백질 상층액을 분리한다. 단백질을 BCA 검정을 사용하여 정량화하고, PBS 중 1 mg/ml에서 재현탁시킨다. 단백질 용해물을 사용하여 액체 질소 속에서 즉시 또는 스냅 동결시키고 -80℃에 저장한다.

[0313] 조직 배양 세포로부터 제조된, 휴스(Hughs) 등의 SP3 프로토콜을 기반으로 한, 변성 단백질 용해물 프로토콜의 예는 다음과 같다: 부착성 세포를 트립신처리(PBS중 0.05% 트립신-EDTA)하고, 원심분리(5분 동안 200g)하여 수집하고, 빙냉 PBS속에서 2회 세척한다. 프로테아제/포스파타제 억제제 및 첨가제(예컨대 1X cOmplete 프로테아제 억제제 칵테일(Roche))가 보충된 빙냉-변성된 용해 완충액(10^7개의 세포/100 mm의 디쉬 또는 150 cm^2의 플라스크당 ~1 mL)을 가한다. 수득되는 세포 현탁액을 95℃에서 5분 동안 항온처리하고, 빙상에 5분 동안 둔다. 벤조나제 뉴클레아제(500 U/ml)를 용해물에 가하고 37℃에서 30분 동안 항온처리하여 DNA 및 RNA를 제거한다.

[0314] 단백질을 100 uL의 용해물 당 5 μL의 200 mM DTT를 가하여 환원시키고 45℃에서 30분 동안 항온처리한다. 단백질 시스테인 그룹의 알킬화는 10 uL의 400 mM 요오도아세트아미드를 100 uL의 용해물당 가하여 달성하고 암실에서 24° 로 30분 동안 항온처리한다. 반응물을 10 uL의 200 mM DTT를 100 uL의 용해물당 가하여 퀀칭시킨다. 단백질을 2 ul의 산 무수물 및 100 ul의 1 M Na2CO3(pH 8.5)를 100 ul의 용해물 당 가하여 임의로 아실화한다. 실온에서 30분 동안 항온처리한다. 발레르산, 벤조산, 및 프로티온산 무수물이 "생체내(in vivo)" 아세틸화된 라이신을 아실화에 의한 라이신 그룹의 "반응계내" 차단으로부터 구별될 수 있도록 하기 위해 아세트산 무수물보다도 추천된다(Sidoli, Yuan et al. 2015). 반응물을 5 mg의 트리스(2-아미노에틸)아민, 중합체(Sigma)를 가하고 실온에서 30분 동안 항온처리하여 퀀칭시킨다. 중합체 수지를 용해물을 2000g에서 10분 동안 0.45 um 셀룰로즈 아세테이트 Spin-X 튜브(Corning)를 통해 원심분리하여 제거한다. 단백질을 BCA 검정을 사용하여 정량화하고, PBS 중 1 mg/ml에서 재현탁한다.

[0315] 추가의 예에서, 표지된 펩타이드는 에르데(Erde) 등이 기술한 바와 같이, 필터-보조된 샘플 제조(FASP) 프로토콜을 사용하여 생성하며, 여기서 MWCO 여과 장치를 단백질 포획, 알킬화, 및 펩티다제 소화를 위해 사용한다(Erde, Loo et al. 2014, Feist and Hummon 2015).

[0316] 실시예 18: 분배-태그된 펩타이드의 생성.

[0317] DNA 태그(임의의 샘플 바코드, 및 직각의 부착 모이어티를 지님)를 사용하여 표준 생접합 방법(Hermanson 2013)을 사용하여 변형된 폴리펩타이드의 라이신 상에서 ε-아미노 그룹을 표지시키는데 사용하거나, 대안적으로 벤조페논과 같은 광친화성 표지화(PAL) 방법(Li, Liu et al. 2013)을 사용하여 폴리펩타이드에 부착시킨다. 폴리펩타이드를 라이신 그룹에서 DNA 태그로 또는 CH 그룹 상에서 무작위로(PAL을 통해) 표지하고 표지되지 않은 그룹을 아실 무수물을 사용한 아실화를 통해 차단한 후, DNA-태그 표지되고, 아실화된 폴리펩타이드를 보편적인 프라이밍 서열, 구획 바코드, 임의의 UMI, 및 폴리펩타이드에 부착된 DNA 태그의 부위에 대해 상보성인 프라이머 서열을 포함하는 DNA 올리고뉴클레오타이드가 부착된 구획 비드에 어닐링한다. 다수의 DNA 하이브리드화 태그의 협동성으로 인하여, 단일의 폴리펩타이드 분자는 폴리펩타이드 분자의 모든 DNA 태그에 대해 동일한 구획 바코드의 쓰기를 가능하도록 하는 단일 비드와 주로 상호작용한다. 어닐링 후, 폴리펩타이드-결합된 DNA 태그는 어닐링된 비드-결합된 DNA 서열 상에서 폴리머라제 연장 반응을 프라이밍한다. 이러한 방식으로, 구획 바코드 및 다른 기능성 요소가 결합된 폴리펩타이드에 부착된 DNA 태그 상에 쓰여진다. 이러한 단계의 완료 후, 폴리펩타이드는 부착된 다수의 기록 태그를 가지며, 여기서 기록 태그는 일반적인 스페이서 서열, 바코드 서열(예컨대, 샘플, 분획, 구획, 공간 등), 임의의 UMI 및 다른 기능성 요소를 갖는다. 이러한 표지된 폴리펩타이드는 트립신, GluC, 프로테이나제 K, 등과 같은 표준 엔도프로테아제를 사용하여 펩타이드 단편으로 소화할 수 있다. 주석: 트립신이 라이신-표지된 폴리펩타이드의 소화에 사용되는 경우, 폴리펩타이드는 Arg 잔기에서만 절단되고 Lys 잔기에서는 절단되지 않는다(Lys 잔기가 표지되어 있기 때문임). 프로테아제 소화는 비드 상에서 직접 또는 바코드화된 비드로부터 표지된 폴리펩타이드의 제거 후 수행될 수 있다.

[0318] 실시예 19: 모델 시스템에 대한 DNA 기록 태그-펩타이드 접합체의 제조

[0319] 기록 태그 올리고뉴클레오타이드를 5' NH_2 그룹, 및 비드에 대한 이후 커플링을 위한 내부 m테트라진 그룹으로 합성한다(알킨-dT는 mTet-PEG-N_3 헤테로이기능성 가교결합제를 통해 m테트라진-dT로 전환된다). 올리고뉴클레오타이드의 5' NH_2를 펩타이드 상에서 반응성 시스테인에 LC-SMCC(ThermoFisher Scientific)와 같은 NHS/말레이미드 헤테로이기능성 가교결합제를 사용하여, 윌리엄스(Williams) 등이 기술한 바와 같이(Williams and Chaput 2010) 커플링한다. 특히, 20 nmol의 5' NH_2-표지된 올리고뉴클레오타이드는 에탄올 침전시키며 실리콘처리된 튜브 속에서 180 ul의 포스페이트 커플링 완충액(0.1 M 인산칼륨 완충액, pH 7.2) 속에 재현탁시킨다. 5 mg의 LC-SMCC를 1 mL의 DMF(5 mg/ml)(분취량으로 -20에 저장) 속에 재현탁시킨다. 20 ul의 LC-SMCC(5 mg/ml)의 분취량을 180 ul의 재현탁된 올리고뉴클레오타이드에 가하고, 혼합하고 실온에서 1시간 동안 항온처리한다. 혼합물은 2X 에탄올 침전시킨다. 수득된 말레이미드-유도체화된 올리고뉴클레오타이드를 200 ul의 포스페이트 커플링 완충액 속에 재현탁시킨다. 시스테인 잔기(>95% 순도, 탈염됨)를 함유하는 펩타이드를 DMSO 중 1 mg/ml(~0.5 mM)에서 재현탁시킨다. 대략 50 nmol의 펩타이드(100 ul)를 반응 혼합물에 가하고 실온에서 밤새 항온처리한다. 수득되는 DNA 기록 태그-펩타이드 접합체를 천연의-PAGE를 사용하여 윌리암(William) 등이 기술한 바와 같이(Williams and Chaput 2010) 정제한다. 접합체를 인산염 커플링 완충액 속에서 100 uM 농도로 실리콘처리된 튜브 속에서 재현탁시킨다.

[0320] 실시예 20: DNA-펩타이드 고정화를 위한 기질의 개발

[0321] 클릭-화학 고정화에 적합한 자기 비드는 M-270 아민 자기 다이나비드(Dynabead)를 아지드 또는 각각 알킨 또는 메틸 테트라진-표지된 올리고-펩타이드 접합체에 대해 커플링할 수 있는 TCO-유도체화된 비드로 전환시켜 생성한다(참고: 예컨대, 도 29의 D 및 E; 도 30의 D 및 E). 즉, 10 mg의 M-270 비드를 세척하고 500 ul의 붕산염 완충액(100 mM 붕산나트륨, pH 8.5) 속에 재현탁시킨다. TCO-PEG(12-120)-NHS(Nanocs) 및 메틸-PEG(12-120)-NHS의 혼합물을 DMSO 속에서 1 mM로 재현탁시키고 M-270 아민 비드와 함께 실온에서 밤새 항온처리한다. 메틸대 TCO PEG의 비를 적정하여 비드 상에서 최종 TCO 표면 밀도로 조정하여 < 100개의 TCO 모이어티/um^2가 존재하도록 한다(참고: 예컨대 도 31의 E; 도 34). 반응하지 않은 아민 그룹을 DMF(10 mg의 비드에 대해 500 ul) 중 0.1M 아세트산 무수물 및 0.1M DIEA의 혼합물로 실온에서 2시간 동안 캡핑한다. 캡핑 및 DMF 속에서 3회 세척 후, 비드를 인산염 커플링 완충액 속에서 10 mg/ml에서 재현탁시킨다.

[0322] 실시예 21: 기질에 대한 기록 태그 표지된 펩타이드의 고정화

[0323] 기록 태그 표지된 펩타이드를 기록 태그 상의 mTet 그룹 및 활성화된 비드 또는 기질의 표면 상의 TCO 그룹을 사용하는 IEDDA 클릭 화학 반응을 통해 기질 상에 고정시킨다. 이러한 반응은 반응물의 낮은 투입 농도에서조차

신속하고 효율적이다. 더욱이, 메틸 테트라진의 사용은 결합에 대해 보다 큰 안정성을 부여한다(Selvaraj and Fox 2013, Knall, Hollauf et al. 2014, Wu and Devaraj 2016). 200 ng의 M-270 TCO 비드를 100 ul의 인산염 커플링 완충액 속에 재현탁시킨다. 기록 태그 상에 mTet 모이어티를 포함하는 5 pmol의 DNA 기록 태그 표지된 펩타이드를 ~ 50 nM의 최종 농도를 위해 비드에 가한다. 반응물을 1시간 동안 실온에서 항온처리한다. 고정화 후, 기질 상의 반응하지 않은 TCO 그룹을 인산염 커플링 완충액 중 1 mM 메틸 테트라진 산으로 1시간 동안 실온에서 퀀칭시킨다.

[0324] 실시예 22: N-말단 아미노산(NTAA) 변형

[0325] 화학적 NTAA 아세틸화:

[0326] 펩타이드의 NTAA를 유기 용액 또는 수용액(설포-NHS-아세테이트) 중 아세트산 무수물 또는 NHS-아세테이트를 사용하여 아세틸화한다. 아세트산 무수물 유도체화를 위해, DMF 중 10 mM의 아세트산 무수물을 펩타이드와 함께 30분 동안 실온에서 항온처리한다(Halpin, Lee et al. 2004). 대안적으로, 펩타이드를 100 mM 2-(N-모르폴리노)에탄설포네이트(MES) 완충액(pH 6.0) 및 1M NaCl 중 50 mM 아세트산 무수물을 사용하여 수용액 속에서 실온으로 30분 동안 아세틸화한다(Tse, Snyder et al. 2008). NHS-아세테이트 유도체화를 위해, 설포-NHS-아세테이트(DMSO중 100 mM)의 스톡 용액을 제조하고 100 mM 인산나트륨 완충액(pH 8.0) 또는 100 mM 붕산염 완충액(pH 9.4) 중 5 내지 10 mM의 최종 농도에서 가하고 10 내지 30분 동안 실온에서 항온처리한다(Goodnow 2014).

[0327] 효소적 NTAA 아세틸화:

[0328] 펩타이드의 NTAA는 N-아세틸 트랜스퍼라제(설폴로부스 솔파타리쿠스(*Sulfolobus solfataricus*)로부터의 SsArd 1)에 노출시켜 다음 조건을 사용하여 효소적으로 아세틸화한다: 펩타이드를 2 μM SsArd1과 함께 NAT 완충액(20 mM 트리스-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM 아세틸-CoA) 속에서 65℃에서 10분 동안 항온처리한다(Chang and Hsu 2015).

[0329] 화학적 NTAA 아미드화(구아니딘화):

[0330] 펩타이드를 MDF 중 10 mM N,N-비스(3급-부톡시카보닐)티오우레아, 20 mM 트리메틸아민, 및 12 mM 무카야마 시약(Mukayama's regent)(2-클로로-1-메틸피리디늄 요오다이드)과 함께 실온에서 30분 동안 항온처리한다. 대안적으로, 펩타이드를 DMF중 10 mM 1H-피라졸-1-카복스아미딘 하이드로클로라이드, 10 mM DIEA와 함께 실온에서 30분 동안 항온처리한다. 표준 탈차단 방법을 사용하여 보호 그룹을 제거한다. 대안적으로, 펩타이드를 10 mM S-메틸 이소티오우레아와 함께 PBS 완충액(pH 8.0) 또는 100 mM 붕산염 완충액(pH 8.0) 속에서 30분 동안 10℃에서 항온처리한다(Tse, Snyder et al. 2008).

[0331] PITC 표지화:

[0332] 펩타이드를 5%(vol./vol.) PITC와 함께 이온성 액체[Bmim][BF4] 속에서 실온으로 5분 동안 항온처리한다. 반응 시간은 연장된 DNA 기록 태그 속에 존재하는 뉴클레오타이드 염기 상의 환외 아민(exocyclic amine)의 이소성 표지화(ectopic labeling)를 최소화하면서, NTAA의 정량적 PITC 표지화를 위해 최적화한다.

[0333] DNFB 표지화:

[0334] 2,4-디니트로플루오로벤젠(DNFB)을 5 mg/ml 스톡으로서 메탄올 속에서 제조한다. 용액을 광으로부터 보호하고 매일 새로이 제조한다. 펩타이드를 10 mM 붕산염 완충액(pH 8.0)중 0.5 -5.0 ug/ml DNFB 속에서 37℃로 5 내지 30분 동안 항온처리하여 표지한다.

[0335] SNFB 표지화:

[0336] 4-설포닐-2-니트로-플루오로벤젠(SNFB)을 메탄올 중 5 mg/ml 스톡으로서 제조한다. 용액을 광으로부터 보호하고 매일 새로이 제조한다. 펩타이드를 10 mM 붕산염 완충액(pH 8.0)중 0.5-5.0 ug/ml DNFB 속에서 37℃로 5 내지 30분 동안 항온처리하여 표지한다.

[0337] 아세틸화된 NTAA 펩타이드의 절단:

[0338] 아세틸화된 NTAA를 25 mM 트리스-HCl(pH 7.5) 중 10 uM 아실펩타이드 하이드롤라제(APH) 효소(설폴로부스 설파타리쿠스(Sulfolobus solfataricus)로부터, SSO2693)와 함께 90℃에서 10분 동안 항온처리하여 펩타이드로부터 절단한다(Gogliettino, Balestrieri et al. 2012).

[0339] 아미딘화된 NTAA 펩타이드의 절단:

[0340] 아미딘화(구아니딘화)된 NTAA를 0.1N NaOH 속에서 10분 동안 37℃에서 항온처리함으로서 펩타이드로부터 절단한다(Hamada 2016).

[0341] 실시예 23: 모델 시스템을 사용하여 기록 태그에 대한 암호화 태그 정보의 분자내 이전의 입증

[0342] DNA 모델 시스템을 사용하여 비드에 고정된 기록 태그에 대한 암호화 태그 정보의 "분가내" 이전을 시험하였다(참고: 도 36a). 2개의 상이한 유형의 기록 태그 올리고뉴클레오타이드를 사용하였다. saRT_Abc_v2(서열 번호: 141)는 "A" DNA 포획 서열(서열 번호: 153)("A'" 결합제에 대한 모사체 에피토프)을 함유하고 상응하는 "A" 바코드(rtA_BC); saRT_Bbc_V2(서열 번호: 142)는 "B" DNA 포획 서열(서열 번호: 154)("B'" 결합제에 대한 모사체 에피토프) 및 상응하는 "B" 바코드(rtB_BC)를 함유하였다. 이들 바코드는 요소적 65개 세트의 15-머 바코드(서열 번호: 1 내지 65)와 이들의 역 상보성 서열(서열 번호: 66 내지 130)의 조합이었다. rtA_BC는 2개의 바코드의 공선형 조합이고, BC_1 및 BC_2, 및 rtB_BC는 단지 하나의 바코드, BC_3이다. 유사하게, 암호화 태그 상의 바코드(암호기 서열)는 65개의 15-머 바코드(서열 번호: 1 내지 65)의 요소적 세트로부터의 바코드로 또한 구성되었다. CT_A'-bc_1PEG(서열 번호: 144) 및 CT_B'-bc(서열 번호: 147) 암호화 태그는 상보성 포획 서열, A' 및 B' 각각으로 구성되었고, 각각 15-머 바코드, BC_5, 및 BC_5 & BC_6으로 지정되었다. 기록 태그 및 암호화 태그에 대한 이러한 설계 설정은 겔 분석을 용이하게 한다. 바람직한 "분자내" 프라이머 연장은 유사한 크기의 올리고뉴클레오타이드 생성물을 생성하는 반면, 바람직하지 않은 "분자간" 연장은 "분자간" 생성물보다 15개 염기가 더 긴 하나의 올리고 생성물 및 15개 염기가 더 짧은 다른 생성물을 생성한다(도 36b).

[0343] "분자내" 대 "분자간" 정보 이전에 대한 기록 태그 밀도의 영향을 평가하였다. 정확한 정보 이전을 위해, "분자간" 정보 이전(A 기록 태그에 대한 "A'" 암호화 태그; B 기록 태그에 대한 B' 암호화 태그)는 "분자간" 정보 이전보다도 관찰되어야 한다(A 기록 태그에 대한 A' 암호화 태그 결합 그러나 B 기록 태그로의 정보 이전, 및 이의 역). 비드 표면에서 이격된 기록 태그의 효과를 시험하기 위해, 바이오티닐화된 기록 태그 올리고뉴클레오타이드, saRT_Abc_v2(서열 번호: 141) 및 saRT_Bbc_v2(서열 번호: 142)를 1:1의 비로 혼합한 후, saDummy-T10 올리고뉴클레오타이드(서열 번호: 143)에 대해 1:0, 1:10, $1:10^2$, $1:10^3$ 및 $1:10^4$의 비로 적정하였다. 총 20 pmol의 기록 태그 올리고뉴클레오타이드를 5 ul의 M270 스트렙타비딘 비드(Thermo)와 50 ul의 면역화 완충액(5 mM 트리스-Cl(pH 7.5), 0.5 mM EDTA, 1 M NaCl) 속에서 15분 동안 37℃에서 항온처리하였다. 비드를 100 ul의 고정화 완충액으로 실온에서 3회 세척하였다. 대부분의 후속된 세척 단계는 100 ul의 용적으로 사용하였다. 암호화 태그(DupCT 서열을 사용한 듀플렉스 어닐링은 이후 주기에 요구되었다)는 비드를 25 ul의 5X 어닐링 완충액(50 mM 트리스-Cl (pH 7.5), 10 mM MgCl2) 속에 재현탁시키고 암호화 태그 혼합물을 가함으로써 비드 상에 고정된 기록 태그에 어닐링하였다. 암호화 태그는 65℃에서 1분 동안 가열함으로써 기록 태그에 어닐링시킨 후, 실온으로 서서히 냉각되도록 하였다(0.2℃/초). 대안적으로, 암호화 태그를 PBST 완충액 속에서 37℃에서 어닐링시킬 수 있다. 비드를 PBST(PBS + 0.1% 트윈-20)로 실온에서 세척하고 PBST로 37℃에서 5분 동안 2회 세척하며, PBST로 실온에서 1회 세척하고 1X 어닐링 완충액으로 최종 세척하였다. 비드를 19.5 ul의 연장 완충액(50 mM 트리스-Cl (pH 7.5), 2 mM MgSO4, 125 uM dNTPs, 50 mM NaCl, 1 mM 디티오트레이톨, 0.1% 트윈-20, 및 0.1 mg/ml BSA) 속에 재현탁시키고 37℃에서 15분 동안 항온처리하였다. 클레노우 엑소-DNA 폴리머라제(NEB, 5 U/ul)를 비드에 0.125 U/ul의 최종 농도로 가하고, 37℃에서 5분 동안 항온처리하였다. 프라이머 연장 후, 비드를 PBST로 2회, 및 50 ul의 0.1 NaOH로 1회 실온에서 5분 동안, PBST로 3회 및 PBS로 1회 세척하였다. 하부 PCR 어댑터 서열, R1'을 가하기 위해, EndCap2T 올리고(R1(서열 번호: 152)으로 구성됨)를 하이브리드화하고 암호화 태그 올리고뉴클레오타이드에 대해 수행된 바와 같이 비드 상에서 연장시켰다. 어댑터 서열을 가한 후, 최종의 연장된 기록 태그 올리고뉴클레오타이드를 95% 포름아미드/10 mM EDTA 속에서 65℃에서 5분 동안 항온처리함으로써 스트렙타비딘 비드로부터 용출시켰다. 용출된 생성물의 대략 1/100번째를 20 ul 속에서 18 주기 동안 PCR 증폭시키고, 1 ul의 PCR 생성물을 10% 변형 PAGE 겔 위에서 분석하였다. 수득되는 겔은 폴리머라제 연장에 의한 기록 태그로 암호화 태그 정보의 쓰기 원리의 증거(도 36c), 및 비드의 표면에 기록 태그 밀도의 희석시 "분가간" 연장 사건에 대해 주로 "분가간" 연장 사건을 생성하는 능력을 입증한다.

[0344] 이러한 모델 시스템에서, 상응하는 암호기 서열 및 보편적인 역 프라이머 부위를 함유하는 기록 태그 RT_ABC 및 RT_BBC로부터의 PCR 생성물의 크기는 100개 염기쌍인 반면(도 36c), saRT_ABC(서열 번호: 141)/CT_B'BC(서열 번호: 147) 및 saRT_BBC(서열 번호: 142)/CT_A'BC(서열 번호: 144)의 부정확한 쌍화(pairing)에 의한 생성물은 각각 115개 및 85개 염기 쌍이다. 도 36d에 나타난 바와 같이, 비드 상에서 saRT_ABC(서열 번호: 141) 및 saRT_BBC(서열 번호: 142)의 고 밀도에서의 존재가 관찰되었다. 기록 태그는 고 밀도에서 자체(분자간 사건) 또

는 이웃하는 기록 태그(분자간 사건)에 대한 근접한 암호화 태그에서 연장하였음이 예측되었다. 그러나, 부정확한 쌍화에 의한 생성물의 밴드는 더미 올리고뉴클레오타이드 속에서 기록 태그를 희석시킴에 의해 감소하였으며, 1:10000의 비율로 사라졌다. 이러한 결과는 기록 태그가 낮은 밀도에서 비드 표면 위에 이격되어, 감소된 분자간 사건을 야기함을 입증하였다.

[0345] [표 1]

모델 시스템 서열

명칭	서열(5'-3')	서열 번호
saRT_Abc_V2	/5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAGTG	141
saRT_Bbc_V2	/5Biosg/TTTTTTTTTTGACTGGTTCCAATTGACAAGCCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG	142
saDummy-pT10	/5Biosg/TTTTTTTTTT/3SpC3/	143
CT_A'-bc	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/	144
CT_A'-bc_1PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	145
CT_A'bc_5PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18//iSP18//iSP18//iSP18//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	146
CT_B'bc	GCTTGTCAATTGGAACCAGTCTTTT/iSp18/CACTCAGTCCTAACGCGTATACGGGAATCTCGGCAGTTCACTCAGT/3SpC3/	147
EndCap2T	CGATTTGCAAGGATCACTCGTCACTCAGTCCTAACGCGTATACG/3SpC3/	148
Sp	ACTGAGTG	149
Sp'	CACTCAGT	150
P1_f2	CGTAGTCCGCGACACTAG	151
R1	CGATTTGCAAGGATCACTCG	152
dupCT_A'BC	CGTATACGCGTTAGGACTGAGTG/3SpC3/	153
dupCT_B'BC	AACTGCCGAGATTCCCGTATACGCGTTAGGACTGAGTG/3SpC3/	154

[0346]

[0347] /3SpC3/ = 3' C3 (3개 탄소) 스페이서

[0348] /5Biosg/ = 5' 바이오틴

[0349] /iSP18/ = 18-원자 헥사-에틸렌글리콜 스페이서

[0350] 실시예 24: 나노공극 서열에서 서열분석 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 작제물

[0351] DNA 바코드를 설계하여 나노공극-기반 서열분석기와 같은 고-오류 발생(highly-error prone) NGS 서열분석기에 대해 내성이 되도록 할 수 있으며, 여기서 전류 기반 콜 오류율(current base call error rate)은 10% 이상의 정도이다. 다수의 오류 정정 코드 시스템이 문헌에 기술되어 있다. 이들은 해밍 코드, 리드-솔로몬 코드, 및 레벤슈타인 코드, 리 코드 등을 포함한다. 오류-내성 바코드는 선택된 설계 매개변수에 따라서, 삽입, 결실, 및 치환을 수정할 수 있는 "DNA바코드"인, R 바이오컨덕터 패키지(R Bioconductor package)를 사용한 해밍 및 레벤슈타인 코드를 기반으로 하였다(Buschmann and Bystrykh 2013). 65개의 상이한 15-머 해밍 바코드는 도 27a(각각 서열 번호: 1 내지 65 및 서열 번호: 66 내지 130에서 이들의 역 상보성 서열에 제시된 바와 같음)에 나타낸다. 이들 바코드는 해밍 거리가 10이며 4개의 치환 오류 및 2개의 인델 오류(indel error)까지 자가-오류수정되어, 10% 오류율로 나노공극 서열분석기에서 정밀하게 판독되기에 매우 충분하다. 더욱이, 이들 바코드는 예측된 나노공극 전류 신호를 사용하여 77개의 원래의 바코드의 세트로부터 여과되었다(참고: 도 27b). 이들은 바코드를 가로질러서 큰 전류 수준 차이를 가지도록 여과되었으며, 이러한 세트내에서 다른 바코드를 사용하여 최대로 교정되지 않도록 여과되었다. 이러한 방식으로, 이들 바코드를 사용하는 검정으로부터 실제 가공되지 않은 나노공극 전류 수준의 플롯을 염기 콜링 알고리즘(base calling algorithm)을 사용하지 않고 예측된 바코드 신호에 직접 맵핑할 수 있다(Laszlo, Derrington et al. 2014).

[0352] 나노공극 서열분석을 사용하여 연장된 기록 태그, 연장된 암호화 태그, 또는 디-태그 작제물의 분석을 모사하기

위해, PCR 생성물은 4개의 전방 프라이머(DTF1 (서열 번호: 157), DTF2(서열 번호: 158), DTF3(서열 번호: 159), DTF4(서열 번호: 160)) 및 4개의 역방 프라이머(DTR9(서열 번호: 161), DTR10(서열 번호: 162), DTR11(서열 번호: 163), DTR12(서열 번호: 164))를 사용하는 15-머 바코드의 소 세트로 구성되었다(도 27c). 8개 프라이머의 이러한 세트는 플랭킹 전방 프라이머 F1(서열 번호: 165), 및 역방 프라이머 R1(서열 번호: 166)을 따라 PCR 반응에 포함되었다. DTF 및 DTR 프라이머는 상보성 15-머 스페이서 서열(Sp15)(서열 번호: 167)을 통해 어닐링되었다. 4개의 DTF 전방 및 4개의 DTR 역방 프라이머의 조합은 16개의 가능한 PCR 생성물의 세트를 유도한다.

PCR 조건:

시약	최종 농도
F1 (5' 포스포릴화된)(서열 번호: 165)	1 uM
R1(5' 포스포릴화된)(서열 번호: 166)	1 uM
DTF1-4 (서열 번호: 157-160); DTR9-12 (서열 번호: 161-164)	0.3 nM ea
VeraSeq 완충액 2	1X
dNTPs	200 uM
물	
VeraSeq 2.0 Ultra Pol	2 U/100 ul

PCR 주기화:

98 ℃	30 초
50 ℃	2 분
98 ℃	10 초
55 ℃	15 초
72 ℃	15 초
19주기 동안 마지막 3개 단계 반복	
72 ℃	5 분

[0353]

[0354] PCR 후에, 앰플리콘은 다음과 같이 평활 말단 연결(도 27c)에 의해 콘카테네이트되었다: 20 ul의 PCR 생성물을 20 ul의 퀵 리가제 혼합물(Quick Ligase Mix)(NEB)과 직접 혼합하고 밤새 실온에서 항온처리하였다. 길이가 ~0.5 내지 2 kb인 수득되는 연결된 생성물을 자이모 정제 컬럼(Zymo purification column)을 사용하여 정제하고 20 ul의 물 속에 용출시켰다. 약 7 ul의 이러한 정제된 연결 생성물을 MinIon Library Rapid Sequencing Prep kit(SQK-RAD002)에 직접 사용하고 MinION Mk 1B(R9.4) 장치에서 분석하였다. 품질 점수가 7.2인 734 bp의 나노공극 판독물의 예(~80%의 정확도)는 도 27d에 나타낸다. 불량한 서열분석 정확도에도 불구하고, 다수의 바코드가 MinIon 서열 판독물에 대한 바코드의 *lalign-계* 정렬에 의해 나타난 바와 같이 서열내에서 용이하게 판독될 수 있다(도 27d).

[0355] 실시예 25: 겔 비드내에서 캡슐화된 단일 세포

[0356] 단일 세포를 표준 기술(Tamminen and Virta 2015, Spencer, Tamminen et al. 2016)을 사용하여 소적(~50 μm) 내로 캡슐화한다(참고: 도 38). 폴리아크릴아미드(아크릴아미드:비스아크릴아미드(29:1)(30% w/vol.)), 벤조페논 메타크릴아미드(BM), 및 APS가 세포와 함께 불연속 상에 포함되어 연속된 오일 상(소적내로 확산됨)으로 TEMED의 첨가시 중합할 수 있는 소적을 생성한다. 벤조페논은 폴리아크릴아미드 겔 소적의 매트릭스내로 가교결합된다. 이는 폴리아크릴아미드 매트릭스로의 단백질의 후속적인 광친화성 가교결합을 허용한다(Hughes, Spelke et al. 2014, Kang, Yamauchi et al. 2016). 수득되는 단일의 세포 겔 비드내에서 고정된 단백질은 다양한 방법을 사용하여 바코드화된 단일 세포일 수 있다. 일 구현예에서, DNA 태그는 앞서 기술한 바와 같이 아민-반응성 제제 또는 광-활성 벤조페논 DNA 태그를 사용하여 단일 세포 겔 비드내에서 고정된 단백질에 화학적으로 또는 광-화학적으로 부착된다. 단일 세포 겔 비드는 앞서 기술한 바와 같이 바코드화된 비드 및, 단백질로 이전된 DNA 바코드 태그의 동시-캡슐화를 통해 바코드를 함유하는 소적내에서 캡슐화될 수 있거나, 대안적으로 단일 세포 겔 비드내의 단백질은 Amini, Cusanovich,d Gunderson et al. (Amini, Pushkarev et al. 2014, Cusanovich, Daza et al. 2015)(Gunderson, Steemers et al. 2016)에 기술된 바와 같이 일련의 혼주-및-분할 단계를 통해 조합적으로 인덱싱(indexing)될 수 있다. 가장 간단한 실행에서, 단일 세포 겔 비드내 단백질은 "클릭-화학" 모이어티로 우선 표지된 후(참고: 도 40), 조합 DNA 바코드가 혼주-및-분할 시도를 사용하여 단백질

샘플 상으로 클릭된다.

