



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2023년11월27일

(11) 등록번호 10-2607157

(24) 등록일자 2023년11월23일

- (51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) **C07K 16/00** (2006.01)
C40B 40/08 (2006.01) **C40B 50/06** (2017.01)
- (52) CPC특허분류
C12N 15/1086 (2013.01)
C07K 16/00 (2013.01)
- (21) 출원번호 10-2019-7029959
- (22) 출원일자(국제) 2018년03월14일
 심사청구일자 2021년03월09일
- (85) 번역문제출일자 2019년10월11일
- (65) 공개번호 10-2019-0129081
- (43) 공개일자 2019년11월19일
- (86) 국제출원번호 PCT/US2018/022487
- (87) 국제공개번호 WO 2018/170164
 국제공개일자 2018년09월20일
- (30) 우선권주장
 62/471,723 2017년03월15일 미국(US)
 62/578,326 2017년10월27일 미국(US)
- (56) 선행기술조사문헌
 US20080287320 A1*
 W02014035693 A2
 US20110082055 A1
 US20150038373 A1
 *는 심사관에 의하여 인용된 문헌

- (73) 특허권자
트위스트 바이오사이언스 코포레이션
 미국 94080 캘리포니아주 사우스 샌 프란시스코
 게이트웨이 불러바드 681
- (72) 발명자
콕스 안토니
 미국 캘리포니아 94158 샌프란시스코 스위트 545
 사우스 미션 베이 불러바드 455
천 쓰위엔
 미국 캘리포니아 94158 샌프란시스코 스위트 545
 사우스 미션 베이 불러바드 455
 (뒷면에 계속)
- (74) 대리인
김진희, 김태홍

전체 청구항 수 : 총 21 항

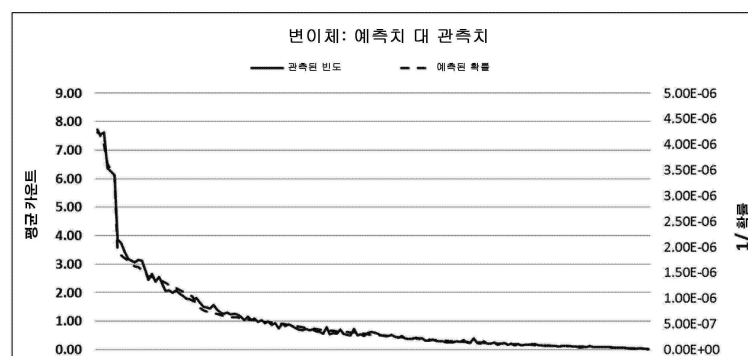
심사관 : 이형곤

(54) 발명의 명칭 드노보 합성된 조합 핵산 라이브러리

(57) 요약

본원은 핵산 서열의 소정의 변이체를 코딩하는 고도로 정확한 핵산 라이브러리를 생성하는 방법을 개시한다. 변이 정도는 포화된 변이체 라이브러리를 야기할 정도로 완전할 수 있거나, 변이체의 불포화 라이브러리를 야기할 정도로 보다 덜 완전할 수 있다. 본원에 기재된 변이체 핵산 라이브러리는 전사 또는 번역에 의해 추가 프로세싱 (뒷면에 계속)

대표도



되도록 디자인될 수 있다. 본원에 기재된 변이체 핵산 라이브러리는 변이체 RNA, DNA 및/또는 단백질 집단을 생성하도록 디자인될 수 있다. 본원은 생물학적 기능의 조절 및 질환의 치료 또는 감소를 위한 치료제의 디자인에 적용되는, 증가되거나 감소된 활성을 가진 변이체 종을 확인하는 방법도 제공한다.

(52) CPC특허분류

C12N 15/1093 (2013.01)

C40B 40/08 (2013.01)

C40B 50/06 (2013.01)

C07K 2317/565 (2013.01)

(72) 발명자

레도가르 찰스

미국 캘리포니아 94158 샌프란시스코 스위트 545

사우스 미션 베이 불러바드 455

토파니 도미니크

미국 캘리포니아 94158 샌프란시스코 스위트 545

사우스 미션 베이 불러바드 455

명세서

청구범위

청구항 1

변이체 핵산 라이브러리의 합성 방법으로서,

- (a) 미리 선택된 불균일 코돈 분포를 가진 복수의 폴리뉴클레오타이드를 코딩하는 소정의 서열을 제공하는 단계;
- (b) 각각의 위치에서 불균일 코돈 분포를 가진 복수의 아미노산을 포함하는 단백질 라이브러리를 코딩하는 복수의 폴리뉴클레오타이드를 합성하는 단계로서, 변이체 핵산 라이브러리에서 적어도 70%의 예측된 다양성이 표시되는 것인 단계;
- (c) 상기 복수의 폴리뉴클레오타이드를 기반으로 번역된 단백질에 대한 세포 활성을 어레이하는 단계; 및
- (d) 음성 또는 무효(null) 결과와 관련된 소정의 서열의 결과를 수집하는 것을 포함하는, 단계 (c)의 어레이로부터 결과를 수집하는 단계

를 포함하는, 변이체 핵산 라이브러리의 합성 방법.

청구항 2

제1항에 있어서, 단계 (d)가 소정의 서열의 적어도 80%에 대한 결과를 수집하는 것을 포함하는 것인 방법.

청구항 3

삭제

청구항 4

제1항에 있어서, 복수의 폴리뉴클레오타이드 서열 중 적어도 80%가 라이브러리 내의 각각의 폴리뉴클레오타이드 서열에 대한 평균 빈도의 2배 이내의 양으로 변이체 핵산 라이브러리에 각각 존재하는 것인 방법.

청구항 5

제1항에 있어서, 세포 활성이 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성, 세포 신호전달, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하는 것인 방법.

청구항 6

제1항에 있어서, 변이체 핵산 라이브러리가 변이체 유전자 또는 이의 단편의 서열을 코딩하는 것인 방법.

청구항 7

제1항에 있어서, 변이체 핵산 라이브러리가 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 것인 방법.

청구항 8

(a) (i) 제1 복수의 폴리뉴클레오타이드, 및 (ii) 제2 복수의 폴리뉴클레오타이드를 코딩하는 소정의 서열을 디자인하는 단계로서, 제1 복수의 폴리뉴클레오타이드의 각각의 폴리뉴클레오타이드가 기준 서열과 비교하여 변이체 서열을 코딩하고, 제2 복수의 폴리뉴클레오타이드의 각각의 폴리뉴클레오타이드가 기준 서열과 비교하여 변이체 서열을 코딩하는 것인 단계;

(b) 제1 복수의 폴리뉴클레오타이드 및 제2 복수의 폴리뉴클레오타이드를 합성하는 단계; 및

(c) 제1 복수의 폴리뉴클레오타이드와 제2 복수의 폴리뉴클레오타이드를 혼합하여 핵산의 조합 라이브러리를 형성하는 단계로서, 핵산의 조합 라이브러리에서 적어도 70%의 예측된 다양성이 표시되고, 상기 핵산의 조합 라이브러리는 각각의 위치에서 아미노산의 불균일 분포를 가진 복수의 아미노산을 포함하는 단백질 라이브러리를 코

당하는 것인 단계

를 포함하는, 핵산의 조합 라이브러리를 생성하는 방법.

청구항 9

제8항에 있어서, 조합 라이브러리가 불포화 조합 라이브러리이고, 불포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수가 포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수보다 적어도 25% 더 적은 것인 방법.

청구항 10

제8항에 있어서, PCR 돌연변이유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는 방법.

청구항 11

제8항에 있어서, 조합 라이브러리가 변이체 유전자 또는 이의 단편의 서열을 코딩하는 것인 방법.

청구항 12

제8항에 있어서, 조합 라이브러리가 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 것인 방법.

청구항 13

제12항에 있어서, 조합 라이브러리가 항체의 가변 영역 또는 불변 영역의 적어도 일부를 코딩하는 것인 방법.

청구항 14

제12항에 있어서, 조합 라이브러리가 항체의 적어도 하나의 CDR 영역을 코딩하는 것인 방법.

청구항 15

제12항에 있어서, 조합 라이브러리가 항체의 중쇄의 CDR1, CDR2 및 CDR3, 및 경쇄의 CDR1, CDR2 및 CDR3을 코딩하는 것인 방법.

청구항 16

(a) 단일 기준 서열과 비교하여 변이체 서열을 포함하는 복수의 폴리뉴클레오타이드에 대한 소정의 서열을 제공하는 단계로서, 폴리뉴클레오타이드는 복수의 코돈을 포함하고, 소정의 서열은 미리 선택된 제1 분포 값을 갖고, 제1 분포 값은 불균일 분포 값인 단계;

(b) 단일 기준 서열로부터 핵산 서열의 세트를 무작위로 생성하도록 기계 지시를 제공하는 단계;

(c) 무작위로 생성된 핵산 서열의 세트로부터 제2 분포 값을 계산하는 단계;

(d) 제2 분포 값이 제1 분포 값과 일치하는지 결정하는 단계로서, 제1 분포 값과 제2 분포 값은 각각 독립적으로 서열의 미리 선택된 위치에 코돈이 존재하는 확률인 단계; 및

(e) 제1 분포 값과 제2 분포 값이 일치하는 경우, 무작위로 생성된 핵산 서열의 세트에 의해 인코딩된 복수의 폴리뉴클레오타이드를 포함하는 변이체 핵산 라이브러리를 합성하는 단계로서, 적어도 70%의 복수의 폴리뉴클레오타이드가 복수의 폴리뉴클레오타이드에 대한 평균 양의 2배 이내의 양으로 합성되는 것인 단계

를 포함하는, 변이체 핵산 라이브러리의 합성 방법.

청구항 17

제16항에 있어서, PCR 돌연변이유발 반응을 위한 프라이머로서 변이체 핵산 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는 방법.

청구항 18

제16항에 있어서, 변이체 서열을 가진 복수의 코돈의 각각의 코돈을 확인하기 위해 코돈 배정을 이용하는 것인

방법.

청구항 19

제18항에 있어서, 코돈 배정이 유기체에서의 코돈 서열의 빈도에 기반을 둔 것인 방법.

청구항 20

제19항에 있어서, 유기체가 동물, 식물, 진균, 원생생물, 고세균 및 세균 중 적어도 하나인 방법.

청구항 21

제18항에 있어서, 코돈 배정이 코돈 서열의 다양성에 기반을 둔 것인 방법.

청구항 22

제16항에 있어서, 변이체 핵산 라이브러리가 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 것인 방법.

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

발명의 설명

기술 분야

[0001] 교차참조

[0002] 본원은 전체로서 본원에 각각 참고로 도입된, 2017년 10월 27일에 출원된 미국 가출원 제62/578,326호; 및 2017년 3월 15에 출원된 미국 가출원 제62/471,723호의 이익을 주장한다.

[0003] 서열목록

[0004] 본원은 ASCII 포맷으로 전자적으로 제출되었고 전체로서 본원에 참고로 도입된 서열목록을 함유한다. 2018년 3월 13일에 생성된 상기 ASCII 사본은 명칭이 44854-729_601_SL.txt이고 크기가 18,419 바이트이다.

배경 기술

[0005] 합성 생물학의 초석은 이 맞춤 경로 및 유기체의 신속하고 적절한 생성 및 최적화를 위해 접근가능해지기 위해 DNA를 요구하는 반복적 과정인 디자인, 구축 및 시험 과정이다. 디자인 단계에서, DNA를 구성하는 A, C, T 및 G 뉴클레오타이드는 관심 있는 좌위 또는 경로를 구성할 다양한 유전자 서열들로 제체화되고, 이때 각각의 서열 변이체는 시험될 특정 가설을 대표한다. 이 변이체 유전자 서열은 진화 생물학으로부터 유래하고 유전자, 게놈, 전사체(transcriptome) 및 단백체(proteome)를 구성하는 전체 서열들과 관련된 개념인 서열 공간의 서브세트를 대표한다.

[0006] 많은 상이한 변이체들은 전형적으로 서열 공간의 적절한 샘플링을 가능하게 하고 최적화된 디자인의 확률을 최대화하기 위해 각각의 디자인-구축-시험 주기용으로 디자인된다. 개념적으로 단순할지라도, 통상적인 합성 방법의 속도, 처리율 및 품질과 관련된 과정 장애물은 이 주기가 진행되는 속도를 약화시켜 개발 시간을 연장한다. 현재 합성 기술의 매우 정확한 DNA의 고비용 및 제한된 처리율로 인해 서열 공간을 충분히 조사할 수 없다는 점은 속도 제한 단계로 남아있다.

[0007] 구축 단계로 시작하는 두 과정들이 주목할 만하다: 핵산 합성 및 유전자 합성. 역사적으로, 상이한 유전자 변이체들의 합성은 분자 클로닝을 통해 달성되었다. 이 방법은 확실히 확장될 수 없다. 초기 화학적 유전자 합성 노력은 중첩 서열 상동성을 가진 다수의 폴리뉴클레오타이드들의 생성에 초점을 맞추었다. 그 다음, 이들을 폴링하였고 다수의 라운드의 중합효소 연쇄 반응(PCR)을 수행하여, 중첩 폴리뉴클레오타이드들을 연쇄적으로

연결하여 전체 길이 이중 가닥 유전자를 만들 수 있었다. 시간 소모적 및 노동 집약적 구축, 높은 부피의 포스포라미다이트의 요구, 비싼 원재료, 및 다운스트림 단계를 위해 요구된 양보다 상당히 더 적은 나노몰 양의 최종 생성물의 생성을 비롯한 다수의 요인들이 이 방법을 방해하고, 다수의 별도의 폴리뉴클레오타이드들이 한 유전자의 합성을 유발하기 위해 1개의 96웰 플레이트를 요구하였다.

[0008]

마이크로어레이에서의 폴리뉴클레오타이드의 합성은 유전자 합성의 처리율을 상당히 증가시켰다. 다수의 폴리뉴클레오타이드들이 마이크로어레이 표면에서 합성된 후, 절단되고 함께 폴딩될 수 있었다. 특정 유전자가 될 각각의 폴리뉴클레오타이드는 폴리뉴클레오타이드의 그 특정 하위집단이 해체되고 관심 있는 유전자로 조립될 수 있게 하는 고유 바코드 서열을 함유하다. 상기 과정의 이 단계에서, 각각의 서브풀은 96웰 플레이트 내의 한 웰 내로 전달되어, 처리율을 96개의 유전자까지 증가시킨다. 이것은 처리율 면에서 고전적 방법보다 두 자릿수 더 높지만, 비용 효율의 결여 및 느린 처리소요 시간으로 인해 한 번에 수천 개의 서열들을 요구하는 디자인, 구축 및 시험 주기를 여전히 적절히 뒷받침하지 못한다.

발명의 내용

[0009]

본원은 (a) 미리 선택된 코돈 분포를 가진 적어도 500개의 폴리뉴클레오타이드 서열들을 코딩하는 소정의 서열을 제공하는 단계; (b) 상기 적어도 500개의 폴리뉴클레오타이드 서열들을 코딩하는 복수의 폴리뉴클레오타이드들을 합성하는 단계; (c) 상기 복수의 폴리뉴클레오타이드들에 의해 코딩된 핵산, 또는 상기 복수의 폴리뉴클레오타이드들을 기반으로 번역된 단백질의 활성을 어레이하는 단계; 및 (d) 음성 또는 무효(null) 결과와 관련된 소정의 서열의 결과를 수집하는 단계를 포함하는, 단계 (c)에서의 어레이로부터 결과를 수집하는 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 단계 (d)가 적어도 80%의 소정의 서열에 대한 결과를 수집하는 단계를 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 단계 (d)가 적어도 90%의 소정의 서열에 대한 결과를 수집하는 단계를 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 단계 (d)가 적어도 100%의 소정의 서열에 대한 결과를 수집하는 단계를 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 70%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 500개의 폴리뉴클레오타이드 서열들 중 적어도 80%의 폴리뉴클레오타이드 서열들이 정확한 크기를 가진 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 500개의 폴리뉴클레오타이드 서열들 중 적어도 약 80%의 폴리뉴클레오타이드 서열들이 라이브러리 내의 폴리뉴클레오타이드 서열들 각각에 대한 평균 빈도의 2배 이내의 양으로 변이체 핵산 라이브러리에 각각 존재하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 향상된 또는 감소된 활성과 관련된 소정의 서열에 대한 결과를 단계 (c)의 어레이로부터 수집하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 활성이 세포 활성인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 세포 활성이 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성, 세포 신호전달, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 유전자 또는 이의 단편의 서열을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 가이드 RNA(gRNA)를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 siRNA, shRNA, RNAi 또는 miRNA를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다.

[0010]

본원은 (a) (i) 제1 복수의 폴리뉴클레오타이드들 및 (ii) 제2 복수의 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 디자인하는 단계로서, 상기 제1 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하고, 상기 제2 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하는 것인 단계; (b) 제1 복수의 폴리뉴클레오타이드들 및 제2 복수의 폴리뉴클레오타이드들을 합성하는 단계; 및 (c) 제1 복수의 폴리뉴클레오타이드들과 제2 복수의 폴리뉴클레오타이드들을 혼합하여 핵산의 조합 라이브러리를 형성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 불포화 조합 라이브러리인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 포화 조합 라이브러리인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 적어도 10,000개의 폴리뉴클레오타이드들을 합성하는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 불포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수가 포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수보다 적어도 25% 더 적은

것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 제1 기준 서열 또는 제2 기준 서열을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 번역될 때 단백질 라이브러리를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 조합 라이브러리의 핵산을 벡터 내로 삽입하는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 변이체 유전자 또는 이의 단편의 서열을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 항체의 가변 영역 또는 불변 영역의 적어도 일부를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 항체의 적어도 하나의 CDR 영역을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 항체의 중쇄의 CDR1, CDR2 및 CDR3, 및 경쇄의 CDR1, CDR2 및 CDR3을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 가이드 RNA(gRNA)를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법도 제공한다.

[0011]

본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 복수의 코돈들을 코딩하는 복수의 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계; (b) 소정의 핵산 기준 서열 내의 미리 선택된 위치에서 코돈에 대한 분포 값을 선택하는 단계; (c) 선택된 분포 값과 일치하는 분포 값을 가진 핵산 서열의 세트를 무작위로 생성하도록 기계 지시를 제공하는 단계로서, 상기 핵산 서열의 세트가 포화 코돈 변이체 라이브러리를 생성하기 위해 요구되는 핵산 서열의 양보다 더 적은 것인 단계; 및 (d) 미리 선택된 분포를 가진 변이체 핵산 라이브러리를 합성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 번역될 때 단백질 라이브러리를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 핵산 라이브러리의 핵산을 벡터 내로 삽입하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 변이체 핵산 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정을 이용하여 변이체 서열을 가진 복수의 코돈들의 각각의 코돈을 결정하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정이 유기체에서의 코돈 서열의 빈도에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 유기체가 동물, 식물, 진균, 원생생물, 고세균 및 세균 중 적어도 하나인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정이 코돈 서열의 다양성에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다.

[0012]

본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 코돈을 코딩하는 복수의 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계; (b) 복수의 폴리뉴클레오타이드들을 폴리뉴클레오타이드의 5' 단편 및 폴리뉴클레오타이드의 3' 단편으로 나누는 단계; (c) 소정의 핵산 기준 서열 내의 미리 선택된 위치에서 코돈에 대한 분포 값을 선택하는 단계; (d) 선택된 분포 값과 일치하는 분포 값을 가진 핵산의 세트를 무작위로 생성하도록 기계 지시를 제공하는 단계로서, 상기 핵산의 세트가 포화 핵산 라이브러리를 생성하기 위해 요구되는 핵산의 양보다 더 적은 것인 단계; (e) 폴리뉴클레오타이드의 5' 단편 및 폴리뉴클레오타이드의 3' 단편을 합성하는 단계; 및 (f) 폴리뉴클레오타이드의 5' 단편과 폴리뉴클레오타이드의 3' 단편을 혼합하여 변이체 핵산 라이브러리를 형성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 10,000개의 폴리뉴클레오타이드들을 합성하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 복수의 폴리뉴클레오타이드들을 하나 초과 5' 단편 및 하나 초과 3' 단편 중 적어도 하나로 나누는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 번역될 때 단백질 라이브러리를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 핵산 라이브러리의 핵산을 벡터 내로 삽입하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 변이체 핵산 라이브러리를 사용하여 핵산의

PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 향상된 또는 감소된 활성을 가진 변이체 서열을 확인하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 활성이 세포 활성인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 세포 활성이 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성, 세포 신호전달, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 유전자 또는 이의 단편의 서열을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 항체의 가변 영역 또는 불변 영역의 적어도 일부를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 항체의 적어도 하나의 CDR 영역을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 항체의 중쇄의 CDR1, CDR2 및 CDR3, 및 경쇄의 CDR1, CDR2 및 CDR3을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 핵산 라이브러리에서 합성된 상이한 서열들의 수가 50 내지 1,000,000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 핵산 라이브러리에서 합성된 상이한 서열들의 수가 500 내지 25000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 변이체 핵산 라이브러리에서 합성된 상이한 서열들의 수가 1000 내지 15000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 변이체 핵산 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정을 이용하여 변이체 서열을 가진 코돈을 결정하는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정이 유기체에서의 코돈 서열의 빈도에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 유기체가 동물, 식물, 진균, 원생생물, 고세균 및 세균 중 적어도 하나인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 코돈 배정이 코돈 서열의 다양성에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 가이드 RNA(gRNA)를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법도 제공한다.

[0013] 본원은 (a) (i) 제1 복수의 폴리뉴클레오타이드들 및 (ii) 제2 복수의 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계로서, 상기 제1 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하고 상기 제2 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하는 것인 단계; (b) 표면을 가진 구조물을 제공하는 단계; (c) 제1 복수의 폴리뉴클레오타이드들을 합성하는 단계로서, 상기 제1 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 표면으로부터 연장되는 것인 단계; (d) 제2 복수의 폴리뉴클레오타이드들을 합성하는 단계로서, 상기 제2 복수의 폴리뉴클레오타이드들의 각각의 폴리뉴클레오타이드가 표면으로부터 연장되는 것인 단계; (e) 상기 표면으로부터 제1 복수의 폴리뉴클레오타이드들 및 제2 복수의 폴리뉴클레오타이드들을 방출하는 단계; 및 (f) 제1 복수의 폴리뉴클레오타이드들과 제2 복수의 폴리뉴클레오타이드들을 혼합하여 핵산의 조합 라이브러리를 형성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 핵산의 조합 라이브러리를 생성하는 방법도 제공한다.

[0014] 본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 복수의 코돈들을 코딩하는 복수의 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 디자인하는 단계; (b) 복수의 폴리뉴클레오타이드들을 합성하여 변이체 핵산 라이브러리를 생성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계; (c) 변이체 핵산 라이브러리를 발현시키는 단계; 및 (d) 변이체 핵산 라이브러리와 관련된 활성을 평가하는 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 약 90%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다. 본원은 적어도 약 95%의 예측된 다양성이 표시되는 것인 변이체 핵산 라이브러리를 합성하는 방법도 제공한다.

[0015] 본원은 (a) (i) 제1 복수의 동일하지 않은 폴리뉴클레오타이드들 및 (ii) 제2 복수의 동일하지 않은 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계로서, 제1 복수의 동일하지 않은 폴리뉴클레오타이드들의 각각의 동일하지 않은 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하고, 제2 복수의 동일하지 않은 폴리뉴클레오타이드들의 각각의 동일하지 않은 폴리뉴클레오타이드가 단일 기준 서열과 비교하여 변이체 서열을 코딩하는 것인 단계; (b) 표면을 가진 구조물을 제공하는 단계; (c) 제1 복수의 동일하지 않은 폴리뉴클레오타이드들을 합성하는 단계로서, 상기 제1 복수의 동일하지 않은 폴리뉴클레오타이드들의 각각의 동일하지 않은 폴리뉴클레오타이드가 표면으로부터 연장되는 것인 단계; (d) 제2 복수의 동일하지 않은 폴리뉴클

레오타이드들을 합성하는 단계로서, 상기 제2 복수의 동일하지 않은 폴리뉴클레오타이드들의 각각의 동일하지 않은 폴리뉴클레오타이드가 표면으로부터 연장되는 것인 단계; (e) 표면으로부터 제1 복수의 동일하지 않은 폴리뉴클레오타이드들 및 제2 복수의 동일하지 않은 폴리뉴클레오타이드들을 방출하는 단계; 및 (f) 제1 복수의 폴리뉴클레오타이드들과 제2 복수의 폴리뉴클레오타이드들을 혼합하여 핵산의 조합 라이브러리를 형성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 불포화 조합 라이브러리인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 포화 조합 라이브러리인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 적어도 10,000개의 폴리뉴클레오타이드들을 합성하는 것인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 불포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수가 포화 조합 라이브러리의 생성을 위한 폴리뉴클레오타이드의 총 수보다 적어도 25% 더 적은 것인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 변이체 조합 라이브러리가 제1 기준 서열 또는 제2 기준 서열을 코딩하는 것인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 번역될 때 단백질 라이브러리를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 조합 라이브러리의 핵산을 벡터 내로 삽입하는 것인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 변이체 유전자 또는 이의 단편의 서열을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 항체의 가변 영역 또는 불변 영역의 적어도 일부를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 항체의 적어도 하나의 CDR 영역을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 항체의 중쇄의 CDR1, CDR1 및 CDR3, 및 경쇄의 CDR1, CDR2 및 CDR3을 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 가이드 RNA(gRNA)를 코딩하는 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 소정의 서열에 비해 1/1000개 염기 미만의 총 오류율을 가진 핵산의 조합 라이브러리를 생성하는 방법을 제공한다. 본원은 구조물이 고체 지지체, 겔 또는 비드이고 상기 고체 지지체가 플레이트 또는 컬럼인 핵산의 조합 라이브러리를 생성하는 방법을 제공한다.

[0016] 본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 복수의 코돈들을 코딩하는 복수의 동일하지 않은 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계; (b) 소정의 핵산 기준 서열 내의 미리 선택된 위치에서 코돈에 대한 분포 값을 선택하는 단계; (c) 포화 코돈 변이체 라이브러리를 생성하기 위해 요구되는 핵산의 양보다 더 적은 핵산의 세트를 무작위로 생성하도록 기계 지시를 제공하는 단계; 및 (d) 미리 선택된 분포를 가진 핵산 라이브러리를 합성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 조합 라이브러리가 번역될 때 단백질 라이브러리를 코딩하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 조합 라이브러리의 핵산을 벡터 내로 삽입하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정을 이용하여 변이체 서열을 가진 복수의 코돈들의 각각의 코돈을 결정하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정이 유기체에서의 코돈 서열의 빈도에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 유기체가 동물, 식물, 진균, 원생생물, 고세균 및 세균 중 적어도 하나인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정이 코돈 서열의 다양성에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다.

[0017] 본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 코돈을 코딩하는 복수의 동일하지 않은 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 제공하는 단계; (b) 복수의 동일하지 않은 폴리뉴클레오타이드들을 동일하지 않은 폴리뉴클레오타이드의 5' 단편과 동일하지 않은 폴리뉴클레오타이드의 3' 단편으로 나누는 단계; (c) 소정의 핵산 기준 서열 내의 미리 선택된 위치에서 코돈에 대한 분포 값을 선택하는 단계; (d) 포화 핵산 라이브러리를 생성하기 위해 요구되는 핵산의 양보다 더 적은 핵산의 세트를 무작위로 생성하도록 기계 지시를 제공하는 단계; (e) 동일하지 않은 폴리뉴클레오타이드의 5' 단편 및 동일하지 않은 폴리뉴클레오타이드의 3' 단편을 합성하는 단계; 및 (f) 동일하지 않은 폴리뉴클레오타이드의 5' 단편과 동일하지 않은 폴리뉴클레오타이드의 3' 단편을 혼합하여 변이체 핵산 라이브러리를 형성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 10,000개의 동일하

지 않은 폴리뉴클레오타이드들을 합성하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 적어도 80%의 변이체가 정확한 크기를 가진 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 복수의 동일하지 않은 폴리뉴클레오타이드들을 하나 초과 5' 단편 및 하나 초과 3' 단편 중 적어도 하나로 나누는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 조합 라이브러리가 번역될 때 단백질 라이브러리를 코딩하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 조합 라이브러리의 핵산을 벡터 내로 삽입하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 PCR 돌연변이 유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 향상된 또는 감소된 활성을 가진 변이체 서열을 확인하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 활성이 세포 활성인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 세포 활성이 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성, 세포 신호전달, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 변이체 유전자 또는 이의 단편의 서열을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 항체, 효소 또는 펩타이드의 적어도 일부를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 가이드 RNA(gRNA)를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 항체의 가변 영역 또는 불변 영역의 적어도 일부를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 항체의 적어도 하나의 CDR 영역을 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 항체의 중쇄의 CDR1, CDR2 및 CDR3, 및 경쇄의 CDR1, CDR2 및 CDR3를 코딩하는 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 복수의 동일하지 않은 폴리뉴클레오타이드들에 대한 소정의 서열에 비해 1/1000개 염기 미만의 총 오류율을 가진 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 핵산 라이브러리에서 합성된 상이한 서열들의 수가 약 50 내지 약 1,000,000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 핵산 라이브러리에서 합성된 상이한 서열들의 수가 약 500 내지 약 25000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 핵산 라이브러리에서 합성된 상이한 서열들의 수가 약 1000 내지 약 15000의 범위 내에 있는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 PCR 돌연변이유발 반응을 위한 프라이머로서 조합 라이브러리를 사용하여 핵산의 PCR 돌연변이유발을 수행하는 단계를 추가로 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정을 이용하여 변이체 서열을 가진 코돈을 결정하는 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정이 유기체에서의 코돈 서열의 빈도에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 유기체가 동물, 식물, 진균, 원생생물, 고세균 및 세균 중 적어도 하나인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다. 본원은 코돈 배정이 코돈 서열의 다양성에 기반을 둔 것인 변이체 핵산 라이브러리를 합성하는 방법을 제공한다.

[0018] 본원은 (a) 단일 기준 서열과 비교하여 변이체 서열을 가진 복수의 코돈들을 코딩하는 복수의 동일하지 않은 폴리뉴클레오타이드들을 코딩하는 소정의 서열을 디자인하는 단계; (b) 복수의 동일하지 않은 폴리뉴클레오타이드들을 합성하여 변이체 핵산 라이브러리를 생성하는 단계로서, 적어도 약 70%의 예측된 다양성이 표시되는 것인 단계; (c) 변이체 핵산 라이브러리를 발현시키는 단계; 및 (d) 변이체 핵산 라이브러리와 관련된 활성을 평가하는 단계를 포함하는, 변이체 핵산 라이브러리를 합성하는 방법을 제공한다.

[0019] **참조에 의한 도입**

[0020] 본 명세서에서 언급된 모든 공개문헌들, 특허들 및 특허출원들은 각각의 개별 공개문헌, 특허 또는 특허출원이 참고로 도입되는 것으로 구체적으로 및 개별적으로 표시된 것처럼 동일한 정도로 본원에 참고로 도입된다.

도면의 간단한 설명

[0021] 도 1은 불포화 조합 라이브러리의 생성에 대한 계략도를 보여준다.

도 2는 포화 조합 라이브러리의 생성에 대한 계략도를 보여준다.

도 3a 내지 3d는 PCR 돌연변이유발 단계를 포함하는, 변이체 생물학적 분자의 합성을 위한 과정 워크플로우(workflow)를 보여준다.

도 4a 내지 4d는 단일 소정의 코돈 부위에서 기준 핵산 서열과 상이한 핵산 서열을 포함하는 핵산의 생성을 위한 과정 워크플로우를 보여준다.

도 5a 내지 5f는 주형 핵산으로부터 핵산 변이체의 세트를 생성하기 위한 대안적 워크플로우를 보여주는데, 이

때 각각의 변이체는 단일 코돈 위치에서 상이한 핵산 서열을 포함한다. 각각의 변이체 핵산은 그의 단일 코돈 위치에서 상이한 아미노산을 코딩하고, 상이한 코돈은 X, Y 및 Z로 표시된다.

도 6a 내지 6e는 다수의 아미노산들을 가진 기준 아미노산 서열(도 6a), 및 본원에 기재된 방법을 이용함으로써 생성된 변이체 아미노산 서열(도 6b, 6c, 6d 및 6e)을 보여주는데, 이때 각각의 잔기는 단일 원으로 표시된다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 7a 및 7b는 기준 아미노산 서열(도 7a, 서열번호 24), 및 변이체 아미노산 서열의 라이브러리(도 7b, 각각 등장 순서대로 서열번호 25 내지 31)를 보여주는데, 이때 각각의 변이체는 ("X"로 표시된) 단일 잔기 변이체를 포함한다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 8a 및 8b는 기준 아미노산 서열(도 8a) 및 변이체 아미노산 서열의 라이브러리(도 8b)를 보여주는데, 이때 각각의 변이체는 단일 위치 변이체의 2개 부위들을 포함한다. 각각의 변이체는 상이하게 패턴화된 원으로 표시된다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 9a 및 9b는 기준 아미노산 서열(도 9a) 및 변이체 아미노산 서열의 라이브러리(도 9b)를 보여주는데, 이때 각각의 변이체는 아미노산의 스트레치(원 주위의 상자로 표시됨)를 포함하고, 각각의 스트레치는 서열 면에서 기준 아미노산 서열과 상이한 위치 변이체(히스티딘을 코딩함)의 3개 부위들을 가진다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 10a 및 10b는 기준 아미노산 서열(도 10a) 및 변이체 아미노산 서열의 라이브러리(도 10b)를 보여주는데, 이때 각각의 변이체는 아미노산 서열의 2개 스트레치(원 주위의 상자로 표시됨)를 포함하고, 각각의 스트레치는 서열 면에서 기준 아미노산 서열과 상이한 단일 위치 변이체(패턴화된 원으로 표시됨)의 1개 부위를 가진다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 11a 및 11b는 기준 아미노산 서열(도 11a) 및 아미노산 서열 변이체의 라이브러리(도 11b)를 보여주는데, 이때 각각의 변이체는 아미노산의 스트레치(패턴화된 원으로 표시됨)를 포함하고, 각각의 스트레치는 서열 면에서 기준 아미노산 서열과 상이한 다수의 위치 변이체들의 단일 부위를 가진다. 이 예시에서, 5개의 위치들이 변이 되는데, 이때 제1 위치는 50/50 K/R 비를 갖고; 제2 위치는 50/25/25 V/L/S 비를 갖고; 제3 위치는 50/25/25 Y/R/D 비를 갖고; 제4 위치는 모든 아미노산들에 대한 동등한 비를 갖고; 제5 위치는 G/P에 대한 75/25 비를 가진다. 기준 아미노산 서열 및 변이체 서열은 본원에 기재된 방법에 의해 생성된 핵산 및 그의 변이체에 의해 코딩된다.

도 12는 CDR1, CDR2 및 CDR3 영역을 가진 항체를 코딩하는 주형 핵산을 보여주는데, 이때 각각의 CDR 영역은 변이를 위한 다수의 부위들을 포함하고, 각각의 단일 부위(별모양으로 표시됨)는 주형 핵산 서열과 상이한 임의의 코돈 서열과 상호교환될 수 있는 단일 위치 및/또는 다수의 연속 위치들의 스트레치를 포함한다.

도 13은 예측된 변이체 분포 및 그 결과로서 생긴 변이체 다양성을 도표를 보여준다.

도 14는 발현 카세트의 변이체 라이브러리를 생성하기 위해 2개의 발현 카세트들의 부분(예를 들면, 프로모터, 오픈 리딩 프레임 및 터미네이터)을 교환함으로써 생성된 다수의 예시적 변이체들을 보여준다.

도 15는 본원에 개시된 유전자 합성을 위한 예시적 과정 워크플로우를 입증하는 단계들의 도표를 제공한다.

도 16은 컴퓨터 시스템의 일례를 예시한다.

도 17은 컴퓨터 시스템의 구성을 예시하는 블록 도표이다.

도 18은 복수의 컴퓨터 시스템들, 복수의 휴대전화들 및 개인 데이터 단말기들 및 네트워크 부착된 저장장치(NAS)를 포함하도록 구성된 네트워크를 입증하는 도표이다.

도 19는 공유된 가상 주소 메모리 공간을 이용하는 멀티프로세서 컴퓨터 시스템의 블록 도표이다.

도 20은 겔 전기영동에 의해 해상된 PCR 반응 생성물의 생물분석기 도표를 보여준다.

도 21은 PCR 생성물의 96개 세트를 보여주는 전기영동도를 보여주는데, 이때 PCR 생성물의 각각의 세트는 단일 코돈 위치에서 서열이 야생형 주형 핵산과 상이하고, 각각의 세트의 단일 코돈 위치는 야생형 주형 핵산 서열에

서 상이한 부위에 위치한다. PCR 생성물의 각각의 세트는 19개의 변이체 핵산들을 포함하고, 이때 각각의 변이체는 그의 단일 코돈 위치에서 상이한 아미노산을 코딩한다.

도 22는 변이체의 관측된 빈도와 예측된 확률을 비교하는 도표를 보여준다.

도 23은 확률 빈(bin)당 평균 카운트의 도표를 보여준다.

도 24는 PCR 생성물의 분석의 도표를 보여준다. X 축은 염기 쌍이고, Y 축은 형광 유닛이다.

도 25는 관측된 조합 변이체의 분포의 도표를 보여준다.

도 26a 내지 26d는 불포화 조합 라이브러리의 생성을 예시한다.

도 27a 내지 27c는 단일 또는 다수의 CDR 영역의 변이체의 도식을 보여준다.

도 28a는 단일 또는 다수의 중쇄 및 경쇄 스카폴드의 변이체의 도식을 보여준다.

도 28b는 단일 또는 다수의 프레임워크의 변이체의 도식을 보여준다.

발명을 실시하기 위한 구체적인 내용

[0022] 달리 표시되어 있지 않은 한, 본 개시는 당분야의 기술 내에 있는 통상적인 분자생물학 기법들을 이용한다. 달리 정의되어 있지 않은 한, 본원에서 사용된 모든 기술 용어들과 과학 용어들은 당분야에서 통상의 기술을 가진 자에 의해 통상적으로 이해되는 의미와 동일한 의미를 가진다.

[0023] 정의

[0024] 본 개시 전체에서, 수치 특징은 범위 형식으로 제시된다. 범위 형식의 기재는 편리함 및 간결함을 위한 것일 뿐이고 임의의 실시양태의 범위의 확고한 한정으로서 해석되어서는 안 된다는 것을 이해해야 한다. 따라서, 문맥이 달리 명시하지 않은 한, 범위의 기재는 모든 가능한 하위범위들뿐만 아니라 그 범위 내의 개별 수치 값도 하한의 유닛의 10분의 1까지 구체적으로 개시한 것으로 간주되어야 한다. 예를 들면, 범위, 예컨대, 1 내지 6의 기재는 하위범위, 예컨대, 1 내지 3, 1 내지 4, 1 내지 5, 2 내지 4, 2 내지 6, 3 내지 6 등뿐만 아니라 그 범위 내의 개별 값, 예를 들면, 1.1, 2, 2.3, 5 및 5.9도 구체적으로 개시한 것으로 간주되어야 한다. 이것은 범위의 폭과 관계없이 적용된다. 언급된 범위에서 임의의 한계치가 구체적으로 배제된다는 것을 조건으로, 이 개재 범위의 상한 및 하한은 보다 더 작은 범위 내에 독립적으로 포함될 수 있고, 본 발명 내에도 포함된다. 문맥이 달리 명시하지 않은 한, 언급된 범위가 한계치들 중 하나 또는 둘 다를 포함하는 경우, 이 포함된 한계치들 중 하나 또는 둘 다를 배제하는 범위도 본 발명에 포함된다.

[0025] 본원에서 사용된 용어는 특정 실시양태만을 기술하기 위한 것이고 임의의 실시양태를 한정하기 위한 것이 아니다. 본원에서 사용된 바와 같이, 문맥이 달리 명시하지 않은 한, 단수 형태는 복수 형태도 포함하기 위한 것이다. 본 명세서에서 사용될 때 용어 "포함한다" 및/또는 "포함하는"은 언급된 특징, 정수, 단계, 작업, 요소 및/또는 성분의 존재를 특정하나, 하나 이상의 다른 특징, 정수, 단계, 작업, 요소, 성분 및/또는 이들의 군의 존재 또는 추가를 배제하지 않는다는 것도 이해할 것이다. 본원에서 사용된 바와 같이, 용어 "및/또는"은 관련된 나열된 항목들 중 하나 이상의 항목의 임의의 조합 및 모든 조합을 포함한다.

[0026] 구체적으로 언급되어 있지 않거나 문맥으로부터 자명하지 않은 한, 수 또는 수의 범위와 관련하여 본원에서 사용된 용어 "약"은 언급된 수 및 이의 $\pm 10\%$ 이내의 수, 또는 범위를 위해 나열된 값들에 대한 나열된 하한의 10% 미만 및 나열된 상한의 10% 초과를 의미하는 것으로 이해된다.

[0027] 본원에서 사용된 바와 같이, 용어 "미리 선택된 서열", "미리 규정된 서열" 또는 "소정의 서열"은 상호교환가능하게 사용된다. 상기 용어들은 중합체의 서열이 중합체의 합성 또는 조립 전에 공지되어 있고 선택된다는 것을 의미한다. 구체적으로, 본 발명의 다양한 양태들은 주로 핵산 분자의 제조와 관련하여 본원에 기재되는데, 이때 올리고뉴클레오타이드 또는 폴리뉴클레오타이드의 서열은 핵산 분자의 합성 또는 조립 전에 공지되어 있고 선택된다.

[0028] 본원은 합성(즉, 드노보 합성된 또는 화학적으로 합성된) 폴리뉴클레오타이드를 생성하는 방법 및 조성물을 제공한다. 용어 올리고뉴클레오타이드, 올리고 및 폴리뉴클레오타이드는 본 명세서 전체에서 동의어인 것으로 정의된다. 본원에 기재된 합성된 폴리뉴클레오타이드의 라이브러리는 하나 이상의 유전자 또는 유전자 단편을 일괄적으로 코딩하는 복수의 폴리뉴클레오타이드들을 포함할 수 있다. 일부 경우, 폴리뉴클레오타이드 라이브러리는 코딩 또는 비-코딩 서열을 포함한다. 일부 경우, 폴리뉴클레오타이드 라이브러리는 복수의 cDNA 서열들을 코

딩한다. cDNA 서열의 기반이 되는 기준 유전자 서열은 인트론을 함유할 수 있는 반면, cDNA 서열은 인트론을 배제한다. 본원에 기재된 폴리뉴클레오타이드는 유기체로부터의 유전자 또는 유전자 단편을 코딩할 수 있다. 예시적 유기체는 원핵생물(예를 들면, 세균) 및 진핵생물(예를 들면, 마우스, 토끼, 인간 및 비-인간 영장류)을 포함하나 이들로 한정되지 않는다. 일부 경우, 폴리뉴클레오타이드 라이브러리는 다수의 엑손들의 서열을 각각 코딩하는 하나 이상의 폴리뉴클레오타이드를 포함한다. 본원에 기재된 라이브러리 내의 각각의 폴리뉴클레오타이드는 상이한 서열, 즉 동일하지 않은 서열을 코딩할 수 있다. 일부 경우, 본원에 기재된 라이브러리 내의 각각의 폴리뉴클레오타이드는 상기 라이브러리 내의 또 다른 폴리뉴클레오타이드의 서열에 상보적인 적어도 하나의 부분을 포함한다. 달리 언급되어 있지 않은 한, 본원에 기재된 폴리뉴클레오타이드 서열은 DNA 또는 RNA를 포함할 수 있다.

