



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년04월29일
(11) 등록번호 10-2105352
(24) 등록일자 2020년04월22일

(51) 국제특허분류(Int. Cl.)
C12N 9/50 (2006.01) B01D 15/38 (2006.01)
C07K 1/22 (2006.01) C12N 9/00 (2006.01)
(52) CPC특허분류
C12N 9/50 (2013.01)
B01D 15/3804 (2013.01)
(21) 출원번호 10-2019-7018450(분할)
(22) 출원일자(국제) 2015년10월23일
심사청구일자 2019년06월26일
(85) 번역문제출일자 2019년06월26일
(65) 공개번호 10-2019-0077620
(43) 공개일자 2019년07월03일
(62) 원출원 특허 10-2018-7023540
원출원일자(국제) 2015년10월23일
심사청구일자 2018년08월16일
(86) 국제출원번호 PCT/US2015/057125
(87) 국제공개번호 WO 2016/073228
국제공개일자 2016년05월12일
(30) 우선권주장
62/074,494 2014년11월03일 미국(US)
62/209,010 2015년08월24일 미국(US)
(56) 선행기술조사문헌
KR1020140069313 A*
W02013045632 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
메르크 파텐트 게엠베하
독일 64293 다름스타트 프랑크푸르터 스트라세 250
(72) 발명자
질만, 마틴
미국 01821 매사추세츠주 빌러리카 콘코드 로드 290 이엠티 밀리포어 코퍼레이션 내
올란드, 조
미국 01821 매사추세츠주 빌러리카 콘코드 로드 290 이엠티 밀리포어 코퍼레이션 내
(74) 대리인
양영준, 김영

전체 청구항 수 : 총 24 항

심사관 : 최준호

(54) 발명의 명칭 **생체분자를 정제하기 위한 가용성 인테인 융합 단백질 및 방법**

(57) 요약

본 발명은 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 융합 단백질 및 그러한 융합 단백질을 포함하는 친화성 크로마토그래피 매트릭스, 및 이의 이용 방법에 관한 것이다.

(52) CPC특허분류

C07K 1/22 (2013.01)

C12N 9/93 (2013.01)

C07K 2319/92 (2013.01)

명세서

청구범위

청구항 1

C-인테인 폴리펩티드를 포함하는 제2 용합 단백질과 회합시킴으로써 활성 인테인 복합체를 형성할 수 있는 N-인테인 폴리펩티드를 포함하는 가용성 용합 단백질로서, 여기서 N-인테인 폴리펩티드는 SEQ ID NO: 1 내지 8, 29 내지 35 또는 42 내지 53에 기재된 아미노산 서열을 포함하고, C-인테인 폴리펩티드는 SEQ ID NO: 9, 10, 36 내지 41 또는 54 내지 60에 기재된 아미노산 서열을 포함하며, N-인테인 폴리펩티드는 고체 지지체에 공유적으로 부착된 것인 가용성 용합 단백질.

청구항 2

제1항에 있어서, 고체 지지체가 비드, 중공사, 고휘 섬유, 패드, 겔, 멤브레인, 카세트, 컬럼, 칩, 슬라이드, 플레이트 또는 단일체(monolith)인 가용성 용합 단백질.

청구항 3

제1항에 있어서, 고체 지지체가 크로마토그래피 수지인 가용성 용합 단백질.

청구항 4

고체 지지체에 공유적으로 부착된 N-인테인 폴리펩티드를 포함하며, 여기서 N-인테인 폴리펩티드는 SEQ ID NO: 1 내지 8, 29 내지 35 또는 42 내지 53에 기재된 아미노산 서열을 포함하는 것인 친화성 크로마토그래피 매트릭스.

청구항 5

제4항에 있어서, 고체 지지체가 크로마토그래피 수지인 친화성 크로마토그래피 매트릭스.

청구항 6

제5항에 있어서, 크로마토그래피 수지가 친수성 폴리비닐 에테르 베이스를 포함하는 것인 친화성 크로마토그래피 매트릭스.

청구항 7

제4항 내지 제6항 중 어느 한 항에 있어서, 고체 지지체가 비드, 중공사, 고휘 섬유, 패드, 겔, 멤브레인, 카세트, 컬럼, 칩, 슬라이드, 플레이트 또는 단일체인 친화성 크로마토그래피 매트릭스.

청구항 8

제7항에 있어서, 고체 지지체가 자성 비드인 친화성 크로마토그래피 매트릭스.

청구항 9

제4항 내지 제6항 중 어느 한 항에 있어서, 고체 지지체가 기공 제어 유리, 실리카, 산화지르코늄, 산화티탄, 아가로스, 폴리메타크릴레이트, 폴리아크릴레이트, 폴리아크릴아미드, 폴리비닐 알코올 또는 폴리스티렌을 포함하는 것인 친화성 크로마토그래피 매트릭스.

청구항 10

- a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하고, 여기서 C-인테인 폴리펩티드는 SEQ ID NO: 9, 10, 36 내지 41 또는 54 내지 60에 기재된 아미노산 서열을 포함하는 것인 제1 용합 단백질; 및
- b) 고체 지지체에 공유적으로 부착된 N-인테인 폴리펩티드를 포함하고, 여기서 N-인테인 폴리펩티드는 SEQ ID NO: 1 내지 8, 29 내지 35 또는 42 내지 53에 기재된 아미노산 서열을 포함하는 것인 친화성 크로마토그래피 매트릭스

를 포함하는, 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 11

제10항에 있어서, N-인테인 폴리펩티드가, 펩티드 결합에 의해 연결된 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 제2 용합 단백질 내에 존재하고, 여기서 N-인테인 가용화 파트너는 SEQ ID NO: 11 내지 28에 기재된 아미노산 서열을 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 12

제10항 또는 제11항에 있어서, 고체 지지체가 크로마토그래피 수지인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 13

제12항에 있어서, 크로마토그래피 수지가 친수성 폴리비닐 에테르 베이스를 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 14

제10항 또는 제11항에 있어서, 고체 지지체가 비드, 중공사, 고휘 섬유, 패드, 겔, 멤브레인, 카세트, 컬럼, 칩, 슬라이드, 플레이트 또는 단일체인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 15

제14항에 있어서, 고체 지지체가 자성 비드인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 16

제10항 또는 제11항에 있어서, 고체 지지체가 기공 제어 유리, 실리카, 산화지르코늄, 산화티탄, 아가로스, 폴리메타크릴레이트, 폴리아크릴레이트, 폴리아크릴아미드, 폴리비닐 알코올 또는 폴리스티렌을 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 17

제11항에 있어서, 친화성 크로마토그래피 매트릭스가 용합 단백질과 고체 지지체 사이의 스페이서 분자를 추가로 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 18

제11항에 있어서, 제2 용합 단백질이 N-인테인 가용화 파트너 내의 하나의 부위에서 고체 지지체에 부착되고, 여기서 N-인테인 가용화 파트너는 SEQ ID NO: 11 내지 28에 기재된 아미노산 서열을 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 19

제11항에 있어서, 제2 용합 단백질이 N-인테인 가용화 파트너 내의 하나 초과 부위에서 고체 지지체에 부착되고, 여기서 N-인테인 가용화 파트너는 SEQ ID NO: 11 내지 28에 기재된 아미노산 서열을 포함하는 것인 분할형 인테인-기반의 친화성 크로마토그래피 시스템.

청구항 20

- a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 용합 단백질을 함유하는 시료를 제공하는 단계;
- b) 상기 시료를, 제1 용합 단백질 내의 C-인테인 폴리펩티드가 N-인테인 폴리펩티드에 선택적으로 결합하여 비활성 인테인 복합체를 형성하는 조건하에서, 제1항의 N-인테인을 포함하는 친화성 크로마토그래피 매트릭스 또는 제4항의 친화성 크로마토그래피 매트릭스와 접촉시키는 단계;
- c) 상기 비활성 인테인 복합체를 함유하는 친화성 크로마토그래피 매트릭스를 세척하여 미결합 오염물질을 제거

하는 단계;

d) 상기 인테인 복합체를, 인테인 복합체가 활성이고 표적 분자를 C-인테인 폴리펩티드로부터 절단하는 조건에 노출시키는 단계; 및

e) 상기 절단된 표적 분자를 회수하는 단계

를 포함하는, 시료 내의 표적 분자의 친화성 정제 방법.

청구항 21

제20항에 있어서, 표적 분자가 치료적 표적에 대한 모노클로날 항체인 방법.

청구항 22

제20항에 있어서, 후속 사용을 위해 친화성 크로마토그래피 매트릭스를 세정, 재생 또는 보관하는 단계를 추가로 포함하는 방법.

청구항 23

a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 융합 단백질을, 제1 융합 단백질 내의 C-인테인 폴리펩티드가 N-인테인 폴리펩티드에 선택적으로 결합하여 인테인 복합체를 형성하는 조건하에서, 제1항의 N-인테인과 접촉시키는 단계; 및

b) 상기 표적 분자가 인테인 활성을 지지하는 조건하에서 C-인테인 폴리펩티드로부터 절단되었는지 여부를 결정하는 단계로서, 절단된 표적 분자의 존재가 촉매적 활성 인테인 복합체를 나타내는 것인 단계

를 포함하는, 친화성 정제에 이용하기에 적합한 인테인 복합체에 대한 스크리닝 방법.

청구항 24

a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 융합 단백질을 함유하는 시료를 제공하는 단계;

b) 상기 시료를, 제1 융합 단백질 내의 C-인테인 폴리펩티드가 N-인테인 폴리펩티드에 선택적으로 결합하여 비활성 인테인 복합체를 형성하는 조건하에서, N-인테인을 포함하는 친화성 크로마토그래피 매트릭스와 접촉시키는 단계;

c) 상기 비활성 인테인 복합체를 함유하는 친화성 크로마토그래피 매트릭스를 세척하여 미결합 오염물질을 제거하는 단계;

d) 상기 인테인 복합체를, 인테인 복합체가 활성이고 표적 분자를 C-인테인 폴리펩티드로부터 절단하는 조건에 노출시키는 단계; 및

e) 상기 절단된 표적 분자를 회수하는 단계

를 포함하는, 제10항의 분할형 인테인-기반의 친화성 크로마토그래피 시스템을 이용한 표적 분자의 정제 방법.

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

발명의 설명

기술 분야

- [0001] 관련 출원
- [0002] 본 출원은 2014년 11월 3일자 출원된 미국 가출원 제62/074,494호 및 2015년 8월 24일자 출원된 미국 가출원 제 62/209,010호의 이익을 주장한다. 상기 출원의 전체 교시내용은 본원에 참조로 포함된다.
- [0003] 기술분야
- [0004] 본 발명은 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 융합 단백질, 그러한 융합 단백질을 포함하는 친화성 크로마토그래피 매트릭스, 및 특히 산업적 규모의 단백질 정제 및 펩티드 라이케이션 공정에서의 그러한 융합 단백질의 이용 방법에 관한 것이다.

배경 기술

- [0005] 친화성 태그를 사용하여 관심 단백질을 태깅하는 것을 포함하는 단백질 정제 방법은 연구 개발 응용을 위한 실험실 환경에서 널리 사용되지만, 대규모 제조 작업에 비효율적인 것으로 입증되었다. 생물학적 처리 산업에서, 절단가능한 친화성 태그만을 사용하여, 최종 생성물이 태그를 함유하지 않는 것을 보장하며, 태그는 전형적으로 위치-특이적 프로테아제를 사용하여 생성 동안 제거되어야 한다. 친화성 태그의 제거는 추가의 처리 단계를 필요로 하며, 이는 특히 산업적 규모에서 비용 및 시간을 실질적으로 증가시킨다. 더욱이, 비효율적인, 표적-외 절단은 각각 생물학적 처리 응용에서 허용가능하지 않은 태그 및 절단된 단백질 단편을 보유하는 단백질로의 최종 단백질 생성물의 오염을 야기한다.
- [0006] 따라서, 산업적 조건하에서 단백질의 대규모 정제를 가능하게 하는 개선된 친화성 크로마토그래피 시약 및 방법을 개발하는 것이 필요하다.

발명의 내용

- [0007] 발명의 요약
- [0008] 인테인은 프로테아제 및 리가아제 활성 둘 모두를 함유하는 자가촉매 효소의 부류이다. "분할형 인테인"으로 지칭되는 하나의 부류의 인테인은 선택적으로, 그리고 매우 단단하게 회합하여, 활성 인테인 효소를 형성하는 N-인테인 및 C-인테인으로 지칭되는 2개의 상보적 절반 인테인을 포함한다(문헌[Shah N.H., *et al.*, *J. Amer. Chem. Soc.* 135: 18673-18681]; 문헌[Dassa B., *et al.*, *Nucl. Acids Res.*, 37:2560-2573 (2009)]).
- [0009] 대규모 단백질 정제 공정에서의 분할형 인테인을 포함하는 인테인의 용도는 이전에 종래 분야에 설명되어 있다(예를 들어, WO 2013/045632호 참조). 또한, 미정제 혼합물로부터의 관심 단백질의 크로마토그래피 분리를 위한 분할형 인테인의 용도도 이전에 설명되어 있다(예를 들어, 중국 공개 제CN101884910호; 문헌[Guan D., *et al.*, *Biotech. Bioeng.* 110:2471-2481 (2013)]; 문헌[Lu W., *et al.*, *J. Chrom. A*, 1218: 2553-2560(2011)] 참조).
- [0010] 그러나, 대규모 단백질 정제 공정에서의 인테인의 이용은 통상의 발현 시스템, 예를 들어, 에스케리키아 콜라이(*E. coli*)에서 발현되는 경우 그들의 불량한 가용성에 의해 저해된다. 추가로, 효율적인 산업적 규모의 단백질 정제 공정에 중요한, 고체 지지체에 공유적으로 부착된 인테인-기반의 친화성 리간드를 포함하는 크로마토그래피 매트릭스는 설명된 적이 없다.
- [0011] 본 발명은 C-인테인 폴리펩티드를 포함하는 제2 융합 단백질과 회합시킴으로써 활성 인테인 복합체를 형성할 수 있는 N-인테인 폴리펩티드를 포함하는 가용성 융합 단백질을 제공한다. N-인테인 폴리펩티드를 포함하는 융합 단백질을 고체 지지체에 공유적으로 부착시켜, 대규모 생물학적 처리 응용에 적합한 친화성 크로마토그래피 매트릭스를 생성할 수 있다.
- [0012] 따라서, 일 구현예에서, 본 발명은 펩티드 결합에 의해 연결되는 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 융합 단백질에 관한 것이다. 이러한 구현예의 특정 양태에서, N-인테인 가용화 파트너는 약 15 kDa 미만의 분자량, 약 60 미만의 지방족 지수 값 및 -1 미만의 그랜드 평균 소수성(Grand Average Hydropathy) 값을 가지며, N-인테인 폴리펩티드의 가용성을 증진(예를 들어, 증가 및/또는 촉진)시킨다. 이러한 구현예의 추가의 양태에서, N-인테인 가용화 파트너는 SEQ ID NO: 15를 포함한다. 이러한 구현예의 또 다른 양태에서, N-인테인 폴리펩티드는 GP41-1 N-인테인(SEQ ID NO: 1) 또는 그의 변이체이다.
- [0013] 다른 구현예에서, 본 발명은 고체 지지체에 부착된, N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 융합 단백질을 포함하는 친화성 크로마토그래피 매트릭스에 관한 것이다. 이러한 구현예의 특정 양태에서, 고체 지지체는 친수성 폴리비닐 에테르 베이스를 포함하는 크로마토그래피 수지이다.

[0014] 추가의 구현예에서, 본 발명은 시료 중 표적 분자의 친화성 정제 방법에 관한 것이다. 이러한 구현예의 일 양태에서, 당해 방법은 a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 용합 단백질질을 함유하는 시료를 제공하는 단계; b) 시료를 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여, 비활성 인테인 복합체를 형성하는 조건하에서 제2 용합 단백질질을 포함하는 친화성 크로마토그래피 매트릭스와 접촉시키는 단계로서, 제2 용합 단백질이 펩티드 결합에 의해 N-인테인 폴리펩티드의 가용성을 증진시키는 N-인테인 가용화 파트너에 연결된 N-인테인 폴리펩티드를 포함하는 단계; c) 비활성 인테인 복합체를 함유하는 친화성 크로마토그래피 매트릭스를 세척하여, 미결합 오염물질들을 제거하는 단계; d) 인테인 복합체를 인테인 복합체가 활성이고, 표적 분자를 C-인테인 폴리펩티드로부터 절단하는 조건에 노출시키는 단계; 및 e) 절단된 표적 분자를 회수하는 단계를 포함한다.

[0015] 또 다른 구현예에서, 본 발명은 친화성 정제에 사용하기에 적합한 촉매적 활성 인테인 복합체에 대한 스크리닝 방법에 관한 것이다. 이러한 구현예의 일 양태에서, 당해 방법은 a) (예를 들어, 펩티드 결합 또는 링커 분자에 의해) 표적 분자에 연결되는 C-인테인 폴리펩티드를 포함하는 제1 용합 단백질, 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여 인테인 복합체를 형성하는 조건하에서, (예를 들어, 펩티드 결합 또는 링커 분자에 의해) N-인테인 가용화 파트너에 연결되는 N-인테인 폴리펩티드를 포함하는 제2 용합 단백질과 접촉시키는 단계; 및 b) 표적 분자가 인테인 활성을 지지하는 조건하에서 C-인테인 폴리펩티드로부터 절단되는지 여부를 결정하는 단계로서, 절단된 표적 분자의 존재가 촉매적 활성 인테인 복합체를 나타내는 단계를 포함한다.

[0016] 본 발명의 N-인테인 용합 단백질은 개선된 가용성 및 증진된 촉매 활성을 가지며, 상응하는 C-인테인과 짝지어지는 경우, 대규모 단백질 정제(예를 들어, 친화성 크로마토그래피) 및 변형 공정(예를 들어, 펩티드 절단 및 라이제이션 반응)을 수행하기 위한 시약으로서 유용하다.

도면의 간단한 설명

[0017] 특허 또는 출원 파일은 컬러로 작도된 적어도 하나의 도면을 포함한다. 컬러 도면(들)을 지닌 이러한 특허 또는 특허 출원 공보의 사본은 요청 및 필요한 수수료의 지불 시에 특허청에 의해 제공될 것이다.

도 1은 본 발명의 예시적인 친화성 정제 방법을 도시한 개략도이다. 당해 방법은 고체 지지체(표면)에 부착된 N-인테인 가용화 파트너에 용합된 N-인테인 폴리펩티드를 갖는 용합 단백질질을 포함하는 본 발명의 예시적인 친화성 크로마토그래피 매트릭스를 사용한다. 친화성 크로마토그래피 매트릭스 내의 N-인테인에 상보적인 C-인테인을 포함하는 제2 용합 단백질은 정제될 표적 단백질(관심 단백질) 및 발현에 필요한 임의의 다른 요소, 예를 들어, 분비 신호에 용합된다. 도 1a는 N-인테인 친화성 크로마토그래피 매트릭스로의 C-인테인 용합 단백질의 결합 이전의 다양한 성분을 보여준다. 도 1b는 인테인 회합에 적절한 조건(예를 들어, pH, 염, 산화/환원)하에 N-인테인 친화성 크로마토그래피 매트릭스에 결합된 C-인테인 용합 단백질을 보여준다. 도 1c는 N- 및 C-엑스테인(extein)이 인테인 복합체의 촉매 활성에 적절한 조건하에서 그들의 각각의 용합 단백질로부터 절단된 후의 성분을 보여준다.

도 2A는 GP41-1 N-인테인의 N-말단(SOLP-NINT) 또는 C-말단(NINT-SOLP) 중 어느 하나에 용합된 3개의 후보 N-인테인 가용화 파트너(46, 206, 246; 표 2 참조)에 대한 촉매 활성(절단율)에 대한 용합 극성의 영향을 도시한 그래프이다.

도 2B는 GP41-1 N-인테인의 N-말단(SOLP-NINT) 또는 C-말단(NINT-SOLP) 중 어느 하나에 용합된 3개의 후보 N-인테인 가용화 파트너(46, 206, 246)에 대한 에스케리키아 콜라이에서의 단백질 발현에 대한 용합 극성의 영향을 도시한 그래프이다.

도 3은 GP41-1 N-인테인의 C-말단에 용합된 7개의 후보 가용화 파트너(46, 206, 246, 51, 138, 342, 368)에 대한 기질 절단율 및 가용성 발현 역가를 도시한 그래프이다.

도 4는 후보 가용화 파트너의 계산된 물리적 특성과, N-인테인의 C-말단에 용합된 가용화 파트너를 함유하는 용합 단백질의 에스케리키아 콜라이에서의 전체(역가) 또는 가용성(가용성 역가) 발현 간의 상관관계를 도시한 그래프이다. mw: 분자량; pI: 등전점; AI: 지방족 지수; GRAVY: 그랜드 소수성 평균.

도 5a는 GP41-1 인테인의 위치 65의 시스테인에 상응하는 잔기에서, 특정 아미노산이 대략 100개의 GP41-1 상동체에서 발견되는 빈도를 보여주는 차트이다.

도 5b는 GP41-1 인테인의 위치 89의 시스테인에 상응하는 잔기에서, 특정 아미노산이 대략 100개의 GP41-1 상동

체에서 발견되는 빈도를 보여주는 차트이다.

도 6은 위치 65 및 89의 2개의 중앙 배치된 자연-발생 시스테인 잔기를 함유하는 야생형 GP41-1 N-인테인, 또는 위치 65 및 89의 시스테인 잔기 중 하나 또는 둘 모두에 대한 아미노산 치환을 함유하는 GP41-1 N-인테인의 변이체와, 가용화 파트너 138의 융합 단백질의 촉매 활성(절단율)을 도시한 그래프이다.

도 7은 가용화 파트너 138(단백질 데이터은행 구조 1RYK)에 대한 NMR 용액 구조를 도시한 것이다. 단백질은 4개의 알파 나선 도메인을 포함하며, 구형이며, N-인테인의 카르복시 말단으로의 연결을 형성하는 긴 비구조화 코일(coil)을 갖는다(원형 영역; N-인테인 미도시). 황색 강조표시로 나타난 루프 영역 GKL 및 GYQ를 시스테인 잔기 삽입을 위해 표적화시켜, 신규한 버전(GCKL(SEQ ID NO: 61), GCYQ(SEQ ID NO: 62) 및 GCGYQ(SEQ ID NO: 63))의 가용화 파트너 138(138_GKL22GCKL, 138_GYQ48GCGYQ 및 138_GYQ48GCGYQ)을 생성하였다.

