(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0039290
(43) 공개일자 2021년04월09일

(51) 국제특허분류(Int. Cl.)
C12N 15/67 (2006.01) *C12N 9/10* (2006.01)
(52) CPC특허분류
C12N 15/67 (2013.01)
C12N 9/104 (2013.01)
(21) 출원번호 10-2020-0122078
(22) 출원일자 2020년09월22일
심사청구일자 2020년09월22일
(30) 우선권주장
1020190121683 2019년10월01일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
송현규
서울특별시 서초구 나루터로4길 49, 332동 102호 (잠원동, 신반포17차아파트)
권도훈
인천광역시 연수구 해돋이로 51, 1동 2005호 (송도동)
김이현
충청남도 천안시 동남구 청당2길 118, 109동 1002호 (청당동, 청당벽산블루밍아파트)
(74) 대리인
특허법인 남앤남

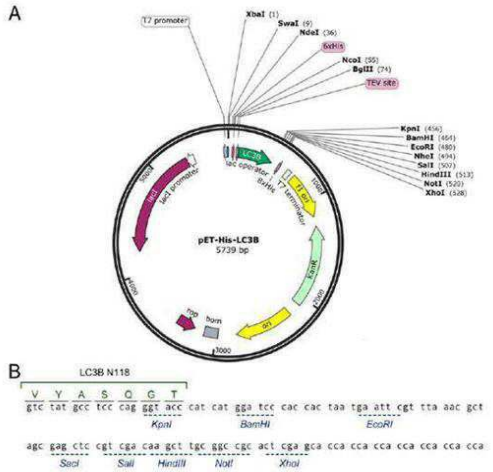
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **아미노 말단 분해 경로에 있는 단백질들의 대량생산을 위한 재조합 벡터 및 이의 용도**

(57) 요 약

본 발명은 아미노 말단 분해 경로에 있는 단백질들의 대량생산을 위한 재조합 벡터 및 이의 용도에 관한 것으로, 구체적으로 본 발명은 서열번호 1의 아미노산으로 이루어진 LC3B 단백질을 코딩하는 염기서열; 아미노산 FG를 코딩하는 염기서열; 및 목적단백질을 코딩하는 염기서열이 순차적으로 포함된 재조합 벡터, 상기 재조합 벡터로 형질전환된 형질전환체, 상기 재조합 벡터를 이용하여 목적단백질을 제조하는 방법 및 목적단백질의 결정화 방법에 관한 것이다.

대 표 도 - 도1



(52) CPC특허분류
C12Y 203/02 (2013.01)
C07K 2319/21 (2013.01)
C07K 2319/23 (2013.01)
C07K 2319/50 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711046966
과제번호	2016R1E1A1A01942623
부처명	미래창조과학부
과제관리(전문)기관명	한국연구재단
연구사업명	전략공모_전략과제
연구과제명	엑스선 자유전자 레이저를 이용한 단백질 구조 및 동력학 연구
기 여 율	1/1
과제수행기관명	고려대학교
연구기간	2016.12.01 ~ 2019.11.30

명 세 서

청구범위

청구항 1

서열번호 1의 아미노산으로 이루어진 LC3B 단백질을 코딩하는 염기서열; 아미노산 FG를 코딩하는 염기서열; 및 목적단백질을 코딩하는 염기서열이 순차적으로 포함된, 재조합 벡터.

청구항 2

제1항에 있어서,

상기 아미노산 FG는 ATG4B 프로테아제에 의해 절단되는 부위인 것을 특징으로 하는, 재조합 벡터.

청구항 3

제1항에 있어서,

상기 목적단백질은 N-데그론(degron) 경로에 관여하는, N-레코그닌(N-recognin); 상기 N-레코그닌에 대한 기질; 또는 N-레코그닌과 기질이 결합된 단백질 복합체인 것을 특징으로 하는, 재조합 벡터.

청구항 4

제3항에 있어서,

상기 N-레코그닌(N-recognin)는 SQSTM1/p62, PRT6, Ubr1 및 NTAQ1으로 이루어진 군 중에서 선택되는 것을 특징으로 하는, 재조합 벡터.

청구항 5

제3항에 있어서,

상기 기질은,

PRT6 또는 Ubr1에 대한 기질인 RLGS 펩타이드 단편,

NTAQ1에 대한 기질인 M1Q 펩타이드 단편 또는

p62 ZZ 도메인에 대한 기질인 GEEED 펩타이드 단편인 것을 특징으로 하는, 재조합 벡터.

청구항 6

제1항에 있어서,

상기 재조합 벡터는 상기 목적단백질의 정제를 위해 글루타티온-S-트랜스퍼라제(Glutathione-S-transferase, GST), MBP(Maltose-binding protein) 또는 5~7개의 히스티딘 아미노산을 코딩하는 유전자가 N 말단 또는 C 말단에 연결된 것을 특징으로 하는, 재조합 벡터.

청구항 7

제1항 내지 제6항 중 어느 한 항의 재조합 벡터로 형질전환된 형질전환체.

청구항 8

(1) 제1항의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계;

(2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계 ; 및

(3) 생성된 목적단백질을 분리 및 정제하는 단계를 포함하되, 상기 목적단백질은 ATG4B 프로테아제 처리에 의해 상기 목적단백질의 N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미

노산 잔기인 것을 특징으로 하는,

제1항의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법.

청구항 9

제8항에 있어서,

상기 목적단백질은 N-데그론(degron) 경로에 관여하는, N-레코그닌(N-recognin); 상기 N-레코그닌에 대한 기질; 또는 N-레코그닌과 기질이 결합된 단백질 복합체인 것을 특징으로 하는, 제1항의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법.

청구항 10

제9항에 있어서,

상기 N-레코그닌(N-recognin)는 SQSTM1/p62, PRT6, Ubr1 및 NTAQ1으로 이루어진 군 중에서 선택되는 것을 특징으로 하는, 제1항의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법.

청구항 11

제9항에 있어서,

상기 기질은,

PRT6 또는 Ubr1에 대한 기질인 RLGS 펩타이드 단편,

NTAQ1에 대한 기질인 M1Q 펩타이드 단편 또는

p62 ZZ 도메인에 대한 기질인 GEEED 펩타이드 단편인 것을 특징으로 하는, 제1항의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법.

청구항 12

(1) 제1항의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계;

(2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계; 및

(3) 생성된 목적단백질을 결정화시키는 단계를 포함하는,

제1항의 재조합 벡터를 이용한 목적단백질의 결정화 방법.

청구항 13

제12항에 있어서,

상기 목적단백질은 ATG4B 프로테아제 처리에 의해 상기 목적단백질의 N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미노산 잔기인 것을 특징으로 하는, 제1항의 재조합 벡터를 이용한 목적단백질의 결정화 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 아미노 말단 분해 경로에 있는 단백질들의 대량생산을 위해 LB3B 유전자를 이용한 재조합 벡터 및 이의 용도에 관한 것이다.

배 경 기 술

[0002] N-말단 법칙 경로(N-end Rule Pathway)는 특정 N-말단 아미노산을 가진 단백질을 분해하는 단백질 분해 경로로서, 상기 단백질의 인 비보 반감기와 관련이 있다. N-말단 아미노산은 N-말단 법칙 경로를 창출하는 분해 시그널(N-degron)이 필수적인 요소로 인식되어 왔다.

[0003] N-말단 법칙 경로는 단백질의 메티오닌 잔여가 잘려나간 후 노출되거나 단백질 중간이 가수분해효소에 의해서 잘린 후 노출되는 첫 아미노산 잔기의 정체에 따라 그 단백질의 분해가 매개되는 것으로, 이러한 분해를 유도하는 노출된 아미노산들을 N-데그론(N-degron)이라 부르고, 이런 N-데그론과 직접적으로 인식하여 결합하는 단백질들을 N-레코그닌(N-recognin)이라 한다.

[0004] 이러한 N-데그론(N-degron)은 단백질 분해작용, 계층적 효소 작용 및 번역 후 변형 과정 등 세포 내 다양한 프로세싱에 의해 생성된다.

[0005] 현재 진핵생물에서는 4개의 N-데그론 경로가 밝혀진 바 있는데, Arg/N-데그론 경로, Ac/N-데그론 경로, Pro/N-데그론 경로 및 formyl-Met/N-데그론 경로이다.

[0006] 한편, 이러한 N-말단 법칙 경로에 대한 연구는 현재 대부분 in vivo에서 진행되고 있을 뿐, in vitro에서 이들의 생화학적 연구 및 고해상도 구조적 연구는 많은 한계를 가지고 있다.