[0357] 참고 문헌:

Harlow, Ed, and David Lane. Using Antibodies. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1999.
Hennessy BT, Lu Y, Gonzalez-Angulo AM, et al. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clinical proteomics*. 2010;6(4):129-151.
Davidson, G. R., S. D. Armstrong and R. J. Beynon (2011). "Positional proteomics at the N-terminus as a means of proteome simplification." Methods Mol Biol **753**: 229-242.
Zhang, L., Luo, S., and Zhang, B. (2016). The use of lectin microarray for assessing glycosylation of therapeutic proteins. mAbs *8*, 524–535.
Akbani, R., K. F. Becker, N. Carragher, T. Goldstein, L. de Koning, U. Korf, L. Liotta, G. B. Mills, S. S. Nishizuka, M. Pawlak, E. F. Petricoin, 3rd, H. B. Pollard, B. Serrels and J. Zhu (2014). "Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society." Mol Cell Proteomics **13**(7): 1625-1643.

[0358]

Amini, S., D. Pushkarev, L. Christiansen, E. Kostem, T. Royce, C. Turk, N. Pignatelli, A. Adey, J. O. Kitzman, K. Vijayan, M. Ronaghi, J. Shendure, K. L. Gunderson and F. J. Steemers (2014). "Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing." Nat Genet **46**(12): 1343-1349.
Assadi, M., J. Lamerz, T. Jarutat, A. Farfsing, H. Paul, B. Gierke, E. Breitinger, M. F. Templin, L. Essioux, S. Arbogast, M. Venturi, M. Pawlak, H. Langen and T. Schindler (2013). "Multiple protein analysis of formalin-fixed and paraffin-embedded tissue samples with reverse phase protein arrays." Mol Cell Proteomics **12**(9): 2615-2622.
Bailey, J. M. and J. E. Shively (1990). "Carboxy-terminal sequencing: formation and hydrolysis of C-terminal peptidylthiohydantoins." Biochemistry **29**(12): 3145-3156.
Bandara, H. M., D. P. Kennedy, E. Akin, C. D. Incarvito and S. C. Burdette (2009). "Photoinduced release of Zn2+ with ZinCleav-1: a nitrobenzyl-based caged complex." Inorg Chem **48**(17): 8445-8455.
Bandara, H. M., T. P. Walsh and S. C. Burdette (2011). "A Second-generation photocage for Zn2+ inspired by TPEN: characterization and insight into the uncaging quantum yields of ZinCleav chelators." Chemistry **17**(14): 3932-3941.
Basle, E., N. Joubert and M. Pucheault (2010). "Protein chemical modification on endogenous amino acids." Chem Biol **17**(3): 213-227.
Bilgicer, B., S. W. Thomas, 3rd, B. F. Shaw, G. K. Kaufman, V. M. Krishnamurthy, L. A. Estroff, J. Yang and G. M. Whitesides (2009). "A non-chromatographic method for the purification of a bivalently active monoclonal IgG antibody from biological fluids." J Am Chem Soc **131**(26): 9361-9367.
Bochman, M. L., K. Paeschke and V. A. Zakian (2012). "DNA secondary structures: stability and function of G-quadruplex structures." Nat Rev Genet **13**(11): 770-780.
Borgo, B. and J. J. Havranek (2014). "Motif-directed redesign of enzyme specificity." Protein Sci **23**(3): 312-320.
Brouzes, E., M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J. B. Hutchison, J. M. Rothberg, D. R. Link, N. Perrimon and M. L. Samuels (2009). "Droplet microfluidic technology for single-cell high-throughput screening." Proc Natl Acad Sci U S A **106**(34): 14195-14200.
Brudno, Y., M. E. Birnbaum, R. E. Kleiner and D. R. Liu (2010). "An in vitro translation, selection and amplification system for peptide nucleic acids." Nat Chem Biol **6**(2): 148-155.
Calcagno, S. and C. D. Klein (2016). "N-Terminal methionine processing by the zinc-activated Plasmodium falciparum methionine aminopeptidase 1b." Appl Microbiol Biotechnol.
Cao, Y., G. K. Nguyen, J. P. Tam and C. F. Liu (2015). "Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation." Chem Commun (Camb) **51**(97): 17289-17292.
Carty, R. P. and C. H. Hirs (1968). "Modification of bovine pancreatic ribonuclease A with 4-sulfonyloxy-2-nitrofluorobenzene. Isolation and identification of modified proteins." J Biol Chem **243**(20): 5244-5253.
Chang, L., D. M. Rissin, D. R. Fournier, T. Piech, P. P. Patel, D. H. Wilson and D. C. Duffy (2012). "Single molecule enzyme-linked immunosorbent assays: theoretical considerations." J Immunol Methods **378**(1-2): 102-115.
Chang, Y. Y. and C. H. Hsu (2015). "Structural basis for substrate-specific acetylation of Nalpha-acetyltransferase Ard1 from Sulfolobus solfataricus." Sci Rep **5**: 8673.

[0359]

Christoforou, A., C. M. Mulvey, L. M. Breckels, A. Geladaki, T. Hurrell, P. C. Hayward, T. Naake, L. Gatto, R. Viner, A. Martinez Arias and K. S. Lilley (2016). "A draft map of the mouse pluripotent stem cell spatial proteome." Nat Commun 7: 8992.
Creighton, C. J. and S. Huang (2015). "Reverse phase protein arrays in signaling pathways: a data integration perspective." Drug Des Devel Ther **9**: 3519-3527.
Crosetto, N., M. Bienko and A. van Oudenaarden (2015). "Spatially resolved transcriptomics and beyond." Nat Rev Genet **16**(1): 57-66.
Cusanovich, D. A., R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell and J. Shendure (2015). "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing." Science **348**(6237): 910-914.
Derrington, I. M., T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis and J. H. Gundlach (2010). "Nanopore DNA sequencing with MspA." Proc Natl Acad Sci U S A **107**(37): 16060-16065.
El-Sagheer, A. H., V. V. Cheong and T. Brown (2011). "Rapid chemical ligation of oligonucleotides by the Diels-Alder reaction." Org Biomol Chem **9**(1): 232-235.
El-Sagheer, A. H., A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown (2011). "Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli." Proc Natl Acad Sci U S A **108**(28): 11338-11343.
Emili, A., M. McLaughlin, K. Zagorovsky, J. B. Olsen, W. C. W. Chan and S. S. Sidhu (2017). Protein Sequencing Method and Reagents. USPTO. USA, The Governing Council of University of Toronto. **9,566,335 B1**.
Erde, J., R. R. Loo and J. A. Loo (2014). "Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments." J Proteome Res **13**(4): 1885-1895.
Farries, T. C., A. Harris, A. D. Auffret and A. Aitken (1991). "Removal of N-acetyl groups from blocked peptides with acylpeptide hydrolase. Stabilization of the enzyme and its application to protein sequencing." Eur J Biochem **196**(3): 679-685.
Feist, P. and A. B. Hummon (2015). "Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples." Int J Mol Sci **16**(2): 3537-3563.
Friedmann, D. R. and R. Marmorstein (2013). "Structure and mechanism of non-histone protein acetyltransferase enzymes." FEBS J **280**(22): 5570-5581.
Frokjaer, S. and D. E. Otzen (2005). "Protein drug stability: a formulation challenge." Nat Rev Drug Discov **4**(4): 298-306.
Fujii, Y., M. Kaneko, M. Neyazaki, T. Nogi, Y. Kato and J. Takagi (2014). "PA tag: a versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin." Protein Expr Purif **95**: 240-247.
Gebauer, M. and A. Skerra (2012). "Anticalins small engineered binding proteins based on the lipocalin scaffold." Methods Enzymol **503**: 157-188.
Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany and F. Barany (1999). "Universal DNA microarray method for multiplex detection of low abundance point mutations." J Mol Biol **292**(2): 251-262.
Gogliettino, M., M. Balestrieri, E. Cocca, S. Mucerino, M. Rossi, M. Petrillo, E. Mazzella and G. Palmieri (2012). "Identification and characterisation of a novel acylpeptide hydrolase from Sulfolobus solfataricus: structural and functional insights." PLoS One 7(5): e37921.

[0360]

Gogliettino, M., A. Riccio, M. Balestrieri, E. Cocca, A. Facchiano, T. M. D'Arco, C. Tesoro, M. Rossi and G. Palmieri (2014). "A novel class of bifunctional acylpeptide hydrolases--potential role in the antioxidant defense systems of the Antarctic fish Trematomus bernacchii." FEBS J **281**(1): 401-415.
Granvogl, B., M. Ploscher and L. A. Eichacker (2007). "Sample preparation by in-gel digestion for mass spectrometry-based proteomics." Anal Bioanal Chem **389**(4): 991-1002.
Gunderson, K. L., X. C. Huang, M. S. Morris, R. J. Lipshutz, D. J. Lockhart and M. S. Chee (1998). "Mutation detection by ligation to complete n-mer DNA arrays." Genome Res **8**(11): 1142-1153.
Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). Methods and Compositions for Analyzing Cellular Components. WIPO, Illumina, Inc.
Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). Methods and compositions for analyzing cellular components, Illumina, Inc.
Guo, H., W. Liu, Z. Ju, P. Tamboli, E. Jonasch, G. B. Mills, Y. Lu, B. T. Hennessy and D. Tsavachidou (2012). "An efficient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for reverse phase protein arrays." Proteome Sci **10**(1): 56.
Hamada, Y. (2016). "A novel N-terminal degradation reaction of peptides via N-amidination." Bioorg Med Chem Lett **26**(7): 1690-1695.
Hermanson, G. (2013). Bioconjugation Techniques, Academic Press.
Hernandez-Moreno, A. V., F. Villasenor, E. Medina-Rivero, N. O. Perez, L. F. Flores-Ortiz, G. Saab-Rincon and G. Luna-Barcenas (2014). "Kinetics and conformational stability studies of recombinant leucine aminopeptidase." Int J Biol Macromol **64**: 306-312.
Hori, M., H. Fukano and Y. Suzuki (2007). "Uniform amplification of multiple DNAs by emulsion PCR." Biochem Biophys Res Commun **352**(2): 323-328.
Horisawa, K. (2014). "Specific and quantitative labeling of biomolecules using click chemistry." Front Physiol **5**: 457.
Hoshika, S., F. Chen, N. A. Leal and S. A. Benner (2010). "Artificial genetic systems: self-avoiding DNA in PCR and multiplexed PCR." Angew Chem Int Ed Engl **49**(32): 5554-5557.
Hughes, A. J., D. P. Spelke, Z. Xu, C. C. Kang, D. V. Schaffer and A. E. Herr (2014). "Single-cell western blotting." Nat Methods **11**(7): 749-755.
Hughes, C. S., S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz and J. Krijgsveld (2014). "Ultrasensitive proteome analysis using paramagnetic bead technology." Mol Syst Biol **10**: 757.
Kang, C. C., K. A. Yamauchi, J. Vlassakis, E. Sinkala, T. A. Duncombe and A. E. Herr (2016). "Single cell-resolution western blotting." Nat Protoc **11**(8): 1508-1530.
Kang, T. S., L. Wang, C. N. Sarkissian, A. Gamez, C. R. Scriver and R. C. Stevens (2010). "Converting an injectable protein therapeutic into an oral form: phenylalanine ammonia lyase for phenylketonuria." Mol Genet Metab **99**(1): 4-9.
Katritzky, A. R. and B. V. Rogovoy (2005). "Recent developments in guanylating agents." ARKIVOC **iv**(Issue in Honor of Prof. Nikolai Zefirov): 49-87.
Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner (2015). "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." Cell **161**(5): 1187-1201.

[0361]

Knall, A. C., M. Hollauf and C. Slugovc (2014). "Kinetic studies of inverse electron demand Diels-Alder reactions (iEDDA) of norbornenes and 3,6-dipyridin-2-yl-1,2,4,5-tetrazine." Tetrahedron Lett **55**(34): 4763-4766.
Le, Z. G., Z. C. Chen, Y. Hu and Q. G. Zheng (2005). "Organic Reactions in Ionic Liquids: Ionic Liquid-promoted Efficient Synthesis of Disubstituted and Trisubstituted Thioureas Derivatives." Chinese Chemical Letters **16**(2): 201-204.
Lesch, V., A. Heuer, V. A. Tatsis, C. Holm and J. Smiatek (2015). "Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants?" Phys Chem Chem Phys **17**(39): 26049-26053.
Li, G., Y. Liu, Y. Liu, L. Chen, S. Wu, Y. Liu and X. Li (2013). "Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry." Angew Chem Int Ed Engl **52**(36): 9544-9549.
Litovchick, A., M. A. Clark and A. D. Keefe (2014). "Universal strategies for the DNA-encoding of libraries of small molecules using the chemical ligation of oligonucleotide tags." Artif DNA PNA XNA **5**(1): e27896.
Liu, Y. and S. Liang (2001). "Chemical carboxyl-terminal sequence analysis of peptides and proteins using tribenzylsilyl isothiocyanate." J Protein Chem **20**(7): 535-541.
Lundblad, R. L. (2014). Chemical reagents for protein modification. Boca Raton, CRC Press, Taylor & Francis Group.
Mashaghi, S. and A. M. van Oijen (2015). "External control of reactions in microdroplets." Sci Rep **5**: 11837.
McCormick, R. M. (1989). "A solid-phase extraction procedure for DNA purification." Anal Biochem **181**(1): 66-74.
Mendoza, V. L. and R. W. Vachet (2009). "Probing protein structure by amino acid-specific covalent labeling and mass spectrometry." Mass Spectrom Rev **28**(5): 785-815.
Mikami, T., T. Takao, K. Yanagi and H. Nakazawa (2012). "N (alpha) Selective Acetylation of Peptides." Mass Spectrom (Tokyo) **1**(2): A0010.
Moghaddam, M. J., L. de Campo, N. Kirby and C. J. Drummond (2012). "Chelating DTPA amphiphiles: ion-tunable self-assembly structures and gadolinium complexes." Phys Chem Chem Phys **14**(37): 12854-12862.
Mukherjee, S., M. Ura, R. J. Hoey and A. A. Kossiakoff (2015). "A New Versatile Immobilization Tag Based on the Ultra High Affinity and Reversibility of the Calmodulin-Calmodulin Binding Peptide Interaction." J Mol Biol **427**(16): 2707-2725.
Namimatsu, S., M. Ghazizadeh and Y. Sugisaki (2005). "Reversing the effects of formalin fixation with citraconic anhydride and heat: a universal antigen retrieval method." J Histochem Cytochem **53**(1): 3-11.
Nguyen, G. K., Y. Cao, W. Wang, C. F. Liu and J. P. Tam (2015). "Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide." Angew Chem Int Ed Engl **54**(52): 15694-15698.
Nguyen, G. K., S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam (2014). "Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis." Nat Chem Biol **10**(9): 732-738.
Nishizuka, S. S. and G. B. Mills (2016). "New era of integrated cancer biomarker discovery using reverse-phase protein arrays." Drug Metab Pharmacokinet **31**(1): 35-45.
Ohkubo, A., R. Kasuya, K. Sakamoto, K. Miyata, H. Taguchi, H. Nagasawa, T. Tsukahara, T. Watanobe, Y. Maki, K. Seio and M. Sekine (2008). "'Protected DNA

[0362]

Probes' capable of strong hybridization without removal of base protecting groups." Nucleic Acids Res **36**(6): 1952-1964.
Ojha, B., A. K. Singh, M. D. Adhikari, A. Ramesh and G. Das (2010). "2-Alkylmalonic acid: amphiphilic chelator and a potent inhibitor of metalloenzyme." J Phys Chem B **114**(33): 10835-10842.
Peng, X., H. Li and M. Seidman (2010). "A Template-Mediated Click-Click Reaction: PNA-DNA, PNA-PNA (or Peptide) Ligation, and Single Nucleotide Discrimination." European J Org Chem **2010**(22): 4194-4197.
Perbandt, M., O. Bruns, M. Vallazza, T. Lamla, C. Betzel and V. A. Erdmann (2007). "High resolution structure of streptavidin in complex with a novel high affinity peptide tag mimicking the biotin binding motif." Proteins **67**(4): 1147-1153.
Rauth, S., D. Hinz, M. Borger, M. Uhrig, M. Mayhaus, M. Riemenschneider and A. Skerra (2016). "High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide." Biochem J **473**(11): 1563-1578.
Ray, A. and B. Norden (2000). "Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future." FASEB J **14**(9): 1041-1060.
Riley, N. M., A. S. Hebert and J. J. Coon (2016). "Proteomics Moves into the Fast Lane." Cell Syst **2**(3): 142-143.
Roloff, A., S. Ficht, C. Dose and O. Seitz (2014). "DNA-templated native chemical ligation of functionalized peptide nucleic acids: a versatile tool for single base-specific detection of nucleic acids." Methods Mol Biol **1050**: 131-141.
Roloff, A. and O. Seitz (2013). "The role of reactivity in DNA templated native chemical PNA ligation during PCR." Bioorg Med Chem **21**(12): 3458-3464.
Sakurai, K., T. M. Snyder and D. R. Liu (2005). "DNA-templated functional group transformations enable sequence-programmed synthesis using small-molecule reagents." J Am Chem Soc **127**(6): 1660-1661.
Schneider, K. and B. T. Chait (1995). "Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry." Nucleic Acids Res **23**(9): 1570-1575.
Selvaraj, R. and J. M. Fox (2013). "trans-Cyclooctene--a stable, voracious dienophile for bioorthogonal labeling." Curr Opin Chem Biol **17**(5): 753-760.
Sharma, A. K., A. D. Kent and J. M. Heemstra (2012). "Enzyme-linked small-molecule detection using split aptamer ligation." Anal Chem **84**(14): 6104-6109.
Shembekar, N., C. Chaipan, R. Utharala and C. A. Merten (2016). "Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics." Lab Chip **16**(8): 1314-1331.
Shenoy, N. R., J. E. Shively and J. M. Bailey (1993). "Studies in C-terminal sequencing: new reagents for the synthesis of peptidylthiohydantoins." J Protein Chem **12**(2): 195-205.
Shim, J. U., R. T. Ranasinghe, C. A. Smith, S. M. Ibrahim, F. Hollfelder, W. T. Huck, D. Klenerman and C. Abell (2013). "Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays." ACS Nano **7**(7): 5955-5964.
Shim, J. W., Q. Tan and L. Q. Gu (2009). "Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity." Nucleic Acids Res **37**(3): 972-982.
Sidoli, S., Z. F. Yuan, S. Lin, K. Karch, X. Wang, N. Bhanu, A. M. Arnaudo, L. M. Britton, X. J. Cao, M. Gonzales-Cope, Y. Han, S. Liu, R. C. Molden, S. Wein, L.

[0363]

Afjehi-Sadat and B. A. Garcia (2015). "Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis." Proteomics **15**(9): 1459-1469.
Sletten, E. M. and C. R. Bertozzi (2009). "Bioorthogonal chemistry: fishing for selectivity in a sea of functionality." Angew Chem Int Ed Engl **48**(38): 6974-6998.
Spencer, S. J., M. V. Tamminen, S. P. Preheim, M. T. Guo, A. W. Briggs, I. L. Brito, A. W. D, L. K. Pitkanen, F. Vigneault, M. P. Juhani Virta and E. J. Alm (2016). "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers." ISME J **10**(2): 427-436.
Spicer, C. D. and B. G. Davis (2014). "Selective chemical protein modification." Nat Commun **5**: 4740.
Spiropulos, N. G. and J. M. Heemstra (2012). "Templating effect in DNA proximity ligation enables use of non-bioorthogonal chemistry in biological fluids." Artif DNA PNA XNA **3**(3): 123-128.
Switzar, L., M. Giera and W. M. Niessen (2013). "Protein digestion: an overview of the available techniques and recent developments." J Proteome Res **12**(3): 1067-1077.
Tamminen, M. V. and M. P. Virta (2015). "Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells." Front Microbiol **6**: 195.
Tessler, L. (2011). Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection. Ph.D., WASHINGTON UNIVERSITY IN ST. LOUIS.
Tyson, J. and J. A. Armour (2012). "Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR." BMC Genomics **13**: 693.
Vauquelin, G. and S. J. Charlton (2013). "Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands." Br J Pharmacol **168**(8): 1771-1785.
Veggiani, G., T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson and M. Howarth (2016). "Programmable polyproteams built using twin peptide superglues." Proc Natl Acad Sci U S A **113**(5): 1202-1207.
Wang, D., S. Fang and R. M. Wohlhueter (2009). "N-terminal derivatization of peptides with isothiocyanate analogues promoting Edman-type cleavage and enhancing sensitivity in electrospray ionization tandem mass spectrometry analysis." Anal Chem **81**(5): 1893-1900.
Williams, B. A. and J. C. Chaput (2010). "Synthesis of peptide-oligonucleotide conjugates using a heterobifunctional crosslinker." Curr Protoc Nucleic Acid Chem **Chapter 4**: Unit4 41.
Wu, H. and N. K. Devaraj (2016). "Inverse Electron-Demand Diels-Alder Bioorthogonal Reactions." Top Curr Chem (J) **374**(1): 3.
Xiong, A. S., R. H. Peng, J. Zhuang, F. Gao, Y. Li, Z. M. Cheng and Q. H. Yao (2008). "Chemical gene synthesis: strategies, softwares, error corrections, and applications." FEMS Microbiol Rev **32**(3): 522-540.
Yao, Y., M. Docter, J. van Ginkel, D. de Ridder and C. Joo (2015). "Single-molecule protein sequencing through fingerprinting: computational assessment." Phys Biol **12**(5): 055003.

[0364]

Zakeri, B., J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth (2012). "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin." Proc Natl Acad Sci U S A **109**(12): E690-697.
Zhang, L., K. Zhang, S. Rauf, D. Dong, Y. Liu and J. Li (2016). "Single-Molecule Analysis of Human Telomere Sequence Interactions with G-quadruplex Ligand." Anal Chem **88**(8): 4533-4540.
Zhou, H., Z. Ning, A. E. Starr, M. Abu-Farha and D. Figeys (2012). "Advancements in top-down proteomics." Anal Chem **84**(2): 720-734.
Zilionis, R., J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein and L. Mazutis (2017). "Single-cell barcoding and sequencing using droplet microfluidics." Nat Protoc **12**(1): 44-73.

[0365]

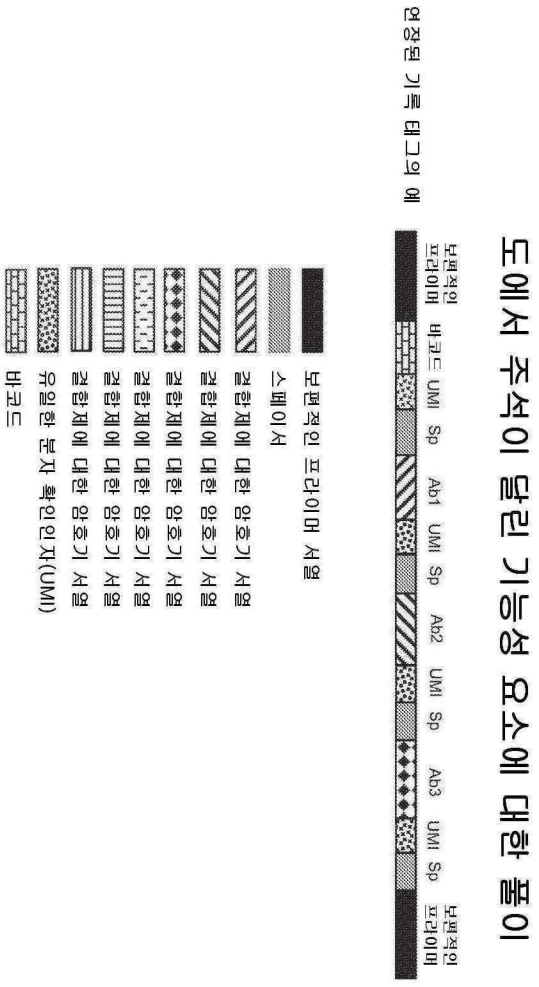
[0366] 이들 및 다른 변화는 상기 상세한 설명의 측면에서 구현예에 대해 이루어질 수 있다. 일반적으로, 다음의 청부 범위에서, 사용된 용어는 청구범위를 명세서 및 청구범위에 개시된 구체적인 구현예로 한정하는 것으로 해석되지 않아야 하지만, 이러한 청구범위가 권리를 부여하는 등가물의 완전한 범위와 함께 모든 가능한 구현예를 포함하는 것으로 해석되어야 한다. 따라서, 청구범위는 본 개시내용에 의해 한정되지 않는다.

[0367] 상기 기술된 다양한 구현예를 조합하여 추가의 구현예를 제공할 수 있다. 본 명세서에서 언급되고/되거나, 이들

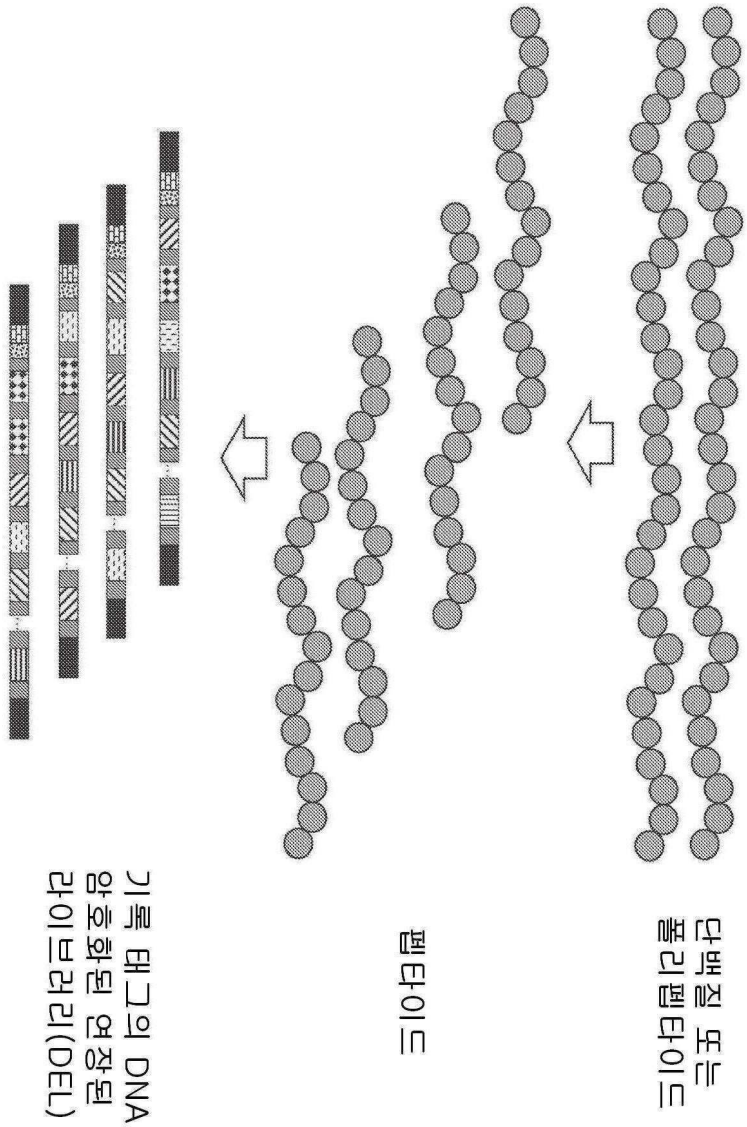
의 전문이 본원에 참고로 포함된, 미국 가특허원 제62/330,841호, 미국 가특허원 제62/339,071호, 및 미국 가특허원 제62/376,886호를 포함하는, 출원 데이타 시이트에 나열된 모든 미국 특허, 미국 특허원 공보, 미국 특허원, 외국 특허, 외국 특허원, 및 비-특허 공보는 이들 전문이 참고로 본원에 포함된다. 구현예의 국면은 필요한 경우, 여전히 추가의 구현예를 제공하기 위해 다양한 특허, 특허원 및 공보의 개념을 사용하여 변형될 수 있다.