발명을 실시하기 위한 구체적인 내용

- [0018] 상세한 설명
- [0019] 본 발명의 예시적인 구현예의 설명이 후술된다.
- [0020] I. 정의
- [0021] 본 발명이 더욱 용이하게 이해될 수 있도록, 특정 용어를 먼저 정의한다. 추가의 정의가 상세한 설명에 기재되어 있다. 다르게 정의되지 않는 한, 본원에 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 분야의 숙련자에 의해 통상적으로 이해되는 것과 동일한 의미를 갖는다.
- [0022] 용어 "관심 생체분자" 및 "표적 분자"는 예를 들어, 혼합물(예를 들어, 미정제 단백질 혼합물)로부터 정제되거나 제거되어야 하는 생물학적 분자(예를 들어, 단백질), 물질 또는 거대분자 어셈블리를 지칭하기 위하여 본원에서 상호교환가능하게 사용된다. 예시적인 관심 생체분자에는 예를 들어, 재조합 펩티드 및 단백질, 예를 들어, 항체(예를 들어, 모노클로날 항체), 백신, 바이러스, 및 다른 거대분자 어셈블리, 예를 들어, 생체분자 및 합성 성분 둘 모두를 혼입할 수 있는 바이러스-유사 입자 및 나노입자가 포함된다. 예를 들어, 관심 생체분자는 단백질 및 생체분자 어셈블리(예를 들어, 재조합 DNA 기술에 의해 생성됨), 예를 들어, 호르몬(예를 들어, 인슐린, 인간 성장 호르몬, 에리트로포이에틴, 인터페론, 과립구 콜로니 자극 인자, 조직 플라스미노겐 활성화제), 모노클로날 항체(mAb) 및 mAb-유도체(예를 들어, 이중특이적 mAb, Fab, scFv, 상어 및 낙타 항체), 스키펴드-유래 치료제(예를 들어, DARPin, 아피바디(Affibody), 안티칼린(anticalin)), 치료적 효소(예를 들어, 알파 갈락토시다제 A, 알파-L-이두로니다제, N-아세틸갈락토사민-4-설파타제, 글루코세레브로시다제), 독소(예를 들어, 보툴리눔, CRM 197, 리신(ricin)), 재조합 백신(예를 들어, 탄저병, 디프테리아, 파상풍, 폐렴, B형 간염 바이러스, 인간 유두종 바이러스), 바이러스-유사 입자(예를 들어, B형 간염, 인간 유두종, 인플루엔자, 파보바이러스, 노르워크 바이러스(Norwalk virus)) 및 산업 효소(예를 들어, 파파인, 브로멜라인, 트립신, 프로테이나제 K, BENZONASE™ 효소, DENERASE™ 효소, 우레아제, 펩신 등) 및 진단 시약(예를 들어, 글루코스 및 락테이트 탈수소효소, DNA 중합효소, 알칼리성 포스포타제, 서양고추냉이 퍼옥시다제, 제한 효소, 하이브리도마-유래 항체 등)을 포함할 수 있다. 특정 구현예에서, 표적 분자는 치료적 표적에 대한 항체(예를 들어, 모노클로날 항체)이다.
- [0023] 용어 "융합 단백질"은 펩티드 결합에 의해 연결되는 2개 이상의 이중 폴리펩티드의 전부 또는 일부를 포함하는 자연 발생, 합성, 반-합성 또는 재조합 단일 단백질 분자를 지칭한다.
- [0024] 본원에 사용되는 용어 "펩티드"는 비-아미노산 분자(예를 들어, 발색단, 약물, 독소, 영상화 조영제 등)도 혼입할 수 있는 2개 초과 아미노산 길이의 펩티드 및 단백질을 지칭한다.
- [0025] 용어 "폴리펩티드"는 아미노산의 폴리머를 지칭하고, 특정 길이를 지칭하는 것이 아니며; 이에 따라, 펩티드, 올리고펩티드 및 단백질이 폴리펩티드의 정의 내에 포함된다.
- [0026] 본원에 사용되는 용어 "분할형 인테인"은 하기의 특성을 갖는, 자연으로부터 단리되거나, 재조합 DNA 기술을 통해 생성되는 단백질을 지칭한다: (1) 단백질은 높은 친화성 및 선택성으로 상호작용하는 2개의 절반으로 발생하며; (2) 2개의 절반은 촉매 활성에 필요한 모든 인테인 서열을 함유해야 하며, 또한 부가된 비-인테인 펩티드 서열을 함유할 수 있고; (3) 단백질은 2개의 절반이 단단히 회합되는 경우에만 효소 활성을 가지고; (4) 효소 활성은 비-인테인 펩티드 서열로부터 인테인 서열을 분리하거나, 비-인테인 펩티드 서열을 인접 선형 또는 원형 단백질로 라이게이션시키는 역할을 하는 위치 선택적 펩티드 절단 또는 라이게이션이다.

- [0027] 용어 "상보적 인테인"은 분할형 인테인 쌍의 N-인테인 및 C-인테인 부분을 지칭하기 위하여 본원에 사용된다.
- [0028] 본원에 사용되는 용어 "N-인테인"은 단일의 인테인 폴리펩티드의 N-말단 부분과 상동성을 갖는 인테인 폴리펩티드를 지칭하며, 이는 상보적 C-인테인과 회합하여 활성 인테인 효소를 형성한다.
- [0029] 본원에 사용되는 용어 "C-인테인"은 단일의 인테인 폴리펩티드의 C-말단 부분과 상동성을 갖는 인테인 폴리펩티드를 지칭하며, 이는 상보적 N-인테인과 회합하여 활성 인테인 효소를 형성한다.
- [0030] 본원에 사용되는 용어 "엑스테인"은 자연에서 N- 및 C-인테인에 융합된 N- 및 C-말단 펩티드 서열을 지칭하며, 분할형 인테인의 효소 작용을 통해 조각된다(예를 들어, 절단되거나 라이게이션된다).
- [0031] 본원에 사용되는 용어 "리간드"는 특히 표면, 예를 들어, 크로마토그래피 수지에 부착되는 경우, 또 다른 것과 강력하고 선택적으로 상호작용할 수 있는 분자를 지칭한다. 본 발명의 일부 구현예에서, 리간드는 본원에 기재된 N-인테인 융합 단백질일 수 있다.
- [0032] 본원에 사용되는 용어 "가용화 파트너"는 N-인테인에 융합되는 경우, 가용화 파트너의 부재하에 발현되는 가용성 N-인테인의 양에 비하여 에스케리키아 콜라이에서 발현되는 가용성 N-인테인의 양을 증진시키는(예를 들어, 증가시키는, 늘리는 또는 유지하는) 단백질을 지칭한다. 예를 들어, 다양한 구현예에서, 가용화 파트너와의 융합 단백질로서 N-인테인을 발현하는 것은 가용화 파트너 없이 발현되는 경우의 인테인의 가용성에 비하여 N-인테인의 가용성을 적어도 약 10%(예를 들어, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90% 이상) 증가시킬 수 있다.
- [0033] 일 구현예에서, 가용화 파트너 E(SEQ ID NO: 25)는 N-인테인에 융합되고, 생성된 융합 단백질의 가용성은 실험 기준선을 제공하기 위하여 사용된다. 이것은 특히 단독의 N-인테인이 가용성이 아니거나, 안정적이지 않은 경우에 유용하다.
- [0034] 본원에 사용되는 용어 "모 분자" 또는 "야생형(wt) 대응물" 또는 "wt 단백질" 또는 "wt 도메인"은 일반적으로 본원에서 대조군으로 사용되는 그의 실질적으로 고유의 형태의 상응하는 단백질(예를 들어, N-인테인, N-인테인 가용화 파트너) 또는 단백질의 도메인을 지칭하는 것으로 의도된다.
- [0035] 용어 "서열 동일성"은 2개의 뉴클레오티드 또는 아미노산 서열이 예를 들어, 디폴트 갭 가중치를 사용하여 프로그램 GAP 또는 BESTFIT에 의해 최적으로 정렬되는 경우, 적어도 70% 서열 동일성 또는 적어도 80% 서열 동일성 또는 적어도 85% 서열 동일성 또는 적어도 90% 서열 동일성 또는 적어도 95% 서열 동일성 또는 그 이상을 공유하는 것을 의미한다. 서열 비교를 위하여, 전형적으로, 하나의 서열은 시험 서열이 비교되는 참조 서열(예를 들어, 모 서열)로서 역할을 한다. 서열 비교 알고리즘을 사용하는 경우, 시험 및 참조 서열은 컴퓨터에 입력되며, 필요하다면, 하위서열 좌표가 지정되고, 서열 알고리즘 프로그램 파라미터가 지정된다. 그 다음, 서열 비교 알고리즘에 의해, 지정된 프로그램 파라미터에 기초하여, 참조 서열에 비한 시험 서열(들)에 대한 서열 동일성 백분율을 계산한다.
- [0036] 비교를 위한 최적의 서열 정렬은 예를 들어, 문헌[Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981)]의 국소 상동성 알고리즘에 의해, 문헌[Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970)]의 상동성 정렬 알고리즘에 의해, 문헌[Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988)]의 유사성 검색 방법에 의해, 이들 알고리즘(미국 위스콘신주 매디슨 사이언스 드라이브 575 제네틱스 컴퓨터 그룹(Genetics Computer Group), 위스콘신 제네틱스 소프트웨어 패키지(Wisconsin Genetics Software Package)에서 GAP, BESTFIT, FASTA 및 TFASTA)의 컴퓨터 구현에 의해, 또는 육안의 검사에 의해(일반적으로 문헌[Ausubel *et al.*, *Current Protocols in Molecular Biology*] 참조) 행해질 수 있다. 서열 동일성 및 서열 유사성 백분율을 결정하기에 적합한 알고리즘의 하나의 예는 BLAST 알고리즘이며, 이는 문헌[Altschul *et al.*, *J. Mol. Biol.* 215:403 (1990)]에 기술되어 있다. BLAST 분석을 수행하기 위한 소프트웨어는 미국 국립생물공학정보센터(National Center for Biotechnology Information)를 통해 공개적으로 입수가 가능하다(미국 국립보건원 NCBI 인터넷 서버를 통해 공개적으로 접근가능함). 전형적으로, 디폴트 프로그램 파라미터를 사용하여, 서열 비교를 수행할 수 있지만, 맞춤형 파라미터도 또한 사용될 수 있다. 아미노산 서열에 있어서, BLASTP 프로그램은 디폴트로서 3의 단어 길이(W), 10의 기대값 및 BLOSUM62 점수화 매트릭스(문헌[Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)] 참조)를 이용한다.
- [0037] 본원에 사용되는 용어 "크로마토그래피"는 혼합물에서 다른 분자로부터 관심 표적 분자를 분리하고, 그것이 분리되게 하는 역학적 분리 기술을 지칭한다. 전형적으로, 크로마토그래피 방법에서, 이동상(액체 또는 기체)은

고정상(보통 고체) 매질을 가로질러 또는 그를 통해 관심 표적 분자를 함유하는 시료를 운반한다. 고정상에 대한 친화성 또는 분배의 차이는 상이한 분자를 분리시키는 한편, 이동상은 상이한 시간에 상이한 분자를 운반한다.

[0038] 본원에 사용되는 용어 "친화성 크로마토그래피"는 분리되어야 하는 표적 분자가 표적 분자와 특이적으로 상호작용하는 분자(예를 들어, N-인테인 및 N-인테인 가용화 요소를 포함하는 본 발명에 따른 친화성 크로마토그래피 리간드)와 표적 분자의 상호작용에 의해 단리되는 크로마토그래피 방식을 말한다. 일 구현예에서, 친화성 크로마토그래피는 본원에 기재된 바와 같이, 고체 지지체 상에 N-인테인-기반의 리간드를 지니는 고체 지지체에, 표적 분자(예를 들어, 면역글로불린 또는 Fc-함유 단백질)를 함유하는 시료를 첨가하는 것을 수반한다.

[0039] 본원에 상호교환가능하게 사용되는 용어 "친화성 매트릭스" 또는 "친화성 크로마토그래피 매트릭스"는 친화성 크로마토그래피 리간드(예를 들어, N-인테인 융합 단백질 또는 그의 도메인)가 부착되는 크로마토그래피 지지체를 지칭한다. 리간드는 혼합물로부터 정제되거나 제거되어야 하는 관심 분자(예를 들어, 상보적 C-인테인 융합 단백질)에 친화성 상호작용을 통해 결합할 수 있다.

[0040] 용어 "면역글로불린", "Ig" 또는 "항체"(본원에 상호교환가능하게 사용됨)는 2개의 중쇄 및 2개의 경쇄로 이루어진 기본 4-폴리펩티드 쇠 구조를 갖는 단백질을 지칭하며, 상기 쇠는 예를 들어, 쇠간 이황화 결합에 의해 안정화되며, 이는 항원에 특이적으로 결합하는 능력을 갖는다. 용어 "단쇄 면역글로불린" 또는 "단쇄 항체"(본원에 상호교환가능하게 사용됨)는 중쇄 및 경쇄로 이루어진 2-폴리펩티드 쇠 구조를 갖는 단백질을 지칭하며, 상기 쇠는 예를 들어, 쇠간 펩티드 링커에 의해 안정화되며, 이는 항원에 특이적으로 결합하는 능력을 갖는다. 용어 "도메인"은 예를 들어, .베타.-병풍 구조 및/또는 쇠간 이황화 결합에 의해 안정화되는 펩티드 루프를 포함하는(예를 들어, 3 내지 4개의 펩티드 루프를 포함하는) 중쇄 또는 경쇄 폴리펩티드의 구형 영역을 지칭한다. 도메인은 추가로, "불변" 도메인의 경우에 다양한 부류 구성원의 도메인 내의 서열 변이의 상대적 결여에 기초하여, 또는 "가변" 도메인의 경우에 다양한 부류 구성원의 도메인 내의 유의미한 변이에 기초하여, 본원에서 "불변" 또는 "가변"으로 지칭된다. 항체 또는 폴리펩티드 "도메인"은 종종 해당 분야에서 상호교환가능하게 항체 또는 폴리펩티드 "영역"으로 지칭된다. 항체 경쇄의 "불변" 도메인은 상호교환가능하게 "경쇄 불변 영역", "경쇄 불변 도메인", "CL" 영역 또는 "CL" 도메인으로 지칭된다. 항체 중쇄의 "불변" 도메인은 상호교환가능하게 "중쇄 불변 영역", "중쇄 불변 도메인", "CH 영역" 또는 "CH 도메인"으로 지칭된다. 항체 경쇄의 "가변" 도메인은 상호교환가능하게 "경쇄 가변 영역", "경쇄 가변 도메인", "VL" 영역 또는 "VL" 도메인으로 지칭된다. 항체 중쇄의 "가변" 도메인은 상호교환가능하게 "중쇄 가변 영역", "중쇄 가변 도메인", "VH" 영역 또는 "VH" 도메인으로 지칭된다.

[0041] "항체" 또는 "면역글로불린"은 모노클로날 또는 폴리클로날일 수 있으며, 단량체 또는 폴리머 형태, 예를 들어, 오량체 형태로 존재하는 IgM 항체 및/또는 단량체, 이량체 또는 다량체 형태로 존재하는 IgA 항체로 존재할 수 있다. 용어 "단편"은 무손상 또는 완전한 항체 또는 항체 쇠보다 더 적은 아미노산 잔기를 포함하는 항체 또는 항체 쇠의 부분 또는 일부를 지칭한다. 단편은 무손상 또는 완전한 항체 또는 항체 쇠의 화학적 또는 효소적 처리를 통해 획득될 수 있다. 또한, 단편은 재조합 수단에 의해 획득될 수 있다. 예시적인 단편은 Fab, Fab', F(ab')₂, Fc 및/또는 Fv 단편을 포함한다.

[0042] 본원에 상호교환가능하게 사용되는 용어 "폴리뉴클레오티드" 및 "핵산 분자"는 리보뉴클레오티드 또는 데옥시리보뉴클레오티드 중 어느 하나의 임의의 길이의 뉴클레오티드의 폴리머 형태를 지칭한다. 이들 용어는 단일-, 이중- 또는 삼중-가닥 DNA, 게놈 DNA, cDNA, RNA, DNA-RNA 하이브리드, 또는 퓨린 및 피리미딘 염기를 포함하는 폴리머, 또는 다른 친연의, 화학적으로 또는 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기를 포함한다. 폴리뉴클레오티드의 백본은 당 및 포스페이트기(전형적으로 RNA 또는 DNA에서 관찰될 수 있는 바와 같음), 또는 변형된 또는 치환된 당 또는 포스페이트 기를 포함할 수 있다. 또한, 이중-가닥 폴리뉴클레오티드는 상보적 가닥을 합성하고, 적절한 조건하에 가닥을 어닐링시키거나, 또는 적절한 프라이머와 함께 DNA 증합 효소를 사용하여 상보적 가닥을 새로이 합성함으로써 단일 가닥 폴리뉴클레오티드의 화학적 합성 산물로부터 획득될 수 있다. 핵산 분자는 많은 상이한 형태를 취할 수 있으며, 예를 들어, 유전자 또는 유전자 단편, 하나 이상의 엑손, 하나 이상의 인트론, mRNA, cDNA, 재조합 폴리뉴클레오티드, 분지형 폴리뉴클레오티드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브 및 프라이머가 있다. 폴리뉴클레오티드는 변형된 뉴클레오티드, 예를 들어, 메틸화 뉴클레오티드 및 뉴클레오티드 유사체, 우라실, 기타 당 및 연결기, 예를 들어, 플루오로리보스 및 티오에이트, 및 뉴클레오티드 분지를 포함할 수 있다. 본원에 사용되는 바와 같이, "DNA" 또는 "뉴클레오티드 서열"은 염기 A, T, C 및 G 뿐만 아니라, 그들의 유사체 또는 이들 염기의 변형된 형태 중 임의의 것, 예를 들어, 메틸화 뉴클레오티드, 뉴클레오티드간 변형, 예를 들어, 비하전된 연결기

및 티오에이트, 당 유사체의 이용, 및 변형된 및/또는 대체 백본 구조, 예를 들어, 폴리아미드도 포함한다. 특정 구현예에서, 핵산 분자는 본원에 기재된 바와 같은 N-인테인 융합 단백질 또는 그의 변이체를 인코딩하는 뉴클레오티드 서열을 포함한다.

[0043] II. 인테인-기반의 융합 단백질

[0044] 인테인은 이들 분자의 천연 수명 주기에서 기능하는 프로테아제 및 리가아제 활성 둘 모두를 함유하는 1990년에 발견된 자가촉매 효소의 부류이다. 인테인 시약이 펩티드 기질의 절단, 라이게이션 및 환화에 유용성을 갖는 것이 입증되었다. 1998년도에, "분할형 인테인"으로 지칭되는 신규한 부류의 인테인이 발견되었으며, 여기서, 효소는 천연적으로 N-인테인 및 C-인테인으로 지칭되는 2개의 부분(상보적 절반 인테인)으로 발생한다. 분할형 인테인은 매우 다양한 하등 원핵생물에서 관찰되었으며(문헌[Zettler J., et al., *FEBS Letters*, 553:909-914 (2009)]; 문헌[Dassa B., et al., *Biochemistry*, 46:322-330 (2007)]; 문헌[Choi J., et al., *J Mol Biol.* 556: 1093-1106 (2006)]; 문헌[Caspi, et al., *Mol Microbiol.*, 50: 1569-1577 (2003)]; 문헌[Liu X. and Yang J., *J Biol Chem.*, 275:26315-26318 (2003)]; 문헌[Wu H., et al., *Proc Natl Acad Sci USA.* 5:9226-9231 (1998)]), 분할형 인테인은 진핵생물에서 확인되지 않았다(뉴 잉글랜드 바이오랩스(New England Biolabs)에 의해 유지되는 인테인 데이터베이스(<http://tools.neb.com/inbase/list.php>) 참조). 2개의 분할형 인테인은 최근에 매우 신속하고, 엑스테인 서열의 인접에 관하여 상당히 난잡한 것으로 특성화되었다. 하나의 부류는 Npu DnaE 인테인이며(문헌[Iwai I., et al., *FEBS Letters* 550: 1853-1858 (2006)]; 문헌[Zettler J., et al., *FEBS Letters*, 553:909-914 (2009)]), 다른 GP41 분할형 인테인이 메타게놈 데이터로부터 확인되었다(문헌[Carvajal-Vallejos P., et al., *J. Biol. Chem.* 287: 28686-28696 (2012)]; 국제 PCT 공개 제W02013045632호).

[0045] 엑스테인이 부착된 N- 및 C-인테인(인테인 활성에 의해 연결될 2개의 절반의 단백질)은 다수의 도메인간 상호작용을 통해 매우 특이적으로 그리고 단단하게 회합하여, 활성 인테인 효소를 형성한다(문헌[Shah N.H., et al., *J. Amer. Chem. Soc.* 135: 18673-18681; 문헌[Dassa B., et al., *Nucl. Acids Res.*, 37:2560-2573 (2009)]). 제1 부류의 인테인에 존재하는 리가아제 및 프로테아제 활성에 더하여, 분할형 인테인은 N- 및 C-인테인 도메인의 단단하고 선택적인 상호작용으로 인하여 친화성 분리에 유용하다.

[0046] 본 발명은 부분적으로, 본원에 가용화 파트너로 지칭되는 특정 이종 단백질과의 융합 단백질로서의 인테인 폴리펩티드를 발현하는 것이 인테인의 가용성을 증가시키며, 그에 의해, 인테인을 소규모 또는 대규모로 실시될 수 있는 친화성 크로마토그래피 및 다른 단백질 정제 및 변형 응용을 위한 시약으로서 적합하게 한다는 발견에 기초한다. 더욱 구체적으로, 본 발명은 상보적 C-인테인 폴리펩티드와 회합시킴으로써 활성 인테인 복합체를 형성할 수 있는 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 고도의 가용성 융합 단백질을 제공한다. 또한, 본 발명은 C-인테인 폴리펩티드 및 표적 분자를 포함하는 융합 단백질을 제공하며, 여기서, 융합 단백질은 상보적 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 다른 융합 단백질과 회합할 수 있다.

[0047] 따라서, 일 구현예에서, 본 발명은 N-인테인 폴리펩티드 및 N-인테인 가용화 파트너의 전부 또는 일부를 포함하는 융합 단백질에 관한 것이다. 다양한 N-인테인 폴리펩티드가 해당 분야에 알려져 있다. 예시적인 N-인테인은 표 1에 나타난 N-인테인 및 본원의 다른 곳에 기재된 다른 것들을 포함한다. 본원에 개시된 N-인테인 및 해당 분야에 알려져 있는 다른 N-인테인, 및 야생형 N-인테인과 적어도 약 75% 서열 동일성(예를 들어, 적어도 약 80%, 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 적어도 약 99% 서열 동일성)을 갖는 그러한 N-인테인의 변이체가 본원에 기재된 융합 단백질에 포함될 수 있다.

[0048] 인테인 N-말단 도메인 내의 제1 아미노산은 전형적으로 고도로 보존되며, 단백질 스플라이싱 반응에 중요할 수 있다. 그러나, 일부 구현예에서, 인테인 N-말단 도메인 내의 제1 아미노산(예를 들어, 시스테인, 세린)은 인테인과 이종 폴리펩티드 간의 절단을 방지하거나 감소시키는 아미노산(예를 들어, 시스테인 또는 세린 이외의 아미노산)으로 치환될 수 있다. 특정 구현예에서, 인테인 N-말단 도메인 내의 제1 아미노산은 알라닌으로 치환된다.

[0049] 특정 구현예에서, 본원에 기재된 N-인테인 융합 단백질은 야생형 GP41-1 N-인테인(SEQ ID NO: 1 또는 SEQ ID NO: 29) 또는 그의 변이체를 포함한다. 적합한 변이체 GP41-1 N-인테인은 야생형 GP41-1 N-인테인(SEQ ID NO: 1)과 적어도 약 75% 서열 동일성(예를 들어, 적어도 약 80%, 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 적어도 약 99% 서열 동일성)을 가질 수 있다. 본 발명의 융합 단백질에 포함시키기 위한 변이체 GP41-1 N-인테인의 특정 예는 본원에 SEQ ID NO: 2 내지 8로 지정된 GP41-1 변이체를 포함한다.