[0007] 특히 단백질들의 결합 친화도, 효소의 동역학 상수 결정같은 정량적 분석이나 고분해능 구조결정을 위해서는 다량의 단백질이 필요한데, 현재까지 기술로는 N-말단 법칙 경로에 관여하는 단백질을 다량으로 정제 및 수득하기가 어려운 실정이다.

[0008] 또한, N-데그론 경로, 특히 잘 연구된 Arg/N-degron 경로에서는 상이한 N-데그론을 생성하기 위해 몇 가지 변형 효소를 필요로 하는데, in vivo 즉, 세포내에서는 이러한 변형에 필요한 많은 분자들이 존재하고 있으나, in vitro 내에서는 실제적인 세포 환경과 유사한 환경을 모방하기 위해 필요한 모든 효소를 제조하기가 불가능하다.

[0009] 또한, N-데그론 경로에 있어서, N-레코그닌 또는 다른 단백질에 의한 N-데그론의 인식에 대한 명확한 이해를 위해서는 이러한 단백질 복합체들에 대한 고해상도의 구조 분석이 요구된다. 그러나 단백질 분해 이전에 N-데그론이 방출되어야 하기 때문에 표적 단백질과 N-데그론 사이의 결합 친화도가 높지 않으며, 따라서 N-데그론이 결합된 단백질 복합체의 결정 구조물을 얻는 것에는 한계가 있다.

[0010] 이에 본 발명자들은 in vitro에서 N-말단 법칙 경로(N-end Rule Pathway)에 관여하는 단백질들을 다량 정제할 수 있으며, 정제한 단백질의 활성 및 구조적 분석을 용이하게 할 수 있을 뿐만 아니라, 목적 단백질의 N-말단에 위치하는 아미노산을 목적하는 임의의 아미노산이 되도록 디자인할 수 있는, 아미노 말단 분해 경로에 있는 단백질들의 연구를 위한 새로운 벡터 시스템을 개발함으로써 본 발명을 완성하였다.

선행기술문헌

특허문헌

[0011] (특허문헌 0001) 국제공개특허 WO 2017/047846

발명의 내용

해결하려는 과제

[0012] 따라서 본 발명의 목적은 LC3B 단백질을 이용하여 아미노 말단 분해 경로에 있는 단백질을 생산할 수 있는 재조합 벡터를 제공하는 것이다.

[0013] 본 발명의 또 다른 목적은 본 발명의 재조합 벡터로 형질전환된 형질전환체를 제공하는 것이다.

[0014] 본 발명의 또 다른 목적은 본 발명의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법을 제공하는 것이다.

[0015] 본 발명의 또 다른 목적은 본 발명의 재조합 벡터를 이용한 목적단백질의 결정화 방법을 제공하는 것이다.

과제의 해결 수단

[0016] 그러므로 본 발명은 서열번호 1의 아미노산으로 이루어진 LC3B 단백질을 코딩하는 염기서열; 아미노산 FG를 코딩하는 염기서열; 및 목적단백질을 코딩하는 염기서열이 순차적으로 포함된, 재조합 벡터를 제공한다.

[0017] 본 발명의 일실시예에 있어서, 상기 아미노산 FG는 ATG4B 프로테아제에 의해 절단되는 부위일 수 있다.

[0018] 본 발명의 일실시예에 있어서, 상기 목적단백질은 N-데그론(degron) 경로에 관여하는, N-레코그닌(N-recognin); 상기 N-레코그닌에 대한 기질; 또는 N-레코그닌과 기질이 결합된 단백질 복합체일 수 있다.

[0019] 본 발명의 일실시예에 있어서, 상기 N-레코그닌(N-recognin)는 SQSTM1/p62, PRT6, Ubr1 및 NTAQ1으로 이루어진 군 중에서 선택되는 것일 수 있다.

[0020] 본 발명의 일실시예에 있어서, 상기 기질은, PRT6 또는 Ubr1에 대한 기질인 RLGS 펩타이드 단편, NTAQ1에 대한 기질인 M1Q 펩타이드 단편 또는 p62 ZZ 도메인에 대한 기질인 GEEED 펩타이드 단편일 수 있다.

[0021] 본 발명의 일실시예에 있어서, 상기 재조합 벡터는 상기 목적단백질의 정제를 위해 글루타티온-S-트랜스퍼라제(Glutathione-S-transferase, GST) 또는 5~7개의 히스티딘 아미노산을 코딩하는 유전자가 N 말단 또는 C 말단에 연결된 것일 수 있다.

[0022] 또한, 본 발명은 본 발명의 재조합 벡터로 형질전환된 형질전환체를 제공한다.

[0023] 또한, 본 발명은, (1) 본 발명의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계; (2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계; 및 (3) 생성된 목적단백질을 분리 및 정제하는 단계를 포함하되, 상기 목적단백질은 ATG4B 프로테아제 처리에 의해 상기 목적단백질의 N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미노산 잔기인 것을 특징으로 하는, 본 발명의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법을 제공한다.

[0024] 본 발명의 일실시예에 있어서, 상기 목적단백질은 N-데그론(degron) 경로에 관여하는, N-레코그닌(N-recognin); 상기 N-레코그닌에 대한 기질; 또는 N-레코그닌과 기질이 결합된 단백질 복합체일 수 있다.

[0025] 본 발명의 일실시예에 있어서, 상기 N-레코그닌(N-recognin)는 SQSTM1/p62, PRT6, Ubr1 및 NTAQ1으로 이루어진 군 중에서 선택되는 것일 수 있다.

[0026] 본 발명의 일실시예에 있어서, 상기 기질은, PRT6 또는 Ubr1에 대한 기질인 RLGS 펩타이드 단편, NTAQ1에 대한 기질인 M1Q 펩타이드 단편 또는 p62 ZZ 도메인에 대한 기질인 GEEED 펩타이드 단편일 수 있다.

[0027] 또한 본 발명은, (1) 본 발명의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계; (2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계; 및 (3) 생성된 목적단백질을 결정화시키는 단계를 포함하는, 본 발명의 재조합 벡터를 이용한 목적단백질의 결정화 방법을 제공한다.

[0028] 본 발명의 일실시예에 있어서, 상기 목적단백질은 ATG4B 프로테아제 처리에 의해 상기 목적단백질의 N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미노산 잔기일 수 있다.

발명의 효과

[0029] 본 발명에서 고안한 LC3B 벡터 시스템을 사용할 경우, N-데그론 경로에 관여하는 N-recognin, 기질 또는 이들의 복합체를 용이하게 분리 및 정제할 수 있을 뿐만 아니라 분석하고자 하는 목적 단백질의 N 말단에 목적하는 임의의 아미노산 잔기가 형성되도록 디자인할 수 있어, in vitro 시스템을 통해서도 N-데그론 경로에 관여된 단백질들의 활성 및 구조 연구에 기여할 수 있는 효과가 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 LC3B 벡터 시스템에 대한 것으로, 1A는 LC3B 융합 목적단백질을 발현시킬 수 있는 백터 개열지도를 나타낸 것이고, 1B는 본 발명의 벡터가 갖는 멀티 클로닝 사이트, 즉 제한효소 사이트를 나타낸 것으로, 상기 제한효소 사이트를 이용하여 목적 단백질을 코딩하는 유전자 서열을 상기 벡터에 클로닝할 수 있다.

도 2는 본 발명의 LC3B 벡터 시스템을 이용하여 N-데그론 경로에 관여하는 단백질의 정제 과정을 나타낸 것으로, 2A는 본 발명의 LC3B 벡터 시스템의 전반적인 개념을 나타낸 것이고, 2B는 본 발명의 일실시예에서 본 발명의 벡터 시스템을 이용하여 클로닝한 N-degron 경로 관련 단백질들(RLGS-AtPRT6 UBR box His affinity, RLGS-yUbr1 UBR box GST affinity, M1Q-hNTAQ1 C28S GST affinity)의 과발현 및 ATG4B 처리에 따른 절단여부 및 컬럼을 이용한 단백질 정제 결과를 확인한 것이고, 2C는 정제된 단백질들의 oligomeric 상태를 SEC-MALS 로 분석한 결과를 나타낸 것이다. 2D는 GEEED-p62 ZZ도메인에 대한 과발현 및 정제결과를 확인한 것이고, 2E는 GEEED-p62 ZZ도메인의 SEC-MALS 분석 결과를 나타낸 것이다.

도 3은 본 발명의 일실시예에서 분리 및 정제한 융합 단백질들의 결정화 구조를 분석한 결과를 나타낸 것으로, 3A는 RLGS-AtPRT6 UBR box, 3B는 RLGS-yUbr1 UBR box, 3C는 M1Q-hNTAQ1 C28S의 결정 구조에 대한 모식도를 나타낸 것이다.

도 4는 본 발명의 일실시예에서 분리 및 정제한 GEEED-p62 ZZ-도메인의 결정 구조에 대한 모식도를 나타낸 것이다.