[0029] 본원은 합성(즉, 드노보 합성된) 유전자를 생성하는 방법 및 조성물을 제공한다. 합성 유전자를 포함하는 라이브러리는 2개 이상의 이중 가닥 폴리뉴클레오타이드들을 조합하여("연결하여") 더 큰 DNA 유닛(즉, 새시)을 생성하는, 본원의 다른 곳에 더 상세히 기재된 다양한 방법들, 예컨대, PCA, 비-PCA 유전자 조립 방법 또는 계층적 유전자 조립에 의해 구축될 수 있다. 큰 구축물의 라이브러리는 길이가 적어도 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400 또는 500 kb 이상인 폴리뉴클레오타이드를 포함할 수 있다. 큰 구축물은 독립적으로 선택된 상한인 약 5000개, 10000개, 20000개 또는 50000개 염기 쌍에 의해 한정될 수 있다. 비-코딩 DNA 또는 RNA, 예컨대, 조절 서열, 예를 들면, 프로모터, 전사 인자, 인핸서, siRNA, shRNA, RNAi, miRNA, 마이크로RNA로부터 유래한 작은 핵인 RNA, 또는 관심 있는 임의의 기능적 또는 구조적 DNA 또는 RNA를 포함하는, 비-리보솜 펩타이드(NRP)를 코딩하는 서열, 비-리보솜 펩타이드-합성효소(NRPS) 모듈 및 합성 변이체를 코딩하는 서열, 다른 모듈식 단백질, 예컨대, 항체의 폴리펩타이드 분절, 다른 단백질 패밀리로부터의 폴리펩타이드 분절을 포함하는, 임의의 수의 폴리펩타이드 분절 코딩 뉴클레오타이드 서열의 합성. 폴리뉴클레오타이드의 비한정적 예는 다음과 같다: 유전자 또는 유전자 단편의 코딩 또는 비-코딩 영역, 유전자간 DNA, 연관 분석으로부터 규정한 좌위들(좌위), 엑손, 인트론, 메신저 RNA(mRNA), 전달 RNA, 리보솜 RNA, 짧은 간섭 RNA(siRNA), 짧은 헤어핀 RNA(shRNA), 마이크로-RNA(miRNA), 작은 핵인 RNA, 리보자임, 통상적으로 메신저 RNA(mRNA)의 역전사 또는 증폭에 의해 수득되는, mRNA의 DNA 표시인 상보적 DNA(cDNA), 합성 또는 증폭에 의해 생성된 DNA 분자, 게놈 DNA, 재조합 폴리뉴클레오타이드, 분지된 폴리뉴클레오타이드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브, 및 프라이머. 본원에서 언급된 유전자 또는 유전자 단편을 코딩하는 cDNA는 상응하는 게놈 서열에서 발견되는 개재 인트론 서열 없이 엑손 서열(들)을 코딩하는 적어도 하나의 영역을 포함할 수 있다. 대안적으로, cDNA에 상응하는 게놈 서열은 우선 인트론 서열을 결여할 수 있다.

[0030] 변이체 라이브러리 합성

[0031] 본원에 기재된 방법은 적어도 하나의 소정의 기준 핵산 서열의 소정의 변이체를 각각 코딩하는 핵산의 라이브러리의 합성을 제공한다. 일부 경우, 소정의 기준 서열은 단백질을 코딩하는 핵산 서열이고, 변이체 라이브러리는 합성된 핵산에 의해 코딩된 후속 단백질에서 단일 잔기의 복수의 상이한 변이체들이 표준 번역 과정에 의해 생성되도록 적어도 단일 코돈의 변이를 코딩하는 서열을 포함한다. 핵산 서열에서의 합성된 특정 변경은 뉴클레오타이드 변화를 중첩 또는 블런트 말단 폴리뉴클레오타이드 프라이머 내에 포함시킴으로써 도입될 수 있다. 대안적으로, 폴리뉴클레오타이드 집단은 일괄적으로 긴 핵산(예를 들면, 유전자) 및 이의 변이체를 코딩할 수 있다. 이 방식에서, 폴리뉴클레오타이드 집단을 하이브리드화할 수 있고 표준 분자생물학 기법을 수행하여 긴 핵산(예를 들면, 유전자) 및 이의 변이체를 형성할 수 있다. 긴 핵산(예를 들면, 유전자) 및 이의 변이체가 세포에서 발현될 때, 변이체 단백질 라이브러리가 생성될 수 있다. 유사하게, 본원은 RNA 서열(예를 들면, miRNA, shRNA 및 mRNA) 또는 DNA 서열(예를 들면, 인핸서, 프로모터, UTR 및 터미네이터 영역)을 코딩하는 변이체 라이브러리의 합성 방법을 제공한다. 일부 경우, 서열은 엑손 서열 또는 코딩 서열이다. 일부 경우, 서열은 인트론 서열을 포함하지 않는다. 본원은 본원에 기재된 방법을 이용함으로써 합성된 라이브러리로부터 선택된 변이체에 대한 다운스트림 적용도 제공한다. 다운스트림 적용은 향상된 생물학적 관련 기능, 예를 들면, 생화학적 친화성, 효소 활성, 세포 활성의 변화를 가진, 질환 상태의 치료 또는 예방을 위한 변이체 핵산 또는 단백질 서열의 확인을 포함한다.

[0032] 조합 핵산 라이브러리

[0033] 변이체 핵산 라이브러리를 합성하는 매우 정확한 효율적 시스템을 위한 방법이 본원에 기재되어 있다. 본원은 조합 기반 변이체 라이브러리를 합성하는 방법도 제공한다. 본원에서 제공된 방법의 유리한 특징은 조합 라이브러리 내의 조립된 핵산의 생성물 및 빈도를 정확히 예측하여, 음성 또는 무효 결과와 관련된 조합 생성물뿐만

아니라 생화학적 또는 세포 활성화와 관련된 향상과 관련된 조합 생성물도 정확히 이해하면서 조합 라이브러리를 스크리닝할 수 있다는 것이다. 이러한 시스템은 현행 방법들, 즉 효율적 수단이 음성 또는 무효 결과에 대한 정보를 모을 수 없게 하는 과거 디스플레이에 비해 유리하다. 본원에서 제공된 방법의 또 다른 유리한 특징은 대표적 조합 라이브러리가 디자인되고 시험될 때, 제1 세대 조합 라이브러리의 생성물의 스크리닝으로부터 모아진 정보에 기반을 둔 정교한 변화 기준으로 제2 세대 라이브러리 및 제3 세대 라이브러리를 신속히 생성할 수 있게 하면서도 전체 포화 라이브러리에 비해 더 적은 재료 및 관련 비용이 필요하다는 것이다.

[0034] 변이체 핵산 라이브러리를 효율적으로 정확히 합성하는, 본원에 기재된 방법은 균일하고 다양한 라이브러리를 생성할 수 있다. 본원에 기재된 방법을 이용함으로써 생성된 라이브러리는 비-무작위적이다. 본원에 기재된 방법을 이용함으로써 생성된 라이브러리는 원하는 빈도로 각각의 의도된 변이체의 정확한 도입을 제공한다. 본원에 기재된 방법을 이용함으로써 생성된 라이브러리는 표시의 감소된 이탈률, 및 각각의 라이브러리 내의 폴리뉴클레오타이드 또는 더 긴 핵산의 종 전체에 걸친 개선된 균일성으로 인해 높은 정확도를 제공한다. 추가로, 폴리뉴클레오타이드 합성 수준에서의 이러한 정확도로부터의 이점은 다운스트림 적용, 예컨대, 코돈 수준에서 코딩된 소정의 변이를 포함하는 번역 생성물로부터의 단백질 활성의 평가에 있어서 기능적 수준에서 높은 정확도를 허용한다. 일부 경우, 정확한 라이브러리를 생성하는, 본원에 기재된 방법은 후속 라이브러리의 개선된 디자인을 가능하게 한다. 이러한 후속 라이브러리는 제1 라이브러리로부터의 음성 또는 무효 결과에 대해 모아진 정보의 결과로서 디자인에 더 초점이 맞춰질 수 있다. 예를 들면, 본원에 기재된 방법을 이용함으로써 합성된 제1 변이체 핵산 라이브러리를 사용하여, 특정 활성화에 대해 스크리닝되는 기능성 RNA 또는 단백질의 변이체 라이브러리를 생성할 수 있다. 정확히 규정된 비-무작위적 라이브러리와 관련된 양성 결과 및 음성 결과 둘 다의 관측을 기반으로, 특정된 활성화와 관련된 종의 추가 스크린 및 선택을 위한 추가 스크리닝 단계를 위해 추후에 사용되는 제2 변이체 라이브러리의 디자인을 선택한다. 이 과정은 1회, 2회, 3회, 4회, 5회, 6회, 7회, 8회, 9회 또는 10회 이상 반복될 수 있다. 단일 활성 또는 다수의 활성들(예를 들면, 결합 친화성, 안정성 및 발현)과 관련된 향상된 종을 확인하기 위해 라이브러리 디자인, 구축, 스크리닝 및 반복 방법을 수행할 수 있다.

[0035] 인 실리코(in silico) 라이브러리의 생성을 이용할 때, 서열은 공지되어 있을 수 있고 비-무작위적일 수 있다. 일부 경우, 라이브러리는 적어도 또는 약 10^1 개, 10^2 개, 10^3 개, 10^4 개, 10^5 개, 10^6 개, 10^7 개, 10^8 개, 10^9 개, 10^{10} 개, 또는 10^{10} 개 초과 변이체를 포함한다. 일부 경우, 적어도 또는 약 10^1 개, 10^2 개, 10^3 개, 10^4 개, 10^5 개, 10^6 개, 10^7 개, 10^8 개, 10^9 개 또는 10^{10} 개의 변이체를 포함하는 라이브러리의 각각의 변이체의 서열은 공지되어 있다. 일부 경우, 라이브러리는 예측된 다양성의 변이체를 포함한다. 일부 경우, 라이브러리에서 표시된 다양성은 적어도 또는 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 또는 95% 초과 예측된 다양성이다. 일부 경우, 라이브러리에서 표시된 다양성은 적어도 또는 약 70%의 예측된 다양성이다. 일부 경우, 라이브러리에서 표시된 다양성은 적어도 또는 약 80%의 예측된 다양성이다. 일부 경우, 라이브러리에서 표시된 다양성은 적어도 또는 약 90%의 예측된 다양성이다. 일부 경우, 라이브러리에서 표시된 다양성은 적어도 또는 약 99%의 예측된 다양성이다. 본원에 기재된 바와 같이, 용어 "예측된 다양성"은 모든 가능한 변이체들을 포함하는 집단에서의 총 이론적 다양성을 지칭한다.

[0036] 각각의 변이체의 서열이 공지되어 있는 경우 본원에 기재된 매우 균일하고 다양한 라이브러리의 생성은 향상된 또는 감소된 활성화와 관련된 조합 생성물 및 음성 또는 무효 결과와 관련된 조합 생성물을 정확히 이해할 수 있게 한다. 향상된 또는 감소된 활성화와 관련된 생성물 및 음성 또는 무효 결과와 관련된 조합 생성물을 아는 것은 라이브러리가 후속 어세이를 위해 효율적으로 사용될 수 있게 한다. 예를 들면, 큰 스크린을 수행함에 있어서, 향상된 또는 감소된 활성을 야기할 변이체 서열은 공지되어 있다. 후속 스크린을 수행함에 있어서, 향상된 또는 감소된 활성을 야기하는 변이체 서열만이 스크리닝되도록 음성 또는 무효 결과를 초래한 서열을 배제할 수 있다.

[0037] 일부 경우, 향상된 또는 감소된 활성화는 세포 활성화와 관련되어 있다. 세포 활성화는 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성화, 세포 신호전달, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하나 이들로 한정되지 않는다.

[0038] 제1 예시적 과정에서, 불포화 조합 라이브러리를 생성한다. 불포화 조합 라이브러리의 생성은 합성 단계의 수를 감소시킬 수 있다. 도 1을 참조하면, 제1 핵산 집단(110)은 위치 1, 2, 3 및 4에서 다양성을 나타낸다. 제2 핵산 집단(120)은 위치 5, 6, 7 및 8에서 다양성을 나타낸다. 제1 핵산 집단(110)을 제2 핵산 집단(120)과 조합하여 핵산 단편의 16개 조합들을 생성한다. 제1 핵산 집단(110)을 블런트 말단 라이게이션으로 제2 핵산 집단(120)과 조합할 수 있다. 일부 경우, 상기 제1 집단 및 제2 집단은 제한 효소 인식 영역을 포함하는 상보적 중

칩 서열을 가짐으로써, 각각의 집단에서 핵산의 절단 후 제1 집단과 제2 집단이 서로 어닐링할 수 있도록 디자인된다.

[0039] 일부 경우, 2개 이상의 핵산 단편들을 사용하여 핵산 라이브러리를 합성한다. 적어도 2개의 단편들, 적어도 3개의 단편들, 적어도 4개의 단편들, 적어도 5개의 단편들 또는 더 많은 단편들을 사용하여 핵산 라이브러리를 합성할 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 적어도 또는 약 적어도 10개, 15개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 100개, 150개, 200개, 300개, 400개, 500개 또는 2000개 이상의 뉴클레오타이드일 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 기껏해야 또는 약 기껏해야 2000개, 500개, 400개, 300개, 200개, 150개, 100개, 50개, 45개, 35개, 30개, 25개, 20개, 19개, 18개, 17개, 16개, 15개, 14개, 13개, 12개, 11개 또는 10개 이하의 뉴클레오타이드일 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 10개 내지 2000개, 10개 내지 500개, 9개 내지 400개, 11개 내지 300개, 12개 내지 200개, 13개 내지 150개, 14개 내지 100개, 15개 내지 50개, 16개 내지 45개, 17개 내지 40개, 18개 내지 35개, 또는 19개 내지 25개의 뉴클레오타이드일 수 있다.

[0040] 다양한 혼합 방법들, 예컨대, 라이게이션, 및 시약들이 당분야에서 공지되어 있고, 본원에서 제공된 방법을 수행하는 데 유용할 수 있다. 블런트 말단 라이게이션을 이용하여, 하나의 핵산 집단으로부터의 단편을 제2 핵산 집단으로부터의 단편과 연결할 수 있다. 리가제(ligase)는 이. 폴라이 리가제, T4 리가제, 포유동물 리가제(예를 들면, DNA 리가제 I, DNA 리가제 II, DNA 리가제 III, DNA 리가제 IV), 열안정성 리가제 및 신속 리가제를 포함할 수 있으나 이들로 한정되지 않는다. 일부 경우, PCR 연장 증폭 방법을 이용하여 2개의 단편들을 어닐링하고 연결하여 더 긴 핵산을 형성한다. 이러한 방식에서, DNA 중합효소 및 증폭 시약, 예를 들면, dNTP, 완충제 용액 및 ATP의 존재 하에서 각각의 단편이 어닐링의 위치로부터 연장되는 증폭 반응을 위해 다른 단편에 대한 프라이머로서 작용하도록 제1 단편은 제2 단편에 상보적인 영역을 가진다. 일부 경우, 제한 효소 인식 영역의 절단 후 라이게이션으로 하나의 핵산 집단으로부터의 단편을 제2 핵산 집단으로부터의 단편과 연결한다. 일부 경우, 제한 효소는 리가제에 의해 연결되는 돌출부(overhang)를 생성한다. 한 핵산 단편 대 또 다른 핵산 단편의 1:1의 몰 비가 사용될 수 있다. 일부 경우, 상기 몰 비는 적어도 1:1, 적어도 1:2, 적어도 1:3 또는 적어도 1:4 이상이다. 대안적으로, 상기 비는 적어도 2:1, 적어도 3:1 또는 적어도 4:1 이상일 수 있다. 라이게이션된 핵산 단편들의 총 몰 질량 또는 핵산 단편들 각각의 몰 질량은 적어도 또는 적어도 약 1, 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000 또는 100000 피코몰 이상일 수 있다.

[0041] 일부 경우, 본원에 기재된 방법에 의해 생성된 핵산 단편은 라이게이션 전에 블런트 말단을 가진다. T4 DNA 중합효소 또는 클레나우(Klenow) 단편을 사용하여 핵산이 블런트 말단을 갖게 할 수 있다. 대안적으로, 블런트 말단을 직접 생성하는 효소(예를 들면, Sma I, Dpn I, Pvu II, Eco RV)가 사용된다. 일부 경우, DNA 엔도뉴클레아제(endonuclease) 또는 DNA 엑소뉴클레아제(exonuclease)를 사용하여 블런트 말단을 생성한다.

[0042] 제2 예시적 워크플로우에서, 포화 조합 라이브러리가 생성된다. 도 2를 참조하면, 제1 핵산 집단(210)은 위치 1, 2, 3 및 4에서 다양성을 나타낸다. 제2 핵산 집단(220)은 위치 5, 6, 7 및 8에서 다양성을 나타낸다. 도 2에서 알 수 있는 바와 같이, 유전자 단편의 "좌측"에 있는 핵산 집단(210)은 4¹의 다양성을 가진다. 유전자 단편의 "우측"에 있는 핵산 집단(220)은 4¹의 다양성을 가진다. 그 후, 원하는 유전자의 "우측" 절반에 걸쳐 다양성을 가진 또 다른 단편과 조합된, 원하는 유전자의 "좌측" 절반에 걸쳐 다양성을 가진 긴 유전자 단편을 합성하여, 총 4²의 다양성을 생성할 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 적어도 또는 약 적어도 10개, 15개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 100개, 150개, 200개, 300개, 400개, 500개 또는 2000개 이상의 뉴클레오타이드일 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 기껏해야 또는 약 기껏해야 2000개, 500개, 400개, 300개, 200개, 150개, 100개, 50개, 45개, 35개, 30개, 25개, 20개, 19개, 18개, 17개, 16개, 15개, 14개, 13개, 12개, 11개 또는 10개 이하의 뉴클레오타이드일 수 있다. 핵산 단편 각각의 길이 또는 합성된 핵산의 평균 길이는 10개 내지 2000개, 10개 내지 500개, 9개 내지 400개, 11개 내지 300개, 12개 내지 200개, 13개 내지 150개, 14개 내지 100개, 15개 내지 50개, 16개 내지 45개, 17개 내지 40개, 18개 내지 35개, 또는 19개 내지 25개의 뉴클레오타이드일 수 있다.

[0043] 생성된 핵산은 검증될 수 있다. 일부 경우, 핵산은 시퀀싱에 의해 검증된다. 일부 경우, 핵산은 고처리량 시퀀싱, 예컨대, 차세대 시퀀싱에 의해 검증된다. 단일 분자 실시간(SMRT) 시퀀싱, 폴로니(Polony) 시퀀싱, 라이게이션에 의한 시퀀싱, 가역적 터미네이터 시퀀싱, 양성자 검출 시퀀싱, 이온 반도체 시퀀싱, 나노포어 시퀀싱, 전자적 시퀀싱, 피로시퀀싱, 막삼-길버트(Maxam-Gilbert) 시퀀싱, 쇠 중결(예를 들면, 생거(Sanger)) 시퀀싱,

+S 시퀀싱 또는 합성에 의한 시퀀싱을 포함하나 이들로 한정되지 않는 임의의 적절한 시퀀싱 기술로 시퀀싱 라이브러리의 시퀀싱을 수행할 수 있다.

[0044] 본원은 변이의 정도에서 불포화된 또는 포화된 핵산 라이브러리를 매우 정확히 합성하는 방법을 제공한다. 일부 경우, 약 70%의 핵산은 삽입 및 결실을 갖지 않는다. 일부 경우, 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 또는 99% 초과와 핵산은 삽입 및 결실을 갖지 않는다. 일부 경우, 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 또는 99% 초과와 핵산은 삽입 및 결실을 갖지 않는다. 일부 경우, 90% 초과와 핵산은 삽입 및 결실을 갖지 않는다. 일부 경우, 적어도 80%의 핵산은 오류를 갖지 않는다. 일부 경우, 적어도 약 70%, 75%, 80%, 85%, 90%, 95% 또는 99% 이상의 핵산은 오류를 갖지 않는다.

[0045] 본원은 변이의 정도에서 불포화된 또는 포화된 핵산 라이브러리를 매우 정확히 합성하는 방법을 제공한다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 80% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 1.5배 내에서 표시된다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 80% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 1.5배, 2배, 2.5배, 3배, 3.5배 또는 4배 내에서 표시된다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 90% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 1.5배 내에서 표시된다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 90% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 1.5배, 2배, 2.5배, 3배, 3.5배 또는 4배 내에서 표시된다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 80% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 2배 내에서 표시된다. 일부 경우, 본원에 기재된 드노보 합성된 핵산 라이브러리에서 80% 초과와 핵산은 증폭 후 전체 라이브러리에 대한 평균 표시의 적어도 약 2배 이내에서 표시된다.

[0046] 대표적 핵산 라이브러리의 생성

[0047] 변이체 코돈 코딩 영역의 미리 선택된 분포를 가진 핵산 라이브러리를 합성하는 방법이 본원에 기재되어 있다. 더욱이, 이러한 라이브러리는 대표적 분포에 대한 이해를 제공하면서 미리 선택된 분포에 대해 불포화될 수 있다. 본원은 일단 번역되면 특정 위치에서 아미노산의 미리 선택된 분포를 제공하는 핵산의 생성에 관한 방법도 제공한다. 미리 선택된 분포로부터 무작위적 샘플을 생성함으로써, 미리 선택된 집단 분포에 가까운 대표적 분포를 갖도록 덜 포화된 핵산 라이브러리를 디자인한다. 미리 선택된 집단 분포에 가까운 대표적 분포를 가진, 본원에 기재된 핵산 라이브러리는 원하는 미리 선택된 분포에서 각각의 의도된 변이체의 정확한 도입을 추가로 포함할 수 있다.

[0048] 본원에 기재된 전산화 기법은 무작위적 샘플링을 포함하나 이것으로 한정되지 않는다. 첫 번째 과정에서, 각각의 위치에서 코돈 변이의 미리 선택된 분포에 대해, 각각의 위치에 대한 누적 분포 값을 계산한다. 일부 경우, 누적 분포 값은 약 0.0과 1.0 사이의 확률에 맵핑된다. 핵산 집단에 대해, 누적 분포 값은 특정 위치에서 코돈 변이체의 확률을 결정하는 것을 제공한다. 예를 들면, 각각의 위치에서 코돈 변이체가 핵산 집단 전체에 걸쳐 나타나는 횟수를 합산한 후, 각각의 아미노산이 각각의 위치에서 나타나는 퍼센트를 측정할 수 있다. 그 다음, 샘플 핵산 집단에서의 퍼센트를 미리 선택된 분포와 비교한다. 집단 내의 충분한 수의 핵산을 사용하여, 미리 선택된 분포와 일치하는 샘플 분포를 생성한다. 일부 경우, 수행된 샘플링은 균일한 무작위 샘플링을 적용하는 몬테 카를로(Monte Carlo) 샘플링의 형태이다.

[0049] 일부 경우, 미리 선택된 분포를 갖도록 디자인되고 합성된 핵산 라이브러리는 포화 핵산 라이브러리에 비해 약 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 또는 60% 초과와 동일하지 않은 핵산을 코딩한다. 일부 경우, 미리 선택된 분포를 갖도록 디자인되고 합성된 핵산 라이브러리는 포화 핵산 라이브러리에 비해 적어도 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 또는 60% 초과와 동일하지 않은 핵산을 코딩한다.

[0050] 일부 경우, 미리 선택된 분포를 갖도록 디자인되고 합성된 핵산 라이브러리는 더 큰 핵산 라이브러리에 비해 약 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 또는 60% 초과와 동일하지 않은 핵산을 코딩한다. 일부 경우, 미리 선택된 분포를 갖도록 디자인되고 합성된 핵산 라이브러리는 더 큰 핵산 라이브러리에 비해 적어도 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 또는 60% 초과와 동일하지 않은 핵산을 코딩한다.

[0051] 일부 경우, 더 큰 변이체 핵산 라이브러리로부터의 대표적 하위집단에서 디자인되고 합성된 핵산의 수는 약 50개 내지 100000개, 100개 내지 75000개, 250개 내지 50000개, 500개 내지 25000개, 1000개 내지 15000개, 2000

개 내지 10000개, 및 4000개 내지 8000개 서열의 범위 내에 있다. 일부 경우, 핵산 집단은 500개 서열이다. 일부 경우, 핵산 집단은 5000개, 10000개 또는 15000개 서열이다. 일부 경우, 핵산 집단은 적어도 50개, 100개, 150개, 500개, 1000개, 2000개, 5000개, 10000개, 20000개, 50000개, 100000개, 200000개, 400000개, 800000개 또는 1000000개 이상의 상이한 서열들을 가진다. 일부 경우, 각각의 핵산 집단은 최대 50, 100, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 400000, 800000 또는 1000000이다.

[0052] 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 70% 내지 99%의 예측된 다양성을 표시한다. 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 적어도 70%의 예측된 다양성을 표시한다. 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 70% 내지 75%, 70% 내지 80%, 70% 내지 85%, 70% 내지 90%, 70% 내지 95%, 70% 내지 97%, 70% 내지 99%, 75% 내지 80%, 75% 내지 85%, 75% 내지 90%, 75% 내지 95%, 75% 내지 97%, 75% 내지 99%, 80% 내지 85%, 80% 내지 90%, 80% 내지 95%, 80% 내지 97%, 80% 내지 99%, 85% 내지 90%, 85% 내지 95%, 85% 내지 97%, 85% 내지 99%, 90% 내지 95%, 90% 내지 97%, 90% 내지 99%, 95% 내지 97%, 95% 내지 99%, 또는 97% 내지 99%의 예측된 다양성을 표시한다. 일부 경우, 합성된 대표적 핵산 집단의 표시된 다양성은 적어도 또는 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 또는 95% 초과 예측된 다양성이다. 일부 경우, 합성된 대표적 핵산 집단의 표시된 다양성은 99%의 예측된 다양성이다.

[0053] **조합 방법을 이용한 대표적 핵산 라이브러리의 생성**

[0054] 본원은 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위해 조합 방법으로 핵산 라이브러리를 합성하는 방법을 제공한다. 일부 경우, 핵산 집단을 합성하기 위해 변이체에 대한 주형으로서 사용되는 기준 서열은 제1 부분이 제1 변이체 핵산 집단에 대한 기준 서열이고 제2 부분이 제2 변이체 핵산 집단에 대한 기준 서열이도록 분할된다.

[0055] 일부 경우, 본원에 기재된 무작위 샘플링 방법을 이용하여, 더 큰 변이체 라이브러리로부터 부분에 대한 대표적 변이체 분포를 생성한다. 전체 기준 서열의 제1 부분에 대한 변이체를 대표하는 제1 대표적 핵산 집단, 및 전체 기준 서열의 제2 부분에 대한 변이체를 대표하는 제2 대표적 핵산 집단을 합성한 후, 라이게이션, 예컨대, 블런트 말단 라이게이션 또는 당분야에서 공지되어 있는 일부 다른 생화학적 기법으로 조합한다. 일부 경우, 생성된 핵산 라이브러리는 포화 핵산 라이브러리이다. 일부 경우, 생성된 핵산 라이브러리는 불포화 핵산 라이브러리이다.

[0056] 일부 경우, 연결될 때 원하는 더 긴 핵산 변이체 라이브러리를 생성하는 2개 이상의 변이체 핵산 집단들로 핵산 라이브러리를 합성한다. 기준 핵산의 상이한 영역을 각각 코딩하는 적어도 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 또는 10개 초과 집단들로 핵산 라이브러리를 합성할 수 있다. 일부 경우, 각각의 핵산 집단은 약 50개 내지 100000개, 100개 내지 75000개, 250개 내지 50000개, 500개 내지 25000개, 1000개 내지 15000개, 2000개 내지 10000개, 및 4000개 내지 8000개 서열의 범위 내에 있다. 일부 경우, 각각의 핵산 집단은 약 500개, 1000개, 5000개, 10000개 또는 15000개 이상의 서열이다. 일부 경우, 각각의 핵산 집단은 적어도 50개, 100개, 150개, 500개, 1000개, 2000개, 5000개, 10000개, 20000개, 50000개, 100000개, 200000개, 400000개, 800000개 또는 1000000개 이상의 서열이다. 일부 경우, 각각의 핵산 집단은 최대 50개, 100개, 500개, 1000개, 2000개, 5000개, 10000개, 20000개, 50000개, 100000개, 200000개, 400000개, 800000개 및 1000000개의 서열이다.

[0057] 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 70% 내지 99%의 예측된 다양성을 표시한다. 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 적어도 70%의 예측된 다양성을 표시한다. 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 70% 내지 75%, 70% 내지 80%, 70% 내지 85%, 70% 내지 90%, 70% 내지 95%, 70% 내지 97%, 70% 내지 99%, 75% 내지 80%, 75% 내지 85%, 75% 내지 90%, 75% 내지 95%, 75% 내지 97%, 75% 내지 99%, 80% 내지 85%, 80% 내지 90%, 80% 내지 95%, 80% 내지 97%, 80% 내지 99%, 85% 내지 90%, 85% 내지 95%, 85% 내지 97%, 85% 내지 99%, 90% 내지 95%, 90% 내지 97%, 90% 내지 99%, 95% 내지 97%, 95% 내지 99%, 또는 97% 내지 99%의 예측된 다양성을 표시한다. 일부 경우, 변이체 코돈 코딩 영역의 미리 선택된 분포에 도달하기 위한 조합 방법에 의한 핵산 라이브러리의 합성은 적어도 또는 약 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 또는 95% 초과 예측된 다양성을 표시한다. 일부 경우, 합성된 대표적 핵산 집단의 표시된 다양성은 99%의 예측된 다양성이다.

[0058] 합성에 이은 PCR 돌연변이유발

[0059] 본원에 기재된 조합 방법에 의해 생성된 핵산 라이브러리(예를 들면, 포화 또는 불포화)는 PCR 돌연변이유발 방법을 위해 사용될 수 있다. 일부 경우, 미리 선택된 분포를 가진 대표적 핵산 라이브러리는 PCR 돌연변이유발 방법을 위해 사용된다. 이 워크플로우에서, 복수의 폴리뉴클레오타이드들이 합성되고, 이때 각각의 폴리뉴클레오타이드는 기준 핵산 서열의 소정의 변이체인 소정의 서열을 코딩한다. 도면을 참조하건대, 도 3a 내지 3d에 묘사된 예시적 워크플로우에서, 폴리뉴클레오타이드는 표면에서 생성된다. 도 3a는 121개의 좌위들을 가진 표면의 단일 클러스터의 확대도를 보여준다. 도 3b에 묘사된 각각의 핵산은 기준 핵산 서열로부터의 증폭을 위해 사용되어 변이체 긴 핵산의 라이브러리를 생성할 수 있는 프라이머이다. 그 후, 임의적으로 변이체 긴 핵산의 라이브러리에 대한 전사 및/또는 번역을 수행하여 변이체 RNA 또는 단백질 라이브러리를 생성한다(도 3d). 이 예시에서, 실질적으로 평평한 표면을 가진 디바이스가 도 3a에 묘사된 폴리뉴클레오타이드의 드노보 합성을 위해 사용된다. 일부 경우, 상기 디바이스는 좌위들의 클러스터를 포함하고, 이때 각각의 좌위는 폴리뉴클레오타이드 연장을 위한 부위이다. 일부 경우, 단일 클러스터는 원하는 변이체 서열 라이브러리를 생성하기 위해 필요한 모든 폴리뉴클레오타이드 변이체들을 포함한다. 대안적 방식에서, 플레이트는 클러스터로 분리되지 않은 좌위들의 필드(field)를 포함한다.

[0060] 본원은 클러스터 내에서 폴리뉴클레오타이드를 합성한 후(예를 들면, 도 3에 나타냄) 단일 클러스터 내에서 폴리뉴클레오타이드를 증폭하는 방법을 제공한다. 이러한 방식은 클러스터링 방식을 이용하지 않으면서 전체 플레이트에 걸쳐 동일하지 않은 폴리뉴클레오타이드를 증폭하는 것에 비해 개선된 핵산 표시를 제공한다. 일부 경우, 클러스터 내의 좌위의 표면에서 합성된 폴리뉴클레오타이드의 증폭은 과도한 GC 함량을 가진 폴리뉴클레오타이드를 가진 큰 폴리뉴클레오타이드 집단의 반복된 합성으로 인한 표시에 대한 부정적 효과를 극복한다. 일부 경우, 본원에 기재된 클러스터는 약 50개 내지 1000개, 75개 내지 900개, 100개 내지 800개, 125개 내지 700개, 150개 내지 600개, 200개 내지 500개, 또는 300개 내지 400개의 분리된 좌위들을 포함한다. 일부 경우, 좌위는 스폿(spot), 웰(well), 마이크로웰(microwell), 채널(channel) 또는 포스트(post)이다. 일부 경우, 각각의 클러스터는 동일한 서열을 가진 폴리뉴클레오타이드의 연장을 지지하는 분리된 특징형상(feature)의 적어도 1배, 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배 또는 10배 이상의 중복을 가진다. 일부 경우, 1배 중복은 동일한 서열을 가진 폴리뉴클레오타이드를 갖지 않는다는 것을 의미한다.

[0061] 본원에 기재된 드노보 합성된 폴리뉴클레오타이드 라이브러리는 제1 위치인 위치 "x"에서 적어도 하나의 변이체 서열을 각각 가진 복수의 폴리뉴클레오타이드들을 포함할 수 있고, 각각의 변이체 폴리뉴클레오타이드를 PCR의 제1 라운드에서 프라이머로서 사용하여 제1 연장 생성물을 생성한다. 이 예에서, 제1 폴리뉴클레오타이드(420)의 위치 "x"는 변이체 코돈 서열, 즉 기준 서열로부터 가능한 19개의 변이체들 중 하나를 코딩한다. 도 4a를 참조한다. 제1 폴리뉴클레오타이드의 서열과 중첩되는 서열을 포함하는 제2 폴리뉴클레오타이드(425)도 PCR의 별도 라운드에서 프라이머로서 사용하여 제2 연장 생성물을 생성한다. 추가로, 외부 프라이머(415 및 430)를 사용하여 긴 핵산 서열로부터 단편을 증폭할 수 있다. 생성된 증폭 생성물은 긴 핵산 서열의 단편(435 및 440)이다. 도 4b를 참조한다. 그 후, 긴 핵산 서열의 단편(435 및 440)을 하이브리드화하고 연장 반응을 수행하여, 긴 핵산의 변이체(445)를 형성한다. 도 4c를 참조한다. 제1 연장 생성물 및 제2 연장 생성물의 중첩되는 말단을 PCR의 제2 라운드의 프라이머로서 사용함으로써, 상기 변이체를 함유하는 제3 연장 생성물을 생성할 수 있다(도 4d). 수율을 증가시키기 위해, DNA 중합효소, 증폭 시약 및 외부 프라이머(415 및 430)를 포함하는 반응에서 긴 핵산의 변이체를 증폭한다. 일부 경우, 제2 폴리뉴클레오타이드는 변이체 부위에 인접하되 이 부위를 포함하지 않는 서열을 포함한다. 대안적 방식에서, 제2 폴리뉴클레오타이드와 중첩되는 영역을 가진 제1 폴리뉴클레오타이드를 생성한다. 이 시나리오에서, 최대 19개의 변이체들을 위해 단일 코돈에서 변이를 가진 제1 핵산을 합성한다. 제2 핵산은 변이체 서열을 포함하지 않는다. 임의적으로, 제1 집단은 제1 폴리뉴클레오타이드 변이체, 및 상이한 코돈 부위에서 변이체를 코딩하는 추가 폴리뉴클레오타이드를 포함한다. 대안적으로, 블런트 말단 라이게이션을 위해 제1 폴리뉴클레오타이드 및 제2 폴리뉴클레오타이드를 디자인할 수 있다.

[0062] 대안적 돌연변이유발 PCR 방법이 도 5a 내지 5f에 묘사되어 있다. 이러한 과정에서, 제1 가닥 및 제2 가닥(505 및 510)을 포함하는 주형 핵산 분자(500)를, 제1 프라이머(515) 및 제2 프라이머(520)를 함유하는 PCR 반응에서 증폭한다(도 5a). 상기 증폭 반응은 뉴클레오타이드 시약으로서 우라실을 포함한다. 우라실-표지부착된 연장 생성물(525)을 생성하고(도 5b), 임의적으로 정제하고, 제1 폴리뉴클레오타이드(535) 및 복수의 제2 폴리뉴클레오타이드들(530)을 사용한 후속 PCR 반응을 위한 주형으로서 사용하여 제1 연장 생성물(540 및 545)을 생성한다(도 5c 및 5d). 이 과정에서, 복수의 폴리뉴클레오타이드들(530)은 변이체 서열을 코딩하는 폴리뉴클레오타이드를 포함한다(도 5c에서 X, Y 및 Z로서 표시됨). 우라실-표지부착된 주형 핵산을 우라실 특이적 절단 시약, 예를

들면, 뉴 잉글랜드 바이오랩스(New England Biolabs)로부터 상업적으로 입수될 수 있는 USER 분해제로 분해한다. 변이체(535), 및 변이체 X, Y 및 Z를 가진 상이한 코돈들(530)을 첨가하고, 제한된 PCR 단계를 수행하여 도 5d를 생성한다. 우라실 함유 주형이 분해된 후, 연장 생성물의 중첩 말단은 제1 외부 프라이머(550) 및 제2 외부 프라이머(555)와 함께 프라이머로서 작용하는 제1 연장 생성물(540 및 545)을 사용한 PCR 반응을 프라이밍하는 데 사용됨으로써, 변이체 부위에서 복수의 변이체 X, Y 및 Z를 함유하는 핵산 분자의 라이브러리(560)를 생성한다(도 5f).

[0063] 긴 핵산의 변이체 부분 및 비-변이체 부분을 가진 집단의 드노보 합성

[0064] 본원에 기재된 조합 방법에 의해 생성된 핵산 라이브러리(예를 들면, 포화 또는 불포화)는 긴 핵산의 다수의 단편들의 드노보 합성을 위해 사용될 수 있고, 이때 상기 단편들 중 적어도 하나의 단편은 다수의 버전들로 합성되고, 각각의 버전은 상이한 변이체 서열을 가진다. 일부 경우, 미리 선택된 분포를 가진 대표적 핵산 라이브러리는 드노보 합성을 위해 사용되고, 이때 상기 단편들 중 적어도 하나의 단편은 다수의 버전들로 합성되고, 각각의 버전은 상이한 변이체 서열을 가진다. 이 방식에서, 변이체 긴 범위 핵산의 라이브러리를 조립하는 데 필요한 모든 단편들이 드노보 합성된다. 합성된 단편들은 합성 후 단편 라이브러리가 하이브리드화되도록 중첩 서열을 가질 수 있다. 하이브리드화 후, 연장 반응을 수행하여 임의의 상보적 겹을 채울 수 있다.

[0065] 대안적으로, 합성된 단편을 프라이머로 증폭한 후, 블런트 말단 라이게이션 또는 중첩 하이브리드화를 수행할 수 있다. 일부 경우, 디바이스는 좌위들의 클러스터를 포함하고, 이때 각각의 좌위는 폴리뉴클레오타이드 연장을 위한 부위이다. 일부 경우, 단일 클러스터는 원하는 변이체 핵산 서열 라이브러리를 생성하도록 소정의 긴 핵산의 모든 폴리뉴클레오타이드 변이체들 및 다른 단편 서열들을 포함한다. 상기 클러스터는 약 50개 내지 500개의 좌위들을 포함할 수 있다. 일부 방식에서, 클러스터는 500개 초과 좌위들을 포함한다.

[0066] 제1 폴리뉴클레오타이드 집단 내의 각각의 개별 폴리뉴클레오타이드는 클러스터의 분리된 개별적으로 주소지정 가능한 좌위에서 생성될 수 있다. 하나의 폴리뉴클레오타이드 변이체는 복수의 개별적으로 주소지정 가능한 좌위들에 의해 표시될 수 있다. 제1 폴리뉴클레오타이드 집단 내의 각각의 변이체는 1회, 2회, 3회, 4회, 5회, 6회, 7회, 8회, 9회 또는 10회 이상 표시될 수 있다. 일부 경우, 제1 폴리뉴클레오타이드 집단 내의 각각의 변이체는 3개 이하의 좌위들에서 표시된다. 일부 경우, 제1 폴리뉴클레오타이드 집단 내의 각각의 변이체는 2개의 좌위들에서 표시된다. 일부 경우, 제1 폴리뉴클레오타이드 집단 내의 각각의 변이체는 단일 좌위에서만 표시된다.

[0067] 본원은 감소된 중복을 가진 핵산 라이브러리를 생성하는 방법을 제공한다. 일부 경우, 원하는 변이체 핵산을 수득하기 위해 변이체 핵산을 1회 초과 빈도로 합성할 필요 없이 변이체 핵산을 생성할 수 있다. 일부 경우, 본 개시는 원하는 변이체 핵산을 생성하기 위해 변이체 핵산을 1회 초과, 2회, 3회, 4회, 5회, 6회, 7회, 8회, 9회 또는 10회 이상의 빈도로 합성할 필요 없이 변이체 핵산을 생성하는 방법을 제공한다.

[0068] 원하는 변이체 핵산을 수득하기 위해 1개 초과 분리된 부위에서 변이체 핵산을 합성할 필요 없이 변이체 핵산을 생성할 수 있다. 본 개시는 원하는 변이체 핵산을 생성하기 위해 1개 초과 부위, 2개의 부위, 3개의 부위, 4개의 부위, 5개의 부위, 6개의 부위, 7개의 부위, 8개의 부위, 9개의 부위 또는 10개의 부위에서 변이체 핵산을 합성할 필요 없이 변이체 핵산을 생성하는 방법을 제공한다. 일부 경우, 기껏해야 6개, 5개, 4개, 3개, 2개 또는 1개의 분리된 부위에서 핵산을 합성한다. 표면 위의 1개, 2개 또는 3개의 분리된 좌위에서 동일한 핵산을 합성할 수 있다.

[0069] 일부 경우, 단일 변이체 핵산을 표시하는 좌위의 양은 다운스트림 프로세싱, 예를 들면, 증폭 반응 또는 세포 어레이를 위해 요구된 핵산 물질의 양의 함수이다. 일부 경우, 단일 변이체 핵산을 표시하는 좌위의 양은 단일 클러스터에서 이용될 수 있는 좌위의 함수이다.

[0070] 본원은 기준 핵산의 복수의 부위들에서 상이한 변이체 핵산을 포함하는 핵산의 라이브러리를 생성하는 방법을 제공한다. 이러한 경우, 각각의 변이체 라이브러리는 좌위의 클러스터 내의 개별적으로 주소지정 가능한 좌위에서 생성된다. 핵산 라이브러리에 의해 표시된 변이체 부위의 수는 클러스터 내의 개별적으로 주소지정 가능한 좌위의 수 및 각각의 부위에서의 원하는 변이체의 수에 의해 결정될 것이라는 것을 이해할 것이다. 일부 경우, 각각의 클러스터는 약 50개 내지 500개의 좌위들을 포함한다. 일부 경우, 각각의 클러스터는 100개 내지 150개의 좌위들을 포함한다.

[0071] 예시적 방식에서, 19개의 가능한 변이체 아미노산들 각각을 코딩하는 코돈에 상응하는 변이체 부위에서 19개의 변이체들이 표시된다. 또 다른 예시적 경우, 19개의 가능한 변이체 아미노산들 각각을 코딩하는 삼중체(triplet)에 상응하는 변이체 부위에서 61개의 변이체들이 표시된다. 비한정적 예에서, 클러스터는 121개의 개

별적으로 주소지정가능한 좌위들을 포함한다. 이 예에서, 핵산 집단은 단일 부위 변이체 각각의 6개 복제물(6개 복제물 x 1개 변이체 부위 x 19개 변이체 = 114개 좌위), 이중 부위 변이체 각각의 3개 복제물(3개 복제물 x 2개 변이체 부위 x 19개 변이체 = 114개 좌위), 또는 삼중 부위 변이체 각각의 2개 복제물(2개 복제물 x 3개 변이체 부위 x 19개 변이체 = 114개 좌위)을 포함한다. 일부 경우, 핵산 집단은 4개, 5개, 6개, 또는 6개 초과 변이체 부위에서 변이체를 포함한다.