특정 구현예에서, GP41-1 N-인테인 변이체는 시스템 잔기가 결여되어 있다. 특정 구현예에서, GP41-1 N-인테인에서 자연 발생하는 하나 이상의 시스템 잔기가 결실된다. 다른 구현예에서, GP41-1 N-인테인에서 자연 발생하는 하나 이상의 시스템 잔기(SEQ ID NO: 1의 위치 7, 65 및 89)는 다른 아미노산 잔기(예를 들어, 트레오닌, 라이신 또는 아스파라긴)으로 치환된다. 일 구현예에서, SEQ ID NO: 1의 위치 65에서 GP41-1 N-인테인에서 천연 발생하는 시스템 잔기는 다른 아미노산 잔기(예를 들어, 세린, 트레오닌)로 치환된다. 특정 구현예에서, SEQ ID NO: 1의 위치 65에서 시스템 잔기는 트레오닌으로 치환된다. 또 다른 구현예에서, SEQ ID NO: 1의 위치 89에서 GP41-1 N-인테인에서 천연 발생하는 시스템 잔기는 다른 아미노산 잔기(예를 들어, 메티오닌, 티로신)로 치환된다. 특정 구현예에서, SEQ ID NO: 1의 위치 89에서 시스템 잔기는 메티오닌으로 치환된다. 일부 구현예에서, GP41-1 변이체는 GP41-1 NINTΔA_{TM} N-인테인 변이체(SEQ ID NO: 6) 또는 GP41-1 NINTΔA_{TK} N-인테인 변이체(SEQ ID NO: 8)이다.

[0050] 일부 구현예에서, 일부 또는 모든 시스템 잔기가 결여된 GP41-1 N-인테인 변이체는 라이게이션 또는 절단 반응에서 천연 GP41-1 N-인테인보다 적어도 2배, 적어도 3배, 적어도 4배, 적어도 5배, 적어도 6배, 적어도 7배, 적어도 8배, 적어도 9배 또는 적어도 10배 더 활성이다. 인테인 활성, 절단 또는 라이게이션은 일반적으로 환원 조건하에서 SDS 겔 전기영동을 사용하여 분석될 수 있다(예를 들어, 문헌[Zettler J., Schutz V., Mootz H.D., FEBS Letters 583: 909-914, 2009]; 문헌[Aranko A.S., Zuger S, Buchinger E, Iwai H, PLoS ONE 4: e5185, 2009]). 약술하면, 인테인 반응, 일반적으로, 시간 경과를 환원제(예를 들어, 디티오트레이톨 또는 β-머캅토에탄올)를 함유하는 SDS 겔 로딩 완충제의 첨가에 의해 중단시키고, 시료를 비등시켜, 그들을 완전히 변성시킨 다음, SDS를 함유하는 폴리아크릴아미드 겔 상으로 적절한 단백질 크기 마커와 함께 로딩한다. 전기영동이 완료된 후에, 반응물 내의 단백질은 그들의 분자량에 따라 분리되고, 통상의 염료 또는 형광 염료로 염색시킴으로써 가시화될 수 있다. 시간의 함수로서 다양한 중간체 및 생성물의 양은 농도계(densitometry)에 의해 정량화될 수 있으며, 시간의 함수로서 세기는 곡선 피팅 프로그램의 응용을 통해 효소 비율(kobs)로 전환될 수 있다.

표 1

예시적인 GP41-1 분할형 인테인 및 그의 변이체

SEQ ID NO:	명칭	서열
1	GP41-1 N-인테인	mtrsgyCLDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIICSEE HLFPTQTGEMNISGGLKEGMCLYVKEgg
2	GP41-1 NINTAA _{CC} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIICSEE HLFPTQTGEMNISGGLKEGMCLYVKEgg
3	GP41-1 NINTAA _{AC} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIIASE EHLFPTQTGEMNISGGLKEGMCLYVKEgg
4	GP41-1 NINTAA _{CK} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIICSEE HLFPTQTGEMNISGGLKEGMKLYVKEgg
5	GP41-1 NINTAA _{AM} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIIASE EHLFPTQTGEMNISGGLKEGMMLYVKEgg
6	GP41-1 NINTAA _{TM} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIITSEE HLFPTQTGEMNISGGLKEGMMLYVKEgg
7	GP41-1 NINTAA _{AK} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIIASE EHLFPTQTGEMNISGGLKEGMKLYVKEgg
8	GP41-1 NINTAA _{TK} N-인테인 변이체	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLS NTGYNEVLNVFPKSKKKS Y KITLEDGKEIITSEE HLFPTQTGEMNISGGLKEGMKLYVKEgg
9	GP41-1 C-인테인(CINT)	MGKNSMMLKKILKIEELDERELIDIEVSGNHLF YANDILTHN
10	GP41-1 C-인테인- 티오레독신 융합 단백질 (CINT _{TRX}) (티오레독신 부분에 밑줄이 있음)	MGKNSMMLKKILKIEELDERELIDIEVSGNHLF YANDILTHNMSDKIIHLTDDSFDTDLKADGAJ LVDFWAEWCGPCKMIAPILDEIADEYQGKLTV AKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA TKV G ALSKGQ L KEFLDANLAHHHHHH

SEQ ID No: 1 내지 8에 있어서, 비-인테인 서열은 소문자를 사용하여 표기되어 있으며, SEQ ID No: 1 내지 8에 있어서, 인테인 서열은 대문자로 표기된다.

[0051]

[0052] 전형적으로, N-인테인 폴리펩티드는 통상의 발현 시스템, 예를 들어, 에스케리키아 콜라이에서 발현되는 경우 불량한 용해도를 갖는다. 본 발명은 (예를 들어, 에스케리키아 콜라이에서 발현되는 경우) 예를 들어, N-인테인을 N-인테인의 가용성을 증가시키는 N-인테인 가용화 파트너와의 융합 단백질로서 발현함으로써 이러한 문제를

피한다. 바람직하게는, N-인테인 가용화 파트너는 N-인테인 폴리펩티드의 가용성을 증가시켜, 발현 시스템(예를 들어, 에스케리키아 콜라이)에서의 생성 후에 생성된 융합 단백질의 질량 기준으로, 약 25% 미만이 봉입체에 존재하게 한다. 발현 시스템에서의 생성 후에 봉입체에 존재하는 발현된 단백질의 질량 기준의 백분율은 표준 기술 및 시약을 사용하여 해당 분야의 숙련자에 의해 용이하게 결정될 수 있다.

[0053] 해당 분야의 숙련자는 해당 분야에 공지되어 있고 본원에 기재된 기술을 사용하여 주어진 N-인테인의 가용성을 증가시킬 수 있는 잠재적인 가용화 파트너를 용이하게 선택할 수 있다. 예를 들어, 발현 시스템(예를 들어, 에스케리키아 콜라이)에서의 과발현시에, 가용성 생성물의 생성 확률은 윌킨슨(Wilkinson) 및 해리슨(Harrison)의 알고리즘(문헌[Wilkinson DL and Harrison RG, *Bio/Technology*, 9: 443, 1991])을 사용하여 계산될 수 있다. 단백질이 기능성 분비 신호를 함유하는지 여부의 예측은 덴마크 기술 대학의 생물학적 서열 분석 센터로부터 이용가능한 SignalP 4.1 알고리즘(<http://genome.cbs.dtu.dk/services/SignalP/>)을 사용하여 수행될 수 있다. 또한, 본원에 개시된 실시예 1 내지 3에 기재된 방법을 참조한다. 궁극적으로, 최대의 인테인 촉매 활성을 가능하게 하면서, 가용성의 최적의 증진 둘 모두를 제공하는 가용화 파트너가 실험적 스크리닝을 통해 후보 가용화 파트너로부터 선택되어야 한다.

[0054] 특정한 물리적 특성을 갖는 N-인테인 가용화 파트너는 본 발명의 융합 단백질에 포함시키기에 특히 적합하다. 그러한 물리적 특성은 약 15 kDa 미만의 분자량, 약 60 미만의 지방족 지수(AI) 값 및 -1 미만의 GRAVY 값을 포함하지만 이에 한정되지 않는다. 이들 특성의 각각은 표준 검정 및 기술을 사용하여, 예를 들어, 생물정보학 툴의 SwissProt ExPASy 모음의 일부인 온라인 ProtParam 툴(<http://web.expasy.org/tools/protparam/>)을 사용하여, 해당 분야의 숙련자에 의해 주어진 가용화 파트너에 대하여 결정될 수 있다.

[0055] 선형 폴리펩티드 서열의 그랜드 소수성 평균(GRAVY)(문헌[Kyte J and Doolittle RF., *J. Mol. Biol.* 157:105, 1982])은 모든 아미노산의 소수성 값의 합을 서열 내의 잔기의 개수로 나누어 계산된다. 양의 점수의 증가는 더 큰 소수성을 나타낸다. 계산은 Kyte-Doolittle 척도에 기초한다. GRAVY는 단백질의 소수성 특징을 나타내기 위한 단순한 방법이다.

알라닌	1.8	류신	3.8
아르기닌	-4.5	라이신	-3.9
아스파라긴	-3.5	메티오닌	1.9
아스파르트산	-3.5	페닐알라닌	2.8
시스테인	2.5	프롤린	-1.6
글루타민	-3.5	세린	-0.8
글루탐산	-3.5	트레오닌	-0.7
글리신	-0.4	트립토판	-0.9
히스티딘	-3.2	티로신	-1.3
이소류신	4.5	발린	4.2

[0056] 다양한 구현예에서, 본원에 기재된 N-인테인 융합 단백질은 -1 미만의 GRAVY 값을 갖는다.

[0058] 단백질의 지방족 지수(문헌[Ikai, AJ., *J. Biochem.* 88:1895, 1980])는 지방족 측쇄(알라닌, 발린, 이소류신 및 류신)에 의해 점유되는 상대적 부피로 정의된다. 그것은 구형 단백질의 열안정성의 증가에 대한 긍정적인 요인으로 간주될 수 있다. 단백질의 지방족 지수는 하기의 식에 따라 계산된다: 지방족 지수 = $X(\text{Ala}) + a^* X(\text{Val}) + b^* (X(\text{Ile}) + X(\text{Leu}))$. * 계수 a 및 b는 알라닌의 측쇄에 대한 발린 측쇄(a = 2.9) 및 류신/이소류신 측쇄(b = 3.9)의 상대 부피이다. 또한, 에스케리키아 콜라이에서의 과발현시 가용성 생성물의 생성 확률은 윌킨슨 및 해리슨의 알고리즘을 사용하여 계산될 수 있다(문헌[Wilkinson DL and Harrison RG., *Bio/Technology*, 9: 443, 1991]). 다른 이용가능한 알고리즘은 본질적으로 유사한 결과를 제공하지 않는다. 다양한 구현예에서, 본원에 기재된 N-인테인 융합 단백질은 약 60 미만의 지방족 지수(AI) 값 및 -1 미만의 GRAVY 값을 갖는다.

[0059] 바람직하게는, N-인테인 가용화 파트너는 약 15 kDa 미만의 분자량, 약 60 미만의 지방족 지수 값을 갖는다.

[0060] 특정 N-인테인 가용화 파트너의 예는 표 2에 개시되어 있다.

표 2

예시적인 N-인테인 가용화 파트너

SEQ ID NO:	명칭	GID	서열
11	가용화 파트너 46: Qin 프로파지; 저온 충격- 유도 단백질	170081219	MREYPNGEKTHLTVMAAGFPSL TGDHKVIYVAADRHVTSSEEILE AAIRLLS
12	가용화 파트너 206: 가설 단백질 ECDH10B_1576 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	170081120	MSHLDEVIARVDA AIEESVIAH MNELLIALSDDAELSREDRYTQ QQRLRTAIAHHGRKHKEDMEA RHEQLTKGGTIL
13	가용화 파트너 246: 가설 단백질 ECDH10B_1388 [에스케리키아 콜라이 균 주 K-12 하위균주 DH10B]	170080950	MNKETQPIDRETLLEANKIIRE HEDTLAGIEATGVTQRNGVLVF TGDYFLDEQGLPTAKSTAVFNM FKHLAHLVSEKYHLVD
14	가용화 파트너 51: 가설 단백질 ECDH10B_0422 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	170080051	MSLENAPDDVKLAVDLIVLLEE NQIPASTVLRALDIVKRDYEKLL TRDDEAEK
15	가용화 파트너 138: 추정의 스트레스-반응 단백 질[에스케리키아 콜라이 균 주 K-12 하위균주 DH10B]	170083502	MNKDEAGGNWKQFKGKVKKEQ WGKLTDDDMTHIEGKRDQLVG KIQERYGYQKDQAEKEVVDWE TRNEYRW
16	가용화 파트너 138 GKL22GCKL: 추정의 스트레스-반응 단백질 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	NA	MNKDEAGGNWKQFKGKVKKEQ WGKLTDDDMTHIEGKRDQLV GKIQERYGYQKDQAEKEVVDW ETRNEYRW
17	가용화 파트너 138 GYQ48GQYQ: 추정의 스트레스-반응 단백질 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	NA	MNKDEAGGNWKQFKGKVKKEQ WGKLTDDDMTHIEGKRDQLVG KIQERYGQYQKDQAEKEVVDW ETRNEYRW
18	가용화 파트너 138 GYQ48GCGY: 추정의 스트레스-반응 단백질 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	NA	MNKDEAGGNWKQFKGKVKKEQ WGKLTDDDMTHIEGKRDQLVG KIQERYGCGYQKDQAEKEVVD WETRNEYRW
19	가용화 파트너 342: 가설 단백질 ECDH10B_2487 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]	170081941	MIAEFESRILALIDGMVDHASDD ELFASGYLRGHLTLAIAELES GDHSAQAVHTTVS QSLEKAIGAG ELSPRDQALVTD MWENLFQQA SQQ
20	가용화 파트너 368: 추정의 시그마(54)	170082737	MQLNITGNNVEITEALREFVTA KFAKLEQYFDRINQVYVVLKVE

[0061]

	조절 단백질 [에스케리키아 콜라이 균주 K-12 하위균주 DH10B]		KVTHTSDATLHVNGGEIHASAE GQDMYAAIDGLIDKLARQLTKH KDKLKQH
21	가용화 파트너 A: 3형 분비 시스템의 EspA 성분	9626250	MDTSNATSVVNVSSASSTSTIYD LGNMSKDEVVKLFEELGVFQA AILMFSYMYQAQSNLSIAKFAD MNEASKASTTAQKMANLVDAK IADVQSSTDKNAKAKLPQDVID YINDPRNDISVTGISDLSGDLA GDLQTVKAAISAKANNLTTVVN NSQLEIQMSNTLNLLTSARSD VQSLQYRTISAIISLGK
22	가용화 파트너 B: 파시올라 헤파티카 (Fasciola hepatica) 유래 의 fh8 프로테아제(HiTag)	147611	MPSVEVEKLLHVLDRNGDGKV SAEELKAFADDSKYPLDSNLIK AFIKEHDKNKDGKLDLKLVSIL SS
23	가용화 파트너 C: 파시올라 헤파티카 유래의 fh8 프로테아제(HiTag)	387618410	MPSVEVEKLLH
24	가용화 파트너 D: 글루타티온 S-트랜스퍼라제	251787291	MGQLIDGVVHDTWYDTKSTGG KFQRSASAFRNWLTADGAPGPT GKGGFAAEKDRYHLYVSLACP WAHRTLIMRKLKGLEPFISVSV VNPLMLENGWTFDDSFPGATG DTLYQHFLYQLYLHADPHYSG RVTVPVLWDKKNHTIVSNESAE IIRMFNTAFDALGAKAGDYPP ALQPKIDELNGWIYDVTNNGVY KAGFATSQQA YDEAVAKVFESL ARLEQILGQHRYLTGNQLTEADI RLWTTLVRFDPVYVTHFKCDK HRISDYLNLYGFLRDIYQMPGIA ETVNFDIRNHYFRSHKTINPTG IISIGPWQDLDEPHGRDVRFG
25	가용화 파트너 E: 헤드-DNA 안정화 단백질[엔 테로박테리아 파지 람다]	410480759	MASWSHPQFEKASKETFTHYQP QGNSDPAHTATAPGGLSAKAPA MTPLMLDTSRKLVAWDGTTD GAAVGLAVAADQTSSTLLTFYK SGTFRYEDVLWPEAASDETKKR TAFAGTAISIV
26	가용화 파트너 F: 말토스 결합 단백질	218465276	MKIKTGARILALSALTTMMFSA SALAKIEEGKLVWINGDKGYN GLAEVGKKFEKDTGKVTVEHP DKLEEKFPQVAATGDGPDIIFW AHDRFGGYAQSGLLAEITPDKA FQDKLYPFTWDAVRYNGKLIAY PIAVEALSILYNKDLLPNPPKTW EEIPALDKELKAKGKSALMFNL QEPYFTWPLIAADGGYAFKYEN

[0062]

			GKYDIKDVGVNAGAKAGLTF LVDLIKNKHMNADTDYSIAEAA FNKGETAMTINGPWAWSNIDTS KVNYGVTVLPTFKGQPSKPFVG VLSAGINAASPKNELAKEFLEN YLLTDEGLEAVNKDKPLGAVAL KSYEEELAKDPRIAATMENAQK GEIMPNIPQMSAFWYAVRTAVI NAASGRQTVDEALKDAQTRITK
27	가용화 파트너 G: 티오레독신	218465276	MSDKIIHLTDDSFDTDLKADG AILVDFWAEWCGPCKMIAPILD EIADEYQGKLTVAKLNIDQNP TAPKYGIRGIPTLLLKNGEVAA TKVGALSKGQLKEFLDANLA
28	가용화 파트너 H: 전사 종결 인자 NusA	387509083	MNKEILAVVEAVSNEKALPREK IFEALESALATATKKKYEQEIDV RVQIDRKSGDFDTRRWLVVDE VTQPTKEITLEAARYEDESLLG DYVEDQIESVTFDRITTQTAKQV IVQKVREAERAMVVDQFREHE GEIITGVVKKVNRDNISLDLGN AEAVILREDMLPRENFRPGDRV RGVLYSVRPEARQAQLFVTRSK PEMLIELFRIEVEIGEEVIEKAA ARDPGSRAKIAVKTNDRIDPV GACVGMRGARVQAVSTELGGE RIDIVLWDDNPAQFVINAMAPA DVASIVVDEDKHTMDIAVEAGN LAQAIGRNGQNVRLASQLSGWE LNVMTVDDLQAKHQAEHAHAI DTFTKYLDDIDFATVLEEGFS TLEELAYVPMKELLEIEGLDEPT VEALRERAKNALATIAQAQEE LGDNKPADDLLNLEGVDRDLAF KLAARGVCTLEDLAEQGIDDLA DIEGLTDEKAGALIMAARNICW FGDEA

"GID"는 유전자은행 식별번호를 지칭한다(<http://www.ncbi.nlm.nih.gov/genbank>). 가용화 파트너 A 내지 H는 공지되어 있는 가용화 파트너이며, 이들 중 다수를 구매가능한 융합 시스템 내로 혼입시켜, 에스케리키아 콜라이에서 생성되는 제조합 단백질의 수율 및 용해도를 증가시킨다. 인공적으로 치환되거나 모 서열 내로 삽입되는 아미노산은 볼드체 및 밑줄로 강조표시된다.

[0063]

[0064]

특정 구현예에서, N-인테인 가용화 파트너는 가용화 파트너 138(SEQ ID NO: 15) 또는 그의 변이체(예를 들어, 가용화 파트너 138 GKL22GCKL(SEQ ID NO: 16); 가용화 파트너 138 GYQ48GCYQ(SEQ ID NO: 17); 가용화 파트너 138 GYQ48GCGY(SEQ ID NO: 18))의 전부 또는 그의 일부이거나, 그를 포함한다.

[0065]

표준 제조합 DNA 기술을 포함하나 이에 한정되지 않는 융합 또는 키메라 단백질의 제조 방법이 해당 분야에 널리 알려져 있다. 예를 들어, 상이한 단백질 서열(예를 들어, N-인테인 및 N-인테인 가용화 파트너; C-인테인 및 표적 분자)을 코딩하는 DNA 단편을 통상의 기술에 따라 함께 프레임내 라이게이션시킨다. 다른 구현예에서, 융합 유전자는 자동화 DNA 합성기를 포함하는 통상의 기술에 의해 합성될 수 있다. 대안적으로, 핵산 단편의 PCR 증폭은 2개의 연속 핵산 단편 사이에 상보적 오버행을 초래하는 앵커(anchor) 프라이머를 사용하여 수행될 수 있으며, 2개의 연속 핵산 단편을 이후에 어닐링시키고 재증폭시켜, 키메라 핵산 서열을 생성할 수 있다(문헌 [Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992] 참조). 더욱이, 이미 융합 모이어티(예를 들어, GST 모이어티, Fc 모이어티)를 인코딩하고 있는 많은 발현 벡터가 구매가능하다.

[0066]

바람직하게는, 융합 단백질은 일시적으로 또는 안정적으로 트랜스펙션되거나 형질전환된 원핵 또는 진핵 숙주 세포 또는 유기체에서 인코딩 핵산으로부터 발현된다. 제조합 단백질의 발현을 위한 통상의 숙주 세포 또는 유기체는 예를 들어, 에스케리키아 콜라이, 코리네박테리움 글루타미쿰(*Corynebacterium glutamicum*), 슈도모나스 플루오레스센스(*Pseudomonas fluorescens*), 락토코커스 락티스(*Lactococcus lactis*), 피키아 파스토리스(*Pichia pastoris*), 사카로마이세스 세레비지애(*Saccharomyces cerevisiae*), 제아 메이즈(*Zea maize*), 니코티니아 타바쿰(*Nicotinia tabacum*), 다우쿠스 카로타(*Daucus carota*), SF9 세포, CHO 세포(예를 들어, CHO DG44 세포, CHO DXB11 세포), NS0 세포, HEK 293 세포 및 전체 동물, 예를 들어, 소 및 염소를 포함한다. 일 구현예에서, N-인테인 융합 단백질은 에스케리키아 콜라이에서 발현된다. 그 다음, 발현되는 N-인테인 융합 단백질은 통상의 분리 및 크로마토그래피 방법, 예를 들어, 심층 여과에 의한 정화, 음이온 및 양이온 교환 크로마토그래피에 의한

정제 및 한외여과에 의한 농축을 사용하여 오염 세포 단백질로부터 정제될 수 있다.