도 5는 서로 다른 종(A, Arabidopsis thaliana; h, human; y, yeast) 유래의 N-recognin에 존재하는 UBR 박스의 서열 정렬 결과를 나타낸 것으로, 채워진 원은 아연 이온이 배치된 잔기를 나타낸 것으로, Zn1은 주황색, Zn2는 녹색, Zn3는 하늘색, Zn1 및 Zn2의 경우 검정으로 나타내었고, N-데그론의 결합을 위한 주요 결정 인자는 서열 아래에 자홍색으로 채워진 삼각형으로 표시하였다.

도 6은 본 발명의 LC3B 벡터 시스템의 활용과 관련된 것으로, 6A는 N-degron 기질인 R-BiP의 제조 및 항체를 이용한 R-BiP 검출과정을 나타낸 것으로, MBP-PB1-ZZ domain과 R-BiP의 복합체를 확인한 것을 나타낸 것이고, 6B는 신규로 2형의 N-데그론 기질을 동정하기 위해 ClpS-접합된 친화성 컬럼을 수행한 결과를 나타낸 것으로, 본 발명의 벡터 시스템으로 제조된 GFP에 융합된 LAPPY 서열을 사용할 수 있음을 확인한 것이며, 6C는 N-데그론 서열을 부착하는 효소 활성을 갖는 야생형 인간 NTAQ1 또는 효모 Nta1을 정제하고 3차 불안정화 잔기들(글루타민 또는 아스파라진)을 2차 불안정화 잔기들(글루탐산 또는 아스팔트산)로 완전히 전환시키는 활성이 있음을 확인한 결과를 나타낸 것이다.

도 7은 본 발명의 또 다른 LC3B 벡터 시스템에 대한 것으로, MBP 태그된 벡터를 토대로 LC3B 융합 목적단백질을 발현시킬 수 있는 재조합 백터(pACYC-MBP-LC3B 벡터)의 개열지도를 나타낸 것이다.

도 8은 본 발명의 또 다른 LC3B 벡터 시스템에 대한 것으로, GST 태그된 벡터를 토대로 LC3B 융합 목적단백질을 발현시킬 수 있는 재조합 백터(pGEX-GST-LC3B 벡터)의 개열지도를 나타낸 것이다.

도 9는 본 발명의 발현벡터인 pACYC-MBP-LC3B, pGEX-GST-LC3B 및 pET-His-LC3B 벡터를 각각 이용하여 LC3B 융합 목적단백질을 발현시키고 정제한 것을 확인한 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0031] 본 발명은 LC3B 단백질을 이용하여 N-데그론 경로에 관여하는 단백질들을 in vitro에서 대량생산하기 위한 재조합 벡터를 고안한 것으로, 특히 본 발명의 재조합 벡터로 목적 단백질을 발현시킬 경우, 목적 단백질의 N 말단에 임의의 목적하는 아미노산이 노출되도록 함으로써, N-데그론 경로에 관여하는 단백질들의 구조 및 활성 연구를 위한 단백질의 대량생산을 가능하게 할 수 있도록 하였다.

[0032] 종래 원하는 아미노산을 N 말단에 노출시키는 방법으로는 유비퀴틴과 유비퀴틴 유사 단백질인 SUMO를 태그로 사용하는 방법이 있다. 그러나 유비퀴틴을 태그(tag)로 사용할 경우, 유비퀴틴과 원하는 서열 사이를 절단하기 위해 DUB (Deubiquitylating enzyme)을 사용해야 하는데, 대부분의 DUB는 상당히 크며 발현 양에 문제가 있고, 유비퀴틴의 C-말단 링커의 LC3B보다 짧기 때문에 프로테아제의 활성이 줄어 더 많은 양을 사용해야 한다는 문제점이 있다. 또한 SUMO를 태그로 사용할 경우, SUMO 프로테아제의 활성은 높지만 SUMO 단백질은 불용체(inclusion body)가 형성되어 제대로 폴딩된 단백질을 생성하기 어려운 문제점이 있다. 그러나 본 발명에서는 LC3B를 태그로 사용하여 이러한 유비퀴틴과 SUMO가 갖는 문제점을 해결할 수 있도록 하였다.

[0033] 그러므로 본 발명은, 서열번호 1의 아미노산으로 이루어진 LC3B 단백질을 코딩하는 염기서열; 아미노산 FG를 코딩하는 염기서열; 및 목적단백질을 코딩하는 염기서열이 순차적으로 포함된, 재조합 벡터를 제공함에 특징이 있다.

[0034] 앞서 종래기술에서도 언급한 바와 같이, N-말단 법칙 경로에 대한 연구에 있어서, in vitro에서 N-말단 법칙 경로 관련 단백질들의 생화학적 연구 및 고해상도 구조 연구에 많은 한계를 가지고 있는데, 이는 단백질들의 결합친화도, 효소의 동역학 파라미터 및 고분해능 구조결정을 보다 정확하게 분석하기 위해서는 다량의 단백질이 필요한데, 현재까지 기술로는 N-말단 법칙 경로에 관여하는 단백질을 다량으로 정제 및 수득하기가 어려운 문제점이 있다.

[0035] 뿐만 아니라, 종래 방법으로 in vitro에서 목적 단백질을 발현시키면 발현된 단백질의 N 말단에 위치하는 아미노산은 메티오닌(Met)으로 고정되어 있어 N 말단의 다양한 아미노산을 갖는 단백질에 대한 연구를 하기에는 한계가 있으며, N-말단의 메티오닌(Met)을 절단 또는 분해해야하는 별도의 과정이 필요한 문제가 있었다.

[0036] 한편, 본 발명에서 고안한 재조합 벡터는 종래 이러한 문제점을 해결할 수 있는 방법으로서, N-말단에 임의의 원하는 아미노산 잔기를 갖는 목적 단백질을 in vitro에서 대량 생산할 수 있는 특징이 있다.

[0037] 구체적으로 본 발명에 따른 상기 재조합 벡터는 서열번호 1의 아미노산으로 이루어진 LC3B 단백질을 코딩하는 염기서열; 아미노산 FG를 코딩하는 염기서열; 및 목적단백질을 코딩하는 염기서열을 포함하며, 이들 염기서열이 순차적으로 연결되어 있다.

[0038] 본 발명의 재조합 벡터에 삽입되어 있는 상기 LC3B 단백질을 코딩하는 염기서열의 C-말단은 KpnI 제한효소 사이트가 되도록 변형시킨 서열로서, 서열번호 1의 아미노산 서열로 이루어져 있으며, 서열번호 1의 아미노산을 코딩하는 염기서열은 서열번호 2에 나타내었다.

[0039] 또한 서열번호 1의 LC3B 단백질을 코딩하는 염기서열 다음에는 FG의 2개의 아미노산 서열을 코딩하는 염기서열이 순차적으로 연결되어 있는데, 상기 FG의 G (Glycine)는 ATG4B 프로테아제에 의해 절단될 수 있도록 하였다.

[0040] 상기 FG 2개의 아미노산 서열을 코딩하는 염기서열은 TTC GGG 일 수 있다.

[0041] 또한, 상기 목적단백질은 본 발명의 재조합 벡터를 이용하여 N-데그론(degron) 경로에 관여하는 연구대상의 목적단백질로서, N-레코그닌(N-recognin); 상기 N-레코그닌에 대한 기질; 또는 N-레코그닌과 기질이 결합된 단백질 복합체가 해당될 수 있다.

[0042] 본 발명의 일실시예에서는, 상기 N-레코그닌(N-recognin)으로 SQSTM1/p62, PRT6, Ubr1 및 NTAQ1을 사용하였고, 상기 기질로는 PRT6 또는 Ubr1에 대한 기질인 RLGS 펩타이드 단편, NTAQ1에 대한 기질인 M1Q 펩타이드 단편 또는 p62 ZZ 도메인에 대한 기질인 GEEED 펩타이드 단편을 코딩하는 염기서열을 사용하였다.

[0043] 또한 본 발명의 재조합 벡터는 상기 목적단백질의 용이한 정제 및 분리를 위해, 글루타티온-S-트랜스퍼라제(Glutathione-S-transferase, GST), MBP(Maltose binding protein) 또는 5~7개의 히스티딘(Histidine) 아미노산을 코딩하는 유전자를 N 말단 또는 C 말단에 연결시킬 수 있으며, 본 발명의 일실시예에서는 N 말단에 연결시켰다. 또한, 히스티딘 태그를 사용하는 경우, 다른 두 종류의 태그에 비해 크기가 작아 정제 과정이 용이한 장점이 있고, GST 또는 MBP 유전자의 경우, 히스티딘 태그와 비교했을 때 더 많은 단백질을 발현시키는 효과가 있어 단백질의 사용 목적에 따라 히스티딘, GST 또는 MBP 태그를 적절하게 이용할 수 있다.