[0072] 본원은 합성(즉, 드노보 합성된 또는 화학적으로 합성된) 핵산을 생성하는 방법 및 조성물을 제공한다. 본원에 기재된 합성된 핵산의 라이브러리는 하나 이상의 유전자 또는 유전자 단편을 일괄적으로 코딩하는 복수의 핵산들을 포함할 수 있다. 일부 경우, 핵산 라이브러리는 코딩 또는 비-코딩 서열을 포함한다. 일부 경우, 핵산 라이브러리는 복수의 cDNA 서열들을 코딩한다. 일부 경우, 핵산 라이브러리는 다수의 엑손들의 서열을 각각 코딩하는 하나 이상의 핵산을 포함한다. 본원에 기재된 라이브러리 내의 각각의 핵산은 상이한 서열, 즉 동일하지 않은 서열을 코딩할 수 있다. 일부 경우, 본원에 기재된 라이브러리 내의 각각의 핵산은 상기 라이브러리 내의 또 다른 핵산의 서열에 상보적인 적어도 하나의 부분을 포함한다. 달리 언급되어 있지 않은 한, 본원에 기재된 핵산 서열은 DNA 또는 RNA를 포함할 수 있다.

[0073] 본원은 합성(즉, 드노보 합성된) 유전자를 생성하는 방법 및 조성물을 제공한다. 합성 유전자를 포함하는 라이브러리는 2개 이상의 이중 가닥 핵산을 조합하여("연결하여") 더 큰 DNA 유닛(즉, 새시)을 생성하는, 본원의 다른 곳에 더 상세히 기재된 다양한 방법들, 예컨대, PCA, 비-PCA 유전자 조립 방법 또는 계층적 유전자 조립에 의해 구축될 수 있다. 큰 구축물의 라이브러리는 길이가 적어도 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400 또는 500 kb 이상인 핵산을 포함할 수 있다. 상기 큰 구축물은 독립적으로 선택된 상한인 약 5000개, 10000개, 20000개 또는 50000개의 염기 쌍에 의해 제한될 수 있다. 임의의 수의 폴리펩타이드 분절 코딩 뉴클레오타이드 서열의 합성은 비-코딩 DNA 또는 RNA, 예컨대, 조절 서열, 예를 들면, 프로모터, 전사 인자, 인핸서, siRNA, shRNA, RNAi, miRNA, 마이크로RNA로부터 유래한 작은 핵인 RNA, 또는 관심 있는 임의의 기능적 또는 구조적 DNA 또는 RNA 유닛을 포함하는, 비-리보솜 펩타이드(NRP)를 코딩하는 서열, 비-리보솜 펩타이드-합성효소(NRPS) 모듈 및 합성 변이체를 코딩하는 서열, 다른 모듈식 단백질, 예컨대, 항체의 폴리펩타이드 분절, 다른 단백질 패밀리로부터의 폴리펩타이드 분절을 포함할 수 있다. 핵산의 비한정적 예는 다음과 같다: 유전자 또는 유전자 단편의 코딩 또는 비-코딩 영역, 유전자간 DNA, 연관 분석으로부터 규정된 좌위들(좌위), 엑손, 인트론, 메신저 RNA(mRNA), 전달 RNA, 리보솜 RNA, 짧은 간섭 RNA(siRNA), 짧은 헤어핀 RNA(shRNA), 마이크로-RNA(miRNA), 작은 핵인 RNA, 리보자임, 통상적으로 메신저 RNA(mRNA)의 역전사 또는 증폭에 의해 수득되는, mRNA의 DNA 표지인 cDNA, 합성 또는 증폭에 의해 생성된 DNA 분자, 게놈 DNA, 재조합 폴리뉴클레오타이드, 분지된 폴리뉴클레오타이드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브, 및 프라이머. cDNA와 관련하여, 용어 유전자 또는 유전자 단편은 개재 인트론 서열 없이 엑손 서열을 코딩하는 적어도 하나의 영역을 포함하는 DNA 핵산 서열을 지칭한다.

[0074] 다양한 실시양태들에서, 본원에 기재된 방법 및 조성물은 유전자의 라이브러리와 관련된다. 유전자 라이브러리는 복수의 하위분절들을 포함할 수 있다. 하나 이상의 하위분절에서, 라이브러리의 유전자들은 함께 공유결합될 수 있다. 하나 이상의 하위분절에서, 라이브러리의 유전자는 하나 이상의 대사 최종 생성물을 가진 제1 대사 경로의 성분을 코딩할 수 있다. 하나 이상의 하위분절에서, 하나 이상의 표적화된 대사 최종 생성물의 제조 과정을 기반으로 라이브러리의 유전자를 선택할 수 있다. 하나 이상의 대사 최종 생성물은 생물연료를 포함할 수 있다. 하나 이상의 하위분절에서, 라이브러리의 유전자는 하나 이상의 대사 최종 생성물을 가진 제2 대사 경로의 성분을 코딩할 수 있다. 제1 대사 경로 및 제2 대사 경로의 하나 이상의 최종 생성물은 하나 이상의 공유된 최종 생성물을 포함할 수 있다. 일부 경우, 제1 대사 경로는 제2 대사 경로에서 조작되는 최종 생성물을 포함한다.

[0075] *유기체를 위한 변이체 핵산 라이브러리*

[0076] 본원에 기재된 방법에 의해 생성된 변이체 핵산 라이브러리는 유기체의 적어도 하나의 유전자를 코딩할 수 있다. 일부 경우, 핵산 라이브러리는 유기체의 단일 유전자, 경로 또는 전체 게놈을 코딩한다. 일부 경우, 변이체 핵산 라이브러리는 유기체의 유전자(예를 들면, 1000개의 염기 쌍), 부분(예를 들면, 3개 내지 10개의 유전자), 경로(예를 들면, 10개 내지 100개의 유전자) 또는 새시(예를 들면, 100개 내지 1000개의 유전자) 중 적어도 하나를 코딩한다. 모델 유기체의 비한정적 예시적 목록은 표 1에서 제공되어 있다.

표 1

모델 유기체 및 유전자 수

모델 유기체	단백질 코딩 유전자*
아라비돕시스 탈리아나(<i>Arabidopsis thaliana</i>)	27000
캐노랍디티스 엘레강스(<i>Caenorhabditis elegans</i>)	20000
캐니스 루푸스 패밀리아리스(<i>Canis lupus familiaris</i>)	19000
클라미도모나스 레인하르트티(<i>Chlamydomonas reinhardtii</i>)	14000
다니오 레리오(<i>Danio rerio</i>)	26000
딕티오스텔리움 디스코이데움(<i>Dictyostelium discoideum</i>)	13000
드로소필라 멜라노가스터(<i>Drosophila melanogaster</i>)	14000
에스케리키아 콜라이(<i>Escherichia coli</i>)	4300
마카카 무라타(<i>Macaca mulatta</i>)	22000
무스 무스쿨루스(<i>Mus musculus</i>)	20000
오릭툴라구스 쿠니쿨루스(<i>Oryctolagus cuniculus</i>)	27000
라투스 노르베기쿠스(<i>Rattus norvegicus</i>)	22000
사카로마이세스 세레비시iae(<i>Saccharomyces cerevisiae</i>)	6600
수스 스크로파(<i>Sus scrofa</i>)	21000

* 숫자는 단백질 코딩 유전자의 수를 반영하고 tRNA 및 비-코딩 RNA 를 배제한다. Ron Milo & Rob Phillips, Cell Biology by the Numbers 286 (2015).

[0077]

[0078]

코돈 변이

[0079]

본원에 기재된 변이체 핵산 라이브러리는 기준 핵산 서열에 비해 변이체 코돈 서열을 각각 코딩하는 복수의 핵산들을 포함할 수 있다. 일부 경우, 제1 핵산 집단의 각각의 핵산은 단일 변이체 부위에서 변이체를 함유한다. 일부 경우, 제1 핵산 집단은 제1 핵산 집단이 동일한 변이체 부위에서 하나 초과와 변이체를 함유하도록 단일 변이체 부위에서 복수의 변이체들을 함유한다. 제1 핵산 집단은 동일한 변이체 부위에서 다수의 코돈 변이체들을 일괄적으로 코딩하는 핵산을 포함할 수 있다. 제1 핵산 집단은 동일한 위치에서 최대 19개 이상의 코돈들을 일괄적으로 코딩하는 핵산을 포함할 수 있다. 제1 핵산 집단은 동일한 위치에서 최대 60개의 변이체 삼중체들을 일괄적으로 코딩하는 핵산을 포함할 수 있거나, 제1 핵산 집단은 동일한 위치에서 최대 61개의 상이한 코돈 삼중체들을 일괄적으로 코딩하는 핵산을 포함할 수 있다. 각각의 변이체는 번역 동안 상이한 아미노산을 야기하는 코돈을 코딩할 수 있다. 표 2는 변이체 부위에 대한 가능한 각각의 코돈의 목록(및 대표적 아미노산)을 제공한다.

표 2

코돈 및 아미노산의 목록

아미노산	1 문자 코드	3 문자 코드	코돈					
알라닌	A	Ala	GCA	GCC	GCG	GCT		
시스테인	C	Cys	TGC	TGT				
아스파르트산	D	Asp	GAC	GAT				
글루탐산	E	Glu	GAA	GAG				
페닐알라닌	F	Phe	TTC	TTT				
글리신	G	Gly	GGA	GGC	GGG	GGT		
히스티딘	H	His	CAC	CAT				
이소류신	I	Iso	ATA	ATC	ATT			
라이신	K	Lys	AAA	AAG				
류신	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
메티오닌	M	Met	ATG					
아스파라긴	N	Asn	AAC	AAT				
프롤린	P	Pro	CCA	CCC	CCG	CCT		
글루타민	Q	Gln	CAA	CAG				
아르기닌	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
세린	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT
쓰레오닌	T	Thr	ACA	ACC	ACG	ACT		
발린	V	Val	GTA	GTC	GTG	GTT		
트립토판	W	Trp	TGG					
티로신	Y	Tyr	TAC	TAT				

[0080]

[0081]

본원은 기준 핵산 서열에 비해 코돈 변이체 코돈 서열을 코딩하는 핵산을 포함하는 변이체 핵산 라이브러리로서, 상기 변이체 코돈 서열이 코돈 배정을 기반으로 선택된 것인 변이체 핵산 라이브러리를 제공한다. 예시적 코돈 배정은 표 3에서 볼 수 있는데, 이때 변이체 코돈 서열은 먼저 좌측에서 우측으로 선택된다. 일부 경우, 코돈 배정은 유기체에서의 코돈의 빈도에 기반을 둔다. 예시적 유기체는 동물, 식물, 진균, 원생생물, 고세균 또는 세균을 포함하나 이들로 한정되지 않는다. 예를 들면, 코돈 배정은 에스케리키아 콜라이(*Escherichia coli*) 또는 호모 사피엔스(*Homo sapiens*)에 기반을 둔다.

표 3

코돈 배정

아미노산	1 문자 코드	3 문자 코드	코돈					
알라닌	A	Ala	GCT	GCA	GCC	GCG		
시스테인	C	Cys	TGC	TGT				
아스파르트산	D	Asp	GAT	GAC				
글루탐산	E	Glu	GAG	GAA				
페닐알라닌	F	Phe	TTC	TTT				
글리신	G	Gly	GGT	GGA	GGC	GGG		
히스티딘	H	His	CAC	CAT				
이소류신	I	Iso	ATC	ATT	ATA			
라이신	K	Lys	AAG	AAA				
류신	L	Leu	CTG	CTC	CTT	TTG	TTA	CTA
메티오닌	M	Met	ATG					
아스파라긴	N	Asn	AAC	AAT				
프롤린	P	Pro	CCT	CCA	CCG	CCC		
글루타민	Q	Gln	CAG	CAA				
아르기닌	R	Arg	AGA	CGT	AGG	CGA	CGC	CGG
세린	S	Ser	AGC	TCT	TCC	AGT	TCA	TCG
쓰레오닌	T	Thr	ACC	ACA	ACT	ACG		
발린	V	Val	GTG	GTT	GTC	GTA		
트립토판	W	Trp	TGG					
티로신	Y	Tyr	TAC	TAT				
정지 코돈			TGA	TAA	TAG			

[0082]

[0083]

본원은 기준 핵산 서열에 비해 변이체 코돈 서열을 코딩하는 핵산을 포함하는 변이체 핵산 라이브러리로서, 코돈 배정에 기반을 둔 상기 변이체 코돈 서열이 다양한 요인들에 의해 결정되는 것인 변이체 핵산 라이브러리를 제공한다. 일부 경우, 변이체 코돈 서열은 코돈 서열의 복잡성 또는 다양성을 기반으로 선택된다. 예를 들면, 2개의 상이한 뉴클레오타이드를 포함하는 코돈 서열 또는 동일한 뉴클레오타이드를 포함하는 코돈 서열 대신에 3개의 상이한 뉴클레오타이드를 포함하는 코돈 서열이 선택된다. 일부 경우, 코돈 서열은 다운스트림 적용을 기반으로 선택된다. 다운스트림 적용은 단백질 번역 후 발현 수준에 대한 영향의 최소화 또는 차세대 시퀀싱에 의한 변이체 코돈 서열의 검출의 개선을 포함하나 이들로 한정되지 않는다. 차세대 시퀀싱에 의한 변이체 코돈 서열의 검출의 개선은 높은 오류율을 가진 동중중합체를 피하는 것을 포함할 수 있다. 일부 경우, 코돈 서열이 서열의 파괴를 초래하는 부위, 예컨대, 제한 효소 부위를 야기하지 않는 한, 이 코돈 서열이 선택된다.

[0084]

본원에 기재된 코돈 배정에 기반을 둔 변이체 부위의 코돈 서열은 무작위화될 수 있다. 일부 경우, 코돈 서열은 무작위화되지 않는다. 예를 들면, 펩타이드당 하나의 돌연변이가 선택되는 단일 변이체 라이브러리의 경우, 코돈 서열은 무작위화되지 않는다. 일부 경우, 다수의 변이체 라이브러리들은 무작위화된 코돈 서열을 포함한다.

[0085]

핵산 집단은 다수의 위치들에서 최대 20개의 코돈 변이를 일괄적으로 코딩하는 변이된 핵산을 포함할 수 있다. 이러한 경우, 상기 집단 내의 각각의 핵산은 동일한 핵산의 하나 초과와 위치에서 코돈에 대한 변이를 포함한다. 일부 경우, 상기 집단 내의 각각의 핵산은 단일 핵산의 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9

개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개 또는 20개 이상의 코돈에서 코돈에 대한 변이를 포함한다. 일부 경우, 각각의 변이체 긴 핵산은 단일 긴 핵산의 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개 이상의 코돈에서 코돈에 대한 변이를 포함한다. 일부 경우, 변이체 핵산 집단은 단일 핵산의 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개 또는 30개 이상의 코돈에서 코돈에 대한 변이를 포함한다. 일부 경우, 변이체 핵산 집단은 단일 긴 핵산의 적어도 약 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개, 100개, 125개, 150개, 175개, 200개, 225개, 250개, 275개 또는 300개 이상의 코돈에서 코돈에 대한 변이를 포함한다.

[0086] 본원은 복수의 개별적으로 주소지정가능한 죄위들을 함유하는 제2 클러스터에서 제2 핵산 집단을 생성하는 방법을 제공한다. 제2 핵산 집단은 각각의 코돈 위치에 대해 일정한(즉, 각각의 위치에서 동일한 아미노산을 코딩하는) 복수의 제2 핵산들을 포함할 수 있다. 제2 핵산은 제1 핵산의 적어도 일부와 중첩될 수 있다. 일부 경우, 제2 핵산은 제1 핵산에 표시된 변이체 부위를 함유하지 않는다. 대안적으로, 제2 핵산 집단은 하나 이상의 코돈 위치에 대한 적어도 하나의 변이체를 함유하는 복수의 제2 핵산들을 포함할 수 있다.

[0087] 본원은 다수의 코돈 위치들에서 변이체를 포함하는 단일 핵산 집단을 생성하는, 핵산의 라이브러리를 합성하는 방법을 제공한다. 제1 핵산 집단은 복수의 개별적으로 주소지정가능한 죄위들을 함유하는 제1 클러스터에서 생성될 수 있다. 이러한 경우, 제1 핵산 집단은 상이한 코돈 위치에서 변이체를 포함한다. 일부 경우, 상이한 위치는 연속적이다(즉, 연속적 아미노산을 코딩한다). 예를 들면, 제1 핵산 집단은 2개의 연속 코돈 위치에서 변이체를 포함함으로써, 한 위치에서 최대 19개의 변이체들을 코딩한다. 일부 경우, 제1 핵산 집단은 2개의 연속 코돈 위치에서 변이체를 포함함으로써, 한 위치에서 약 1개의 변이체 내지 약 19개의 변이체들을 코딩한다. 일부 경우, 약 38개의 핵산들이 합성된다. 제1 핵산 집단은 동일한 또는 추가 변이체 위치에서 최대 19개의 코돈 변이체들을 일관적으로 코딩하는 변이된 핵산을 포함할 수 있다. 제1 핵산 집단은 위치 x에서 최대 19개의 변이체들을 함유하고 위치 y에서 최대 19개의 변이체들을 함유하고 위치 z에서 최대 19개의 변이체들을 함유하는 복수의 제1 핵산들을 포함할 수 있다. 이러한 방식에서, 각각의 변이체는 최대 19개의 아미노산 변이체들이 상이한 변이체 부위 각각에서 코딩되도록 상이한 아미노산을 코딩한다. 추가 경우, 제2 핵산 집단은 복수의 개별적으로 주소지정가능한 죄위들을 함유하는 제2 클러스터에서 생성된다. 제2 핵산 집단은 각각의 코돈 위치에 대해 일정한(즉, 각각의 위치에서 동일한 아미노산을 코딩하는) 복수의 제2 핵산들을 포함할 수 있다. 제2 핵산은 제1 핵산의 적어도 일부와 중첩될 수 있다. 제2 핵산은 제1 핵산에 표시된 변이체 부위를 함유하지 않을 수 있다.

[0088] 본원에 기재된 방법에 의해 생성된 변이체 핵산 라이브러리는 변이체 단백질 라이브러리의 생성을 제공한다. 제1 예시적 방식에서, 주형 핵산은 전사되고 번역될 때 단일 원으로 표시된, 다수의 코돈 위치들을 가진 기준 아미노산 서열(도 6a)을 야기하는 서열을 코딩한다. 본원에 기재된 방법을 이용하여 상기 주형의 핵산 변이체를 생성할 수 있다. 일부 경우, 단일 변이체가 핵산에 존재하여, 단일 아미노산 서열을 야기한다(도 6b). 일부 경우, 하나 이상의 코돈에 의해 분리된 하나 초과 변이체들이 핵산에 존재함으로써, 변이체 잔기들 사이의 공간을 가진 단백질이 생성된다(도 6c). 일부 경우, 순차적이고 서로 인접하거나 연속적인 하나 초과 변이체들이 핵산에 존재함으로써, 잔기의 이격된 변이체 스트레치가 생성된다(도 6d). 일부 경우, 순차적이고 인접하거나 연속적인 변이체들을 각각 포함하는 2개의 변이체 스트레치들이 핵산에 존재한다(도 6e).

[0089] 본원은 단일 위치 코돈 변이체를 각각 포함하는 핵산 변이체들의 라이브러리를 생성하는 방법을 제공한다. 한 경우, 주형 핵산은 다수의 코돈 위치들을 갖고, 이때 예시적 아미노산 잔기들은 그들 각각의 1 문자 코드 단백질 코돈과 함께 원으로 표시되어 있다(도 7a). 도 7b는 변이체 핵산의 라이브러리에 의해 코딩된 아미노산 변이체의 라이브러리를 묘사하고, 이때 각각의 변이체는 상이한 단일 부위에 위치하는, "X"로 표시된 단일 위치 변이체를 포함한다. 제1 위치 변이체는 알라닌을 대체하기 위해 임의의 코돈을 갖고, 제2 변이체는 트립토판을 대체하기 위해 변이체 핵산의 라이브러리에 의해 코딩된 임의의 코돈을 갖고, 제3 변이체는 이소류신을 대체하기 위해 임의의 코돈을 갖고, 제4 변이체는 라이신을 대체하기 위해 임의의 코돈을 갖고, 제5 변이체는 아르기닌을 대체하기 위해 임의의 코돈을 갖고, 제6 변이체는 글루탐산을 대체하기 위해 임의의 코돈을 갖고, 제7 변이체는 글루타민을 대체하기 위해 임의의 코돈을 가진다. 모든 코돈 변이체들 또는 더 적은 코돈 변이체들이 변이체 핵산 라이브러리에 의해 코딩될 때, 단백질 발현(즉, 표준 세포 사건인 DNA 전사에 이은 번역 및 프로세싱 사건) 후 아미노산 서열 변이체의 상응하는 집단이 생성된다.

[0090] 일부 방식에서, 단일 위치 변이체의 다수의 부위들을 가진 라이브러리가 생성된다. 도 8a에 묘사된 바와 같이, 야생형 주형이 제공된다. 도 8b는 단일 위치 코돈 변이체의 2개 부위들을 가진 생성된 아미노산 서열을 보여주

는데, 이때 상이한 아미노산을 코딩하는 각각의 코돈 변이체는 상이하게 패턴화된 원으로 표시되어 있다.

[0091] 본원은 다중 부위 단일 위치 변이체의 스트레치를 가진 라이브러리를 생성하는 방법을 제공한다. 핵산의 각각의 스트레치는 1개, 2개, 3개, 4개 또는 5개 이상의 변이체를 가질 수 있다. 핵산의 각각의 스트레치는 적어도 1개의 변이체를 가질 수 있다. 핵산의 각각의 스트레치는 적어도 2개의 변이체를 가질 수 있다. 핵산의 각각의 스트레치는 적어도 3개의 변이체를 가질 수 있다. 예를 들면, 5개의 핵산의 스트레치는 1개의 변이체를 가질 수 있다. 5개의 핵산의 스트레치는 2개의 변이체를 가질 수 있다. 5개의 핵산의 스트레치는 3개의 변이체를 가질 수 있다. 5개의 핵산의 스트레치는 4개의 변이체를 가질 수 있다. 예를 들면, 4개의 핵산의 스트레치는 1개의 변이체를 가질 수 있다. 4개의 핵산의 스트레치는 2개의 변이체를 가질 수 있다. 4개의 핵산의 스트레치는 3개의 변이체를 가질 수 있다. 4개의 핵산의 스트레치는 4개의 변이체를 가질 수 있다.

[0092] 일부 경우, 단일 위치 변이체는 모두 동일한 아미노산, 예를 들면, 히스티딘을 코딩할 수 있다. 도 9a에 표시된 바와 같이, 기준 아미노산 서열이 제공된다. 이 방식에서, 핵산의 스트레치는 단일 위치 변이체의 다수의 부위들을 코딩하고, 발현될 때, 히스티딘을 코딩하는 모든 단일 위치 변이체들을 가진 아미노산 서열을 생성한다(도 9b). 일부 실시양태들에서, 본원에 기재된 방법에 의해 합성된 변이체 라이브러리는 생성된 아미노산 서열에서 4개 초과인 히스티딘 잔기들을 코딩하지 않는다.

[0093] 일부 경우, 본원에 기재된 방법에 의해 생성된 핵산의 변이체 라이브러리는 변이의 분리된 스트레치를 가진 아미노산 서열의 발현을 제공한다. 주형 아미노산 서열은 도 10a에 표시되어 있다. 핵산의 스트레치는 2개의 스트레치에서 1개의 변이체 코돈만을 가질 수 있고, 발현될 때, 도 10b에 표시된 아미노산 서열을 생성할 수 있다. 도 10b에서 변이체는 단일 스트레치 내의 상이한 위치에서 아미노산의 변이를 표시하기 위해 상이하게 패턴화된 원으로 표시되어 있다.

[0094] 본원은 1개, 2개 또는 3개 이상의 코돈 변이체를 가진 핵산 라이브러리를 합성하는 방법 및 디바이스로서, 각각의 부위에 대한 변이체가 선택적으로 조절되는 것인 방법 및 디바이스를 제공한다. 단일 부위 변이체에 대한 2개의 아미노산들의 비는 약 1:100, 1:50, 1:10, 1:5, 1:3, 1:2 또는 1:1일 수 있다. 단일 부위 변이체에 대한 3개의 아미노산들의 비는 약 1:1:100, 1:1:50, 1:1:20, 1:1:10, 1:1:5, 1:1:3, 1:1:2, 1:1:1, 1:10:10, 1:5:5, 1:3:3 또는 1:2:2일 수 있다. 도 11a는 야생형 핵산 서열에 의해 코딩된 야생형 기준 아미노산 서열을 보여준다. 도 11b는 (패턴화된 원에 의해 표시된) 서열의 스트레치를 각각 포함하는 아미노산 변이체들의 라이브러리를 보여주는데, 이때 각각의 위치는 생성된 변이체 단백질 라이브러리에서 특정 비의 아미노산들을 가질 수 있다. 생성된 변이체 단백질 라이브러리는 본원에 기재된 방법에 의해 생성된 변이체 핵산 라이브러리에 의해 코딩된다. 이 예시에서, 5개의 위치가 변이된다: 제1 위치(1100)는 50/50 K/R 비를 갖고, 제2 위치(1110)는 50/25/25 V/L/S 비를 갖고, 제3 위치(1120)는 50/25/25 Y/R/D 비를 갖고, 제4 위치(1130)는 모든 아미노산들에 대해 동등한 비를 갖고, 제5 위치(1140)는 G/P에 대해 75/25 비를 가진다. 본원에 기재된 비는 예일 뿐이다.

[0095] 일부 경우, 궁극적으로 단백질의 아미노산 서열로 번역되는 핵산 서열을 코딩하는 합성된 변이체 라이브러리를 생성한다. 예시적 아미노산 서열은 작은 펩타이드뿐만 아니라 큰 펩타이드, 예를 들면, 항체 서열의 적어도 일부를 코딩하는 아미노산 서열을 포함한다. 일부 경우, 합성된 핵산은 항체 서열의 부분에서 변이체 코돈을 각각 코딩한다. 변이체 합성된 핵산의 부분에 의해 코딩되는 예시적 항체 서열은 이의 항원 결합 또는 가변 영역 또는 이의 단편을 포함한다. 본원에 기재된 핵산에 의해 코딩되는 항체 단편의 예로는 Fab, Fab', F(ab')₂ 및 Fv 단편, 디아바디, 선형 항체, 단일 쇄 항체 분자, 및 항체 단편으로부터 형성된 다중특이적 항체가 있으나 이들로 한정되지 않는다. 본원에 기재된 핵산에 의해 코딩되는 항체 영역의 예로는 Fc 영역, Fab 영역, Fab 영역의 가변 영역, Fab 영역의 불변 영역, 중쇄 또는 경쇄의 가변 도메인(V_H 또는 V_L), 또는 V_H 또는 V_L의 특이적 상보성 결정 영역(CDR)이 있으나 이들로 한정되지 않는다. 본원에 개시된 방법에 의해 생성된 변이체 라이브러리는 본원에 기재된 항체 영역들 중 하나 이상의 항체 영역의 변이를 야기할 수 있다. 한 예시적 과정에서, 여러 CDR들을 코딩하는 핵산의 변이체 라이브러리가 생성된다. 도 12를 참조한다. 본원에 기재된 방법으로 CDR1(1210), CDR2(1220) 및 CDR3(1230) 영역을 가진 항체를 코딩하는 주형 핵산을 변형하고, 이때 각각의 CDR 영역은 변이를 위한 다수의 부위들을 포함한다. 중쇄 또는 경쇄의 단일 가변 도메인 내의 3개의 CDR들(1215, 1225 및 1235) 각각에 대한 변이가 생성된다. 별모양으로 표시된 각각의 부위는 주형 핵산 서열과 상이한 임의의 코돈 서열과 상호교환가능한 단일 위치, 다수의 연속 위치들의 스트레치, 또는 이들 둘 다를 포함할 수 있다. 본원에서 제공된 방법을 이용하여 변이체 라이브러리의 다양성을 최대 약 10¹⁰ 이상의 다양성까지 현저히 증가시킬 수 있다.

[0096] 일부 경우, 변이체 라이브러리는 중쇄 또는 경쇄의 가변 도메인(V_H 또는 V_L)의 단일 또는 다수의 변이체를 포함

한다. 일부 경우, 변이체 라이브러리는 V_H 영역에서 단일 또는 다수의 변이체를 포함한다. 예시적 V_H 영역은 IGHV1, IGHV2, IGHV3, IGHV4, IGHV5, IGHV6 및 IGHV7을 포함하나 이들로 한정되지 않는다. 일부 경우, 변이체 라이브러리는 V_L 영역에서 단일 또는 다수의 변이체를 포함한다. 예시적 V_L 영역은 IGKV1, IGKV2, IGKV3, IGKV4, IGKV5, IGLV1, IGLV2 및 IGLV3을 포함하나 이들로 한정되지 않는다.

[0097] 발현 카세트의 변이

[0098] 일부 경우, 발현 구축물의 부분을 코딩하는 합성된 변이체 라이브러리가 생성된다. 발현 구축물의 예시적 부분은 프로모터, 오픈 리딩 프레임 및 종결 영역을 포함한다. 일부 경우, 발현 구축물은 1개, 2개 또는 3개 이상의 발현 카세트를 코딩한다. 도 14에 묘사된 바와 같이, 발현 구축물 카세트의 부분을 구성하는 단일 부위 또는 다중 부위 분리된 영역에서 코돈 변이를 코딩하는 핵산 라이브러리가 생성될 수 있다. 두 구축물 발현 카세트를 생성하기 위해, 제1 프로모터(1410), 제1 오픈 리딩 프레임(1420), 제1 터미네이터(1430), 제2 프로모터(1440), 제2 오픈 리딩 프레임(1450) 또는 제2 터미네이터 서열(1460)의 변이체 서열의 적어도 일부를 코딩하는 변이체 핵산을 합성하였다. 증폭의 라운드 후, 앞의 예에 기재된 바와 같이, 1,024개의 발현 구축물들의 라이브러리가 생성되었다. 도 14는 단지 하나의 예시적 방식을 제공한다. 일부 경우, 추가 조절제 서열, 예컨대, 번역되지 않는 조절 영역(UTR) 또는 인핸서 영역도 본원에서 언급된 발현 카세트에 포함된다. 발현 카세트는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이상의 성분을 포함할 수 있고, 이 성분들에 대한 변이체 서열은 본원에 기재된 방법에 의해 생성된다. 일부 경우, 발현 구축물은 멀티시스트론성 벡터 내에 하나 초과와 유전자를 포함한다. 한 예에서, 합성된 DNA 핵산을 세포 내로 형질도입하기 위해 바이러스 벡터(예를 들면, 렌티바이러스) 내에 삽입한 후 포장하거나, 세포 내로 전달하기 위해 비-바이러스 벡터 내에 삽입한 후, 스크리닝 및 분석을 수행한다.

[0099] 본원에 개시된 핵산을 삽입하기 위한 발현 벡터는 진핵(예를 들면, 세균 및 진균) 벡터 및 원핵 벡터(예를 들면, 포유동물, 식물 및 곤충 발현 벡터)를 포함한다. 예시적 발현 벡터는 하기 발현 벡터들을 포함하나 이들로 한정되지 않는다: 포유동물 발현 벡터: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His(서열번호 32로서 개시된 "6His"), pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-mCherry-N1 벡터, pEF1a-tdTomato 벡터, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m), 및 pSF-CMV-PURO-NH2-CMYC; 세균 발현 벡터: pSF-OXB20-BetaGal, pSF-OXB20-Fluc, pSF-OXB20, 및 pSF-Tac; 식물 발현 벡터: pRI 101-AN DNA 및 pCambia2301; 및 효모 발현 벡터: pTYB21 및 pKLAC2, 및 곤충 벡터: pAc5.1/V5-His A 및 pDEST8. 예시적 세포는 원핵 세포 및 진핵 세포를 포함하나 이들로 한정되지 않는다. 예시적 진핵 세포는 동물 세포, 식물 세포 및 진균 세포를 포함하나 이들로 한정되지 않는다. 예시적 동물 세포는 곤충 세포, 어류 세포 및 포유동물 세포를 포함하나 이들로 한정되지 않는다. 예시적 포유동물 세포는 마우스 세포, 인간 세포 및 영장류 세포를 포함한다. 본원에 기재된 방법에 의해 합성된 핵산은 형질감염, 형질도입 및 전기천공을 포함하나 이들로 한정되지 않는, 당분야에서 공지되어 있는 다양한 방법들에 의해 세포 내로 전달될 수 있다. 시험되는 예시적 세포 기능은 세포 증식, 이동/부착, 대사 및 세포 신호전달 활성의 변화를 포함하나 이들로 한정되지 않는다.

[0100] 대량 동시적 핵산 합성

[0101] 본원은 혁신적 합성 플랫폼을 생성하기 위해 규소 위의 나노웰 내에서 폴리뉴클레오타이드 합성부터 유전자 조립까지 엔드-투-엔드(end-to-end) 과정의 소형화, 동시화 및 수직 통합을 이용하는 플랫폼 방법을 제공한다. 본원에 기재된 디바이스는 단일 대량 동시화된 실시에서 최대 대략 1,000,000개 이상의 폴리뉴클레오타이드 또는 10,000개 이상의 유전자들을 생성하면서 전통적 합성 방법에 비해 최대 1,000배 이상까지 처리율을 증가시킬 수 있는 규소 합성 플랫폼을 96웰 플레이트와 동일한 풋프린트(footprint)로 제공한다.

[0102] 차세대 시퀀싱의 출현으로 인해, 고해상 게놈 데이터는 정상 생물학 및 질환 발병기전 둘 다에서 다양한 유전자들의 생물학적 역할을 탐구하는 연구를 위한 중요한 요인이 되었다. 이 연구의 핵심에는 분자생물학의 센트럴 도그마(central dogma) 및 "순차적 정보의 잔기에 의한 잔기(residue-by-residue) 전달"이라는 개념이 있다. DNA에 코딩된 유전 정보는 메시지로 전사되고, 그 후 상기 메시지는 주어진 생물학적 경로 내에서 활성 생성물인 단백질로 번역된다.

[0103] 또 다른 흥미로운 연구 분야는 고도 특이적 세포 표적에 초점이 맞춰진 치료 분자의 발견, 개발 및 제조이다. 높은 다양성 DNA 서열 라이브러리는 표적화된 치료제를 위한 개발 파이프라인의 핵심에 있다. 이상적으로는 그의 치료 표적에 대한 높은 친화성을 가진 단백질의 고도 발현을 위해 최적화된 유전자를 이끌어내는 디자인, 구

축 및 시험 단백질 조작 주기에서 단백질을 발현시키기 위해 유전자 돌연변이체를 사용한다. 한 예로서, 수용체의 결합 포켓을 생각한다. 결합 포켓 내의 모든 잔기들의 모든 서열 순열들을 시험하는 능력은 동시적으로 철저한 조사를 가능하게 하여, 성공 기회를 증가시킬 것이다. 연구자가 수용체 내의 특정 부위에서 모든 가능한 돌연변이들을 생성하고자 시도하는 포화 돌연변이유발은 이 개발 과제에 대한 한 방법을 대표한다. 이 방법은 비싸고 시간 및 노동 집약적일지라도, 각각의 변이체로 하여금 각각의 위치 내로 도입될 수 있게 한다. 대조적으로, DNA의 몇몇 선택된 위치 또는 짧은 스트레치가 집중적으로 변형될 수 있는 조합 돌연변이유발은 편향된 표시를 가진 변이체의 불완전한 레퍼토리를 생성한다.

[0104] 약물 개발 파이프라인을 가속화하기 위해, 시험을 위해 이용될 수 있는 우측 위치에서 의도된 빈도로 이용될 수 있는 원하는 변이체를 가진 라이브러리, 다시 말해, 정확한 라이브러리는 비용뿐만 아니라 스크리닝을 위한 처리소요 시간도 감소시킬 수 있다. 본원은 원하는 빈도로 각각의 의도된 변이체의 정확한 도입을 제공하는 핵산 합성 변이체 라이브러리를 합성하는 방법을 제공한다. 최종 사용자에게, 이것은 서열 공간을 철저히 샘플링할 뿐만 아니라 효율적 방식으로 이 가설들을 절의하여 시간 및 스크리닝 시간을 감소시킬 수도 있는 능력으로 해석된다. 게놈 범위 편집은 중요한 경로를 설명할 수 있고, 최적 기능성을 위해 각각의 변이체 및 서열 순열이 시험될 수 있는 라이브러리, 및 수천 개의 유전자들을 사용하여, 전체 경로 및 게놈을 재구축함으로써, 약물 전달을 위한 생물학적 시스템을 재조작할 수 있다.

[0105] 첫 번째 예에서, 본원에 기재된 방법을 이용하여 약물 자체를 최적화할 수 있다. 예를 들면, 항체의 특정된 기능을 개선하기 위해, 상기 항체의 부분을 코딩하는 변이체 핵산 라이브러리를 디자인하고 합성한다. 그 다음, 상기 항체에 대한 변이체 핵산 라이브러리를 본원에 기재된 방법(예를 들면, PCR 돌연변이유발에 이은 벡터 내로의 삽입)으로 생성할 수 있다. 그 다음, 생산 세포주에서 항체를 발현시키고 향상된 활성에 대해 스크리닝한다. 예시적 스크린은 항원에 대한 결합 친화성, 안정성 또는 이펙터 기능(예를 들면, ADCC, 보체 또는 아포토시스)의 조절을 조사하는 것을 포함한다. 항체를 최적화하기 위한 예시적 영역은 Fc 영역, Fab 영역, Fab 영역의 가변 영역, Fab 영역의 불변 영역, 중쇄 또는 경쇄의 가변 도메인(V_H 또는 V_L), 및 V_H 또는 V_L 의 특이적 상보성 결정 영역(CDR)을 포함하나 이들로 한정되지 않는다.

[0106] 대안적으로, 최적화될 분자는 활성화제 또는 경쟁 억제제로서 사용되는 수용체 결합 에피토프이다. 핵산의 변이체 라이브러리의 합성 후, 핵산의 변이체 라이브러리를 벡터 서열 내에 삽입한 후 세포에서 발현시킬 수 있다. 수용체 항원을 세포(예를 들면, 곤충, 포유동물 또는 세균)에서 발현시킨 후 정제할 수 있거나, 세포(예를 들면, 포유동물)에서 발현시켜 서열의 변이로부터 기능적 결과를 조사할 수 있다. 기능적 결과는 단백질 발현, 결합 친화성 및 안정성의 변화를 포함하나 이들로 한정되지 않는다. 세포의 기능적 결과는 생식, 성장, 부착, 사멸, 이동, 에너지 생성, 산소 사용, 대사 활성, 세포 신호전달, 노화, 유리 라디칼 손상에 대한 반응 또는 이들의 임의의 조합의 변화를 포함하나 이들로 한정되지 않는다. 일부 실시양태들에서, 최적화를 위해 선택된 단백질의 유형은 효소, 수송자 단백질, G-단백질과 커플링된 수용체, 전압-게이팅된 이온 채널, 전사 인자, 중합 효소, 어댑터 단백질(2개의 다른 단백질들을 함께 유지하는 데 사용되는, 효소 활성을 갖지 않는 단백질), 및 세포골격 단백질이다. 효소의 예시적 유형은 신호전달 효소들(예컨대, 단백질 키나제(kinase), 단백질 포스파타제(phosphatase), 포스포디에스터라제(phosphodiesterase), 히스톤 데아세틸라제(deacetylase) 및 GTPase)을 포함하나 이들로 한정되지 않는다.

[0107] 본원은 전체 경로 또는 전체 게놈에 포함된 분자에 대한 변이체를 포함하는 변이체 핵산 라이브러리를 제공한다. 예시적 경로는 대사, 세포 사멸, 세포 주기 진행, 면역 세포 활성화, 염증 반응, 혈관신생, 림프발생, 저산소증 및 산화적 스트레스 반응, 또는 세포 부착/이동 경로를 포함하나 이들로 한정되지 않는다. 세포 사멸 경로의 예시적 단백질은 Fas, Cadd, 캐스파제(Caspase) 3, 캐스파제 6, 캐스파제 8, 캐스파제 9, 캐스파제 10, IAP, TNFR1, TNF, TNFR2, NF-kB, TRAFs, ASK, BAD 및 Akt를 포함하나 이들로 한정되지 않는다. 세포 주기 경로의 예시적 단백질은 NFkB, E2F, Rb, p53, p21, 사이클린 A, 사이클린 B, 사이클린 D, 사이클린 E 및 cdc 25를 포함하나 이들로 한정되지 않는다. 세포 이동 경로의 예시적 단백질은 Ras, Raf, PLC, 코필린(cofilin), MEK, ERK, MLP, LIMK, ROCK, RhoA, Src, Rac, 미오신(Myosin) II, ARP2/3, MAPK, PIP2, 인테그린(integrin), 탈린(talin), 킨들린(kindlin), 미그필린(migfilin) 및 필라민(filamin)을 포함하나 이들로 한정되지 않는다.

[0108] 본원에 기재된 방법에 의해 합성된 핵산 라이브러리는 다양한 유형의 세포들에서 발현될 수 있다. 예시적 유형의 세포는 원핵생물(예를 들면, 세균 및 진균) 및 진핵생물(예를 들면, 식물 및 동물)을 포함한다. 예시적 동물은 마우스, 래트, 영장류, 어류 및 곤충을 포함하나 이들로 한정되지 않는다. 예시적 식물은 단자엽식물 및 쌍

자엽식물을 포함하나 이들로 한정되지 않는다. 예시적 식물은 미세조류, 켈프, 시아노박테리아, 녹색 조류, 갈색 조류, 적색 조류, 밀, 담배, 및 옥수수, 쌀, 목화, 야채 및 과일도 포함하나 이들로 한정되지 않는다.

[0109] 본원에 기재된 방법에 의해 합성된 핵산 라이브러리는 질환 상태와 관련된 다양한 세포들에서 발현될 수 있다. 질환 상태와 관련된 세포는 세포주, 조직 샘플, 대상체로부터의 일차 세포, 대상체로부터 증폭된 배양된 세포, 또는 모델 시스템 내의 세포를 포함한다. 예시적 모델 시스템은 질환 상태의 식물 모델 및 동물 모델을 포함하나 이들로 한정되지 않는다.

[0110] 본원에 기재된 방법에 의해 합성된 핵산 라이브러리는 세포 활성의 변화를 평가하기 위해 다양한 유형의 세포들에서 발현될 수 있다. 예시적 세포 활성은 증식, 주기 진행, 세포 사멸, 부착, 이동, 생식, 세포 신호전달, 에너지 생성, 산소 사용, 대사 활성, 및 노화, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하나 이들로 한정되지 않는다.

[0111] 질환 상태의 예방, 감소 또는 치료와 관련된 변이체 분자를 확인하기 위해, 본원에 기재된 변이체 핵산 라이브러리는 질환 상태와 관련된 세포, 또는 질환 상태가 유도될 수 있는 세포에서 발현된다. 일부 경우, 세포에서 질환 상태를 유도하는 물질이 사용된다. 질환 상태 유도를 위한 예시적 수단은 Cre/Lox 재조합 시스템, LPS 염증 유도, 및 저혈당증을 유도하는 스트렙토조토신을 포함하나 이들로 한정되지 않는다. 질환 상태와 관련된 세포는 모델 시스템 또는 배양된 세포로부터의 세포일 수 있을 뿐만 아니라, 특정 질환 상태를 가진 대상체로부터의 세포일 수도 있다. 예시적 질환 상태는 세균, 진균, 바이러스, 자가면역 또는 증식 장애(예를 들면, 암)를 포함한다. 일부 경우, 변이체 핵산 라이브러리는 모델 시스템, 세포주, 또는 대상체로부터 유래한 일차 세포에서 발현되고, 적어도 하나의 세포 활성의 변화에 대해 스크리닝된다. 예시적 세포 활성은 증식, 주기 진행, 세포 사멸, 부착, 이동, 생식, 세포 신호전달, 에너지 생성, 산소 사용, 대사 활성, 및 노화, 유리 라디칼 손상에 대한 반응, 또는 이들의 임의의 조합을 포함하나 이들로 한정되지 않는다.