- [0067] 이종 단백질(예를 들어, N-인테인 가용화 파트너, 표적 분자)은 인테인 폴리펩티드의 어느 하나의 말단에 융합될 수 있다. 일 구현예에서, N-인테인 가용화 파트너는 N-인테인 폴리펩티드의 N-말단에 연결된다. 다른 구현예에서, N-인테인 가용화 파트너는 N-인테인 폴리펩티드의 C-말단에 연결된다.
- [0068] 일부 구현예에서, 인테인 폴리펩티드(예를 들어, N-인테인, C-인테인) 및 이종 단백질(예를 들어, N-인테인 가용화 파트너, 표적 분자)은 펩티드 결합을 통해 직접 연결된다. 다른 구현예에서, 융합 단백질은 스페이서 또는 링커, 인테인 폴리펩티드(예를 들어, N-인테인, C-인테인)와 이종 단백질(예를 들어, N-인테인 가용화 파트너, 표적 분자) 사이의 분자를 포함한다. 적합한 스페이서/링커 분자는 해당 분야에 알려져 있다.
- [0069] 본원에 기재된 융합 단백질에서, 인테인 N-말단 도메인은 이종 폴리펩티드에 직접적으로(예를 들어, 펩티드 결합을 통해) 또는 간접적으로(예를 들어, 링커 아미노산 서열을 통해) 융합될 수 있다. 따라서, 일부 구현예에서, 이종 폴리펩티드는 인테인 N-말단 도메인의 N-말단에 직접적으로 또는 간접적으로 융합된다. 특정 구현예에서, 이종 폴리펩티드의 제1 아미노산은 Met, Cys, Thr, Arg, Lys, Ser, Gln, His, Ala, Tyr, Phe, Asn, Trp, Val, Leu, Asp, He, Gly, Glu 및 Pro으로 이루어진 군으로부터 선택된다.
- [0070] 일부 구현예에서, 융합 단백질은 이종 폴리펩티드와 인테인 서열 사이에 링커를 포함한다. 예를 들어, 융합 단백질은 이종 단백질의 C-말단과 인테인의 N-말단 도메인의 N-말단 사이에 링커를 포함할 수 있다. 링커는 예를 들어, 약 1 내지 약 10개 아미노산 길이일 수 있다. 일부 구현예에서, 링커는 약 1 내지 약 5개 아미노산 길이일 수 있다. 예를 들어, 링커는 1, 2, 3, 4 또는 5개의 아미노산을 함유할 수 있다. 일부 구현예에서, 이종 폴리펩티드와 인테인의 N-말단 도메인의 N-말단과 접촉하는 링커의 마지막 아미노산은 Met, Cys, Thr, Arg, Lys, Ser, Gln, His, Ala, Tyr, Phe, Asn, Trp, Val, Leu, Asp, Ile, Gly, Glu 및 Pro으로 이루어진 군으로부터 선택된다.
- [0071] 일부 구현예에서, 링커는 엑스테인 서열을 포함할 수 있다. 일부 구현예에서, 링커는 고유 엑스테인 서열을 포함할 수 있다. 일부 구현예에서, 엑스테인은 W0201345632호로부터의 SEQ ID NO: 4, 8, 13, 17, 21, 25, 35 및 39로 이루어진 군으로부터 선택되는 서열을 포함한다. 일부 구현예에서, 엑스테인의 아미노산을 포함하는 링커는 예를 들어, SEQ ID NO: 4, 8, 13, 17, 21, 25, 35 및 39로 이루어진 군으로부터 선택되는 서열의 처음(즉, N-말단) 약 1 내지 약 5개의 아미노산을 포함한다. 일부 구현예에서, 링커는 SEQ ID NO: 4, 8, 13, 17, 21, 25, 35 및 39로 이루어진 군으로부터 선택되는 서열의 약 1, 2, 3, 4 또는 5개의 아미노산을 포함한다. 일부 구현예에서, 융합 단백질은 천연에서 함께 관찰되는 인테인 도메인 및 엑스테인 도메인을 포함한다(예를 들어, GP41-1 N-인테인 및 GP41-1 C-인테인). 다른 구현예에서, 융합 단백질은 인테인 도메인, 및 본원에서 "이종 엑스테인 도메인"으로도 지칭되는 자연에서 특정 인테인 도메인과 함께 관찰되지 않는 엑스테인 도메인을 포함한다. 예를 들어, 융합 단백질은 GP41-1 인테인 도메인 및 IMPDH 엑스테인 도메인을 포함할 수 있다.
- [0072] 본 발명의 융합 단백질은 선택적으로 하나 이상의 검출가능한 표지를 추가로 포함할 수 있다. 본 발명에 따라 사용하기에 적합한 표지는 해당 분야에 알려져 있으며, 일반적으로 그의 화학적 성질에 의해, 그리고 직접적 수단이든지 또는 간접적 수단이든지, 단백질의 검출을 가능하게 하는 식별가능한 신호를 제공하는 임의의 분자를 포함한다. 따라서, 예를 들어, 융합 단백질은 통상적인 방식으로, 예를 들어, 특정 리포터 분자, 형광단, 방사성 물질 또는 효소(예를 들어, 피옥시다제, 포스포타제)를 사용하여 표지될 수 있다. 특정 구현예에서, 융합 단백질은 검출가능한 표지로서 하나 이상의 형광 염료를 포함한다. 검출가능한 표지를 포함하도록 단백질을 변형시키는 표준 방법은 해당 분야에 알려져 있다.
- [0073] 다양한 구현예에서, 본 발명은 추가로 본 발명의 융합 단백질을 인코딩하는 뉴클레오티드 서열을 포함하는 단리된 핵산, 그러한 핵산을 포함하는 발현 벡터 및 그러한 발현 벡터를 지니는 숙주 세포에 관한 것이다.
- [0074] III. N-인테인 융합 단백질을 포함하는 친화성 크로마토그래피 매트릭스
- [0075] N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 함유하는 본원에 기재된 융합 단백질은 특히, 친화성 크로마토그래피 응용을 위한 리간드로서 유용하다. 따라서, 특정 구현예에서, 본 발명은 고체 지지체에 부착된, N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 포함하는 융합 단백질을 포함하는 크로마토그래피 매트릭스를 제공한다.
- [0076] 특정 구현예에서, 고체 지지체는 크로마토그래피 수지이다. 특정 구현예에서, 크로마토그래피 수지는 친수성 폴리비닐 에테르 베이스를 포함한다. 친수성 폴리비닐 에테르 베이스를 갖는 적합한 크로마토그래피 수지는

ESHMUNO® 수지(이엠디 밀리포어 코포레이션(EMD Millipore Corporation))를 포함하나 이에 한정되지 않는다.

- [0077] 다른 구현예에서, 크로마토그래피 수지는 합성 메타크릴레이트계 폴리머 매트릭스(예를 들어, 약 20 내지 40 μm 또는 약 40 내지 90 μm 의 범위의 특정 크기를 갖는 비드)이다. 일부 구현예에서, 크로마토그래피 수지는 카르복실산 작용기를 갖는다. 카르복실산 작용기를 갖는 적합한 크로마토그래피 수지는 FRACTOGEL® COO 수지(이엠디 밀리포어 코포레이션)를 포함하나 이에 한정되지 않는다.
- [0078] 본 발명의 친화성 크로마토그래피 매트릭스를 위한 다른 적합한 고체 지지체는 예를 들어, 기공 제어 유리, 실리카, 산화지르코늄, 산화티탄, 아가로스, 폴리메타크릴레이트, 폴리아크릴레이트, 폴리아크릴아미드, 폴리비닐알코올 및 폴리스티렌 및 그들의 유도체(예를 들어, 그들의 합금)를 포함할 수 있다.
- [0079] 고체 지지체로서 사용되는 다공성 물질은 친수성 화합물, 소수성 화합물, 소유성 화합물, 친유성 화합물 또는 그들의 임의의 조합으로 이루어질 수 있다. 다공성 물질은 폴리머 또는 코폴리머로 이루어질 수 있다. 적합한 다공성 물질의 예에는 폴리에테르 설폰, 폴리아미드, 예를 들어, 나일론, 다당류, 예를 들어, 아가로스 및 셀룰로스, 폴리아크릴레이트, 폴리메타크릴레이트, 폴리아크릴아미드, 폴리메타크릴아미드, 폴리테트라플루오로에틸렌, 폴리설폰, 폴리에스테르, 폴리비닐리덴 플루오라이드, 폴리프로필렌, 폴리에틸렌, 폴리비닐알코올, 폴리카르보네이트, 플루오로카본의 폴리머, 예를 들어, 폴리(테트라플루오로에틸렌-코-피플루오로(알킬 비닐 에테르)), 유리, 실리카, 지르코니아, 티탄, 세라믹, 금속 및 그들의 합금이 포함되나 이들에 한정되지 않는다.
- [0080] 다공성 물질은 유기 또는 무기 분자 또는 유기 및 무기 분자의 조합으로 이루어질 수 있으며, 단백질에 공유 결합되도록 추가의 화학적 변형을 위해 반응하기에, 예를 들어, 공유 결합을 형성하기에 적합한 하나 이상의 작용기, 예를 들어, 하이드록실기, 티올기, 아미노기, 카르보닐기 또는 카르복실산기로 이루어질 수 있다. 다른 구현예에서, 다공성 물질은 작용기를 갖지 않을 수 있고, 작용기, 예를 들어, 하이드록실기, 티올기, 아미노산기, 카르보닐기 또는 카르복실산기를 지니는 물질의 층으로 코팅될 수 있다.
- [0081] 일부 구현예에서, 예를 들어, 유기적 성질이며, 사용되는 수성 매트릭스 친수성 표면을 노출시키는, 예를 들어, 폴리머의 외측 및 존재한다면, 내부 표면상의 하이드록시(-OH), 카르복시(-COOH), 카르보닐(-CHO 또는 RCO-R'), 카르복사미도(아마도 N-치환된 형태의 -CONH₂), 아미노(아마도 치환된 형태의 -NH₂), 올리고- 또는 폴리에틸렌 옥시기를 노출시키는 폴리머에 기초한 통상적인 친화성 분리 매트릭스가 사용된다. 일 구현예에서, 폴리머는 예를 들어, 적합한 다공성 및 강성을 제공하기 위해 유리하게는 예를 들어, 비스에폭시드, 에피할로하이드린, 알릴 브로마이드, 알릴글리시딜 에테르, 1,2,3-트리할로 치환된 저급 탄화수소와 가교된 다당류, 예를 들어, 덱스트란, 전분, 셀룰로스, 풀룰란, 아가로스 등에 기초할 수 있다. 다른 구현예에서, 고체 지지체는 다공성 아가로스 비드를 포함한다. 본 발명에 사용되는 다양한 지지체는 해당 분야에 알려져 있는 표준 방법, 예를 들어, 문헌[Hjerten, *Biochim Biophys Acta* 79(2), 393-398 (1964)]에 기재된 역 현탁액 겔화에 따라 용이하게 제조될 수 있다. 대안적으로, 베이스 매트릭스는 구매가능한 제품, 예를 들어, SEPHAROSE™ FastFlow 매트릭스(스웨덴 업살라 소재의 지이 헬스케어(GE Healthcare))일 수 있다. 대규모 분리에 특히 유리한 일부 구현예에서, 지지체를 조정하여, 그의 강성을 증가시켜, 매트릭스를 높은 유속에 더욱 적합하게 만든다.
- [0082] 대안적으로, 고체 지지체는 합성 폴리머, 예를 들어, 폴리비닐알코올, 폴리하이드록시알킬 아크릴레이트, 폴리하이드록시알킬 메타크릴레이트, 폴리아크릴아미드, 폴리메타크릴아미드 등에 기초할 수 있다. 소수성 폴리머, 예를 들어, 디비닐 및 모노비닐-치환된 벤젠에 기초한 매트릭스의 경우에, 매트릭스의 표면을 종종 친수성화시켜, 상기 정의된 바와 같은 친수성 기를 주변의 수성 액체에 노출시킨다. 그러한 폴리머는 표준 방법에 따라 용이하게 생성될 수 있으며, 예를 들어, 문헌[Arshady, *Chimica e L'Industria* 70(9), 70-75 (1988)]을 참조한다. 대안적으로, 구매가능한 제품, 예를 들어, SOURCE™(스웨덴 업살라 소재의 지이 헬스케어) 및 POROS 수지(미국 캘리포니아주 포스터 시티 소재의 어플라이드 바이오시스템즈(Applied Biosystems))가 사용될 수 있다.
- [0083] 또 다른 구현예에서, 고체 지지체는 무기 성질, 예를 들어, 실리카, 산화지르코늄, 산화티탄 및 그들의 합금의 지지체를 포함한다. 무기 매트릭스의 표면은 종종 적합한 반응성 기를 포함하도록 변형된다. 예에는 CM 지르코니아(사이퍼젠-바이오셉라(Ciphergen-BioSeptra)(프랑스 세르지퐁뜨와즈)) 및 CPG® 지지체(밀리포어 코포레이션)가 포함된다.
- [0084] 일부 구현예에서, 고체 지지체는 예를 들어, 기공 제어 유리의 형태의 지르코니아, 티타니아 또는 실리카에 기초할 수 있으며, 반응성 기를 함유하고/거나 가성 소킹(caustic soaking)을 유지하도록 이를 변형시켜, 리간드에 커플링시킬 수 있다.

- [0085] 예시적인 고체 지지체 형태는 비드(구형 또는 비정형), 중공사, 고휘 섬유, 패드, 겔, 멤브레인, 카세트, 컬럼, 칩, 슬라이드, 플레이트 또는 단일체(monolith)를 포함하나 이들에 한정되지 않는다.
- [0086] 일 구현예에서, 매트릭스의 형태에 관하여, 그것은 다공성 단일체의 형태로 존재한다. 대안적인 구현예에서, 매트릭스는 다공성이거나 비다공성일 수 있는 비드 또는 입자 형태로 존재한다. 비드 또는 입자 형태의 매트릭스는 팩킹층(packed bed)으로서 또는 현탁화된 형태로 사용될 수 있다. 현탁화 형태는 팽창층(expanded bed) 및 순수한 현탁액으로 알려져 있는 형태를 포함하며, 여기서, 입자 또는 비드는 자유롭게 이동한다. 단일체, 팩킹층 및 팽창층의 경우에, 분리 절차는 통상적으로 농도 기울기를 사용한 통상적인 크로마토그래피를 따른다. 순수한 현탁액의 경우에, 회분식이 사용될 것이다. 또한, 표면, 칩, 모세관 또는 필터와 같은 형태의 고체 지지체가 사용될 수 있다.
- [0087] 매트릭스는 또한 카트리지 내의 멤브레인의 형태일 수 있다. 멤브레인은 평판, 나선형 또는 중공사 형태일 수 있다.
- [0088] 특정 구현예에서, 고체 지지체는 가용성 지지체, 예를 들어, 가용성 폴리머 또는 수용성 폴리머일 수 있다. 예시적인 가용성 지지체는 바이오-폴리머, 예를 들어, 단백질 또는 핵산을 포함하나 이들에 한정되지 않는다. 또한, 폴리머는 예를 들어, 음으로 하전된 기(카르복실 또는 설펜), 양으로 하전된 기(4차 아민, 3차 아민, 2차 또는 1차 기), 소수성 기(페닐 또는 부틸기), 친수성 기(하이드록실 또는 아미노기) 또는 상기의 것의 조합을 포함하나 이들에 한정되지 않는 합성 가용성 폴리머일 수 있다. 예시적인 합성 가용성 폴리머는 국제 PCT 공개 제W02008091740호 및 미국 공개 제US20080255027호에서 찾을 수 있으며, 그의 각각의 전체 교시내용은 본원에 참조로 포함된다.
- [0089] 일부 구현예에서, 고체 지지체는 아비딘 분자(예를 들어, 스트렙타아비딘)를 포함할 수 있으며, N-인테인 융합 단백질은 비오틴 태그(예를 들어, 융합 단백질 내의 가용화 파트너에 공유적으로 부착된 비오틴 분자)를 포함하여, 고체 지지체로의 융합 단백질의 결합이 아비딘 및 비오틴 분자의 상호작용을 통해 달성되게 할 수 있다.
- [0090] 본 발명의 N-인테인 융합 단백질은 융합 단백질 내의 오직 하나의 부위(단일 점 부착)에서 또는 융합 단백질 내의 하나 초과 부위(다점 부착)에서 고체 지지체에 부착될 수 있다. 바람직하게는, 융합 단백질이 고체 지지체에 부착되는 경우 융합 단백질 내의 N-인테인 폴리펩티드는 고체 지지체로부터 멀리 배향된다. 예를 들어, 독특한 반응성 아미노산 기(예를 들어, 시스테인 잔기)를 N-인테인 도메인의 활성 영역에 대해 원위 위치의 가용화 파트너에 배치하여, N-인테인이 고체 지지체로부터 멀리 지향되게 보장할 수 있다.
- [0091] 바람직하게는, 고체 지지체로의 부착에 수반되는 융합 단백질 내의 부위(들)(예를 들어, 독특한 반응성 아미노산 기)는 배타적으로 N-인테인 가용화 파트너에 위치한다. 따라서, 이를 달성하기 위하여, N-인테인 폴리펩티드를 변형시켜, 예를 들어, 독특한 반응성 부위를 제공하는 아미노산이 N-인테인에서 발생하는 곳 어디에서나 그러한 아미노산을 결실시키거나 치환시켜, 독특한 반응 부위를 제공하는 아미노산(예를 들어, 시스테인)을 제거하는 것이 필요할 수 있다. 단백질에서의 아미노산의 결실 또는 치환 방법은 해당 분야에 널리 알려져 있다.
- [0092] 고정화된 N-인테인 융합 단백질은 컬럼 또는 다중-웰 크로마토그래피 분리에 적합할 수 있거나, 상자성이어서, 자기장의 인가에 의해 그것이 용액으로부터 포획될 수 있게 할 수 있다.
- [0093] 본원에 기재된 융합 단백질을 지지체, 예를 들어, 해당 분야에 널리 알려져 있고 본원에 기재된 것들을 포함하는 고체 지지체에 부착시키기 위한 임의의 적합한 기술이 사용될 수 있다. 예를 들어, 일부 구현예에서, 융합 단백질은 예를 들어, 융합 단백질에 존재하는 티올, 아미노 및/또는 카르복시기를 사용하는 통상의 커플링 기술을 통해 지지체에 부착될 수 있다. 예를 들어, 비스에폭시드, 에피클로로하이드린, CNBr, N-하이드록시석신이미드(NHS) 등은 널리 알려져 있는 커플링 시약이다. 일부 구현예에서, 스페이서는 지지체와 융합 단백질 사이에 도입되며, 이는 융합 단백질의 이용가능성을 개선시키고, 지지체로의 융합 단백질의 화학적 커플링을 용이하게 한다.
- [0094] 고체 지지체로의 N-인테인 융합 단백질의 부착은 대부분이 해당 분야에 잘 알려져 있는 많은 상이한 방식 및 본원에 기재된 방식을 통해 달성될 수 있다. 예를 들어, 문헌[Hermanson et al., *Immobilized Affinity Ligand Techniques*, Academic Press, pp. 51-136 (1992)]을 참조한다. 예를 들어, 단백질 리간드는 고체 지지체의 표면 또는 단백질 리간드 상의 활성 기, 예를 들어, 하이드록실, 티올, 에폭시드, 아미노, 카르보닐, 에폭시드 또는 카르복실산 기를 통해 고체 지지체에 커플링될 수 있다. 부착은 시아노겐 브로마이드(CNBr), N-하이드록시석신이미드 에스테르, 에폭시(비스옥시란) 활성화 및 환원성 아민화의 이용을 포함하나 이들에 한정되지 않는 알려져 있는 화학을 사용하여 달성될 수 있다.

- [0095] 특정 구현예에서, 카르복실산(-COOH) 또는 아미노(-NH₂) 기를 갖는 크로마토그래피 수지(예를 들어, 비드)가 사용된다. 추가의 구현예에서, 크로마토그래피 수지는 또한 하이드록실기 및/또는 -COOH 또는 -NH₂ 또는 -OH로 전환될 수 있는 다른 작용기를 갖는다.
- [0096] 일부 구현예에서, 티올-유도 단백질 커플링을 사용하여, 고체 지지체에 본 발명의 N-인테인 융합 단백질을 부착시킬 수 있다. 티올-유도 단백질 커플링은 문헌에 설명되어 있다. 예를 들어, 문헌[Ljungquist, *et al.*, *Eur. J. Biochem.* Vol 186, pp. 558-561 (1989)]을 참조한다. 말레이미드는 pH 7.0 내지 7.5에서 티올기와 선택적으로 반응하는 것으로 알려져 있다. 8 초과인 pH에서, 그들은 또한 아민기와 반응할 수 있으며, 또한, 가수분해하는 경향이 있다(문헌[Greg T. Hermanson, *Bioconjugation Techniques*, Academic Press, 2008]; 문헌[Ian Johnson, Michelle T.Z. Spence, *Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies*, 2010]). pH 8 미만에서, 아이오도아세트아미드는 또한 티올기에 대하여 고도로 선택적이다(문헌[Greg T. Hermanson, *Bioconjugation Techniques*, Academic Press, 2008]; 문헌[Ian Johnson, Michelle T.Z. Spence, *Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies*, 2010]). 그러나, 아이오도아세트아미드는 광에서 본질적으로 불안정하며, 대부분의 구매가능한 링커는 수 가용성이 아니고/거나 매우 고가이다. 티올기에 대한 아이오도아세트아미드의 선택성이 말레이미드에 대하여 뛰어나지 않기 때문에, 말레이미드는 일반적으로 대규모 제조를 위한 더 나은 선택이다.
- [0097] 일부 구현예에서, N-인테인 리간드는 가용화 도메인 내의 단일의 이용가능한 설프하이드릴기를 통해 AMP 또는 아이오도아세트아미드 활성화 FG-COO에 커플링될 수 있다. 유도체화된 수지의 리간드 농도는 용액으로부터 C-인테인을 지니는 융합 단백질의 고갈을 측정함으로써 계산될 수 있다. 현재까지, 1 gm/리터(FG-COO)의 비최적화 N-인테인 리간드 농도가 달성되었다.
- [0098] 또한, 다수의 단백질은 에폭시 활성화 수지, 예를 들어, FRACTOGEL® 에폭시에 성공적으로 커플링되었다. 에폭시드는 일차 아미노기, 하이드록실 및 설프하이드릴기와 반응하며, 매우 안정적인 친화성 매트릭스를 제공한다(문헌[PV Kuznetsov 1993. *Pharmaceutical Chemistry Journal* 27:439-52]).
- [0099] 일부 구현예에서, 단백질 리간드는 개재 링커를 통해 고체 지지체에 커플링될 수 있다. 링커는 연결 모이어티에 커플링된 적어도 하나의 작용기를 포함할 수 있다. 연결 모이어티는 작용기에 커플링될 수 있는 임의의 분자를 포함할 수 있다. 예를 들어, 연결 모이어티는 알킬, 알케닐 또는 알킬닐기 중 임의의 것을 포함할 수 있다. 연결 모이어티는 1 내지 30개의 탄소 원자 범위의 탄소쇄를 포함할 수 있다. 일부 구현예에서, 링커는 30개 초과인 탄소 원자로 이루어질 수 있다. 연결 모이어티는 적어도 하나의 헤테로-원자, 예를 들어, 질소, 산소 및 황을 포함할 수 있다. 연결 모이어티는 분지쇄, 비분지쇄 또는 환형쇄로 이루어질 수 있다. 연결 모이어티는 둘 이상의 작용기로 치환될 수 있다.
- [0100] 고체 지지체로의 단백질 리간드의 커플링에 적절한 완충 조건의 선택은 숙련자의 능력 이내이다. 적합한 완충제는 임의의 비-아민 함유 완충제, 예를 들어, 탄산염, 중탄산염, 황산염, 인산염 및 아세트산염 완충제를 포함한다. 회합성 화학물질이 사용되는 경우, 완충제의 염 농도는 사용되는 회합성 기에 좌우될 것이다. 예를 들어, 염 농도는 5 nM 내지 100 mM의 범위일 수 있다. 하전된 종이 사용되는 경우, 염 농도는 5 nM 이상, 0.1 M 미만, 5 nM 이상 0.01 M 미만, 5 nM 이상 0.001 M 미만일 수 있다. 특정 구현예에서, 염 농도는 0.01 M일 수 있다. 소수성 종이 사용되는 경우, 높은 염 농도가 통상 바람직하다. 따라서, 염 농도는 0.001 M 초과, 0.01 M 초과 또는 0.1 M 초과일 수 있다.
- [0101] 일부 구현예에서, 회합성 화학물질이 사용되는 경우, 반응은 0°C 내지 99°C 범위의 온도에서 수행된다. 특정 구현예에서, 반응 방법은 60°C 미만, 40°C 미만, 20°C 미만 또는 10°C 미만의 온도에서 실시된다. 일부 구현예에서, 본 발명의 방법은 약 4°C의 온도에서 실시된다. 다른 구현예에서, 본 발명의 방법은 20°C의 온도에서 실시된다.
- [0102] 다른 구현예에서, N-인테인 융합 단백질은 적절한 가교제 또는 축합 화학물질과 함께 다양한 개질제(멤브레인, 폴리머 표면, 형광 또는 기타 검출 표지)와 병용되어, N-인테인 융합 단백질 및 개질제를 포함하는 공유적 부가물을 형성할 수 있다.
- [0103] IV. 본 발명의 인테인-기반의 융합 단백질의 이용 방법
- [0104] N-인테인 폴리펩티드 및 N-인테인 가용화 파트너를 함유하는 본원에 기재된 융합 단백질, 및 그러한 융합 단백질을 포함하는 친화성 크로마토그래피 매트릭스는 특히, 친화성 정제 방법, 친화성 정제 방법에 사용하기에 적

합한 분할형 활성 인테인 복합체에 대한 스크리닝 방법 및 본원에 추가로 기재된 바와 같은 펩티드 절단 및 라이제이션 방법에 유용하다.