[0044] 본 발명에서 "벡터(vector)"는 적합한 숙주 내에서 DNA를 발현시킬 수 있는 적합한 조절 서열에 작동가능하게 연결된 DNA 서열을 함유하는 DNA 제조물을 의미한다. 벡터는 플라스미드, 파지 입자, 또는 간단하게 잠재적 게놈 삽입물일 수 있다. 적당한 숙주로 형질전환 되면, 벡터는 숙주 게놈과 무관하게 복제하고 기능할 수 있거나, 또는 일부 경우에 게놈 그 자체에 통합될 수 있다. 플라스미드가 현재 벡터의 가장 통상적으로 사용되는 형태이므로, 본 발명의 명세서에서 "플라스미드 (plasmid)" 및 "벡터 (vector)"는 때로 상호 교환적으로 사용된다.

[0045] 본 발명은 목적상 플라스미드 벡터를 사용하는 것이 바람직하다. 이러한 목적에 사용될 수 있는 전형적인 플라스미드 벡터는 (a) 숙주세포 당 수백 개의 플라스미드 벡터를 포함하도록 복제가 효율적으로 이루어지도록 하는 복제 개시점, (b) 플라스미드 벡터로 형질전환된 숙주세포가 선발될 수 있도록 하는 항생제 내성 유전자 및 (c) 외래 DNA 절편이 삽입될 수 있는 제한효소 절단부위를 포함하는 구조를 지니고 있다. 적절한 제한효소 절단부위가 존재하지 않을지라도, 통상의 방법에 따른 합성 올리고뉴클레오타이드 어댑터(oligonucleotide adaptor) 또는 링커(linker)를 사용하면 벡터와 외래 DNA를 용이하게 라이게이션(ligation)할 수 있다.

[0046] 본 발명에서 목적 유전자의 과발현을 위하여 사용되는 벡터는 당업계에 공지된 발현 벡터가 사용될 수 있으며, 바람직하게는 pET-His 벡터(Novagen), pACYC-MBP 벡터 또는 pGEX-GST 벡터를 사용할 수 있으나, 이에 한정되는 것은 아니다. 상기 벡터들을 사용하여 클로닝을 수행하면, 발현되는 단백질의 말단에 히스티딘기 tag, GST tag 또는 MBP tag이 결합되어 발현되므로, 상기 단백질을 효과적으로 정제할 수 있다. 클로닝된 유전자로부터 발현된 융합단백질을 분리하기 위해서는 당업계에 공지된 일반적인 방법이 이용될 수 있으며, 구체적으로, 본 발명에서는 Ni-NTA His-결합 레진(Novagen)을 사용하는 크로마토그래피 방법을 이용하여 분리할 수 있다.

[0047] 본 발명의 재조합 벡터는 적절한 숙주세포로 형질전환되어야 한다. 본 발명에 있어서, 선호되는 숙주세포는 원핵세포이다. 적합한 원핵 숙주세포는 E. coli균주 BL21(DE3), E. coli균주 DH5a, E. coli균주 JM101, E. coli K12균주 294, E. coli균주 W3110, E. coli균주 X1776, E. coli XL-1Blue (Stratagene) 및 E. coli B 등을 포함한다. 그러나 FMB101, NM522, NM538 및 NM539와 같은 E. coli균주 및 다른 원핵생물의 종(speices) 및 속(genera)등이 또한 사용될 수 있다.

[0048] 상기 재조합 벡터의 숙주세포로의 형질전환은 당업계에 공지된 칼슘 클로라이드 방법을 사용해서 용이하게 달성

될 수 있다. 선택적으로, 전기천공법(electroporation)(Neumann et al., EMBO J., 1: 841(1982))이 또한 이러한 세포들을 형질전환하는데 사용될 수 있다.

[0049] 또한 "발현 조절 서열 (expression control sequence)"이라는 표현은 특정한 숙주 생물에서 작동가능하게 연결된 코딩 서열의 발현에 필수적인 DNA 서열을 의미한다. 그러한 조절 서열은 전사를 실시하기 위한 프로모터, 그러한 전사를 조절하기 위한 임의의 오퍼레이터 서열, 적합한 mRNA 리보좀 결합 부위를 코딩하는 서열 및 전사 및 해독의 종결을 조절하는 서열을 포함한다. 예를 들면, 원핵생물에 적합한 조절 서열은 프로모터, 임의로 오퍼레이터 서열 및 리보좀 결합 부위를 포함한다. 진핵세포는 프로모터, 폴리아데닐화 시그날 및 인핸서가 이에 포함된다. 플라스미드에서 유전자의 발현 양에 가장 영향을 미치는 인자는 프로모터이다. 고 발현용의 프로모터로서 SRα α 프로모터와 사이토메가로바이러스(cytomegalovirus) 유래 프로모터 등이 바람직하게 사용된다.

[0050] 본 발명의 DNA 서열을 발현시키기 위하여, 매우 다양한 발현 조절 서열중 어느 것이라도 벡터에 사용될 수 있다. 유용한 발현 조절서열의 예에는, 예를 들어, SV40 또는 아데노바이러스의 초기 및 후기 프로모터들, lac 시스템, trp 시스템, TAC 또는 TRC 시스템, T3 및 T7 프로모터들, 파지 람다의 주요 오퍼레이터 및 프로모터 영역, fd 코드 단백질의 조절 영역, 3-포스포글리세레이트 키나제 또는 다른 글리콜분해 효소에 대한 프로모터, 상기 포스파타제의 프로모터들, 예를 들어 Pho5, 효모 알파-교배 시스템의 프로모터 및 원핵세포 또는 진핵 세포 또는 이들의 바이러스의 유전자의 발현을 조절하는 것으로 알려진 구성과 유도의 기타 다른 서열 및 이들의 여러 조합이 포함된다. T7 RNA 폴리메라아제 프로모터 Φ10은 E. coli에서 단백질 NSP를 발현시키는데 유용하게 사용될 수 있다.

[0051] 또한 재조합 벡터에 포함되는 염기서열, 즉 핵산은 다른 핵산 서열과 기능적 관계로 배치될 때 "작동가능하게 연결(operably linked)" 되어야 한다. 이것은 적절한 분자 (예를 들면, 전사 활성화 단백질)가 조절 서열(들)에 결합될 때 유전자 발현을 가능하게 하는 방식으로 연결된 유전자 및 조절 서열(들)일 수 있다. 예를 들면, 전서열(pre-sequence) 또는 분비 리더(leader)에 대한 DNA는 폴리펩타이드의 분비에 참여하는 전단백질로서 발현되는 경우 폴리펩타이드에 대한 DNA에 작동가능하게 연결되고 프로모터 또는 인핸서(enhancer)는 서열의 전사에 영향을 끼치는 경우 코딩서열에 작동가능하게 연결되거나 또는 리보좀 결합 부위는 서열의 전사에 영향을 끼치는 경우 코딩 서열에 작동가능하게 연결되거나 또는 리보좀 결합 부위는 번역을 용이하게 하도록 배치되는 경우 코딩 서열에 작동가능하게 연결된다. 일반적으로, "작동가능하게 연결된"은 연결된 DNA 서열이 접촉하고, 또한 분비 리더의 경우 접촉하고 리딩 프레임 내에 존재하는 것을 의미한다. 그러나, 인핸서는 접촉할 필요가 없다. 이들 서열의 연결은 편리한 제한 효소 부위에서 라이게이션(연결)에 의해 수행된다. 그러한 부위가 존재하지 않는 경우, 통상의 방법에 따른 합성 올리고뉴클레오티드 어댑터(oligonucleotide adaptor) 또는 링커(linker)를 사용한다.

[0052] 나아가 본 발명은 본 발명의 재조합 벡터로 형질전환된 형질전환체를 제공할 수 있으며, 상기 형질전환체는 상기 기술된 숙주 세포를 의미한다.

[0053] 또한 본 발명은 본 발명의 재조합 벡터를 이용하여 목적단백질을 제조하는 방법을 제공할 수 있는데, 상기 방법은 바람직하게, (1) 본 발명의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계; (2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계 ; 및 (3) 생성된 목적단백질을 분리 및 정제하는 단계를 포함하며, 상기 목적단백질은 ATG4B 프로테아제 처리에 의해 상기 목적단백질의 N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미노산 잔기로 노출된다.

[0054] 따라서 본 발명의 재조합 벡터로 목적 단백질을 수득할 경우, N-말단의 아미노산 잔기가 프롤린(proline)을 제외한 19개의 아미노산 중 어느 하나의 아미노산 잔기를 갖는 목적 단백질을 대량 생산할 수 있다.