도면

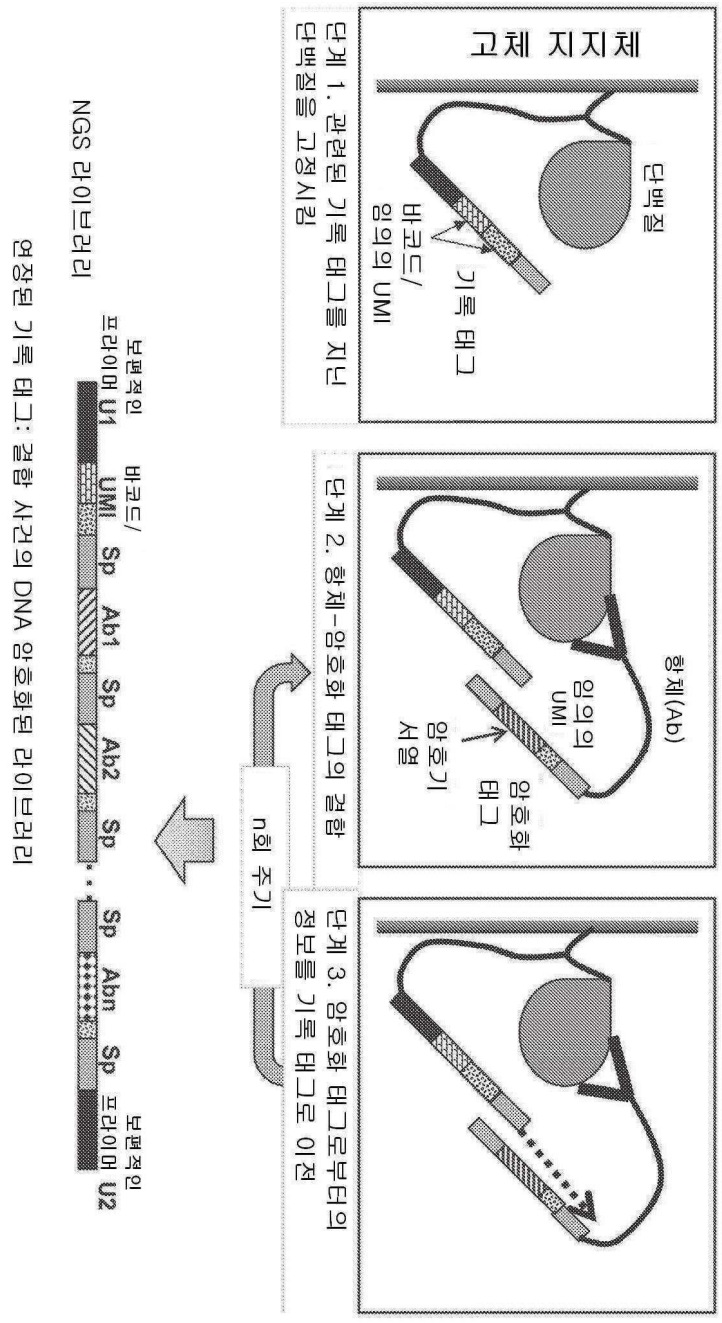
도면1a



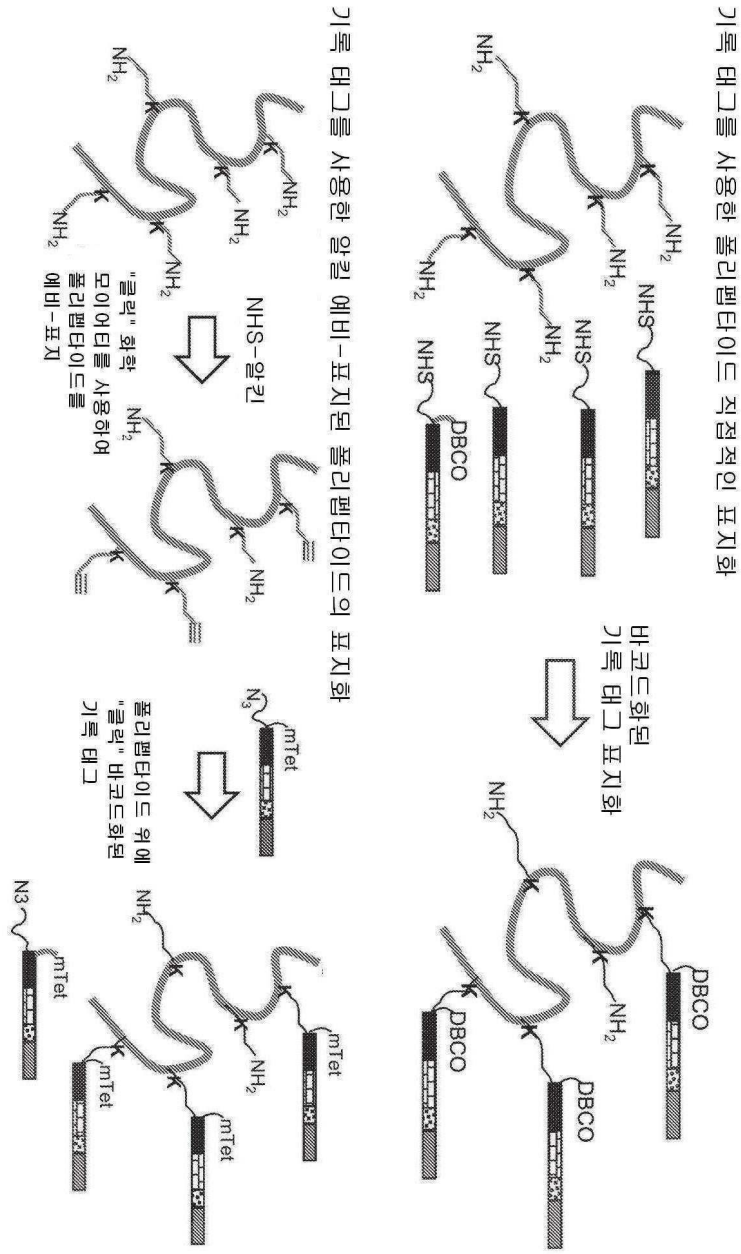
도면1b



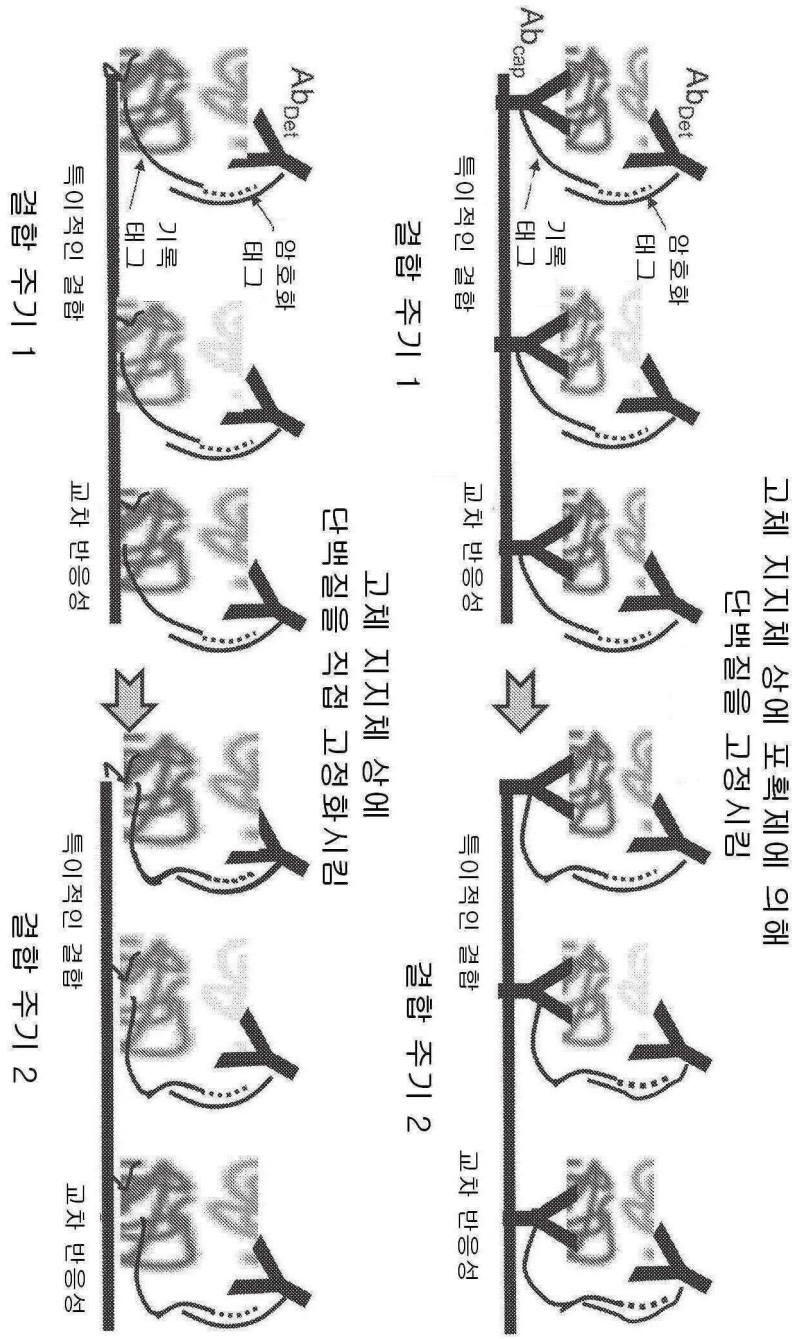
도면2a


고체 지지체
단백질
바코드/
임의의 UMI
기록 태그
단계 1. 관련된 기록 태그를 지닌
단백질을 고정시킴
항체(Ab)
임의의
UMI
암호기
서열
암호화
태그
단계 2. 항체-암호화 태그의 결합
n회 주기
단계 3. 암호화 태그로부터의
정보를 기록 태그로 이전
연장된 기록 태그: 결합 사건의 DNA 암호화된 라이브러리
NGS 라이브러리
보편적인
프라이머 U1
바코드/
UMI
Sp
Ab1
Sp
Ab2
Sp
Sp
Abn
Sp
보편적인
프라이머 U2

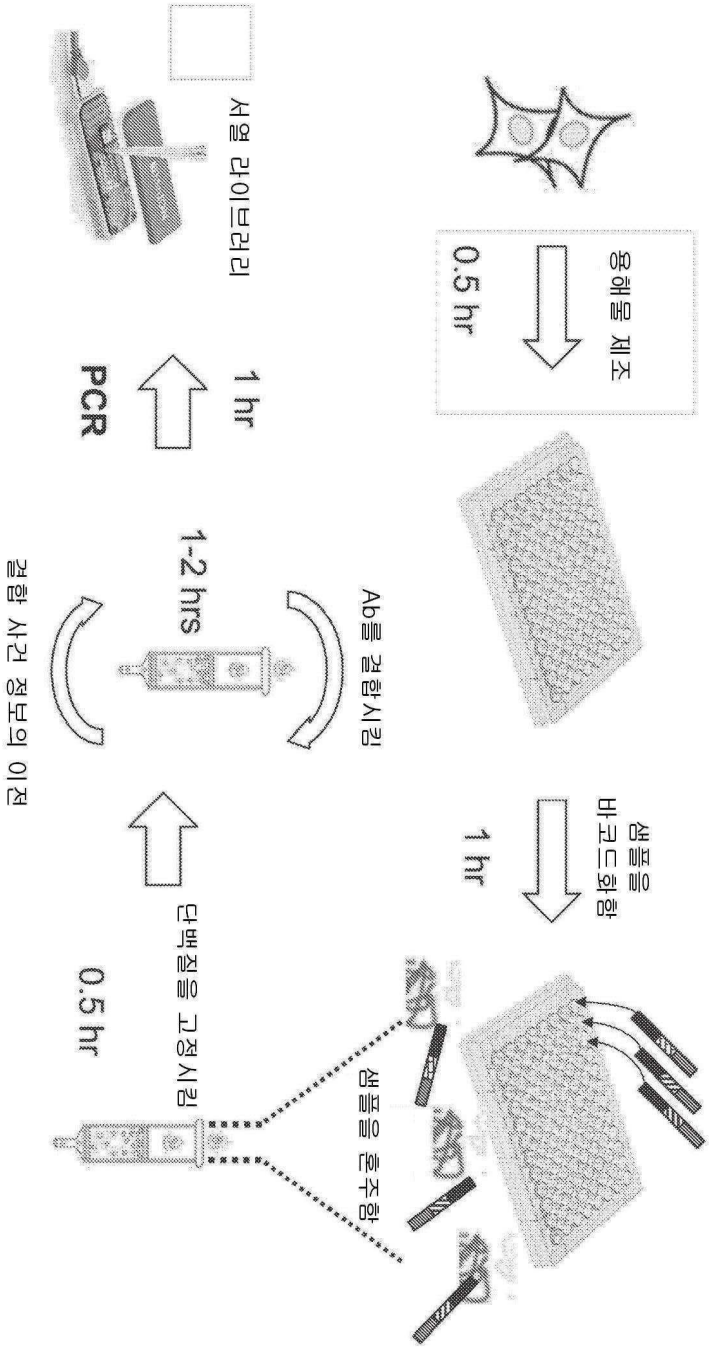
도면2b


기록 태그를 사용한 폴리펩타이드 직접적인 표지화
NHS
DBCO
바코드화된
기록 태그 표지화
기록 태그를 사용한 알킬 에비-표지된 폴리펩타이드의 표지화
NHS-알킬
"클릭" 화학
모이어티를 사용하여
폴리펩타이드를
에비-표지
폴리펩타이드 위에
"클릭" 바코드화된
기록 태그
mTet
N3

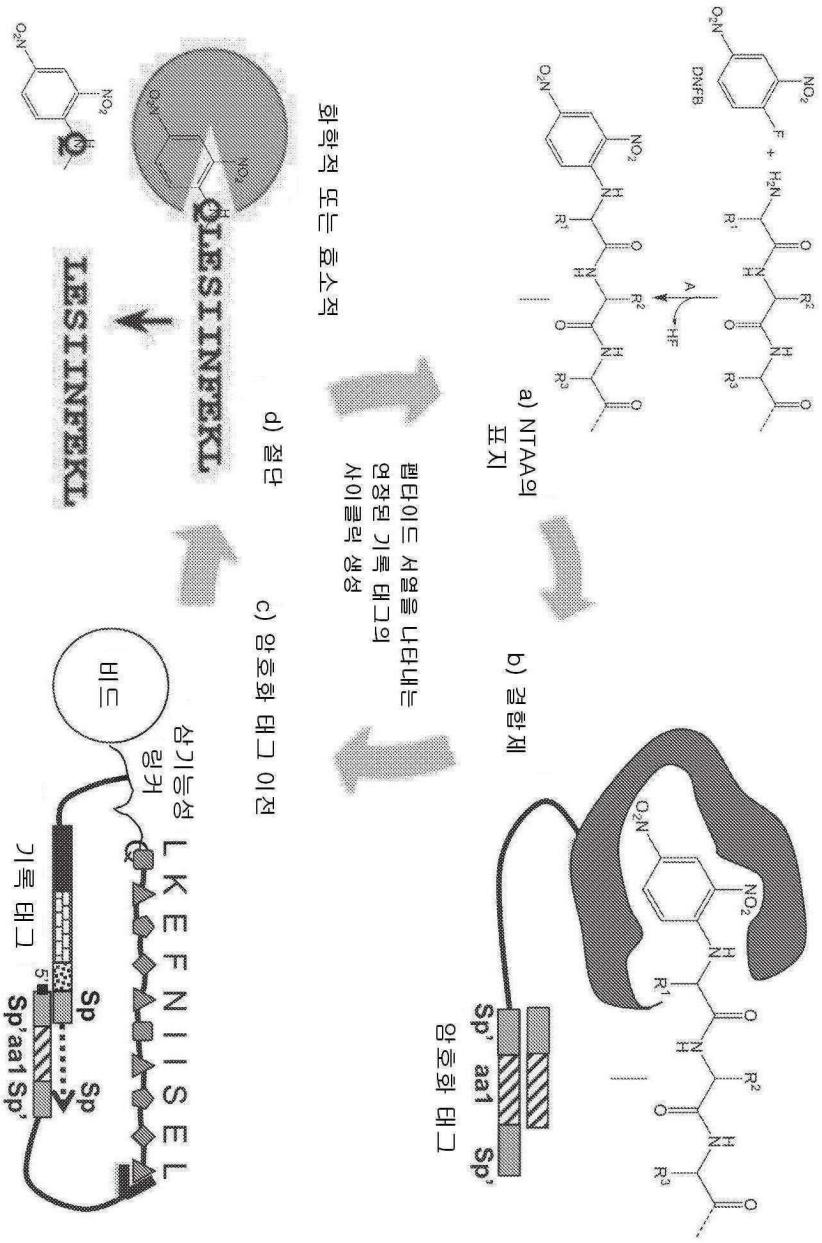
도면2c


고체 지지체 상에 포획제에 의해
단백질을 고정시킴
결합 주기 1
결합 주기 2
고체 지지체 상에
단백질을 직접 고정화시킴
결합 주기 1
결합 주기 2
Ab Det
Ab cap
암호화 태그
기록 태그
특이적인 결합
교차 반응성

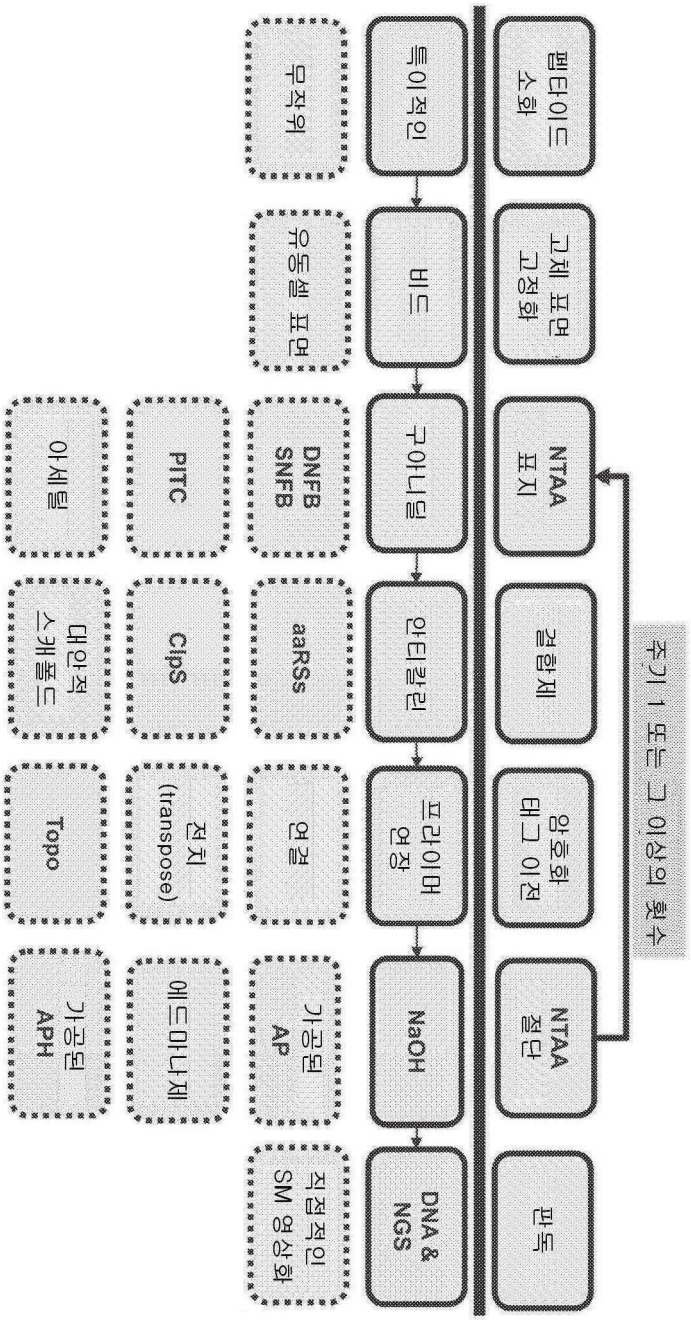
도면2d


용해물 제조
0.5 hr
샘플을 바코드화함
1 hr
샘플을 혼주함
단백질을 고정시킴
0.5 hr
Ab를 결합시킴
1-2 hrs
결합 사건 정보의 이전
PCR
1 hr
서열 라이브러리

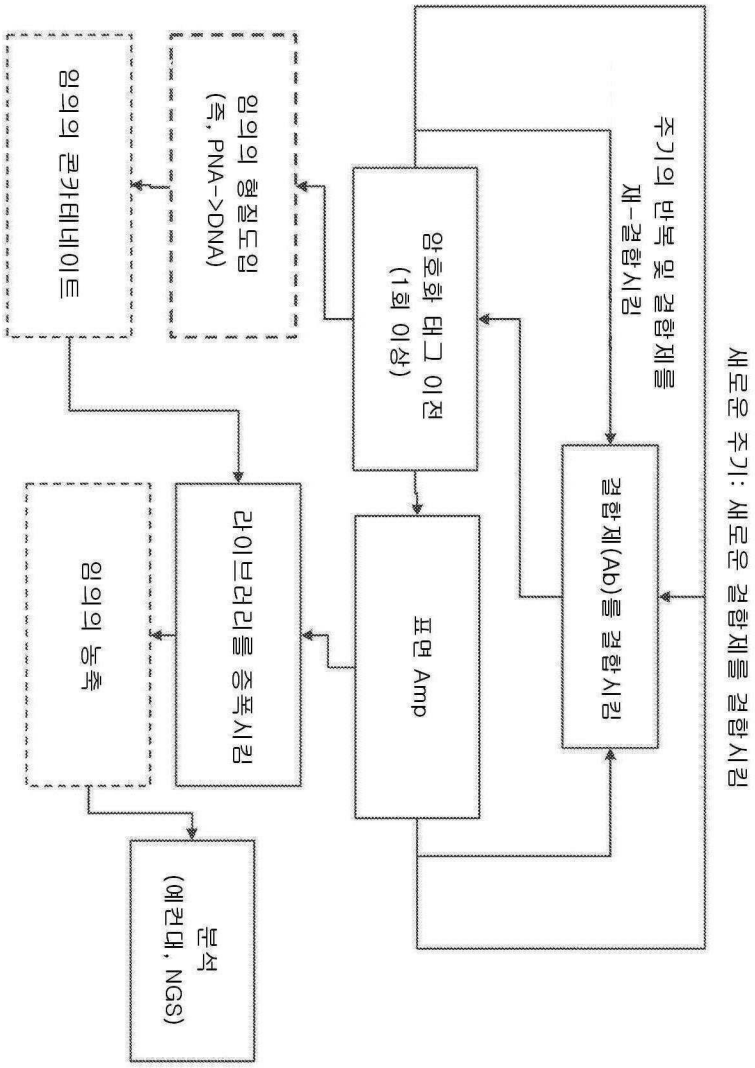
도면3


a) NTAA의 표지
b) 결합제
펩타이드 서열을 나타내는 연장된 기록 태그의 사이클릭 생성
화학적 또는 효소적
c) 암호화 태그 이전
d) 절단
비드
삼기능성 링커
기록 태그
Sp
Sp
Sp'aa1Sp'
암호화 태그
Sp' aa1 Sp'
LESINFEKL
OLESINFEKL
DNFB
HF

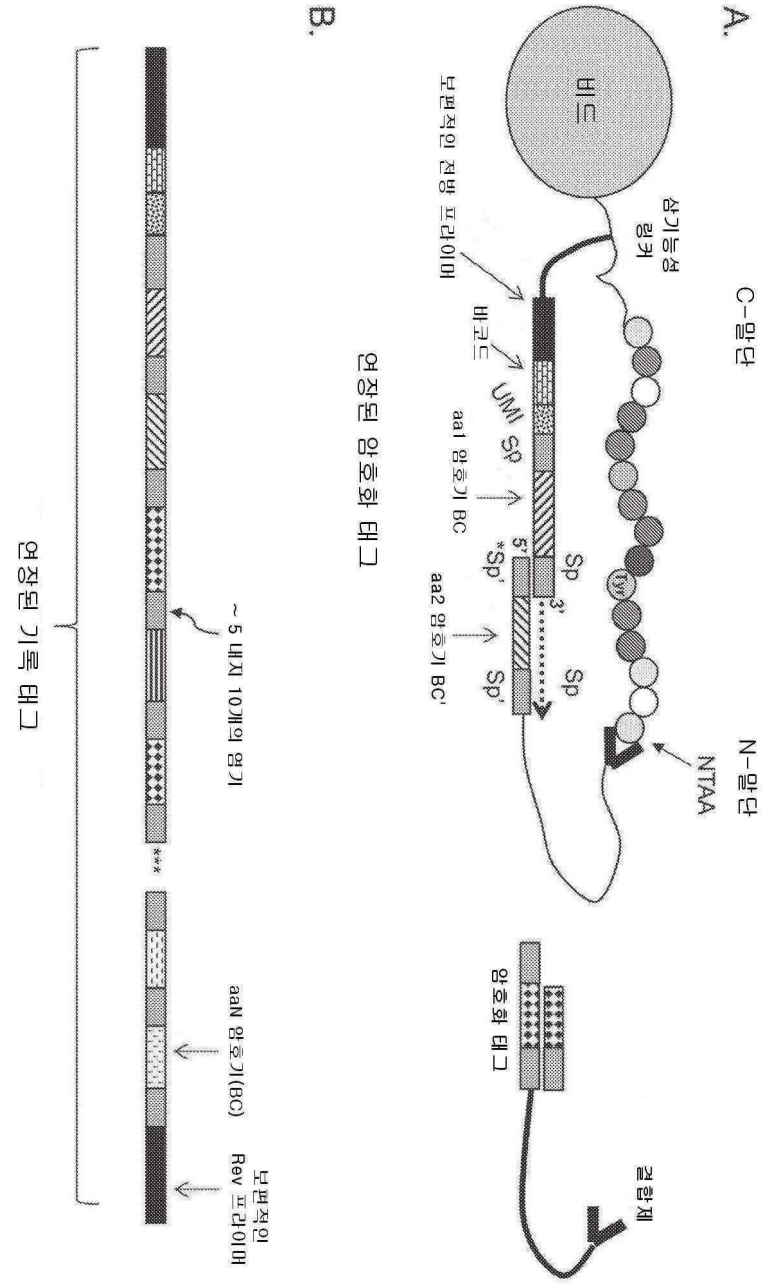
도면4a


펩타이드 소화
고체 표면 고정화
NTAA 표지
결합제
암호화 태그 이전
NTAA 절단
판독
특이적인
비드
구아니딜
안티칼린
프라이머 연장
NaOH
DNA & NGS
무작위
유동셀 표면
DNFB SNFB
aaRSs
연결
가공된 AP
직접적인 SM 영상화
아세틸
PITC
ClpS
전치 (transpose)
에드마나제
대안적 스캐폴드
Topo
가공된 APH
주기 1 또는 그 이상의 횟수

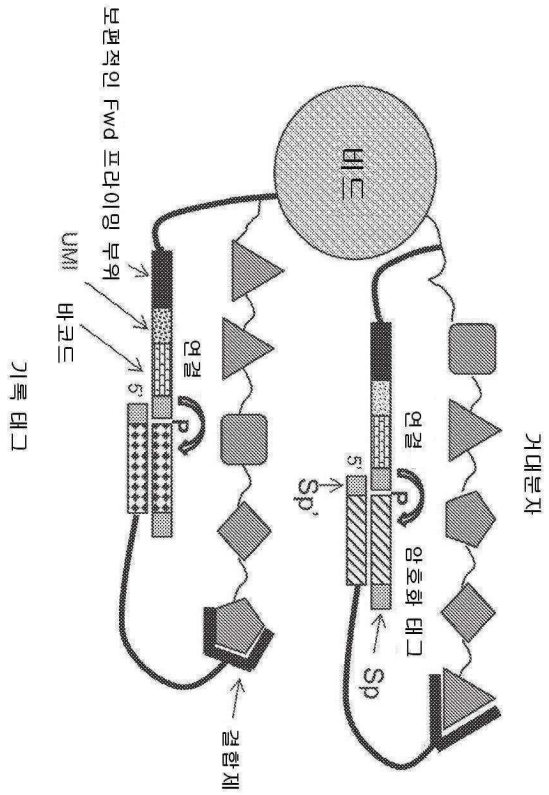
도면4b


임의의 코카테네이트
임의의 올절도입
(즉, PNA->DNA)
암호화 태그 이전
(1회 이상)
주기의 반복 및 결합제를
재-결합시킴
새로운 주기: 새로운 결합제를 결합시킴
결합제(Ab)를 결합시킴
표면 Amp
라이브러리를 증폭시킴
임의의 농축
분석
(예컨대, NGS)

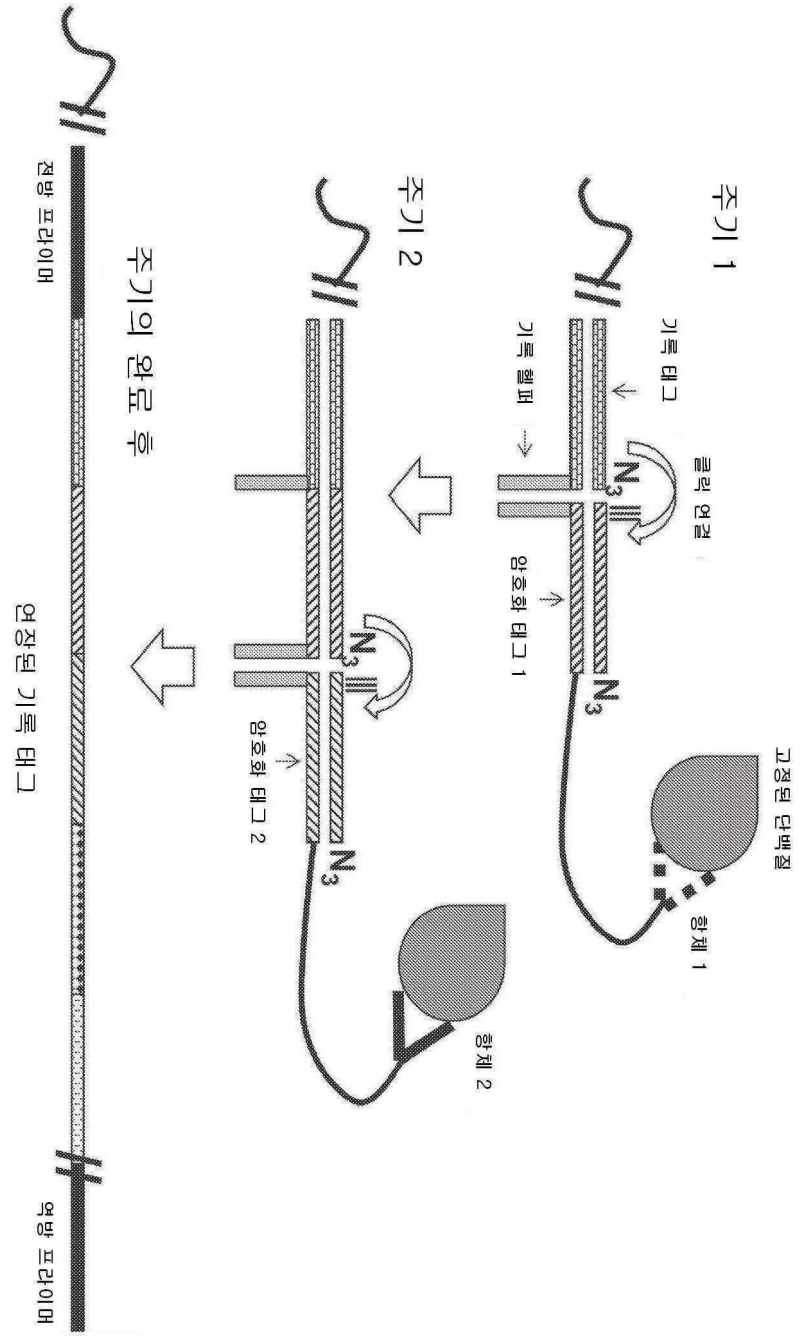
도면5


A.
C-말단
삼기능성 링커
비드
보편적인 전방 프라이머
바코드
UMI
Sp
aa1 암호기 BC
5'
*Sp'
Sp
3'
Sp
aa2 암호기 BC'
Sp'
Uv
N-말단
NTAA
암호화 태그
결합제
B.
연장된 암호화 태그
연장된 기록 태그
~ 5 내지 10개의 염기
***
aaN 암호기(BC)
보편적인 Rev 프라이머

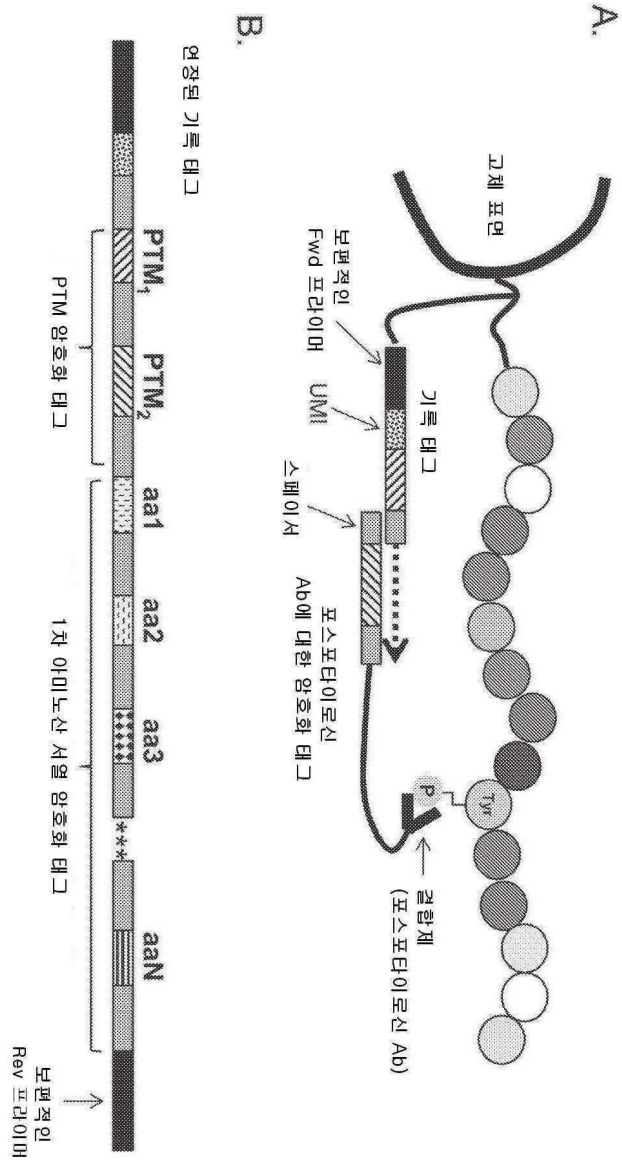
도면6



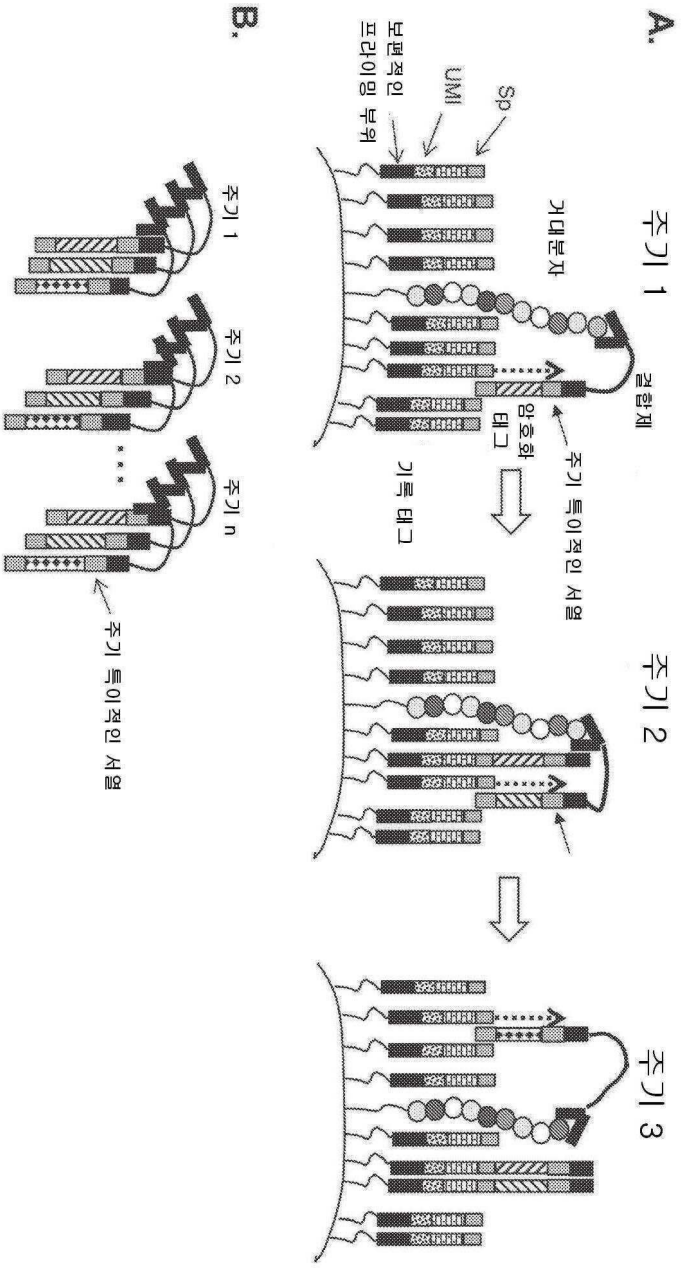
도면7


주기 1
기록 태그
기록 헬퍼
클릭 연결
암호화 태그 1
고정된 단백질
항체 1
주기 2
암호화 태그 2
항체 2
주기의 완료 후
전방 프라이머
연장된 기록 태그
역방 프라이머