- [0105] 따라서, 특정 구현예에서, 본 발명은 시료 중 표적 분자의 친화성 정제 방법에 관한 것이다. 이러한 구현예의 일 양태에서, 당해 방법은 a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 용합 단백질을 함유하는 시료를 제공하는 단계; b) 시료를 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여 비활성 인테인 복합체를 형성하는 조건하에서 제2 용합 단백질을 포함하는 친화성 크로마토그래피 매트릭스와 접촉시키는 단계로서, 제2 용합 단백질이 펩티드 결합에 의해 N-인테인 폴리펩티드의 가용성을 증진시키는 N-인테인 가용화 파트너에 연결된 N-인테인 폴리펩티드를 포함하는 단계; c) 비활성 인테인 복합체를 함유하는 친화성 크로마토그래피 매트릭스를 세척하여, 미결합 오염 물질을 제거하는 단계; d) 인테인 복합체를 인테인 복합체가 활성이고, 표적 분자를 C-인테인 폴리펩티드로부터 절단하는 조건에 노출시키는 단계; 및 e) 절단된 표적 분자를 회수하는 단계를 포함한다.
- [0106] N-인테인 가용화 파트너에 연결된 N-인테인 폴리펩티드를 포함하는 용합 단백질은 본원의 다른 곳에 기재된 N-인테인 용합 단백질 중 임의의 것일 수 있다.
- [0107] 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 용합 단백질을 함유하는 시료는 임의의 적합한 시료(예를 들어, 생물학적 시료)일 수 있다. 일 구현예에서, 시료는 미정제 단백질 제제 또는 혼합물(예를 들어, 세포 추출물)이다.
- [0108] 표적 분자는 임의의 관심 생체분자일 수 있다. 예를 들어, 관심 생체분자는 단백질 및 생체분자 어셈블리(예를 들어, 재조합 DNA 기술에 의해 생성), 예를 들어, 호르몬(예를 들어, 인슐린, 인간 성장 호르몬, 에리트르포이에틴, 인터페론, 과립구 콜로니 자극 인자, 조직 플라스미노겐 활성화제), 모노클로날 항체(mAb) 및 mAb-유도체(예를 들어, 이중-특이적 mAb, Fab, scFv, 상어 및 낙타 항체), 스캐폴드-유래 치료제(예를 들어, DARPin, 아피바디(Affibody), 안티칼린(anticalin)), 치료적 효소(예를 들어, 알파 갈락토시다제 A, 알파-L-이두로니다제, N-아세틸갈락토사민-4-설파타제, 글루코세레브로시다제), 독소(예를 들어, 보툴리눔, CRM 197, 리신(ricin)), 재조합 백신(예를 들어, 탄저병, 디프테리아, 과상풍, 폐렴, B형 간염 바이러스, 인간 유두종 바이러스), 바이러스-유사 입자(예를 들어, B형 간염, 인간 유두종, 인플루엔자, 파보바이러스, 노르윌크 바이러스) 및 산업적 효소(예를 들어, 과파인, 브로멜라인, 트립신, 프로테이나제 K, BENZONASE™ 효소, DENERASE™ 효소, 우레아제, 펩신 등) 및 진단 시약(예를 들어, 글루코스 및 락트산염 탈수소효소, DNA 중합효소, 알칼리성 포스파타제, 서양고추냉이 퍼옥시다제, 제한 효소, 하이브리도마-유래 항체 등)을 포함할 수 있다. 특정 구현예에서, 표적 분자는 치료적 표적에 대한 항체(예를 들어, 모노클로날 항체)이다.
- [0109] 사용되는 특정 인테인(예를 들어, 시네코시스티스(Synechocystis) 종(Ssp) DnaB, 노스토크 펀크티포르메(Nostoc punctiforme; Npu) DnaE, GP41-1)에 따라, 로딩, 세척, 절단 및 용리 조건은 유의미하게 상이하다. 그럼에도 불구하고, 특정 인테인에 적절한 로딩, 세척, 절단 및 용리 조건은 해당 분야의 숙련자에 의해 용이하게 결정될 수 있다. 특정 인테인에 적합한 조건(예를 들어, 카오트로픽제 및 환원/산화제의 농도, 금속 이온(예를 들어, 아연, 칼슘, 스트론튬, 마그네슘, 망간), 용적 배제제(volume excluding agent)(예를 들어, PEG, PVP, 텍스트란), 세제, 염, 온도 및 pH)은 본원에 하기 기재된 조건을 포함하나 이에 한정되지 않는다. 특히 GP41-1 인테인에 있어서, 활성이 상대적으로 6 내지 10 범위의 pH에 의해 영향을 받지 않는 것이 알려져 있다(문헌[Carvajal-Vallejos P., et al., J. Biol. Chem. 287: 28686-28696 (2012)]).
- [0110] 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여, 촉매적 비활성 인테인 복합체를 형성하는 조건은 주어진 인테인에 대하여 해당 분야의 숙련자에 의해 결정될 수 있다. 일반적으로, 산업 규모 공정에 있어서, 크로마토그래피 분리 동안의 온도의 변화는 컬럼 및 팩킹 온도를 내내 균일하게 보장하기 위해 너무 긴 평형화 단계를 초래하기 때문에, 크로마토그래피 분리 동안의 온도의 변화는 비실용적인 것으로 간주된다. 예시적인 결합 조건은 a) 약 4 내지 25℃의 범위의 온도 및 50 mM Tris/HCl, 300 mM NaCl, 1 mM EDTA, 10%(v/v) 글리세롤, 2 mM DTT을 포함하고 pH가 7인 완충제(예를 들어, GP41-1에 있어서, 문헌[Carvajal-Vallejos P., et al., J. Biol. Chem. 287: 28686-28696 (2012)] 참조); b) 약 4 내지 25℃의 범위의 온도 및 50 mM NaAc, 0.5 M NaCl을 포함하고 pH가 5인 완충제(예를 들어, DnaB 인테인에 있어서, 문헌[Lu W., et al., J. Chrom. A, 1218: 2553-2560(2011)] 참조); 및 c) 약 4 내지 25℃의 범위의 온도 및 0.5 M NaCl, 10 mM Tris-HCl, 0.5 mM 염화아연을 포함하고 pH가 8인 완충제(예를 들어, Npu DnaE에 있어서, 문헌[Guan D., et al., Biotech. Bioeng. 110:2471-2481 (2013)] 참조)을 포함한다.
- [0111] 유사하게, 인테인 복합체의 촉매 활성을 증진시키는 조건은 사용되는 인테인에 따라 달라질 수 있으며, 해당 분

야의 숙련자에 의해 결정될 수 있다. 촉매적 인테인 활성을 증진시키기 위한 예시적인 조건은 a) 50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA를 포함하고 pH가 7.0인 완충제; b) 0.3 M L-아르기닌, 5 mM EDTA, 50 mM 인산염 완충제를 포함하고 pH가 6.5인 완충제; 및 c) 0.5 M NaCl, 10 mM Tris-HCl, 50 mM DTT를 포함하고 pH가 8.0인 완충제를 포함한다.

- [0112] 이러한 구현예의 일 양태에서, 당해 방법은 추가로 본 발명의 친화성 크로마토그래피 매트릭스의 세정, 재생 및/또는 보관을 포함할 수 있다. 전형적으로, 친화성 크로마토그래피 매트릭스는 매트릭스의 조성에 따라, 알칼리성 또는 산성 조건하에서 세정될 수 있다. 친화성 매트릭스를 세정, 재생, 복원 및/또는 보관하기에 적합한 조건은 해당 분야의 숙련자에 의해 결정될 수 있다.
- [0113] 본 발명의 예시적인 친화성 정제 방법은 본원에 개시된 도 1 및 실시예 10에 제공되어 있다.
- [0114] 또 다른 구현예에서, 본 발명은 친화성 정제에 사용하기에 적합한 촉매적 활성 인테인 복합체에 대한 스크리닝 방법에 관한 것이다. 이러한 구현예의 일 양태에서, 당해 방법은 a) 펩티드 결합에 의해 표적 분자에 연결된 C-인테인 폴리펩티드를 포함하는 제1 용합 단백질, 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여 인테인 복합체를 형성하는 조건하에서, 펩티드 결합에 의해 N-인테인 가용화 파트너에 연결된 N-인테인 폴리펩티드를 포함하는 제2 용합 단백질과 접촉시키는 단계; 및 b) 표적 분자가 인테인 활성을 지지하는 조건하에서 C-인테인 폴리펩티드로부터 절단되는지 여부를 결정하는 단계로서, 절단된 표적 분자의 존재가 촉매적 활성 인테인 복합체를 나타내는 단계를 포함한다.
- [0115] 이러한 구현예의 방법에서 사용되는 N- 및 C-인테인은 상보적인 분할형 인테인의 임의의 쌍, 예를 들어, 본원에 개시된 분할형 인테인 쌍(예를 들어, GP41-1 N-인테인 및 C-인테인)일 수 있다.
- [0116] 제1 용합 단백질 내의 C-인테인 폴리펩티드가 제2 용합 단백질 내의 N-인테인 폴리펩티드에 선택적으로 결합하여, 촉매적 비활성 인테인 복합체를 형성하는 조건은 사용되는 인테인에 따라 달라질 수 있으며, 해당 분야의 숙련자에 의해 결정될 수 있다. 예시적인 결합 조건은 a) 약 4 내지 25°C 범위의 온도 및 100 mM Tris-HCl, 25 mM NaCl, 0.1 mM 염화아연을 포함하고 pH가 9인 완충제; b) 약 4 내지 25°C 범위의 온도 및 50 mM NaAc, 0.5 M NaCl을 포함하고 pH가 5인 완충제; 및 c) 약 4 내지 25°C 범위의 온도 및 0.5 M NaCl, 10 mM Tris-HCl을 포함하고 pH가 8인 완충제를 포함한다.
- [0117] 표적 분자는 본원에 개시된 표적 분자 중 임의의 것을 포함하나 이에 한정되지 않는 임의의 적합한 표적 분자일 수 있다.
- [0118] 실시예
- [0119] 실시예 1: GP41-1 N-인테인에 대한 후보 가용화 파트너의 특성화의 선택
- [0120] 출발점으로서 4000개 이상의 알려져 있는 에스케리키아 단백질의 세트를 사용하여, 7개의 가용화 파트너(표 2, SEQ ID NO: 11 내지 15, 19, 20 참조)를 시험을 위해 하기의 기준을 사용하여 선택하였다:
- [0121] (1) 선택된 단백질은 시스테인 잔기가 결여되었으며;
- [0122] (2) 선택된 단백질은 에스케리키아 콜라이에서 과발현되는 경우 가상에서 가용성인 것으로 예측되었고;
- [0123] (3) 선택된 단백질은 11 kDa 미만의 분자량을 가졌고;
- [0124] (4) 선택된 단백질은 분비되지 않는 것으로 가상에서 예측되거나, 알려져 있고;
- [0125] (5) 선택된 단백질은 단백질 상호작용에 관한 정보가 이용가능한 경우, 다량체라기보다는 단량체였고;
- [0126] (6) 선택된 단백질은 단백질 기능에 관한 정보가 이용가능한 경우, 그들(예를 들어, 뉴클레아제, 중합효소 등)을 과발현하는 에스케리키아 콜라이의 사멸을 야기할 가능성이 있거나, 성질이 조절성 또는 독성이 아니었으며, 이는 그들이 주요 세포 경로의 조절에 관여하지 않는 것을 의미하고;
- [0127] (7) 알려져 있는 NMR 또는 X-선 결정학적 구조를 갖는 단백질이 선호된다.
- [0128] 표 3은 이러한 연구에서 평가되는 인테인 및 가용화 파트너에 대한 물리적 특성(분자량(mw), 등전 pH(pI), 에스케리키아 콜라이에서의 가용성 발현의 확률, 단백질이 에스케리키아 콜라이에서 분비되는 것으로 예측되는지 여부, 그랜드 평균 소수성(GRAVY) 및 지방족 지수(AI))을 제공하며, 이는 공개 이용가능한 알고리즘을 사용하여 계산되었다. 에스케리키아 콜라이에서의 과발현시의 가용성의 확률 및 분비 가능성의 예측을 제외하고, 물리적 파라미터의 전부를 생물정보학 툴의 SwissProt ExPASy 모음의 일부인 온라인 ProtParam 툴

(<http://web.expasy.org/tools/protparam/>)을 사용하여 계산하였다. 분자량은 달톤으로 제공되어 있다. 각 단백질에 대한 pI는 단백질이 순 전하를 갖지 않는 pH 값이다. 등전점(pI)은 단백질의 순 전하가 0인 pH이다. 단백질의 경우에, 등전점은 대부분 7개의 하전된 아미노산에 좌우된다: 글루탐산염(δ-카르복실기), 아스파르트산염(β-카르복실기), 시스테인(티올기), 티로신(페놀기), 히스티딘(이미다졸 측쇄), 라이신(ε-아미노기) 및 아르기닌(구아니디늄기). 또한, 단백질 말단기(NH₂ i COOH)의 전하를 고려해야 한다. 그들의 각각은 pK로 지칭되는 그의 독특한 산 해리 상수를 갖는다. 더욱이, 단백질의 순 전하는 용액(완충제) pH와 밀접하게 관련이 있다. 이를 유념하여, 본 발명자들은 헨더슨-하셀바하(Henderson-Hasselbach) 식을 사용하여, 특정 pH에서의 단백질 전하를 계산할 수 있다:

아미노산	NH ₂	COOH	C	D	E	H	K	R	Y
pKa(위키피디아)	8.2	3.65	8.18	3.9	4.07	6.04	10.54	12.48	10.46

[0129]

[0130]

선형 폴리펩티드 서열의 그랜드 평균 소수성(GRAVY)(문헌[Kyte J and Doolittle RF., J. Mol. Biol. 157:105, 1982])은 모든 아미노산의 소수성 값의 합을 서열의 잔기의 수로 나누어 계산한다. 양의 점수의 증가는 더 큰 소수성을 나타낸다. 계산은 카이트-둘리틀(Kyte-Doolittle) 척도에 기초한다. GRAVY는 단백질의 소수성 특징을 나타내기 위한 간단한 방법이다.

알라닌	1.8	류신	3.8
아르기닌	-4.5	라이신	-3.9
아스파라긴	-3.5	메티오닌	1.9
아스파르트산	-3.5	페닐알라닌	2.8
시스테인	2.5	프롤린	-1.6
글루타민	-3.5	세린	-0.8
글루탐산	-3.5	트레오닌	-0.7

[0131]

글리신	-0.4	트립토판	-0.9
히스티딘	-3.2	티로신	-1.3
이소류신	4.5	발린	4.2

[0132]

[0133]

단백질의 지방족 지수(문헌[Ikai, AJ., J. Biochem. 88:1895, 1980])는 지방족 측쇄(알라닌, 발린, 이소류신 및 류신)가 점유하는 상대 부피로 정의된다. 그것은 구형 단백질의 열안정성의 증가에 대한 긍정적인 요인으로 간주될 수 있다. 단백질의 지방족 지수는 하기의 식에 따라 계산한다: 지방족 지수 = X(Ala) + a * X(Val) + b * (X(Ile) + X(Leu)). * 계수 a 및 b는 알라닌의 측쇄에 비한 발린 측쇄(a = 2.9) 및 류신/이소류신 측쇄(b = 3.9)의 상대 부피이다. 또한, 에스케리키아 콜라이에서의 과발현시 가용성 생성물의 생성 확률은 윌킨슨 및 헤리슨의 알고리즘을 사용하여 계산될 수 있다(문헌[Wilkinson DL and Harrison RG., Bio/Technology, 9: 443, 1991]). 다른 이용가능한 알고리즘이 반드시 유사한 결과를 제공하는 것은 아니다.

[0134]

단백질이 기능적 분비 신호를 함유하는지 여부의 예측을 덴마크 기술 대학(Technical University of Denmark)의 생물학적 서열 분석 센터(Center for Biological Sequence Analysis)(<http://genome.cbs.dtu.dk/services/SignalP/>)로부터 이용가능한 SignalP 4.1 알고리즘을 사용하여 수행하였다.

표 3

본 연구에 사용되는 인테인 및 N-인테인 가용화 파트너에 대하여 측정되는 물리적 파라미터

단백질	분자량(D)	PI	에스케리키아 콜라이에서의 가용성 확률	분비	GRAVY	지방족 지수
가용화 파트너 46	5639.4	5.8	74	아니요	-0.086	95.7
가용화 파트너 206	8799.8	5.6	97	아니요	-0.636	93.9
가용화 파트너 246	9385.6	5.4	84	아니요	-0.38	89.3
가용화 파트너 51	6024.8	4.5	98	아니요	-0.391	117.7
가용화 파트너 138	8325.2	5.4	69	아니요	-1.526	48.0
가용화 파트너 138_	8428.3	5.4	60	아니요	-1.469	47.3
GKL22GCKL						
가용화 파트너 138_	8428.3	5.4	60	아니요	-1.469	47.3
GYQ48GICYQ						
가용화 파트너 138_	8485.4	5.4	60	아니요	-1.454	46.6
GYQ48GICYQ						
가용화 파트너 342	10000	4.3	98	아니요	-0.148	93.4

[0135]

가용화 파트너 368	10750.2	6.5	72	아니요	-0.398	95.5
GP41-1 N-인테인 변이체 NINTΔA_CC	10589.1	5.8	21	아니요	-0.44	80.1
GP41-1 C-인테인 (CINT)	4916.7	5.1	86	아니요	-0.312	111.4
티오레독신과의 GP41-1 C-인테인 융합체(CINT_TRX)	17528.1	5.5	78	아니요	-0.189	101.35
가용화 파트너 A	20504.9	4.6	34	아니요	-0.14	91.5
가용화 파트너 B	7597.6	5.6	51	아니요	-0.606	96.0
가용화 파트너 C	1281.5	5.4	94	아니요	0.127	123.6
가용화 파트너 D	37389.1	6.3	42	아니요	-0.388	77.9
가용화 파트너 E	12740.2	5.8	41	아니요	-0.299	60.4
가용화 파트너 F	43387.6	5.5	40	아니요	-0.251	85.4
가용화 파트너 G	12629.4	5.7	74	아니요	-0.143	97.6
가용화 파트너 H	54870.9	4.5	95	아니요	-0.278	98.0

[0136]

[0137]

실시예 2: 에스케리키아 콜라이 단백질 발현 작제물의 생성

[0138]

잠재적인 가용화 파트너 46, 206 및 246에 대한 코딩 서열을 지니는 플라스미드 작제물을 NINTΔA_CC의 아미노 또는 카르복시 말단 아미노산을 통해 NINTΔA_CC에 대한 코딩 서열에 융합시키고, DNA2.0으로부터의 pJ414의 버전 내로 삽입하였다. 이들 작제물을 통상적인 방법을 사용하여 컴피턴트(competent) BL21 DE3 에스케리키아 콜라이 세포 내로 형질전환시키고, 앰피실린 내성 콜로니를 단리하였다. 예상되는 크기의 단백질의 생성을 SDS 폴리아크릴아미드 전기영동(SDS PAGE)을 사용하여 확인하였다.

[0139]

6개의 작제물의 각각의 형질전환체를 상응하는 작제물로 형질전환된 BL21 DE3 에스케리키아 콜라이 세포의 글리세롤 스톱으로부터 100 μg 앰피실린/ml를 함유하는 2 ml의 LB(LB+Amp)에서 배양하였다. 이러한 예비-접종물을 37°C 및 250 rpm에서 하룻밤 성장시키고, 이를 사용하여, 200 ml의 LB+Amp(1% 접종물)를 접종시켰다. 배양물을 0.5 내지 0.6의 OD₆₀₀까지 37°C 및 250 rpm에서 인큐베이션시켰다. 0.4 mM IPTG의 첨가에 의해 단백질 발현을 유도하였다. 온도를 30°C로 내리고, 배양물을 이러한 온도 및 250 rpm에서 5시간 동안 인큐베이션시켰다. 상기 시간 후에, 세포를 원심분리(4500 g, 25분, 4°C)에 의해 수집하고, 상청액을 폐기하고, 세포 펠렛을 추가의 단백

질 정제를 위해 -80℃에 보관하였다.