[0055] 또한 본 발명은 본 발명의 재조합 벡터를 이용한 목적단백질의 결정화 방법을 제공할 수 있는데, 상기 방법은 (1) 본 발명의 재조합 벡터로 형질전환된 재조합 미생물을 배양하여 LC3B와 목적단백질의 융합단백질을 과발현시키는 단계; (2) ATG4B 프로테아제를 처리하여 LC3B와 목적단백질을 절단시키는 단계; 및 (3) 생성된 목적단백질을 결정화시키는 단계를 포함할 수 있다.

[0056] 본 발명자들은 LC3B 유전자를 이용한 본 발명의 재조합 벡터를 이용하여 목적 단백질로서, RLGS-AtPRT6 UBR box, RLGS-yUbr1 UBR box, M1Q-hNTAQ1 C28S 및 GEEED-p62 ZZ 도메인을 각각 대장균에서 발현시킨 후, 이들 목적 단백질을 각각 정제한 결과, 목적 단백질의 활성 및 구조 분석을 위한 충분한 양의 단백질을 생산할 수 있었고, 결정화시킨 단백질들의 구조분석을 통해 단백질 또는 융합단백질들의 결합관계, 활성 분석이 가능하다는 것

을 하기 실시예들의 실험을 통해 확인할 수 있었다.

[0057] 본 발명의 일실시예에서 본 발명의 재조합 벡터에 삽입된 목적 단백질인 상기 RLGS-AtPRT6 UBR box은 바람직하게 서열번호 3의 아미노산 서열로 이루어진 것일 수 있고, 서열번호 3의 아미노산 서열을 코딩하는 염기서열은 서열번호 4에 나타내었다.

[0058] 또한, 본 발명의 일실시예에서 본 발명의 재조합 벡터에 삽입된 목적 단백질인 상기 RLGS-yUbr1 UBR box은 바람직하게 서열번호 5의 아미노산 서열로 이루어진 것일 수 있고, 서열번호 5의 아미노산 서열을 코딩하는 염기서열은 서열번호 6에 나타내었다.

[0059] 또한, 본 발명의 일실시예에서 본 발명의 재조합 벡터에 삽입된 목적 단백질인 상기 M1Q-hNTAQ1 C28S는 바람직하게 서열번호 7의 아미노산 서열로 이루어진 것일 수 있고, 서열번호 7의 아미노산 서열을 코딩하는 염기서열은 서열번호 8에 나타내었다.

[0060] 또한, 본 발명의 일실시예에서 본 발명의 재조합 벡터에 삽입된 목적 단백질인 상기 GEEED-p62 ZZ 도메인은 바람직하게 서열번호 9의 아미노산 서열로 이루어진 것일 수 있고, 서열번호 9의 아미노산 서열을 코딩하는 염기서열은 서열번호 10에 나타내었다.

[0061] 따라서 본 발명에서 제공하는 재조합 벡터는 N-데그론 경로에 관여하는 N-recognin, 기질 또는 이들의 복합체를 용이하게 분리 및 정제할 수 있을 뿐만 아니라 분석하고자 하는 목적 단백질의 N 말단에 목적하는 임의의 아미노산 잔기가 형성되도록 디자인할 수 있어, in vitro 시스템을 통해서도 N-데그론 경로에 관여된 단백질들의 활성 및 구조 연구를 정확하고 용이하게 수행할 수 있다.

[0063] 이하, 본 발명을 실시예에 의해 상세히 설명하기로 한다. 그러나 이들 실시예는 본 발명을 보다 구체적으로 설명하기 위한 것으로서, 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[0065] **<실험방법>**

[0066] ① LC3B 융합 단백질 제조를 위한 재조합 DNA 구조물의 제조

[0067] LC3B(1-120 아미노산) 및 N- 데그론에 관여하는 단백질이 융합된 융합단백질의 서열은 gBlock 유전자 절편(Integrated DNA Technologies)을 토대로 PCR을 수행하여 증폭시켰다. 이후 N-말단 헥사-히스티딘이 태그된 단백질 구조물의 제조를 위해 상기 증폭된 산물은 변형된 pET 벡터(tobacco etch virus 및 thrombin cleavage sites를 포함)의 BamHI 및 XhoI의 제한효소 부위로 클로닝하였고, 클로닝된 플라스미드를 이용하여 대장균 BL21(DE3) 세포를 형질전환시켰다.

[0068] 또한, LC3B 서열을 포함하는 표적 단백질의 용이한 클로닝을 위해 다수의 클로닝 사이트를 이용하여 LC3B-융합단백질을 발현하는 발현 벡터(pET-His-LC3B 벡터)를 제조하였다(도 1a 참조).

[0069] 또한, LC3B 서열을 포함하는 표적 단백질의 용이한 클로닝을 위해 다수의 클로닝 사이트를 이용하여 LC3B-융합단백질을 발현하는 발현 벡터로서 도 1a 의 pET-His-LC3B 벡터 이외에 다른 태그(MBP, GST)가 도입된 재조합 발현벡터인 pACYC-MBP-LC3B 벡터(도 7a) 및 pGEX-GST-LC3B 벡터(도 8a)도 각각 제조하였다.

[0071] ② 단백질의 정제

[0072] 히스티딘이 태그된 융합단백질, 즉 결정화를 위한 융합단백질인 LC3B-RLGS-AtPRT6 UBR box, LC3B-M1Q-hNTAQ1 및 LC3B-GEEED-ZZ 단백질과, ClpS 컬럼을 위한 E. coli ClpS 및 LC3B-LAPPS-GFP의 발현은 상기 실험방법 ①에서 제조한 각 융합단백질의 발현벡터로 형질 전환된 대장균의 세포 배양액에 최종 농도가 0.5 mM이 되도록 IPTG(isopropyl-β-D-thiogalactoside)를 처리하고 20시간 동안 18℃에서 배양하여 융합 단백질의 발현을 유도하였다. 이후 세포를 원심분리로 수집하였고 200 mM NaCl이 함유된 버퍼 A((50 mM Tris-HCl pH 8.0 and 0.5 mM tris[2-carboxyethyl]phosphine [TCEP])로 혼탁시킨 후, 초음파 처리를 수행하여 세포를 파쇄하였고, 불용성 물질들은 27,216g의 속도로 1시간 동안 원심분리하여 제거하였다. 이후 상등액을 HisTrapTM 컬럼(GE Healthcare, 17-5255-01)에 로딩하였고, 100 mM NaCl 및 500mM 이미다졸을 함유한 버퍼 A의 농도 구배를 이용하여 단백질을 용출시켰다. 용출된 시료들은 다시 HiTrapTM Q HP 컬럼(GE Healthcare, 17-1154-01)을 이용하여

정제하였다. 이후 human ATB4B 프로테아제로 25℃에서 밤새도록 반응시켜 His-LC3B-tag를 제거하였다. 그런 뒤, 정제한 단백질들은 한외여과(ultrafiltration; Amicon Ultra 30 K NMWL, Millipore)를 이용하여 농축한 다음, HiLoadTM 16/600 SuperdexTM 75pg (GE Healthcare, 28-9893-33) 컬럼에 로딩하였고 버퍼 B(20 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM TCEP)로 평형화 시킨 후, 단백질을 정제하였다.

[0073] 또한, GST 태그된 융합 단백질(결정화를 위한 LC3B-RLGS-yUbr1 UBR box 및 항체 인지 확인을 위한 LC3B-R-BiP 융합단백질)의 발현은 상기 융합단백질의 발현벡터로 형질전환된 세포 배양액에 최종 농도가 0.5 mM이 되도록 IPTG (isopropyl-β-D-thiogalactoside)를 처리하고 25℃에서 20시간 동안 단백질 발현을 유도시켰다. 이후 원심분리로 세포들을 수집하였고 1X PBS 용액으로 세포들을 현택시킨 다음, 초음파로 세포를 파쇄하였다. 이후 불용성 물질들은 27,216g의 속도로 1시간 동안 원심분리하여 제거하였다. 상등액을 취하여 GST 컬럼(GE Healthcare)에 로딩하였고, 50 mM Tris-HCl pH 8.0, 1 mM TCEP, 100 mM NaCl 및 25 mM glutathione이 함유된 GST 용출 버퍼로 단백질을 용출시켰다. 용출된 시료들은 HiTrapTM Q HP 컬럼 (GE Healthcare, 17-1154-01)을 이용하여 정제하였고 human ATB4B 프로테아제로 25℃에서 밤새도록 반응시켜 GST-LC3B-tag를 제거하였다. 정제된 단백질들은 한외여과(Amicon Ultra 3 K NMWL, Millipore)한 후, HiLoadTM 16/600 SuperdexTM 75pg 컬럼(GE Healthcare, 28-9893-33)에 로딩한 다음, 20 mM Tris-HCl pH 8.0, 200 mM NaCl 및 1 mM TCEP을 함유한 버퍼로 평형화한 후, 단백질을 정제하였다.