[0112] **기관**

[0113] 본원은 폴리뉴클레오타이드의 부착 및 합성을 지지하는 복수의 좌위들을 각각 포함하는 복수의 클러스터들을 포함하는 기관을 제공한다. 본원에서 사용된 용어 "좌위"는 표면으로부터 연장하도록 단일 소정의 서열을 코딩하는 폴리뉴클레오타이드를 지지하는 구조물의 분리된 영역을 지칭한다. 일부 경우, 좌위는 2차원 표면, 예를 들면, 실질적으로 평평한 표면 위에 있다. 일부 경우, 좌위는 표면, 예를 들면, 웰, 마이크로웰, 채널 또는 포트 위의 분리된 용기된 또는 함몰된 부위를 지칭한다. 일부 경우, 좌위의 표면은 폴리뉴클레오타이드 합성을 위해 적어도 하나의 뉴클레오타이드를 부착하거나, 바람직하게는 폴리뉴클레오타이드들의 집단의 합성을 위해 동일한 뉴클레오타이드들의 집단을 부착하도록 능동적으로 작용화되는 물질을 포함한다. 일부 경우, 폴리뉴클레오타이드는 동일한 핵산 서열을 코딩하는 폴리뉴클레오타이드들의 집단을 지칭한다. 일부 경우, 디바이스의 표면은 기관의 하나의 표면 또는 복수의 표면들을 포함한다.

[0114] 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 평균 오류율은 종 종 1/1000 미만, 1/1250 미만, 1/1500 미만, 1/2000 미만 또는 1/3000 미만일 수 있다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 평균 오류율은 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000 또는 1/3000 미만이다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 평균 오류율은 1/1000 미만이다.

[0115] 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 총 오류율은 소정의 서열에 비해 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000 또는 1/3000 미만이다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 총 오류율은 1/500, 1/600, 1/700, 1/800, 1/900 또는 1/1000 미만이다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 총 오류율은 소정의 서열에 비해 1/500 미만이다.

[0116] 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드를 위해 오류 보정 효소가 사용될 수 있다. 일부 경우, 오류 보정 시 폴리뉴클레오타이드에 대한 총 오류율은 소정의 서열에 비해 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000 및 1/3000 미만일 수 있다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 오류 보정 시 총 오류율은 1/500, 1/600, 1/700,

1/800, 1/900 또는 1/1000 미만일 수 있다. 일부 경우, 제공된 시스템 및 방법을 이용함으로써 라이브러리 내에서 합성된 폴리뉴클레오타이드에 대한 오류 보정 시 총 오류율은 1/1000 미만일 수 있다.

[0117]

오류율은 유전자 변이체의 라이브러리의 생성을 위한 유전자 합성의 값을 한정할 수 있다. 오류율이 1/300일 때, 1500개 염기 쌍 유전자에서 약 0.7%의 클론이 정확할 것이다. 폴리뉴클레오타이드 합성으로부터의 대부분의 오류가 프레임-시프트 돌연변이를 야기하기 때문에, 이러한 라이브러리에서 99% 이상의 클론이 전체 길이 단백질을 생성하지 않을 것이다. 오류율을 75%까지 감소시키는 것은 정확한 클론의 비율을 40배까지 증가시킬 것이다. 본 개시의 방법 및 조성물은 합성의 개선된 품질, 및 대량 동시적 및 시간 효율적 방식으로 가능해진 오류 보정 방법의 적용가능성으로 인해 통상적으로 관측된 유전자 합성 방법보다 더 낮은 오류율로 큰 핵산 및 유전자 라이브러리의 신속한 드노보 합성을 가능하게 한다. 따라서, 염기 삽입, 결실, 치환, 또는 라이브러리 전체에 걸쳐 또는 라이브러리의 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98% 또는 99.99% 이상에 걸쳐 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000 또는 1/1000000 미만의 총 오류율을 가진 라이브러리를 합성할 수 있다. 본 개시의 방법 및 조성물은 소정의/미리 선택된 서열에 비해 오류 부재 서열을 언급하기 위해 라이브러리의 적어도 서브세트 내의 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98% 또는 99.99% 이상의 폴리뉴클레오타이드들 또는 유전자들과 관련된 낮은 오류율을 가진 큰 합성 핵산 및 유전자 라이브러리도 언급한다. 일부 경우, 상기 라이브러리 내의 단리된 부피 내의 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98% 또는 99.99% 이상의 폴리뉴클레오타이드들 또는 유전자들은 동일한 서열을 가진다. 일부 경우, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8% 또는 99.9% 이상의 유사성 또는 동일성을 가진 관련된 임의의 폴리뉴클레오타이드들 또는 유전자들 중 적어도 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98% 또는 99.99% 이상의 임의의 폴리뉴클레오타이드들 또는 유전자들은 동일한 서열을 가진다. 일부 경우, 폴리뉴클레오타이드 또는 유전자의 특정된 좌위와 관련된 오류율은 최적화된다. 따라서, 큰 라이브러리의 부분으로서 하나 이상의 폴리뉴클레오타이드 또는 유전자의 주어진 좌위 또는 복수의 선택된 좌위들은 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000 또는 1/1000000 미만의 오류율을 각각 가질 수 있다. 다양한 경우, 이러한 오류 최적화된 좌위는 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 60개, 70개, 80개, 90개, 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개, 1000개, 1500개, 2000개, 2500개, 3000개, 4000개, 5000개, 6000개, 7000개, 8000개, 9000개, 10000개, 30000개, 50000개, 75000개, 100000개, 500000개, 1000000개, 2000000개 또는 3000000개 이상의 좌위를 포함할 수 있다. 오류 최적화된 좌위는 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 60개, 70개, 80개, 90개, 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개, 1000개, 1500개, 2000개, 2500개, 3000개, 4000개, 5000개, 6000개, 7000개, 8000개, 9000개, 10000개, 30000개, 75000개, 100000개, 500000개, 1000000개, 2000000개 또는 3000000개 이상의 폴리뉴클레오타이드 또는 유전자에 분포될 수 있다.

[0118]

오류율은 오류 보정에 의해, 또는 오류 보정 없이 달성될 수 있다. 오류율은 라이브러리 전체에 걸쳐, 또는 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.9%, 99.95%, 99.98% 또는 99.99% 이상의 라이브러리에 걸쳐 달성될 수 있다.

[0119]

본원은 통상의 지지체 위의 주소지정가능한 위치에서 상이한 소정의 서열들을 가진 복수의 폴리뉴클레오타이드들의 합성을 지지하는 표면을 포함할 수 있는 구조물을 제공한다. 일부 경우, 디바이스는 2,000개, 5,000개, 10,000개, 20,000개, 30,000개, 50,000개, 75,000개, 100,000개, 200,000개, 300,000개, 400,000개, 500,000개, 600,000개, 700,000개, 800,000개, 900,000개, 1,000,000개, 1,200,000개, 1,400,000개, 1,600,000개, 1,800,000개, 2,000,000개, 2,500,000개, 3,000,000개, 3,500,000개, 4,000,000개, 4,500,000개, 5,000,000개 또는 10,000,000개 이상의 동일하지 않은 폴리뉴클레오타이드들의 합성을 위한 지지체를 제공한다. 일부 경우, 디바이스는 상이한 서열들을 코딩하는 2,000개, 5,000개, 10,000개, 20,000개, 30,000개, 50,000개, 75,000개,

100,000개, 200,000개, 300,000개, 400,000개, 500,000개, 600,000개, 700,000개, 800,000개, 900,000개, 1,000,000개, 1,200,000개, 1,400,000개, 1,600,000개, 1,800,000개, 2,000,000개, 2,500,000개, 3,000,000개, 3,500,000개, 4,000,000개, 4,500,000개, 5,000,000개 또는 10,000,000개 이상의 폴리뉴클레오타이드들의 합성을 위한 지지체를 제공한다. 일부 경우, 상기 폴리뉴클레오타이드들의 적어도 일부는 동일한 서열을 갖거나 동일한 서열을 가지면서 합성되도록 구성된다.

[0120] 본원은 길이가 약 5개, 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개, 100개, 125개, 150개, 175개, 200개, 225개, 250개, 275개, 300개, 325개, 350개, 375개, 400개, 425개, 450개, 475개, 500개, 600개, 700개, 800개, 900개, 1000개, 1100개, 1200개, 1300개, 1400개, 1500개, 1600개, 1700개, 1800개, 1900개 또는 2000개 염기인 폴리뉴클레오타이드를 제조하고 성장시키는 방법 및 디바이스를 제공한다. 일부 경우, 형성된 폴리뉴클레오타이드의 길이는 약 5개, 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개, 100개, 125개, 150개, 175개, 200개 또는 225개 염기이다. 폴리뉴클레오타이드는 길이가 적어도 5개, 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개 또는 100개 염기일 수 있다. 폴리뉴클레오타이드는 길이가 10개 내지 225개 염기, 12개 내지 100개 염기, 20개 내지 150개 염기, 20개 내지 130개 염기, 또는 30개 내지 100개 염기일 수 있다.

[0121] 일부 경우, 폴리뉴클레오타이드는 기관의 상이한 좌위에서 합성되고, 이때 각각의 좌위는 폴리뉴클레오타이드들의 집단의 합성을 지지한다. 일부 경우, 각각의 좌위는 또 다른 좌위에서 성장된 폴리뉴클레오타이드들의 집단과 상이한 서열을 가진 폴리뉴클레오타이드들의 집단의 합성을 지지한다. 일부 경우, 디바이스의 좌위는 복수의 클러스터들 내에 위치한다. 일부 경우, 디바이스는 적어도 10개, 500개, 1000개, 2000개, 3000개, 4000개, 5000개, 6000개, 7000개, 8000개, 9000개, 10000개, 11000개, 12000개, 13000개, 14000개, 15000개, 20000개, 30000개, 40000개 또는 50000개 이상의 클러스터를 포함한다. 일부 경우, 디바이스는 2,000개, 5,000개, 10,000개, 100,000개, 200,000개, 300,000개, 400,000개, 500,000개, 600,000개, 700,000개, 800,000개, 900,000개, 1,000,000개, 1,100,000개, 1,200,000개, 1,300,000개, 1,400,000개, 1,500,000개, 1,600,000개, 1,700,000개, 1,800,000개, 1,900,000개, 2,000,000개, 300,000개, 400,000개, 500,000개, 600,000개, 700,000개, 800,000개, 900,000개, 1,000,000개, 1,200,000개, 1,400,000개, 1,600,000개, 1,800,000개, 2,000,000개, 2,500,000개, 3,000,000개, 3,500,000개, 4,000,000개, 4,500,000개, 5,000,000개 또는 10,000,000개 이상의 상이한 좌위들을 포함한다. 일부 경우, 디바이스는 약 10,000개의 상이한 좌위를 포함한다. 상이한 경우, 단일 클러스터 내의 좌위의 양은 변경된다. 일부 경우, 각각의 클러스터는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개, 100개, 120개, 130개, 150개, 200개, 300개, 400개, 500개 또는 1000개 이상의 좌위를 포함한다. 일부 경우, 각각의 클러스터는 약 50개 내지 500개의 좌위를 포함한다. 일부 경우, 각각의 클러스터는 약 100개 내지 200개의 좌위를 포함한다. 일부 경우, 각각의 클러스터는 약 100개 내지 150개의 좌위를 포함한다. 일부 경우, 각각의 클러스터는 약 109개, 121개, 130개 또는 137개의 좌위를 포함한다. 일부 경우, 각각의 클러스터는 약 19개, 20개, 61개 또는 64개 이상의 좌위를 포함한다.

[0122] 디바이스에서 합성된 상이한 폴리뉴클레오타이드의 수는 기관에서 이용될 수 있는 상이한 좌위의 수에 의해 좌우될 수 있다. 일부 경우, 디바이스의 클러스터 내의 좌위의 밀도는 적어도 또는 약 mm^2 당 1개 좌위, mm^2 당 10개 좌위, mm^2 당 25개 좌위, mm^2 당 50개 좌위, mm^2 당 65개 좌위, mm^2 당 75개 좌위, mm^2 당 100개 좌위, mm^2 당 130개 좌위, mm^2 당 150개 좌위, mm^2 당 175개 좌위, mm^2 당 200개 좌위, mm^2 당 300개 좌위, mm^2 당 400개 좌위, mm^2 당 500개 좌위 또는 mm^2 당 1,000개 좌위 이상이다. 일부 경우, 디바이스는 약 mm^2 당 10개 좌위 내지 약 mm^2 당 500개 좌위, 약 mm^2 당 25개 좌위 내지 약 mm^2 당 400개 좌위, 약 mm^2 당 50개 좌위 내지 약 mm^2 당 500개 좌위, 약 mm^2 당 100개 좌위 내지 약 mm^2 당 500개 좌위, 약 mm^2 당 150개 좌위 내지 약 mm^2 당 500개 좌위, 약 mm^2 당 10개 좌위 내지 약 mm^2 당 250개 좌위, 약 mm^2 당 50개 좌위 내지 약 mm^2 당 250개 좌위, 약 mm^2 당 10개 좌위 내지 약 mm^2 당 200개 좌위, 또는 약 mm^2 당 50개 좌위 내지 약 mm^2 당 200개 좌위를 포함한다. 일부 경우, 한 클러스터 내의 2개의 인접 좌위들의 중심으로부터의 거리는 약 $10\ \mu\text{m}$ 내지 약 $500\ \mu\text{m}$, 약 $10\ \mu\text{m}$ 내지 약 $200\ \mu\text{m}$, 또는 약 $10\ \mu\text{m}$ 내지 약 $100\ \mu\text{m}$ 이다. 일부 경우, 인접 좌위들의 두 중심으로부터의 거리는 약 $10\ \mu\text{m}$, $20\ \mu\text{m}$, $30\ \mu\text{m}$, $40\ \mu\text{m}$, $50\ \mu\text{m}$, $60\ \mu\text{m}$, $70\ \mu\text{m}$, $80\ \mu\text{m}$, $90\ \mu\text{m}$ 또는 $100\ \mu\text{m}$ 를 초과한다. 일부 경우, 2개의 인접 좌위들의 중심으로부터의 거리는 약 $200\ \mu\text{m}$, $150\ \mu\text{m}$, $100\ \mu\text{m}$, $80\ \mu\text{m}$, $70\ \mu\text{m}$, $60\ \mu\text{m}$, $50\ \mu\text{m}$, $40\ \mu\text{m}$, $30\ \mu\text{m}$, $20\ \mu\text{m}$ 또는 $10\ \mu\text{m}$ 미만이다. 일부 경우, 각각의 좌위는 약 $0.5\ \mu\text{m}$, $1\ \mu\text{m}$, $2\ \mu\text{m}$, $3\ \mu\text{m}$, $4\ \mu\text{m}$, $5\ \mu\text{m}$, $6\ \mu\text{m}$, $7\ \mu\text{m}$, $8\ \mu\text{m}$, $9\ \mu\text{m}$, $10\ \mu\text{m}$, $20\ \mu\text{m}$, $30\ \mu\text{m}$,

40 μm , 50 μm , 60 μm , 70 μm , 80 μm , 90 μm 또는 100 μm 의 폭을 가진다. 일부 경우, 각각의 좌위는 약 0.5 μm 내지 100 μm , 약 0.5 μm 내지 50 μm , 약 10 μm 내지 75 μm , 또는 약 0.5 μm 내지 50 μm 의 폭을 가진다.

[0123] 일부 경우, 디바이스 내의 클러스터의 밀도는 적어도 또는 약 100 mm^2 당 1개 클러스터, 10 mm^2 당 1개 클러스터, 5 mm^2 당 1개 클러스터, 4 mm^2 당 1개 클러스터, 3 mm^2 당 1개 클러스터, 2 mm^2 당 1개 클러스터, 1 mm^2 당 1개 클러스터, 1 mm^2 당 2개 클러스터, 1 mm^2 당 3개 클러스터, 1 mm^2 당 4개 클러스터, 1 mm^2 당 5개 클러스터, 1 mm^2 당 10개 클러스터 또는 1 mm^2 당 50개 클러스터 이상이다. 일부 경우, 디바이스는 약 10 mm^2 당 1개 클러스터 내지 약 1 mm^2 당 10개 클러스터를 포함한다. 일부 경우, 2개의 인접 클러스터들의 중심으로부터의 거리는 약 50 μm , 100 μm , 200 μm , 500 μm , 1000 μm , 2000 μm 또는 5000 μm 미만이다. 일부 경우, 2개의 인접 클러스터들의 중심으로부터의 거리는 약 50 μm 내지 약 100 μm , 약 50 μm 내지 약 200 μm , 약 50 μm 내지 약 300 μm , 약 50 μm 내지 약 500 μm , 및 약 100 μm 내지 약 2000 μm 이다. 일부 경우, 2개의 인접 클러스터들의 중심으로부터의 거리는 약 0.05 mm 내지 약 50 mm, 약 0.05 mm 내지 약 10 mm, 약 0.05 mm 내지 약 5 mm, 약 0.05 mm 내지 약 4 mm, 약 0.05 mm 내지 약 3 mm, 약 0.05 mm 내지 약 2 mm, 약 0.1 mm 내지 약 10 mm, 약 0.2 mm 내지 약 10 mm, 약 0.3 mm 내지 약 10 mm, 약 0.4 mm 내지 약 10 mm, 약 0.5 mm 내지 약 10 mm, 약 0.5 mm 내지 약 5 mm, 또는 약 0.5 mm 내지 약 2 mm이다. 일부 경우, 각각의 클러스터는 한 차원을 따라 약 0.5 내지 2 mm, 약 0.5 내지 1 mm, 또는 약 1 내지 2 mm의 직경 또는 폭을 가진다. 일부 경우, 각각의 클러스터는 한 차원을 따라 약 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 또는 2 mm의 직경 또는 폭을 가진다. 일부 경우, 각각의 클러스터는 한 차원을 따라 약 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 또는 2 mm의 내부 직경 또는 폭을 가진다.

[0124] 디바이스는 약 표준 96웰 플레이트의 크기, 예를 들면, 약 100 내지 200 mm의 폭과 약 50 내지 150 mm의 길이를 가질 수 있다. 일부 경우, 디바이스는 약 1000 mm, 500 mm, 450 mm, 400 mm, 300 mm, 250 mm, 200 mm, 150 mm, 100 mm 또는 50 mm 이하의 직경을 가진다. 일부 경우, 디바이스의 직경은 약 25 mm 내지 1000 mm, 약 25 mm 내지 약 800 mm, 약 25 mm 내지 약 600 mm, 약 25 mm 내지 약 500 mm, 약 25 mm 내지 약 400 mm, 약 25 mm 내지 약 300 mm, 또는 약 25 mm 내지 약 200 mm이다. 디바이스 크기의 비한정적 예로는 약 300 mm, 200 mm, 150 mm, 130 mm, 100 mm, 76 mm, 51 mm 및 25 mm가 있다. 일부 경우, 디바이스는 적어도 약 100 mm^2 , 200 mm^2 , 500 mm^2 , 1,000 mm^2 , 2,000 mm^2 , 5,000 mm^2 , 10,000 mm^2 , 12,000 mm^2 , 15,000 mm^2 , 20,000 mm^2 , 30,000 mm^2 , 40,000 mm^2 또는 50,000 mm^2 이상의 평면 표면적을 가진다. 일부 경우, 디바이스의 두께는 약 50 mm 내지 약 2000 mm, 약 50 mm 내지 약 1000 mm, 약 100 mm 내지 약 1000 mm, 약 200 mm 내지 약 1000 mm, 또는 약 250 mm 내지 약 1000 mm이다. 디바이스 두께의 비한정적 예는 275 mm, 375 mm, 525 mm, 625 mm, 675 mm, 725 mm, 775 mm 및 925 mm를 포함한다. 일부 경우, 디바이스의 두께는 직경에 따라 달라지고 기관의 조성에 의해 좌우된다. 예를 들면, 규소 이외의 물질을 포함하는 디바이스는 동일한 직경의 규소 디바이스와 상이한 두께를 가진다. 디바이스 두께는 사용된 물질의 기계적 강도에 의해 결정될 수 있고, 디바이스는 취급 동안 균열 없이 그 자신의 중량을 지지하기에 충분한 두께를 가져야 한다. 일부 경우, 구조물은 본원에 기재된 복수의 디바이스들을 포함한다.

[0125] 표면 물질

[0126] 본원은 소정의 위치에서 폴리뉴클레오타이드 합성을 지지하도록 변형되어 있고 그 결과 낮은 오류율, 낮은 이탈률, 높은 수율 및 높은 올리고 표시를 가진 표면을 포함하는 디바이스를 제공한다. 일부 실시양태들에서, 본원에서 제공된 폴리뉴클레오타이드 합성을 위한 디바이스의 표면은 드노보 폴리뉴클레오타이드 합성 반응을 지지하도록 변형될 수 있는 다양한 물질들로부터 제작된다. 일부 경우, 상기 디바이스는 충분한 전도성을 가진다, 예를 들면, 디바이스의 전체 또는 일부에 걸쳐 균일한 전기장을 형성할 수 있다. 본원에 기재된 디바이스는 유연성 물질을 포함할 수 있다. 예시적 유연성 물질은 변형된 나일론, 변형되지 않은 나일론, 니트로셀룰로스 및 폴리프로필렌을 포함하나 이들로 한정되지 않는다. 본원에 기재된 디바이스는 강성 물질을 포함할 수 있다. 예시적 강성 물질은 유리, 혼연 실리카, 규소, 이산화규소, 질화규소, 플라스틱(예를 들면, 폴리테트라플루오로에틸렌, 폴리프로필렌, 폴리스티렌, 폴리카보네이트 및 이들의 블렌드), 및 금속(예를 들면, 금, 백금)을 포함하나 이들로 한정되지 않는다. 본원에 개시된 디바이스는 규소, 폴리스티렌, 아가로스, 텍스트란, 셀룰로스성 중합체, 폴리아크릴아미드, 폴리디메틸실록산(PDMS), 유리 또는 이들의 임의의 조합을 포함하는 물질로부터 제작될 수 있다. 일부 경우, 본원에 개시된 디바이스는 본원에 나열된 물질 또는 당분야에서 공지되어 있는 임의의 다른 적합한 물질의 병용에 의해 제작된다.

- [0127] 본원에 기재된 예시적 물질에 대한 인장 강도의 목록은 다음과 같이 제공된다: 나일론(70 MPa), 니트로셀룰로스(1.5 MPa), 폴리프로필렌(40 MPa), 규소(268 MPa), 폴리스티렌(40 MPa), 아가로스(1-10 MPa), 폴리아크릴아미드(1-10 MPa), 폴리디메틸실록산(PDMS)(3.9-10.8 MPa). 본원에 기재된 고체 지지체는 1 내지 300, 1 내지 40, 1 내지 10, 1 내지 5, 또는 3 내지 11 MPa의 인장 강도를 가질 수 있다. 본원에 기재된 고체 지지체는 약 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 또는 270 MPa 이상의 인장 강도를 가질 수 있다. 일부 경우, 본원에 기재된 디바이스는 연속 루프 또는 얼레(reel)에 저장될 수 있는 유연성 물질, 예컨대, 테이프 또는 유연성 시트의 형태로 존재하는, 폴리뉴클레오타이드 합성을 위한 고체 지지체를 포함한다.
- [0128] 영물은 하중 하에서 탄성(회복가능한) 변형에 대한 물질의 저항성을 측정한다. 본원에 기재된 예시적 물질의 강도에 대한 영물의 목록은 다음과 같이 제공된다: 나일론(3 GPa), 니트로셀룰로스(1.5 GPa), 폴리프로필렌(2 GPa), 규소(150 GPa), 폴리스티렌(3 GPa), 아가로스(1-10 GPa), 폴리아크릴아미드(1-10 GPa), 폴리디메틸실록산(PDMS)(1-10 GPa). 본원에 기재된 고체 지지체는 1 내지 500, 1 내지 40, 1 내지 10, 1 내지 5, 또는 3 내지 11 GPa의 영물을 가질 수 있다. 본원에 기재된 고체 지지체는 약 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 400 또는 500 GPa 이상의 영물을 가질 수 있다. 유연성과 강성 사이의 관계가 서로 반비례하기 때문에, 유연성 물질은 낮은 영물을 갖고 하중 하에서 그의 형태를 상당히 변화시킨다. 일부 경우, 본원에 기재된 고체 지지체는 적어도 나일론의 유연성을 가진 표면을 가진다.
- [0129] 일부 경우, 본원에 기재된 디바이스는 이산화규소 기부(base) 및 이산화규소의 표면층을 포함한다. 대안적으로, 디바이스는 이산화규소의 기부를 가질 수 있다. 본원에서 제공된 디바이스의 표면은 폴리뉴클레오타이드 합성을 위한 전체 표면적을 증가시키도록 질감을 살릴 수 있다. 본원에 개시된 디바이스는 적어도 5%, 10%, 25%, 50%, 80%, 90%, 95% 또는 99%의 규소를 포함할 수 있다. 본원에 개시된 디바이스는 절연체상 규소(SOI) 웨이퍼로부터 제작될 수 있다.
- [0130] **표면 구성**
- [0131] 본원은 용기된 특징형상 및/또는 함몰된 특징형상을 포함하는 디바이스를 제공한다. 이러한 특징형상들을 가짐으로써 얻는 한 이점은 폴리뉴클레오타이드 합성을 지지하는 표면적의 증가이다. 일부 경우, 용기된 특징형상 및/또는 함몰된 특징형상을 가진 디바이스는 3차원 기관으로서 지칭된다. 일부 경우, 3차원 디바이스는 하나 이상의 채널을 포함한다. 일부 경우, 하나 이상의 좌위는 채널을 포함한다. 일부 경우, 채널은 침착 디바이스, 예컨대, 물질 침착 디바이스를 통한 시약 침착을 위해 이용될 수 있다. 일부 경우, 시약 및/또는 유체는 하나 이상의 채널과 유체를 주고받는 더 큰 웰 내에서 모아진다. 예를 들면, 디바이스는 클러스터와 함께 복수의 좌위들에 상응하는 복수의 채널들을 포함하고, 복수의 채널들은 클러스터의 한 웰과 유체를 주고받는다. 일부 방법들에서, 폴리뉴클레오타이드의 라이브러리는 클러스터의 복수의 좌위들에서 합성된다.
- [0132] 일부 경우, 구조물은 표면에서의 폴리뉴클레오타이드 합성을 위해 유동 및 질량 전달 경로를 제어할 수 있도록 구성된다. 일부 경우, 디바이스의 구성은 폴리뉴클레오타이드 합성 동안 질량 전달 경로, 화학적 노출 시간 및/또는 세척 효능의 제어된 균일한 분포를 가능하게 한다. 일부 경우, 디바이스의 구성은 예를 들면, 성장하는 폴리뉴클레오타이드에 의해 배제된 부피가 폴리뉴클레오타이드의 성장에 이용될 수 있거나 적합한 초기 이용가능한 부피의 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% 또는 1% 이하의 비율보다 더 많이 차지하지 않도록 성장하는 폴리뉴클레오타이드를 위한 충분한 부피를 제공함으로써 스위프(sweep) 효율을 증가시킬 수 있게 한다. 일부 경우, 3차원 구조물은 화학적 노출의 신속한 교환을 가능하게 하도록 유체의 유동을 관리할 수 있게 한다.
- [0133] 본원은 1 fM, 5 fM, 10 fM, 25 fM, 50 fM, 75 fM, 100 fM, 200 fM, 300 fM, 400 fM, 500 fM, 600 fM, 700 fM, 800 fM, 900 fM, 1 pM, 5 pM, 10 pM, 25 pM, 50 pM, 75 pM, 100 pM, 200 pM, 300 pM, 400 pM, 500 pM, 600 pM, 700 pM, 800 pM 또는 900 pM 이상의 양의 DNA를 합성하는 방법을 제공한다. 일부 경우, 폴리뉴클레오타이드 라이브러리는 유전자의 약 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 100%의 길이를 포괄할 수 있다. 유전자는 약 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% 또는 100%까지 변이될 수 있다.
- [0134] 동일하지 않은 폴리뉴클레오타이드들은 유전자의 적어도 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% 또는 100%의 서열을 일괄적으로 코딩할 수 있다. 일부 경우, 폴리뉴클레오타이드는 유전자의 50%, 60%, 70%, 80%, 85%, 90% 또는 95% 이상의 서열을 코딩할 수 있다. 일부 경우, 폴리뉴클레오

타이드는 유전자의 80%, 85%, 90% 또는 95% 이상의 서열을 코딩할 수 있다.

[0135] 일부 경우, 분리는 물리적 구조물에 의해 달성된다. 일부 경우, 분리는 폴리뉴클레오타이드 합성을 위한 능동 영역 및 수동 영역을 생성하는 표면의 상이한 작용화에 의해 달성된다. 상이한 작용화는 디바이스 표면에 걸쳐 소수성을 교대로 엇갈리게 함으로써, 침착된 시약의 비딩(beading) 또는 습윤(wetting)을 야기하는 물 접촉각 효과를 생성함으로써도 달성된다. 보다 더 큰 구조물의 사용은 인접 스폿들의 시약에 의한 상이한 폴리뉴클레오타이드 합성 위치의 더럽혀짐 및 교차오염을 감소시킬 수 있다. 일부 경우, 디바이스, 예컨대, 폴리뉴클레오타이드 합성기를 이용하여 시약을 상이한 폴리뉴클레오타이드 합성 위치에 침착시킨다. 3차원 특징형상을 가진 기관은 낮은 오류율(예를 들면, 약 1:500, 1:1000, 1:1500, 1:2,000, 1:3,000, 1:5,000 또는 1:10,000 미만)로 다수(예를 들면, 약 10,000개 초과)의 폴리뉴클레오타이드들을 합성할 수 있게 하는 방식으로 구성된다. 일부 경우, 디바이스는 약 mm^2 당 1개, 5개, 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 100개, 110개, 120개, 130개, 140개, 150개, 160개, 170개, 180개, 190개, 200개, 300개, 400개 또는 500개 이상의 특징형상의 밀도로 특징형상을 포함한다.

[0136] 디바이스의 웰은 기관의 또 다른 웰과 동일한 또는 상이한 폭, 높이 및/또는 부피를 가질 수 있다. 디바이스의 채널은 기관의 또 다른 채널과 동일한 또는 상이한 폭, 높이 및/또는 부피를 가질 수 있다. 일부 경우, 클러스터의 폭은 약 0.05 mm 내지 약 50 mm, 약 0.05 mm 내지 약 10 mm, 약 0.05 mm 내지 약 5 mm, 약 0.05 mm 내지 약 4 mm, 약 0.05 mm 내지 약 3 mm, 약 0.05 mm 내지 약 2 mm, 약 0.05 mm 내지 약 1 mm, 약 0.05 mm 내지 약 0.5 mm, 약 0.05 mm 내지 약 0.1 mm, 약 0.1 mm 내지 약 10 mm, 약 0.2 mm 내지 약 10 mm, 약 0.3 mm 내지 약 10 mm, 약 0.4 mm 내지 약 10 mm, 약 0.5 mm 내지 약 10 mm, 약 0.5 mm 내지 약 5 mm, 또는 약 0.5 mm 내지 약 2 mm이다. 일부 경우, 클러스터를 포함하는 웰의 폭은 약 0.05 mm 내지 약 50 mm, 약 0.05 mm 내지 약 10 mm, 약 0.05 mm 내지 약 5 mm, 약 0.05 mm 내지 약 4 mm, 약 0.05 mm 내지 약 3 mm, 약 0.05 mm 내지 약 2 mm, 약 0.05 mm 내지 약 1 mm, 약 0.05 mm 내지 약 0.5 mm, 약 0.05 mm 내지 약 0.1 mm, 약 0.1 mm 내지 약 10 mm, 약 0.2 mm 내지 약 10 mm, 약 0.3 mm 내지 약 10 mm, 약 0.4 mm 내지 약 10 mm, 약 0.5 mm 내지 약 10 mm, 약 0.5 mm 내지 약 5 mm, 또는 약 0.5 mm 내지 약 2 mm이다. 일부 경우, 클러스터의 폭은 약 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm 또는 0.05 mm 이하이다. 일부 경우, 클러스터의 폭은 약 1.0 내지 약 1.3 mm이다. 일부 경우, 클러스터의 폭은 약 1.150 mm이다. 일부 경우, 웰의 폭은 약 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm 또는 0.05 mm 이하이다. 일부 경우, 웰의 폭은 약 1.0 내지 1.3 mm이다. 일부 경우, 웰의 폭은 약 1.150 mm이다. 일부 경우, 클러스터의 폭은 약 0.08 mm이다. 일부 경우, 웰의 폭은 약 0.08 mm이다. 클러스터의 폭은 2차원 또는 3차원 기관 내의 클러스터를 지칭할 수 있다.

[0137] 일부 경우, 웰의 높이는 약 20 μm 내지 약 1000 μm , 약 50 μm 내지 약 1000 μm , 약 100 μm 내지 약 1000 μm , 약 200 μm 내지 약 1000 μm , 약 300 μm 내지 약 1000 μm , 약 400 μm 내지 약 1000 μm , 또는 약 500 μm 내지 약 1000 μm 이다. 일부 경우, 웰의 높이는 약 1000 μm 미만, 약 900 μm 미만, 약 800 μm 미만, 약 700 μm 미만 또는 약 600 μm 미만이다.

[0138] 일부 경우, 디바이스는 클러스터 내의 복수의 좌위들에 상응하는 복수의 채널들을 포함하고, 이때 채널의 높이 또는 깊이는 약 5 μm 내지 약 500 μm , 약 5 μm 내지 약 400 μm , 약 5 μm 내지 약 300 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm , 또는 약 10 μm 내지 약 50 μm 이다. 일부 경우, 채널의 높이는 100 μm 미만, 80 μm 미만, 60 μm 미만, 40 μm 미만 또는 20 μm 미만이다.

[0139] 일부 경우, (예를 들면, 실질적으로 평평한 기관에서) 채널, 좌위, 또는 (예를 들면, 좌위가 채널에 상응하는 3차원 디바이스에서) 채널 및 좌위 둘 다의 직경은 약 1 μm 내지 약 1000 μm , 약 1 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 100 μm , 또는 약 10 μm 내지 약 100 μm , 예를 들면, 약 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 이다. 일부 경우, 채널, 좌위, 또는 채널 및 좌위 둘 다의 직경은 약 100 μm , 90 μm , 80 μm , 70 μm , 60 μm , 50 μm , 40 μm , 30 μm , 20 μm 또는 10 μm 미만이다. 일부 경우, 2개의 인접 채널들, 좌위들, 또는 채널들 및 좌위들의 중심으로부터의 거리는 약 1 μm 내지 약 500 μm , 약 1 μm 내지 약 200 μm , 약 1 μm 내지 약 100 μm , 약 5 μm 내지 약 200 μm , 약 5 μm 내지 약 100 μm , 약 5 μm 내지 약 50 μm , 또는 약 5 μm 내지 약 30 μm , 예를 들면, 약 20 μm 이다.

[0140] **표면 변형**

[0141] 다양한 경우, 디바이스 표면, 또는 디바이스 표면의 선택된 부위 또는 영역의 하나 이상의 화학적 성질 및/또는 물리적 성질을 변화시키기 위한 적층 또는 절삭 공정에 의한 표면의 화학적 변경 및/또는 물리적 변경을 위해

표면 변형을 이용한다. 예를 들면, 표면 변형은 (1) 표면의 습윤 성질의 변화, (2) 표면의 작용화, 즉 표면 작용기의 제공, 변형 또는 치환, (3) 표면의 탈작용화, 즉 표면 작용기의 제거, (4) 다른 방식, 예를 들면, 에칭을 통한 표면의 화학조성의 변경, (5) 표면 조도의 증가 또는 감소, (6) 표면에의 코팅, 예를 들면, 표면의 습윤 성질과 상이한 습윤 성질을 나타내는 코팅의 제공, 및/또는 (7) 표면에의 미립자의 침착을 포함하나 이들로 한정되지 않는다.

[0142] 일부 경우, (부착 촉진제로서 지칭되는) 화학적 층을 표면 위에 추가하는 것은 기관의 표면에서 좌위의 구조화된 패턴화를 용이하게 한다. 부착 촉진의 적용을 위한 예시적 표면은 유리, 규소, 이산화규소 및 질화규소를 포함하나 이들로 한정되지 않는다. 일부 경우, 부착 촉진제는 높은 표면 에너지를 가진 화학물질이다. 일부 경우, 제2 화학적 층은 기관의 표면 위에 침착된다. 일부 경우, 제2 화학적 층은 낮은 표면 에너지를 가진다. 일부 경우, 표면 위에 코팅된 화학적 층의 표면 에너지는 소적이 표면 위에 위치하는 것을 지지한다. 선택된 패턴화 방식에 따라, 좌위의 인접성 및/또는 좌위에서의 유체 접촉의 면적은 변경될 수 있다.

[0143] 일부 경우, 예를 들면, 폴리뉴클레오타이드 합성을 위해 폴리뉴클레오타이드 또는 다른 모이어티가 침착되는 디바이스 표면 또는 분리된 좌위는 매끄럽거나 실질적으로 평면이거나(예를 들면, 2차원), 불규칙한 구조, 예를 들면, 용기된 특징형상 또는 함몰된 특징형상(예를 들면, 3차원 특징형상)을 가진다. 일부 경우, 디바이스 표면은 화합물의 하나 이상의 상이한 층에 의해 변형된다. 관심 있는 층의 이러한 변형은 무기 층 및 유기 층, 예컨대, 금속, 산화금속, 중합체, 작은 유기 분자 등을 포함하나 이들로 한정되지 않는다. 비한정적 중합체성 층은 펩타이드, 단백질, 핵산 또는 이들의 모방물질(예를 들면, 펩타이드 핵산 등), 폴리사카라이드, 인지질, 폴리우레탄, 폴리에스테르, 폴리카보네이트, 폴리우레아, 폴리아미드, 폴리에틸렌아민, 폴리아릴렌 설파이드, 폴리실록산, 폴리이미드, 폴리아세테이트, 및 본원에 기재되어 있거나 당분야에서 다른 방식으로 공지되어 있는 임의의 다른 적합한 화합물을 포함한다. 일부 경우, 중합체는 이중중합체이다. 일부 경우, 중합체는 동중중합체이다. 일부 경우, 중합체는 작용성 모이어티를 포함하거나 접합된다.

[0144] 일부 경우, 디바이스의 분리된 좌위는 표면 에너지를 증가시키고/시거나 감소시키는 하나 이상의 모이어티에 의해 작용화된다. 일부 경우, 모이어티는 화학적 불활성 모이어티이다. 일부 경우, 모이어티는 원하는 화학반응, 예를 들면, 폴리뉴클레오타이드 합성 반응에서의 하나 이상의 과정을 지지하도록 구성된다. 표면의 표면 에너지 또는 소수성은 표면에 부착시킬 뉴클레오타이드의 친화성을 결정하는 데 있어서 한 요인이다. 일부 경우, 디바이스를 작용화하는 방법은 (a) 이산화규소를 포함하는 표면을 가진 디바이스를 제공하는 단계; 및 (b) 본원에 기재되어 있거나 당분야에서 다른 방식으로 공지되어 있는 적합한 실란화제, 예를 들면, 유기작용성 알콕시실란 분자를 사용하여 표면을 실란화하는 단계를 포함할 수 있다.

[0145] 일부 경우, 유기작용성 알콕시실란 분자는 디메틸클로로-옥토데실-실란, 메틸디클로로-옥토데실-실란, 트리클로로-옥토데실-실란, 트리메틸-옥토데실-실란, 트리에틸-옥토데실-실란, 또는 이들의 임의의 조합을 포함한다. 일부 경우, 디바이스 표면은 (감마 방사선조사 또는 크로산 산화, 및 하이드록시알킬 표면으로의 환원에 의해 작용화된) 폴리에틸렌/폴리프로필렌, (클로로메틸화에 의해 유도체화되고 벤질아민 작용성 표면으로 아민화된) 고도로 교차결합된 폴리스티렌-디비닐벤젠 또는 나일론(말단 아미노핵심기가 직접적으로 반응함)에 의해 작용화된 표면, 또는 환원된 폴리테트라플루오로에틸렌에 의해 에칭된 표면을 포함한다. 다른 방법 및 작용화제는 전체로서 본원에 참고로 도입된 미국 특허 제5474796호에 기재되어 있다.

[0146] 일부 경우, 디바이스 표면은 실란을 디바이스 표면에 커플링시키기에 효과적인 반응 조건 하에서 전형적으로 디바이스 표면에 존재하는 반응성 친수성 모이어티를 통해 실란의 혼합물을 함유하는 유도체화 조성물과 접촉시킴으로써 작용화된다. 실란화는 일반적으로 유기작용성 알콕시실란 분자를 사용한 자가조립을 통해 표면을 덮는다.

[0147] 예를 들면, 표면 에너지를 낮추거나 증가시키기 위해, 당분야에서 현재 공지되어 있는 다양한 실록산 작용화 시약들도 사용할 수 있다. 유기작용성 알콕시실란은 그의 유기 작용기 따라 분류될 수 있다.

[0148] 본원은 뉴클레오사이드에 커플링될 수 있는 물질의 패턴화를 함유할 수 있는 디바이스를 제공한다. 일부 경우, 디바이스를 활성 물질로 코팅할 수 있다. 일부 경우, 디바이스를 비활성 물질로 코팅할 수 있다. 본원에 기재된 코팅 물질에 포함될 예시적 활성 물질은 N-(3-트리에톡시실릴프로필)-4-하이드록시부티라미드(HAPS), 11-아세톡시운데실트리에톡시실란, n-데실트리에톡시실란, (3-아미노프로필)트리에톡시실란, (3-아미노프로필)트리에톡시실란, 3-글리시독시프로필트리에톡시실란(GOPS), 3-요오도-프로필트리에톡시실란, 부틸-알데하이드-트리에톡시실란, 이량체성 이차 아미노알킬 실록산, (3-아미노프로필)-디에톡시-메틸실란, (3-아미노프로필)-디메틸-에톡시실란, 및 (3-아미노프로필)-트리에톡시실란, (3-글리시독시프로필)-디메틸-에톡시실란, 글리시독시-트리에

톡시실란, (3-머캅토프로필)-트리메톡시실란, 3-4 에폭시사이클로헥실-에틸트리메톡시실란, 및 (3-머캅토프로필)-메틸-디메톡시실란, 알릴 트리클로로로실란, 7-옥트-1-에닐 트리클로로로실란, 또는 비스(3-트리메톡시실틸프로필)아민을 포함하나 이들로 한정되지 않는다.

[0149] 본원에 기재된 코팅 물질에 포함될 예시적 비활성 물질은 퍼플루오로옥틸트리클로로실란; 트리데카플루오로-1,1,2,2-테트라하이드로옥틸)트리클로로실란; 1H,1H,2H,2H-플루오로옥틸트리에톡시실란(FOS); 트리클로로(1H,1H,2H,2H-퍼플루오로옥틸)실란; tert-부틸-[5-플루오로-4-(4,4,5,5-테트라메틸-1,3,2-디옥사보롤란-2-일)인돌-1-일]-디메틸-실란; CYTOP™; Fluorinert™; 퍼플루오로옥틸트리클로로실란(PFOTCS); 퍼플루오로옥틸디메틸클로로실란(PFODCS); 퍼플루오로테실트리에톡시실란(PFDTES); 펜타플루오로페닐-디메틸프로필클로로-실란(PFPTES); 퍼플루오로옥틸트리에톡시실란; 퍼플루오로옥틸트리메톡시실란; 옥틸클로로실란; 디메틸클로로-옥토데실-실란; 메틸디클로로-옥토데실-실란; 트리클로로-옥토데실-실란; 트리메틸-옥토데실-실란; 트리에틸-옥토데실-실란; 또는 옥타데실트리클로로실란을 포함하나 이들로 한정되지 않는다.

[0150] 일부 경우, 작용화제는 탄화수소 실란, 예컨대, 옥타데실트리클로로실란을 포함한다. 일부 경우, 작용화제는 11-아세톡시우데실트리에톡시실란, n-데실트리에톡시실란, (3-아미노프로필)트리메톡시실란, (3-아미노프로필)트리에톡시실란, 글리시딜옥시프로필/트리메톡시실란 및 N-(3-트리에톡시실틸프로필)-4-하이드록시부티라미드를 포함한다.