- [0140] 시험 기질 단백질 CINT_TRX에 대한 코딩 영역을 pSABAD92A(유전자은행 수탁 번호 HM070247) 내로 클로닝하고, 컴피턴트 BL21 DE3 세포 내로 형질전환시켰다. 성공적인 형질전환체를 루리아(Luria) 브로쓰 + 50 µg/ml의 카르베네실린(carbenicillin)(LB+C) 상에서 단리하였다. 예상되는 크기의 단백질의 생성을 SDS PAGE를 사용하여 확인하였다. 3개의 BL21 클론/작제물의 각각의 글리세롤 스톱을 -80℃에 보관한다.
- [0141] 소량의 동결된 BL21 글리세롤 스톱을 사용하여, 37℃, 250 rpm에서, LB+C에서 5 ml 배양물을 접종시켰다. 다음 날, 0.1 ml의 밤샘 성장 배양물을 사용하여, 10 ml의 LB+C를 접종시키고, 이러한 배양물을 0.6 내지 0.9의 OD₆₀₀ 까지 37℃, 250 rpm에서 성장시킨다. 배양물을 28℃, 250 rpm에서 5시간 동안 0.02% 아라비노스로 유도하였다. 유도 후에, 세포를 원심분리(4500 g, 25분, 4℃)에 의해 수집하고, 상청액을 폐기하고, 세포 펠렛을 추가의 단백질 정제를 위해 -80℃에 보관하였다.
- [0142] 실시예 3: 발현된 단백질의 가용성 대 불용성 비 및 총량의 결정
- [0143] 각 작제물에 대한 발현 수율 및 가용성:불용성 비를 결정하기 위하여, 상기 나타낸 바와 같이 배양된 등가의 바이오매스에 상응하는 성장된 배양물의 분취물을 5000g, 4℃에서 15분 원심분리하였다. 배양 상청액을 폐기한 후에, 세포를 50 mM 트리스 pH 8, 300 mM NaCl, 0.5% 트리톤 X-100으로 이루어진 가용화 완충제 200 µl에 재현탁화시켰다. 세포를 초음파분해에 의해 파괴하였다(10회 버스트(burst) x 3, 브란슨(Branson) 250 음파처리기 (Sonifier), 각 시리즈 사이에, 시료가 냉각되게 하는 시간 이용). 가용성 및 불용성 분획을 분리하기 위하여, 시료를 16000 g 및 4℃에서 10분 동안 원심분리하였다. 가용성 분획을 개별 튜브에 옮기는 한편, 불용성 분획을 초음파분해에 의해 동일한 가용화 완충제 200 µl 중에 재현탁화시켰다(이전의 초음파분해에서와 동일한 파라미터 사용).
- [0144] 세포 용해물 중 재조합 단백질을 SDS PAGE 겔 이후에 쿠마시 염색 후에 참조물질로서 정량화된 BSA의 곡선을 사용한 농도 분석을 사용하여 정량화시켰다. 클론마다 3가지 상이한 시료 부피를 BSA 표준 곡선(0.2에서 1.2 µg까지 6점)과 함께 로딩하였다. 단백질 밴드의 세기를 "퀀티티 원(Quantity One)"(바이오라드(Biorad)) 소프트웨어를 사용하여 농도계에 의해 결정하였다. BSA/재조합 단백질 분자량 비를 고려하여 각 단백질에 대하여 보정물을 적용한다.
- [0145] 정제된 단백질에 대한 농도를 계산된 흡광 계수 및 280 nm에서 계산된 그들의 흡광도를 사용하여 결정하였다.
- [0146] 실시예 4: 발현된 단백질의 정제
- [0147] 절단 기질로서 사용되는 C-인테인 융합 단백질 CINT_TRX를 완전히 정제하기 위하여, 단백질을 발현하는 에스케리키아 콜라이 세포를 50 mM Tris-HCl, pH=8.0, 300 mM NaCl, 0.5X CellLytic B(시그마-알드리치(Sigma-Aldrich)) 및 20 mM 이미다졸을 함유하는 완충제 중에 재현탁화시켰다. 세포를 30% 펄스 활성 사이클을 사용하여 20분 동안 얼음 상에서 음파분해하고(브란슨 250 음파처리기), 34500 g, 4℃에서 30분 동안 원심분리하였다. 가용성 C-인테인 융합체를 제조자의 지침에 따라 His-Trap HP(지이 헬스케어(GE Healthcare)) 컬럼 상에서 상청액으로부터 정제하였다. 정제된 C-인테인 융합 단백질을 함유하는 용리된 분획을 풀링하고, 2 mM DTT의 존재하에 절단 완충제(50 mM Tris-HCl, pH=7.0, 300 mM NaCl, 1 mM EDTA, 10%(v/v) 글리세롤; CB)에 대하여 투석하고, -80℃에서 분취물로 보관하였다.
- [0148] N-인테인 융합체를 발현 작제물을 지니는 에스케리키아 콜라이 세포로부터 고유 조건하에 정제하였다. 세포 펠렛을 100 mM Tris-HCl, pH=8.0, 150 mM NaCl 및 1 mM EDTA를 함유하는 완충제 중에 재현탁화시켰다. 그 다음, 세포를 30% 펄스 활성 사이클을 사용하여 20분 동안 얼음 상에서 음파분해하고(브란슨 250 음파처리기), 34500 g, 4℃에서 30분 동안 원심분리하였다. N-인테인 융합체의 가용성 분획을 상기 기재된 바와 같이 His-Trap HP 컬럼에서의 크로마토그래피에 의해 정제하였다. 정제된 단백질을 함유하는 용리된 분획을 풀링하고, 2 mM DTT의 존재하에 CB에 대하여 투석하고, -80℃에서 분취물로 보관하였다.
- [0149] 실시예 5: 발현된 단백질에 대한 절단 역학의 결정
- [0150] 시험관내 반응을 이전에 기재된 바와 같이 수행하였다(문헌[Carvajal-Vallejos P., et al., J. Biol. Chem. 287: 28686, 2012]). 요약하면, 정제된 N- 및 C-융합 단백질을 간단히 상응하는 시험 조건에서 개별적으로 사전-인큐베이션시켰다. 절단 완충제 중에 상보적 N- 및 C-인테인 융합 단백질을 5 µM의 등몰 농도로 혼합함으로써 절단 반응을 개시하고, 25℃ 및 37℃에서 인큐베이션시켰다. 본 발명과 관련된 실험을 위하여, 절단 파트너는 항상 CINT_TRX였다. 분취물을 특정 시간 간격으로 제거하고, 8% SDS(w/v) 및 20% β-머캅토에탄올(v/v)을 함유

하는 SDS PAGE 완충제를 첨가함으로써 반응을 중단시킨 다음, 5분 비등시켰다. 반응 생성물을 SDS PAGE(노벡스(Novex)로부터의 4 내지 12% 비스-트리스 겔, 미국 칼스배드 소재의 인비트로젠(Invitrogen))에 이어서, 쿠마시브릴리언트 블루(Coomassie Brilliant Blue)(시그마) 염색으로 정량화시켰다. 단백질 밴드의 상대적 세기를 퀀티티 윈(바이오라드) 프로그램을 사용하여 농도계에 의해 결정하였다. 상이한 절단 생성물을 그들의 상응하는 분자량에 따라 정규화시켰다. 단백질 절단 백분율을 절단된 산물 및 인테인-태깅된 전구체 CINT_TRX의 비로부터 계산하였다. 고정 비율(kobs)을 GraFit 소프트웨어(영국 서리 소재의 에리타쿠스(Erithacus))를 사용하여 데이터를 식 $P=P_0(1-e^{-kt})$ 에 피팅시킴으로써 결정하였으며, 여기서, P는 시간 t에 형성된 절단된 C-인테인 용합 생성물의 양이며, P₀은 수득될 수 있는 절단 생성물의 최대량(수율)이며, e는 오일러 상수이며, k는 관찰되는 비율이다. 2개의 상보적인 인테인 단편의 신속한 회합 후에, C-인테인 용합 단백질의 절단이 단일-분자 반응과 같이 진행된다는 가정하에, 모든 반응을 비가역적인, 사전-정상 상태 및 1차 공정으로서 처리하였다.

[0151] 실시예 6: NINT 가용화 파트너에 대한 최적의 배치 및 특성의 결정

[0152] 잠재적인 가용화 파트너에 대한 NINTΔA_{CC} 용합체를 모든 가용화 파트너에 대하여 가능한 둘 모두의 배향으로(즉, NINTΔA_{CC}의 N- 또는 C-말단에 용합됨) 생성하였다. 6개의 생성된 작제물을 에스케리키아 콜라이에서 발현시켰으며, 생성된 단백질을 이전에 기재된 바와 같이 생성된 총량 및 가용성에 관하여 분석하였다. 또한, 각각의 작제물로부터 단백질을 정제하고, 기질로서 정제된 CINT_TRX를 사용하여 절단물을 특성화하였다. 이러한 분석의 결과는 도 2A 및 도 2B에 나타나 있다. NINTΔA_{CC}의 N-말단으로의 가용화 파트너의 용합이 시험된 모든 작제물에 대하여 에스케리키아 콜라이에서 더 많은 양의 단백질을 생성하는 한편, C-말단으로의 가용화 파트너의 용합에서, 절단물을 측정하는 경우, 이러한 경향이 역전이 관찰된다. 이들 연구가 진행 중인 동안 상이한 분할형 인테인 시스템을 사용하여 공개된 연구는 가용화 파트너와 N-인테인의 위치와, N-인테인 활성 간의 유사한 관계를 입증하였으며, 이는 용합이 반대 극성에서 이루어지는 경우, 엑스테인 도메인 사이의 더 큰 입체 방해 확률을 나타내는 알려져 있는 구조적 정보를 참조함으로써 설명되었다(문헌[Guan D, Ramirez M, Chen Z., *Biotechnol Bioeng.* 110:2471, 2013]).

[0153] 이러한 연구를 확대시켜, 크기 및 등전점(pI)에 관하여 이전에 특성화된 가용화 파트너와 별개의 추가의 가용화 파트너 51, 138, 342 및 368(표 2 및 표 3 참조)을 포함시켰다. 이들 전부를 가용화 파트너 46, 206 및 246을 사용하여 이전에 나타난 바와 같이 NINTΔA_{CC}의 카복실 말단에 용합시켜, 가장 높은 촉매 활성을 갖는 용합체를 제공하였다. 이들 작제물을 이전에 기재된 바와 같이, 에스케리키아 콜라이에서 발현시키고, 정제하고, 절단물에 대하여 분석하였다. 이들 분석의 결과는 도 3에 제시되어 있다. 가용화 파트너 246이 명백하게 가장 높은 활성을 갖지만, 촉매 활성과 가용성 발현 간의 최적의 절충이 가용화 파트너 138에 대하여 관찰되었다.

[0154] 에스케리키아 콜라이에서의 발현 동안 N-인테인의 가용화를 위해 효율적이게 하는 가용화 파트너 138의 특성을 이해하기 위하여, 표 3의 후보 가용화의 각각에 대하여 계산된 단백질 파라미터를 도 4에서 가용성 역가와 상호 관련시켰다. 이들 파라미터 중 어느 것도 전반적인 발현과 강력하게 관련되지 않지만, AI 및 GRAVY 값 둘 모두는 가용성 역가와 음의 상관관계를 보였다.

[0155] 실시예 7: NINTΔA_{CC} 내의 시스테인 잔기의 대체를 위한 아미노산의 선택

[0156] 천연 공급원으로부터 단리된 GP41-1 N-인테인은 3개의 시스테인 잔기를 함유하지만, 하나를 이전에 대체하여, 본 발명에 대한 모 작제물인 NINTΔA_{CC}를 제공하였다. NINTΔA_{CC} 내에 함유된 나머지 2개의 시스테인 잔기를 대체를 위해 표적화시켜, 독특한 반응성 시스테인 잔기가 이후의 고정화 또는 기타 변형을 위해 가용화 도메인 내로 도입될 수 있게 하였다.

[0157] NINTΔA_{CC} 내의 2개의 시스테인 잔기에 대해 치환될 수 있고, 여전히 안정한 기능성 인테인 단백질을 제공하는 아미노산을 확인하기 위해, 몇몇의 계통발생 분석을 수행하였으며, 여기서, 단백질 서열을 정렬하고, SEQ ID NO: 1의 위치 65 및 89에서의 천연 발생 아미노산 변이체를 시험하였다. GP41-1에서의 천연 발생 내부 시스테인을 유사한 인테인에서 이들 위치에 발생하는 다른 아미노산으로 대체하여, 자연 선택은 이들 변이체가 자연에서 지속되게 함에 따라, 기능성 및/또는 안정성 GP41-1 변이체 단백질을 제공하는 것이 예상될 것이다. 그러한 분석이 GP41 인테인 부류(1, 2, 3, 4, 5, 6; 문헌[Dassa B., *et al.*, *Nucl. Acids Res.*, 37:2560-2573 (2009)])의 N-인테인을 사용하여 수행되는 경우, SEQ ID NO: 1에서 위치 65 및 89의 2개의 시스테인 잔기가 고도로 보존되는 것으로 관찰되며, 이는 이들 시스테인의 치환이 GP41-1 인테인의 활성 및/또는 안정성에 불리하게 영향을 미칠 것을 시사한다. 그러나, 인테인 기능을 갖거나, 갖지 않을 수 있는 약간 더 분기된 단백질을 포함하도록 분석을 확대한다면, 에스케리키아 콜라이 인산염 레굴론(regulon)의 phoH 유전자에 대하여 상동성을 갖는

많은 단백질이 확인된다. 대략 100개의 상동체를 BLAST 검색 툴을 사용하여 유전자은행으로부터 획득하고, 프리웨어(freeware) 툴, BioEdit(문헌[Hall TA., *Nucl. Acids. Symp. Ser.* **41**:95, 1999])를 사용하여 CLUSTAL 알고리즘으로 정렬하였다. 이러한 분석의 결과는 도 5에 나타나 있으며, 여기서, 위치 넘버링은 NINTΔA_CC(SEQ ID NO: 2)에 기초한다. 이러한 분석으로부터, 트레오닌 및 알라닌이 위치 65에서 빈번하게 발생하고, 라이신, 메티오닌 및 아스파라긴이 위치 89에서 빈번하게 발생하는 것이 명백하며, 이는 이들 천연 발생 아미노산으로의 천연 시스템의 치환이 안정한 단백질을 제공할 것을 나타낸다.

[0158] 실시예 8: 최적의 특성에 대한 NINTΔA_CC 아미노산 변이체의 스크리닝

[0159] 표 4에 나타난 아미노산 치환을 함유하는 NINTΔA_CC(SEQ ID NO: 2)에 기초한 작제물을 이전에 기재된 바와 같이 생성하고, 발현시키고, 정제하고, 촉매 활성에 대하여 특성화시켰다.

[0160] 각 작제물로부터 이루어진 N-인테인 융합 단백질에 대한 절단률 측정치는 도 6에 제공되어 있다. NINTΔA_CC 모체(+cnt)는 비교를 위해 좌측에 나타내었다. 위치 65 및 89의 아미노산은 도면의 하측에 나타내었다. 위치 65의 트레오닌 잔기는 모체보다 유의미하게 더 많은 활성을 갖는 N-인테인 융합체를 제공한다. 시험된 작제물 중에, 위치 65에 트레오닌 및 위치 89에 메티오닌을 갖는 N-인테인 융합체는 모체 작제물보다 약 3배 더 신속한 촉매 속도를 갖는 작제물을 제공하였다.

표 4

GP41-1 N-인테인 변이체

GP41-1 N-인테인 변이체	서열
NINTΔA_CC (SEQ ID NO:2)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII C SEEHLFPQTGEMNISGGLKEGM C LYVKEgg
NINTΔA_AC (SEQ ID NO:3)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII A SEEHLFPQTGEMNISGGLKEGM C LYVKEgg
NINTΔA_CK (SEQ ID NO:4)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII C SEEHLFPQTGEMNISGGLKEGM K LYVKEgg
NINTΔA_AM (SEQ ID NO:5)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII A SEEHLFPQTGEMNISGGLKEGM M LYVKEgg
NINTΔA_TM (SEQ ID NO:6)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII T SEEHLFPQTGEMNISGGLKEGM M LYVKEgg
NINTΔA_AK (SEQ ID NO:7)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII A SEEHLFPQTGEMNISGGLKEGM K LYVKEgg
NINTΔA_TK (SEQ ID NO:8)	mtrsgyALDLKTQVQTPQGMKEISNIQVGDVLSNTGYNEVLNVFPKSKKK SYKITLEDGKEII T SEEHLFPQTGEMNISGGLKEGM K LYVKEgg

SEQ ID No: 2 내지 8에 있어서, 비-인테인 서열은 소문자를 사용하여 표기되어 있으며, 인테인 서열은 대문자로 표기되어 있다.

[0161]

[0162] 실시예 9: 가용화 파트너 138 내로의 독특한 시스템인 잔기의 도입을 위한 전략

[0163] N-인테인 융합 단백질의 촉매 활성의 감소 없이, N-인테인 융합 단백질의 화학적 변형을 가능하게 하기 위하여, 궁극적인 변형 부위는 N-인테인의 활성 부위로부터 가능한 한 멀리 떨어져야 한다. 가용화 파트너 138에 대한 구조 정보의 부재하에, 합리적인 접근법은 GP41-1 N-인테인에 대하여 상기 기재된 바와 같은 계통발생 분석을 수행하고(실시예 7 참조), 높은 가변성을 나타내는 단백질의 영역을 결정하고, 시스템의 삽입에 의해 이들을 변형시킨 다음, 생성된 작제물의 전부를 시험하는 것일 것이다. 그러나, 도 7에 나타난 가용화 파트너 138(단백질 데이터은행 구조 1RYK)에 대하여 이용가능한 NMR 용액 구조가 존재한다. 단백질은 4개의 알파 헬릭스 도메인을 함유하며, 구형이며, N-인테인의 카르복시 말단으로의 연결을 형성하는 긴 비구조화 코일(원형 영역; N-인테인 미도시)을 갖는다. 황색 강조표시에 의해 나타난 루프 영역 GKL 및 GYQ를 시스템 잔기 삽입을 위해 표적화시켜, 신규한 버전(GCKL(SEQ ID NO: 61), GCYQ(SEQ ID NO: 62) 및 GCGYQ(SEQ ID NO: 63))의 가용화 파트너 138(138_GKL22GCKL(SEQ ID NO: 16), 138_GYQ48GKYQ(SEQ ID NO: 17) 및 138_GYQ48GCGYQ(SEQ ID NO: 18))을 생성하였다.

[0164] 실시예 10: 크로마토그래피 수지로의 N-인테인 융합 단백질(리간드)의 커플링

- [0165] GP41-1 변이체 NINTΔA_{TM}(SEQ ID NO: 6)의 카르복실-말단에 융합된 가용화 파트너 138_GYQ48GCGYQ(SEQ ID NO: 18)를 함유하는 가용성 융합 단백질을 에스케리키아 콜라이에서 인코딩 핵산으로부터 발현하고, 이후에, 통상의 분리 방법을 사용하여 오염 세포 단백질을로부터 분리한다.
- [0166] 그 다음, 정제된 N-인테인 융합 단백질을 표준 기술을 사용하여 융합 단백질의 가용화 파트너 도메인 내의 독특한 반응성 시스테인 부위를 통해 FRACTOGEL® 또는 ESHMUNO® 크로마토그래피 수지(이엠티 밀리포어 코포레이션)에 커플링시킨다.
- [0167] 활성화를 위한 제제에서, 5 ml의 습윤 FRACTOGEL® COO(FG-COO) 수지를 뷔히너 깔때기(Buechner funnel)에서 탈이온수로 1회 세척하고, 150 mM 2-(N-모르폴리노)에탄설펜산, pH=6.5(MES 완충제)로 3회 세척하고, 스코트(Schott) 유리병으로 옮긴다. 0.1035 gm의 1-에틸-3-(3-디메틸아미노-프로필)카르보디이미드(EDC)를 3 ml의 MES 완충제에 용해시키고, FG-COO에 첨가한다. 혼합물을 실온에서 2분 동안 인큐베이션시킨다. 4 ml의 MES 완충제 중 0.1372 gm의 N-(3-아미노프로필)말레이미드 트리플루오로아세트산(APM)의 용액을 첨가하고, 혼합물을 교반하면서, 하룻밤 실온에 둔다. 1 M NaOH를 사용한 적정을 통해 pH를 6.5로 유지한다. 보관을 위하여, 활성화 수지를 20% 에탄올을 함유하는 150 mM NaCl 중에 재현탁화시키고, 냉장고에 보관한다. 기능성의 분석을 위하여, 50 v/v% 활성화 수지 용액을 150 mM NaCl을 함유하는 100 mM 인산염 완충제, pH=7.2(PO 완충제)에서 제조한다. 0.5 ml의 활성화 FG-COO를 PO 완충제 중 204 μM 시스테인 하이드로클로라이드 용액 1 ml과 혼합하고, 1시간 동안 인큐베이션시킨다. AMP로 활성화되지 않은 FG-COO의 시료를 음성 대조군으로서 병행하여 처리한다. 그 다음, 엘만 시약(5,5'-디티오-비스-(2-니트로벤조산)) 및 알려져 있는 방법을 사용한 유리 설피하이드릴기의 분석을 위해, 수지를 PO 완충제, 0.5 M NaCl로 대대적으로 세척하고, 0.5 M NaCl 중에 재현탁화시켰다. 건조 수지 그램 당 최대 400 μmol의 리간드 농도를 이러한 분석을 사용하여 결정한다.
- [0168] **실시예 11: 인테인 융합 단백질을 사용한 티오레독신의 친화성 정제**
- [0169] 본원의 실시예 10에 따라 제조된 고정화된 N-인테인 융합 단백질을 함유하는 수지를 표준 크로마토그래피 컬럼 내로 팩킹하고, GP41-1 C-인테인의 카르복시-말단에 융합된 표적 분자 티오레독신을 포함하는 CINT_TRX 융합 단백질(SEQ ID NO: 10)을 함유하는 미정제 단백질 혼합물을 4 내지 25°C 범위의 온도에서, 그리고 100 mM Tris-HCl, 25 mM NaCl, 0.1 mM 엽화아연, pH=9를 함유하는 로딩 완충제를 사용하여, 고정화된 N-인테인 융합 단백질을 함유하는 컬럼에 첨가하여, 인테인 촉매작용을 가능하게 하지 않으면서, GP41-1 N- 및 C-인테인 도메인 사이에 강력한 상호작용을 가능하게 한다.
- [0170] 그 다음, 로딩된 컬럼을 세척하여, 세제(예를 들어, 트리톤 X100, ND40) 또는 염(예를 들어, 나트륨, 암모늄 또는 칼륨의 아세트산염, 인산염, 염산염, 황산염)을 함유하는 세척 완충제를 사용하여 미결합된 및 약하게 결합된 오염물질을 제거한다.
- [0171] C-인테인 융합 단백질의 티오레독신 부분의 절단 및 용리는 절단 완충제(50 mM Tris-HCl, pH=7.0, 300 mM NaCl, 1 mM EDTA)의 첨가에 의해 달성된다. 그 다음, 절단된 티오레독신을 용리액에서 회수한다.

표 5

예시적인 분할형 인테인 및 그들의 서열

인테인	SEQ ID NO:	서열
GP41.1의 N-말단 도메인	29	CLDLKTQVQT PQGMKEISNI QVGDVLVSNT GYNEVLNVFP KSKKKSYKIT LEDGKEIICS EEHLFPTQTG EMNISGGLKE GMCLYVKE
GP41.8의 N-말단 도메인	30	CLSLDTMVVT NGKAIEIRDV KVGDWLESEC GPVQVTEVLP IIKQPVFEIV LKSGKKIRVS ANHKFPTKDG LKTINSGLKV GDFLRSRA
NrdJ1의 N-말단 도메인	31	CLVGSSEIIT RNYGKTTIKE VVEIFDNDKN IQVLAFNTHT DNIEWAPIKA AQLTRPNAEL VELEINTLHG VKTIRCTPDH PVYTKNRDYV RADELTDDE LVVAI
IMPDH1의 N-말단 도메인	32	CFVPGTLVNT ENGLKKIEEI KVGDKVFSHT GKLQEVVDTL IFDRDEEIIIS INGIDCTKNH EFYVIDKENA NRVNEDNIHL FARWVHAEEL DMKKHLLIEL E
NrdA-2의 N-말단 도메인	33	CLTGDAKIDV LIDNIPISQI SLEEVVNLFN EGKEIYVLSY NIDTKEVEYK EISDAGLISE SAEVLEIIDE ETGQKIVCTP DHKVYTLNRG YVSAKDLKED DELVFS
Npu DNA-E의 N-말단 도메인 (Genbank 수탁 번호 ZP_00111398)	34	CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR GEQEVFEYCL EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL PN
Ssp DNA-B의 N-말단 도메인 (Genbank 수탁 번호 Q55418)	35	CISGDSLISL ASTGKRVSİK DLLDEKDFEI WAINEQTMKL ESAKVS RVFC TGKLVYILK TRLGRTİKAT ANHRFLTİDG WKRLDELSLK EHIALPRKLE SSSLQ
GP41.1의 C-말단 도메인	9	MMLKKILKIE ELDERELIDI EVSGNHLFYA <u>NDILTHN</u>
GP41.8의 C-말단 도메인	36	MCEIFENEID WDEIASIEYV GVEETIDINV TNDRLFF <u>ANG ILTHN</u>
NrdJ1의 C-말단 도메인	37	MEAKTYIGKL KSRKIVSNED TYDIQTSTHN FF <u>ANDILVHN</u>
IMPDH1의 C-말단 도메인	38	MKFKLKEITS IETKHYKGKV HDLTVNQDHS YN <u>VRGTVVHN</u>

[0172]

NrdA-2의 C-말단 도메인	39	MGLKIIKRES KEPVFDITVK DNSNFF <u>ANNI LVHN</u>
Npu DNA-E의 C-말단 도메인 (Genbank 수탁 번호 ZP_00111398)	40	MIKIATRKYL GKQNVYDIGV ERDHNFAL <u>KN GFIASN</u>
Ssp DNA-B의 C-말단 도메인 (Genbank 수탁 번호 Q55418)	41	SPEIEKLSQS DIYWDSIVSI TETGVVEEVF DLTVPGPHNF <u>VANDIIVHN</u>

밑줄이 있는 서열은 인테인 N-말단 도메인의 N1 박스에 상응한다. 밑줄 2줄이 있는 서열은 인테인 C-말단 도메인의 C1 박스(예를 들어, 엑스테인의 제1 아미노산 결어)에 상응한다.