[0075] ③ 결정화 및 데이터 수집

[0076] 정제된 RLGS-AtPRT6 UBR box 단백질은 10 mg/ml의 농도로 농축하였고 22℃에서 시팅 드롭 증기 확산법(sitting drop vapor diffusion method)을 이용하여 결정화하였으며, 이후 결정화된 단백질과 동일 부피의 충진 용매(reservoir solution: 0.2 M sodium formate pH 7.0 (Hampton Research, HR2-765), 18.5-22.5% (w/v) PEG 3350 (Hampton Research, HR2-527)를 함께 혼합하였다.

[0077] 또한, 정제된 RLGS-yUbr1 UBR box 단백질은 9 mg/ml의 농도로 농축하였고 22℃에서 시팅 드롭 증기 확산법(sitting drop vapor diffusion method)을 이용하여 결정화하였으며, 이후 결정화된 단백질과 동일 부피의 충진 용매(0.1 M BIS-Tris pH 6.5 및 2.7-3.0 M NaCl 함유)를 함께 혼합하였다.

[0078] 또한 정제된 M1Q-hNTAQ1 C28S 단백질은 16 mg/ml의 농도로 농축하였고 22℃에서 시팅 드롭 증기 확산법(sitting drop vapor diffusion method)을 이용하여 결정화하였으며, 이후 결정화된 단백질과 동일 부피의 충진 용매(0.4 M ammonium sulfate(Hampton Research, HR2-541), 0.1 M sodium citrate tribasic dihydrate pH 5.6(Hampton Research, HR2-735) 및 1.0 M lithium sulfate monohydrate (Hampton Research, HR2-545) 함유)를 함께 혼합하였다.

[0079] 또한 정제된 GEEED-p62 ZZ 융합 단백질은 농축 후 22℃에서 시팅 드롭 증기 확산법(sitting drop vapor diffusion method)을 이용하여 결정화하였으며, 이후 결정화된 단백질과 동일 부피의 충진 용매(0.17 M ammonium sulfate, 0.085 M sodium cacodylate trihydrate pH 6.5 및 22-30% (w/v) polyethylene glycol 8000 함유)를 함께 혼합하였다.

[0080] 또한 단백질 결정체들은 동결 방지제로서 20-30% (v/v) 글리세롤을 사용하여 액체 질소에서 급속 동결시켰다.

[0081] 단백질 결정화의 데이터 수집은 일본의 Photon Factory와 Spring-8 및 한국의 PAL (Pohang Accelerator Laboratory)에서 빔라인을 사용하여 기본 단백질 결정 데이터를 수집하였고, 회절 데이터는 HKL2000 소프트웨어를 사용하여 색인화, 통합화 및 등급화 하였다. 데이터 수집 통계결과는 하기 표 1에 기재한 바와 같다.

표 1

	RLGS-AtPRT6 UBR box (119-189)	RLGS-yUbr1 UBR box (113-194)	M1Q-hNTAQ1 C28S (FL, 1-205)
Data collection			
Beamline	PAL-5C	PF-17A	PAL-5C
Space group	$C\,2\,2\,2_1$	$P\,4_3\,2_1\,2$	$I\,2\,2\,2$
Cell dimensions			
a, b, c (Å)	74.2, 104.2, 87.7	60.6, 60.6, 69.8	95.6, 105.8, 140.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.28294	0.90000	0.97957
Resolution (Å)	50.00 - 2.80 (2.85 - 2.80)[a]	50.00 - 1.04 (1.06 - 1.04)[a]	50.00 - 1.80 (1.84 - 1.80)[a]
R_{merge}	0.126 (0.834)[a]	0.064 (0.770)[a]	0.080 (0.542)[a]
I/σI	36.1 (3.91)[a]	43.9 (2.1)[a]	25.9 (2.56)[a]
Completeness (%)	99.9 (100.0)[a]	100.0 (99.9)[a]	100.0 (100.0)[a]
Redundancy	6.9 (7.2)[a]	12.5 (11.3)[a]	6.7 (6.4)[a]
Zn-Phasing			
No. of Zn atoms	9		
Initial figure of merit (FOM)	0.46		
FOM after density modification	0.87		
Refinement			
Resolution (Å)	35.5 - 2.79 (2.89 - 2.79)[a]	30.25 - 1.04 (1.077 - 1.04)[a]	35.47 - 1.80 (1.87 - 1.80)[a]
No. reflections	8,725 (836)[a]	62,978 (6,199)[a]	65,202 (6,176)[a]
R_{work} / R_{free}	0.215 / 0.281	0.189 / 0.202	0.216 / 0.218
No. atoms	1,638	704	3,426
Protein	1,629	669	3,270
Ligand	9	3	
Water		32	156
B-factors (Å^2)	60.12	16.31	26.70
Protein	60.04	16.25	26.68
Ligand	74.41	10.26	
Water		18.11	27.02
R.m.s deviations			
Bond lengths (Å)	0.014	0.005	0.031
Bond angles (°)	1.59	0.77	1.87
Ramachandran (%)			
Favor	91.94	97.62	97.25
Allow	8.06	2.38	2.75
Outliers	0.00	0.00	0.00
PDB ID	6KGG	6KGI	6KGJ

[a] Highest resolution shell

[0082]

[0084] ④ 구조결정

[0085] AtPRT6 UBR box 단백질을 제외한 다른 단백질들은 검색모델로서 p62 ZZ domain(PDB ID: 5YP7), yUbr1 UBR box(PDB ID: 3NIT) 및 Ntaq1 structures(PDB ID: 4W79)을 이용하여 MR로 분석하였다. MR 계산에 의해 얻어진 모델은 Coot를 사용하여 반복적인 사이클로 재구성하였다. AtPRT6 UBR box의 초기 단계는 PAL 설비에서 빔라인 5C로 아연 원자의 흡수 에지(λ = 1.282282A)에서 수집된 2.7-A 해상도 SAD 데이터 세트로 결정되었다. Zn-사이트의 결정, 위상 및 자동 모델 구축은 Phenix 소프트웨어 패키지로 구현된 SAD 위상 모듈을 사용하여 수행하였다. 초기 모델은 Coot를 사용하여 반복적인 주기로 재구성 및 정제하였고, 라마반드란 값(Ramachandran values)은 Molprobity 프로그램을 사용하여 계산하였다. 수집된 데이터 및 정제된 값의 통계는 상기 표 1에 기재되어 있으며, 모든 구조적 형상은 PyMOL (http://www.pymol.org/)을 이용하여 나타내었다.

[0087] ⑤ 다중-앵글 빛 산란과 결합된 크기 배제 크로마토그래피

[0088] SEC-MALS 실험은 Wyatt MiniDAWN TREOS MALS 기기 및 Wyatt Optilab rEX 시차 굴절계에 연결된 고속 단백질 액체 크로마토그래피 시스템(GE Healthcare)을 사용하여 수행하였고, 사용된 완충액은 최종 단백질 정제 단계에서 사용된 완충액과 동일한 용액을 사용하였다. 정제된 단백질은 SuperdexTM 75 Increase 10/300 GL gel filtration 컬럼(GE Healthcare, 29-1487-21)에 로딩하였다. 검출기 정규화를위한 등방성 산란제(isotropic scatterer)로 난백알부민을 사용하였다. 빛 산란은 ASTRA V software (Wyatt)를 이용하여 측정 및 분석하였다.

[0090] ⑥ R-BiP 항체를 이용한 검출

[0091] 정제된 R-BiP 단백질은 2×SDS-PAGE 로딩버퍼(100 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol 및 200 mM DTT)와 혼합하여 웨스턴 블럿을 수행하였다. 이때 토끼 폴리클로날 anti-R-BiP 항체 (Abfrontier, AR05-PA0001, 1:1000)를 이용하여 아르기닐화된 BiP(arginylated BiP)를 검출하였다.

[0093] ⑦ ClpS-접합된 컬럼분석

[0094] 버퍼 교환을 위해, 정제된 E. coli ClpS 단백질은 Superdex 75 16/600 pg 컬럼(GE Healthcare)에 로딩하였고, 20 mM HEPES pH 7.4, 50 mM NaCl, 1 mM DTT 및 10% (v/v) glycerol을 함유한 용액으로 평형화하였다. ClpS 단백질은 Affi-Gel 15 (Biorad)와의 커플화를 위해 4℃에서 4시간 동안 반응시켰다. Affi-Gel 15 컬럼에 ClpS 단백질의 커플링을 확인하기 위하여 정제된 LAPPS-GFP 단백질을 친화성 컬럼에 로딩한 후 녹색으로 변하는지를 분석하였다. 수회 세척한 후, 결합된 GFP를 6M 구아니딘 -HCl로 용리하여 컬럼에서 모든 녹색을 제거하였다.