[0151] 폴리뉴클레오타이드 합성

[0152] 본 개시의 폴리뉴클레오타이드 합성 방법은 포스포라미다이트 화학반응을 포함하는 과정을 포함할 수 있다. 일부 경우, 폴리뉴클레오타이드 합성은 염기를 포스포라미다이트와 커플링시키는 단계를 포함한다. 폴리뉴클레오타이드 합성은 커플링 조건 하에서 포스포라미다이트를 침착시킴으로써 염기를 커플링시키는 단계를 포함할 수 있고, 이때 동일한 염기가 임의적으로 포스포라미다이트에 의해 1회 초과된 빈도로 침착된다(즉, 이중 커플링). 폴리뉴클레오타이드 합성은 반응하지 않은 부위의 캡핑을 포함할 수 있다. 일부 경우, 캡핑은 임의적이다. 폴리뉴클레오타이드 합성은 산화 또는 산화 단계 또는 산화 단계들도 포함할 수 있다. 폴리뉴클레오타이드 합성은 탈블록화, 탈트리틸화 및 황화를 포함할 수 있다. 일부 경우, 폴리뉴클레오타이드 합성은 산화 또는 황화를 포함한다. 일부 경우, 폴리뉴클레오타이드 합성 반응 동안 한 단계 또는 각각의 단계 사이에, 예를 들면, 테트라졸 또는 아세트니트릴을 사용하여 디바이스를 세척한다. 포스포라미다이트 합성 방법에서 어느 한 단계를 위한 시간은 약 2분, 1분, 50초, 40초, 30초, 20초 및 10초 미만일 수 있다.

[0153] 포스포라미다이트 방법을 이용한 폴리뉴클레오타이드 합성은 포스파이트 트리에스테르 결합을 형성하기 위해 포스포라미다이트 구축 블록(예를 들면, 뉴클레오사이드 포스포라미다이트)을 성장하는 폴리뉴클레오타이드 쇠에 후속 추가하는 단계를 포함할 수 있다. 포스포라미다이트 폴리뉴클레오타이드 합성은 3'에서 5' 방향으로 진행한다. 포스포라미다이트 폴리뉴클레오타이드 합성은 합성 주기당 1개의 뉴클레오타이드를 성장하는 폴리뉴클레오타이드 쇠에 추가하는 것을 조절할 수 있게 한다. 일부 경우, 각각의 합성 주기는 커플링 단계를 포함한다. 포스포라미다이트 커플링은 예를 들면, 링커를 통해 활성화된 뉴클레오사이드 포스포라미다이트와 기관에 결합된 뉴클레오사이드 사이의 포스파이트 트리에스테르 결합을 형성하는 단계를 포함한다. 일부 경우, 뉴클레오사이드 포스포라미다이트는 활성화된 디바이스에 제공된다. 일부 경우, 뉴클레오사이드 포스포라미다이트는 활성화된 화제를 가진 디바이스에 제공된다. 일부 경우, 뉴클레오사이드 포스포라미다이트는 기관에 결합된 뉴클레오사이드에 비해 1.5배, 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배, 20배, 25배, 30배, 35배, 40배, 50배, 60배, 70배, 80배, 90배 또는 100배 이상 더 많이 디바이스에 제공된다. 일부 경우, 뉴클레오사이드 포스포라미다이트의 추가를 무수 환경, 예를 들면, 무수 아세트니트릴에서 수행한다. 뉴클레오사이드 포스포라미다이트의 추가 후, 디바이스를 임의적으로 세척한다. 일부 경우, 뉴클레오사이드 포스포라미다이트를 기관에 추가하는 단계들 사이에 임의적으로 세척 단계를 수행하면서 커플링 단계를 추가 1회 이상 반복한다. 일부 경우, 본원에서 이용된 폴리뉴클레오타이드 합성 방법은 1개, 2개 또는 3개 이상의 순차적 커플링 단계를 포함한다. 커플링 전에, 많은 경우, 중합을 방지하는 작용을 하는 보호기를 제거함으로써 디바이스에 결합된 뉴클레오사이드를 탈보호한다. 통상의 보호기는 4,4'-디메톡시트리틸(DMT)이다.

[0154] 커플링 후, 포스포라미다이트 폴리뉴클레오타이드 합성 방법은 임의적으로 캡핑 단계를 포함한다. 캡핑 단계에서, 성장하는 폴리뉴클레오타이드를 캡핑제로 처리한다. 캡핑 단계는 추가 쇠 연장으로부터 커플링 후 기관에 결합된 반응하지 않은 5'-OH 기를 차단하여, 내부 염기 결실을 가진 폴리뉴클레오타이드의 형성을 방지하는 데 유용하다. 추가로, 1H-테트라졸에 의해 활성화된 포스포라미다이트는 구아노신의 O6 위치와 어느 정도 반응할

수 있다. 이론에 의해 구속받고자 하지는 않지만, I_2 /물을 사용한 산화 시, 이 부산물은 아마도 O6-N7 이동을 통해 탈퓨린화를 겪을 수 있다. 탈퓨린 부위는 결국 폴리뉴클레오타이드의 최종 탈보호의 과정에서 절단됨으로써, 전체 길이 생성물의 수율을 감소시킬 수 있다. O6 변형은 I_2 /물을 사용한 산화 전에 캡핑 시약으로 처리함으로써 제거될 수 있다. 일부 경우, 폴리뉴클레오타이드 합성 동안 캡핑 단계의 포함은 캡핑 없는 합성에 비해 오류율을 감소시킨다. 한 예로서, 캡핑 단계는 기관에 결합된 폴리뉴클레오타이드를 아세트산 무수물과 1-메틸이미다졸의 혼합물로 처리하는 것을 포함한다. 캡핑 단계 후, 임의적으로 디바이스를 세척한다.

[0155] 일부 경우, 뉴클레오사이드 포스포라미다이트의 추가 후, 및 임의적으로 캡핑 및 하나 이상의 세척 단계 후, 디바이스에 결합된 성장하는 폴리뉴클레오타이드를 산화시킨다. 산화 단계는 포스포이트 트리에스테르를, 천연 생성 포스포이트 디에스테르 뉴클레오사이드간 결합의 보호된 전구체인 사베위결합된 포스포이트 트리에스테르로 산화시키는 것을 포함한다. 일부 경우, 성장하는 폴리뉴클레오타이드의 산화는 임의적으로 약염기(예를 들면, 피리딘, 루티딘, 콜리딘)의 존재 하에서 요오드 및 물로 처리함으로써 달성된다. 예를 들면, tert-부틸 하이드로퍼옥사이드 또는 (1S)-(+)-(10-캄포르설포닐)-옥사지리딘(CSO)을 사용함으로써 무수 조건 하에서 산화를 수행할 수 있다. 일부 방법들에서, 산화 후 캡핑 단계를 수행한다. 지속할 수 있는 산화로부터의 잔류 물이 후속 커플링을 억제할 수 있기 때문에, 제2 캡핑 단계는 디바이스 건조를 허용한다. 산화 후, 임의적으로 디바이스 및 성장하는 폴리뉴클레오타이드를 세척한다. 일부 경우, 산화 단계를 황화 단계로 치환시켜 폴리뉴클레오타이드 포스포로티오에이트를 획득하는데, 이때 황화 후 임의의 캡핑 단계를 수행할 수 있다. 3-(디메틸아미노메틸렌)아미노-3H-1,2,4-디티아졸-3-티온, DDTT, 보카주(Beaucage) 시약으로서도 공지되어 있는 3H-1,2-벤조디티올-3-온 1,1-디옥사이드, 및 N,N,N',N'-테트라에틸티우람 디설파이드(TETD)를 포함하나 이들로 한정되지 않는 많은 시약들은 효율적 황 전달을 수행할 수 있다.

[0156] 뉴클레오사이드 도입의 후속 주기가 커플링을 통해 일어나기 위해, 일차 하이드록실기가 다음 뉴클레오사이드 포스포라미다이트와 반응하도록 디바이스에 결합된 성장하는 폴리뉴클레오타이드의 보호된 5' 말단을 제거한다. 일부 경우, 보호기는 DMT이고 탈블록화는 디클로로메탄에서 트리클로로아세트산에 의해 일어난다. 탈트리티화를 연장된 시간 동안 수행하거나 권장된 산 용액보다 더 강한 산 용액으로 수행하여 고체 지지체에 결합된 폴리뉴클레오타이드의 탈퓨린화를 증가시킴으로써, 원하는 전체 길이 생성물의 수율을 감소시킬 수 있다. 본원에 기재된 본 개시의 방법 및 조성물은 원치 않는 탈퓨린화 반응을 제한하는 제어된 탈블록화 조건을 제공한다. 일부 경우, 디바이스에 결합된 폴리뉴클레오타이드를 탈블록화 후 세척한다. 일부 경우, 탈블록화 후 효율적 세척은 낮은 오류율을 가진 합성된 폴리뉴클레오타이드에 기여한다.

[0157] 폴리뉴클레오타이드를 합성하는 방법은 전형적으로 반복되는 일련의 하기 단계들을 포함한다: 보호된 단량체를 능동적으로 작용화된 표면(예를 들면, 좌위)에 적용하여 활성화된 표면, 링커 또는 미리 탈보호된 단량체와 연결하는 단계; 적용된 단량체가 후속 적용된 보호된 단량체와 반응하도록 탈보호하는 단계; 및 연결을 위해 또 다른 보호된 단량체를 적용하는 단계. 하나 이상의 중간 단계는 산화 또는 황화를 포함한다. 일부 경우, 하나 이상의 세척 단계는 한 단계 또는 모든 단계들 전에 또는 후에 수행된다.

[0158] 포스포라미다이트 기반 폴리뉴클레오타이드 합성 방법은 일련의 화학적 단계들을 포함한다. 일부 경우, 합성 방법의 하나 이상의 단계는 시약 순환을 포함하고, 이때 상기 방법의 하나 이상의 단계는 단계에 유용한 시약을 디바이스에 적용하는 것을 포함한다. 예를 들면, 시약은 일련의 액체 침착 및 진공 건조 단계에 의해 순환된다. 3차원 특징형상, 예컨대, 웰, 마이크로웰, 채널 등을 포함하는 기관의 경우, 임의적으로 웰 및/또는 채널을 통해 시약을 디바이스의 하나 이상의 영역에 통과시킨다.

[0159] 본원에 기재된 방법 및 시스템은 폴리뉴클레오타이드를 합성하기 위한 폴리뉴클레오타이드 합성 디바이스와 관련된다. 합성은 동시적일 수 있다. 예를 들면, 적어도 또는 약 적어도 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 30개, 35개, 40개, 45개, 50개, 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 1000개, 10000개, 50000개, 75000개 또는 100000개 이상의 폴리뉴클레오타이드들을 동시에 합성할 수 있다. 동시에 합성될 수 있는 폴리뉴클레오타이드의 총 수는 2개 내지 100000개, 3개 내지 50000개, 4개 내지 10000개, 5개 내지 1000개, 6개 내지 900개, 7개 내지 850개, 8개 내지 800개, 9개 내지 750개, 10개 내지 700개, 11개 내지 650개, 12개 내지 600개, 13개 내지 550개, 14개 내지 500개, 15개 내지 450개, 16개 내지 400개, 17개 내지 350개, 18개 내지 300개, 19개 내지 250개, 20개 내지 200개, 21개 내지 150개, 22개 내지 100개, 23개 내지 50개, 24개 내지 45개, 25개 내지 40개, 또는 30개 내지 35개일 수 있다. 당분야에서 숙련된 자는 동시적으로 합성된 폴리뉴클레오타이드의 총 수가 이 값들 중 임

의의 값에 의해 한정된 임의의 범위, 예를 들면, 25개 내지 100개의 범위 내에 속할 수 있다는 것을 인식한다. 동시에 합성된 폴리뉴클레오타이드의 총 수는 범위의 중점으로서 사용되는 값들 중 임의의 값에 의해 한정된 임의의 범위 내에 속할 수 있다. 디바이스 내에서 합성된 폴리뉴클레오타이드의 총 몰 질량 또는 각각의 폴리뉴클레오타이드의 몰 질량은 적어도 또는 적어도 약 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000 또는 100000 피코몰 이상일 수 있다. 디바이스 내의 각각의 폴리뉴클레오타이드의 길이 또는 폴리뉴클레오타이드의 평균 길이는 적어도 또는 약 적어도 10개, 15개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 100개, 150개, 200개, 300개, 400개 또는 500개 이상의 뉴클레오타이드일 수 있다. 디바이스 내의 각각의 폴리뉴클레오타이드의 길이 또는 폴리뉴클레오타이드의 평균 길이는 기껏해야 또는 약 기껏해야 500개, 400개, 300개, 200개, 150개, 100개, 50개, 45개, 35개, 30개, 25개, 20개, 19개, 18개, 17개, 16개, 15개, 14개, 13개, 12개, 11개 또는 10개 이하의 뉴클레오타이드일 수 있다. 디바이스 내의 각각의 폴리뉴클레오타이드의 길이 또는 폴리뉴클레오타이드의 평균 길이는 10개 내지 500개, 9개 내지 400개, 11개 내지 300개, 12개 내지 200개, 13개 내지 150개, 14개 내지 100개, 15개 내지 50개, 16개 내지 45개, 17개 내지 40개, 18개 내지 35개, 또는 19개 내지 25개의 뉴클레오타이드일 수 있다. 당분야에서 숙련된 자는 디바이스 내의 각각의 폴리뉴클레오타이드의 길이 또는 폴리뉴클레오타이드의 평균 길이가 이 값들 중 임의의 값에 의해 한정된 임의의 범위, 예를 들면, 100개 내지 300개의 뉴클레오타이드일 수 있다는 것을 인식한다. 디바이스 내의 각각의 폴리뉴클레오타이드의 길이 또는 폴리뉴클레오타이드의 평균 길이는 범위의 중점으로서 사용되는 값들 중 임의의 값에 의해 한정된 임의의 범위 내에 속할 수 있다.

[0160] 본원에서 제공된 표면에서 폴리뉴클레오타이드를 합성하는 방법은 빠른 속도로 합성할 수 있게 한다. 한 예로서, 시간당 적어도 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 35개, 40개, 45개, 50개, 55개, 60개, 70개, 80개, 90개, 100개, 125개, 150개, 175개 또는 200개 이상의 뉴클레오타이드가 합성된다. 뉴클레오타이드는 아데닌, 구아닌, 타이민, 사이토신, 유리딘 구축 블록, 또는 이들의 유사체/변형된 버전을 포함한다. 일부 경우, 폴리뉴클레오타이드의 라이브러리는 기관에서 동시에 합성된다. 예를 들면, 약 또는 적어도 약 100개, 1,000개, 10,000개, 30,000개, 75,000개, 100,000개, 1,000,000개, 2,000,000개, 3,000,000개, 4,000,000개 또는 5,000,000개의 분리된 좌위들을 포함하는 디바이스는 적어도 동일한 수의 상이한 폴리뉴클레오타이드들의 합성을 지지할 수 있고, 이때 상이한 서열을 코딩하는 폴리뉴클레오타이드가 분리된 좌위에서 합성된다. 일부 경우, 폴리뉴클레오타이드의 라이브러리는 약 3개월, 2개월, 1개월, 3주, 15일, 14일, 13일, 12일, 11일, 10일, 9일, 8일, 7일, 6일, 5일, 4일, 3일, 2일 또는 24시간 이하의 시간 이내에 본원에 기재된 낮은 오류율로 디바이스에서 합성된다. 일부 경우, 본원에 기재된 기관 및 방법을 이용함으로써 낮은 오류율로 합성된 폴리뉴클레오타이드 라이브러리로부터 조립된 더 큰 핵산은 약 3개월, 2개월, 1개월, 3주, 15일, 14일, 13일, 12일, 11일, 10일, 9일, 8일, 7일, 6일, 5일, 4일, 3일, 2일 또는 24시간 이하의 시간 이내에 제조된다.

[0161] 일부 경우, 본원에 기재된 방법은 복수의 코돈 부위들에서 상이한 변이체 핵산을 포함하는 핵산의 라이브러리의 생성을 제공한다. 일부 경우, 핵산은 1개 부위, 2개 부위, 3개 부위, 4개 부위, 5개 부위, 6개 부위, 7개 부위, 8개 부위, 9개 부위, 10개 부위, 11개 부위, 12개 부위, 13개 부위, 14개 부위, 15개 부위, 16개 부위, 17개 부위, 18개 부위, 19개 부위, 20개 부위, 30개 부위, 40개 부위 또는 50개 부위 이상의 변이체 코돈 부위를 가질 수 있다.

[0162] 일부 경우, 변이체 코돈 부위들 중 하나 이상의 부위는 인접할 수 있다. 일부 경우, 변이체 코돈 부위들 중 하나 이상의 부위는 인접하지 않을 수 있고 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개 또는 10개 이상의 코돈에 의해 분리될 수 있다.

[0163] 일부 경우, 핵산은 변이체 코돈 부위들 중 다수의 부위들을 포함할 수 있고, 이때 모든 변이체 코돈 부위들은 서로 인접하여, 변이체 코돈 부위의 스트레치를 형성한다. 일부 경우, 핵산은 변이체 코돈 부위들 중 다수의 부위들을 포함할 수 있고, 이때 변이체 코돈 부위들 중 어느 부위도 서로 인접하지 않는다. 일부 경우, 핵산은 변이체 코돈 부위들 중 다수의 부위들을 포함할 수 있고, 이때 변이체 코돈 부위들 중 일부는 서로 인접하여, 변이체 코돈 부위의 스트레치를 형성하고, 변이체 코돈 부위들 중 일부는 서로 인접하지 않는다.

[0164] 도면을 참조하면, 도 15는 더 짧은 폴리뉴클레오타이드로부터 핵산(예를 들면, 유전자)을 합성하는 예시적 과정 워크플로우를 보여준다. 상기 워크플로우는 일반적으로 하기 단계들로 나뉘어진다: (1) 단일 가닥 폴리뉴클레오타이드 핵산 라이브러리의 드노보 합성, (2) 더 큰 단편을 형성하기 위한 폴리뉴클레오타이드들의 연결, (3) 오류 보정, (4) 품질 관리, 및 (5) 운반. 드노보 합성 전에, 의도된 핵산 서열 또는 핵산 서열의 군을 미리

선택한다. 예를 들면, 생성할 유전자의 군을 미리 선택한다.

- [0165] 일단 생성할 큰 핵산이 선택되면, 드노보 합성을 위해 폴리뉴클레오타이드의 예정된 라이브러리를 디자인한다. 고밀도 폴리뉴클레오타이드 어레이를 생성하는 다양한 적합한 방법들이 공지되어 있다. 워크플로우 예에서, 디바이스 표면 층(1501)이 제공된다. 상기 예에서, 폴리뉴클레오타이드 합성 과정을 개선하기 위해 표면의 화학반응을 변경시킨다. 액체를 밀어내기 위해 낮은 표면 에너지의 영역을 생성하는 반면, 액체를 끌어당기기 위해 높은 표면 에너지의 영역을 생성한다. 표면 그 자체는 평평한 표면의 형태로 존재할 수 있거나 형태의 변경, 예컨대, 표면적을 증가시키는 돌출부 또는 마이크로웰을 함유할 수 있다. 상기 워크플로우 예에서, 전체로서 본원에 참고로 도입된 국제 특허출원 공보 제WO/2015/021080호에 개시된 바와 같이, 선택된 높은 표면 에너지 분자는 DNA 화학반응을 지지하는 이중 작용을 한다.
- [0166] 폴리뉴클레오타이드 어레이의 제자리 제조는 고체 지지체에서 일어나고 단일 뉴클레오타이드 연장 과정을 이용하여 다수의 올리고머들을 동시에 연장한다. 침착 디바이스, 예컨대, 물질 침착 디바이스는 단계적 방식으로 시약을 방출함으로써, 다수의 폴리뉴클레오타이드들이 동시에 한 번에 1개의 잔기를 연장하여 소정의 핵산 서열을 가진 올리고머를 생성하도록 디자인된다(1502). 일부 경우, 폴리뉴클레오타이드는 이 단계에서 표면으로부터 절단된다. 절단은 예를 들면, 암모니아 또는 메틸아민을 사용한 기체 절단을 포함한다.
- [0167] 생성된 폴리뉴클레오타이드 라이브러리는 반응 챔버 내에 배치된다. 이 예시적 워크플로우에서, 반응 챔버("나노반응기"로서도 지칭됨)는 PCR 시약을 함유하고 폴리뉴클레오타이드 라이브러리에서 함유된 실리콘 코팅된 웰이다(1503). 폴리뉴클레오타이드의 밀봉(1504) 전 또는 후에, 시약을 첨가하여 기관으로부터 폴리뉴클레오타이드를 방출한다. 예시적 워크플로우에서, 나노반응기의 밀봉 후 폴리뉴클레오타이드를 방출한다(1505). 일단 방출되면, 단일 가닥 폴리뉴클레오타이드의 단편은 DNA의 전체 긴 범위 서열을 포괄하도록 하이브리드화한다. 각각의 합성된 폴리뉴클레오타이드가 집단 내의 적어도 하나의 다른 폴리뉴클레오타이드와 중첩되는 작은 부분을 갖도록 디자인되기 때문에 부분적 하이브리드화(1505)가 가능하다.
- [0168] 하이브리드화 후, PCA 반응을 시작한다. 중합효소 주기 동안, 폴리뉴클레오타이드는 상보적 단편에 어닐링하고 갭은 중합효소에 의해 채워진다. 각각의 주기는 어느 폴리뉴클레오타이드가 서로 발견되는 지에 따라 다양한 단편들의 길이를 무작위로 증가시킨다. 단편들 사이의 상보성은 완전한 큰 폭의 이중 가닥 DNA를 형성할 수 있게 한다(1506).
- [0169] PCA가 완료된 후, 나노반응기를 디바이스로부터 분리하고(1507), PCR을 위한 프라이머를 가진 디바이스와 상호작용하도록 위치시킨다(1508). 밀봉 후, 나노반응기에 대해 PCR을 수행하고(1509), 더 큰 핵산을 증폭시킨다. PCR 후(1510), 나노챔버를 개방하고(1511), 오류 보정 시약을 첨가하고(1512), 챔버를 밀봉하고(1513), 이중 가닥 PCR 증폭 생성물로부터 상보성이 좋지 않은 불일치된 염기 쌍 및/또는 가닥을 제거하도록 오류 보정 반응이 일어난다(1514). 나노반응기를 개방하고 분리한다(1515). 그 다음, 오류 보정된 생성물에 대해 추가 프로세싱 단계, 예컨대, PCR 및 분자 바 코딩을 수행한 후, 운반(1523)을 위해 포장한다(1522).
- [0170] 일부 경우, 품질 관리 측정을 수행한다. 오류 보정 후, 품질 관리 단계는 예를 들면, 오류 보정된 생성물의 증폭을 위해 시퀀싱 프라이머를 가진 웨이퍼와 상호작용시키는 단계(1516), 오류 보정된 증폭 생성물을 함유하는 챔버에 웨이퍼를 밀봉시키는 단계(1517), 및 추가 라운드의 증폭을 수행하는 단계(1518)를 포함한다. 나노반응기를 개방하고(1519), 생성물을 풀링하고(1520) 시퀀싱한다(1521). 허용가능한 품질 관리 측정을 수행한 후, 포장된 생성물(1522)은 운반(1523)을 위해 승인받는다.
- [0171] 일부 경우, 본원에 개시된 중첩 프라이머를 사용한 돌연변이유발을 워크플로우, 예컨대, 도 15의 워크플로우에 의해 생성된 폴리뉴클레오타이드에 대해 수행한다. 일부 경우, 프라이머의 라이브러리는 고체 지지체에서의 제자리 제조에 의해 생성되고 단일 뉴클레오타이드 연장 과정을 이용하여 다수의 올리고머들을 동시에 연장한다. 침착 디바이스, 예컨대, 물질 침착 디바이스는 단계적 방식으로 시약을 방출함으로써, 다수의 폴리뉴클레오타이드들이 동시에 한 번에 1개의 잔기를 연장하여 소정의 핵산 서열을 가진 올리고머를 생성하도록 디자인된다(1502).
- [0172] **컴퓨터 시스템**
- [0173] 본원에 기재된 시스템들 중 임의의 시스템은 컴퓨터에 작동가능하게 연결될 수 있고, 근거리 또는 원거리에서 컴퓨터를 통해 자동화될 수 있다. 다양한 경우, 본 개시의 방법 및 시스템은 컴퓨터 시스템의 소프트웨어 프로그램 및 이의 사용도 포함할 수 있다. 따라서, 분배/진공/재충전 기능의 동시화를 위한 전산화된 제어, 예컨대, 물질 침착 디바이스 이동, 분배 작용 및 진공 구동의 조화 및 동시화는 본 개시의 범위 내에 있다. 컴퓨터 시스

템은 사용자 특정된 염기 서열과 물질 침착 디바이스의 위치 사이에 접속하여 정확한 시약을 기관의 특정된 영역에 전달하도록 프로그래밍될 수 있다.

- [0174] 도 16에 예시된 컴퓨터 시스템(1600)은 임의적으로 고정된 매체(1612)를 가진 서버(1609)에 연결될 수 있는 매체(1611) 및/또는 네트워크 포트(1605)로부터 지시를 관독할 수 있는 논리 장치로서 이해될 수 있다. 예컨대, 도 16에 표시된 시스템은 CPU(1601), 디스크 드라이브(1603), 임의적 입력 디바이스, 예컨대, 키보드(1615) 및/또는 마우스(1616) 및 임의적 모니터(1607)를 포함할 수 있다. 서버로의 데이터 통신은 근거리 또는 원거리 위치에서 표시된 통신 매체를 통해 달성될 수 있다. 통신 매체는 데이터를 전송하고/하거나 제공받는 임의의 수단을 포함할 수 있다. 예를 들면, 통신 매체는 네트워크 연결, 무선 연결 또는 인터넷 연결일 수 있다. 이러한 연결은 월드 와이드 웹 상에서 통신을 제공할 수 있다. 본 개시와 관련된 데이터는 도 16에 예시된 바와 같이 당사자(1622)에 의한 수용 및/또는 검토를 위해 이러한 네트워크 또는 연결을 통해 전송될 수 있다는 것이 예측된다.
- [0175] 도 17은 본 개시의 예시적 경우와 관련하여 사용될 수 있는 컴퓨터 시스템(1700)의 첫 번째 예시적 구성을 보여주는 블록 도표이다. 도 17에 묘사된 바와 같이, 예시적 컴퓨터 시스템은 지시를 프로세싱하기 위한 프로세서(1702)를 포함할 수 있다. 프로세서의 비한정적 예로는 인텔(Intel) Xeon™ 프로세서, AMD Opteron™ 프로세서, 삼성(Samsung) 32-bit RISC ARM 1176JZ(F)-S v1.0™ 프로세서, ARM Cortex-A8 삼성 S5PC100™ 프로세서, ARM Cortex-A8 애플(Apple) A4™ 프로세서, 마블(Marvell) PXA 930™ 프로세서, 또는 기능적으로 동등한 프로세서가 있다. 병렬 프로세싱을 위해 다수의 실행 스레드(thread)를 사용할 수 있다. 일부 경우, 단일 컴퓨터 시스템 내에 있든, 클러스터 내에 있든, 아니면 다수의 컴퓨터들, 휴대전화들 및/또는 개인 데이터 단말기 디바이스들을 포함하는 네트워크 상에서 시스템 전체에 걸쳐 분포되어 있든 관계없이 다수의 프로세서들 또는 다수의 코어들을 가진 프로세서도 사용할 수 있다.
- [0176] 도 17에 예시된 바와 같이, 고속 캐시(1704)를 프로세서(1702)에 연결하거나 도입하여, 프로세서(1702)에 의해 최근에 사용되었거나 종종 사용되는 지시 또는 데이터에 대한 고속 메모리를 제공할 수 있다. 프로세서(1702)는 프로세서 버스(1708)에 의해 노쓰 브리지(1706)에 연결된다. 노쓰 브리지(1706)는 메모리 버스(1712)에 의해 무작위 접속 메모리(RAM)(1710)에 연결되고 프로세서(1702)에 의해 RAM(1710)에의 접속을 관리한다. 노쓰 브리지(1706)는 칩세트 버스(1716)에 의해 사우스 브리지(1714)에도 연결된다. 그 다음, 사우스 브리지(1714)는 주변 버스(1718)에 연결된다. 주변 버스는 예를 들면, PCI, PCI-X, PCI Express 또는 다른 주변 버스일 수 있다. 노쓰 브리지 및 사우스 브리지는 종종 프로세서 칩셋으로서 지칭되고 프로세서, RAM, 및 주변 버스(1718)의 주변 부품 사이의 데이터 전달을 관리한다. 일부 대안적 구성에서, 별도의 노쓰 브리지 칩을 사용하는 대신에 노쓰 브리지의 기능을 프로세서 내에 도입할 수 있다. 일부 경우, 시스템(1700)은 주변 버스(1718)에 부착된 액셀러레이터 카드(1722)를 포함할 수 있다. 액셀러레이터는 특정 프로세싱을 가속화하기 위해 필드 프로그래밍 가능 게이트 어레이(FPGA) 또는 다른 하드웨어를 포함할 수 있다. 예를 들면, 액셀러레이터는 적응 데이터 재구조화, 또는 확장된 세트 프로세싱에서 사용된 대수식의 평가를 위해 사용될 수 있다.
- [0177] 소프트웨어 및 데이터는 외부 저장장치(1724)에 저장되고 프로세서에 의한 사용을 위해 RAM(1710) 및/또는 캐시(1704) 내에 적재될 수 있다. 시스템(1700)은 시스템 자원을 관리하기 위한 운영 시스템을 포함하고, 운영 시스템의 비한정적 예로는 Linux, Windows™, MACOS™, BlackBerry OST™, iOS™, 및 다른 기능적으로 동등한 운영 시스템뿐만 아니라, 본 개시의 예시적 경우에 따라 데이터 저장 및 최적화를 관리하기 위해 운영 시스템의 상부에서 실행되는 응용 소프트웨어도 있다. 이 예에서, 시스템(1700)은 네트워크 인터페이스를 외부 저장장치, 예컨대, 네트워크 부착된 저장장치(NAS), 및 분포된 병렬 프로세싱을 위해 사용될 수 있는 다른 컴퓨터 시스템에게 제공하기 위한 주변 버스에 연결된 네트워크 인터페이스 카드(NIC)(1720 및 1721)도 포함한다.
- [0178] 도 18은 복수의 컴퓨터 시스템들(1802a 및 1802b), 복수의 휴대전화들 및 개인 데이터 단말기들(1802c), 및 네트워크 부착된 저장장치(NAS)(1804a 및 1804b)를 가진 네트워크(1800)를 보여주는 도표이다. 일부 경우, 시스템(1802a, 1802b 및 1802c)은 데이터 저장을 관리할 수 있고 네트워크 부착된 저장장치(NAS)(1804a 및 1804b)에 저장된 데이터에 대한 데이터 접속을 최적화할 수 있다. 데이터를 위해 수학적 모델을 사용할 수 있고 컴퓨터 시스템(1802a 및 1802b), 및 휴대전화 및 개인 데이터 단말기 시스템(1802c) 전체에 걸쳐 분포된 병렬 프로세싱을 이용하여 평가할 수 있다. 컴퓨터 시스템(1802a 및 1802b), 및 휴대전화 및 개인 데이터 단말기 시스템(1802c)은 네트워크 부착된 저장장치(NAS)(1804a 및 1804b)에 저장된 데이터의 적응 데이터 재구조화를 위한 병렬 프로세싱도 제공할 수 있다. 도 18은 한 예만을 보여주었고, 본 개시의 다양한 경우와 관련하여 매우 다양한 다른 컴퓨터 구성 및 시스템들이 사용될 수 있다. 예를 들면, 블레이드 서버를 이용하여 병렬 프로세싱을 제공할 수 있다. 뒤판을 통해 프로세서 블레이드를 연결하여 병렬 프로세싱을 제공할 수 있다. 별도의 네트워크

인터페이스를 통해 저장장치도 뒤판에 연결할 수 있거나 네트워크 부착된 저장장치(NAS)로서 연결할 수 있다. 일부 예시적 경우, 프로세서는 별도의 메모리 공간을 유지할 수 있고 다른 프로세서에 의한 병렬 프로세싱을 위해 네트워크 인터페이스, 뒤판 또는 다른 연결기를 통해 데이터를 전송할 수 있다. 다른 경우, 일부 또는 모든 프로세서들이 공유된 가상 주소 메모리 공간을 사용할 수 있다.

[0179] 도 19는 예시적 경우에 따라 공유된 가상 주소 메모리 공간을 사용하는 멀티프로세서 컴퓨터 시스템(1900)의 블록 도표이다. 상기 시스템은 공유된 메모리 서브시스템(1904)에 접속할 수 있는 복수의 프로세서들(1902a-f)을 포함한다. 상기 시스템은 메모리 서브시스템(1904) 내에 복수의 프로그래밍가능한 하드웨어 메모리 알고리즘 프로세서들(MAP)(1906a-f)을 포함한다. 각각의 MAP(1906a-f)는 메모리(1908a-f) 및 하나 이상의 필드 프로그래밍가능한 게이트 어레이(FPGA)(1910a-f)를 포함할 수 있다. MAP는 환경 설정될 수 있는 기능적 유닛을 제공하고, 특정 알고리즘 또는 알고리즘의 부분이 각각의 프로세서에 의해 세심하게 프로세싱되도록 FPGA(1910a-f)에게 제공될 수 있다. 예를 들면, MAP를 사용하여, 데이터 모델에 대한 대수식을 평가할 수 있고 일부 경우에서 적응 데이터 재구조화를 수행할 수 있다. 이 예에서, 각각의 MAP는 이 목적을 위해 모든 프로세서들에 의해 전반적으로 접속될 수 있다. 한 환경 설정에서, 각각의 MAP는 직접 메모리 접속(DMA)을 이용하여 관련 메모리(1908a-f)에 접속함으로써, 이 메모리가 각각의 마이크로프로세서(1902a-f)와 무관하게, 그리고 이러한 마이크로프로세서로부터 비동기적으로 과제를 실행할 수 있게 한다. 이 환경 설정에서, MAP는 알고리즘의 파이프라이닝 및 병렬 실행을 위해 결과를 또 다른 MAP에게 직접 공급할 수 있다.

[0180] 상기 컴퓨터 구성 및 시스템은 예일 뿐이고, 일반 프로세서, 보조프로세서, FPGA 및 다른 프로그래밍가능한 논리 디바이스, 칩 상의 시스템(SOC), 응용 특정 통합 회로(ASIC), 및 다른 프로세싱 및 논리 요소의 임의의 조합을 사용하는 시스템들을 포함하는 매우 다양한 다른 컴퓨터, 휴대전화 및 개인 데이터 단말기 구성들 및 시스템들이 예시적 경우와 관련하여 사용될 수 있다. 일부 경우, 컴퓨터 시스템의 전부 또는 일부가 소프트웨어 또는 하드웨어에서 실행될 수 있다. 무작위 접속 메모리, 하드 드라이브, 플래시 메모리, 테이프 드라이브, 디스크 어레이, 네트워크 부착된 저장장치(NAS) 및 다른 근거리 또는 분포된 데이터 저장 디바이스 및 시스템을 포함하는 임의의 다양한 데이터 저장 매체들이 예시적 경우와 관련하여 사용될 수 있다.

[0181] 예시적 경우, 상기 또는 다른 컴퓨터 구성들 및 시스템들 중 임의의 컴퓨터 구성 또는 시스템에서 실행되는 소프트웨어 모듈을 사용하여 컴퓨터 시스템을 구현할 수 있다. 다른 경우, 펌웨어, 프로그래밍가능한 논리 디바이스, 예컨대, 도 19에서 언급된 필드 프로그래밍가능한 게이트 어레이(FPGA), 칩 상의 시스템(SOC), 응용 특정 통합 회로(ASIC), 또는 다른 프로세싱 및 논리 요소에서 시스템의 기능을 부분적으로 또는 전체적으로 구현할 수 있다. 예를 들면, 하드웨어 액셀러레이터 카드, 예컨대, 도 17에 예시된 액셀러레이터 카드(1722)의 이용을 통한 하드웨어 가속화로 세트 프로세서 및 옵티마이저를 구현할 수 있다.

[0182] 하기 실시예는 당분야에서 숙련된 자에게 본원에 개시된 실시양태의 원리 및 실시를 더 명확히 예시하기 위해 기재되어 있고 임의의 청구된 실시양태의 범위를 한정하는 것으로서 해석되어서는 안 된다. 달리 언급되어 있지 않은 한, 모든 부 및 퍼센트는 중량을 기준으로 한 것이다.

[0183] 실시예

[0184] 하기 실시예는 본 개시의 다양한 실시양태들을 예시하기 위해 제공되고 본 개시를 어떠한 방식으로든 한정하기 위한 것이 아니다. 본원에 기재된 방법과 함께 본 실시예는 현재 바람직한 실시양태를 대표하고, 예시하고, 본 개시의 범위를 한정하기 위한 것이 아니다. 청구범위에 의해 정의된 본 개시의 사상 내에 포함되는 변화 및 다른 용도는 당분야에서 숙련된 자에 의해 인식될 것이다.

[0185] 실시예 1: 디바이스 표면의 작용화

[0186] 폴리뉴클레오타이드의 라이브러리의 부착 및 합성을 지지하기 위해 디바이스를 작용화하였다. 먼저, 90% H₂SO₄ 및 10% H₂O₂를 포함하는 피라냐(piranha) 용액을 사용하여 디바이스 표면을 20분 동안 습식 세척하였다. 여러 바이커 내에서 디바이스를 탈이온수로 린싱하고, 탈이온수 거위목 수전 아래에 5분 동안 놓아두고 N₂로 건조하였다. 그 후, 디바이스를 5분 동안 NH₄OH(1:100; 3 mL:300 mL)에 담구고, 핸드건을 이용하여 탈이온수로 린싱하고, 3개의 연속 바이커에서 탈이온수로 각각 1분 동안 담군 후, 핸드건을 이용하여 탈이온수로 다시 린싱하였다. 그 후, 디바이스 표면을 O₂에 노출시켜 디바이스를 플라즈마 세척하였다. SAMCO PC-300 기계를 이용하여 다운스트림 모드로 1분 동안 250 와트에서 O₂로 플라즈마 에칭하였다.

- [0187] 하기 파라미터를 가진 YES-1224P 증착 오븐 시스템을 이용하여 N-(3-트리에톡시실릴프로필)-4-하이드록시부티라미드를 포함하는 용액으로 상기 세척된 디바이스 표면을 능동적으로 작용화하였다: 0.5 내지 1 토르, 60분, 70℃, 135℃ 증발기. Brewer Science 200X 스핀 코팅기를 이용하여 디바이스 표면을 레지스트 코팅하였다. 디바이스를 40초 동안 2500 rpm에서 SPR™ 3612 포토레지스트로 회전 코팅하였다. 디바이스를 90℃의 Brewer 고온 플레이트 위에서 30분 동안 미리 소성하였다. Karl Suss MA6 마스크 얼라이너 기계를 이용하여 디바이스에 대해 포토리쓰그래피를 수행하였다. 디바이스를 2.2초 동안 노출시키고 MSF 26A에서 1분 동안 현상하였다. 남은 현상제를 핸드건으로 린싱하였고 디바이스를 5분 동안 물에 담갔다. 디바이스를 100℃의 오븐 내에서 30분 동안 소성한 후, Nikon L200을 이용하여 리쓰그래피 결함에 대해 시각적으로 검사하였다. SAMCO PC-300 기계를 이용하여 1분 동안 250 와트에서 O₂로 플라즈마 에칭하기 위해 세척 과정을 이용하여 잔류 레지스트를 제거하였다.
- [0188] 10 μ l 경질 미네랄 오일과 혼합된 100 μ l의 퍼플루오로옥틸트리클로로실란 용액으로 디바이스 표면을 수동적으로 작용화하였다. 디바이스를 챔버 내에 배치하고 10분 동안 펌핑한 후, 밸브를 펌프에 가깝게 위치시키고 10분 동안 방치하였다. 공기가 통하도록 챔버를 환기하였다. 최대 출력(크레스트(Crest) 시스템에서 9)으로 초음파처리기를 이용하여 70℃에서 500 ml의 NMP에 5분 동안 2회 담그기를 수행함으로써 디바이스를 레지스트 스트립핑하였다. 그 후, 최대 출력으로 초음파처리를 이용하여 디바이스를 실온에서 500 ml의 이소프로판올에 5분 동안 담갔다. 디바이스를 300 ml의 200 프루프 에탄올에 담구고 N₂로 취입 건조하였다. 작용화된 표면을 활성화시켜 폴리뉴클레오타이드 합성을 위한 지지체로서 사용하였다.
- [0189] **실시예 2: 50머 서열의 합성**
- [0190] 2차원 올리고뉴클레오타이드 합성 디바이스를 플로우셀(flowcell)(Applied Biosystems(ABI394 DNA 합성기))에 연결된 플로우셀로 조립하였다. 2차원 올리고뉴클레오타이드 합성 디바이스를 N-(3-트리에톡시실릴프로필)-4-하이드록시부티라미드(Gelest)로 균일하게 작용화하고 본원에 기재된 폴리뉴클레오타이드 합성 방법을 이용하여 50 bp의 예시적 폴리뉴클레오타이드("50머 폴리뉴클레오타이드")를 합성하는 데 사용하였다.
- [0191] 50머의 서열은 서열번호 20에 기재된 바와 같았다:
5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTTTTT3'(서열번호 20), 이때 #은 탈보호 동안 표면으로부터 폴리뉴클레오타이드를 방출할 수 있게 하는 절단가능한 링커인 타이미딘-석시닐 헥사아미드 CED 포스포라미다이트(켄진스(ChemGenes)의 CLP-2244)를 표시한다.
- [0192] 표 4의 프로토콜에 따른 표준 DNA 합성 화학반응(커플링, 캡핑, 산화 및 탈블록화) 및 ABI 합성기를 이용하여 합성을 수행하였다.

표 4

합성 프로토콜

일반적 DNA 합성 과정 명칭	과정 단계	시간(초)
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	23
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
DNA 염기 첨가(포스포라미다이트 + 활성화제 유동)	활성화제 매니폴드 플러시	2
	플로우셀로의 활성화제	6
	플로우셀로의 활성화제 + 포스포라미다이트	6
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포라미다이트	5
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포라미다이트	5
	플로우셀로의 활성화제	0.5
	플로우셀로의 활성화제 + 포스포라미다이트	5
	25 초 동안 항온처리	25
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
DNA 염기 첨가(포스포라미다이트 + 활성화제 유동)	활성화제 매니폴드 플러시	2
	플로우셀로의 활성화제	5
	플로우셀로의 활성화제 + 포스포라미다이트	18
	25 초 동안 항온처리	25
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4

[0193]

일반적 DNA 합성 과정 명칭	과정 단계	시간(초)
	아세트오니트릴 시스템 플러시	4
캡핑(CapA+ B, 1:1, 유동)	플로우셀로의 CapA+ B	15
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	아세트오니트릴 시스템 플러시	4
산화(산화제 유동)	플로우셀로의 산화제	18
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	23
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	15
탈블록화(탈블록 유동)	플로우셀로의 탈블록	36
세척(아세트오니트릴 세척 유동)	아세트오니트릴 시스템 플러시	4
	N2 시스템 플러시	4
	아세트오니트릴 시스템 플러시	4
	플로우셀로의 아세트오니트릴	18
	N2 시스템 플러시	4.13
	아세트오니트릴 시스템 플러시	4.13
	플로우셀로의 아세트오니트릴	15

[0194]

[0195]

포스포라미다이트/활성화제 조합을 플로우셀을 통한 대용량 시약의 전달과 유사하게 전달하였다. 환경이 전체 시간 동안 시약에 의해 "젖은" 상태로 유지되기 때문에 건조 단계를 수행하지 않았다.