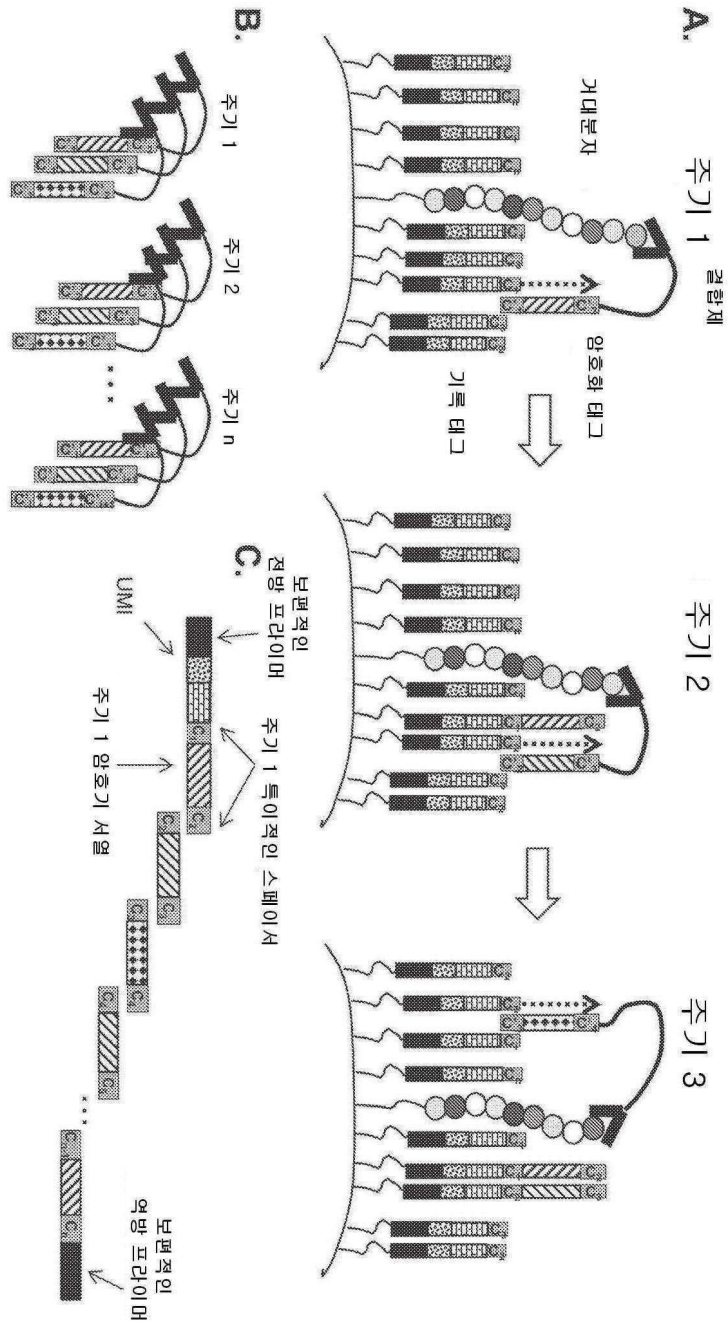
도면8


A.
고체 표면
보편적인
Fwd 프라이머
UMI
스페이서
기록 태그
포스포타이로신
Ab에 대한 암호화 태그
P
Tyr
결합제
(포스포타이로신 Ab)
B.
연장된 기록 태그
PTM암호화 태그
PTM1
PTM2
aa1
aa2
aa3
***
aaN
1차 아미노산 서열 암호화 태그
보편적인
Rev 프라이머

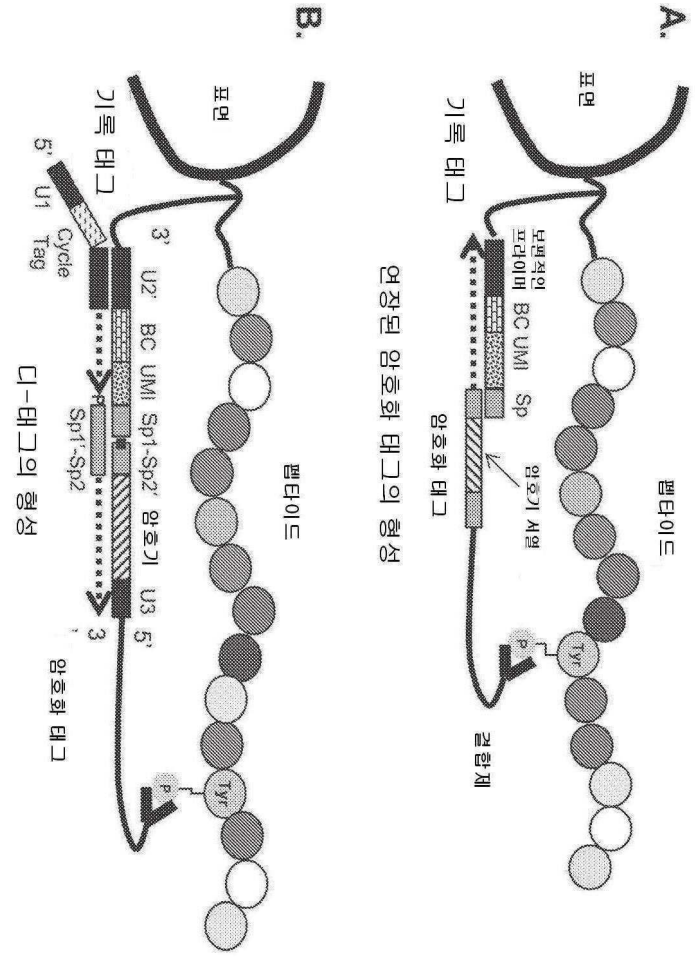
도면9


A.
주기 1
보편적인 프라이밍 부위
UMI
Sp
겔대분자
결합제
암호화 태그
주기 특이적인 서열
기록 태그
주기 2
주기 3
B.
주기 1
주기 2
주기 n
주기 특이적인 서열

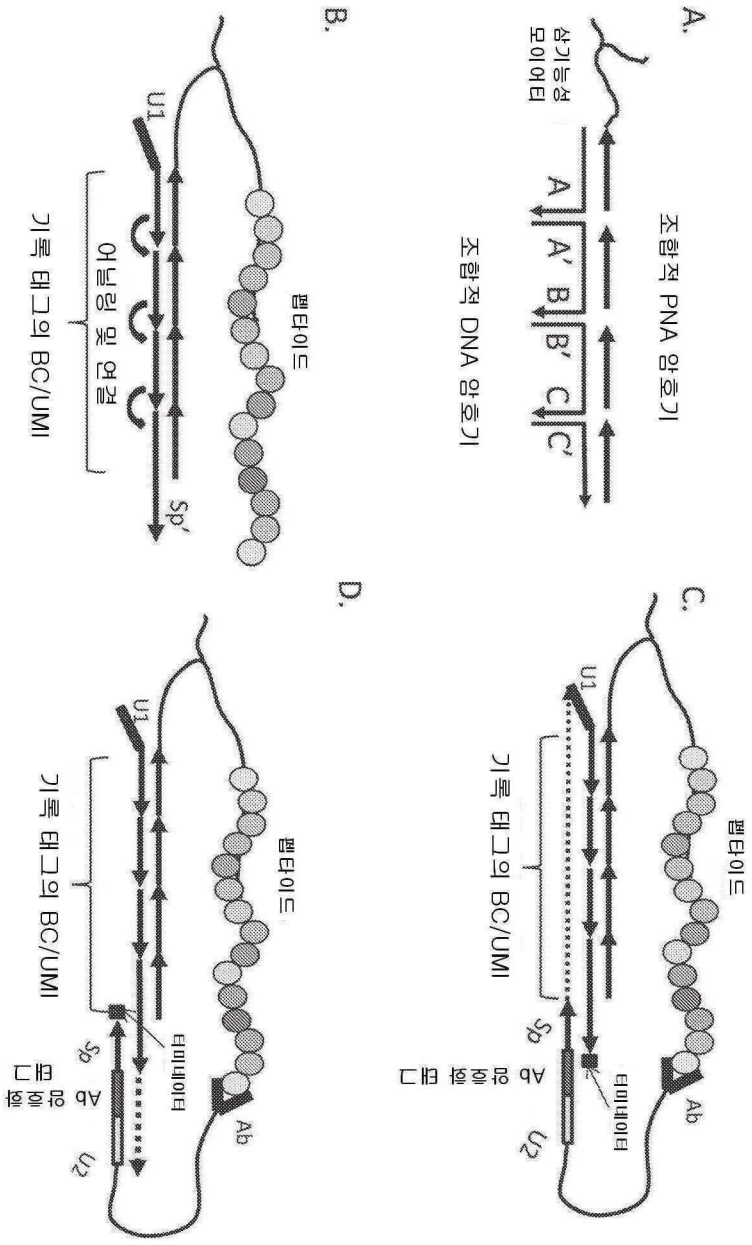
도면10


A.
주기 1
결합제
거대분자
암호화 태그
기록 태그
주기 2
주기 3
B.
주기 1
주기 2
주기 n
C.
보편적인
전방 프라이머
UMI
주기 1 암호기 서열
주기 1 특이적인 스페이서
보편적인
역방 프라이머

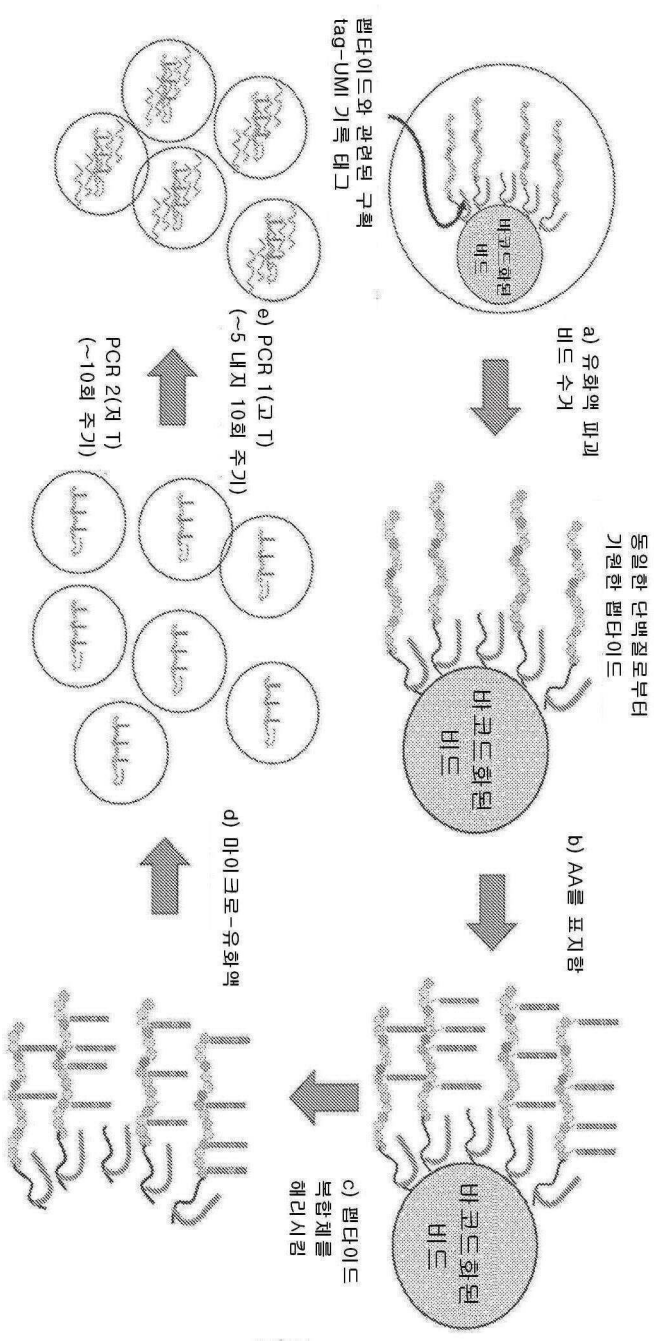
도면11


A.
표면
기록 태그
모멘적인 프라이머
BC UMI
Sp
암호기 서열
암호화 태그
펩타이드
Tyr
P
절합제
연장된 암호화 태그의 형성
B.
표면
기록 태그
5'
U1
Cycle Tag
3'
U2'
BC
UMI
Sp1-Sp2'
암호기
U3
5'
Sp1'-Sp2
3'
펩타이드
Tyr
P
디-태그의 형성
암호화 태그

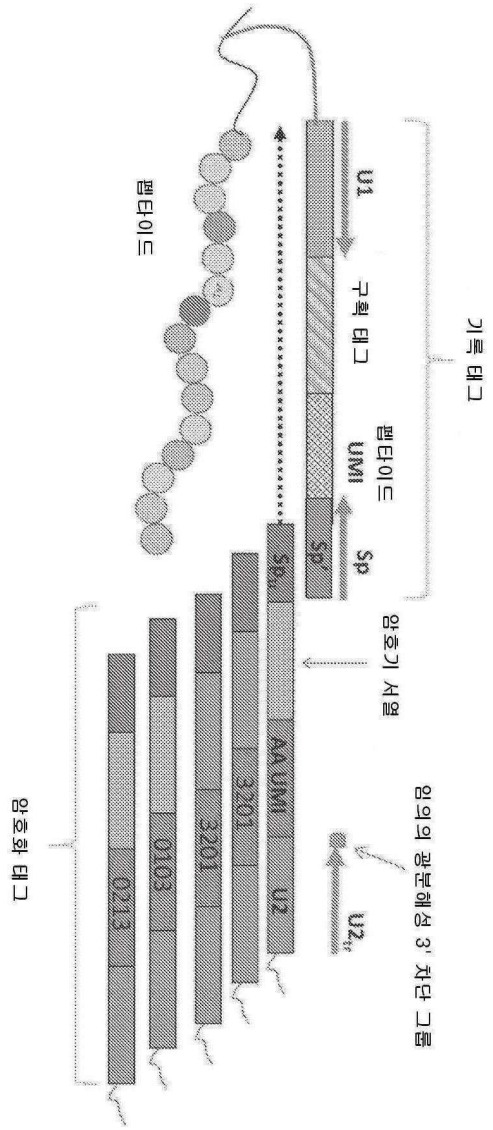
도면12


A.
조합적 PNA 암호기
삼기능성 모이어티
A
A′
B
B′
C
C′
조합적 DNA 암호기
B.
기록 태그의 BC/UMI
어닐링 및 연결
U1
Sp′
펩타이드
C.
기록 태그의 BC/UMI
U1
Sp
Ab 암호화 태그
U2
테이퍼이터
Ab
펩타이드
D.
기록 태그의 BC/UMI
U1
Sp
Ab 암호화 태그
U2
테이퍼이터
Ab
펩타이드

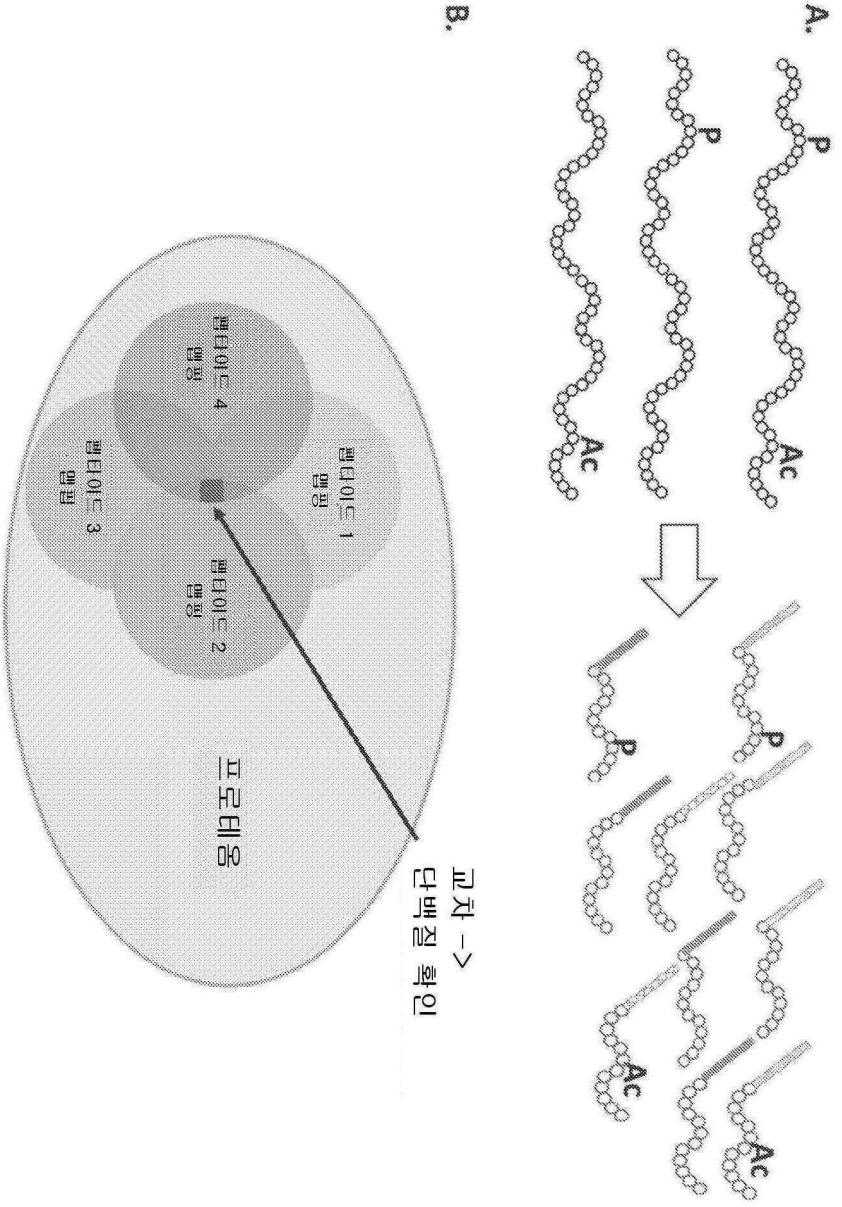
도면13


바코드화된 비드
a) 유화액 파괴 비드 수거
펩타이드와 관련된 구별 tag-UMI 기록 태그
e) PCR 1(고 T) (~5 내지 10회 주기)
PCR 2(저 T) (~10회 주기)
동일한 단백질로부터 기원한 펩타이드
바코드화된 비드
b) AA를 표지함
바코드화된 비드
c) 펩타이드 복합체를 해리시킴
d) 마이크로-유화액

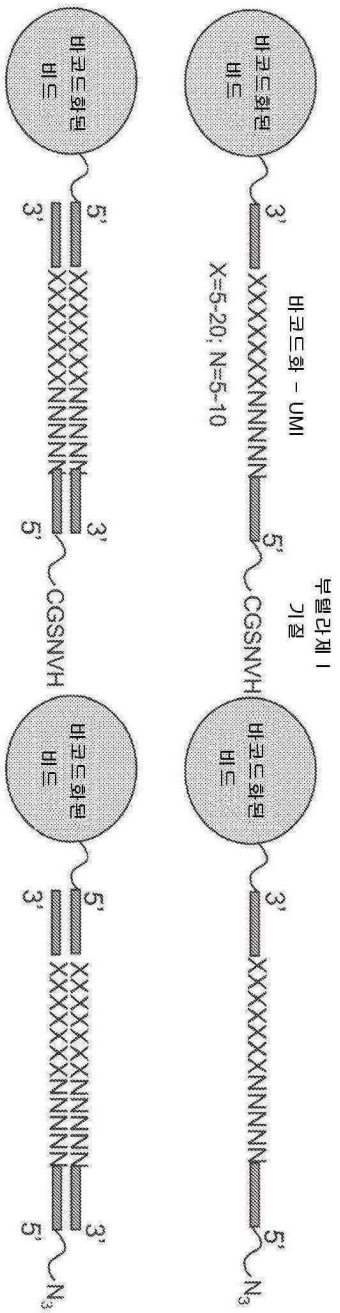
도면14


U1
구획 태그
펩타이드
UMI
Sp
기록 태그
펩타이드
Sp
Sp
암호기 서열
AA UMI
3201
3201
0103
0213
U2
$U2_{r'}$
임의의 공분해성 3' 차단 그룹
암호화 태그

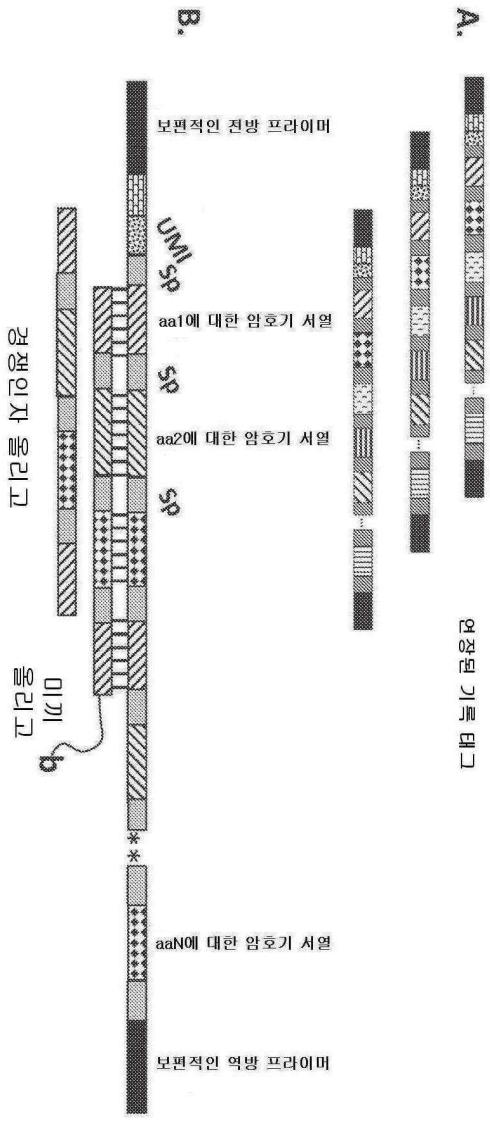
도면15



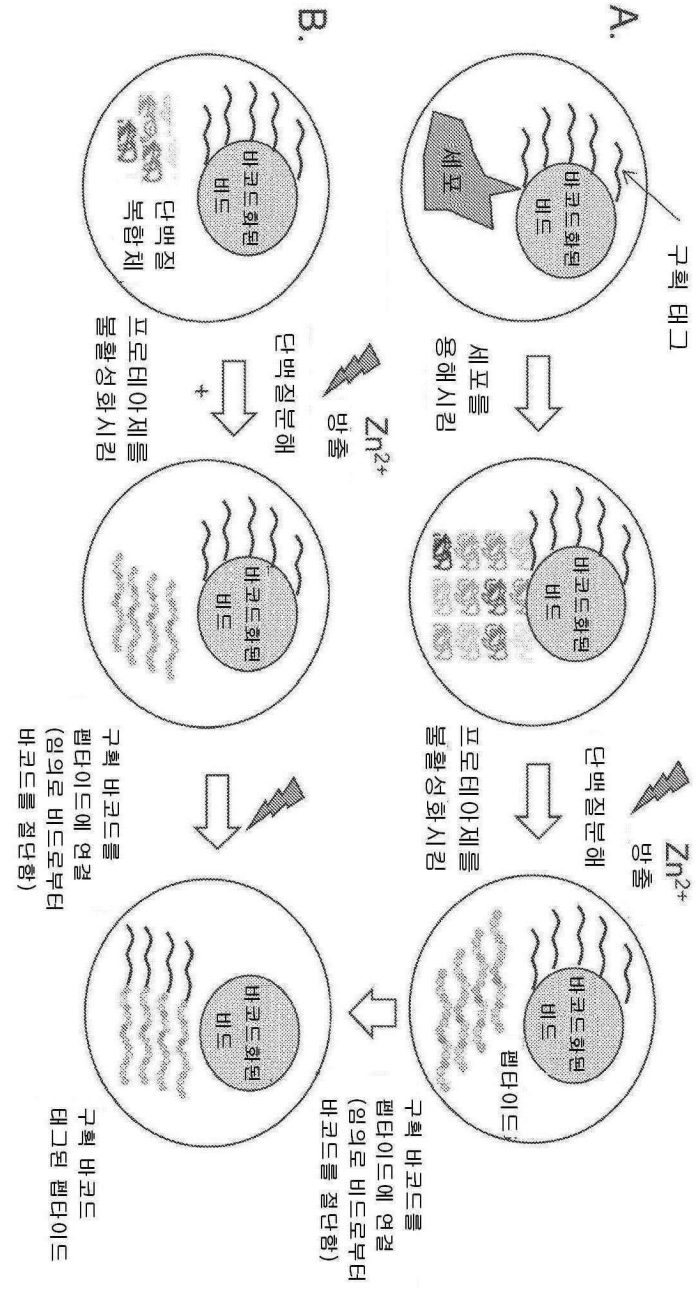
도면16


바코드화된 비드
3'
5'
XXXXXXXNNNNNN
바코드화 - UMI
X=5-20; N=5-10
CGSNVH
부텔라제 I 기질
N3

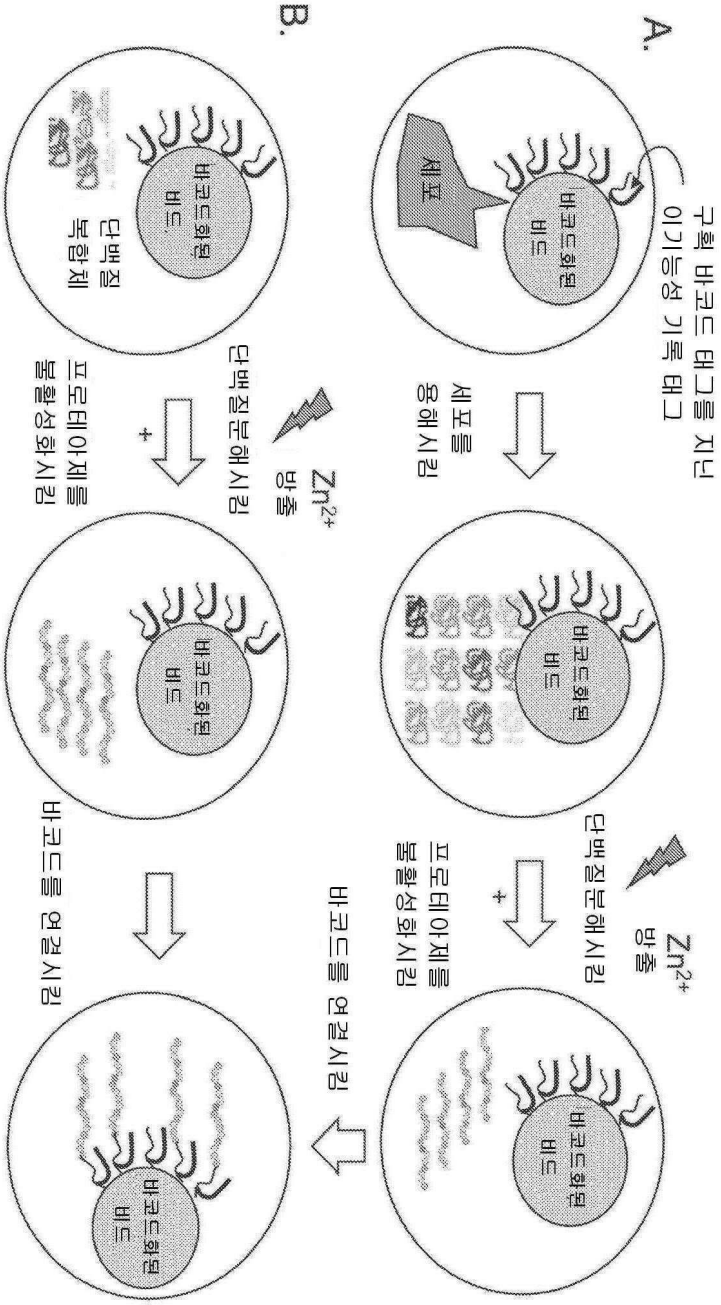
도면17


B.
보편적인 전방 프라이머
UMI
SP
aa1에 대한 암호기 서열
SP
aa2에 대한 암호기 서열
SP
경쟁인자 올리고
미끼 올리고
aaN에 대한 암호기 서열
보편적인 역방 프라이머
A.
연장된 기질 태그

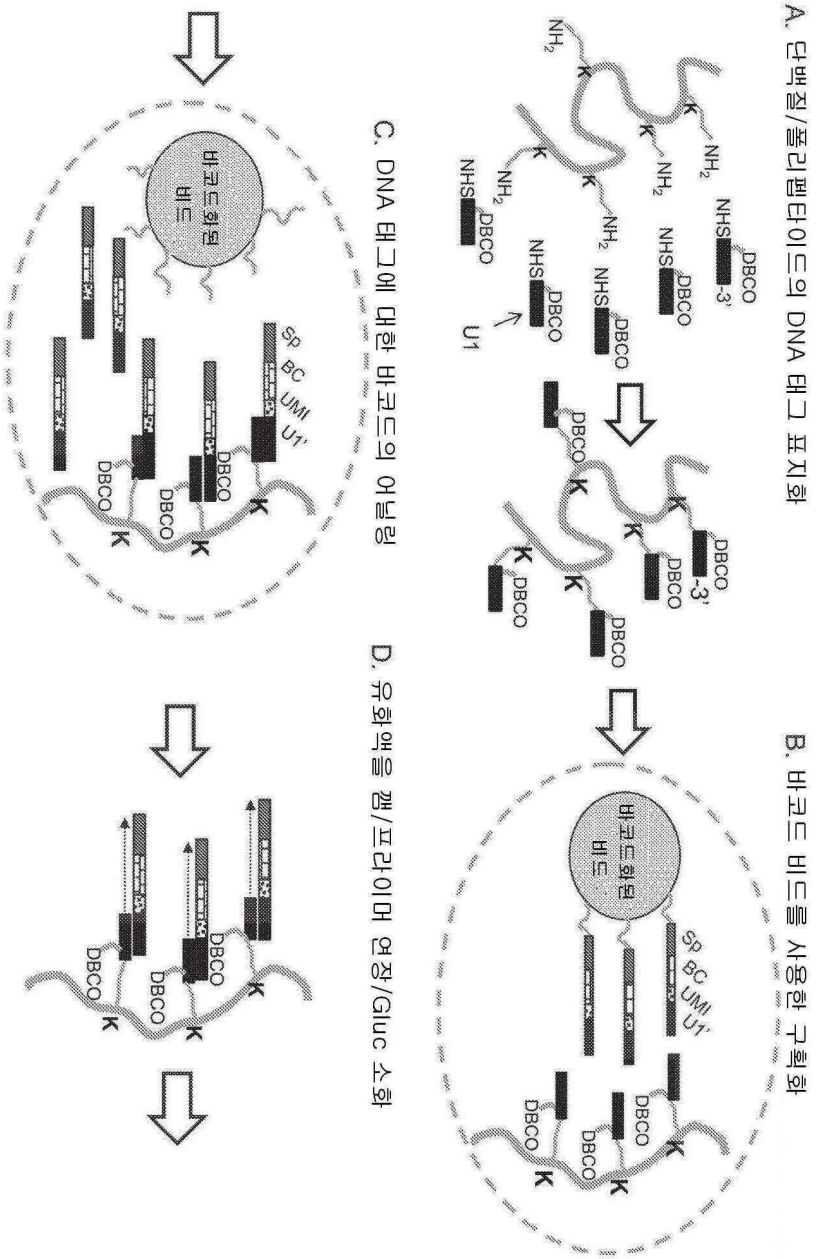
도면18


A.
구획 태그
세포
바코드화 비드
세포를 용해시킴
단백질 분해
Zn^{2+} 방출
프로테아제를 불활성화시킴
펩타이드
구획 바코드를 펩타이드에 연결 (임의로 바코드를 절단함)
구획 바코드 펩타이드 태그
B.
단백질 분해
Zn^{2+} 방출
프로테아제를 불활성화시킴
+
단백질 복합체
구획 바코드를 펩타이드에 연결 (임의로 바코드를 절단함)