[0096] 참고로, 원자 좌표 및 구조적 요소들은 PDB ID 코드 6KGG(RLGS-AtPRT6 UBR box), 6KGI (RLGS-yUbr1 UBR box), 6KGJ (M1Q-hNTAQ1 C28S) 및 6KHZ (GEEED- p62 ZZ)로 Protein Data Bank에 저장되어 있다.

[0098] **<실시예 1>**

[0099] **다양한 N-말단 잔기를 갖는 목적 단백질의 제조**

[0100] <1-1> 본 발명에 따른 LC3B 재조합 벡터를 이용한 다양한 N-말단 잔기를 갖는 목적단백질의 정제

[0101] 본 발명자들은 상기 실험방법에서 제조한 본 발명의 벡터 시스템을 이용하여 인비트로(in vitro)에서 메티오닌이 아닌 목적하는 다양한 아미노산 잔기를 N-말단에 갖는 N-데그론 경로에 관여하는 목적 단백질을 용이하게 분리 및 정제할 수 있는지를 분석하였다. 이를 위해 LC3B와 목적 단백질의 융합 단백질은 도 1A, 도 7A 및 도 8A의 벡터 시스템을 통해 발현될 수 있도록 재조합 벡터를 각각 제조하였으며, 이때 LC3B와 목적 단백질 사이에는 ATG4B 프로테아제에 의해 절단될 수 있는 부위로서 글라이신(Gly)이 오도록 설계하였다. 즉, ATG4B 프로테아제는 LC3B의 C-말단의 Gly을 인식하여 이 부분을 절단함으로써 LC3B가 제거된 목적 단백질만을 용이하게 분리할 수 있도록 하였다. 또한 목적 단백질의 용이한 정제를 위해 재조합 벡터에는 hexa-histidine 잔기 또는 GST 테그를 포함하도록 하였다.

[0102] 상기와 같은 본 발명에서 고안한 재조합 벡터맵은 도 1A, 도 7A 및 도 8A에 각각 나타내었고, 각 벡터맵에 존재하는 다양한 제한효소 사이트를 이용하여 목적 단백질을 암호화하는 유전자들이 삽입되도록 클로닝하였으며, 118개의 아미노산으로 구성된 LC3B 서열 다음에 ATG4B 프로테아제의 절단 인식부위인 FG 아미노산이 위치되도록 하였고, 그 뒤로 목적 단백질이 오도록 위치시키되, 목적 단백질의 N-말단에는 프롤린(proline)을 제외한 임의의 19개 중 어느 하나의 목적하는 아미노산을 갖는 목적 단백질이 오도록 해당 유전자를 클로닝하였다.

[0103] 구체적으로, 본 발명자들은 상기 목적 단백질로서, N 말단잔기로 Gly을 갖는 p62/SQSTM1, N 말단잔기로 Arg를 갖는 PRT6, N 말단잔기로 Arg를 갖는 Ubr1 및 N 말단잔기로 Gln을 갖는 NTAQ1이 되도록 상기 벡터에 클로닝하였다.

[0104] 또한, 상기 Ubr1은 E3 Ub ligase로서 yeast에서 최초로 발견되었고 N-데그론 경로 연구에서 많이 연구되고 있으며, UBR 박스를 통해 1형 N-데그론을 인식하는 것으로 알려져 있다. 상기 PRT6(Ub protein E3 ligase proteolysis 6) 식물에서 Arg/N-데그론 경로에 관여하는 것으로 알려져 있고 UBR 박스를 통해 베이직 N-말단 잔기를 인식하는 것으로 알려져 있다. 상기 NTAQ1(human N-terminal glutamine amidohydrolase)은 아미노기의 제거를 통해 N-말단의 글루타민(glutamine)을 글루탐산(glutamate)으로 전환하고 Arg/N-데그론 경로의 첫 번째 단계의 촉매작용을 하는 것으로 알려져 있다. 이와 같이 본 발명자들은 본 발명의 LC3B를 이용한 재조합 벡터 시스템을 이용하여 하기 표 1의 단백질들을 분리 및 정제하였고, 각 목적 단백질의 N 말단에 다양한 아미노산 잔기가 오도록 하였고, 하기 표에서 박스로 표기한 것이 본 발명의 실시예에서 실제 목적 단백질로 사용한

것이다.

표 2

Protein	N-terminal residue	Origin	Space group	PDB ID
SQSTM1/p62	Arg	*Homo sapiens*	$P\,2_1\,2_1\,2_1$	5YP8
SQSTM1/p62	Lys	*Homo sapiens*	$P\,2_1\,2_1\,2_1$	5YPA
SQSTM1/p62	His	*Homo sapiens*	$I\,2\,3$	5YPB
SQSTM1/p62	Tyr	*Homo sapiens*	$I\,2\,3$	5YPE
SQSTM1/p62	Trp	*Homo sapiens*	$I\,2\,3$	5YPF
SQSTM1/p62	Phe	*Homo sapiens*	$C\,1\,2\,1$	5YPC
SQSTM1/p62	Leu	*Homo sapiens*	$P\,2_1\,2_1\,2_1$	5YPG
SQSTM1/p62	Ile	*Homo sapiens*	$P\,1\,2_1\,1$	5YPH
SQSTM1/p62	Gly	*Homo sapiens*	$I\,2\,3$	6KHZ
PRT6	Arg	*Arabidopsis thaliana*	$C\,2\,2\,2_1$	6KGG
Ubr1	Gly	*Saccharomyces cerevisiae*	$P\,3_2$	3NIS
Ubr1	Arg	*Saccharomyces cerevisiae*	$P\,4_3\,2_1\,2$	6KGI
UBR1	Gly	*Homo sapiens*	$P\,1\,2_1\,1$	3NY1
ClpS	Leu	*Escherichia coli*	$C\,1\,2\,1$	3O2O
NTAQ1	Ser	*Homo sapiens*	$P\,2_1\,2_1\,2_1$	4W79
NTAQ1	Gln	*Homo sapiens*	$I\,2\,2\,2$	6KGJ

[0105]

[0106] 또한, p62의 기질에 해당하는 BiP/GRP78(Arg-Glu-Glu-Glu-Asp---)에 대하여, 본 발명자들은 p62의 ZZ-도메인의 N 말단에 p62의 기질에 해당하는 Xaa-Glu-Glu-Glu-Asp (상기 Xaa는 다양한 아미노산 잔기가 올 수 있음) 펩타이드가 오도록 하였다. 또한, Arabidopsis thaliana PRT6(AtPRT6) 및 Saccharomyces cerevisiae Ubr1 (yUbr1)의 UBR 박스는 N- 말단에 Arg-Gly-Leu-Ser 서열을 포함하도록 하였고, Homo sapiens NTAQ1 (hNTAQ1)의 첫 번째 메티오닌이 글루타민으로 변환된 단백질을 수득할 수 있도록 제조하였다.

[0107] 재조합 벡터의 제조가 완료된 후, E. coli에서 발현시킨 다음, 컬럼 크로마토그래피(SEC-MALS; size-exclusion chromatography column coupled with a multi-angle light scattering device) 및 Coomassie blue 염색을 통해 목적 단백질의 발현 여부 및 분리를 수행하였다.

[0109] 그 결과, 도 2에 나타낸 바와 같이, 상기 본 발명의 벡터 시스템(pET-His-LC3B 벡터)을 이용할 경우, 대장균에서 목적 단백질들이 LC3B와 융합단백질의 형태로 잘 발현되는 것으로 나타났고, ATG4B 프로테아제 처리 시, LC3B의 C-말단의 글라이신(Gly)을 인식하여 절단 현상이 일어남을 확인하였으며, 특히 N 말단에 다양한 아미노산 잔기를 갖는 목적단백질들이 기질과 결합된 상태의 복합체 형태로 정제된 것을 확인할 수 있었다. 또한, 다른 벡터인 pACYC-MBP-LC3B 벡터 및 pGEX-GST-LC3B 벡터를 사용한 경우에도 목적 단백질이 LC3B와 융합단백질의 형태로 잘 발현된다는 것을 확인할 수 있었다(도 9 참조).

[0111] <1-2> 본 발명의 방법으로 정제된 목적단백질의 N 말단 서열분석

[0112] 상기 <1-1>에서 분리 및 정제한 목적 단백질의 N 말단 서열을 시퀀싱을 통해 분석하였다.