[0196]

ABI 394 합성기로부터 유동 제한장치를 제거하여 더 빨리 유동할 수 있게 하였다. 유동 제한장치를 이용하지 않는 경우, 아미다이트(ACN 중의 0.1 M), 활성화제(ACN 중의 0.25 M 벤조일티오테트라졸("BT"; 글렌리서치(GlenResearch)로부터의 30-3070-xx)) 및 Ox(20% 피리딘 중의 0.02 M I₂, 10% 물 및 70% THF)에 대한 유속은 대략 100 μ l/초이었고, 아세트오니트릴("ACN") 및 캡핑 시약(CapA와 CapB의 1:1 혼합물, 이때 CapA는 THF/피리딘 중의 아세트산 무수물이고, CapB는 THF 중의 16% 1-메틸이미디졸임)에 대한 유속은 대략 200 μ l/초이었고, 탈블록화제(톨루엔 중의 3% 디클로로아세트산)에 대한 유속은 대략 300 μ l/초이었다(유동 제한장치를 이용한 경우 모든 시약들에 대한 유속 약 50 μ l/초와 비교됨). 산화제를 완전히 밀어내기 위한 시간을 관측하였고, 화학적 유동 시간에 대한 타이밍을 그에 맞게 조절하였고, 상이한 화학물질들 사이에 가외의 ACN 세척을 도입하였다. 폴리뉴클레오타이드 합성 후, 75 psi에서 하룻밤 동안 암모니아 기체로 칩을 탈보호하였다. 5 방울의 물을 표면에 적용하여 폴리뉴클레오타이드를 회수하였다. 그 다음, 회수된 폴리뉴클레오타이드를 생물분석기 작은 RNA 칩에서 분석하였다(데이터는 표시되어 있지 않음).

[0197]

실시예 3: 100머 서열의 합성

[0198] 실시예 2에 기재된 50머 서열의 합성 과정과 동일한 과정을 이용하여 2개의 상이한 규소 칩, 즉 N-(3-트리에톡시실릴프로필)-4-하이드록시부티라미드에 의해 균일하게 작용화된 제1 칩, 및 11-아세톡시운데실트리에톡시실란과 n-데실트리에톡시실란의 5/95 혼합물에 의해 작용화된 제2 칩에서 100머 폴리뉴클레오타이드("100머 폴리뉴클레오타이드"; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCATTTTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT## TTTTTTTTTT3'(서열번호 21), 이때 #은 타이미딘-석시닐 헥사아미드 CED 포스포라미다이트(캠진스의 CLP-2244)를 표시함)를 합성하였고, 표면으로부터 추출된 폴리뉴클레오타이드를 생물분석기 기계에서 분석하였다(데이터는 표시되어 있지 않음).

[0199] 하기 열순환 프로그램을 이용하여 50 μ l PCR 혼합물(25 μ l NEB Q5 마스터믹스, 2.5 μ l 10 μ M 정방향 프라이머, 2.5 μ l 10 μ M 역방향 프라이머, 표면으로부터 추출된 1 μ l 폴리뉴클레오타이드, 및 50 μ l까지 물)에서 정방향 프라이머(5'ATGCGGGTTCTCATCATC3'; 서열번호 22) 및 역방향 프라이머(5'CGGATCCTTATCGTCATCG3'; 서열번호 23)를 사용하여 상기 2개의 칩으로부터의 모든 10개의 샘플들을 PCR로 더 증폭하였다:

[0200] 98℃, 30초

[0201] 98℃, 10초; 63℃, 10초; 72℃, 10초; 12 주기 반복

[0202] 72℃, 2분

[0203] PCR 생성물을 생물분석기에서도 분석하여(데이터는 제시되어 있지 않음), 100머 위치에서 날카로운 피크를 입증하였다. 그 다음, PCR에 의해 증폭된 샘플을 클로닝하고 생거 시퀀싱하였다. 표 5는 칩 1의 스폿 1 내지 5로부터 채취된 샘플 및 칩 2의 스폿 6 내지 10으로부터 채취된 샘플에 대한 생거 시퀀싱으로부터의 결과를 요약한다.

5 5

시퀀싱 결과

스폿	오류율	주기 효율
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[0204]

[0205] 따라서, 상이한 표면 화학반응을 이용하여 2개의 칩에서 고품질의 균일한 합성된 폴리뉴클레오타이드를 반복하였다. 종합하건대, 시퀀싱된 262개의 100머들 중 233개에 상응하는 89%의 서열들은 오류를 갖지 않는 완벽한 서열들이었다. 최종적으로, 표 6은 스폿 1 내지 10으로부터의 폴리뉴클레오타이드 샘플로부터 수득된 서열에 대한 오류 특성을 요약한다.

표 6

오류 특성

샘플 ID/스팟 번호	OSA_0 046/1	OSA_0 047/2	OSA_0 048/3	OSA_0 049/4	OSA_0 050/5	OSA_0 051/6	OSA_0 052/7	OSA_0 053/8	OSA_0 054/9	OSA_00 55/10
총 서열	32	32	32	32	32	32	32	32	32	32
시퀀싱 품질	28 중 25	27 중 27	30 중 26	23 중 21	26 중 25	30 중 29	31 중 27	31 중 29	29 중 28	28 중 25
올리고 품질	25 중 23	27 중 25	26 중 22	21 중 18	25 중 24	29 중 25	27 중 22	29 중 28	28 중 26	25 중 20
ROI 일치 카운트	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI 돌연변이	2	2	1	3	1	0	2	1	2	1
ROI 다중 염기 결실	0	0	0	0	0	0	0	0	0	0
ROI 작은 삽입	1	0	0	0	0	0	0	0	0	0
ROI 단일 염기 결실	0	0	0	0	0	0	0	0	0	0
큰 결실 카운트	0	0	1	0	0	1	1	0	0	0
돌연변이: G>A	2	2	1	2	1	0	2	1	2	1
돌연변이: T>C	0	0	0	1	0	0	0	0	0	0
ROI 오류 카운트	3	2	2	3	1	1	3	1	2	1
ROI 오류율	오류율: 약 1/834	오류율: 약 1/1350	오류율: 약 1/1282	오류율: 약 1/708	오류율: 약 1/2500	오류율: 약 1/2667	오류율: 약 1/876	오류율: 약 1/2900	오류율: 약 1/1400	오류율: 약 1/2349
ROI 마이너스 프라이머 오류율	MP 오류율: 약 1/763	MP 오류율: 약 1/824	MP 오류율: 약 1/780	MP 오류율: 약 1/429	MP 오류율: 약 1/1525	MP 오류율: 약 1/1615	MP 오류율: 약 1/531	MP 오류율: 약 1/1769	MP 오류율: 약 1/854	MP 오류율: 약 1/1451

[0206]

[0207]

실시예 4: 단일 부위 단일 위치 돌연변이유발에 의한 핵산 라이브러리의 생성

[0208]

주형 핵산의 핵산 변이체의 라이브러리를 생성하기 위해 일련의 PCR 반응들에서 사용할 폴리뉴클레오타이드 프라이머를 드노보 합성하였다(도 4a 내지 4d 참조). 도 4a에서 4종의 프라이머들이 생성되었다: 외부 5' 프라이머(415), 외부 3' 프라이머(430), 내부 5' 프라이머(425) 및 내부 3' 프라이머(420). 표 4에 일반적으로 요약된 폴리뉴클레오타이드 합성 방법을 이용하여 내부 5' 프라이머/제1 폴리뉴클레오타이드(420) 및 내부 3' 프라이머/제2 폴리뉴클레오타이드(425)를 생성하였다. 내부 5' 프라이머/제1 폴리뉴클레오타이드(420)는 소정의 서열의 최대 19개 프라이머의 세트를 대표하고, 이때 상기 세트 내의 각각의 프라이머는 서열의 단일 부위에서 단일 코돈에서 서로 상이하다.

[0209]

121개의 개별적으로 주소지정가능한 좌위들을 각각 가진 적어도 2개의 클러스터들을 가진 디바이스에서 폴리뉴클레오타이드 합성을 수행하였다.

[0210]

내부 5' 프라이머(425) 및 내부 3' 프라이머(420)를 별도의 클러스터에서 합성하였다. 단일 클러스터 내의 121개의 좌위들에서 연장하도록 내부 5' 프라이머(425)를 121회 복제하였다. 내부 3' 프라이머(420)의 경우, 변이체 서열의 19개의 프라이머들 각각을 6개의 상이한 좌위들에서 각각 연장하여, 114개의 상이한 좌위들에서 114개의 폴리뉴클레오타이드들을 연장하였다.

[0211]

합성된 폴리뉴클레오타이드를 디바이스의 표면으로부터 절단하고 플라스틱 바이알로 옮겼다. 도 4b에 예시된 바와 같이, 긴 핵산 서열의 단편(435 및 440)을 사용하여 제1 PCR 반응을 수행함으로써, 주형 핵산을 증폭하였다. 도 4c 및 4d에 예시된 바와 같이, 프라이머 조합, 및 주형으로서 제1 PCR 반응의 생성물을 사용하여 제2 PCR 반응을 수행하였다. 도 20의 추적선에 표시된 바와 같이, 제2 PCR 생성물의 분석을 생물분석기에서 수행하였다.

[0212]

실시예 5: 96개의 상이한 단일 위치 변이체 세트들을 포함하는 핵산 라이브러리의 생성

[0213] 도 4a에 일반적으로 표시되어 있고 실시예 2에서 다루어진 바와 같이, 드노보 폴리뉴클레오타이드 합성을 이용하여 4개의 프라이머 세트들을 생성하였다. 내부 5' 프라이머(420)에 대해, 주형 핵산의 단일 부위 내에 위치한 상이한 단일 코돈을 각각 표적화하는 96개의 상이한 프라이머 세트들을 생성하였다. 각각의 프라이머 세트에 대해, 단일 부위에서 상이한 아미노산을 코딩하는 코돈을 각각 포함하는 19개의 상이한 변이체들을 생성하였다. 도 4a 내지 4d에 일반적으로 표시되어 있고 실시예 2에 기재된 바와 같이, 생성된 프라이머를 사용하여 2 라운드의 PCR을 수행하였다. 100% 증폭 성공률을 계산하는 데 사용된 전기영동도에서 96개의 증폭 생성물 세트들을 가시화하였다(도 21).

[0214] **실시예 6: 500개의 상이한 단일 위치 변이체 세트들을 포함하는 핵산 라이브러리의 생성**

[0215] 도 4a에 일반적으로 표시되어 있고 실시예 2에서 다루어진 바와 같이, 드노보 폴리뉴클레오타이드 합성을 이용하여 4개의 프라이머 세트들을 생성하였다. 내부 5' 프라이머(420)에 대해, 주형 핵산의 단일 부위 내에 위치한 상이한 단일 코돈을 각각 표적화하는 500개의 상이한 프라이머 세트들을 생성하였다. 각각의 프라이머 세트에 대해, 단일 부위에서 상이한 아미노산을 코딩하는 코돈을 각각 포함하는 19개의 상이한 변이체들을 생성하였다. 도 4a에 일반적으로 표시되어 있고 실시예 2에 기재된 바와 같이, 생성된 프라이머를 사용하여 2 라운드의 PCR을 수행하였다. 전기영동도는 상이한 단일 부위에서 19개의 변이체들을 가진 핵산 집단을 가진 500개의 PCR 생성물 세트들 각각을 표시한다(데이터는 제시되어 있지 않음). 라이브러리의 포괄적 시퀀싱 분석은 미리 선택된 코돈 돌연변이 전체에 걸쳐 99% 초과 성공률을 보여주었다(서열 추적 및 분석 데이터는 제시되어 있지 않음).

[0216] **실시예 7: 1개의 위치에 대한 단일 부위 돌연변이 유발 프라이머**

[0217] 황색 형광 단백질에 대한 코돈 변이 디자인의 한 예는 표 7에 제공되어 있다. 이 경우, 50개의 서열로부터의 단일 코돈을 19회 변이시킨다. 변이체 핵산 서열은 굵은 글자체로 표시되어 있다. 야생형 프라이머 서열은 ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT(서열번호 1)이다. 이 경우, 야생형 코돈은 서열번호 1에서 밑줄로 표시된 발린을 코딩한다. 따라서, 하기 19개의 변이체들은 발린을 코딩하는 코돈을 배제한다. 대안적 예에서, 모든 삼중체들이 고려되어야 하는 경우, 야생형 코돈에 대한 대안적 서열을 포함하는 모든 60개의 변이체들이 생성될 것이다.

표 7

변이체 서열

서열 번호	변이체 서열	변이체 코돈
2	atg TT TAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	F
3	atg TTA AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	L
4	atg ATT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	I
5	atg TCT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	S
6	atg CCT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	P
7	atg ACT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	T
8	atg GCT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	A
9	atg TAT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	Y
10	atg CAT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	H
11	atg CAA AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	Q
12	atg AAT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	N
13	atg AAA AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	K
14	atg GAT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	D
15	atg GAA AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	E
16	atg TGT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	C
17	atg TGG AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	W
18	atg CGT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	R
19	atg GGT AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT	G

[0218]

[0219] **실시예 8: 단일 부위 이중 위치 핵산 변이체**

[0220] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 단일 부위에서 2개의 연속 코돈 위치들에 대한 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성하였고, 이때 각각의 위치는 아미노산을 코딩하는 코돈이다. 이 방식에서, 각각의 핵산의 3개의 복제물을 사용하

여 2개의 위치에 대해 위치당 19개의 변이체들을 생성함으로써, 114개의 핵산들을 합성하였다.

[0221] **실시예 9: 다중 부위 이중 위치 핵산 변이체**

[0222] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 2개의 불연속 코돈 위치들에 대한 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성하였고, 이때 각각의 위치는 아미노산을 코딩하는 코돈이다. 이 방식에서, 2개의 위치에 대해 위치당 19개의 변이체들을 생성하였다.

[0223] **실시예 10: 단일 스트레치 삼중 위치 핵산 변이체**

[0224] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 3개의 연속 코돈 위치들에 대한 기준 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성하였다. 3개의 연속 코돈 위치 방식에서, 각각의 핵산의 2개의 복제물을 사용하여 3개의 위치에 대해 위치당 19개의 변이체들을 생성함으로써, 114개의 핵산들을 합성하였다.

[0225] **실시예 11: 다중 부위 삼중 위치 핵산 변이체**

[0226] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 적어도 3개의 불연속 코돈 위치들에 대한 기준 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성하였다. 소정의 영역 내에서, 3개의 히스티딘 잔기들을 코딩하는 코돈의 위치를 변이시켰다.

[0227] **실시예 12: 다중 부위 다중 위치 핵산 변이체**

[0228] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 하나 이상의 스트레치에서 하나 이상의 코돈 위치에 대한 기준 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성하였다. 5개의 위치를 라이브러리에서 변이시켰다. 제1 위치는 발현된 단백질에서 50/50 K/R 비로 코돈을 코딩하였고; 제2 위치는 발현된 단백질에서 50/25/25 V/L/S 비로 코돈을 코딩하였고; 제3 위치는 발현된 단백질에서 50/25/25 Y/R/D 비로 코돈을 코딩하였고; 제4 위치는 발현된 단백질에서 모든 아미노산들에 대해 동등한 비로 코돈을 코딩하였고; 제5 위치는 발현된 단백질에서 75/25 G/P 비로 코돈을 코딩하였다.

[0229] **실시예 13: 샘플링에 의한 핵산 라이브러리의 생성**

[0230] 미리 선택된 분포를 가진 핵산 집단을 생성하기 위해, 전산 기법을 이용하였다. 미리 선택된 분포의 한 예는 하기 표 8에 제공되어 있고, 이 표에서 숫자는 각각의 위치에서 각각의 아미노산의 원하는 퍼센트를 표시한다. 표 9에 나타낸 바와 같이, 누적 분포 값을 먼저 계산하여 0.0 내지 1.0의 값을 획득하였다. 엑셀과 같은 프로그램에서, 균일한 무작위 숫자 생성기를 이용하여, 샘플링 집단으로서 사용된 500개의 핵산에 대해 10개의 아미노산 위치 각각에 대한 0 내지 1의 값을 생성하였다. 예를 들면, 위치 1에 대한 "0.95"의 균일한 무작위 값은 "S" 버킷에 속할 것이므로, 아미노산 "S"를 표시한다. 이 기법은 "롤렛-휠(roulette-wheel)" 선택으로서 지칭된다. 각각의 디자인된 올리고뉴클레오타이드에 대한 10개의 구별된 분포로부터 10개의 무작위 숫자를 생성하였고; 이 과정을 500회 반복하여 500개 핵산들의 샘플 집단을 생성하였다. 생성된 샘플 집단을 검증하기 위해, 각각의 아미노산이 그 위치에서 나타나는 빈도의 합계를 상기 집단 전체에 걸쳐 측정한 후 퍼센트로서 표현하였다. 예를 들면, 500개 핵산들의 상기 샘플에서 아미노산 C가 위치 1에서 나타나는 퍼센트를 계산하였다. 값은 집단에서의 대략적인 분포를 표시한다. 상기 집단에서 충분한 수의 핵산을 사용하였을 때, 샘플 분포는 미리 선택된 분포에 가까웠다.

표 8

아미노산의 미리 선택된 분포

	위치									
아미노산	1	2	3	4	5	6	7	8	9	10
% C	0.1	0.2	0.1	0.1	0.0	0.0	0.1	6.0	2.0	0.1
% A	13.7	2.4	4.0	8.9	4.8	7.1	5.1	6.0	5.1	13.1
% R	13.7	16.7	6.7	13.3	6.0	8.3	5.1	6.0	3.8	6.6
% N	1.1	2.4	2.7	4.4	2.4	3.6	6.3	6.0	3.8	4.9
% D	15.8	16.7	5.3	4.4	1.2	13.1	21.5	6.0	11.4	8.2
% Q	7.4	6.0	4.0	2.2	1.2	1.2	2.5	6.0	3.8	1.6
% E	4.2	3.6	1.3	6.7	2.4	13.1	5.1	6.0	11.4	3.3
% G	14.7	15.5	20.0	13.3	27.4	9.5	11.4	6.0	11.4	3.3
% H	1.1	1.2	1.3	2.2	1.2	2.4	2.5	6.0	3.8	3.3
% I	1.1	1.2	10.7	3.3	1.2	6.0	5.1	6.0	3.8	3.3
% L	7.4	2.4	2.7	4.4	3.6	4.8	6.3	6.0	3.8	3.3
% K	2.1	4.8	1.3	2.2	6.0	1.2	1.3	6.0	2.5	9.8
% M	2.1	1.2	1.3	2.2	2.4	1.2	1.3	6.0	10.1	2.0
% F	1.1	1.2	4.0	6.7	3.6	2.4	1.3	6.0	2.5	3.3
% P	7.4	6.0	8.0	12.2	3.6	4.8	2.5	6.0	10.1	16.4
% S	5.3	9.5	10.7	10.0	19.0	19.0	19.0	6.0	8.9	11.5
% T	2.1	9.5	16.0	3.3	14.3	2.4	3.8	6.0	3.8	6.6

[0231]

표 9

누적 표준화 분포

	위치									
아미노산	1	2	3	4	5	6	7	8	9	10
C	.00	.00	.00	.00	.00	.00	.00	.06	.02	.00
A	.14	.03	.04	.09	.05	.07	.05	.12	.07	.13
R	.27	.19	.11	.22	.11	.15	.10	.18	.11	.20
N	.28	.22	.13	.27	.13	.19	.17	.24	.14	.25
D	.44	.38	.19	.31	.14	.32	.38	.29	.26	.33
Q	.52	.44	.23	.33	.16	.33	.41	.35	.29	.34
E	.56	.48	.24	.40	.18	.46	.46	.41	.40	.38
G	.70	.63	.44	.53	.45	.56	.57	.47	.52	.41
H	.72	.64	.45	.56	.46	.58	.59	.53	.55	.44
I	.73	.66	.56	.59	.48	.64	.65	.59	.59	.47
L	.80	.68	.59	.63	.51	.69	.71	.65	.63	.51
K	.82	.73	.60	.66	.57	.70	.72	.71	.65	.60
M	.84	.74	.61	.68	.60	.71	.73	.76	.75	.62
F	.85	.75	.65	.74	.63	.74	.75	.82	.78	.66
P	.93	.81	.73	.87	.67	.79	.77	.88	.88	.82
S	.98	.91	.84	.97	.86	.98	.96	.94	.96	.93
T	1	1	1	1	1	1	1	1	1	1

[0232]

[0233] 실시예 14: 여과된 샘플링에 의한 핵산 라이브러리의 생성

[0234] 실시예 13에 기재된 방법을 이용하여, 집단의 재샘플링을 수행함으로써 원치 않는 조합을 제거하고 이를 집단으로부터 여과하여 제거하였다. 예를 들면, 임의의 위치에서 4개의 "H"(히스티딘) 아미노산을 가진 조합은 생물학적 목적에 적합하지 않은 것으로서 간주되었다. 따라서, 이 경우, 500번째 올리고뉴클레오타이드가 "HHHCCHHCHH(서열번호 55)"로서 생성되었을 때, 조합은 8개의 H를 가졌기 때문에 바람직하지 않았다. 결과적으로, 실시예 13에 기재된 방법 후 또 다른 무작위로 생성된 조합이 그 자리에서 생성되었다. 임의의 수의 기준을 이용하여 미리 선택된 분포를 생성하였다. 예를 들면, 임의의 위치에서 각각의 올리고뉴클레오타이드 내에 적어

도 하나의 "A"(알라닌) 아미노산을 포함하도록 집단을 생성하였다. 또한, 생성된 조합이 서로 인접한 2개의 "M"(메티오닌) 아미노산을 갖지 않도록 집단을 생성하였다. 따라서, 미리 선택된 분포 및 구체적인 기준이 충족될 때까지 무작위 샘플링을 수행하였다.

[0235] **실시예 15: 균일한 분포를 가진 조합 라이브러리**

[0236] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 집단을 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성하였고, 이때 변이체는 각각의 위치에서 미리 선택되었고 미리 선택된 분포를 가진다.

[0237] 조합 방법으로 균일한 변이체 분포 라이브러리를 생성하기 위해, 변이체 라이브러리에 대한 기준 서열을 2개의 부분으로 분할하였다. 본원에서 사용된 균일한 변이체 분포는 각각의 변이체가 대략 동등한 양으로 합성될 것임을 의미하기 위한 것이다. 분할의 한 측면은 5' 측면으로서 지칭되었고, 분할의 제2 측면은 3' 측면으로서 지칭되었다. 어닐링될 때 원하는 핵산 라이브러리가 합성되도록 기준 서열의 각각의 측면에 대한 서열을 디자인하고 합성하였다. 표 10과 유사한 변이를 가진 균일한 라이브러리의 경우, 5' 측면에서의 다양성은 2548($14 \times 14 \times 13$)이다. 3' 측면에서의 다양성은 546($3 \times 13 \times 14$)이다. 5' 측면 및 3' 측면을 어닐링으로 합성하여, 1,391,208(2548×546)의 총 다양성을 야기하였다. 변이체를 차세대 시퀀싱으로 분석하였다(테이터는 제시되어 있지 않음).

표 10

균일한 라이브러리의 변이

	5' 변이체			3' 변이체		
N	N	N	N		N	N
P						
Q						
R	R	R	R		R	R
S	S	S	S		S	S
T						
V	V	V	V		V	V
W	W	W	W	W	W	W
Y	Y	Y	Y		Y	Y
A	A	A	A	A	A	A
C						
D						
E						
F	F	F	F		F	F
G	G	G	G		G	G
H	H	H	H		H	H
I	I	I				I
K	K	K	K	K	K	K
L	L	L	L		L	L
M	M	M	M		M	M
다양성	2548			546		
총 다양성	1391208					

[0238]

[0239] **실시예 16: 불균일한 분포를 가진 조합 라이브러리**

[0240] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 집단을 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성하였고, 이때 변이체는 각각의 위치에서 미리 선택되었고 미리 선택된 분포를 가진다.

[0241] 표 11에서 확인되는 분포와 유사한 미리 선택된 분포를 가진, 불균일한 변이체 분포를 가진 라이브러리도 생성하였다. 기준 서열을 다시 절반으로 분할하였고, 각각의 부분에 대한 변이체를 생성하였다. 분할의 한 측면은 5' 측면으로서 지칭되었고, 분할의 제2 측면은 3' 측면으로서 지칭되었다. 그 변이체에 대한 치환의 이론적 빈도를 곱함으로써 5' 변이체 및 3' 변이체의 예측된 확률을 계산하였다. 예를 들면, 서열 NRS의 5' 변이체에 대해, 예측된 확률은 $0.0677\%(9.9\% \times 7.6\% \times 9.0\%)$ 이었다. 5' 변이체 및 3' 변이체에 대해, 변이체들의 일부는

동일한 확률을 가졌고 함께, 즉 동일한 확률 "빈(bin)" 내로 분류되었다. 따라서, 동일한 빈 내의 모든 변이체들은 동일한 이론적 발생 빈도를 가진다. 1,391,208개의 총 이론적 변이체들에 대해, 162개의 상이한 확률이 있었으므로, 162개의 상이한 확률 빈이 있었다.

표 11

변이 분포

	5' 변이체			3' 변이체		
N	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
P	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Q	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
R	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
S	3.0%	9.0%	9.0%	0.0%	9.0%	3.0%
T	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
V	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
W	4.0%	4.0%	4.0%	20.0%	4.0%	4.0%
Y	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
A	4.0%	4.0%	4.0%	60.0%	4.0%	4.0%
C	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
D	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
E	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
F	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
G	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
H	9.9%	7.6%	8.7%	0.0%	8.7%	9.9%
I	9.9%	7.6%	0.0%	0.0%	0.0%	9.9%
K	4.0%	4.0%	4.0%	20.0%	4.0%	4.0%
L	3.0%	9.0%	9.0%	0.0%	9.0%	3.0%
M	3.0%	9.0%	9.0%	0.0%	9.0%	3.0%

[0242]

[0243]

그 다음, 차세대 시퀀싱(NGS)을 수행하여, 생성된 변이체에서 얼마나 많은 이론적 다양성이 표시되는 지를 확인하였다. 시퀀싱을 10^6 개의 리드(read)로 수행하였기 때문에, 단지 30%의 실제 다양성이 관측되었다. 따라서, 원하는 빈도로 표시된 총 실제 다양성을 측정하였다.

[0244]

동일한 빈도로 다수의 변이체들을 표시하는 162개의 상이한 확률 빈을 사용하여 NGS 데이터를 분석하였다. 162개의 상이한 확률 빈에 대해, 도 22에서 확인되는 바와 같이 NGS로부터의 리드를 그의 예측된 발생 확률(파선)로 분류하였다. 그 다음, 관측된 빈도(직선)를 예측된 확률과 비교하였다. 162개의 빈 각각에 대해, 변이체의 총 수를 그 빈 내의 변이체의 수로 나눔으로써 관측된 빈도를 측정하였다. 각각의 빈에 대해 이 값을 계산하였고 도 23에 나타난 바와 같이 평균 카운트로서 표시한다. 이 값을 관측된 빈도로서 그래프로 작성하였고 도 22에 나타난 바와 같이 예측된 확률과 비교하였다.

[0245]

도 22에 나타난 바와 같이, 변이체의 관측된 빈도(직선)와 변이체의 예측된 확률(파선)의 비교는 관측된 다양성이 원하는 빈도로 표시되는 지를 시사한다. 도 22에서 확인된 바와 같이, 관측된 다양성은 예측된 확률과 잘 일치하고, 99% 초과인 이론적 다양성이 표시되었다.

[0246]

추가로, 높은 빈도 조합뿐만 아니라 소정의 낮은 빈도 조합도 관측되었다. 다양성의 39개 염기 쌍 영역에 걸쳐 있는 89.9%의 NGS 리드는 정확한 크기를 가졌고, 70% 초과인 완전한 126개 염기 쌍 구축물은 삽입 및 결실을 갖지 않는 것으로 평가되었다. 도 24를 참조하건대, 단일 피크에 의해 표시된 바와 같이 높은 퍼센트의 전체 길이 단편이 생성되었다.

[0247]

실시예 17: 8개의 위치들 각각에서 144개의 단일 코돈 변이체들 및 9072개의 이중 코돈 변이체들을 포함하는 조합 라이브러리

[0248]

실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행하였다. 핵산 집단을 실시예 4 내지 6 및 8 내지 12와 유사하게 생성하였다. 핵산 집합은 144개의 단일 코돈 변이체들 및 9072개의 이중 코돈 변이체들을 포함하였고(9216의 다양성), 이때 변이체는 8개의 위치에서 미리 선택되었다.

[0249]

그 다음, 차세대 시퀀싱(NGS)을 수행하여 관측된 조합 변이체의 분포를 확인하였다. 10^5 초과인 리드 적용범위(coverage)로 시퀀싱을 수행하였다. 도 25에서 확인된 바와 같이, 99% 초과인 관측된 변이체가 균일한 분포를

가진 것으로 NGS에 의해 검출되었다. 90% 초과인 관측된 변이체는 삽입 및 결실을 갖지 않았고, 5% 미만의 표적 이탈 서열이 검출되었다. 1% 미만의 야생형 서열이 관측되었다.

[0250] **실시예 18: 어레이 기반 방법을 이용한 대표적 변이체 라이브러리의 생성**

[0251] 실시예 1 내지 3과 유사한 어레이 기반 방법을 이용하여 변이체 라이브러리를 드노보 합성하였다. 그 다음, 어레이 기반 방법을 이용함으로써 생성된 변이체 라이브러리를, PCR 기반 방법을 이용함으로써 생성된 변이체 라이브러리와 비교하였다.

[0252] 변이체 라이브러리 구축 후, 2개의 라이브러리들로부터 콜로니를 샘플링하고 시퀀싱하였다. 데이터는 표 12에 제시되어 있다. 실패한 시퀀싱의 수("실패한 시퀀싱 수")는 시퀀싱이 가능하지 않았던 콜로니의 수로서 측정되었다. 시퀀싱 후 수득된 돌연변이체의 수 대 예측된 이론적으로 가능한 돌연변이체의 수의 비로부터 퍼센트 다양성(다양성(%))을 측정하였다. 퍼센트 정확도("정확도(%))는 정확한 DNA 서열을 가진 돌연변이체의 수 대 시퀀싱을 위해 사용된 돌연변이체의 수의 비에 의해 측정되었다. 표 12로부터, 어레이 기반 방법을 이용함으로써 생성된 변이체 라이브러리는 개선된 다양성 및 품질과 상관관계를 가진 더 높은 "정확도"를 나타내었다.

[0253] 또한, 2개의 라이브러리들을 샘플링으로 단백질 수준에서 비교하였다. 어레이 기반 방법을 이용함으로써 생성된 변이체 라이브러리는 PCR 기반 방법을 이용함으로써 생성된 변이체 라이브러리보다 증가된 수의 이론적으로 예측된 생성된 돌연변이체를 가진 더 대표적인 변이체 집단을 가졌다.

표 12

변이체 라이브러리 데이터

라이브러리	결실을 가진 돌연변이체의 수	삽입을 가진 돌연변이체의 수	실패한 시퀀싱의 수	야생형의 수	상당한 돌연변이체의 수	다양성 (%)	정확도 (%)
PCR 기반 라이브러리	14	4	18	4	109	42.5	86.4
어레이 기반 라이브러리	12	4	2	2	164	64.1	93.2

[0254]

[0255] **실시예 19: 코돈 배정 체계**

[0256] 코돈 배정을 이용하여 폴리뉴클레오타이드 라이브러리를 디자인하였다. 코돈 배정을 이용하여 각각의 부위에서 디자인될 코돈 서열을 결정하였다.

[0257] 표 13에 나열된 야생형(WT) 아미노산 서열 및 WT DNA 서열을 가진 인간 종양 단백질 p53(TP53)에 대한 코돈 변이를 생성하였다. 코돈 변이를 생성할 때, 디자인될 변이체 코돈 서열은 상기 표 3의 코돈 배정에 기반을 두었다. 구체적으로, WT 아미노산으로부터 변이체 아미노산을 생성할 때, 변이체 아미노산을 코딩하는 변이체 코돈 서열을 표 3에 나열된 코돈 서열로부터 좌측에서 우측으로 먼저 선택하였다.

[0258] 표 13을 참조하건대, 펩타이드의 위치 2에 있는 WT 아미노산은 "F"(붉은 글자체)이다. 위치 2에서 변이를 생성하기 위해, WT 서열의 변이체를 디자인하였는데, 이때 "F"는 다른 19개의 아미노산들 중 임의의 아미노산으로 교체되었다. 그 다음, 표 3에 따른 코돈 배정을 이용하여, 그 위치에서 변이체 아미노산을 생성하기 위해 어떤 변이체 코돈 서열을 디자인할 지를 결정하였다. "F"가 "A"로 교체되어 있는 변이체를 생성하기 위해, 먼저 표 3에 따라 선택된 변이체 코돈 서열은 모두 "A"를 코딩하는 "GCA", "GCC" 또는 "GCG" 대신에 "GCT"이었다. 표 14는 위치 2에서 "F"의 모든 가능한 변이체 아미노산들을 나열하고, 변이체 아미노산을 생성하기 위해 어떤 변이체 코돈 서열을 디자인하였는 지를 보여준다.

표 13

변이에 대한 서열

서열 번호	아미노산 또는 DNA	서열
33	아미노산	MFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTVEVVRCPH HERCSDSDGLAPPQHLIRVEGNLRVEYLDNRNTRFRHSVVVPYEPPEVG SDCTTIHYNMNCSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEVVRVC ACPGRRDRRTEENLRKKGEPHLEPPGSTKRALPNNTSSSPQPKKKPL DGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPGGSGRAHSSHLKS KKQSTSRHKLMFKTEGPDS
34	DNA	TGAGGCCAGGAGATGGAGGCTGCAGTGAGCTGTGATCACACCACT GTGCTCCAGCCTGAGTGACAGAGCAAGACCCTATCTCAAAAAA AAAAAAGAAAGCTCCTGAGGTGTAGACGCCAACTCTCTCT AGCTCGCTAGTGGGTGCAGGAGGTGCTTACGCATGTTTGTCTTT GCTGCCGTCTCCAGTTGCTTTATCTGTTCACTTGTGCCCTGACTTT CAACTCTGTCTCCTTCTCTTCTACAGTACTCCCTGCCCTCAACA AGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGT TGATTCCACACCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATC TACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCC CACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCCCCTCCTCAGC ATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGA CAGAAACACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCCT GAGGTTGGCTCTGACTGTACCACTCCACTACAACATACATTTGTA ACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCATCCTCACCAT CATCACACTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAG CTTTGAGGTGCGTGTGTGTGCTGTCTGGGAGAGACCGGCGCACA GAGGAAGAGAATCTCCGCAAGAAAGGGGAGCCTCACCACGAGCTG CCCCCAGGGAGCACTAAGCGAGCACTGCCCAACAACACAGCTCC TCTCCCCAGCCAAGAAGAAACCACTGGATGGAGAATATTTACAC CTTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGA ATGAGGCCTTGGAATCAAGGATGCCAGGCTGGGAAGGAGCGAG GGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGTC AGTCTACCTCCCGCCATAAAAACTCATGTTCAAGACAGAAGGGC CTGACTCAGACTGACATTCTCCACTTCTGTTCCCCACTGACAGCCT CCCACCCCACTCTCTCCCTCCCTGCCATTTTGGGTTTTGGGTCTTT GAACCCTTGCTTGAATAGGTGTGCGTCAGAAGCACCCAGGACTTC CATTGCTTTGTCCCGGGGCTCCACTGAACAAGTTGGCCTGCAGTG GTGTTTTGTGGGGAGGAGGATGGGGAGTAGGACATACAGCCTT AGATTTAAGGTTTTTACTGTGAGGGATGTTTGGGAGATGTAAGAA ATGTTCTTGCAAGTAAAGGTTAGTTTACAATCAGCCACATTCTAGG TAGGGGCCCACTTACCGTACTAACCAGGGAAGCTGTCCCTCACTG TTGAATTTTCTCTAATCTCAAGGCCATATCTGTGAAATGCTGGCAT TTGCACCTACCTACAGAGTGCAATGTGAGGGTTAATGAAATATG TACATCTGGCCTTGAAACCACTTTTATTACATGGGGTCTAGAACT TGACCCCTTGAGGGTGCTTGTCCCTCTCCCTGTTGGTCCGTGGGT TGGTAGTTTCTACAGTTGGGCAGCTGGTTAGGTAGAGGGAGTTGTC AAGTCTCTGCTGGCCAGCCAAACCTGTCTGACAACCTCTTGGTG AACCTTAGTACCTAAAAGGAAATCTCACCCATCCACACCCCTGGA GGATTTCATCTCTGTATATGATGATCTGGATCCACCAAGACTTGT TTATGCTCAGGGTCAATTCTTTTTCTTTTTTTTTTTTTTTTTCTTT TTCTTTGAGACTGGGTCTCGCTTTGTTGCCAGGCTGGAGTGGAGT GGCGTGATCTTGGCTTACTGCAGCCTTGCCTCCCCGGCTCGAGCA GTCCTGCCTCAGCCTCCGGAGTAGCTGGGACCACAGGTTTCATGCCA

[0259]

	CCATGGCCAGCCAACCTTTGCATGTTTGTAGAGATGGGGTCTCAC AGTGTGCCCAGGCTGGTCTCAAACCTCTGGGCTCAGGCGATCCAC CTGTCTCAGCCTCCCAGAGTGCTGGGATTACAATTGTGAGCCACCA CGTCCAGCTGGAAGGGTCAACATCTTTACATTCTGCAAGCACATC TGCATTTTCACCCACCCCTTCCCCTCTCTCCCTTTTATATCCCAT TTTTATATCGATCTCTATTTTACAATAAACTTTGCTGCCACCTGT GTGCTGAGGGGTG
--	--

[0260]

표 14

변이체 아미노산

서열 번호	야생형 아미노산	야생형 코돈	변이체 아미노산	변이체 코돈	DNA 위치	아미노산 위치	예시적 하위서열 (소문자로 기재된 변이체 코돈)
35	F	TTT	A	GCT	282	2	CCCCTGCCCTCAACAAGATG getTGCCAACCTGGCCAA
36	F	TTT	C	TGC	282	2	CCCCTGCCCTCAACAAGATG tgcTGCCAACCTGGCCAA
37	F	TTT	D	GAT	282	2	CCCCTGCCCTCAACAAGATG gatTGCCAACCTGGCCAA
38	F	TTT	E	GAG	282	2	CCCCTGCCCTCAACAAGATG gagTGCCAACCTGGCCAA
39	F	TTT	F	TTC	282	2	CCCCTGCCCTCAACAAGATG tteTGCCAACCTGGCCAA
40	F	TTT	G	GGT	282	2	CCCCTGCCCTCAACAAGATG ggfTGCCAACCTGGCCAA
41	F	TTT	H	CAC	282	2	CCCCTGCCCTCAACAAGATG cacTGCCAACCTGGCCAA
42	F	TTT	I	ATC	282	2	CCCCTGCCCTCAACAAGATG atcTGCCAACCTGGCCAA
43	F	TTT	K	AAG	282	2	CCCCTGCCCTCAACAAGATG aagTGCCAACCTGGCCAA
44	F	TTT	L	CTG	282	2	CCCCTGCCCTCAACAAGATG ctgTGCCAACCTGGCCAA
45	F	TTT	M	ATG	282	2	CCCCTGCCCTCAACAAGATG atgTGCCAACCTGGCCAA
46	F	TTT	N	AAC	282	2	CCCCTGCCCTCAACAAGATG aacTGCCAACCTGGCCAA
47	F	TTT	P	CCT	282	2	CCCCTGCCCTCAACAAGATG cetTGCCAACCTGGCCAA
48	F	TTT	Q	CAG	282	2	CCCCTGCCCTCAACAAGATG cagTGCCAACCTGGCCAA
49	F	TTT	R	AGA	282	2	CCCCTGCCCTCAACAAGATG agaTGCCAACCTGGCCAA
50	F	TTT	S	AGC	282	2	CCCCTGCCCTCAACAAGATG agcTGCCAACCTGGCCAA
51	F	TTT	T	ACC	282	2	CCCCTGCCCTCAACAAGATG accTGCCAACCTGGCCAA
52	F	TTT	V	GTG	282	2	CCCCTGCCCTCAACAAGATG gtgTGCCAACCTGGCCAA
53	F	TTT	W	TGG	282	2	CCCCTGCCCTCAACAAGATG tggTGCCAACCTGGCCAA
54	F	TTT	Y	TAC	282	2	CCCCTGCCCTCAACAAGATG tacTGCCAACCTGGCCAA

[0261]

[0262] 실시예 20: 다수의 변이체 부위들을 가진 CDR 내의 스트레치

[0263] 단일 부위 또는 다수의 부위들에서 코돈 변이체를 코딩하는 핵산 라이브러리를 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성하는데, 이때 변이체는 각각의 위치에서 미리 선택된다. 변이체 영역은 CDR의 적어도 일부를 코딩한다. 예를 들면, 도 12를 참조한다. 합성된 핵산을 디바이스 표면으로부터 방출하고 프라이머로서 사용하여 핵산 라이브러리를 생성하고, 이 핵산 라이브러리를 세포에서 발현시켜 변이체 단백질 라이브러리를 생성한다. 에피토프에 대한 증가된 결합 친화성에 대해 변이체 항체를 평가한다.

[0264] 실시예 21: 변이체 항체 라이브러리의 생성

[0265] 상기 실시예들에 기재된 바와 같이 핵산 라이브러리를 생성한다. 도 12로부터의 대표적 CDR을 코딩하는 핵산에 대한 변이체 라이브러리를 생성한다. 대표적 CDR은 변형되는데, 이때 CDR 영역은 도 13에 나타난 바와 같이 변이될 다수의 위치들을 포함한다. 도 13에 나타난 바와 같이, 상이한 수의 코돈 변이체 및 변이체의 위치를 선택한다. 도 13에서, 생성될 수 있는 변이체 라이브러리의 다양성은 1,152이다. 차세대 시퀀싱에 의한 분석은 의도된 변이체가 우측 분획 및 우측 위치에 존재한다는 것을 입증한다.

[0266] 실시예 22: 다양한 펩타이드들을 발현시키기 위한 모듈식 플라스미드 성분

[0267] 도 14에 묘사된 바와 같이, 발현 구축물 카세트의 부분을 구성하는 별도의 영역들 각각에 대해 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 라이브러리를 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성한다. 2 구축물 발현 카세트를 생성하기 위해, 제1 프로모터(1410), 제1 오픈 리딩 프레임(1420), 제1 터미네이터(1430), 제2 프로모터(1440), 제2 오픈 리딩 프레임(1450) 또는 제2 터미네이터 서열(1460)의 변이체 서

열의 적어도 일부를 코딩하는 변이체 핵산들을 합성하였다. 증폭의 라운드 후, 앞의 실시예에 기재된 바와 같이, 1,024개의 발현 구축물들의 라이브러리가 생성된다.

[0268] **실시예 23: 다중 부위 단일 위치 변이체**

[0269] 핵산의 적어도 일부를 코딩하는 영역 내의 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 라이브러리를 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성한다. 다중 부위 단일 위치 변이체로 구성된, 핵산 변이체의 라이브러리가 생성된다. 예를 들면, 도 8b를 참조한다.

[0270] **실시예 24: 변이체 라이브러리 합성**

[0271] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행한다. 적어도 약 30,000개의 동일하지 않은 폴리뉴클레오타이드들이 드노보 합성되는데, 이때 각각의 동일하지 않은 폴리뉴클레오타이드는 아미노산 서열의 상이한 코돈 변이체를 코딩한다. 합성된 적어도 30,000개의 동일하지 않은 폴리뉴클레오타이드들은 적어도 약 30,000개의 동일하지 않은 폴리뉴클레오타이드들 각각에 대한 소정의 서열에 비해 1/1,000개 염기 미만의 총 오류율을 가진다. 상기 라이브러리는 긴 핵산의 PCR 돌연변이유발을 위해 사용되고, 적어도 약 30,000개의 동일하지 않은 변이체 폴리뉴클레오타이드들이 형성된다.