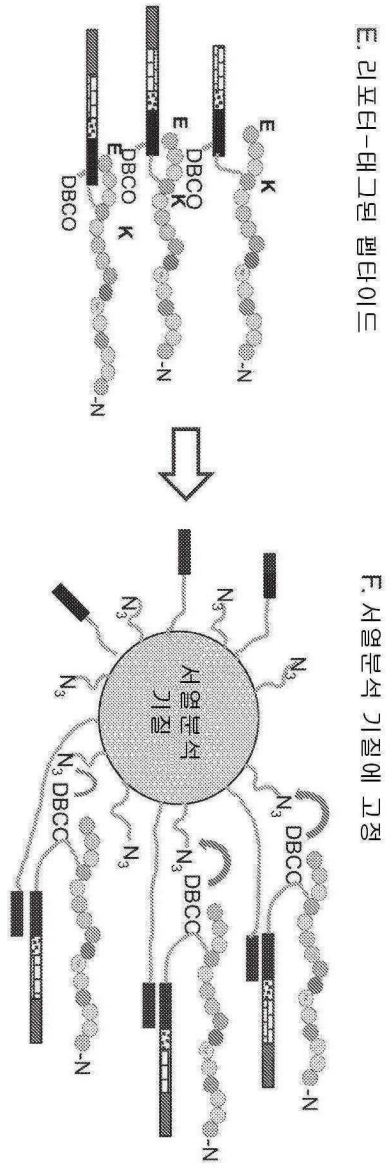
도면19


A.
구획 바코드 태그를 지닌
이기능성 기록 태그
세포
바코드화합
비드
세포를
융해시킴
단백질
복합체
바코드화합
비드
B.
단백질분해시킴
Zn2+
방출
프로테아제를
불활성화시킴
바코드화합
비드
바코드를 연결시킴
바코드화합
비드
단백질분해시킴
Zn2+
방출
프로테아제를
불활성화시킴
바코드화합
비드
바코드를 연결시킴
바코드화합
비드

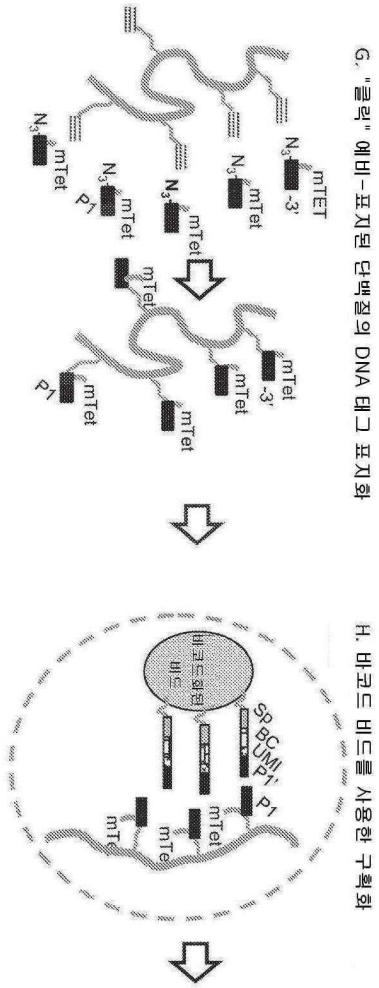
도면20a


A. 단백질/폴리펩타이드의 DNA 태그 표지화
B. 바코드 비드를 사용한 구획화
C. DNA 태그에 대한 바코드의 어닐링
D. 유화액을 깸/프라이머 연장/Gluc 소화
바코드화된 비드
SP
BC
UMI
U1'
U1
NHS
DBCO
NH2
K

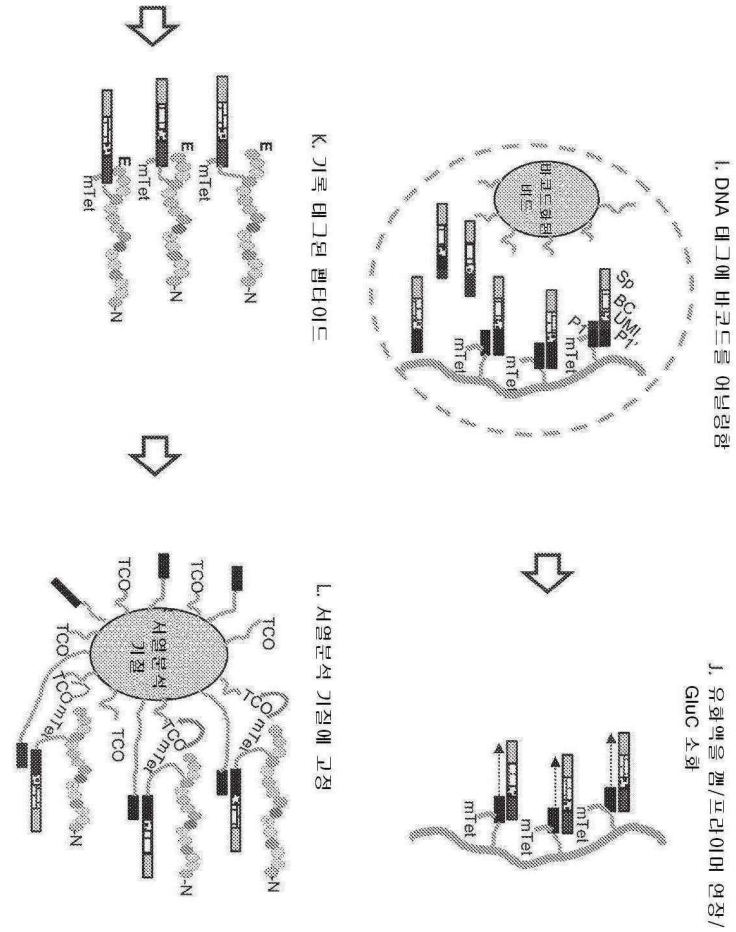
도면20b



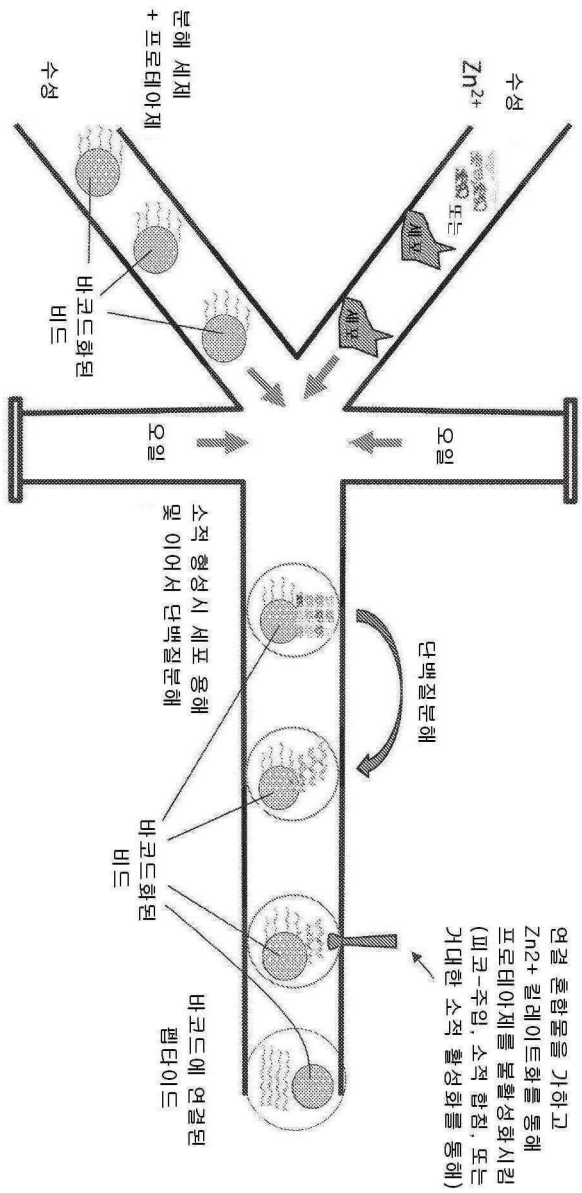
도면20c


G. "클릭" 에비-표지된 단백질의 DNA 태그 표지화
H. 바코드 비드를 사용한 구획화
mTET
-3'
mTet
P1
SP
BC
UMI
P1'
바코드 비드

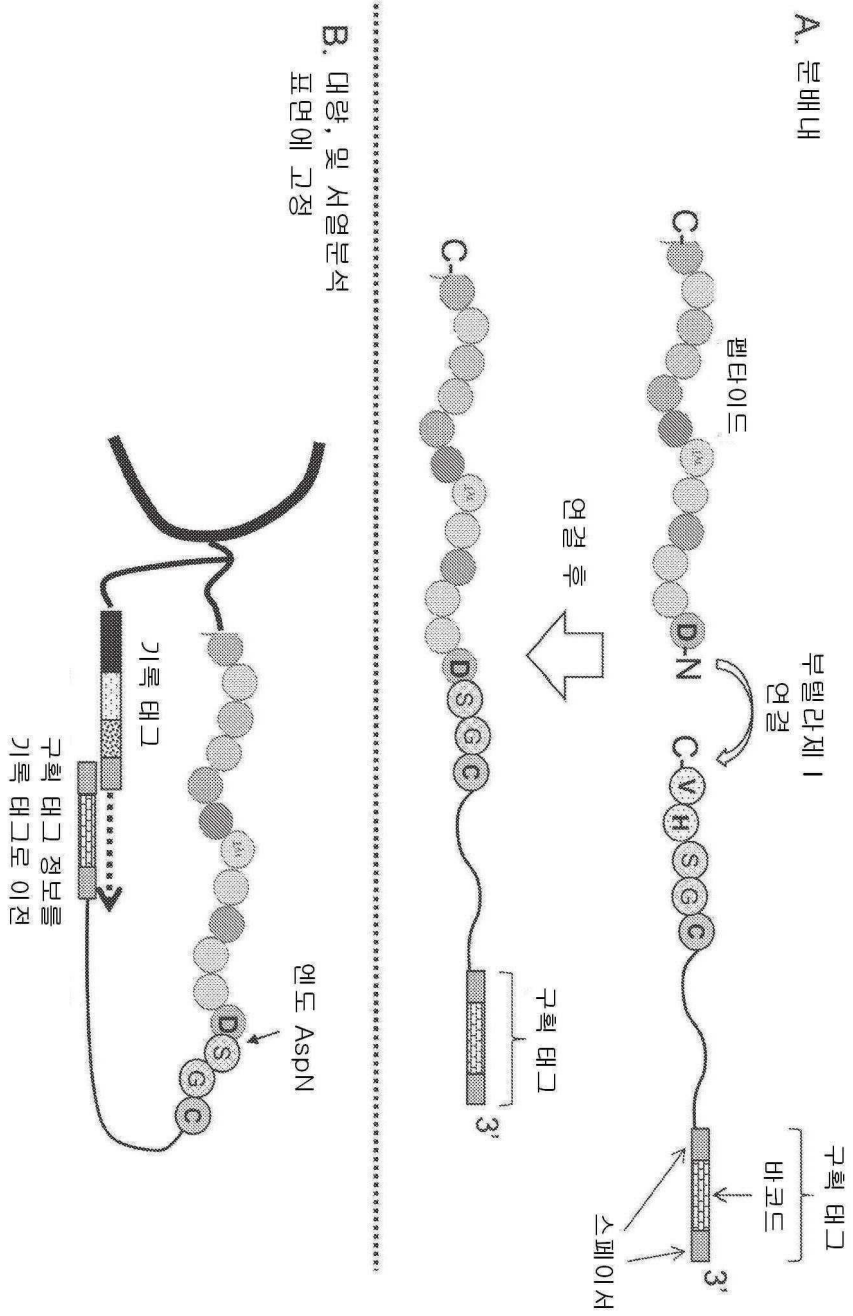
도면20d


I, DNA 태그에 바코드를 어닐링함
바코드 올리고 비드
SP
BC
UMI
P1'
P1
mTet
J, 유효 엔드 랩/프라이머 연장/
GluC 소화
K, 기록 태그된 펩타이드
L, 서열분석 기질에 고정
서열분석 기질
TCO
N

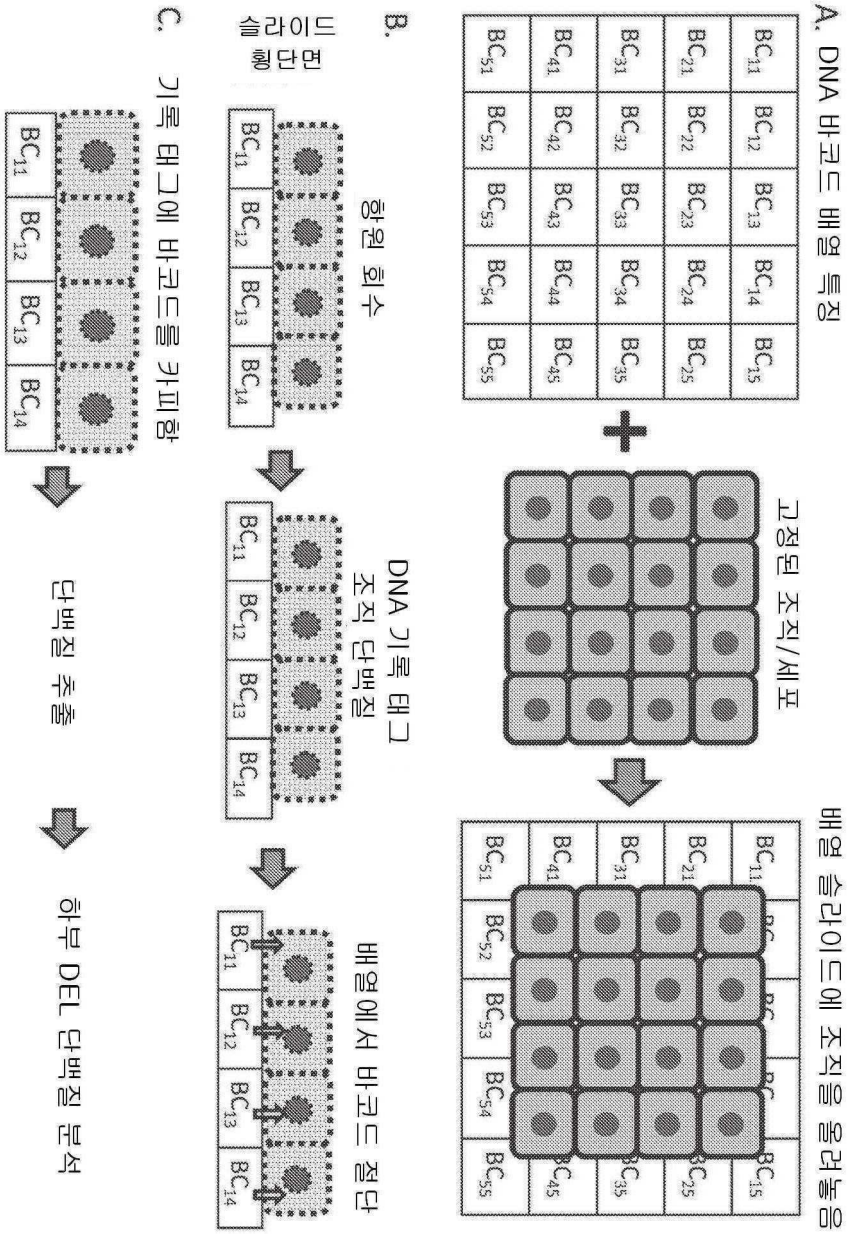
도면21


수성
분해 세제 + 프로테아제
수성
Zn^{2+}
바코드형 펩타이드
바코드형 펩타이드
세포
세포
입구
입구
소적 형성 시 세포 용해에 의해 방출된 단백질분해 효소
단백질분해
바코드형 펩타이드
바코드형 펩타이드
연결 혼합물을 가하고 Zn2+ 킬레이트화를 통해 프로테아제를 불활성화시킴 (피코-주입, 소적 합천, 또는 거대한 소적 활성화를 통해)

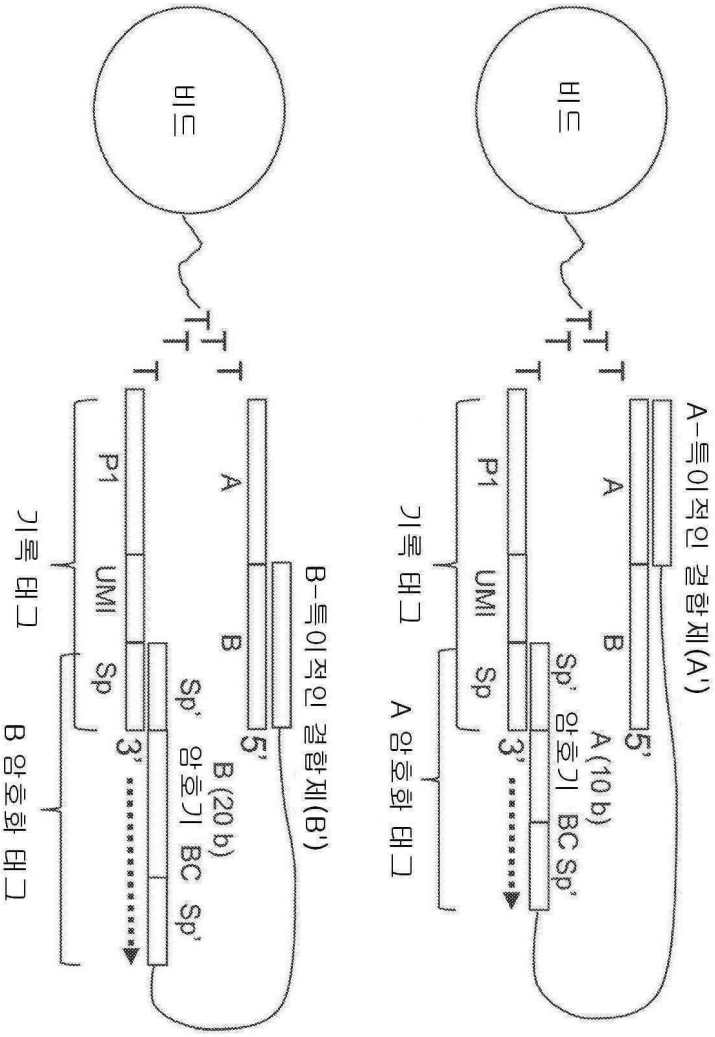
도면22


A. 큐베나
펩타이드
C
D-N
C
V
H
S
G
C
큐텔라제 I
연결
구획 태그
바코드
스페이서
3'
연결 후
C
D
S
G
C
구획 태그
3'
B. 대량, 및 서열분석
표면에 고정
기록 태그
구획 태그 정보를
기록 태그로 이전
엔도 AspN
D
S
G
C

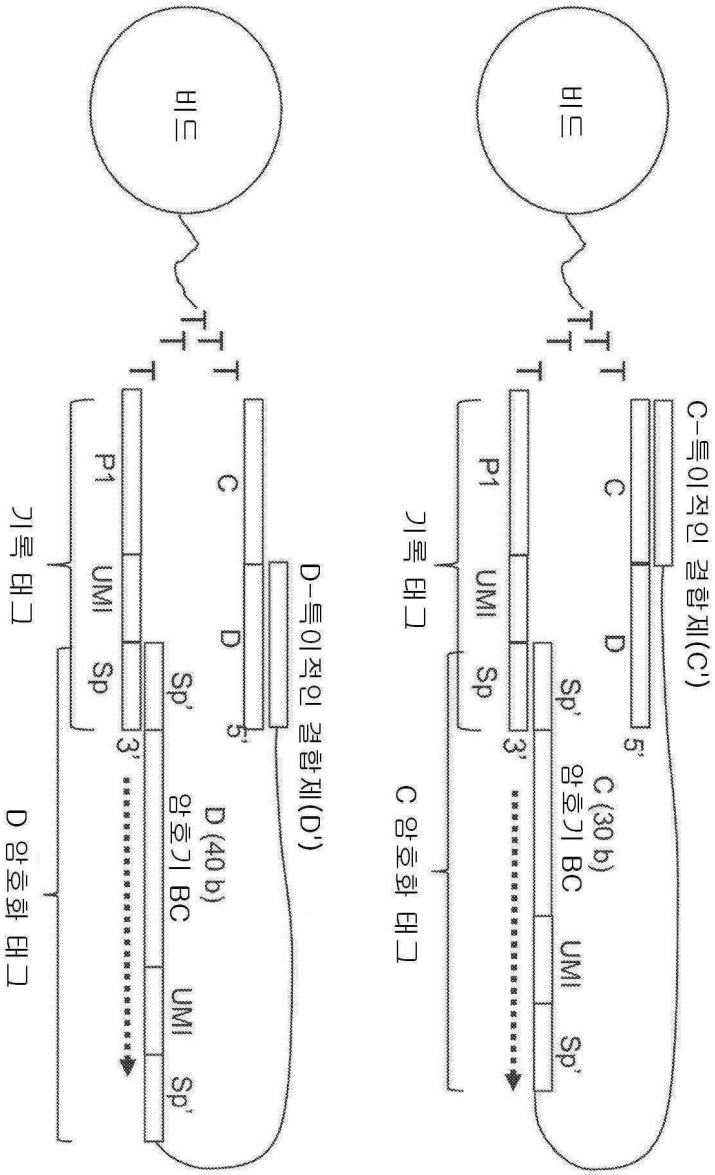
도면23


A. DNA 바코드 배열 특징
고정된 조직/세포
배열 슬라이드에 조직을 올려놓음
B.
슬라이드 횡단면
항원 회수
DNA 기록 태그
조직 단백질
배열에서 바코드 절단
C. 기록 태그에 바코드를 카피함
단백질 추출
하부 DEL 단백질 분석

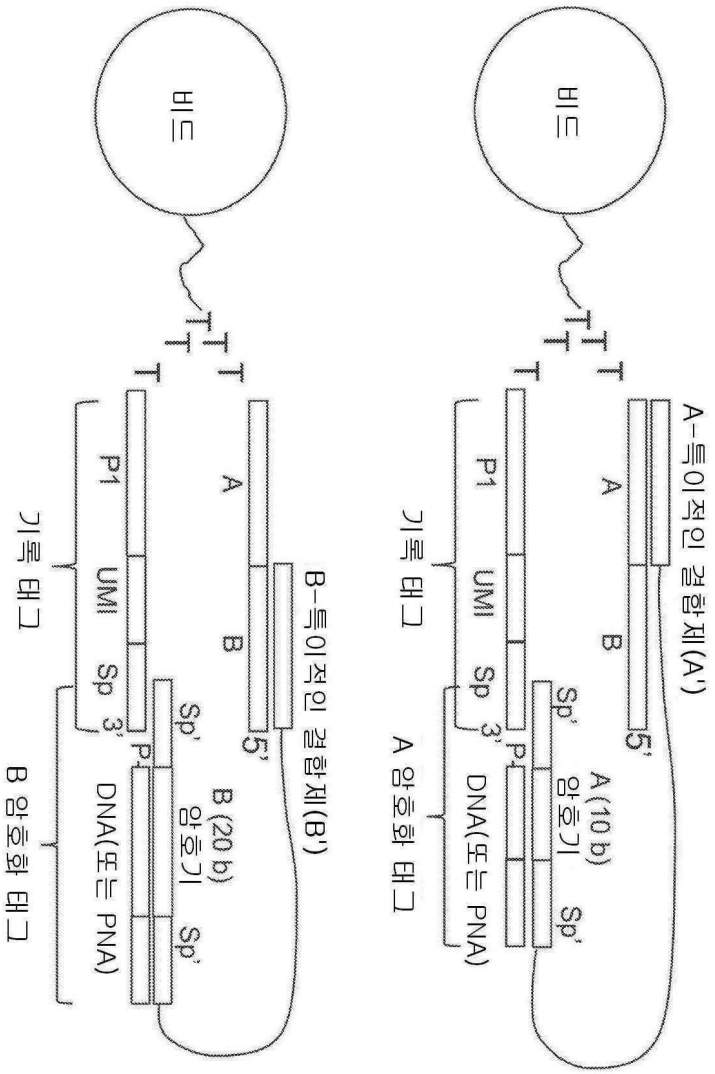
도면24a


비드
T
T
T
T
T
A
B
5'
A (10 b)
Sp' 암호기 BC Sp'
3'
P1
UMI
Sp
기록 태그
A 암호화 태그
A-특이적인 결합제(A')
비드
T
T
T
T
T
A
B
5'
B (20 b)
Sp' 암호기 BC Sp'
3'
P1
UMI
Sp
기록 태그
B 암호화 태그
B-특이적인 결합제(B')

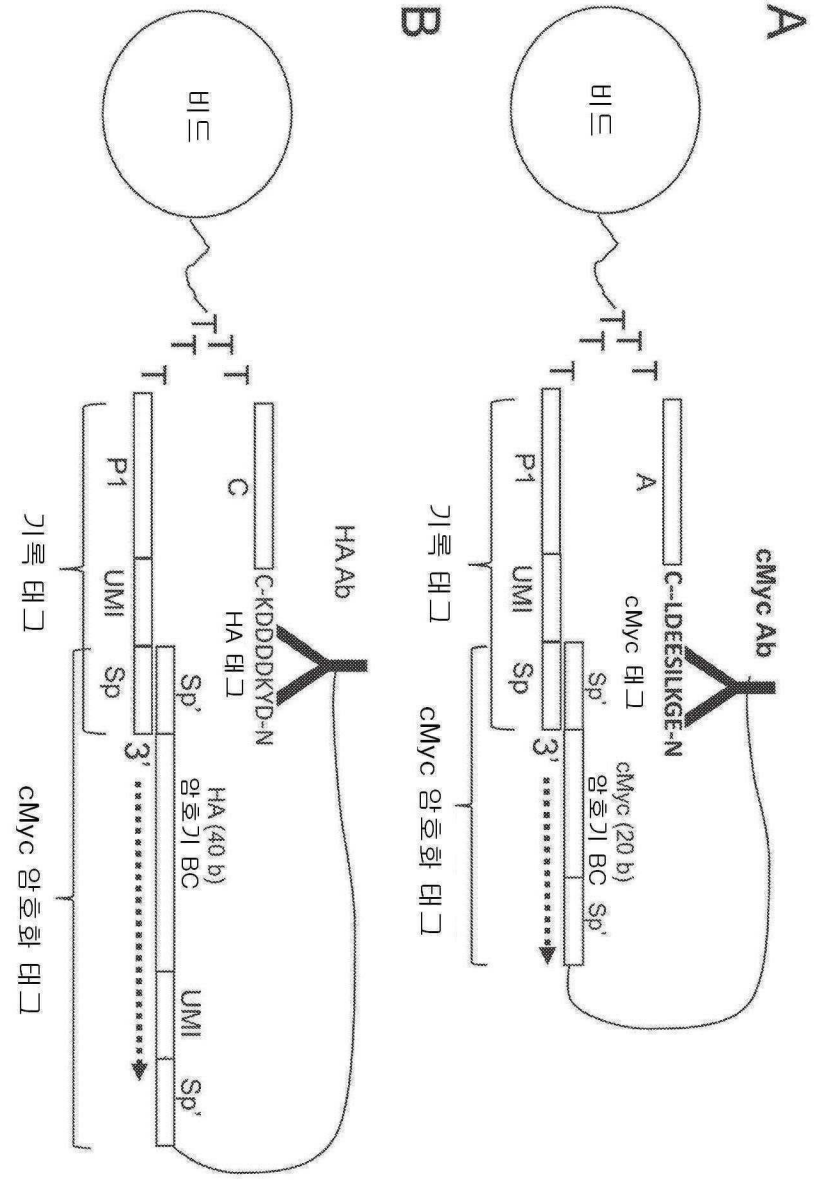
도면24b


비드
T
T
T
T
C-특이적인 결합제(C')
C
D
5'
Sp'
C (30 b)
암호기 BC
UMI
Sp'
3'
P1
UMI
Sp
기록 태그
C 암호화 태그
비드
T
T
T
T
D-특이적인 결합제(D')
C
D
5'
Sp'
D (40 b)
암호기 BC
UMI
Sp'
3'
P1
UMI
Sp
기록 태그
D 암호화 태그

도면25


비드
A-특이적인 결합제(A')
A
B
5'
A (10 b)
암호기
Sp'
Sp'
P1
UMI
Sp
3'
P
DNA(또는 PNA)
기록 태그
A 암호화 태그
비드
B-특이적인 결합제(B')
A
B
5'
B (20 b)
암호기
Sp'
Sp'
P1
UMI
Sp
3'
P
DNA(또는 PNA)
기록 태그
B 암호화 태그

도면26

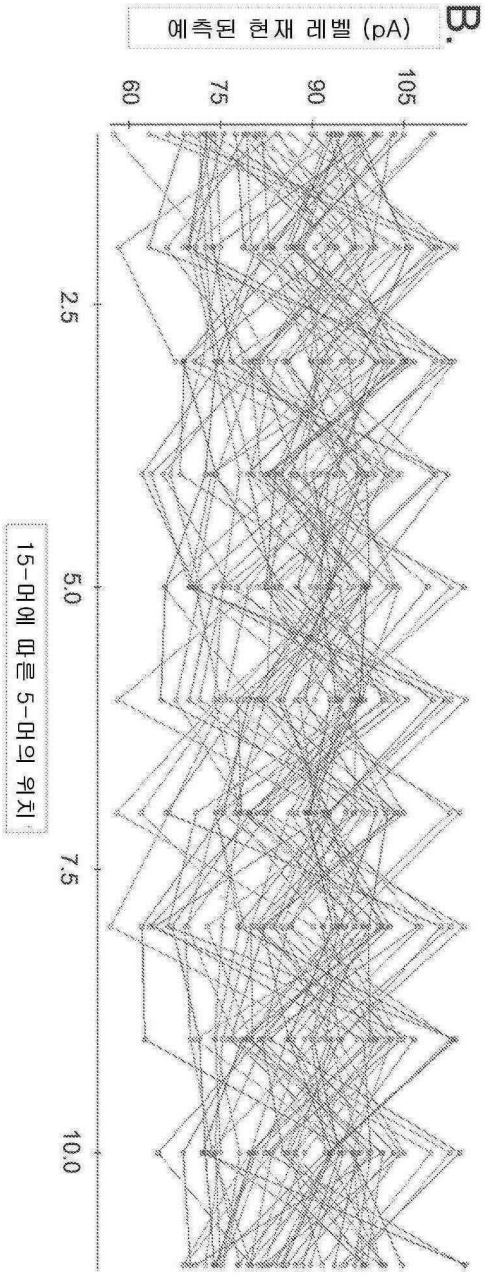

A
비드
T
T
T
A
C--LDEESILKGE-N
cMyc Ab
cMyc 태그
P1
UMI
Sp
Sp'
3'
cMyc (20 b)
암호기 BC
Sp'
기록 태그
cMyc 암호화 태그
B
비드
T
T
T
C
C-KDDDDKYD-N
HA 태그
HA Ab
P1
UMI
Sp
Sp'
3'
HA (40 b)
암호기 BC
UMI
Sp'
기록 태그
cMyc 암호화 태그

도면27a

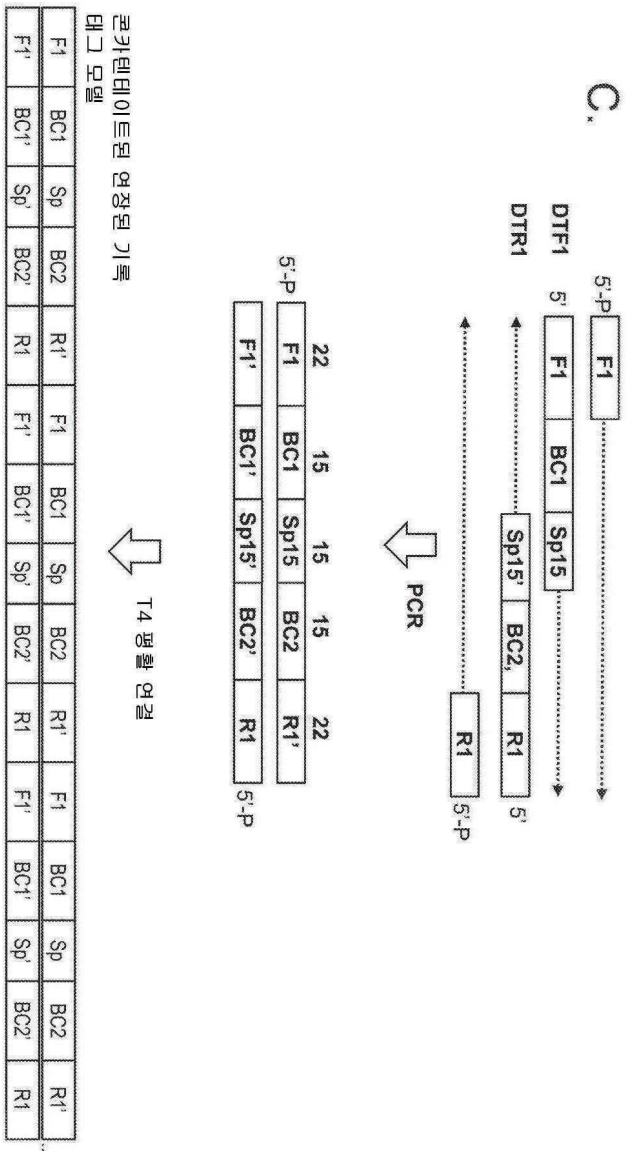
A.