[0173]

[0174]

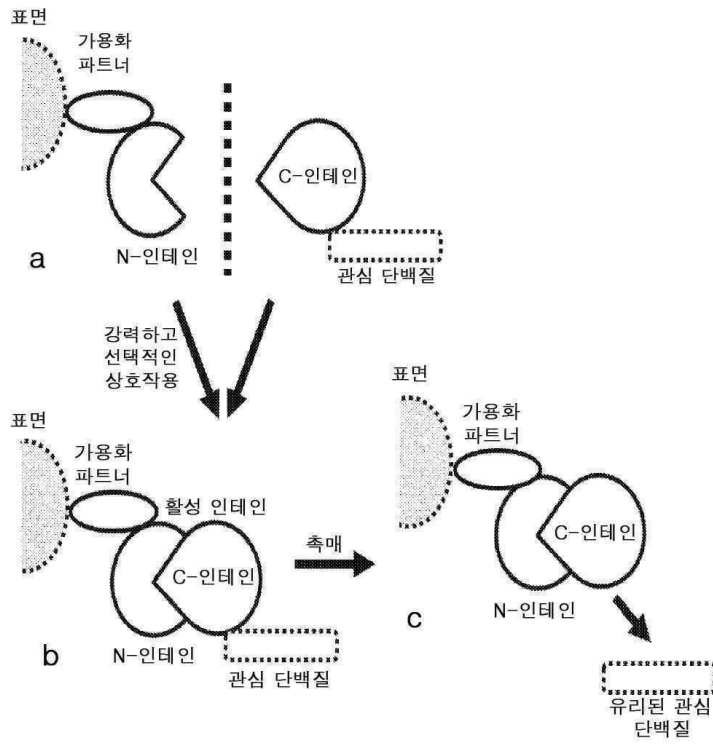
추가 예시적인 N-인테인 서열

- [0175] gp 41-2
- [0176] CLDLKTQVQTQQGLKDISNIQVGDLVL (SEQ ID NO:42)
- [0177] gp 41-3
- [0178] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKS (SEQ ID NO:43)
- [0179] gp 41-4
- [0180] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYPKITLEDGKEII
CSEEHLPFTQTGEMNISGGLKEGMCLYVKE (SEQ ID NO:44)
- [0181] gp 41-5
- [0182] CLDLKTQVQTPQGMKEISNIQVGDLVLSNTGYNEVLNVFPKSKKKSYPKITLEDGKEII
CSEEHLPFTQTGEMNISGGLKEGMCLYVKE (SEQ ID NO:45)
- [0183] gp 41-6
- [0184] SYKITLEDGKEIICSEEHLPFTQNGEVNIKGGGLKEGMCLYVKE (SEQ ID NO:46)
- [0185] gp 41-7
- [0186] MMLKKILKIEELDERELIDIEVSGNH (SEQ ID NO:47)
- [0187] NrdA-1
- [0188] CVAGDTKIKIKYPESVGDQYGTWYWNVLEKEIQIEDLEDYIIMRECEIYDSNAPQIEV
LSYNIETGEQEWKPITAFQAQTSPKAKVMKITDEESGKSIVVTPEHQVFTKNRGYVMA
KDLIETDEPIIVNKDMNF (SEQ ID NO:48)
- [0189] NrdA-4
- [0190] CLAGDTTIVTVLEGDIVFEMTLENLVSLYKNVFSVSVLSFNPETQKQEFKPVNTAALM
NPESKVLKITDSDTGKSIVCTPDHKVFTKNRGYVIASELNAEDILEIK (SEQ ID NO:49)
- [0191] NrdA-5
- [0192] HTETVRRVGTITAFQAQTSPKSKVMKITDEESGNSIVVTPEHKVFTKNRGYVMAKNLV
ETDELVIN (SEQ IDNO:50)
- [0193] NrdA-6
- [0194] YVCSRDDTTGFKLICTPDHMIYTKNRGYIMAKYLKEDDELLINEIHLPT (SEQ ID
NO:51)
- [0195] NrdJ-1
- [0196] CLVGSSEIITRNYGKTTIKEVVEIFDNDKNIQVLA FNTHTDNIEWAPIKAAQLTRPNAE
LVELEIDTLHGVK TIRCTPDHPVYTKNRGYVRADEL TDDDELVVAI (SEQ ID NO:52)
- [0197] NrdJ-2
- [0198] CLVGSSEIITRNYGKTTIKEVVEIFDNDKNIQVLA FNTHTDNIEWAPIKAAQLTRPNAE
LVELEINTLHGVK TIRCTPDHPVYTKNRDYVRADEL TDDDELVVAI (SEQ ID NO:53)
- [0199] **추가된 예시적인 C-인테인 서열**
- [0200] gp 41-9
- [0201] MIMKNRERFITEKILNIEEIDDDLTVDIGMDNEDHYFVANDILTHNT (SEQ ID NO:54)

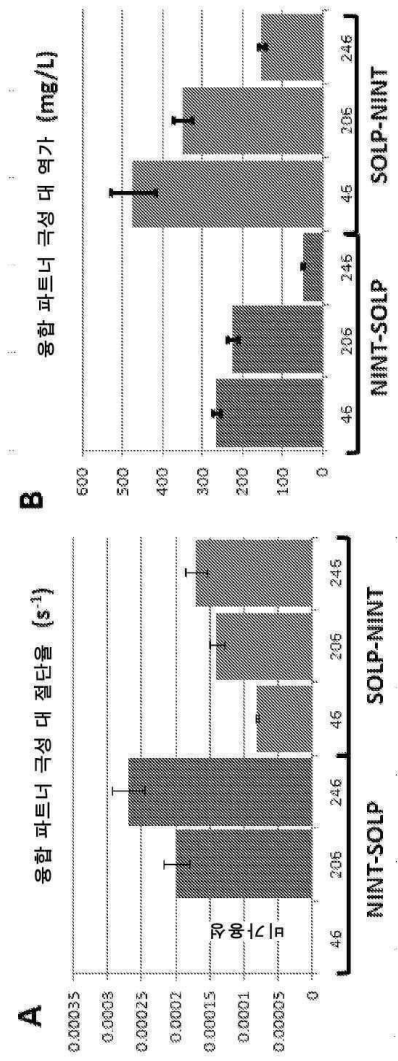
- [0202] IMPDH-2
- [0203] MKFTLEPITKIDSYEVTAEPVYDIEVENDHSFCVcNGFVVHNS (SEQ ID NO:55)
- [0204] IMPDH-3
- [0205] MKFKLVEITSKETFNYSGQ-VHDLTVEDDHSYSI-NNIVVHNS (SEQ ID NO:56),
- [0206] NrdA-3
- [0207] MLKIEYLEEEIPVYDITVEETHNFFANDILIHNC (SEQ ID NO:57),
- [0208] NrdA-5
- [0209] MLKIEYLEEEIPVYDITVEGTHNLAYSL (SEQ ID NO:58),
- [0210] NrdA-6
- [0211] MGIKIRKLEQNRVYDIKVEKIIIFCNNILVHNC (SEQ ID NO:59), 및
- [0212] NrdJ-1
- [0213] MEAKTYIGKLSRKIVSNEDTYDIQTSTHNFFANDILVHNS (SEQ ID NO:60)
- [0214] 본원에 열거된 모든 특허, 공개된 출원 및 참고문헌의 관련 교시는 그들 전문이 참조로 포함된다.
- [0215] 달리 나타내지 않는 한, 청구범위를 포함하는 명세서에 사용되는 성분의 양, 발현 조건, 처리 조건 등을 표현하는 모든 숫자는 모든 예에서 용어 "약"에 의해 수식되는 것으로 이해되어야 한다. 따라서, 반대로 나타나지 않는 한, 수치 파라미터는 근사치이며, 본 발명에 의해 수득되어야 하는 요망되는 특성에 따라 달라질 수 있다. 달리 나타내지 않는 한, 일련의 요소 앞의 용어 "적어도"는 일련의 모든 요소를 지칭하는 것으로 이해된다. 해당 분야의 숙련자는 단지 통상적인 실험을 사용하여 본원에 기재된 본 발명의 특정 구현예에 대한 많은 등가물을 인식하거나 확인할 수 있을 것이다. 그러한 등가물은 하기의 청구범위에 포함되는 것으로 의도된다.
- [0216] 본 발명이 특히 본 발명의 예시적인 구현예를 참조하여 나타나 있고, 기재되어 있지만, 첨부된 청구범위에 의해 포함되는 본 발명의 범주로부터 벗어나지 않고, 형태 및 상세사항의 다양한 변경이 그 안에서 이루어질 수 있음이 해당 분야의 숙련자에 의해 이해될 것이다.

도면

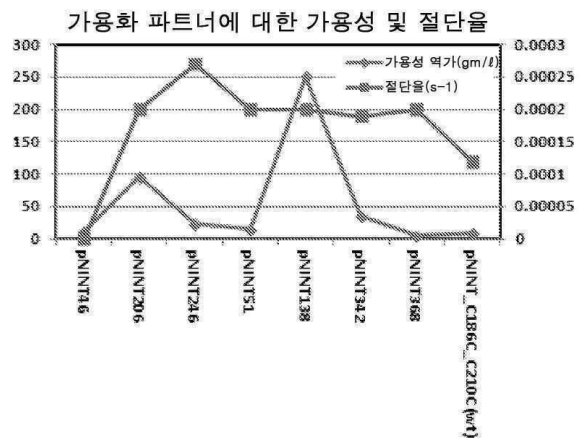
도면1



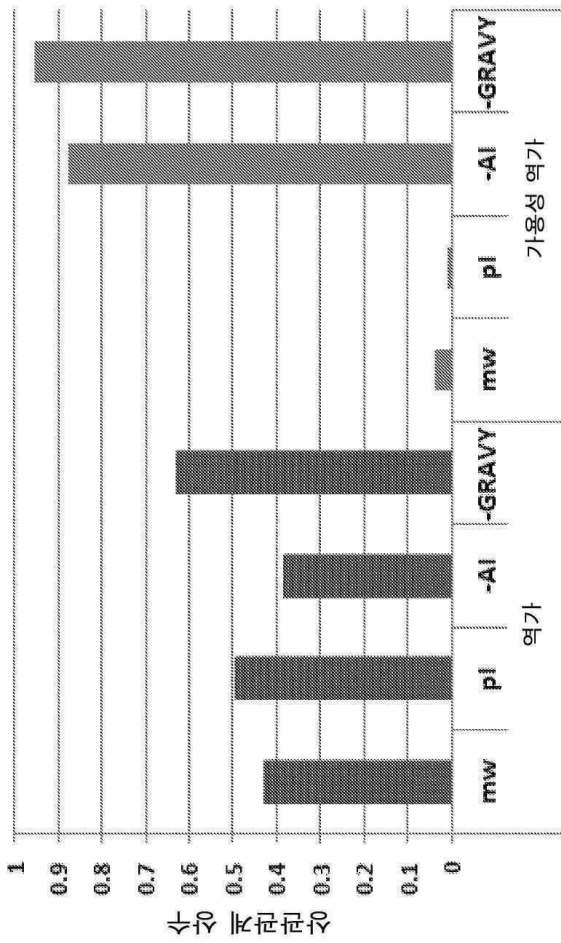
도면2



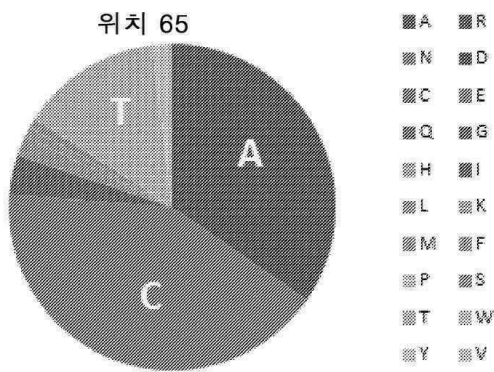
도면3



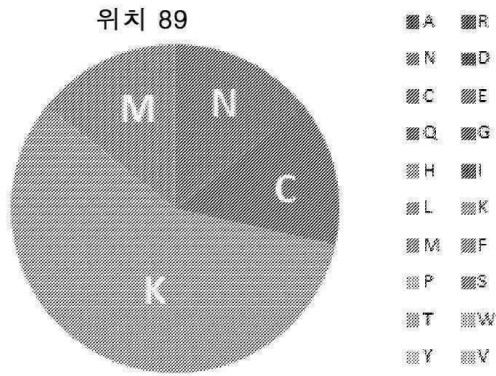
도면4



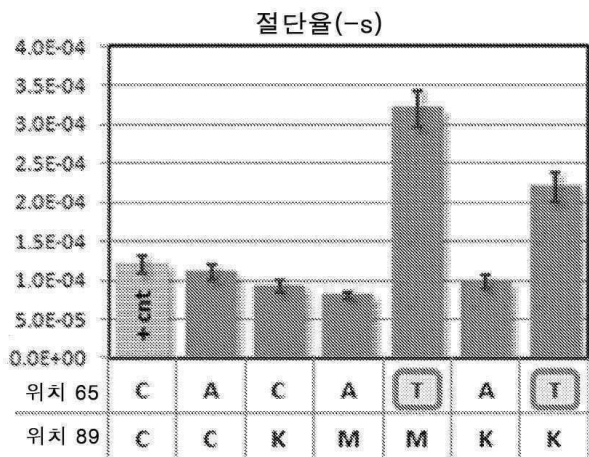
도면5a



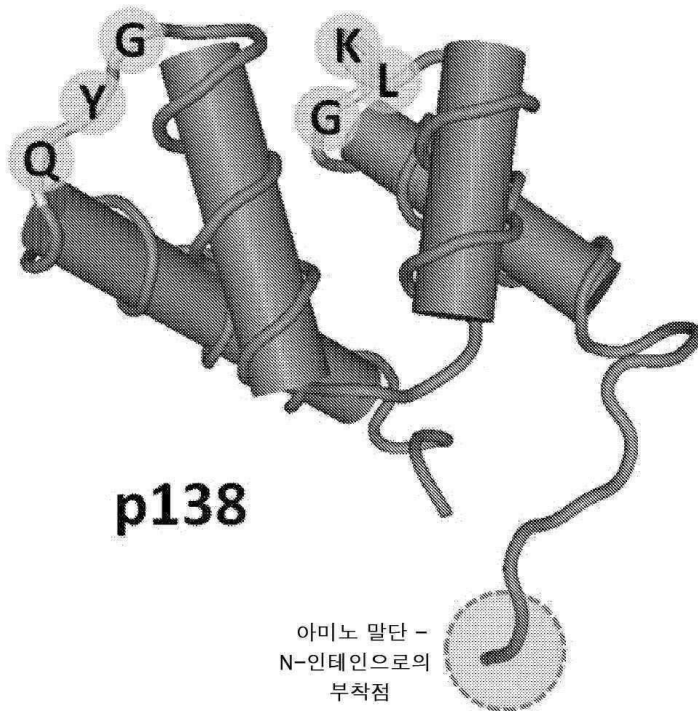
도면5b



도면6



도면7



서열목록

SEQUENCE LISTING

<110> Merck Patent GmbH

Zillmann, Martin

Orlando, Joe

<120> Soluble Intein Fusion Proteins And

Methods For Purifying Biomolecules

<130> 0046.2053-003

<140> 15/511,908

<141> 2017-03-16

<150> PCT/US2015/057125

<151> 2015-10-23

<150> US 62/074,494

<151> 2014-11-03

<150> US 62/209,010

<151> 2015-08-24

<160> 63

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein with flanking non-intein sequences

<400> 1

Met Thr Arg Ser Gly Tyr Cys Leu Asp Leu Lys Thr Gln Val Gln Thr

1 5 10 15

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val

20 25 30

Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser

35 40 45

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile

50 55 60

Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile

65 70 75 80

Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly

85 90 95

<210> 2

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein
sequences

<400> 2

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr

1 5 10 15

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val

20 25 30

Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser

35 40 45

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60

Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile

65 70 75 80

Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly

85 90 95

<210> 3

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein
 sequences

<400> 3

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr

1 5 10 15

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val

20 25 30

Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser

35 40 45

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile

50 55 60

Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile

65 70 75 80

Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly

85 90 95

<210> 4

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein

sequences

<400> 4

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1 5 10 15
 Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
 20 25 30
 Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
 35 40 45
 Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60
 Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
 65 70 75 80
 Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
 85 90 95

<210> 5

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein

sequences

<400> 5

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1 5 10 15
 Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
 20 25 30
 Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
 35 40 45
 Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60
 Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
 65 70 75 80

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val

20 25 30

Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser

35 40 45

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile

50 55 60

Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile

65 70 75 80

Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly

85 90 95

<210> 8

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein sequences

<400> 8

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr

1 5 10 15

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val

20 25 30

Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser

35 40 45

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile

50 55 60

Thr Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile

65 70 75 80

Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly

85 90 95

<210> 9

<211> 42

<212> PRT

<213> cyanophage

<400> 9

Met Gly Lys Asn Ser Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu
 1 5 10 15
 Leu Asp Glu Arg Glu Leu Ile Asp Ile Glu Val Ser Gly Asn His Leu
 20 25 30
 Phe Tyr Ala Asn Asp Ile Leu Thr His Asn
 35 40

<210> 10

<211> 157

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 C-intein-thioredoxin fusion protein

<400> 10

Met Gly Lys Asn Ser Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu

 1 5 10 15
 Leu Asp Glu Arg Glu Leu Ile Asp Ile Glu Val Ser Gly Asn His Leu
 20 25 30
 Phe Tyr Ala Asn Asp Ile Leu Thr His Asn Met Ser Asp Lys Ile Ile
 35 40 45
 His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 50 55 60
 Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met

 65 70 75 80
 Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
 85 90 95
 Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
 100 105 110
 Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
 115 120 125

Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu

130 135 140

Phe Leu Asp Ala Asn Leu Ala His His His His His His

145 150 155

<210> 11

<211> 51

<212> PRT

<213> E. coli

<400> 11

Met Arg Glu Tyr Pro Asn Gly Glu Lys Thr His Leu Thr Val Met Ala

1 5 10 15

Ala Gly Phe Pro Ser Leu Thr Gly Asp His Lys Val Ile Tyr Val Ala

20 25 30

Ala Asp Arg His Val Thr Ser Glu Glu Ile Leu Glu Ala Ala Ile Arg

35 40 45

Leu Leu Ser

50

<210> 12

<211> 77

<212> PRT

<213> E. coli

<400> 12

Met Ser His Leu Asp Glu Val Ile Ala Arg Val Asp Ala Ala Ile Glu

1 5 10 15

Glu Ser Val Ile Ala His Met Asn Glu Leu Leu Ile Ala Leu Ser Asp

20 25 30

Asp Ala Glu Leu Ser Arg Glu Asp Arg Tyr Thr Gln Gln Gln Arg Leu

35 40 45

Arg Thr Ala Ile Ala His His Gly Arg Lys His Lys Glu Asp Met Glu

50 55 60

Ala Arg His Glu Gln Leu Thr Lys Gly Gly Thr Ile Leu

65 70 75

<210> 13

<211> 83

<212> PRT

<213> E. coli

<400> 13

Met Asn Lys Glu Thr Gln Pro Ile Asp Arg Glu Thr Leu Leu Lys Glu

1 5 10 15

Ala Asn Lys Ile Ile Arg Glu His Glu Asp Thr Leu Ala Gly Ile Glu

 20 25 30

Ala Thr Gly Val Thr Gln Arg Asn Gly Val Leu Val Phe Thr Gly Asp

 35 40 45

Tyr Phe Leu Asp Glu Gln Gly Leu Pro Thr Ala Lys Ser Thr Ala Val

 50 55 60

Phe Asn Met Phe Lys His Leu Ala His Val Leu Ser Glu Lys Tyr His

65 70 75 80

Leu Val Asp

<210> 14

<211> 53

<212> PRT

<213> E. coli

<400> 14

Met Ser Leu Glu Asn Ala Pro Asp Asp Val Lys Leu Ala Val Asp Leu

1 5 10 15

Ile Val Leu Leu Glu Glu Asn Gln Ile Pro Ala Ser Thr Val Leu Arg

 20 25 30

Ala Leu Asp Ile Val Lys Arg Asp Tyr Glu Lys Lys Leu Thr Arg Asp

 35 40 45

Asp Glu Ala Glu Lys

50

<210> 15

<211> 69

<212> PRT

<213> E. coli

<400> 15

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys

1 5 10 15

Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile

20 25 30

Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly

35 40 45

Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr Arg

50 55 60

Asn Glu Tyr Arg Trp

65

<210> 16

<211> 70

<212> PRT

<213> E. coli

<400> 16

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys

1 5 10 15

Val Lys Glu Gln Trp Gly Cys Lys Leu Thr Asp Asp Asp Met Thr Ile

20 25 30

Ile Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr

35 40 45

Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr

50 55 60

Arg Asn Glu Tyr Arg Trp

65 70

<210> 17

<211> 70

<212> PRT

<213> E. coli

<400> 17

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1 5 10 15
 Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
 20 25 30
 Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
 35 40 45
 Cys Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr

 50 55 60
 Arg Asn Glu Tyr Arg Trp
 65 70
 <210> 18
 <211> 71
 <212> PRT
 <213> E. coli
 <400> 18

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1 5 10 15
 Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
 20 25 30
 Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
 35 40 45

Cys Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu
 50 55 60
 Thr Arg Asn Glu Tyr Arg Trp
 65 70
 <210> 19
 <211> 92
 <212> PRT
 <213> E. coli
 <400> 19

Met Ile Ala Glu Phe Glu Ser Arg Ile Leu Ala Leu Ile Asp Gly Met
 1 5 10 15

Val Asp His Ala Ser Asp Asp Glu Leu Phe Ala Ser Gly Tyr Leu Arg
 20 25 30
 Gly His Leu Thr Leu Ala Ile Ala Glu Leu Glu Ser Gly Asp Asp His
 35 40 45
 Ser Ala Gln Ala Val His Thr Thr Val Ser Gln Ser Leu Glu Lys Ala
 50 55 60
 Ile Gly Ala Gly Glu Leu Ser Pro Arg Asp Gln Ala Leu Val Thr Asp
 65 70 75 80
 Met Trp Glu Asn Leu Phe Gln Gln Ala Ser Gln Gln

85 90

<210> 20

<211> 95

<212> PRT

<213> E. coli

<400> 20

Met Gln Leu Asn Ile Thr Gly Asn Asn Val Glu Ile Thr Glu Ala Leu
 1 5 10 15
 Arg Glu Phe Val Thr Ala Lys Phe Ala Lys Leu Glu Gln Tyr Phe Asp
 20 25 30
 Arg Ile Asn Gln Val Tyr Val Val Leu Lys Val Glu Lys Val Thr His
 35 40 45
 Thr Ser Asp Ala Thr Leu His Val Asn Gly Gly Glu Ile His Ala Ser
 50 55 60
 Ala Glu Gly Gln Asp Met Tyr Ala Ala Ile Asp Gly Leu Ile Asp Lys
 65 70 75 80
 Leu Ala Arg Gln Leu Thr Lys His Lys Asp Lys Leu Lys Gln His
 85 90 95