[0272] **실시예 25: 클러스터 기반 변이체 라이브러리 합성**

[0273] 실시예 2에 기재된 조건과 유사한 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행한다. 2개의 코돈 위치들에 대한 기준 핵산의 합성된 소정의 변이체를 함유하는 단일 클러스터를 디바이스에서 생성한다. 2개의 연속 코돈 위치 방식에서, 각각의 핵산의 2개의 복제물을 사용하여 2개의 위치에 대해 위치당 19개의 변이체들을 생성함으로써, 38개의 핵산들을 합성하였다. 각각의 변이체 서열은 길이가 40개 염기이다. 동일한 클러스터에서, 추가 비-변이체 핵산 서열이 생성되는데, 이때 추가 비-변이체 핵산 및 변이체 핵산은 유전자의 코딩 서열의 38개의 변이체들을 일괄적으로 코딩한다. 각각의 핵산은 또 다른 핵산에 상보적인 적어도 하나의 영역을 가진다. 클러스터 내의 핵산은 암모니아 기체 절단에 의해 방출된다. 물을 포함하는 핀(pin)은 클러스터와 접촉하고 핵산을 뽑아내고 핵산을 작은 바이알로 옮긴다. 상기 바이알은 중합효소 순환 조립(PCR) 반응을 위한 DNA 중합효소 시약도 함유한다. 핵산은 어닐링하고 갭은 연장 반응에 의해 채워지고 생성된 이중 가닥 DNA 분자가 형성됨으로써, 변이체 핵산 라이브러리가 형성된다. 임의적으로, 변이체 핵산 라이브러리를 제한 효소에 노출시킨 후 발현 벡터에 라이게이션시킨다.

[0274] **실시예 26: 단백질 결합 친화성의 변화에 대한 변이체 핵산 라이브러리의 스크리닝**

[0275] 복수의 발현 벡터들을 실시예 13 내지 16에 기재된 바와 같이 생성한다. 이 실시예에서, 발현 벡터는 HIS-태그 부착된 세균 발현 벡터이다. 벡터 라이브러리를 세균 세포 내로 전기천공한 후, HIS-태그부착된 변이체 단백질의 발현 및 정제를 위해 클론을 선택한다. 변이체 단백질을 표적 분자에 대한 결합 친화성의 변화에 대해 스크리닝한다.

[0276] 금속 이온으로 코팅된 수지(예를 들면, IDA-아가로스 또는 NTA-아가로스)를 이용하여 HIS-태그부착된 단백질을 분리하는 금속 친화성 크로마토그래피(IMAC)의 이용과 같은 방법으로 친화성을 조사한다. 일련의 히스티딘 잔기들은 특정 완충제 조건 하에서 니켈, 코발트 및 구리를 포함하는 여러 유형의 고정된 금속 이온에 결합하기 때문에 발현된 His-태그부착된 단백질을 정제할 수 있고 검출할 수 있다. 예시적 결합/세척 완충제는 10 내지 25 mM 이미다졸을 함유하는 Tris-완충제 식염수(TBS)(pH 7.2)로 구성된다. IMAC 컬럼으로부터의 포획된 HIS-태그부착된 단백질의 용출 및 회수는 고농도의 이미다졸(적어도 200 mM)(용출제), 낮은 pH(예를 들면, 0.1 M 글리신-HCl, pH 2.5) 또는 과량의 강한 킬레이터(예를 들면, EDTA)에 의해 달성된다.

[0277] 대안적으로, HIS-태그부착된 단백질을 사용하는 어세이 방법, 예컨대, HIS-태그부착된 단백질을 분리하기 위한 풀-다운 어세이 또는 HIS-태그부착된 단백질을 검출하기 위한 면역블롯팅 어세이에서 사용하기 위한 항-HIS-태그 항체는 상업적으로 입수될 수 있다.

[0278] **실시예 27: 세포 부착 및 이동의 조절제에 대한 활성의 변화에 대한 변이체 핵산 라이브러리의 스크리닝**

[0279] 실시예 13 내지 16에 기재된 바와 같이 생성된 변이체 핵산 라이브러리를 GFP-태그부착된 포유동물 발현 벡터 내로 삽입한다. 상기 라이브러리로부터 분리된 클론을 포유동물 세포 내로 일시적으로 형질감염시킨다. 대안적으로, 발현 구축물을 함유하는 세포로부터 단백질을 발현시키고 분리한 후, 추가 측정을 위해 상기 단백질을 세포에게 전달한다. 면역형광 어세이를 수행하여, GFP-태그부착된 변이체 발현 생성물의 세포 국소화의 변화를 평가한다. FACS 어세이를 수행하여, GFP-태그부착된 변이체 단백질 발현 생성물의 비-변이체 버전과 상호작용하는

막횡단 단백질의 입체구조적 상태의 변화를 평가한다. 상처 치유 어세이를 수행하여, 세포 배양 접시에서 스크래치에 의해 생성된 공간을 침습하는 GFP-태그부착된 변이체 단백질을 발현하는 세포의 능력의 변화를 평가한다. 형광 광원 및 카메라를 이용하여 GFP-태그부착된 단백질을 발현하는 세포를 확인하고 추적한다.

[0280] 실시예 28: 바이러스 진행을 억제하는 펩타이드에 대한 변이체 핵산 라이브러리의 스크리닝

[0281] 실시예 13 내지 16에 기재된 바와 같이 생성된 변이체 핵산 라이브러리를 FLAG-태그부착된 포유동물 발현 벡터 내로 삽입하고, 상기 변이체 핵산 라이브러리는 펩타이드 서열을 코딩한다. 바이러스 장애를 앓고 있는 대상체로부터 일차 포유동물 세포를 수득한다. 대안적으로, 건강한 대상체로부터의 일차 세포를 바이러스로 감염시킨다. 세포를 일련의 마이크로웰 접시들 위에 플레이팅한다. 변이체 라이브러리로부터 단리된 클론을 세포에 일시적으로 형질감염시킨다. 대안적으로, 발현 구축물을 함유하는 세포로부터 단백질을 발현시키고 단리한 후, 추가 측정을 위해 상기 단백질을 세포에게 전달한다. 세포 생존 어세이를 수행하여, 변이체 펩타이드와 관련된 향상된 생존에 대해 감염된 세포를 평가한다. 예시적 바이러스는 조류 독감, 지카 바이러스, 한타바이러스, C형 간염 및 천연두를 포함하나 이들로 한정되지 않는다.

[0282] 한 예시적 어세이는 세포에 첨가될 때 원형질막을 가로질러 확산하고 중성 레드의 약한 양이온성으로 인해 산성 라이소좀 구획에서 축적되는 중성 레드 염료를 사용하는 중성 레드 세포독성 어세이다. 바이러스에 의해 유도된 세포 변성은 막 단편화 및 라이소좀 ATP에 의해 유발된 양성자 전위 활성화의 손실로 이어진다. 세포내 중성 레드의 결과적 감소를 멀티웰 플레이트 포맷에서 분광광도계로 평가할 수 있다. 변이체 펩타이드를 발현하는 세포는 신호 획득 색채 어세이에서의 세포내 중성 레드의 증가에 의해 점수화된다. 바이러스에 의해 유도된 세포 변성을 억제하는 펩타이드에 대해 세포를 평가한다.

[0283] 실시예 29: 세포의 대사 활성을 증가시키거나 감소시키는 변이체 단백질에 대한 스크리닝

[0284] 세포의 대사 활성의 변화를 야기하는 발현 생성물을 확인하기 위해 실시예 13 내지 16에 기재된 바와 같이 복수의 발현 벡터들을 생성한다. 이 실시예에서, (예를 들면, 형질감염 또는 형질도입을 통해) 발현 벡터를 일련의 마이크로웰 접시 위에 플레이팅된 세포 내로 전달한다. 그 다음, 대사 활성의 하나 이상의 변화에 대해 세포를 스크리닝한다. 대안적으로, 발현 구축물을 함유하는 세포로부터 단백질을 발현시키고 단리한 후, 대사 활성을 측정하기 위해 상기 단백질을 세포에게 전달한다. 임의적으로, 대사 활성을 측정하기 위해 세포를 대사 활성의 하나 이상의 변화에 대해 스크리닝하기 전에 독소로 처리한다. 투여된 예시적 독소는 보툴리눔(*botulinum*) 독소 (면역학적 유형: A, B, C1, C2, D, E, F 및 G를 포함함), 스타필로코커스(*staphylococcus*) 장독소 B, 예르시니아 페스티스(*Yersinia pestis*), C형 간염, 머스타드 물질, 중금속, 시아나이드, 내독소, 바실러스 안쓰라시스(*Bacillus anthracis*), 지카 바이러스, 조류 독감, 제초제, 살충제, 수은, 유기포스페이트 및 리신을 포함하나 이들로 한정되지 않는다.

[0285] 기초 에너지 요구는 호기성 트리카복실산(TCA) 또는 크렙(Kreb)의 주기 또는 혐기성 해당작용을 수반하는 산화적 인산화에 의해 대사 기질, 예를 들면, 글루코스의 산화로부터 유도된다. 해당작용이 에너지의 주공급원일 때, 세포의 대사 활성은 세포가 대사의 산성 생성물, 예를 들면, 락테이트 및 CO₂를 배출하는 속도를 모니터링함으로써 추정될 수 있다. 호기성 대사의 경우, 세포의 산소의 소비 및 산화적 유리 라디칼의 생성은 세포의 에너지 요구를 반영한다. 세포내 산화-환원력은 NADH 및 NAD⁺의 자가형광 측정에 의해 측정될 수 있다. 세포에 의해 방출된 에너지, 예를 들면, 열의 양은 정상 환경 하에서 소비된 산소의 양(예를 들면, 4.8 kcal/l O₂)으로부터 예측될 수 있는, 대사 동안 생성되고/되거나 소비된 물질에 대한 분석 값으로부터 유도된다. 열 생성과 산소 사용 사이의 커플링은 독소에 의해 방해될 수 있다. 직접적인 마이크로열량측정은 열적으로 단리된 샘플의 온도 상승을 측정한다. 따라서, 산소 소비의 측정과 조합될 때 열량측정을 이용하여 독소의 엔커플링 활성을 검출할 수 있다.

[0286] 대사 활성의 다양한 마커들의 변화를 측정하는 다양한 방법들 및 디바이스들은 당분야에서 공지되어 있다. 예를 들면, 이러한 방법, 디바이스 및 마커는 전체로서 본원에 참고로 도입된 미국 특허 제7,704,745호에서 논의되어 있다. 요약하건대, 각각의 세포 집단에 대해 하기 특성들 중 임의의 특성의 측정치가 기록된다: 글루코스, 락테이트, CO₂, NADH 및 NAD⁺ 비, 열, O₂ 소비, 및 유리 라디칼 생성. 스크리닝된 세포는 간세포, 대식세포 또는 신경모세포종 세포를 포함할 수 있다. 스크리닝된 세포는 세포주, 대상체로부터의 일차 세포, 또는 모델 시스템(예를 들면, 마우스 모델)로부터의 세포일 수 있다.

[0287] 멀티웰 플레이트의 챔버 내에 위치한 단일 세포 또는 세포 집단의 산소 소비율의 측정을 위해 다양한 기법들을

이용할 수 있다. 예를 들면, 세포를 포함하는 챔버는 온도, 전류 또는 형광의 변화를 기록하기 위한 센서뿐만 아니라, 형광 광을 모니터링하기 위해 각각의 챔버에 커플링된 광학 시스템, 예를 들면, 섬유-커플링된 광학 시스템도 가질 수 있다. 이 실시예에서, 각각의 챔버는 챔버 내부의 분자를 여기시키기 위해 조명 공급원에 대한 윈도우(window)를 가진다. 섬유-커플링된 광학 시스템은 자가형광을 검출하여 세포내 NADH/NAD 비를 측정할 수 있고 전압 및 칼슘 민감성 염료를 검출하여 막횡단 전위 및 세포내 칼슘을 측정할 수 있다. 추가로, CO₂ 및/또는 O₂ 민감성 형광 염료 신호의 변화를 검출한다.

[0288] 실시예 30: 암 세포의 선택적 표적화에 대한 변이체 핵산 라이브러리의 스크리닝

[0289] 실시예 13 내지 16에 기재된 바와 같이 생성된 변이체 핵산 라이브러리를 FLAG-태그부착된 포유동물 발현 벡터 내로 삽입하고, 상기 변이체 핵산 라이브러리는 펩타이드 서열을 코딩한다. 변이체 라이브러리로부터 단리된 클론을 암 세포 및 비-암 세포 내로 따로 일시적으로 형질감염시킨다. 변이체 핵산에 의해 코딩된 변이체 펩타이드를 각각 발현하는 암 세포 및 비-암 세포 둘 다에 대한 세포 생존 및 세포 사멸 어세이를 수행한다. 변이체 펩타이드와 관련된 선택적 암 세포 사멸에 대해 세포를 평가한다. 임의적으로, 암 세포는 암 세포주, 또는 암으로 진단받은 대상체로부터의 일차 암 세포이다. 암으로 진단받은 대상체로부터의 일차 암 세포의 경우, 임의적으로 대상체에게 투여하기 위해 스크리닝 어세이에서 확인된 변이체 펩타이드를 선택한다. 대안적으로, 단백질 발현 구축물을 함유하는 세포로부터 단백질을 발현시키고 단리한 후, 추가 측정을 위해 상기 단백질을 암 세포 및 비-암 세포에게 전달한다.

[0290] 실시예 31: 조합 라이브러리의 생성

[0291] 실시예 2에 일반적으로 기재된 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행한다. 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 집단을 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성하는데, 이때 변이체는 각각의 위치에서 미리 선택된다. 제1 집단의 핵산을 제2 집단으로부터의 핵산과 조합하여 조합 라이브러리를 생성한다. 도 1에 나타난 바와 같이, 4개 핵산의 집단(110)을 4개 핵산의 또 다른 집단(120)과 조합하여 16개의 조합을 생성한다.

[0292] 핵산을 블런트 말단 라이게이션으로 어닐링한다. 1.5 ml 바이알 내에서 한 핵산의 50 ng의 DNA를 또 다른 핵산의 50 ng의 DNA와 혼합한다. 그 다음, 1 μ l의 T4 DNA 리가제(New England BioLabs)를 20 μ l의 라이게이션 완충제 및 20 μ l의 뉴클레아제 무함유 물과 함께 첨가한다. 그 다음, 반응 혼합물을 항온처리한다. 항온처리 후, 라이게이션 생성물을 시퀀싱으로 분석한다.

[0293] 실시예 32: 샘플링에 의한 조합 라이브러리의 생성

[0294] 실시예 2에 일반적으로 기재된 조건 하에서 드노보 폴리뉴클레오타이드 합성을 수행한다. 단일 부위 또는 다수의 부위들에서 코돈 변이를 코딩하는 핵산 집단을 실시예 4 내지 6 및 8 내지 12에 기재된 바와 같이 생성하는데, 이때 변이체는 각각의 위치에서 미리 선택된다.

[0295] 도 26a를 참조하면, 실시예 13 내지 16에 기재된 방법과 유사한 방법으로 미리 선택된 분포를 가진, 불균일한 변이체 분포를 가진 라이브러리를 생성한다. 영상의 각각의 패턴화된 부분은 각각의 위치(A1, A2, A3, B1, B2 및 B3)에서 상이한 미리 선택된 분포를 가진 4개의 상이한 아미노산들 중 1개의 아미노산을 표시한다. 흑색 원은 각각의 위치 내에서의 무작위 선택을 표시한다. 도 26b를 참조하면, A에 대한 5개의 무작위로 생성된 샘플들 및 B에 대한 5개의 무작위로 생성된 샘플들을 독립적으로 생성한다. 그 다음, 도 26c에서 확인되는 바와 같이, A에서의 5개의 무작위로 생성된 샘플들 및 B에서의 5개의 무작위로 생성된 샘플들을 예를 들면, 블런트 말단 라이게이션으로 함께 어닐링한다. 이것은 25개의 조합($n^2 = 5^2$)을 야기한다. 도 26d를 참조하면, 통계학적 비교는 생성된 분포가 미리 선택된 분포와 일치한다는 것을 입증한다.

[0296] 실시예 33: 조합 항체 라이브러리의 생성

[0297] 상기 실시예에 기재된 바와 같이 핵산 라이브러리를 생성한다. 도 27a에 나타난 바와 같이 단일 CDR 영역을 코딩하거나, 도 27b에 나타난 바와 같이 2개의 CDR 영역들을 코딩하거나, 도 27c에 나타난 바와 같이 다수의 CDR 영역들을 코딩하는 핵산에 대한 변이체 라이브러리를 생성한다.

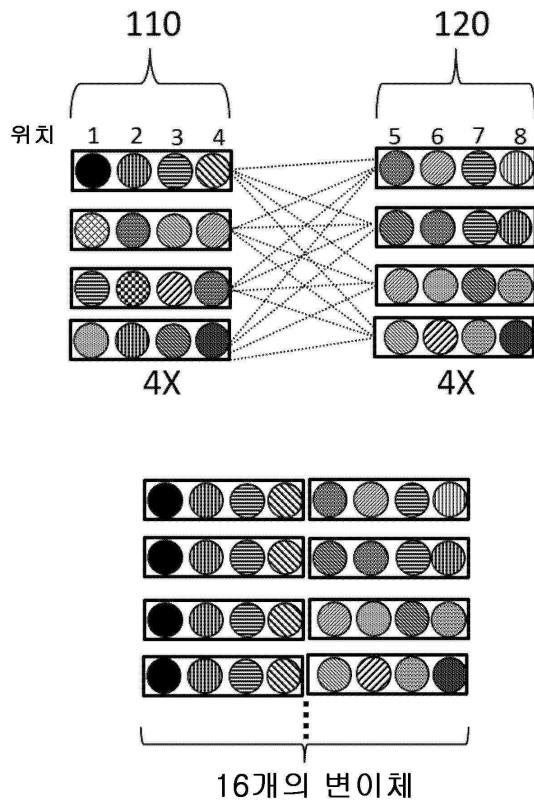
[0298] 또한, 도 28a에 나타난 바와 같이 단일 또는 다수의 중쇄 스카폴드 및 경쇄 스카폴드 내에 변이체를 포함하거나 도 28b에 나타난 바와 같이 단일 또는 다수의 프레임워크 내에 변이체를 포함하도록 변이체 항체 라이브러리를 생성한다.

[0299]

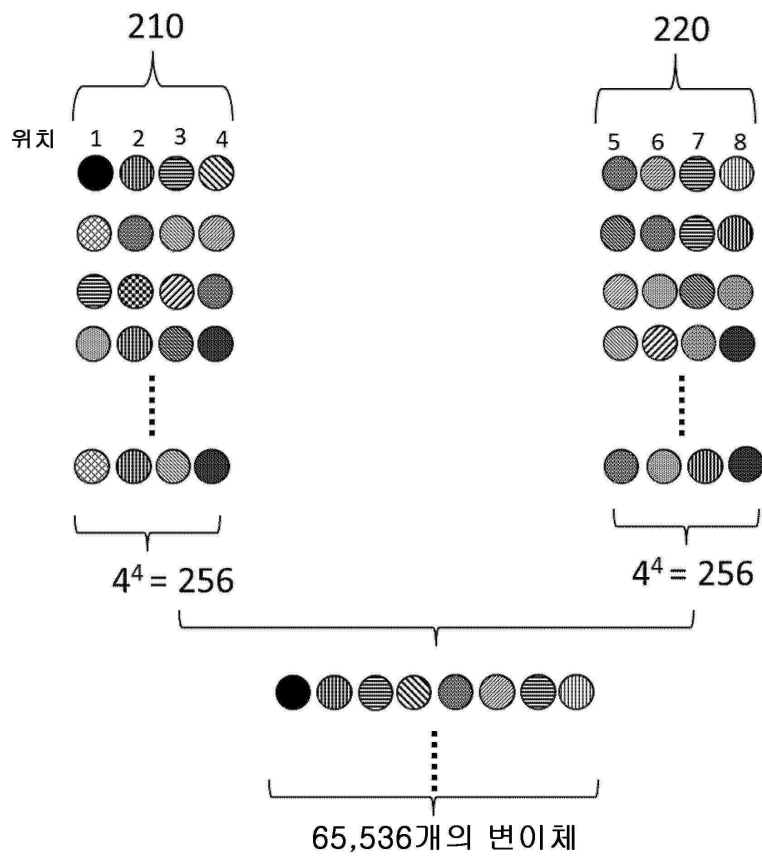
본 발명의 바람직한 실시양태가 본원에 제시되어 있고 기재되어 있지만, 이러한 실시양태가 예로써 제공된다는 것은 당분야에서 숙련된 자에게 자명할 것이다. 본 발명을 벗어나지 않으면서 다수의 변경, 변화 및 치환이 당분야에서 숙련된 자에게 인식될 것이다. 본원에 기재된 본 발명의 실시양태에 대한 다양한 대안들이 본 발명을 실시하는 데 사용될 수 있다는 것을 이해해야 한다. 하기 청구범위는 본 발명의 범위를 정의하기 위한 것이고, 이 청구범위 내의 방법 및 구조, 및 이들의 균등물은 이 청구범위에 의해 커버된다.

도면

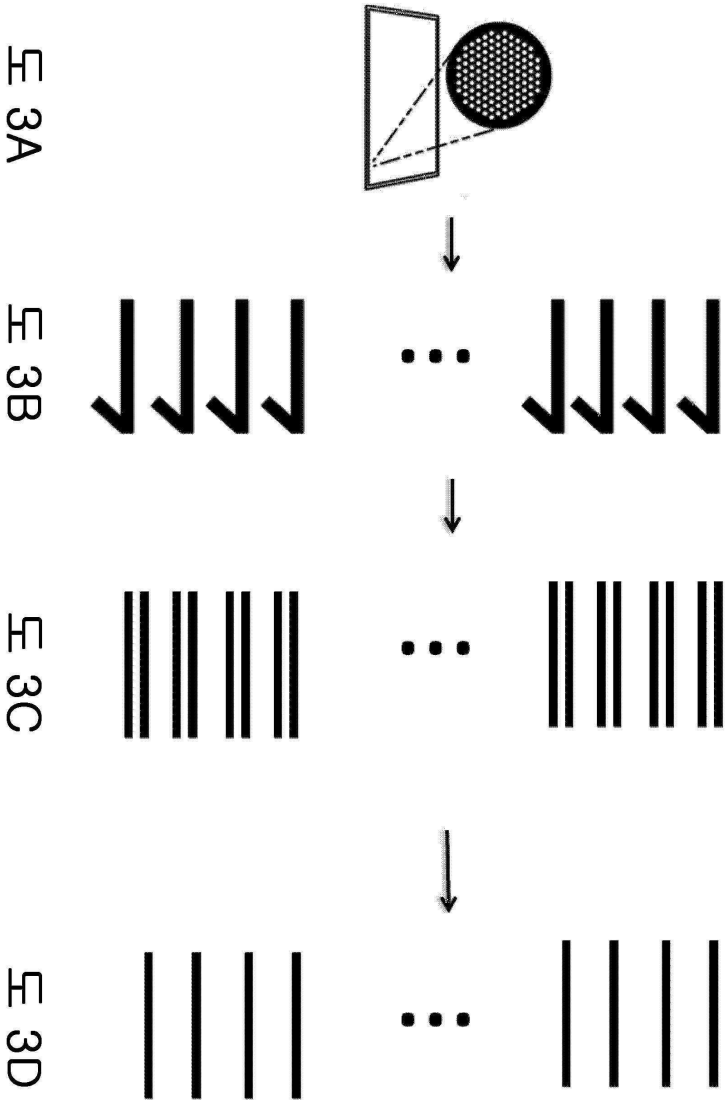
도면1



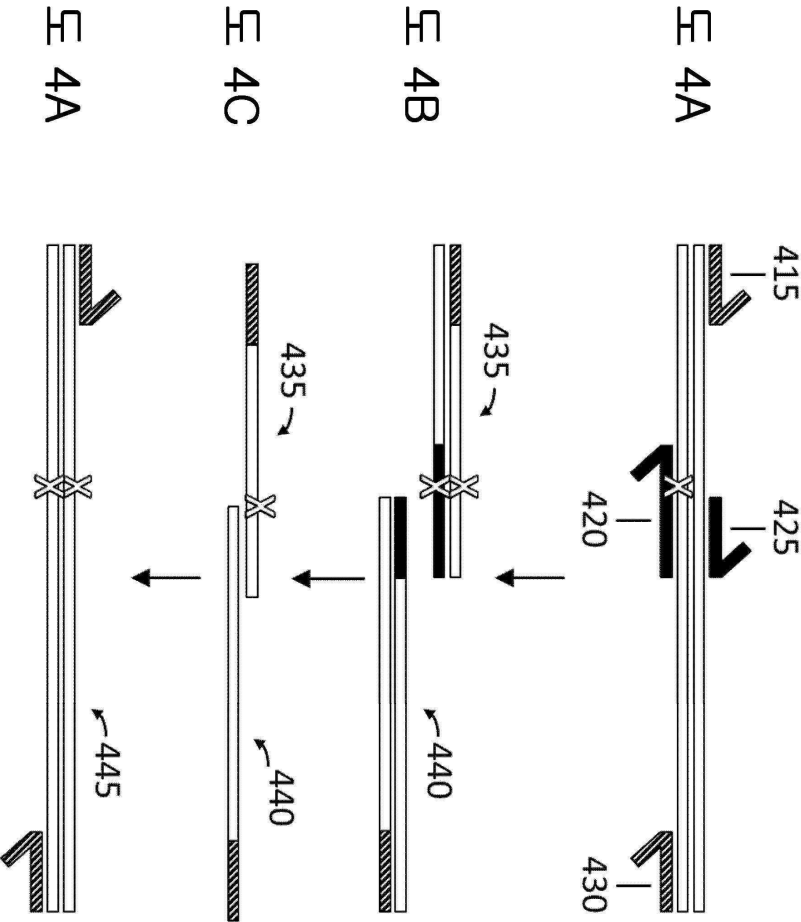
도면2



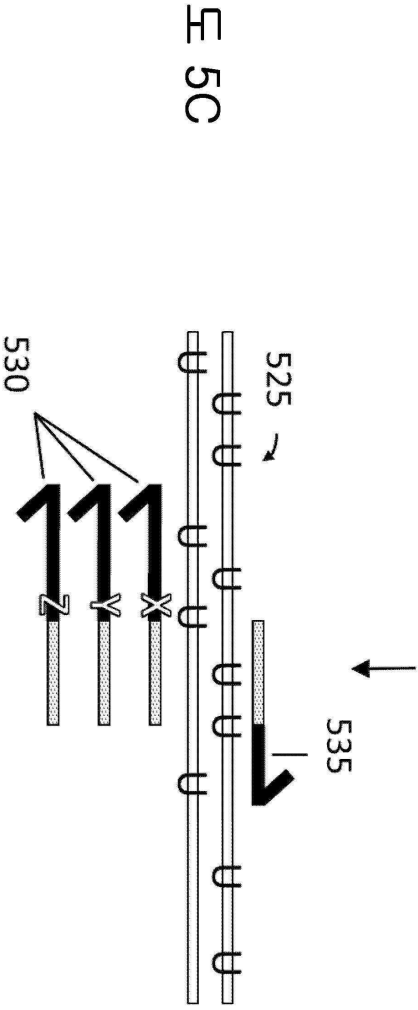
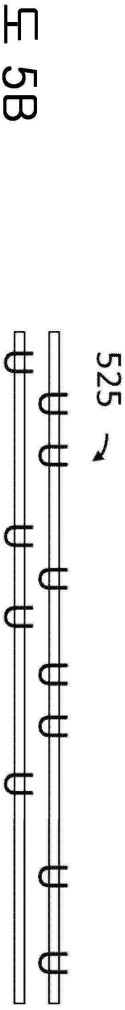
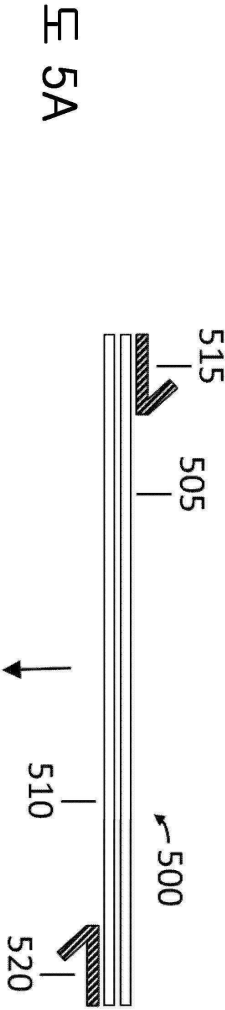
도면3



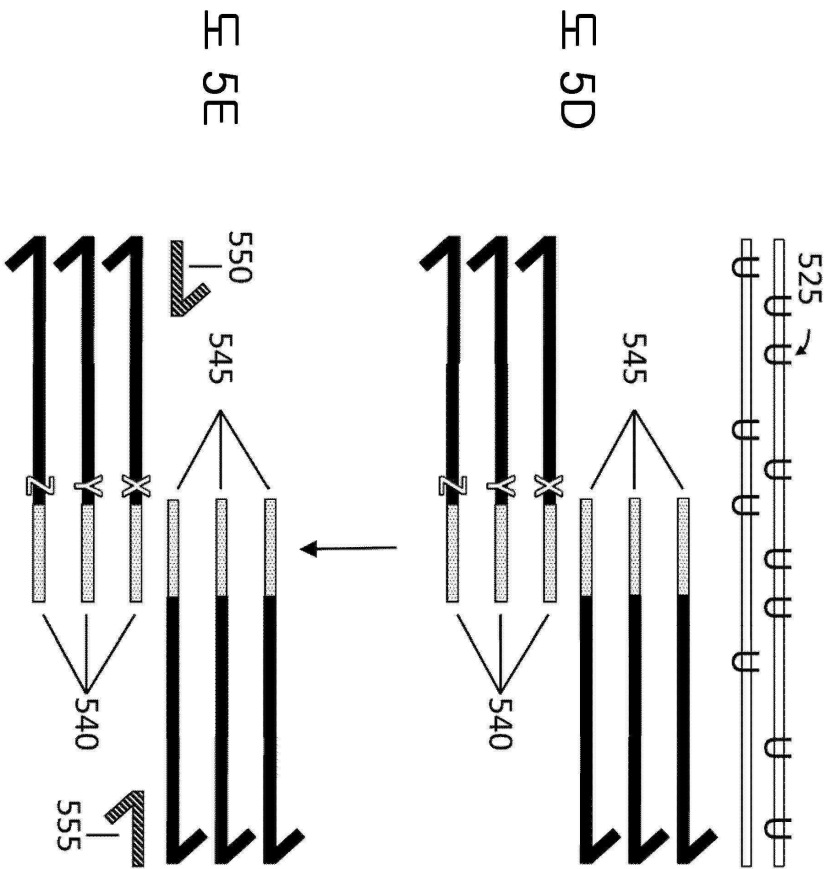
도면4



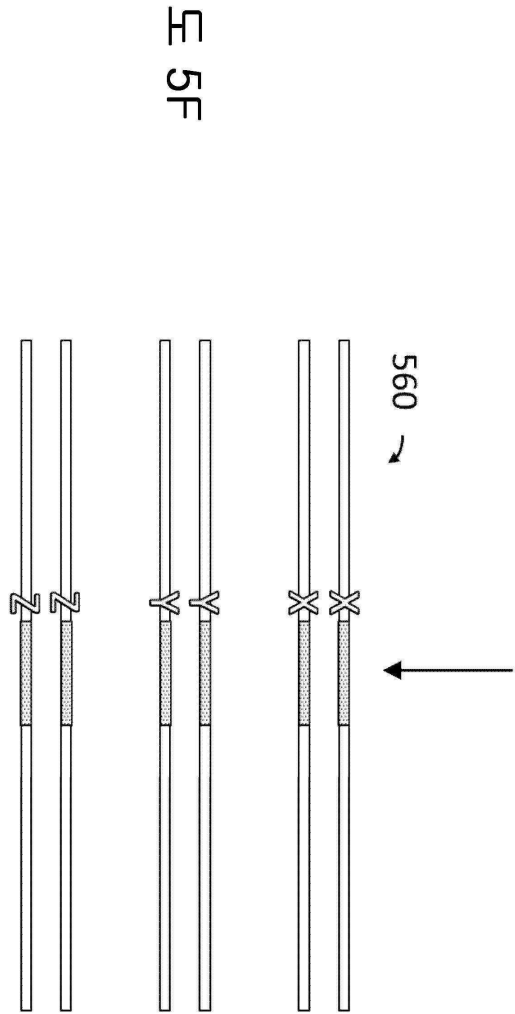
도면5a



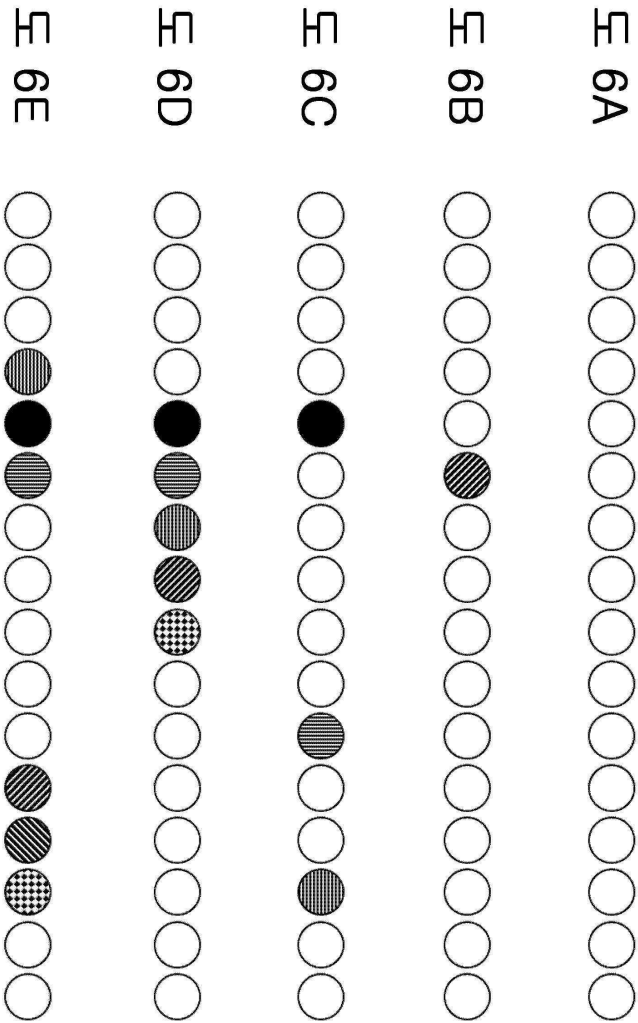
도면5d



도면5f



도면6



도면7

도 7A (A) (W) (I) (K) (R) (E) (Q)

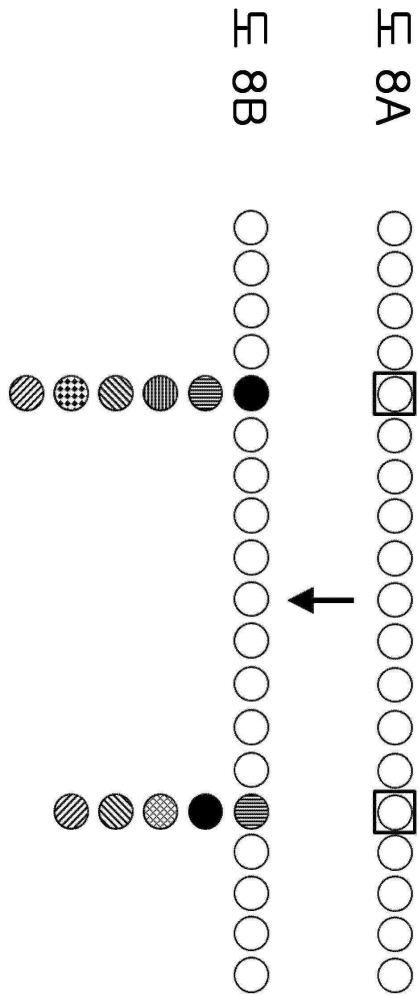


도 7B

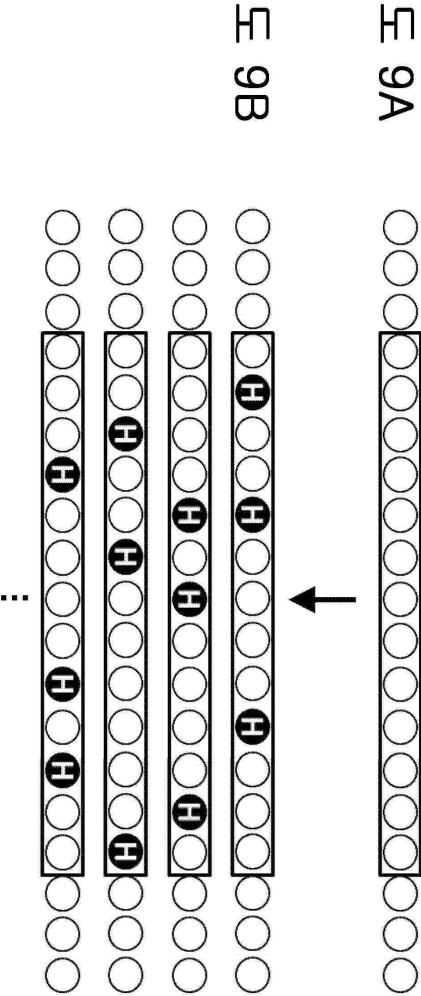
(X)	(W)	(I)	(K)	(R)	(E)	(Q)
(A)	(X)	(I)	(K)	(R)	(E)	(Q)
(A)	(W)	(X)	(K)	(R)	(E)	(Q)
(A)	(W)	(I)	(X)	(R)	(E)	(Q)
(A)	(W)	(I)	(K)	(X)	(E)	(Q)
(A)	(W)	(I)	(K)	(R)	(X)	(Q)
(A)	(W)	(I)	(K)	(R)	(E)	(X)