ATGTCTAGCATGCCG	CCGCTTGTGATCTGG	CTGGATGGTTGTCGA	GATTCAACGGAGAAG	GCCATCAGTAAGAGA	GTCCTCGTCTATCCT	CGTTACAGGTCTGCG	GGCCAACTAAGGTGC
CCGTGTCATGTGGAA	AGATAGCGTACCGGA	ACTGCACCGTTCCAA	ACGAACCTCGCACCA	GCAACGTGAATTGAG	TAGGATTCCGTTACC	GATCATTGGCCAATT	GCACCTATTCGACAA
TAAGCCGGTATATCA	TCCAGGCTCATCATC	CGAGAGATTGGTCCTT	AGGACTTCAAGAAGA	CTAAGTAGAGCCACA	TCTGACCACCGGAAG	TTCAAGGCTGAGTTG	TGGACACGATCGGCT
TTCGATATGACGGAA	GAGTACTAGAGCCAA	TCTTGAGAGACAAGA	GGTTGAATCCTCGCA	TGTCTGTTTGGAAGCG	AGAGTCACCTCTGTGG	TGGCTCGATTGAATC	CTATAATTCCAACGG
CGTATCCGCGTTAGG	GAGCGTCAATAACGG	AATTCGCACTGTGTT	AACCAACCTCTAGCG	TTAATAGACAGCGCG	CTGATGTAGTCGAAG	GTAAGCCATCCGCTC	AACGTGGTTAGTAAG
AACTGCCGAGATTCC	GCGGTATCTACACTG	GTAGTGCCGCTAAGA	ACGCGAATATCTAAC	CGACGCTCTAACAAG	GTCGGTTGCGGATAG	ACACATCGCTAGACA	CAAGGAACGAGTGGC
TCATCTTAGCTGTGC	CTTCTCCGAAGAGAA	CCTATAGCACAATCC	GTTCAGAATTACACC	CATGGCTTATTGAGA	TCCTCCTCCTAAGAA	TGCTATGGATTCAAG	CACCAGAACCGAAGA
GAGTCCGGTACCTTCA	TCAACCGAAGGTTCGA	CTCTCTCTGTGAACC	ACTAGGTATGGCCCGG	ATTCGGTCCACTTCA	CCACGAGGCTTAGTT	CGTACGGTCAAGCAA	TCGGTGACAGGCTAA

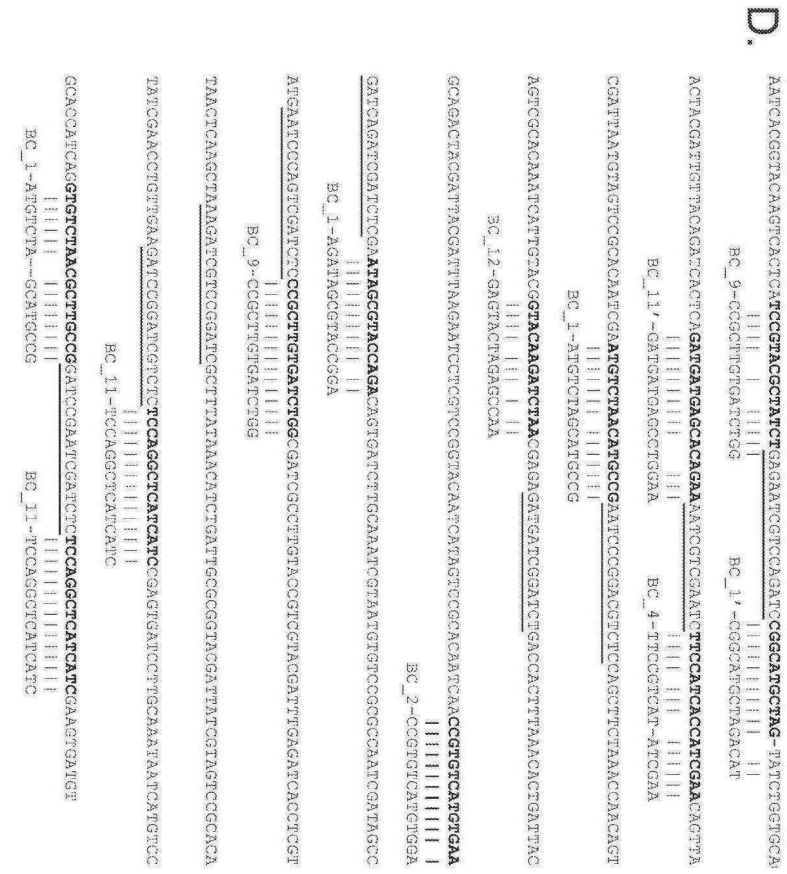
도면27b



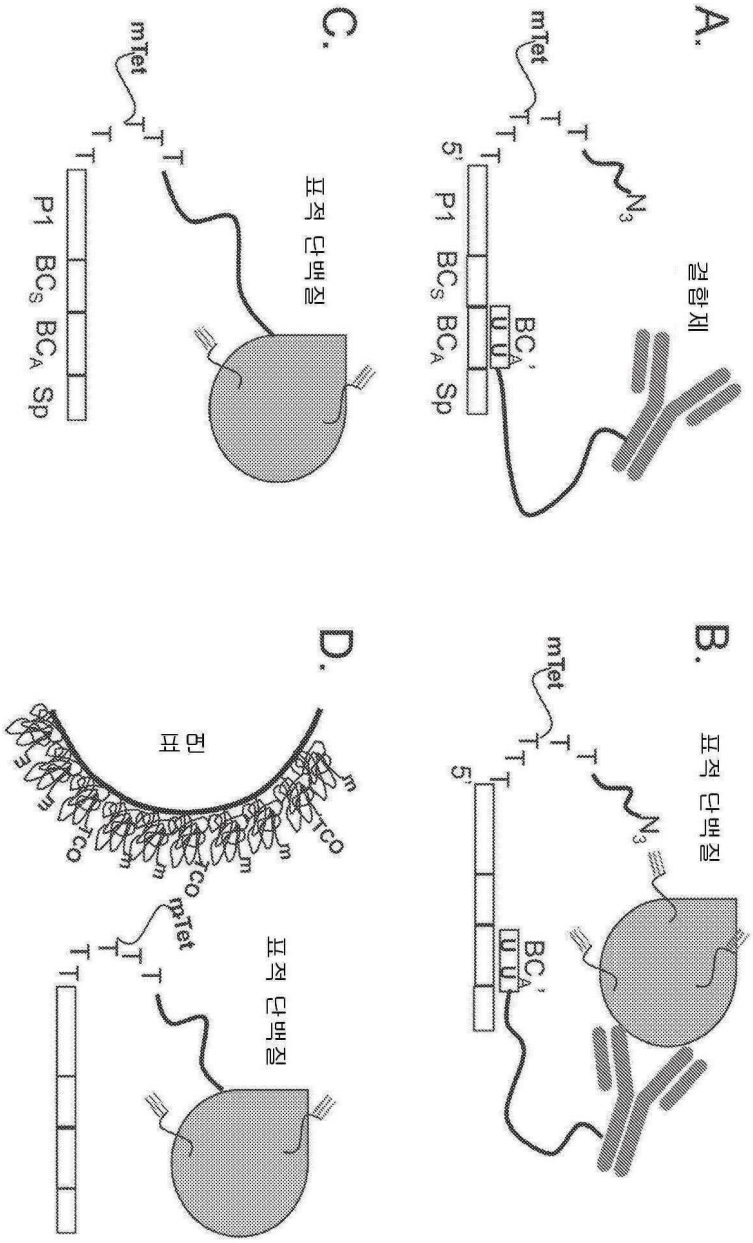
도면27c


콘카테네이트된 연장된 기록
태그 모델
T4 폴리 연결
PCR
DTF1
DTR1
5'-P
F1
BC1
Sp15
Sp15'
BC2'
R1
22
15
15
15
22
F1'
BC1'
Sp15'
BC2'
R1'
Sp
BC2
Sp'
BC2'
R1'