<210> 21

<211> 192

<212> PRT

<213> E. coli

<400> 21

Met Asp Thr Ser Asn Ala Thr Ser Val Val Asn Val Ser Ala Ser Ser
 1 5 10 15

Ser Thr Ser Thr Ile Tyr Asp Leu Gly Asn Met Ser Lys Asp Glu Val
 20 25 30

Val Lys Leu Phe Glu Glu Leu Gly Val Phe Gln Ala Ala Ile Leu Met
 35 40 45

Phe Ser Tyr Met Tyr Gln Ala Gln Ser Asn Leu Ser Ile Ala Lys Phe
 50 55 60

Ala Asp Met Asn Glu Ala Ser Lys Ala Ser Thr Thr Ala Gln Lys Met
 65 70 75 80

Ala Asn Leu Val Asp Ala Lys Ile Ala Asp Val Gln Ser Ser Thr Asp
 85 90 95

Lys Asn Ala Lys Ala Lys Leu Pro Gln Asp Val Ile Asp Tyr Ile Asn
 100 105 110

Asp Pro Arg Asn Asp Ile Ser Val Thr Gly Ile Ser Asp Leu Ser Gly
 115 120 125

Asp Leu Ser Ala Gly Asp Leu Gln Thr Val Lys Ala Ala Ile Ser Ala
 130 135 140

Lys Ala Asn Asn Leu Thr Thr Val Val Asn Asn Ser Gln Leu Glu Ile
 145 150 155 160

Gln Gln Met Ser Asn Thr Leu Asn Leu Leu Thr Ser Ala Arg Ser Asp
 165 170 175

Val Gln Ser Leu Gln Tyr Arg Thr Ile Ser Ala Ile Ser Leu Gly Lys
 180 185 190

<210> 22

<211> 68

<212> PRT

<213> Fasciola hepatica

<400> 22

Met Pro Ser Val Glu Val Glu Lys Leu Leu His Val Leu Asp Arg Asn

1 5 10 15

Gly Asp Gly Lys Val Ser Ala Glu Glu Leu Lys Ala Phe Ala Asp Asp
 20 25 30
 Ser Lys Tyr Pro Leu Asp Ser Asn Lys Ile Lys Ala Phe Ile Lys Glu
 35 40 45
 His Asp Lys Asn Lys Asp Gly Lys Leu Asp Leu Lys Glu Leu Val Ser
 50 55 60
 Ile Leu Ser Ser
 65
 <210> 23

<211> 11
 <212> PRT
 <213> Fasciola hepatica
 <400> 23

Met Pro Ser Val Glu Val Glu Lys Leu Leu His
 1 5 10
 <210> 24
 <211> 328
 <212> PRT
 <213> E. coli
 <400> 24

Met Gly Gln Leu Ile Asp Gly Val Trp His Asp Thr Trp Tyr Asp Thr
 1 5 10 15
 Lys Ser Thr Gly Gly Lys Phe Gln Arg Ser Ala Ser Ala Phe Arg Asn
 20 25 30
 Trp Leu Thr Ala Asp Gly Ala Pro Gly Pro Thr Gly Lys Gly Gly Phe
 35 40 45
 Ala Ala Glu Lys Asp Arg Tyr His Leu Tyr Val Ser Leu Ala Cys Pro
 50 55 60
 Trp Ala His Arg Thr Leu Ile Met Arg Lys Leu Lys Gly Leu Glu Pro
 65 70 75 80
 Phe Ile Ser Val Ser Val Val Asn Pro Leu Met Leu Glu Asn Gly Trp
 85 90 95

Thr Phe Asp Asp Ser Phe Pro Gly Ala Thr Gly Asp Thr Leu Tyr Gln
 100 105 110
 His Glu Phe Leu Tyr Gln Leu Tyr Leu His Ala Asp Pro His Tyr Ser
 115 120 125
 Gly Arg Val Thr Val Pro Val Leu Trp Asp Lys Lys Asn His Thr Ile
 130 135 140
 Val Ser Asn Glu Ser Ala Glu Ile Ile Arg Met Phe Asn Thr Ala Phe
 145 150 155 160
 Asp Ala Leu Gly Ala Lys Ala Gly Asp Tyr Tyr Pro Pro Ala Leu Gln
 165 170 175
 Pro Lys Ile Asp Glu Leu Asn Gly Trp Ile Tyr Asp Thr Val Asn Asn
 180 185 190
 Gly Val Tyr Lys Ala Gly Phe Ala Thr Ser Gln Gln Ala Tyr Asp Glu
 195 200 205
 Ala Val Ala Lys Val Phe Glu Ser Leu Ala Arg Leu Glu Gln Ile Leu
 210 215 220
 Gly Gln His Arg Tyr Leu Thr Gly Asn Gln Leu Thr Glu Ala Asp Ile
 225 230 235 240
 Arg Leu Trp Thr Thr Leu Val Arg Phe Asp Pro Val Tyr Val Thr His
 245 250 255
 Phe Lys Cys Asp Lys His Arg Ile Ser Asp Tyr Leu Asn Leu Tyr Gly
 260 265 270
 Phe Leu Arg Asp Ile Tyr Gln Met Pro Gly Ile Ala Glu Thr Val Asn
 275 280 285
 Phe Asp His Ile Arg Asn His Tyr Phe Arg Ser His Lys Thr Ile Asn
 290 295 300
 Pro Thr Gly Ile Ile Ser Ile Gly Pro Trp Gln Asp Leu Asp Glu Pro
 305 310 315 320
 His Gly Arg Asp Val Arg Phe Gly
 325

<210> 25

<211> 120

<212> PRT

<213> Enterobacteria phage lambda

<400> 25

Met Ala Ser Trp Ser His Pro Gln Phe Glu Lys Ala Ser Lys Glu Thr

1 5 10 15

Phe Thr His Tyr Gln Pro Gln Gly Asn Ser Asp Pro Ala His Thr Ala

20 25 30

Thr Ala Pro Gly Gly Leu Ser Ala Lys Ala Pro Ala Met Thr Pro Leu

35 40 45

Met Leu Asp Thr Ser Ser Arg Lys Leu Val Ala Trp Asp Gly Thr Thr

50 55 60

Asp Gly Ala Ala Val Gly Ile Leu Ala Val Ala Ala Asp Gln Thr Ser

65 70 75 80

Thr Thr Leu Thr Phe Tyr Lys Ser Gly Thr Phe Arg Tyr Glu Asp Val

85 90 95

Leu Trp Pro Glu Ala Ala Ser Asp Glu Thr Lys Lys Arg Thr Ala Phe

100 105 110

Ala Gly Thr Ala Ile Ser Ile Val

115 120

<210> 26

<211> 396

<212> PRT

<213> E. coli

<400> 26

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr

1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys

20 25 30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu

35 40 45

Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
 305 310 315 320

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
 325 330 335

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys

 340 345 350

Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
 355 360 365

Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
 370 375 380

Glu Ala Leu Lys Asp Ala Gln Thr Arg Ile Thr Lys

385 390 395

<210> 27

<211> 109

<212> PRT

<213> E. coli

<400> 27

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp

1 5 10 15

Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
 20 25 30

Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
 35 40 45

Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
 50 55 60

Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu

65 70 75 80

Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
 85 90 95

Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala

100 105

<210> 28

<211> 495

<212> PRT

<213> E. coli

<400> 28

Met Asn Lys Glu Ile Leu Ala Val Val Glu Ala Val Ser Asn Glu Lys

1 5 10 15

Ala Leu Pro Arg Glu Lys Ile Phe Glu Ala Leu Glu Ser Ala Leu Ala

20 25 30

Thr Ala Thr Lys Lys Lys Tyr Glu Gln Glu Ile Asp Val Arg Val Gln

35 40 45

Ile Asp Arg Lys Ser Gly Asp Phe Asp Thr Phe Arg Arg Trp Leu Val

50 55 60

Val Asp Glu Val Thr Gln Pro Thr Lys Glu Ile Thr Leu Glu Ala Ala

65 70 75 80

Arg Tyr Glu Asp Glu Ser Leu Asn Leu Gly Asp Tyr Val Glu Asp Gln

85 90 95

Ile Glu Ser Val Thr Phe Asp Arg Ile Thr Thr Gln Thr Ala Lys Gln

100 105 110

Val Ile Val Gln Lys Val Arg Glu Ala Glu Arg Ala Met Val Val Asp

115 120 125

Gln Phe Arg Glu His Glu Gly Glu Ile Ile Thr Gly Val Val Lys Lys

130 135 140

Val Asn Arg Asp Asn Ile Ser Leu Asp Leu Gly Asn Asn Ala Glu Ala

145 150 155 160

Val Ile Leu Arg Glu Asp Met Leu Pro Arg Glu Asn Phe Arg Pro Gly

165 170 175

Asp Arg Val Arg Gly Val Leu Tyr Ser Val Arg Pro Glu Ala Arg Gly

180 185 190

Ala Gln Leu Phe Val Thr Arg Ser Lys Pro Glu Met Leu Ile Glu Leu

195 200 205

Phe Arg Ile Glu Val Pro Glu Ile Gly Glu Glu Val Ile Glu Ile Lys

Asp Asp Leu Ala Asp Ile Glu Gly Leu Thr Asp Glu Lys Ala Gly Ala

465 470 475 480

Leu Ile Met Ala Ala Arg Asn Ile Cys Trp Phe Gly Asp Glu Ala

 485 490 495

<210> 29

<211> 88

<212> PRT

<213> GP41-1 N-intein (cyanophage)

<400> 29

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu

1 5 10 15

Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr

 20 25 30

Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys

 35 40 45

Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu

 50 55 60

Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu

65 70 75 80

Gly Met Cys Leu Tyr Val Lys Glu

 85

<210> 30

<211> 88

<212> PRT

<213> Unknown

<220>

<223> N-terminal domain of GP41.8

<400> 30

Cys Leu Ser Leu Asp Thr Met Val Val Thr Asn Gly Lys Ala Ile Glu

1 5 10 15

Ile Arg Asp Val Lys Val Gly Asp Trp Leu Glu Ser Glu Cys Gly Pro

 20 25 30

Val Gln Val Thr Glu Val Leu Pro Ile Ile Lys Gln Pro Val Phe Glu

35 40 45

Ile Val Leu Lys Ser Gly Lys Lys Ile Arg Val Ser Ala Asn His Lys

50 55 60

Phe Pro Thr Lys Asp Gly Leu Lys Thr Ile Asn Ser Gly Leu Lys Val

65 70 75 80

Gly Asp Phe Leu Arg Ser Arg Ala

85

<210> 31

<211> 105

<212> PRT

<213> Unknown

<220>

<223> N-terminal domain of NrdJ1

<400> 31

Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr

1 5 10 15

Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln

20 25 30

Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile

35 40 45

Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu

50 55 60

Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His

65 70 75 80

Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr

85 90 95

Asp Asp Asp Glu Leu Val Val Ala Ile

100 105

<210> 32

<211> 101

<212> PRT

<213> Unknown

<220>

<223> N-terminal domain of IMPDH1

<400> 32

Cys Phe Val Pro Gly Thr Leu Val Asn Thr Glu Asn Gly Leu Lys Lys

1 5 10 15

Ile Glu Glu Ile Lys Val Gly Asp Lys Val Phe Ser His Thr Gly Lys

 20 25 30

Leu Gln Glu Val Val Asp Thr Leu Ile Phe Asp Arg Asp Glu Glu Ile

 35 40 45

Ile Ser Ile Asn Gly Ile Asp Cys Thr Lys Asn His Glu Phe Tyr Val

 50 55 60

Ile Asp Lys Glu Asn Ala Asn Arg Val Asn Glu Asp Asn Ile His Leu

65 70 75 80

Phe Ala Arg Trp Val His Ala Glu Glu Leu Asp Met Lys Lys His Leu

 85 90 95

Leu Ile Glu Leu Glu

 100

<210> 33

<211> 106

<212> PRT

<213> Unknown

<220>

<223> N-terminal domain of NrdA-2

<400> 33

Cys Leu Thr Gly Asp Ala Lys Ile Asp Val Leu Ile Asp Asn Ile Pro

1 5 10 15

Ile Ser Gln Ile Ser Leu Glu Glu Val Val Asn Leu Phe Asn Glu Gly

 20 25 30

Lys Glu Ile Tyr Val Leu Ser Tyr Asn Ile Asp Thr Lys Glu Val Glu

 35 40 45

Tyr Lys Glu Ile Ser Asp Ala Gly Leu Ile Ser Glu Ser Ala Glu Val

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

<210> 34

<211> 102

<212> PRT

<213> Nostoc punctiforme

<400> 34

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu
 1 5 10 15
 Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
 20 25 30
 Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
 35 40 45
 Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
 50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
 65 70 75 80
 Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
 85 90 95
 Val Asp Asn Leu Pro Asn
 100

<210> 35

<211> 105

<212> PRT

<213> Synechocystis

<400> 35

Cys Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser Thr Gly Lys Arg

1 5 10 15

Val Ser Ile Lys Asp Leu Leu Asp Glu Lys Asp Phe Glu Ile Trp Ala

 20 25 30

Ile Asn Glu Gln Thr Met Lys Leu Glu Ser Ala Lys Val Ser Arg Val

 35 40 45

Phe Cys Thr Gly Lys Lys Leu Val Tyr Ile Leu Lys Thr Arg Leu Gly

 50 55 60

Arg Thr Ile Lys Ala Thr Ala Asn His Arg Phe Leu Thr Ile Asp Gly

65 70 75 80

Trp Lys Arg Leu Asp Glu Leu Ser Leu Lys Glu His Ile Ala Leu Pro

 85 90 95

Arg Lys Leu Glu Ser Ser Ser Leu Gln

 100 105

<210> 36

<211> 45

<212> PRT

<213> Unknown

<220>

<223> C-terminal domain of GP41.8

<400> 36

Met Cys Glu Ile Phe Glu Asn Glu Ile Asp Trp Asp Glu Ile Ala Ser

1 5 10 15

Ile Glu Tyr Val Gly Val Glu Glu Thr Ile Asp Ile Asn Val Thr Asn

 20 25 30

Asp Arg Leu Phe Phe Ala Asn Gly Ile Leu Thr His Asn

 35 40 45

<210> 37

<211> 40

<212> PRT

<213> Unknown

<220>

<223> C-terminal domain of NrdJ1

<400> 37

Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val
 1 5 10 15
 Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe
 20 25 30

Ala Asn Asp Ile Leu Val His Asn
 35 40

<210> 38

<211> 40

<212> PRT

<213> Unknown

<220>

<223> C-terminal domain of IMPDH1

<400> 38

Met Lys Phe Lys Leu Lys Glu Ile Thr Ser Ile Glu Thr Lys His Tyr
 1 5 10 15
 Lys Gly Lys Val His Asp Leu Thr Val Asn Gln Asp His Ser Tyr Asn
 20 25 30

Val Arg Gly Thr Val Val His Asn
 35 40

<210>

> 39

<211> 34

<212> PRT

<213> Unknown

<220>

<223> C-terminal domain of NrdA-2

<400> 39

Met Gly Leu Lys Ile Ile Lys Arg Glu Ser Lys Glu Pro Val Phe Asp
 1 5 10 15
 Ile Thr Val Lys Asp Asn Ser Asn Phe Phe Ala Asn Asn Ile Leu Val
 20 25 30

His Asn

<210> 40

<211> 36

<212> PRT

<213> Nostoc punctiforme

<400> 40

Met Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr

1 5 10 15
 Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe
 20 25 30
 Ile Ala Ser Asn
 35

<210> 41

<211> 48

<212> PRT

<213> Synechocystis

<400> 41

Ser Pro Glu Ile Glu Lys Leu Ser Gln Ser Asp Ile Tyr Trp Asp Ser

1 5 10 15
 Ile Val Ser Ile Thr Glu Thr Gly Val Glu Glu Val Phe Asp Leu Thr
 20 25 30

Val Pro Gly Pro His Asn Phe Val Ala Asn Asp Ile Ile Val His Asn

35 40 45

<210> 42

<211> 27

<212> PRT

<213> Unknown

<220>

<223> GP41-2 N-intein sequence

<400> 42

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Gln Gln Gly Leu Lys Asp

1 5 10 15
 Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu
 20 25

<210> 43

<211> 46

<212> PRT

<213> Unknown

<

<220>

<223> GP41-3 N-intein sequence

<400> 43

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu

1 5 10 15

Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr

 20 25 30

Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser

 35 40 45

<210> 44

<211> 88

<212> PRT

<213> Unknown

<220>

<223> GP41-4 N-intein sequence

<400> 44

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu

1 5 10 15

Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr

 20 25 30

Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys

 35 40 45

Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu

50 55 60

Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu

65 70 75 80

Gly Met Cys Leu Tyr Val Lys Glu

85

<210> 45

<211> 88

<212> PRT

<213> Unknown

<220>

<223> GP41-5 N-intein sequence

<400> 45

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu

1 5 10 15

Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr

20 25 30

Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Lys Ser Tyr Lys

35 40 45

Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu

50 55 60

Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu

65 70 75 80

Gly Met Cys Leu Tyr Val Lys Glu

85

<210> 46

<211> 43

<212> PRT

<213> Unknown

<220>

<223> GP41-6 N-intein sequence

<400> 46

Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu

1 5 10 15

Glu His Leu Phe Pro Thr Gln Asn Gly Glu Val Asn Ile Lys Gly Gly

20 25 30

Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu

35 40

<210> 47

<211> 26

<212> PRT

<213> Unknown

<220>

<223> GP41-7 N-intein sequence

<400> 47

Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu Leu Asp Glu Arg Glu

1 5 10 15

Leu Ile Asp Ile Glu Val Ser Gly Asn His

20 25

<210> 48

<211> 133

<212> PRT

<213> Unknown

<220>

<223> NrdA-1 N-intein sequence

<400> 48

Cys Val Ala Gly Asp Thr Lys Ile Lys Ile Lys Tyr Pro Glu Ser Val

1 5 10 15

Gly Asp Gln Tyr Gly Thr Trp Tyr Trp Asn Val Leu Glu Lys Glu Ile

20 25 30

Gln Ile Glu Asp Leu Glu Asp Tyr Ile Ile Met Arg Glu Cys Glu Ile

35 40 45

Tyr Asp Ser Asn Ala Pro Gln Ile Glu Val Leu Ser Tyr Asn Ile Glu

50 55 60

Thr Gly Glu Gln Glu Trp Lys Pro Ile Thr Ala Phe Ala Gln Thr Ser

65 70 75 80

Pro Lys Ala Lys Val Met Lys Ile Thr Asp Glu Glu Ser Gly Lys Ser

85 90 95

Ile Val Val Thr Pro Glu His Gln Val Phe Thr Lys Asn Arg Gly Tyr

100 105 110

Val Met Ala Lys Asp Leu Ile Glu Thr Asp Glu Pro Ile Ile Val Asn

Thr Ser Pro Lys Ser Lys Val Met Lys Ile Thr Asp Glu Glu Ser Gly

20 25 30

Asn Ser Ile Val Val Thr Pro Glu His Lys Val Phe Thr Lys Asn Arg

35 40 45

Gly Tyr Val Met Ala Lys Asn Leu Val Glu Thr Asp Glu Leu Val Ile

50 55 60

Asn

65

<210> 51

<211> 49

<212> PRT

<213> Unknown

<220>

<223> NrdA-6 N-intein sequence

<400> 51

Tyr Val Cys Ser Arg Asp Asp Thr Thr Gly Phe Lys Leu Ile Cys Thr

1 5 10 15

Pro Asp His Met Ile Tyr Thr Lys Asn Arg Gly Tyr Ile Met Ala Lys

20 25 30

Tyr Leu Lys Glu Asp Asp Glu Leu Leu Ile Asn Glu Ile His Leu Pro

35 40 45

Thr

<210> 52

<211> 105

<212> PRT

<213> Unknown

<220>

<223> NrdJ-1 N-intein sequence

<400> 52

Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr

1 5 10 15

Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
 20 25 30
 Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
 35 40 45
 Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
 50 55 60
 Ile Asp Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His

65 70 75 80
 Pro Val Tyr Thr Lys Asn Arg Gly Tyr Val Arg Ala Asp Glu Leu Thr
 85 90 95
 Asp Asp Asp Glu Leu Val Val Ala Ile
 100 105

<210> 53

<211> 105

<212> PRT

<213> Unknown

<220>

<223> NrdJ2 N-intein sequence

<400> 53

Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
 1 5 10 15
 Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln

20 25 30
 Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
 35 40 45
 Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
 50 55 60
 Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
 65 70 75 80

Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr
 85 90 95

Asp Asp Asp Glu Leu Val Val Ala Ile

100 105

<210> 54

<211> 47

<212> PRT

<213> Unknown

<220>

<223> GP41-9 C-intein sequence

<400> 54

Met Ile Met Lys Asn Arg Glu Arg Phe Ile Thr Glu Lys Ile Leu Asn

1 5 10 15

Ile Glu Glu Ile Asp Asp Asp Leu Thr Val Asp Ile Gly Met Asp Asn

20 25 30

Glu Asp His Tyr Phe Val Ala Asn Asp Ile Leu Thr His Asn Thr

35 40 45

<210> 55

<211> 42

<212> PRT

<213> Unknown

<220>

<223> IMPDH-2 C-intein sequence

<400> 55

Met Lys Phe Thr Leu Glu Pro Ile Thr Lys Ile Asp Ser Tyr Glu Val

1 5 10 15

Thr Ala Glu Pro Val Tyr Asp Ile Glu Val Glu Asn Asp His Ser Phe

20 25 30

Cys Val Asn Gly Phe Val Val His Asn Ser

35 40

<210> 56

<211> 41

<212> PRT

<213> Unknown

<220>

<223> IMPDH-3 C-intein sequence

<400> 56

Met Lys Phe Lys Leu Val Glu Ile Thr Ser Lys Glu Thr Phe Asn Tyr
 1 5 10 15
 Ser Gly Gln Val His Asp Leu Thr Val Glu Asp Asp His Ser Tyr Ser
 20 25 30
 Ile Asn Asn Ile Val Val His Asn Ser
 35 40

<210> 57

<211> 34

<212> PRT

<213> Unknown

<220>

<223> NrdA-3 C-intein sequence

<400> 57

Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
 1 5 10 15
 Thr Val Glu Glu Thr His Asn Phe Phe Ala Asn Asp Ile Leu Ile His
 20 25 30
 Asn Cys

<210> 58

<211> 28

<212> PRT

<213> Unknown

<220>

<223> NrdA-5 C-intein sequence

<400> 58

Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
 1 5 10 15
 Thr Val Glu Gly Thr His Asn Leu Ala Tyr Ser Leu
 20 25

<210> 59

<211> 33

<212> PRT

<213> Unknown

<220>

<223> NrdA-6 C-intein sequence

<400> 59

Met Gly Ile Lys Ile Arg Lys Leu Glu Gln Asn Arg Val Tyr Asp Ile

1 5 10 15

Lys Val Glu Lys Ile Ile Ile Phe Cys Asn Asn Ile Leu Val His Asn

20 25 30

Cys

<210> 60

<211> 41

<212> PRT

<213> Unknown

<220>

<223> NrdJ-1 C-intein sequence

<400> 60

Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val

1 5 10 15

Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe

20 25 30

Ala Asn Asp Ile Leu Val His Asn Ser

35 40

<210> 61

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Loop region of E. coli

<400> 61

Gly Cys Lys Leu

1

<210> 62

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Loop region of E. coli

<400> 62

Gly Cys Tyr Gln

1

<210> 63

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Loop region of E. coli

<400> 63

Gly Cys Gly Tyr Gln

1

5