(X) = 임의의 코돈

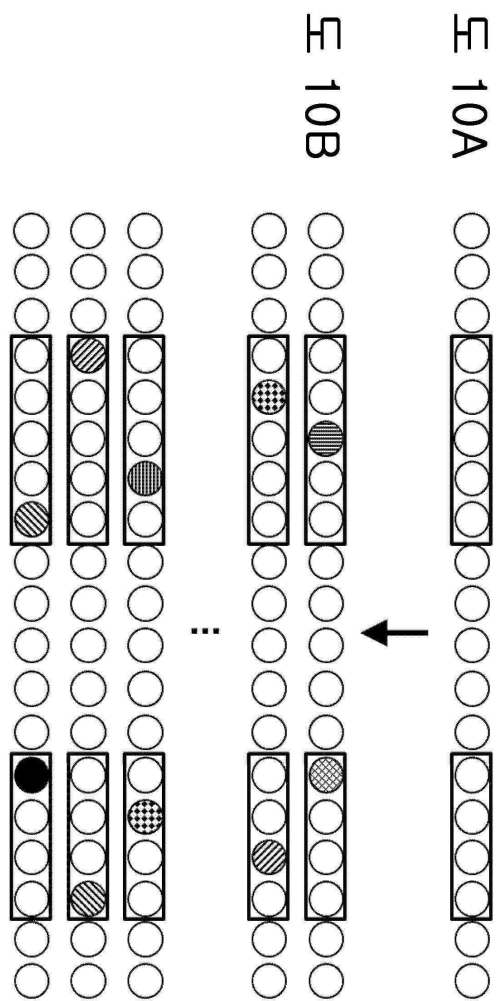
도면8



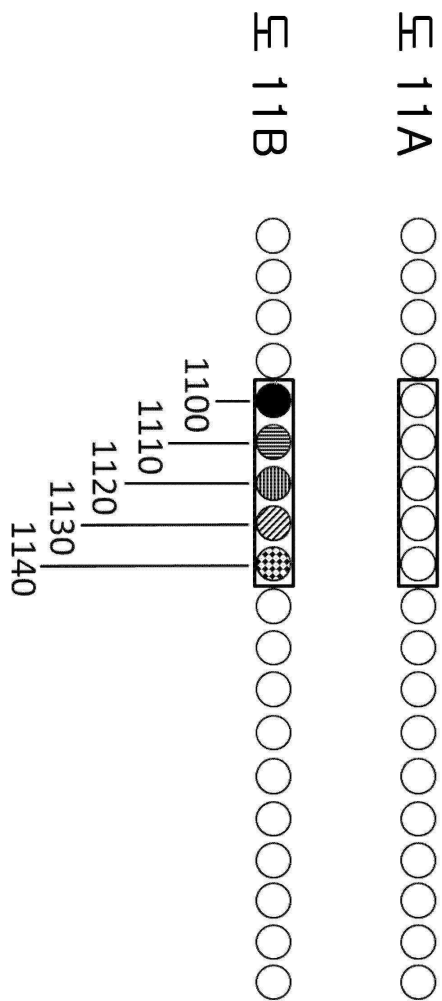
도면9



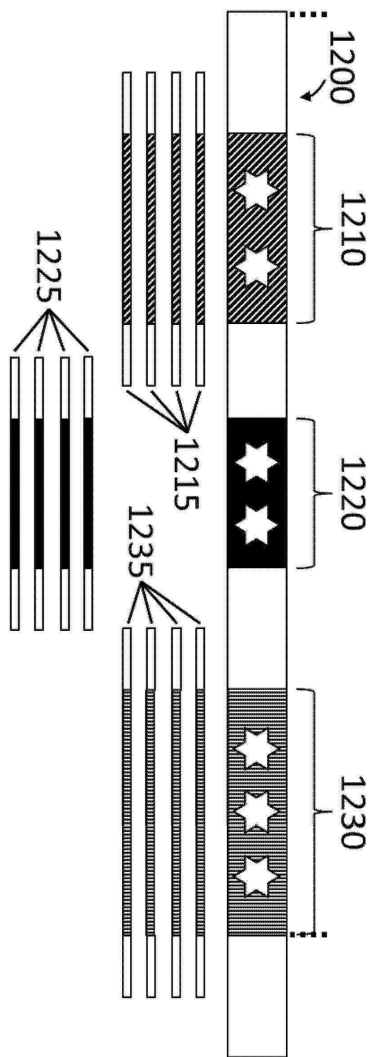
도면10



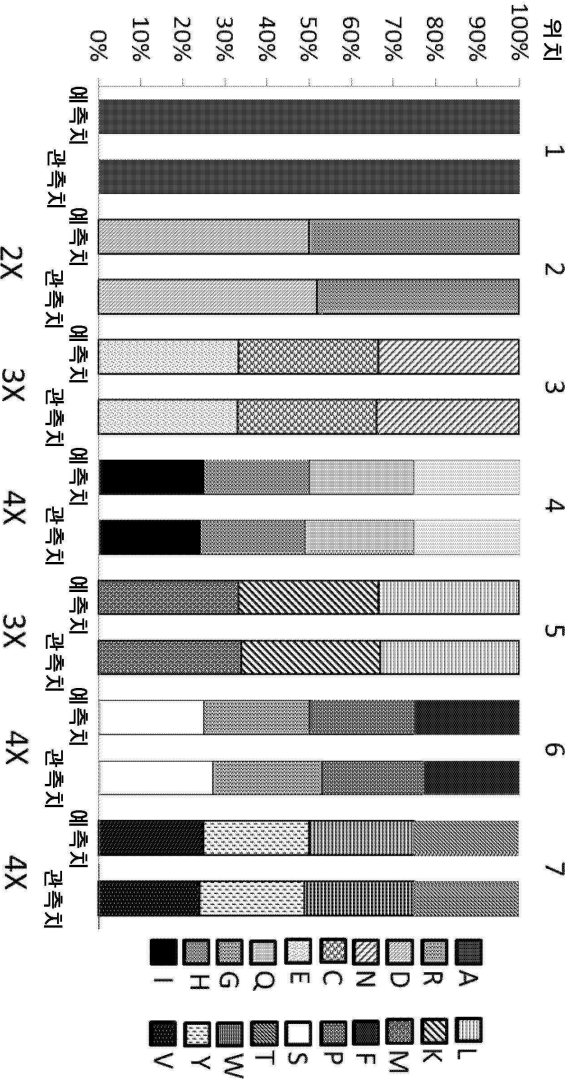
도면11



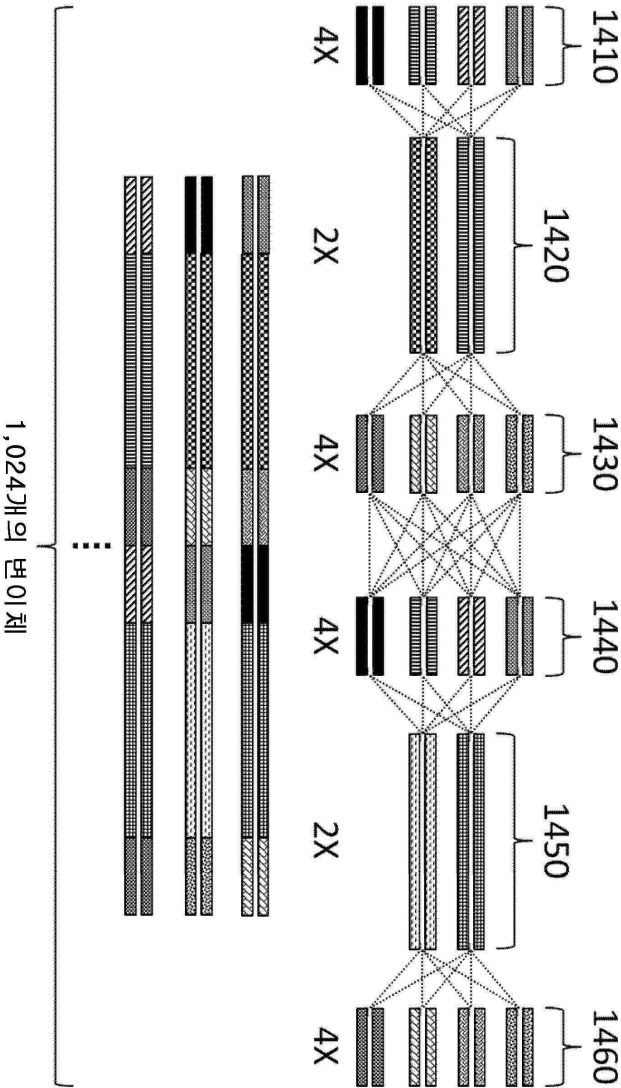
도면12



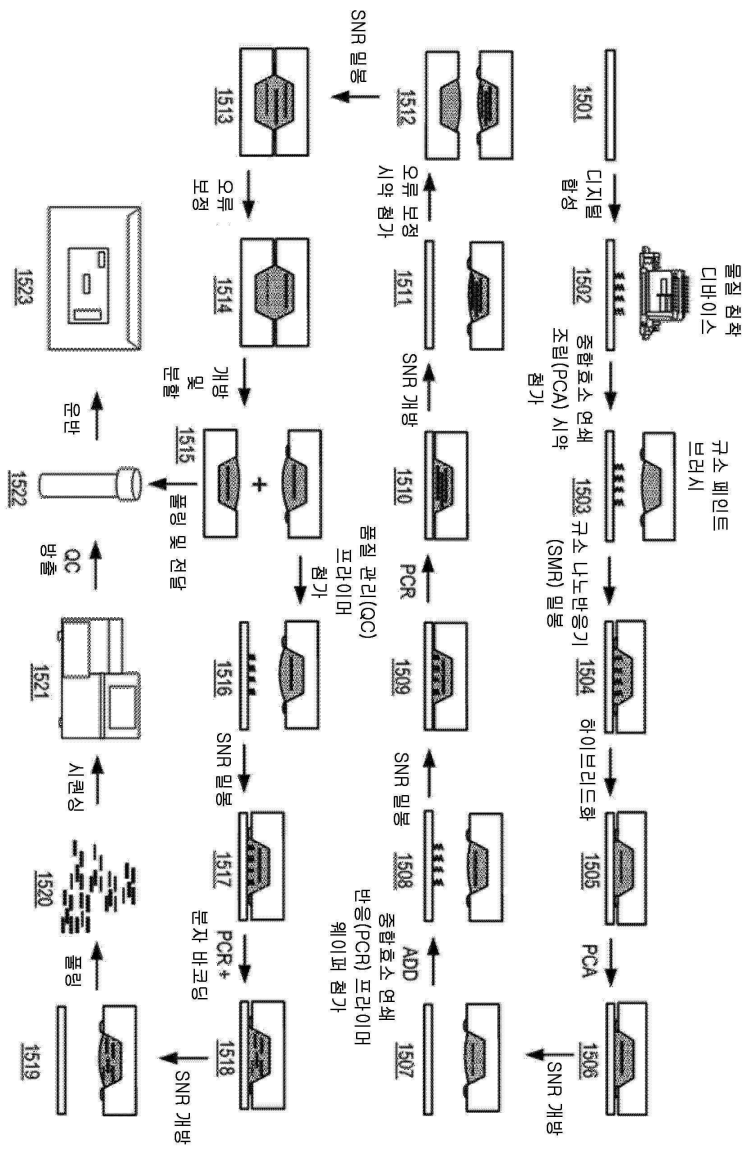
도면13



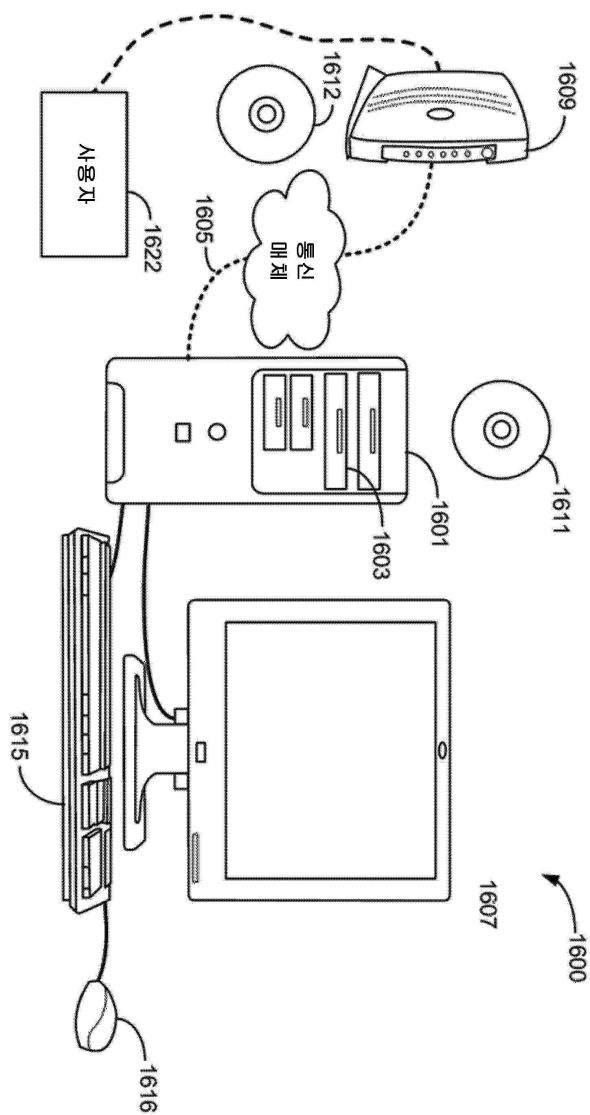
도면14



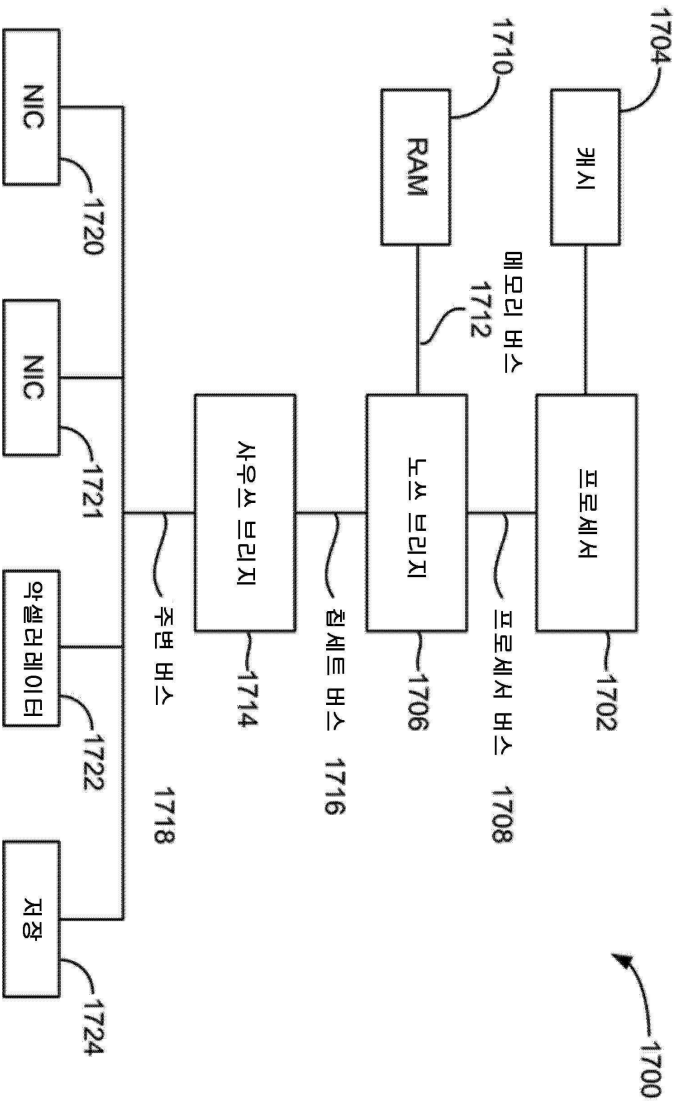
도면15



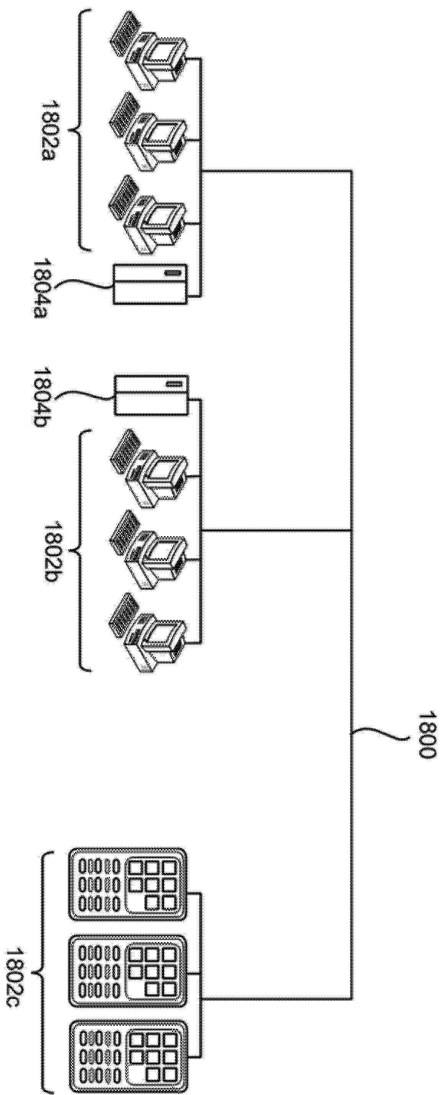
도면16



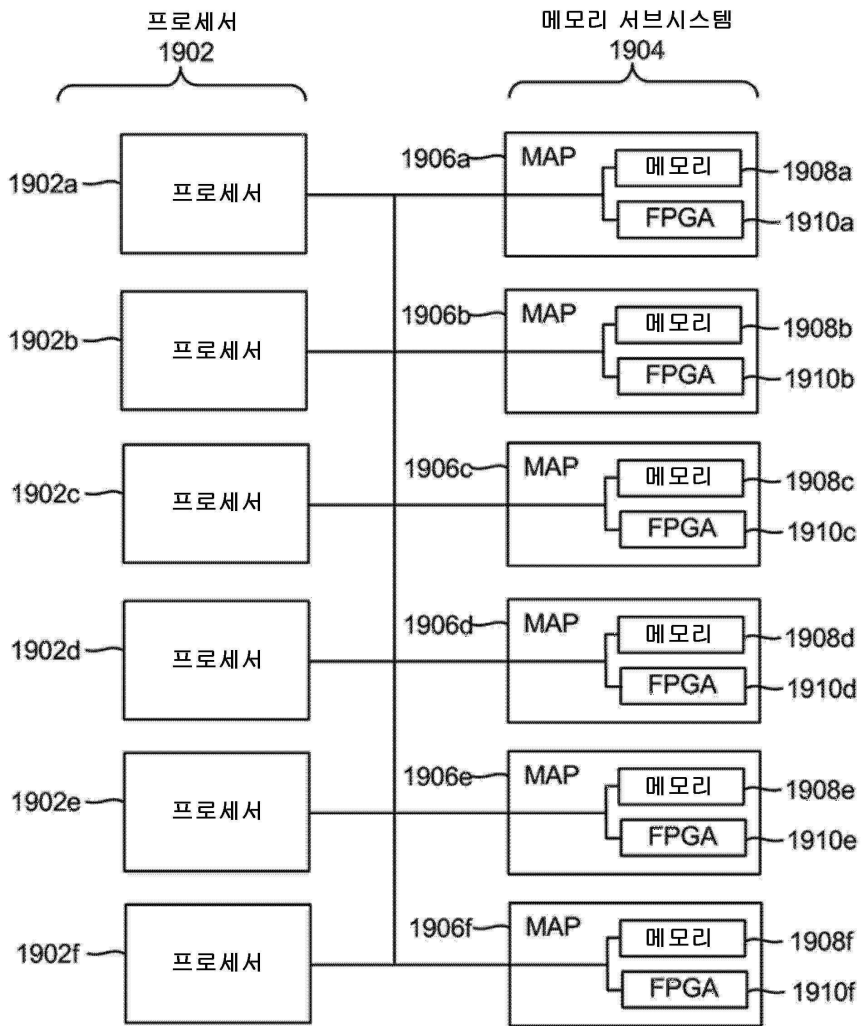
도면17



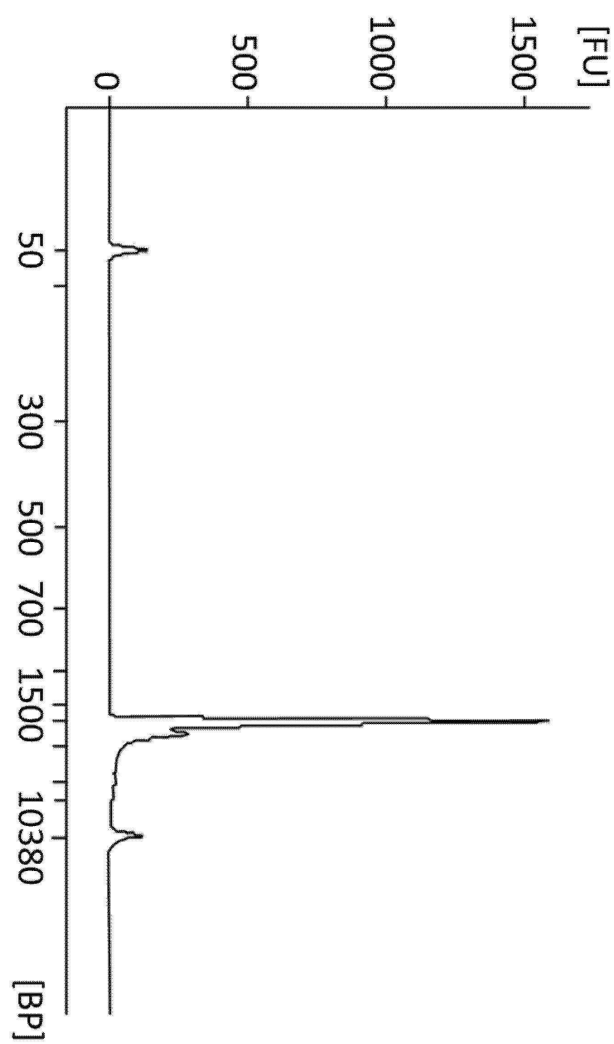
도면18



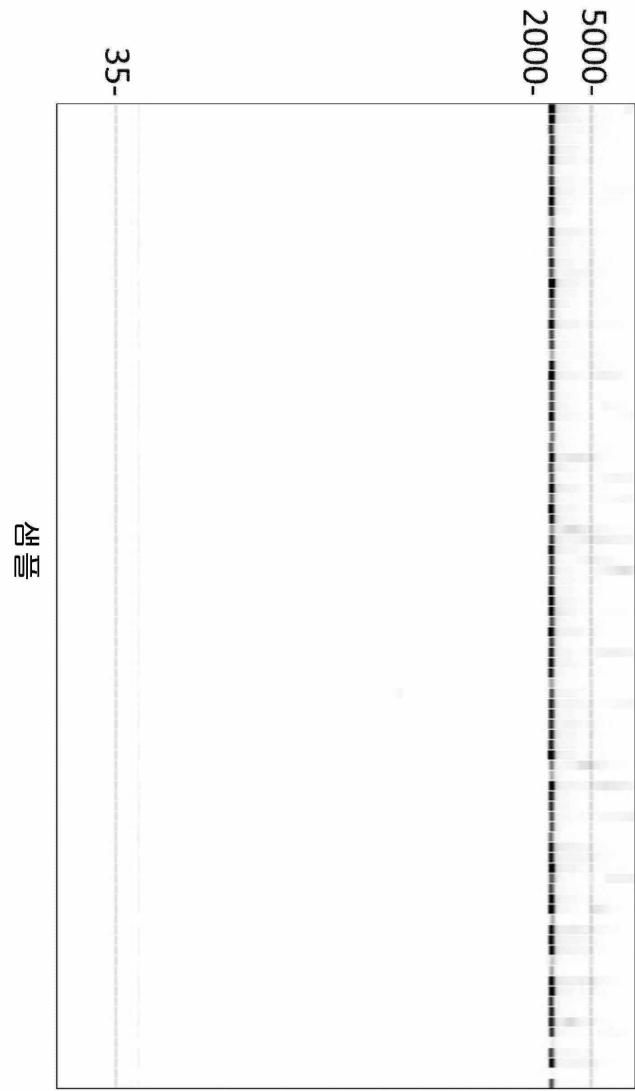
도면19



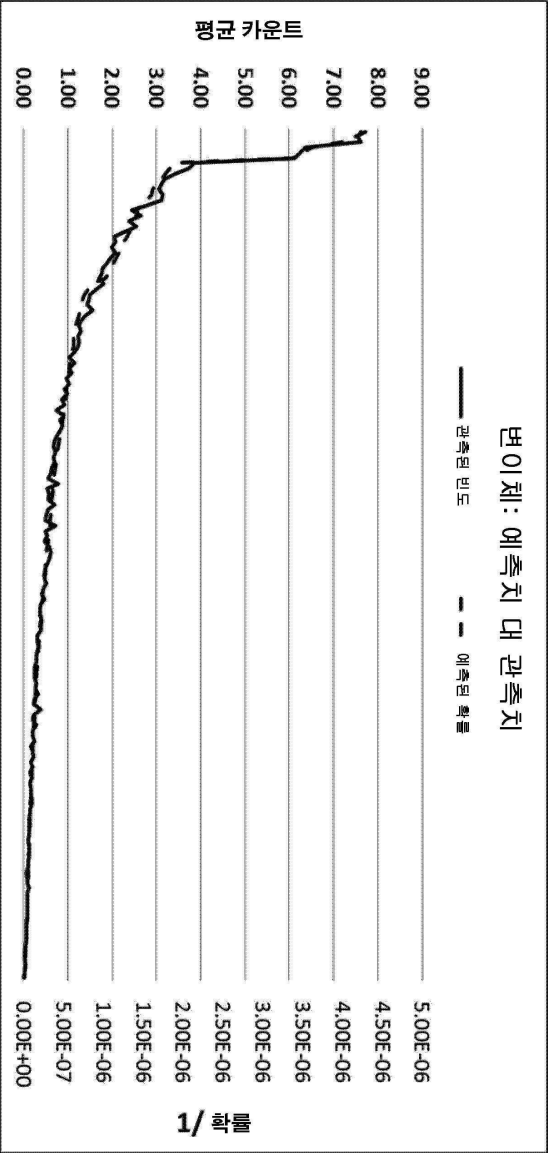
도면20



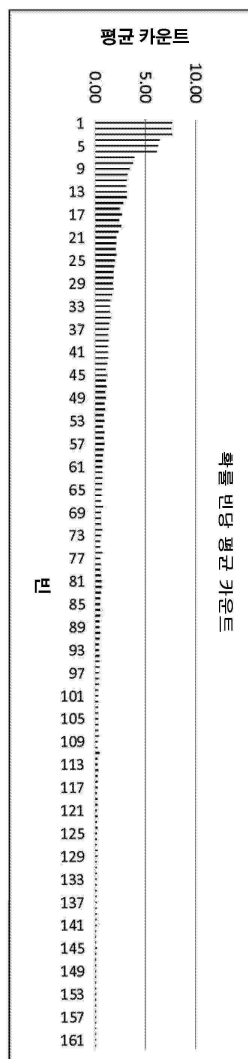
도면21



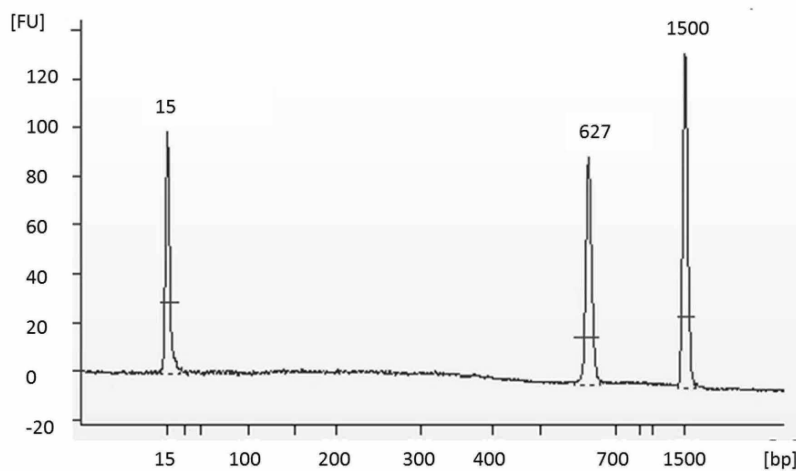
도면22



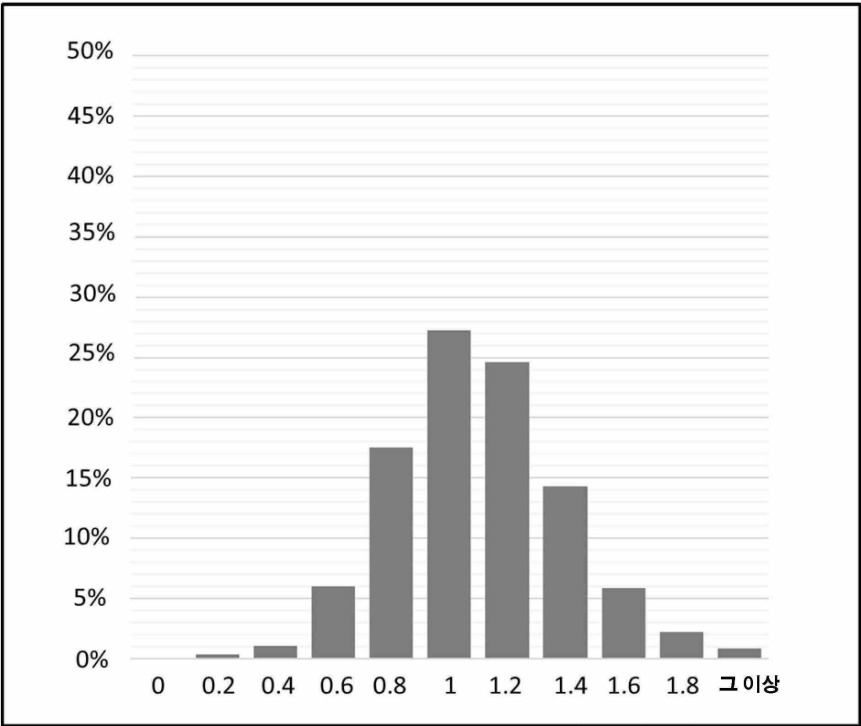
도면23



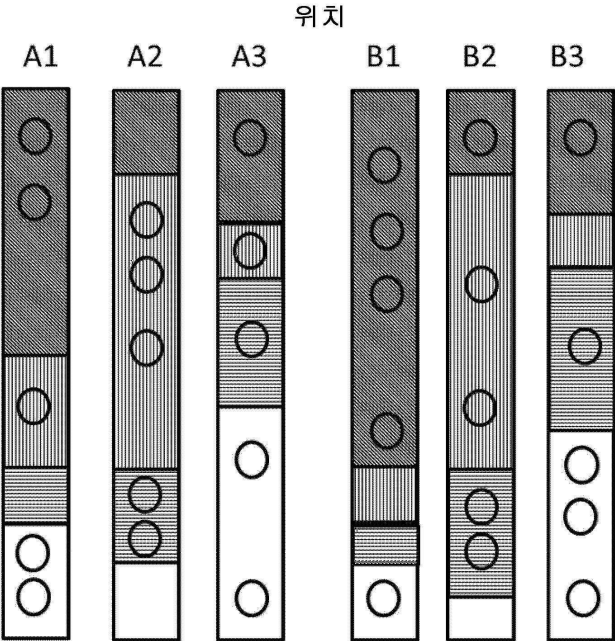
도면24



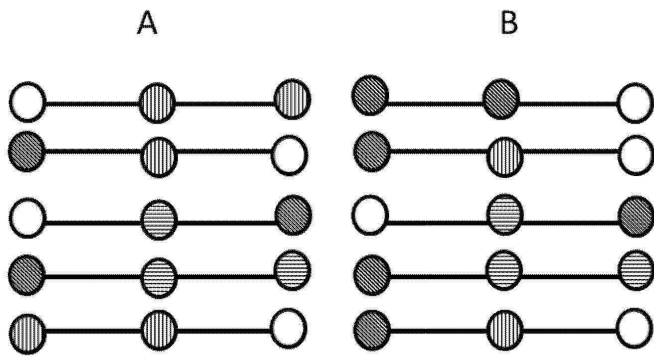
도면25



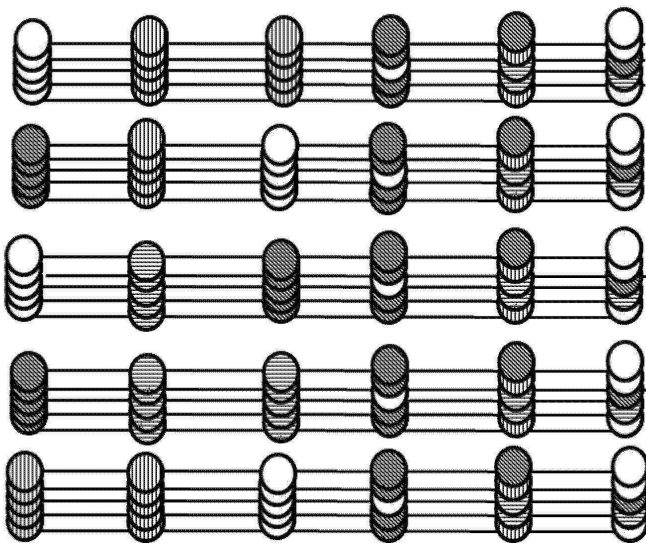
도면26a



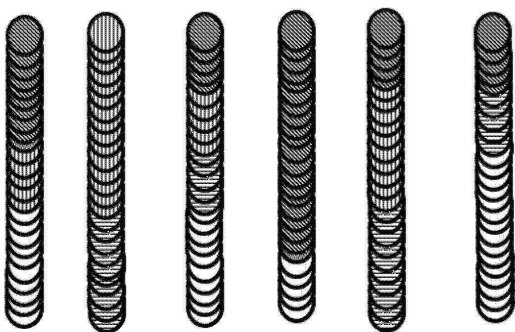
도면26b



도면26c



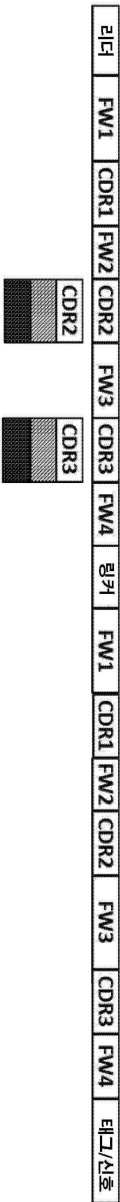
도면26d



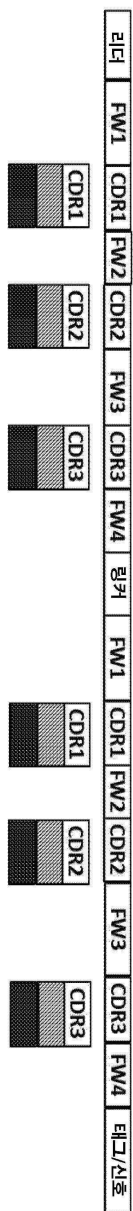
도면27a



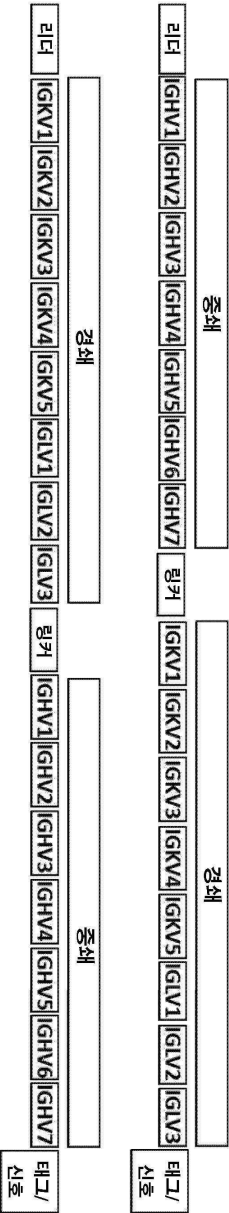
도면27b



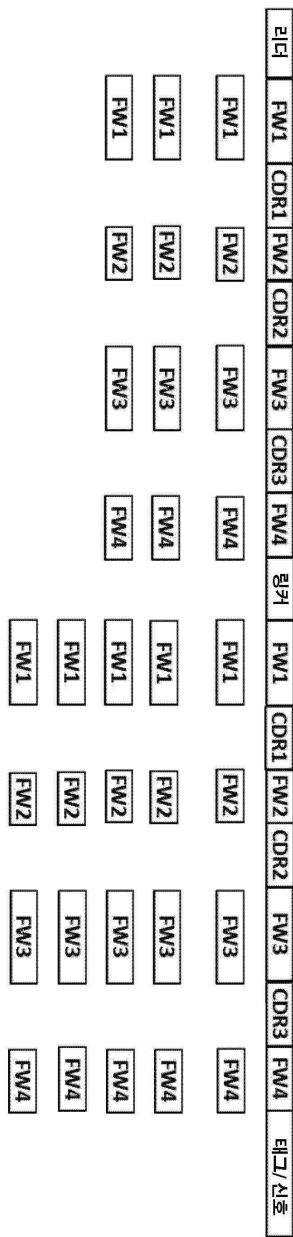
도면27c



도면28a



도면28b



서열 목록

SEQUENCE LISTING

- <110> TWIST BIOSCIENCE CORPORATION
- <120> DE NOVO SYNTHESIZED COMBINATORIAL NUCLEIC ACID LIBRARIES
- <130> 44854-729.601
- <140><141><150> 62/578,326
- <151> 2017-10-27
- <150> 62/471,723
- <151> 2017-03-15

<160> 55

<170> PatentIn version 3.5

<210> 1

<211> 44

<212> DNA

<213> Artificial Sequence

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

<400> 1

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 2

<211>

> 44

<212> DNA

<213> Artificial Sequence

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

<400> 2

atgtttagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 3

<211> 44

<212> DNA

<213> Artificial Sequence

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

<400> 3

atgttaagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 4

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 4

atgattagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 5

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 5

atgtctagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 6

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 6

atgcctagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 7

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 7

atgactagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 8

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 8

atggctagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 9
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 9
 atgtatagca agggcgagga gctgttcacc ggggtggtgc ccat 44
 <210> 10
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 10
 atgcatagca agggcgagga gctgttcacc ggggtggtgc ccat 44
 <210> 11
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 11
 atgcaaagca agggcgagga gctgttcacc ggggtggtgc ccat 44
 <210> 12
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 12
 atgaatagca agggcgagga gctgttcacc ggggtggtgc ccat 44
 <210> 13
 <211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 13

atgaaaagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 14

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 14

atggatagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 15

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 15

atggaaagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 16

<211> 44

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 16

atgttagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 17

<211> 44

<212> DNA

<213> Artificial Sequence

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

<400> 17

atgtggagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 18

<211> 44

<212> DNA

<213> Artificial Sequence

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

<400> 18

atgcgtagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 19

<211> 44

<212> DNA

<213> Artificial Sequence

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

<400> 19

atgggtagca agggcgagga gctgttcacc ggggtggtgc ccat 44

<210> 20

<211> 62

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 20

agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat tttttttttt 60

tt 62

<210> 21

<211> 112

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 21

cgggatacctt atcgatcatg tcgtacagat cccgacccat ttgctgtcca ccagtcatgc 60

tagccataacc atgatgatga tgatgatgag aaccccgcat tttttttttt tt 112

<210> 22

<211>

> 19

<212> DNA

<213> Artificial Sequence

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

primer

<400> 22

atgcgggggtt ctcatcatc 19

<210> 23

<211> 20

<212> DNA

<213> Artificial Sequence

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

primer

<400> 23

cgggatacctt atcgatcatg 20

<210> 24

<211> 7

<212> PRT

<213> Artificial Sequence

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

peptide

<400> 24

Ala Trp Ile Lys Arg Glu Gln

1 5

<210> 25

<211> 7

<212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 peptide
 <220><221> MOD_RES
 <222> (1)..(1)
 <223> Any amino acid
 <400> 25

Xaa Trp Ile Lys Arg Glu Gln

1 5

<210> 26

<211> 7

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD_RES

<222> (2)..(2)

<223> Any amino acid

<400> 26

Ala Xaa Ile Lys Arg Glu Gln

1 5

<210> 27

<211> 7

<212> PRT

<213> Artificial Sequence

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

<220><221> MOD_RES

<222> (3)..(3)

<223> Any amino acid

<400> 27

Ala Trp Xaa Lys Arg Glu Gln

1 5

<210> 28
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 peptide

<220><221> MOD_RES
 <222> (4)..(4)
 <223> Any amino acid
 <400> 28

Ala Trp Ile Xaa Arg Glu Gln

1 5

<210> 29
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 peptide

<220><221> MOD_RES
 <222> (5)..(5)
 <223> Any amino acid
 <400> 29

Ala Trp Ile Lys Xaa Glu Gln

1 5

<210> 30
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic

peptide

<220><221> MOD_RES
 <222> (6)..(6)
 <223> Any amino acid

<400> 30

Ala Trp Ile Lys Arg Xaa Gln

1 5

<210> 31

<211> 7

<212> PRT

<213> Artificial Sequence

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

peptide

<220><221> MOD_RES

<222> (7)..(7)

<223> Any amino acid

<400> 31

Ala Trp Ile Lys Arg Glu Xaa

1 5

<210> 32

<211> 6

<212> PRT

<213> Artificial Sequence

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

6xHis tag

<400> 32

His His His His His His

1 5

<210> 33

<211> 261

<212> PRT

<213> Homo sapiens

<400> 33

Met Phe Cys Gln Leu Ala Lys Thr Cys Pro Val Gln Leu Trp Val Asp

1 5 10 15

Ser Thr Pro Pro Pro Gly Thr Arg Val Arg Ala Met Ala Ile Tyr Lys

20 25 30

Gln Ser Gln His Met Thr Glu Val Val Arg Arg Cys Pro His His Glu

35 40 45
 Arg Cys Ser Asp Ser Asp Gly Leu Ala Pro Pro Gln His Leu Ile Arg
 50 55 60
 Val Glu Gly Asn Leu Arg Val Glu Tyr Leu Asp Asp Arg Asn Thr Phe
 65 70 75 80
 Arg His Ser Val Val Val Pro Tyr Glu Pro Pro Glu Val Gly Ser Asp
 85 90 95
 Cys Thr Thr Ile His Tyr Asn Tyr Met Cys Asn Ser Ser Cys Met Gly
 100 105 110

 Gly Met Asn Arg Arg Pro Ile Leu Thr Ile Ile Thr Leu Glu Asp Ser
 115 120 125
 Ser Gly Asn Leu Leu Gly Arg Asn Ser Phe Glu Val Arg Val Cys Ala
 130 135 140
 Cys Pro Gly Arg Asp Arg Arg Thr Glu Glu Glu Asn Leu Arg Lys Lys
 145 150 155 160
 Gly Glu Pro His His Glu Leu Pro Pro Gly Ser Thr Lys Arg Ala Leu
 165 170 175

 Pro Asn Asn Thr Ser Ser Ser Pro Gln Pro Lys Lys Lys Pro Leu Asp
 180 185 190
 Gly Glu Tyr Phe Thr Leu Gln Ile Arg Gly Arg Glu Arg Phe Glu Met
 195 200 205
 Phe Arg Glu Leu Asn Glu Ala Leu Glu Leu Lys Asp Ala Gln Ala Gly
 210 215 220
 Lys Glu Pro Gly Gly Ser Arg Ala His Ser Ser His Leu Lys Ser Lys
 225 230 235 240

 Lys Gly Gln Ser Thr Ser Arg His Lys Lys Leu Met Phe Lys Thr Glu
 245 250 255
 Gly Pro Asp Ser Asp
 260

<210> 34

<211> 2271

<212> DNA

<213> Homo sapiens

<400> 34

tgaggccagg agatggaggc tgcagtgagc tgtgatcaca ccactgtgct ccagcctgag	60
tgacagagca agacctatc tcaaaaaaaaa aaaaaaaaaa gaaaagctcc tgaggtgtag	120
acgccaaetc tctctagctc gctagtgggt tgcaggaggt gcttacgcat gtttgtttct	180
ttgctgccgt cticcagttg ctttatctgt tcacttgtgc cctgacttcc aactctgtct	240
ccttcctctt cctacagtac tccccgccc tcaacaagat gttttgcaa ctggccaaga	300
cctgccctgt gcagctgtgg gttgattcca ccccccgcc cggcacccgc gtccgcgcca	360
tggccatcta caagcagtc cagcacatga cggaggttgt gaggcgctgc cccaccatg	420
agcgtgctc agatagcgt ggtctggccc ctctcagca tcttatccga gtggaaggaa	480
atttgcgtgt ggagtatttg gatgacagaa acacttttcg acatagtgtg gtggtgccct	540
atgagccgcc tgaggttggc tctgactgta ccaccatcca ctacaactac atgtgtaaca	600
gttcttgcgt gggcggcatg aaccggaggc ccatcctcac catcatcaca ctggaagact	660
ccagtggtaa tctactggga cggaacagct ttgaggtgcg tgtttgtgcc tgtcctggga	720
gagaccggcg cacagaggaa gagaatctcc gcaagaaagg ggagcctcac cagagctgc	780
ccccaggag cactaagcga gcactgccc acaacaccag ctctctccc cagccaaaga	840
agaaaccact ggatggagaa tatttcaccc ttcagatccg tgggcgtgag cgcttcgaga	900
tgttccgaga gctgaatgag gccttggaa tcaaggatgc ccaggctggg aaggagccag	960
gggggagcag ggctcactcc agccacctga agtccaaaaa gggtcagtct acctcccgc	1020
ataaaaaact catgttcaag acagaagggc ctgactcaga ctgacattct ccacttcttg	1080
ttcccactg acagcctccc acccccatct ctccctcccc tgccattttg gtttttgggt	1140
ctttgaacct ttgcttgaac taggtgtgcg tcagaagcac ccaggacttc catttgcttt	1200
gtcccggggc tccactgaac aagtttgcct gcactggtgt tttgttgtgg ggaggaggat	1260
ggggagtagg acataccagc ttagatttta aggtttttac tgtgagggat gtttgggaga	1320
tgtaaagaaat gtctttgcag ttaagggtta gtttacaatc agccacattc taggtagggg	1380
cccatttcac cgtactaacc aggggaagctg tccctcactg ttgaatttcc tctaacttca	1440
aggcccatat ctgtgaaatg ctggcatttg cacctacctc acagagtgc tttgagggt	1500
taatgaaata atgtacatct ggccttgaac ccacctttta ttacatgggg tctagaactt	1560
gaccccttg aggtgtcttg ttccctctcc ctgttggtcg gtgggttgggt agtttctaca	1620
gttgggcagc tggtaggta gagggagtgt tcaagtctct gctggcccag ccaaaccctg	1680

tctgacaacc tcttggtgaa ccttagtacc taaaaggaaa tctcacccca tcccacaccc 1740
 tggaggattt catctcttgt atatgatgat ctggatccac caagacttgt tttatgctca 1800
 ggggtcaattt cttttttctt tttttttttt ttttttcttt tttcttgaga ctgggtctcg 1860
 ctttgttgcc caggctggag tggagtggcg tgatcttggc ttactgcagc ctttgcctcc 1920

ccggctcgag cagtctgcc tcagcctccg gagtagctgg gaccacaggt tcatgccacc 1980
 atggccagcc aacttttgca tgttttgtag agatggggtc tcacagtgtt gccaggtg 2040
 gtctcaaact cctgggctca ggcgatccac ctgtctcagc ctcccagagt gctgggatta 2100
 caattgtgag ccaccacgtc cagctggaag ggtcaacatc ttttacatc tgcaagcaca 2160
 tctgcatttt caccacccc ttcctctcct tctccctttt tataatccat ttttatatcg 2220
 atctcttatt ttacaataaa actttgctgc cacctgtgtg tctgaggggt g 2271

<210> 35

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 35

cccctgccct caacaagatg gcttgccaac tggccaa 37

<210> 36

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 36

cccctgccct caacaagatg tgctgccaac tggccaa 37

<210> 37

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 37
 ccctgccct caacaagatg gattgccaac tggccaa 37
 <210> 38
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 38
 ccctgccct caacaagatg gagtgccaac tggccaa 37
 <210> 39
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 39
 ccctgccct caacaagatg ttctgccaac tggccaa 37
 <210> 40
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 40
 ccctgccct caacaagatg gggtgccaac tggccaa 37
 <210> 41
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 41
 ccctgccct caacaagatg cactgccaac tggccaa 37

<210> 42

<211> 37

<212> DNA

<213> Artificial Sequence

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

<400> 42

cccctgccct caacaagatg atctgccaac tggccaa 37

<210> 43

<211> 37

<212> DNA

<213> Artificial Sequence

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

<400> 43

cccctgccct caacaagatg aagtgccaac tggccaa 37

<210> 44

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 44

cccctgccct caacaagatg ctgtgccaac tggccaa 37

<210> 45

<211> 37

<212> DNA

<213> Artificial Sequence

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

<400> 45

cccctgccct caacaagatg atgtgccaac tggccaa 37

<210> 46

<211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 46
 cccctgccct caacaagatg aactgccaac tggccaa 37

<210> 47
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 47
 cccctgccct caacaagatg ccttgccaac tggccaa 37

<210> 48
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 48
 cccctgccct caacaagatg cagtgccaac tggccaa 37

<210> 49
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 49
 cccctgccct caacaagatg agatgccaac tggccaa 37

<210> 50
 <211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 50

cccctgccct caacaagatg agctgccaac tggccaa 37

<210> 51

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 51

cccctgccct caacaagatg acctgccaac tggccaa 37

<210> 52

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 52

cccctgccct caacaagatg gtgtgccaac tggccaa 37

<210> 53

<211> 37

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 53

cccctgccct caacaagatg tggcgccaac tggccaa 37

<210> 54

<211> 37

<212> DNA

<213> Artificial Sequence

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

<400> 54

cccctgccct caacaagatg tactgccaac tggccaa

37

<210> 55

<211> 10

<212> PRT

<213> Artificial Sequence

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

<400> 55

His His His Cys Cys His His Cys His His

1

5

10