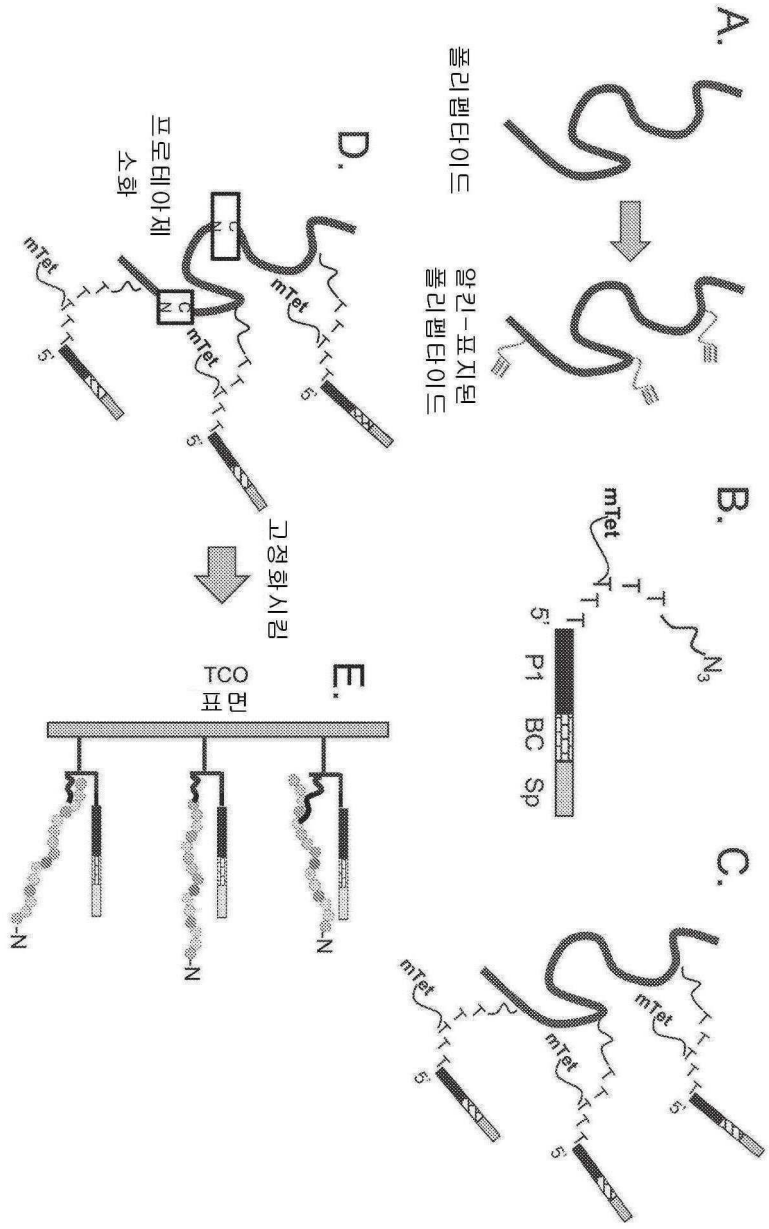
도면27d



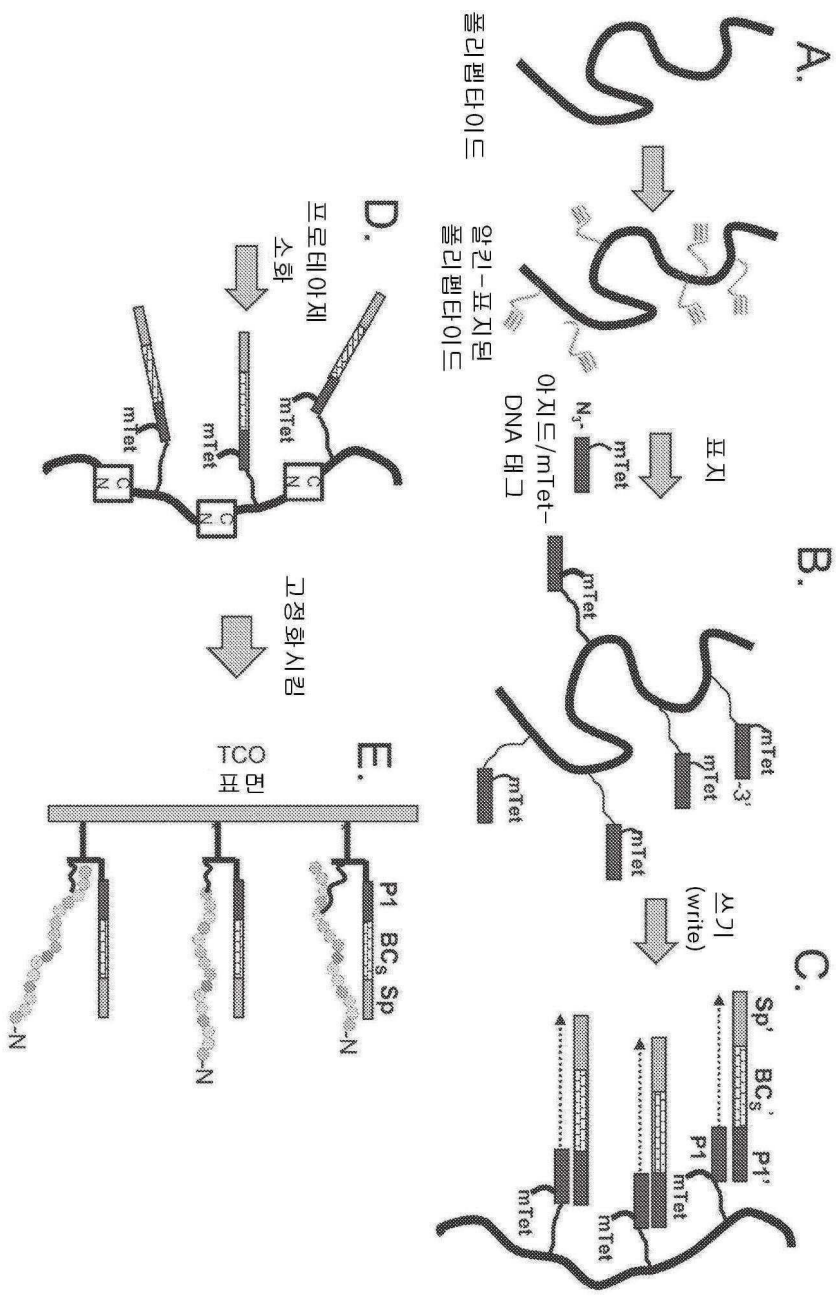
도면28


A.
mTet
T
T
T
N_3
5'
P1
BC_S
BC_A
Sp
BC_A'
결합제
B.
표적 단백질
C.
표적 단백질
P1
BC_S
BC_A
Sp
D.
표면
TCO
표적 단백질

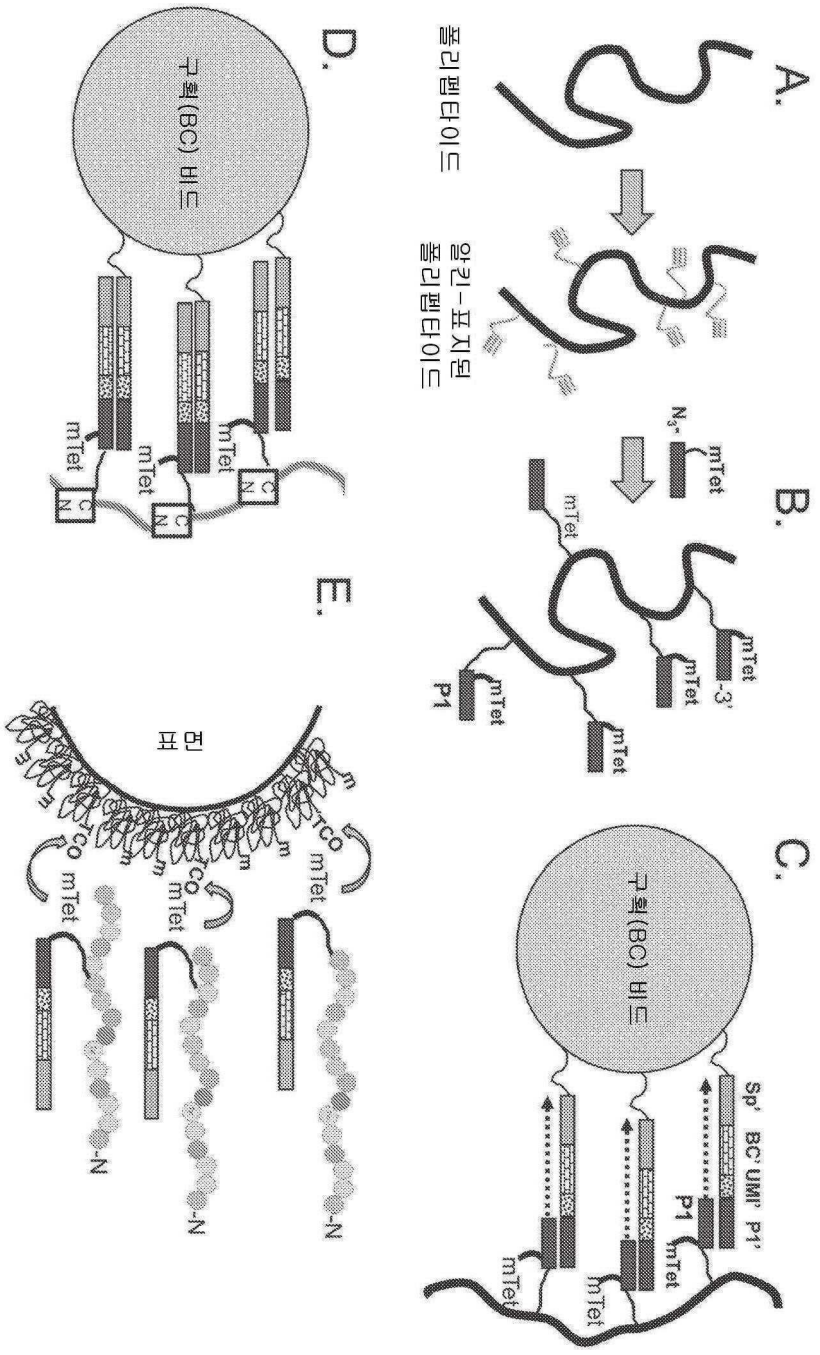
도면29


A.
폴리펩타이드
알킨-표지된
폴리펩타이드
B.
mTet
T
T
T
5'
P1
BC
Sp
N3
C.
mTet
T
5'
D.
프로테아제
소화
C
N
고정화시킴
E.
TCO
표면
N

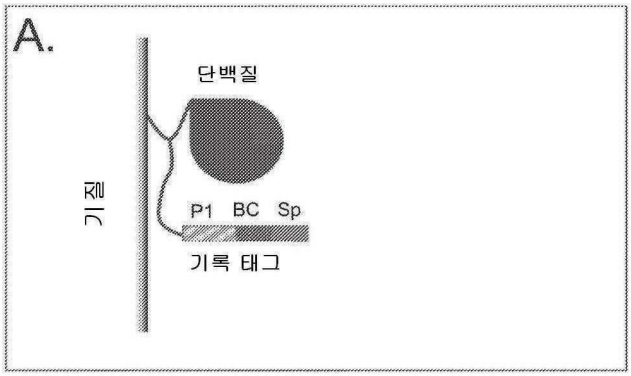
도면30


A.
폴리펩타이드
알킨-표지된
폴리펩타이드
표지
아지드/mTet-
DNA 태그
B.
쓰기
(write)
C.
D.
프로테아제
소화
고정화시킴
E.
TCO
표면
mTet
P1
BCs
Sp
N

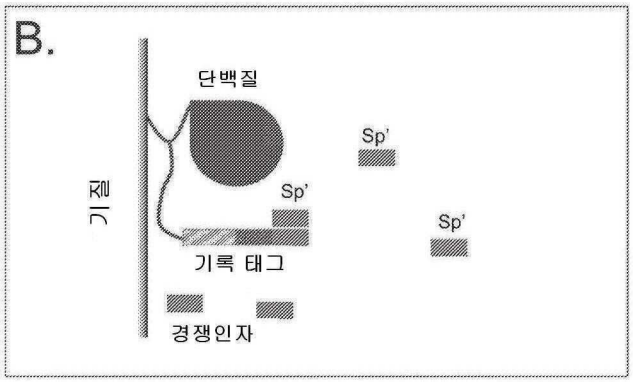
도면31


A.
프로브태깅드 폴리펩타이드
올리고-표지된 폴리펩타이드
B.
mTet
N3-
-3'
P1
C.
구획(BC) 비드
Sp'
BC'
UMI'
P1'
D.
구획(BC) 비드
E.
표면
-N

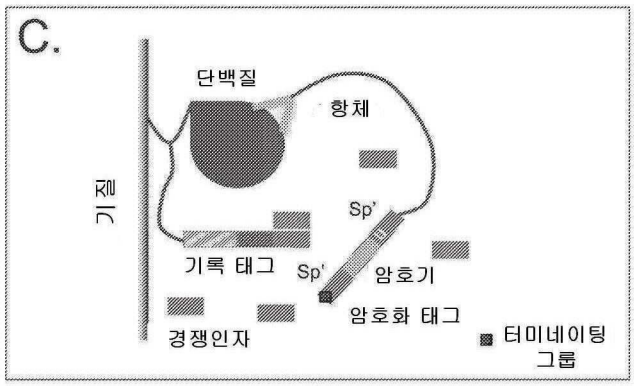
도면32a



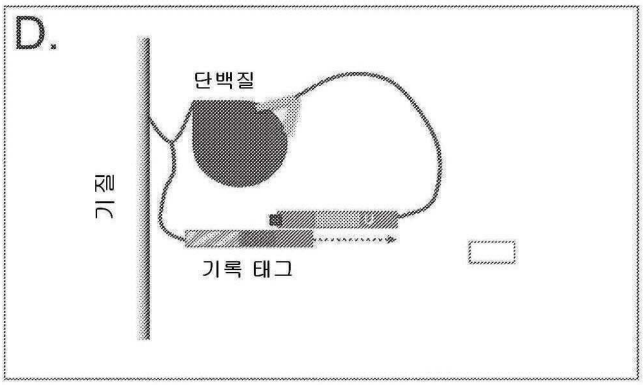
도면32b



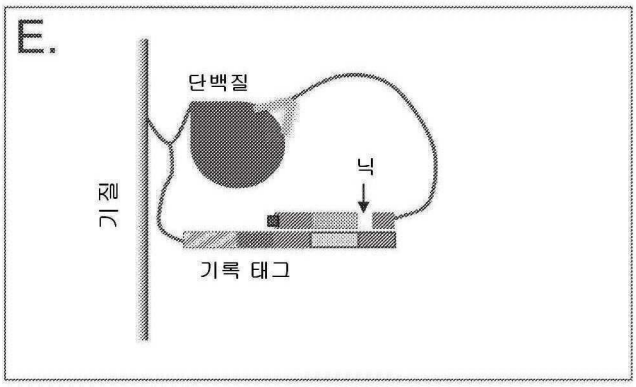
도면32c



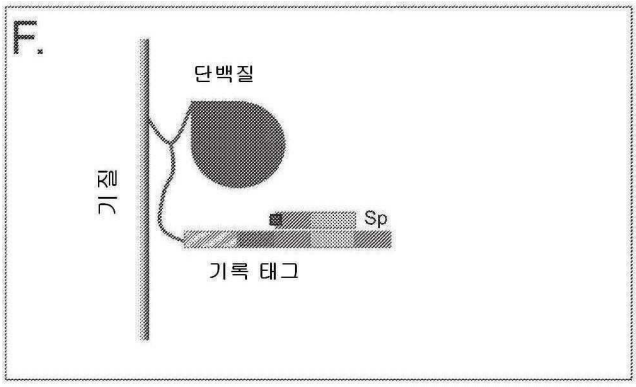
도면32d



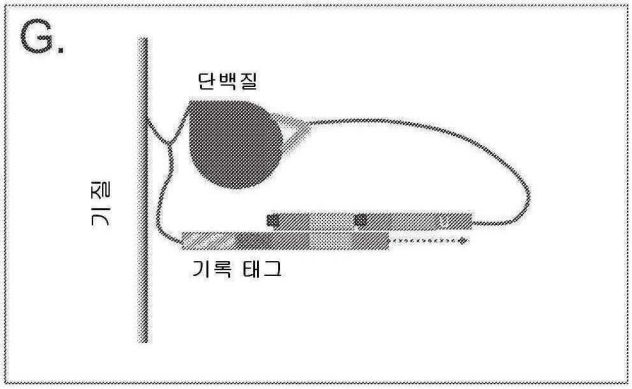
도면32e



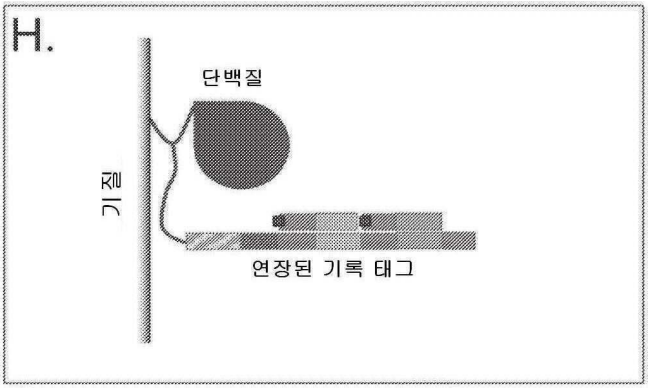
도면32f



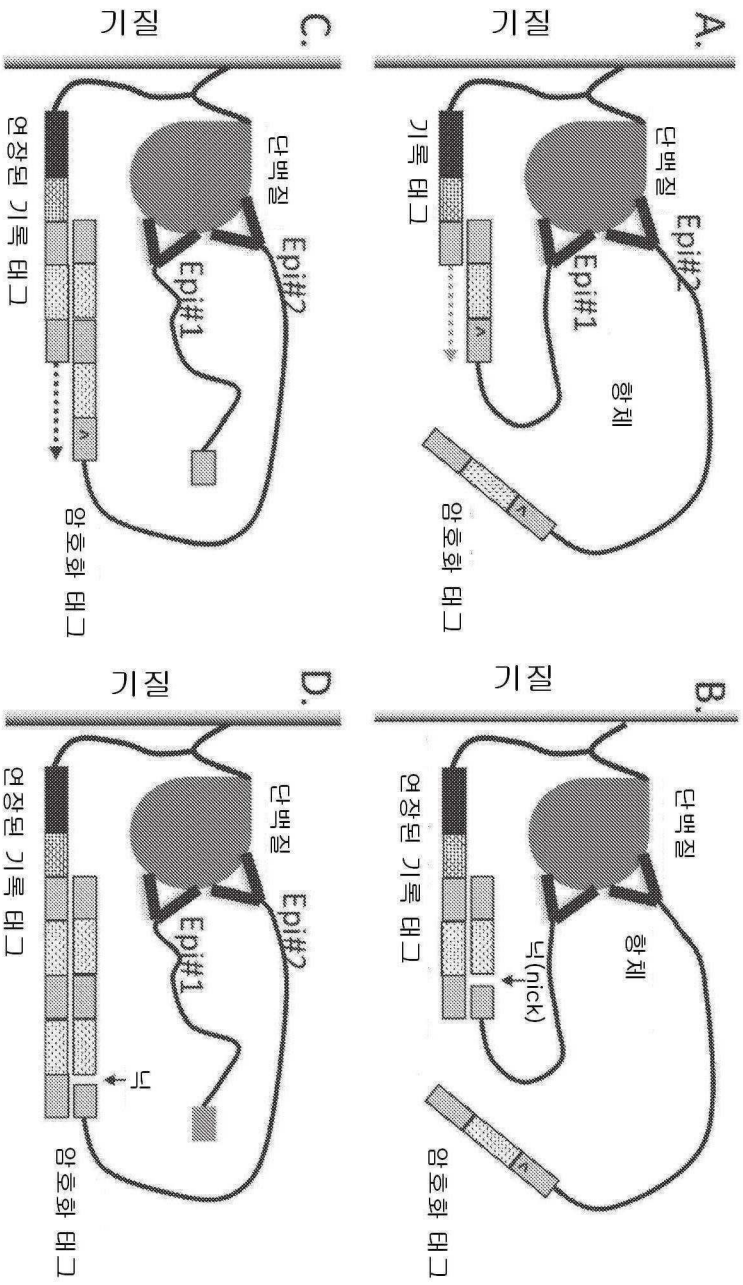
도면32g



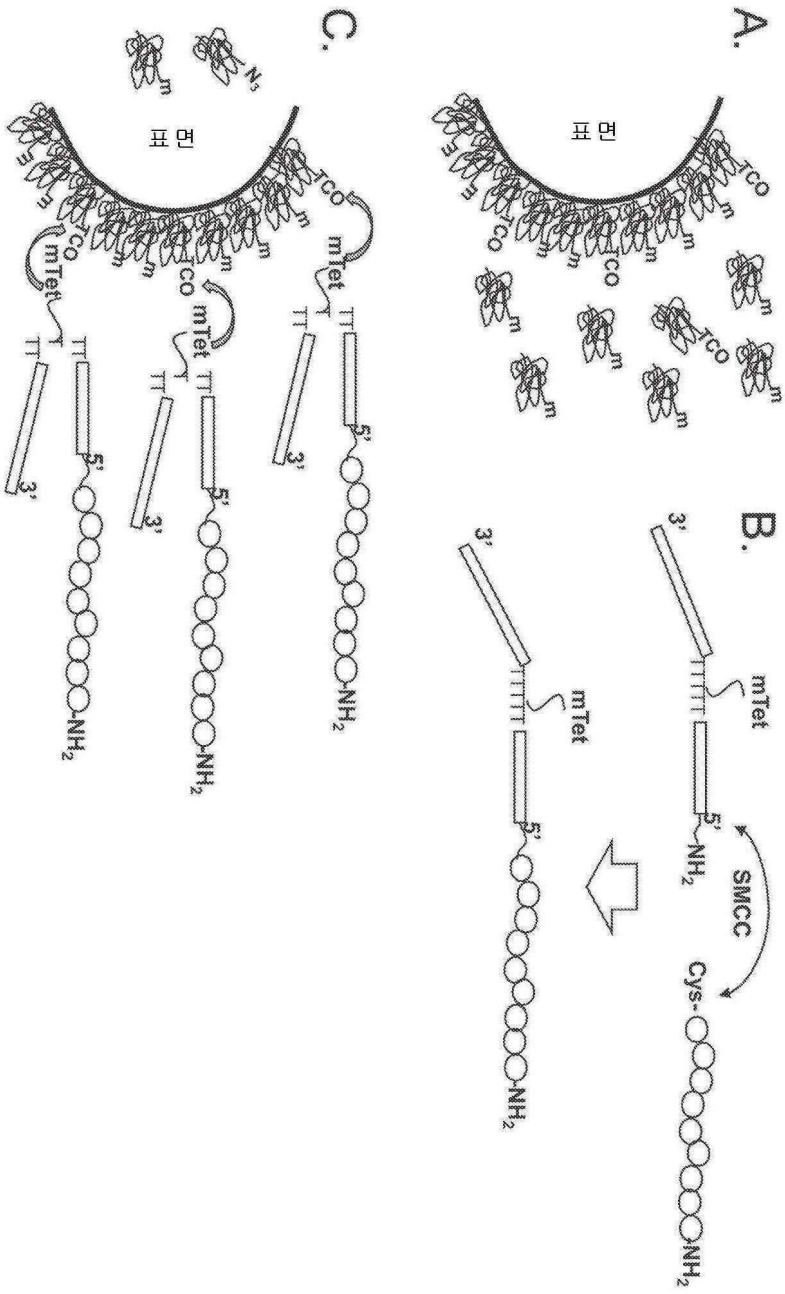
도면32h



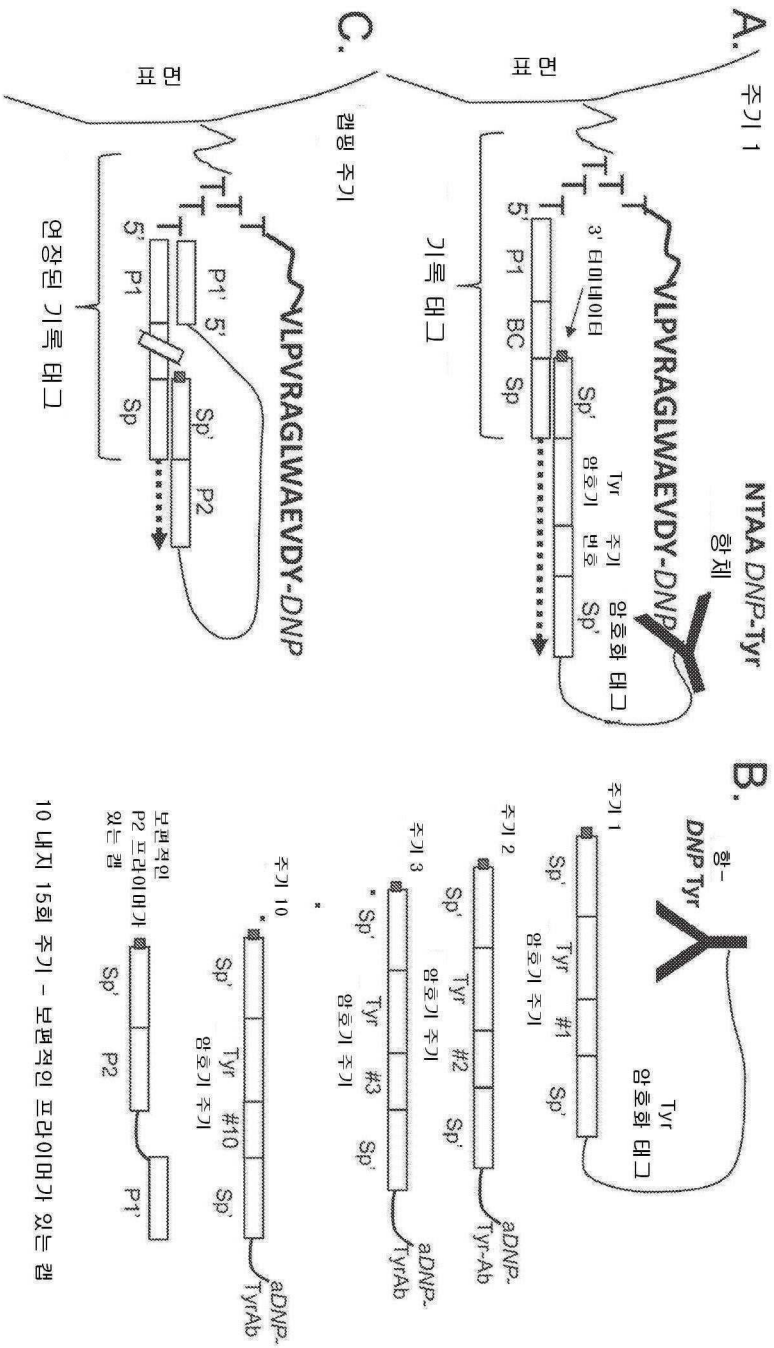
도면33


A.
기질
단백질
Epi#2
Epi#1
항체
기록 태그
암호화 태그
B.
기질
단백질
항체
닉(nick)
연장된 기록 태그
암호화 태그
C.
기질
단백질
Epi#2
Epi#1
연장된 기록 태그
암호화 태그
D.
기질
단백질
Epi#2
Epi#1
닉
연장된 기록 태그
암호화 태그

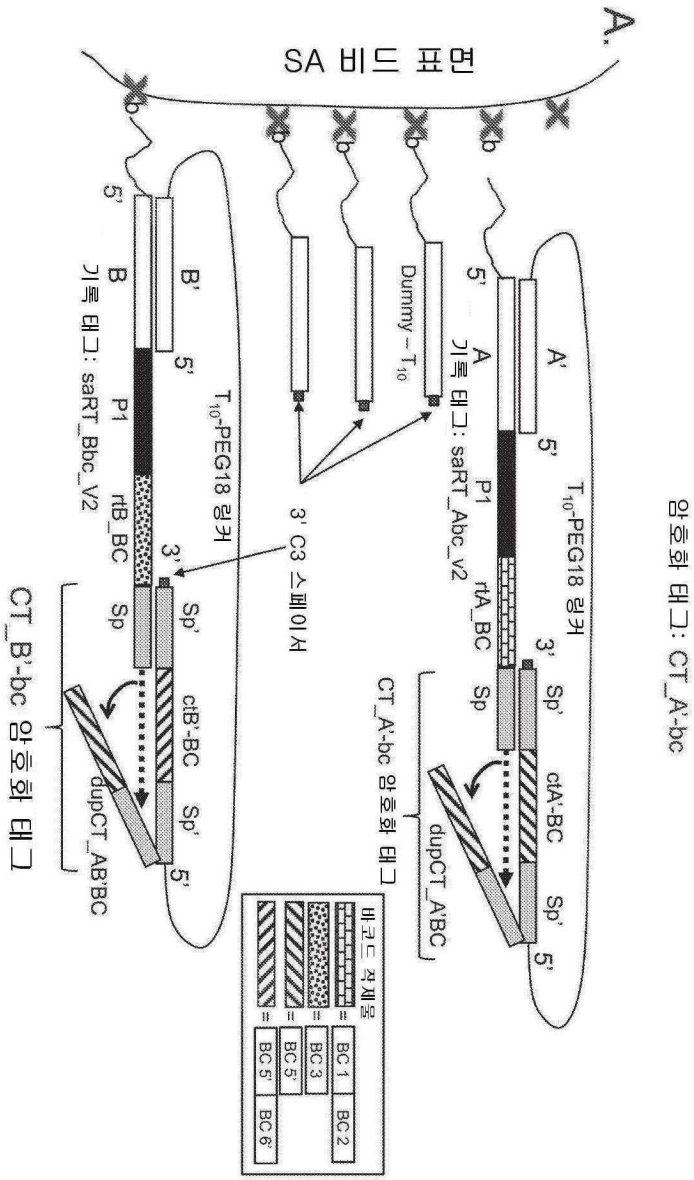
도면34



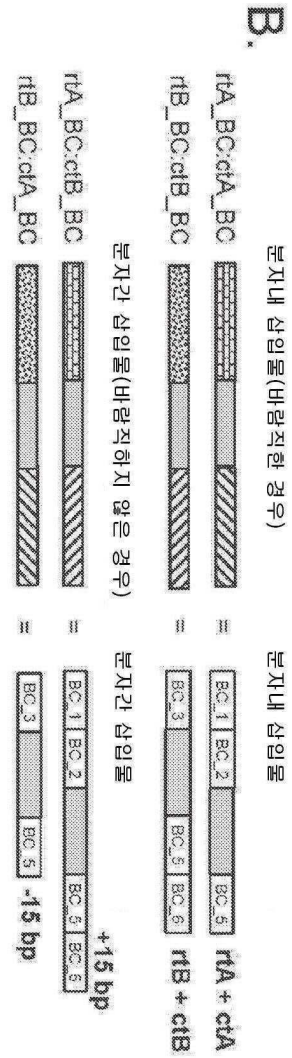
도면35


A.
주기 1
표면
NTAA DNP-Tyr
항체
VLPVRAGLWAEVDY-DNP
3' 터미네이터
5' P1 BC Sp
Sp'
Tyr 암호기
주기 번호
Sp'
암호화 태그
기록 태그
B.
항-DNP Tyr
Tyr 암호화 태그
주기 1
Sp' Tyr #1 Sp'
암호기 주기
주기 2
Sp' Tyr #2 Sp'
암호기 주기
aDNP-Tyr-Ab
주기 3
Sp' Tyr #3 Sp'
암호기 주기
aDNP-TyrAb
주기 10
Sp' Tyr #10 Sp'
암호기 주기
aDNP-TyrAb
보편적인 P2 프라이머가 있는 캡
Sp' P2
P1'
10 내지 15회 주기 - 보편적인 프라이머가 있는 캡
C.
캡핑 주기
표면
연장된 기록 태그
5' P1
P1' 5'
Sp' P2
Sp
VLPVRAGLWAEVDY-DNP

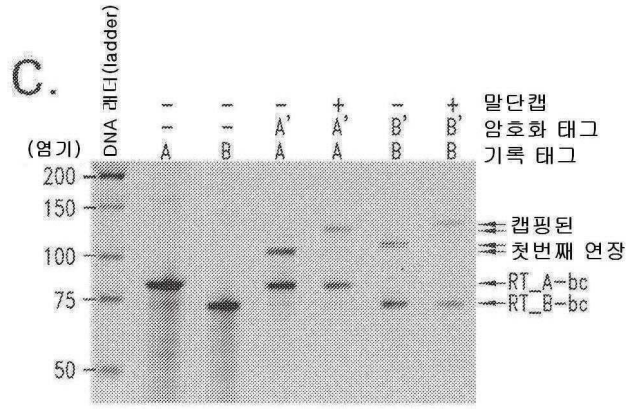
도면36a


A.
SA 비드 표면
Dummy – T10
5'
B
기록 태그: saRT_Bbc_V2
P1
rtB_BC
Sp
3'
B'
5'
T10-PEG18 링커
3' C3 스페이서
Sp'
ctB'-BC
Sp'
5'
dupCT_AB'BC
CT_B'-bc 암호화 태그
A'
5'
T10-PEG18 링커
3'
Sp'
ctA'-BC
Sp'
5'
A
기록 태그: saRT_Abc_v2
P1
rtA_BC
Sp
dupCT_A'BC
CT_A'-bc 암호화 태그
암호화 태그: CT_A'-bc
바코드 삭제물
BC 1
BC 2
BC 3
BC 5'
BC 5'
BC 6'

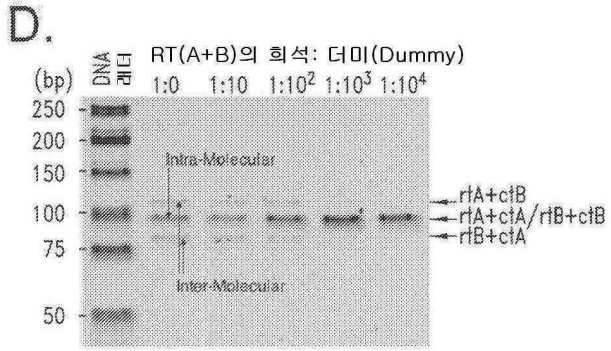
도면36b


B.
rtA_BC:ctA_BC
rtB_BC:ctB_BC
rtA + ctA
rtB + ctB
BC_1
BC_2
BC_3
BC_5
BC_6
콜임상 내 군자금
콜임상 내 군자금
(바람직한 형 상)
콜임상 내 군자금
(바람직하지 않은 형 상)
콜임상 내 군자금
rtA_BC:ctB_BC
rtB_BC:ctA_BC
+15 bp
-15 bp

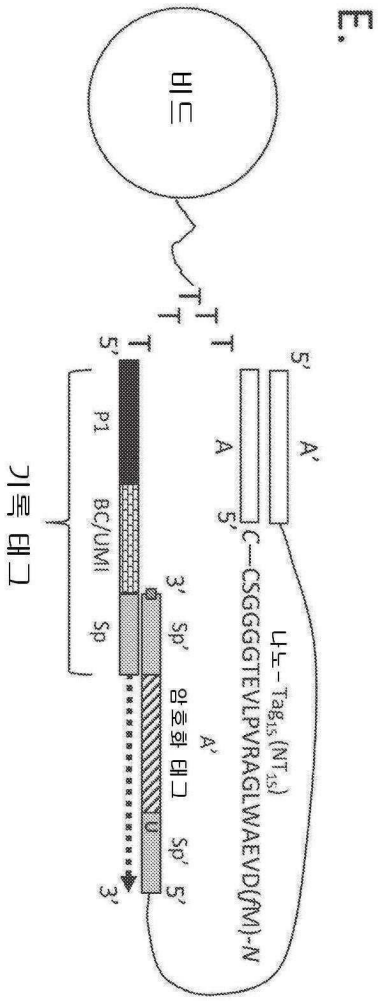
도면36c


C.
DNA 래더(ladder)
(염기)
200
150
100
75
50
-
-
A
-
B
-
A'
+
A'
A
-
B'
B
+
B'
B
말단캡
암호화 태그
기록 태그
캡핑된
첫번째 연장
RT_A-bc
RT_B-bc

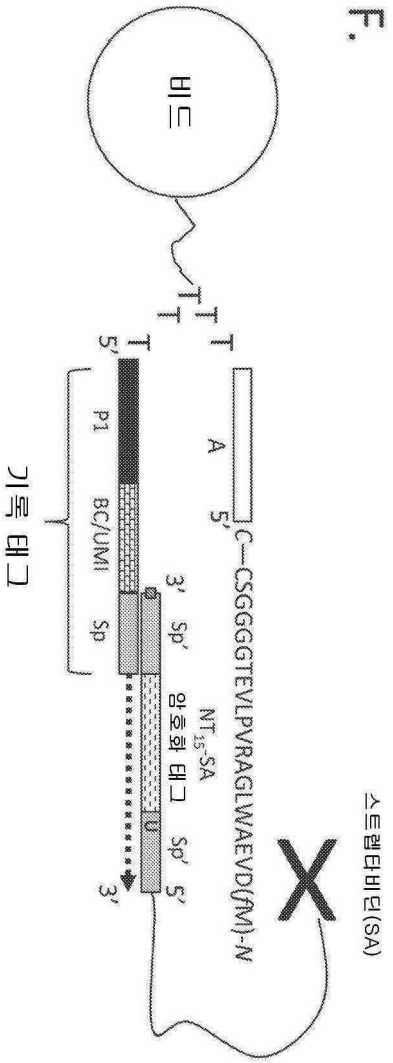
도면36d


D.
RT(A+B)의 희석: 더미(Dummy)
(bp)
DNA 래더
1:0
1:10
1:10^2
1:10^3
1:10^4
250
200
150
100
75
50
Intra-Molecular
Inter-Molecular
rtA+ctB
rtA+ctA/rtB+ctB
rtB+ctA

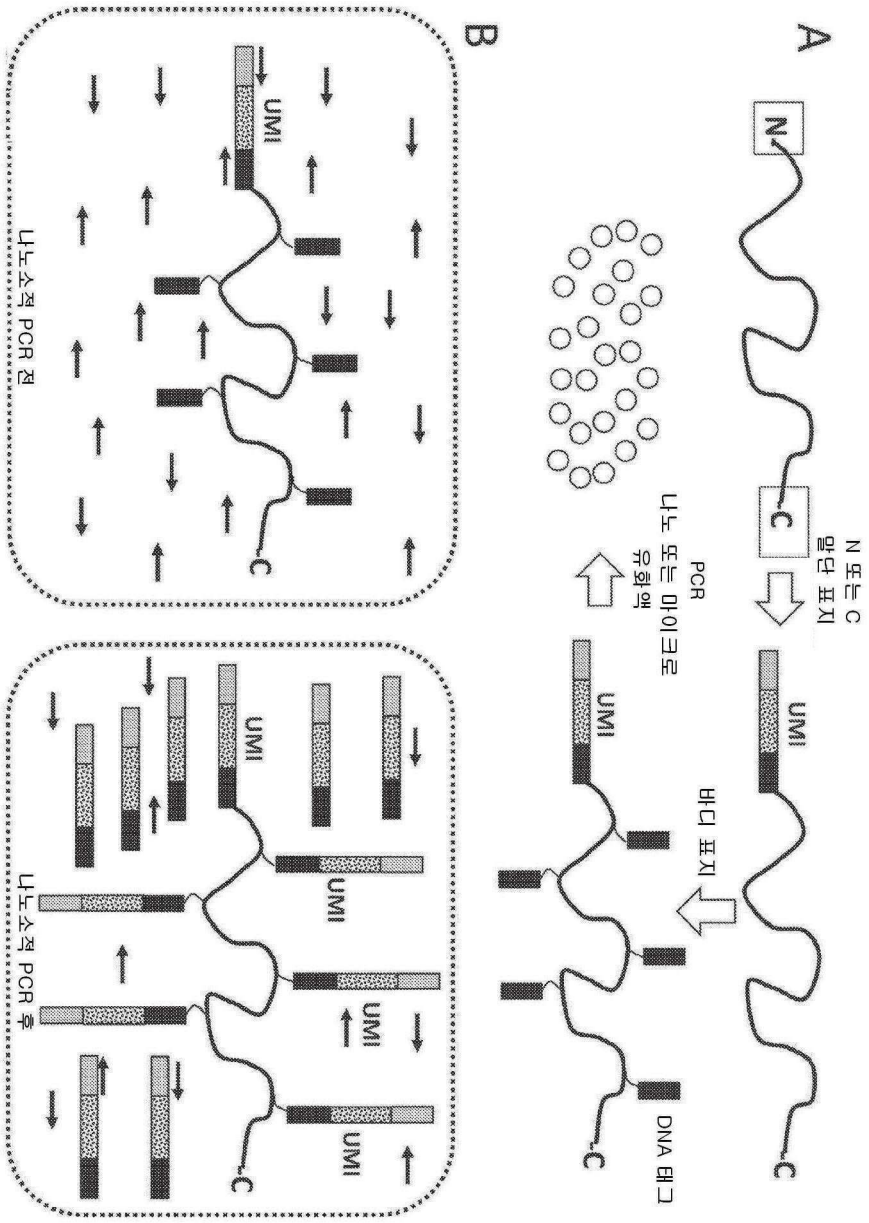
도면36e


E.
비드
5'
T
P1
BC/UMI
Sp
기록 태그
3'
Sp'
A'
앰플화 태그
U
Sp'
5'
A
A'
나노 - Tag15 (NT15)
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N

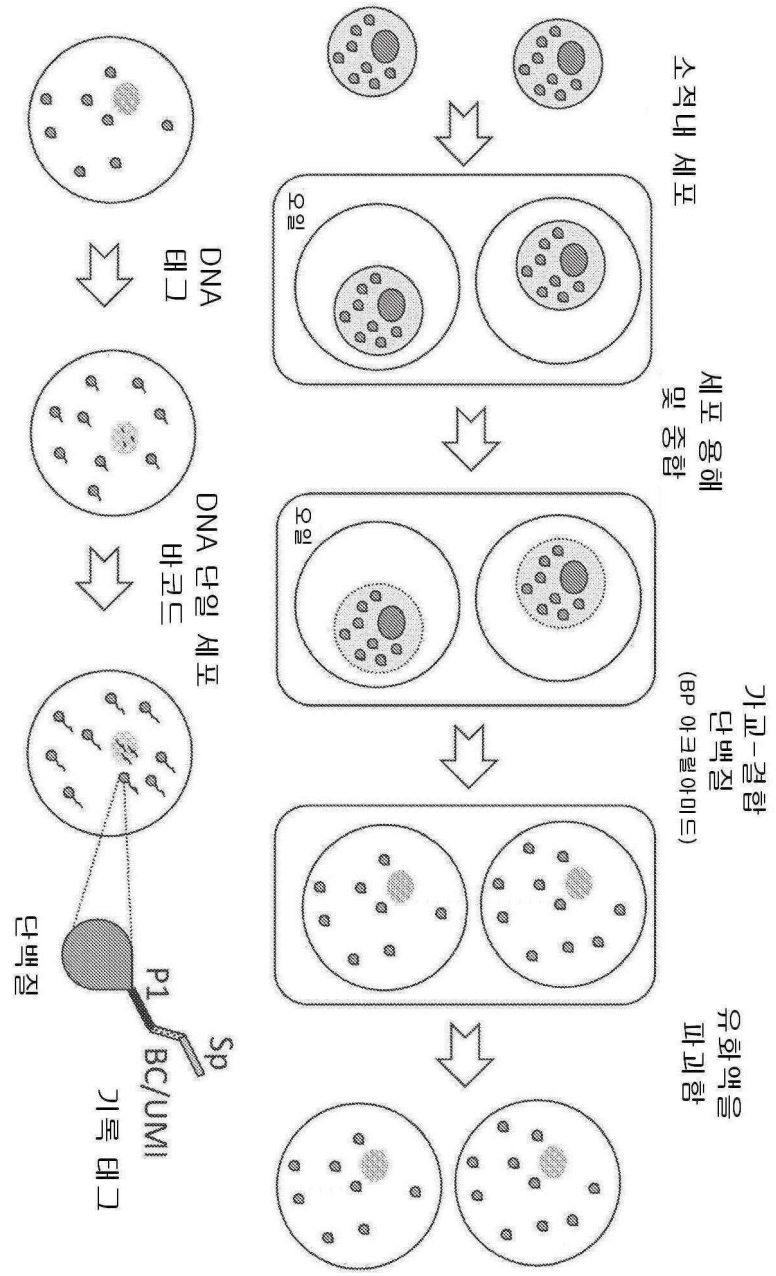
도면36f


F.
비드
5'
P1
BC/UMI
Sp
기록 태그
3'
Sp'
암호화 태그
U
Sp'
5'
NT15-SA
A
5'
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N
스트렙타비딘(SA)

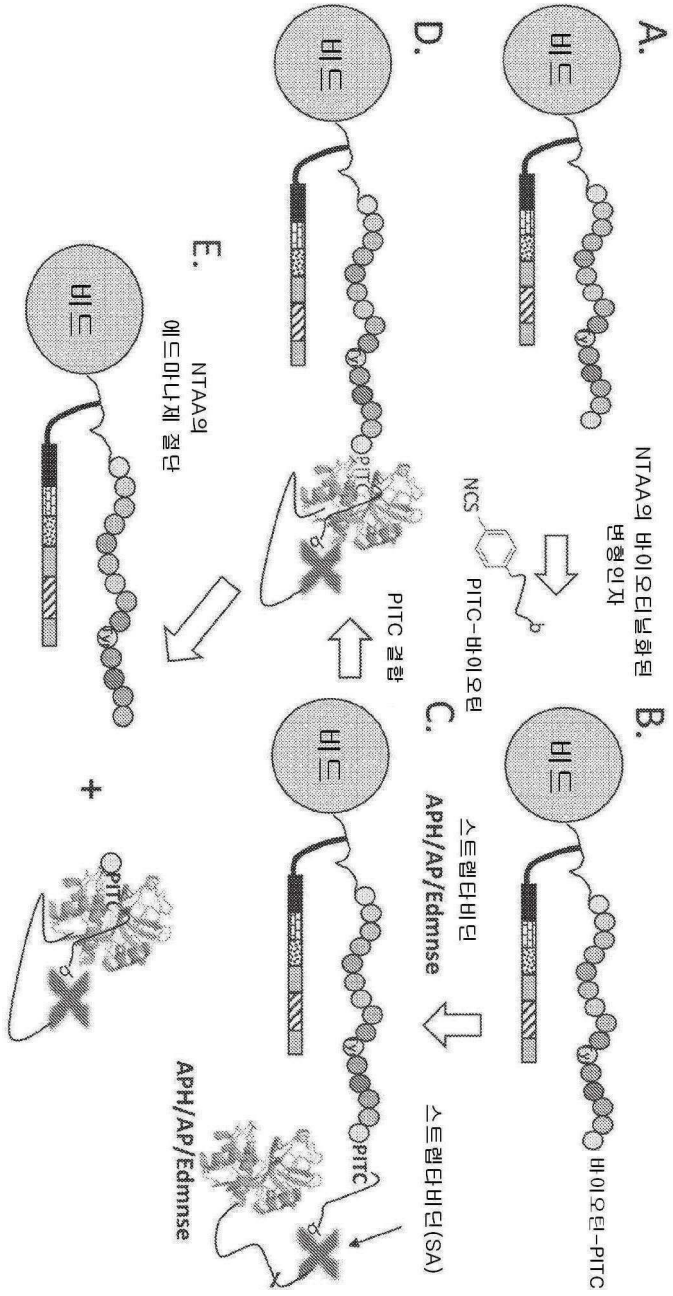
도면37


A
N
C
탐침 표지
N 또는 C
나노 또는 마이크로
PCR
유화액
바디 표지
DNA 태그
UMI
B
나노스적 PCR 전
나노스적 PCR 후

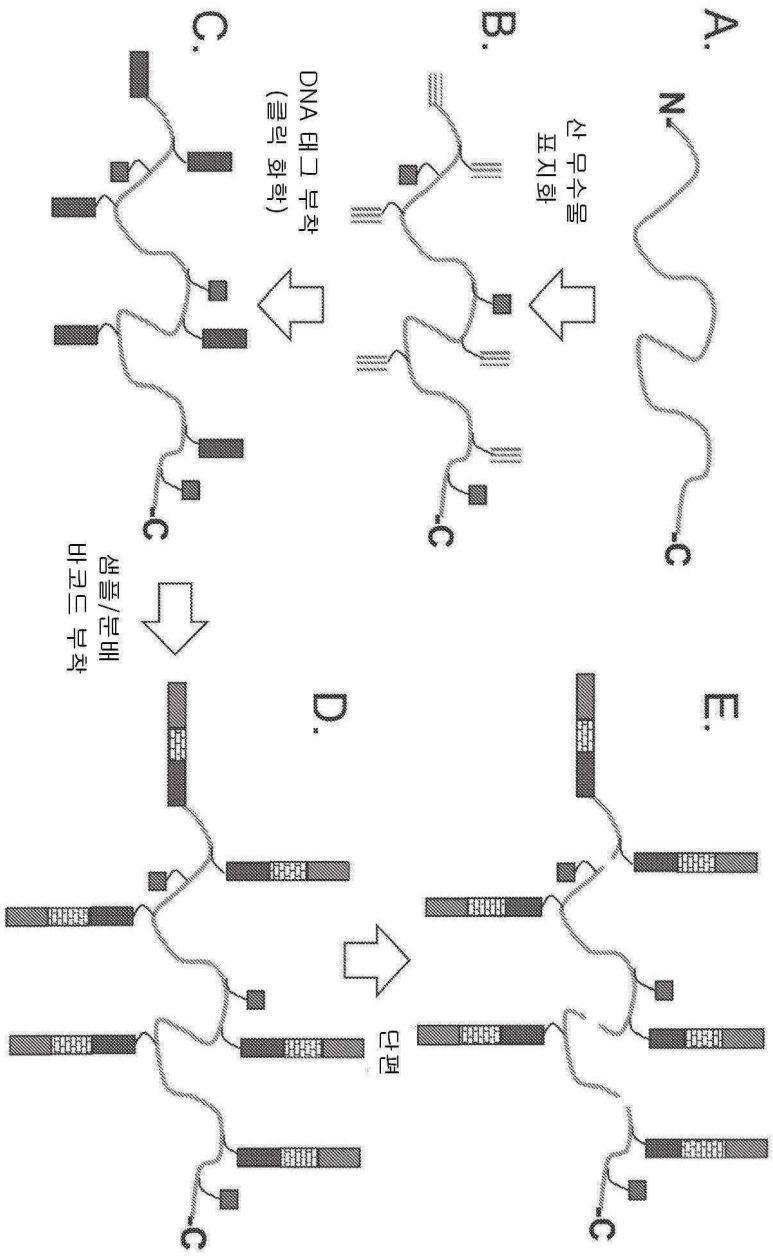
도면38


수득내 세포
세포 용해 및 중합
가교-결합 단백질 (BP 아크릴아미드)
유화액을 파괴함
DNA 태그
DNA 단일 세포 바코드
단백질
P1
Sp
BC/UMI
기록 태그

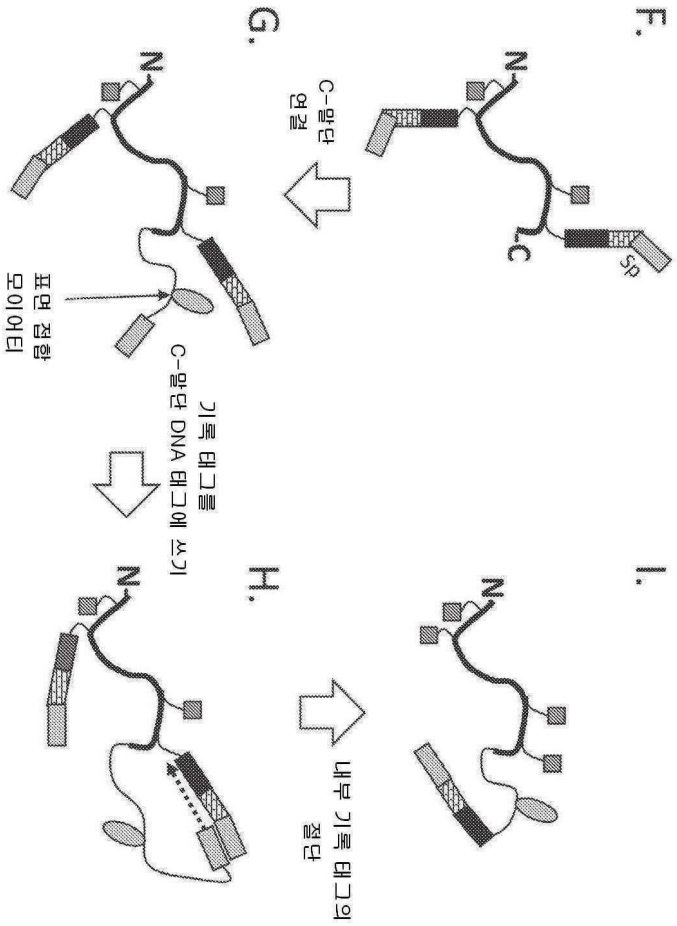
도면39


A.
비드
NTAA의 바이오티닐화를 위한 링커
NCS
PITC-바이오틴
B.
비드
바이오틴-PITC
C.
스트렙타비딘
APH/AP/Edmnse
스트렙타비딘(SA)
D.
비드
PITC 결합
E.
NTAA의
에드마나제 절단
비드
+
APH/AP/Edmnse

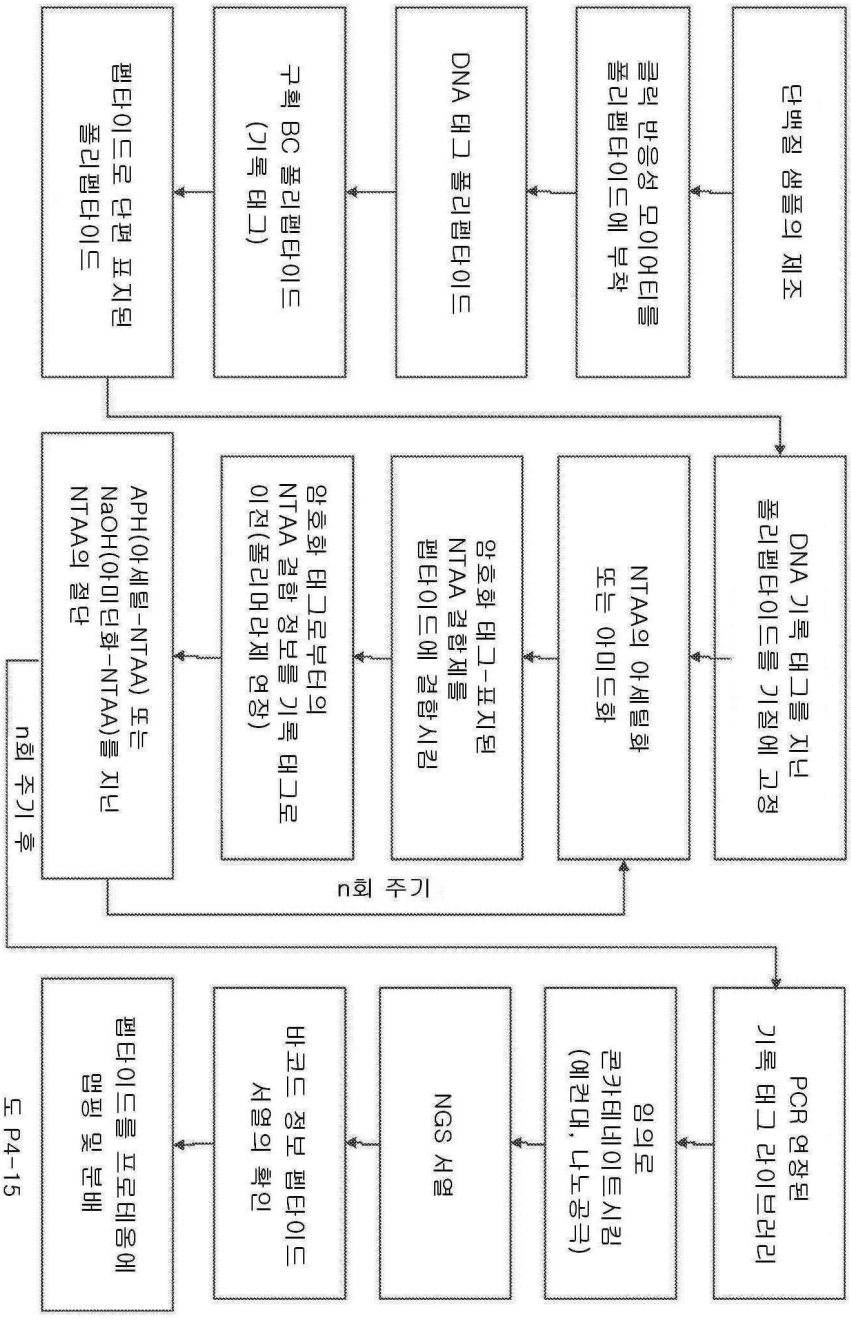
도면40a



도면40b

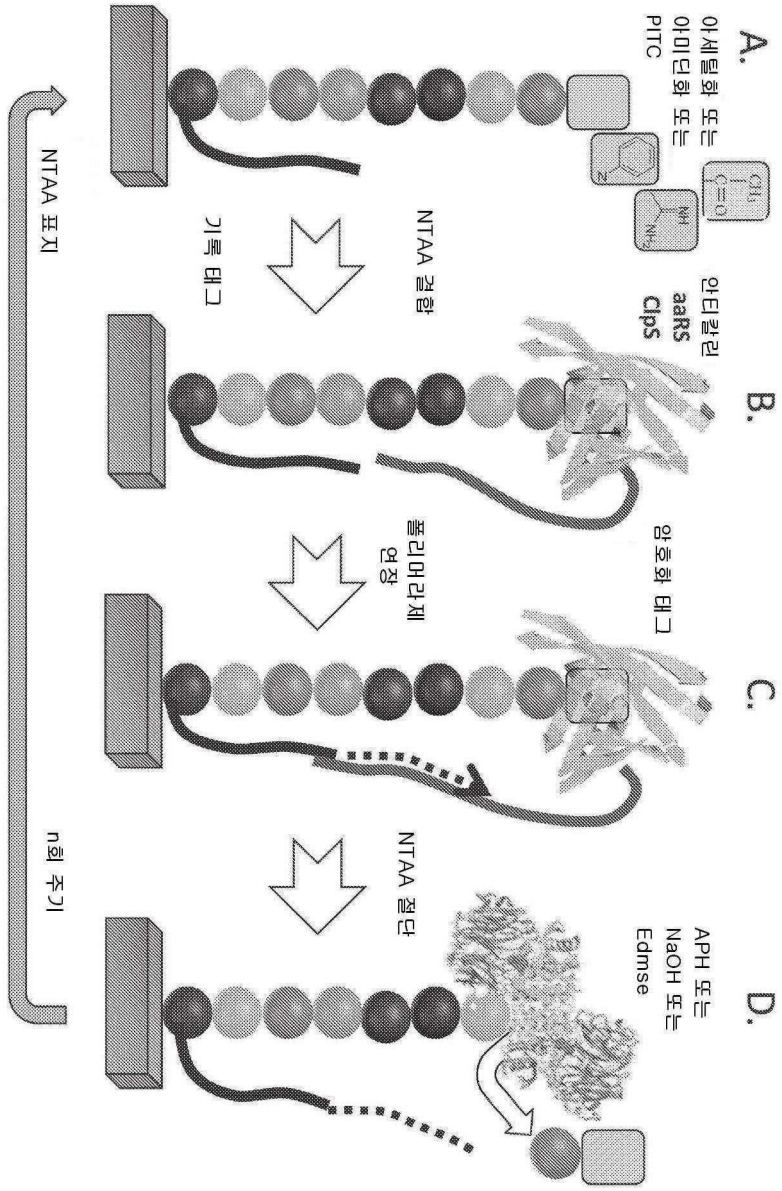


도면41

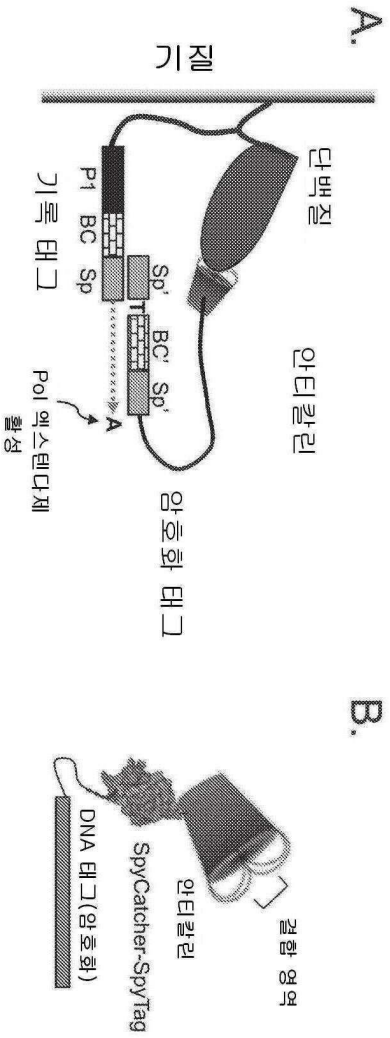

단백질 샘플의 제조
클릭 반응성 모이어티를 폴리펩타이드에 부착
DNA 태그 폴리펩타이드
구획 BC 폴리펩타이드 (기록 태그)
펩타이드로 단편 표지된 폴리펩타이드
DNA 기록 태그를 지닌 폴리펩타이드를 기질에 고정
NTAA의 아세틸화 또는 아미드화
암호화 태그-표지된 NTAA 결합체를 펩타이드에 결합시킴
암호화 태그로부터의 NTAA 결합 정보를 기록 태그로 이전(폴리머라제 연장)
APH(아세틸-NTAA) 또는 NaOH(아미딘화-NTAA)를 지닌 NTAA의 절단
n회 주기
n회 주기 후
PCR 연장된 기록 태그 라이브러리
임의로 콘카테네이트시킴 (에컬대, 나노공극)
NGS 서열
바코드 정보 펩타이드 서열의 확인
펩타이드를 프로테옴에 맵핑 및 분배

도 P4-15

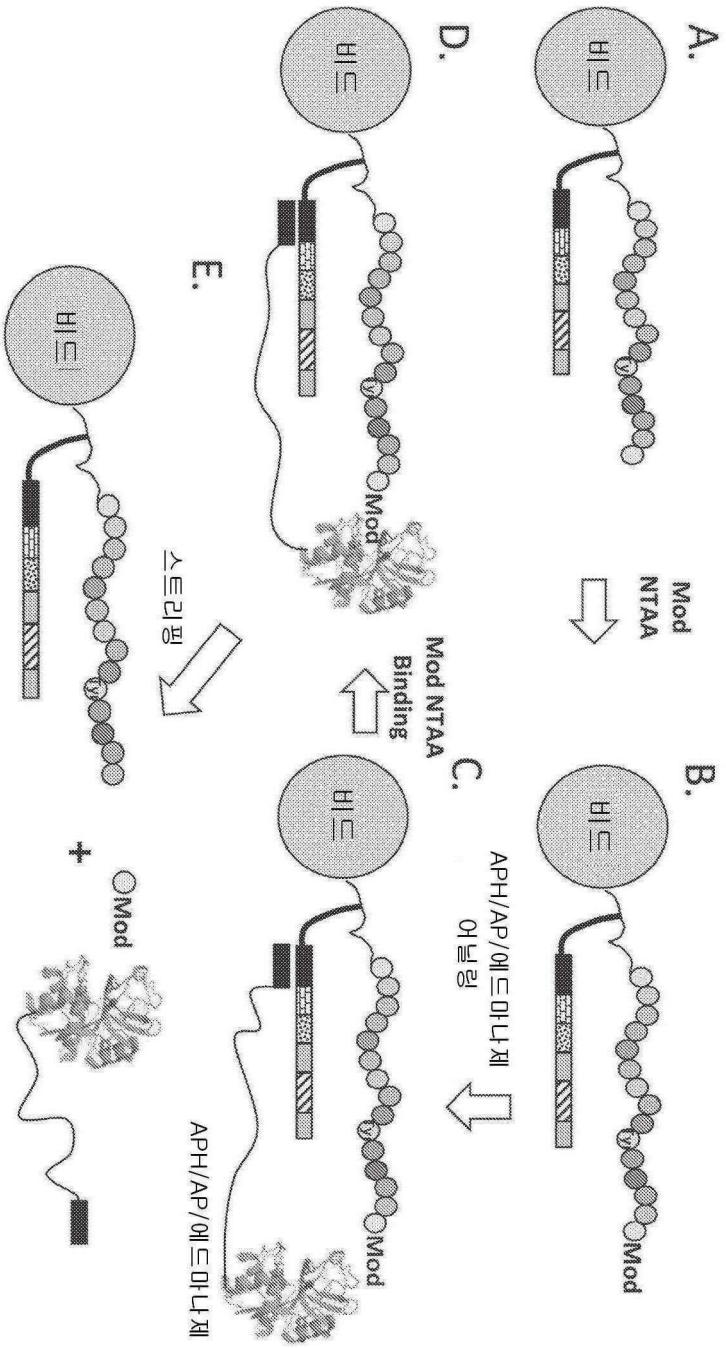
도면42


A.
아세틸화 또는
아미딘화 또는
PITC
NTAA 결합
기록 태그
B.
안티칼린
aaRS
ClpS
폴리머라제
연장
암호화 태그
C.
NTAA 절단
D.
APH 또는
NaOH 또는
Edmse
NTAA 표지
n회 주기

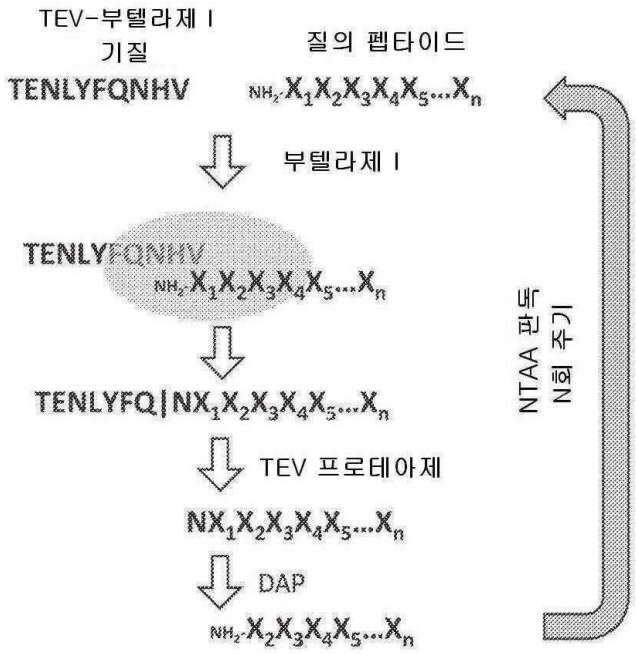
도면43


A.
기질
단백질
안티칼린
암호화 태그
기록 태그
P1 BC Sp
Sp' BC' Sp'
A
PoI 엑스텐다제 활성
B.
결합 영역
안티칼린
SpyCatcher-SpyTag
DNA 태그(암호화)

도면44


A.
비드
Mod
NTAA
B.
비드
Mod
APH/AP/에드마나제
어닐링
C.
비드
Mod
APH/AP/에드마나제
Mod NTAA
Binding
D.
비드
Mod
스트리핑
E.
비드
+
Mod

도면45



서 열 목 록

SEQUENCE LISTING

<110> ProCure Life Sciences Inc.

Chee, Mark

Gunderson, Kevin

Weiner, Michael Phillip

<120> MACROMOLECULE ANAYSIS EMPLOYING NUCLEIC ACID ENCODING

<130> 760229.401WO

<140> PCT

<141> 2017-05-02

<150> US 62/330,841

<151> 2016-05-02

<150> US 62/339,071

<151> 2016-05-19

<150> US 62/376,886

<151> 2016-08-18

<160> 180

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_1

<400> 1

atgtctagca tgccg 15

<210> 2

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_2

<400> 2

ccgtgtcatg tggaa 15

<210> 3

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_3

<400> 3

taagccggta tatca 15

<210> 4

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_4

<400> 4

ttcgatatga cggaa 15

<210> 5

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_5

<400> 5

cgtatacgcg ttagg 15

<210> 6

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_6

<400> 6

aactgccgag attcc 15

<210> 7

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_7

<400> 7

tgatcttagc tgtgc 15

<210> 8

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_8

<400> 8

gagtcggtac cttga 15

<210> 9

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_9

<400> 9

ccgcttgtga tctgg 15

<210> 10

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_10

<400> 10

agatagcgta ccgga 15

<210> 11

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_11

<400> 11

tccaggctca tcatc 15

<210> 12

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_12

<400> 12

gagtactaga gccaa 15

<210> 13

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_13

<400> 13

gagcgtcaat aacgg 15

<210> 14

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_14

<400> 14

gcggtatcta cactg 15

<210> 15

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_15

<400> 15

cttctccgaa gagaa 15

<210> 16

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_16

<400> 16

tgaagcctgt gttaa 15

<210> 17

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_17

<400> 17

ctggatggtt gtcga 15

<210> 18

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_18

<400> 18

actgcacggt tccaa 15

<210> 19

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_19

<400> 19

cgagagatgg tcctt 15

<210> 20

<211>

15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_20

<400> 20

tcttgagaga caaga 15

<210> 21

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_21

<400> 21

aattcgcact gtgtt 15

<210> 22

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_22

<

400> 22

gtagtgccgc taaga 15

<210> 23

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_23

<400> 23

cctatagcac aatcc 15

<210> 24

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_24

<400> 24

atcaccgagg ttgga 15

<210> 25

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_25

<400> 25

gattcaacgg agaag 15

<210> 26

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_26

<400> 26

acgaacctcg cacca 15

<210> 27

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_27

<400> 27

aggacttcaa gaaga 15

<210> 28

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_28

<400> 28

ggttgaatcc tcgca 15

<210> 29

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_29

<400> 29

aaccaacctc tagcg 15

<210> 30

<211> 15

<212>

DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_30

<400> 30

acgcgaatat ctaac 15

<210> 31

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_31

<400> 31

gttgagaatt acacc 15

<210> 32

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_32

<400> 32

ctctctctgt gaacc 15

<210> 33

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_33

<400> 33

gccatcagta agaga 15

<210> 34

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_34

<400> 34

gcaacgtgaa ttgag 15

<210> 35

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_35

<400> 35

ctaagtagag ccaca 15

<210> 36

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_36

<400> 36

tgtctgttgg aagcg 15

<210> 37

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_37

<400> 37

ttaatagaca gcgcg 15

<210> 38

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_38

<400> 38

cgacgctcta acaag 15

<210> 39

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_39

<400> 39

catggcttat tgaga 15

<210> 40

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_40

<400> 40

actaggtatg gccgg 15

<210> 41

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_41

<400> 41

gtcctcgtct atcct 15

<210> 42

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_42

<400> 42

taggattccg ttacc 15

<210> 43

<211>

15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_43

<400> 43

tctgaccacc ggaag 15

<210> 44

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_44

<400> 44

agagtcacct cgtgg 15

<210> 45

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_45

<

400> 45

ctgatgtagt cgaag 15

<210> 46

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_46

<400> 46

gtcggttgcg gatag 15

<210> 47

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_47

<400> 47

tcctcctcct aagaa 15

<210> 48

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_48

<400> 48

attcggtcca cttca 15

<210> 49

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_49

<400> 49

ccttacaggt ctgcg 15

<210> 50

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_50

<400> 50

gatcattggc caatt 15

<210> 51

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_51

<400> 51

ttcaaggctg agttg 15

<210> 52

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_52

<400> 52

tggctcgatt gaatc 15

<210> 53

<211> 15

<212>

DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_53

<400> 53

gtaagccatc cgctc 15

<210> 54

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_54

<400> 54

acacatgcgt agaca 15

<210> 55

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_55

<400> 55

tgctatggat tcaag 15

<210> 56

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_56

<400> 56

ccacgaggct tagtt 15

<210> 57

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_57

<400> 57

ggccaactaa ggtgc 15

<210> 58

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_58

<400> 58

gcacctattc gacaa 15

<210> 59

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_59

<400> 59

tggacacgat cggct 15

<210> 60

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_60

<400> 60

ctataattcc aacgg 15

<210> 61

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_61

<400> 61

aacgtggtta gtaag 15

<210> 62

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_62

<400> 62

caaggaacga gtggc 15

<210> 63

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_63

<400> 63

caccagaacg gaaga 15

<210> 64

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_64

<400> 64

cgtacggtca agcaa 15

<210> 65

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_65

<400> 65
tcggtgacag gctaa 15
<210> 66
<211>
15
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide barcode BC_1 REV

<400> 66
cggcatgcta gacat 15
<210> 67
<211> 15
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide barcode BC_2 REV

<400> 67
ttccacatga cacgg 15
<210> 68
<211> 15
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide barcode BC_3 REV

<400> 68
tgatataccg gctta 15
<210> 69
<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_4 REV

<400> 69

ttccgtcata tcgaa 15

<210> 70

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_5 REV

<400> 70

cctaacgcgt atacg 15

<210> 71

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_6 REV

<400> 71

ggaatctcgg cagtt 15

<210> 72

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_7 REV

<400> 72

gcacagctaa gatca 15

<210> 73

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_8 REV

<400> 73

tcaaggtacc gactc 15

<210> 74

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_9 REV

<400> 74

ccagatcaca agcgg 15

<210> 75

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_10 REV

<400> 75

tccggtacgc tatct 15

<210> 76

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_11 REV

<400> 76

gatgatgagc ctgga 15

<210> 77

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_12 REV

<400> 77

ttggctctag tactc 15

<210> 78

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_13 REV

<400> 78

ccgttattga cgctc 15

<210> 79

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_14 REV

<400> 79

cagtgtagat accgc 15

<210> 80

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_15 REV

<400> 80

ttctcttcgg agaag 15

<210> 81

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_16 REV

<400> 81

ttaacacagg cttca 15

<210> 82

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_17 REV

<400> 82

tcgacaacca tccag 15

<210> 83

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_18 REV

<400> 83

ttggaaccgt gcagt 15

<210> 84

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_19 REV

<400> 84

aaggaccatc tctcg 15

<210> 85

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_20 REV

<400> 85

tcttgtctct caaga 15

<210> 86

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_21 REV

<400> 86

aacacagtgc gaatt 15

<210> 87

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_22 REV

<400> 87

tcttagcggc actac 15

<210> 88

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_23 REV

<400> 88

ggattgtgct atagg 15

<210> 89

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_24 REV

<400> 89

tccaacctcg gtgat 15

<210> 90

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_25 REV

<400> 90

cttctccgtt gaatc 15

<210> 91

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_26 REV

<400> 91

tggtgcgagg ttcgt 15

<210> 92

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_27 REV

<400> 92

tcttcttgaa gtcct 15

<210> 93

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_28 REV

<400> 93

tgcgaggatt caacc 15

<210> 94

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_29 REV

<400> 94

cgctagaggt tggtt 15

<210> 95

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_30 REV

<400> 95

gttagatatt cgcgt 15

<210> 96

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_31 REV

<400> 96

ggtgtaattc tcaac 15

<210> 97

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_32 REV

<400> 97

ggttcacaga gagag 15

<210> 98

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_33 REV

<400> 98

tctcttactg atggc 15

<210> 99

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_34 REV

<400> 99

ctcaattcac gttgc 15

<210> 100

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_35 REV

<400> 100

tgtggctcta cttag 15

<210> 101

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_36 REV

<400> 101

cgcttccaac agaca 15

<210> 102

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_37 REV

<400> 102

cgcgctgtct attaa 15

<210> 103

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_38 REV

<400> 103

cttgttagag cgtcg 15

<210> 104

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_39 REV

<400> 104

tctcaataag ccatg 15

<210> 105

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_40 REV

<400> 105

ccggccatac ctagt 15

<210> 106

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_41 REV

<400> 106

aggatagacg aggac 15

<210> 107

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_42 REV

<400> 107

ggtaacggaa tccta 15

<210> 108

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_43 REV

<400> 108

cttccggtgg tcaga 15

<210> 109

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_44 REV

<400> 109

ccacgaggtg actct 15

<210> 110

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_45 REV

<400> 110

cttcgactac atcag 15

<210> 111

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_46 REV

<400> 111

ctatccgcaa ccgac 15

<210> 112

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_47 REV

<400> 112

ttcttaggag gagga 15

<210> 113

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_48 REV

<400> 113

tgaagtggac cgaat 15

<210> 114

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_49 REV

<400> 114

cgcagacctg taagg 15

<210> 115

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_50 REV

<400> 115

aattggccaa tgatc 15

<210> 116

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_51 REV

<400> 116

caactcagcc ttgaa 15

<210> 117

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_52 REV

<400> 117

gattcaatcg agcca 15

<210> 118

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_53 REV

<400> 118

gagcggatgg cttac 15

<210> 119

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_54 REV

<400> 119

tgtctacgca tgtgt 15

<210> 120

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_55 REV

<400> 120

cttgaatcca tagca 15

<210> 121

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_56 REV

<400> 121

aactaagcct cgtgg 15

<210> 122

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_57 REV

<400> 122

gcaccttagt tggcc 15

<210> 123

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_58 REV

<400> 123

ttgtcgaata ggtgc 15

<210> 124

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_59 REV

<400> 124

agccgatcgt gtcca 15

<210> 125

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_60 REV

<400> 125

ccgttggaat tatag 15

<210> 126

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_61 REV

<400> 126

cttactaacc acgtt 15

<210> 127

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_62 REV

<400> 127

gccactcgtt ccttg 15

<210> 128

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_63 REV

<400> 128

tcttccgttc tggtg 15

<210> 129

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_64 REV

<400> 129

ttgcttgacc gtacg 15

<210> 130

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide barcode BC_65 REV

<400> 130

ttagcctgtc accga 15

<210> 131

<211> 16

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<220><221> MOD_RES

<222> 1

<223> formyl-Methionine

<400> 131

Met Asp Val Glu Ala Trp Leu Gly Ala Arg Val Pro Leu Val Glu Thr

1 5 10 15

<210> 132

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 132

Thr Glu Asn Leu Tyr Phe Gln Asn His Val

1 5 10

<210> 133

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 133

aatgatacgg cgaccaccga 20

<210> 134

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 134

caagcagaag acggcatacg agat 24

<210> 135

<211> 5

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide

<220><221> misc_feature

<222> (1)...(5)

<223> n = A,T,C or G

<400> 135

nnnnn 5

<210> 136

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 136

Asp Tyr Lys Asp Asp Asp Asp Lys

1 5

<210> 137

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 137

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr

1 5 10

<210> 138

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 138

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu

1 5 10

<210> 139

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 139

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1 5

<210> 140

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 140

Asn Trp Ser His Pro Gln Phe Glu Lys

1 5

<210> 141

<211> 82

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 1

<223> biotin

<400> 141

tttttgcaaa tggcattctg acatcccgta gtccgcgaca ctagatgtct agcatgccgc 60

cgtgtcatgt ggaaactgag tg 82

<210> 142

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 1

<223> biotin

<400> 142

tttttttttt gactggttcc aattgacaag ccgtagtccg cgacactagt aagccggtat 60

atcaactgag tg 72

<210> 143

<211> 10

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 1

<223> biotin

<220><221> modified_base

<222> 10

<223> three carbon (3C) spacer

<400> 143

tttttttttt 10

<210> 144

<211> 62

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> (31)...(32)

<223> 18-atom hexa-ethyleneglycol spacer

<220><221> modified_base

<222> 62

<223> three carbon (3C) spacer

<400> 144

ggatgtcaga atgccatttg ctttttttt tcactcagtc ctaacgcgta tacgcactca 60

gt 62

<210> 145

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> (31)...(32)

<223> 18-atom hexa-ethyleneglycol spacer

<220><221> modified_base

<222> 63

<223> three carbon (3C) spacer

<400> 145

ggatgtcaga atgccatttg ctttttttt tcactcagtc ctaacgcgta tacgtcactc 60

agt 63

<210> 146

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> (31)...(32)

<223> five 18-atom hexa-ethyleneglycol spacers

<220><221> modified_base

<222> 63

<223> three carbon (3C) spacer

<400> 146

ggatgtcaga atgccatttg ctttttttt tcactcagtc ctaacgcgta tacgtcactc 60

agt 63

<210> 147

<211> 71

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> (25)...(26)

<223> 18-atom hexa-ethyleneglycol spacer

<220><221> modified_base

<222> 63

<223> three carbon (3C) spacer)

<400> 147

gcttgtcaat tggaaccagt cttttcactc agtcctaacg cgtatacggg aatctcggca 60

gttcactcag t 71

<210> 148

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 44

<223> three carbon (3C) spacer

<400> 148

cgatttgcaa ggatcactcg tcactcagtc ctaacgcgta tacg 44

<210> 149

<211> 8

<212> DNA

<213> Artificial Sequence

<220>

<223> spacer sequence

<400> 149

actgagtg 8

<210> 150

<211> 8

<212> DNA

<213> Artificial Sequence

<220>

<223> spacer sequence

<400> 150

cactcagt 8

<210> 151

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 151

cgtagtccgc gacactag 18

<210> 152

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 152

cgatttgcaa ggatcactcg 20

<210> 153

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 153

gcaaatggca ttctgacatc c 21

<210> 154

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 154

gactggttcc aattgacaag c 21

<210> 155

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 23

<223> three carbon (3C) spacer

<400> 155

cgtatacgcg ttaggactga gtg 23

<210> 156

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<220><221> modified_base

<222> 38

<223> three carbon (3C) spacer

<400> 156

aactgccgag attcccgtat acgcgttagg actgagtg 38

<210> 157

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 157

agtccgcgca atcagatgtc tagcatgccg gatccggatc gatctc 46

<210> 158

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 158

agtccgcgca atcagccgtg tcatgtggaa gatccggatc gatctc 46

<210> 159

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<

400> 159

agtccgcgca atcagtaagc cggtatatca gatccggatc gatctc 46

<210> 160

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 160

agtccgcgca atcagttcga tatgacggaa gatccggatc gatctc 46

<210> 161

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 161

tgcaaggatc actcgccaga tcacaagcgg gagatcgatc cggatc 46

<210> 162

<211> 46

<212

> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 162

tgcaaggatc actcgtccgg tacgctatct gagatcgatc cggatc 46

<210> 163

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 163

tgcaaggatc actcggatga tgagcctgga gagatcgatc cggatc 46

<210> 164

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 164

tgcaaggatc actcgttggc tctagtactc gagatcgatc cggatc 46

<210> 165

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 165

aatcgtagtc cgcgcaatca g 21

<210> 166

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer

<400> 166

acgatttgca aggatcactc g 21

<210> 167

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> spacer sequence

<400> 167

gatccggatc gatctc 16

<210> 168

<211> 734

<212> DNA

<213> Artificial Sequence

<220>

<223> extended recording tag construct

<400> 168

aatcacggta caagtcactc atccgtacgc tatctgagaa tcgtccagat ccggcatgct 60

agtatctggt gcagactacg attgttacag atcactcaga tgatgagcac agaaaatcgt 120

cgaatcttcc atcaccatcg aacagttacg attaatgtag tccgcacaat cgaatgtcta 180

acatgccgaa tcccggacgt ctccagcttc taaaccaaca gtagtcgcac aaatcattgt 240

acggtacaag atctaacgag agatgatcgg atctgaccac tttaaacact gattacgcag 300

actacgatta cgatttaaga atcctcgtcc ggtacaatca tagtccgcac aatcaaccgt 360

gtcatgtgaa gatcagatcg atctcgaata gcgtaccaga cagtgatctt gcaaatcgta 420

atgtgtccgc gccaatcgat agccatgaat cccagtcgat ctcccgcttg tgatctggcg 480

atcgccttgt accgtcgtac gatttgagat cacctcgtta actcaagcta aagatcgtcc 540

ggatcgcttt ataaacatct gattgcgcgg tacgattatc gtagtccgca catatcgaac 600

ctgttgaaga tccggatcgt ctctccaggc tcatcatccg agtgatcctt gcaaataatc 660

atgtccgcac catcaggtgt ctaacgcttg ccggatccga atcgatctct ccaggctcat 720

catcgaagtg atgt 734

<210> 169

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<400> 169

Cys Pro Val Gln Leu Trp Val Asp Ser Thr

1 5 10

<210> 170

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> synthetic peptide

<220><221> VARIANT

<222> (1)...(10)

<223> Xaa = Any Amino Acid

<400> 170

Cys Pro Xaa Gln Xaa Trp Xaa Asp Xaa Thr

1 5 10

<210> 171

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> FLAG epitope peptide

<400> 171

Asp Tyr Lys Asp Asp Asp Asp Lys

1 5

<210> 172

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> V5 epitope peptide

<400> 172

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr

1 5 10

<210> 173

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> c-Myc epitope peptide

<400> 173

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu

1 5 10

<210> 174

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> HA epitope peptide

<400> 174

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1 5

<210> 175

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> V5 epitope peptide

<400> 175

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr

1 5 10

<210> 176

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> StrepTag II peptide

<400> 176

Asn Trp Ser His Pro Gln Phe Glu Lys

1 5

<210> 177

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic nucloetide

<220><221> misc_feature

<222> (11)...(22)

<223> compartment bar code n = A, C, T, or G

<220><221> misc_feature

<222> (23)...(27)

<223> unique molecular identifier n = A, T, C or G

<400> 177

gcgcaatcag nnnnnnnnnn nnnnnnntgc aaggat 36

<210> 178

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic nucloetide

<220><221> misc_feature

<222> (1)...(12)

<223> Compartment barcod n = A, T, C or G

<400> 178

nnnnnnnnnn nn 12

<210> 179

<211> 5

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic nucloetide

<220><221> misc_feature

<222> (1)...(5)

<223> Unique molecular identifier; n = A, T, C or G

<400> 179

nnnnn 5

<210> 180

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> butelase I peptide substrate

<400> 180

Cys Gly Gly Ser Ser Gly Ser Asn His Val

1 5 10