표 3

N-terminal 시퀀싱 결과

Protein	Deduced sequence from DNA	Observed sequence
RLGS-AtPRT6 UBR box (120-188)	Arg – Leu – Gly – Ser –	Arg – Leu – Gly – Ser –
RLGS-yUbr1 UBR box (113-194)	Arg – Leu – Gly – Ser –	Arg – Leu – Gly – Ser –
M1Q-hNTAQ1 C28S (FL, 1-205)	Gln – Glu – Gly – Asn –	Gln (Glu)* – Glu – Gly –
GEEED-p62 ZZ (126-172)	Gly – Glu – Glu – Glu –	Gly – Glu – Glu – Glu –
R-BiP N406 (19-406)	Arg – Glu – Glu – Glu –	Arg – Glu – Glu – Glu –
M1Q-hNTAQ1 (FL, 1-205)	Gln – Glu – Gly – Asn –	Glu – Glu – Gly – Asn –
NAA-yNta1 (FL, 1-457)	Asn – Ala – Ala – Met –	Asp – Ala – Ala – Met –

*Reaction product sequence (Glu--) was also detected partially because M1Q-hNTAQ1 C28S mutant has residual enzymatic activity.

[0113]

[0115] 그 결과, 상기 표 3에 기재된 바와 같이, 해당 목적 단백질들의 N 말단 서열이 메티오닌(Met)이 아닌 다른 아미노산 잔기를 갖는 것을 확인할 수 있었고, 실험을 통해 분리한 단백질의 서열분석 결과가 컴퓨터 상에서 예측된 서열(observed sequence)과 일치하는 것으로 나타나, 본 발명의 재조합 벡터 시스템이 잘 가동된다는 것을 알 수 있었다.

[0116] 따라서 이러한 결과를 통해 본 발명자들은 N-데그론(degron) 경로에 관여하는 단백질들의 연구를 in vitro 시스템을 통해서도 진행할 수 있는 기반을 마련하였고, 특히 대장균과 같은 in vitro 시스템에서 목적 단백질을 발현시킬 경우, 단백질의 N 말단이 메티오닌(Met)으로 고정화될 수밖에 없었던 기존의 문제점을 극복할 수 있다는 것을 확인하였다.

[0117] 또한 본 발명의 LC3B 재조합 벡터 시스템은, 분석을 위한 목적 단백질로서 N-recognin 단독을 사용할 수 있거나, 또는 N-recognin의 타겟 기질 단독을 사용할 수 있고, 또는 N-recognin 및 이의 기질이 결합된 융합 단백질을 사용할 수 있다.

[0119] **<실시예 2>**

[0120] **결정화 분석 결과**

[0121] 실시예 1의 본 발명의 LC3B 재조합 벡터 시스템을 통해 분리 및 정제한 목적 단백질들의 결정화 분석을 수행하였다. 특히 본 발명에서는 N-데그론 경로에 관여하는 N-recognin과 이의 기질들이 결합된 단백질 복합체의 형성 여부 및 단백질 복합체의 결정화 분석이 가능한지를 확인하는 과정을 수행하였다. 본 발명의 실험으로 분리한 단백질 복합체인 RLGS-AtPRT6 UBR box, RLGS-yUbr1 UBR box, M1Q-hNTAQ1 및 GEEED-p62 ZZ들의 결정화 구조 분석은 앞서 기술한 결정화 분석 방법으로 수행하였고, 그 결과를 하기 표 4~5 및 도 3~4에 나타내었다.

표 4

	RLGS-AtPRT6 UBR box (119-189)	RLGS-yUbr1 UBR box (113-194)	M1Q-hNTAQ1 C28S (FL, 1-205)
Data collection			
Beamline	PAL-5C	PF-17A	PAL-5C
Space group	$C\,2\,2\,2_1$	$P\,4_3\,2_1\,2$	$I\,2\,2\,2$
Cell dimensions			
a, b, c (Å)	74.2, 104.2, 87.7	60.6, 60.6, 69.8	95.6, 105.8, 140.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.28294	0.90000	0.97957
Resolution (Å)	50.00 - 2.80 (2.85 - 2.80)[a]	50.00 - 1.04 (1.06 - 1.04)[a]	50.00 - 1.80 (1.84 - 1.80)[a]
R_{merge}	0.126 (0.834)[a]	0.064 (0.770)[a]	0.080 (0.542)[a]
I/σI	36.1 (3.91)[a]	43.9 (2.1)[a]	25.9 (2.56)[a]
Completeness (%)	99.9 (100.0)[a]	100.0 (99.9)[a]	100.0 (100.0)[a]
Redundancy	6.9 (7.2)[a]	12.5 (11.3)[a]	6.7 (6.4)[a]
Zn-Phasing			
No. of Zn atoms	9		
Initial figure of merit (FOM)	0.46		
FOM after density modification	0.87		
Refinement			
Resolution (Å)	35.5 - 2.79 (2.89 - 2.79)[a]	30.25 - 1.04 (1.077 - 1.04)[a]	35.47 - 1.80 (1.87 - 1.80)[a]
No. reflections	8,725 (836)[a]	62,978 (6,199)[a]	65,202 (6,176)[a]
R_{work} / R_{free}	0.215 / 0.281	0.189 / 0.202	0.216 / 0.218
No. atoms	1,638	704	3,426
Protein	1,629	669	3,270
Ligand	9	3	
Water		32	156
B-factors (Å^2)	60.12	16.31	26.70
Protein	60.04	16.25	26.68
Ligand	74.41	10.26	
Water		18.11	27.02
R.m.s deviations			
Bond lengths (Å)	0.014	0.005	0.031
Bond angles (°)	1.59	0.77	1.87
Ramachandran (%)			
Favor	91.94	97.62	97.25
Allow	8.06	2.38	2.75
Outliers	0.00	0.00	0.00
PDB ID	6KGG	6KGI	6KGJ

[a] Highest resolution shell

[0122]

표 5

	GEEED-p62 ZZ (126-172)
Data collection	
X-ray source	Home (CuKα, λ = 1.54178 Å)
Space group	*I* 2 3
Cell parameters [a, b, c (Å) / α, β, γ (°)]	114.0, 114.0, 114.0 / 90, 90, 90
Resolution (Å)	50.00 - 2.80 (2.85 - 2.80)[a]
R_{merge}	0.130 (0.860)[a]
I/σI	19.8 (2.79)[a]
Completeness (%)	99.9 (100.0)[a]
Redundancy	6.7 (6.8)[a]
Refinement	
Resolution (Å)	40.30 - 2.80 (2.90 - 2.80)[a]
No. reflections	6,189 (624)[a]
R_{work} / R_{free}	0.235 / 0.285
No. atoms	1,409
Protein	1,401
Ligand	8
B-factors (Å^2)	59.49
Protein	59.49
Ligand	58.84
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.52
Ramachandran (%)	
Favor	95.58
Allow	4.42
Outliers	0.00
PDB ID	6KHZ

[a] Highest resolution shell

[0123]

[0124] 여기서, RLGS-AtPRT6 UBR box, RLGS-yUbr1 UBR box, M1Q-hNTAQ1 및 GEEED-p62 ZZ의 단백질 복합체들 중에서 AtPRT6 UBR box, yUbr1 UBR box, hNTAQ1 및 p62 ZZ은 N-degron 경로에서 N-recognin에 해당하는 단백질들이며 이들의 특이적인 기질들인 RLGS, RLGS, M1Q 및 GEEED의 펩타이드와 각각 결합된 복합 단백질 형태를 본 발명의 벡터 시스템을 통해 분리하였고, 이들의 결정화 구조를 확인할 수 있었다. 특히 이들 중에서 RLGS-AtPRT6 UBR box의 결정화 구조는 종래 보고된 바 없었던 것으로 본 발명에서 최초로 단백질의 결정화 구조를 제공한다.

[0125] 또한, 애기장대 유래의 AtPRT6 UBR box는 3개의 zinc 결합 사이트를 갖는 heart 형태의 도메인 구조임을 알 수 있었다.

[0126] 또한, yUbr1 UBR box 및 RLGES 펩티드에 의해 형성된 복합체의 구조와 비교하여, yUbr1 UBR box 및 AtPRT6 box는 유사한 방식으로 Arg1 * 및 Leu2 * 잔기를 인식하는 것을 확인할 수 있었다("*"는 N-말단에 부착된 N-degron 서열을 나타냄) (도 3D 참조).

[0127] 이러한 특성은 RLGS-yUbr1 UBR box 구조분석 결과에서도 확인할 수 있었다(도 3B, 3E 참조).

[0128] 일반적으로 기질의 인식 잔기(Substrate recognition residues)는 서로 다른 종 유래의 UBR box에서 일반적으로 대부분 보존되고 있다. 한편, AtPRT6 UBR box에서 zinc-coordination 잔기들은 human UBR1 UBR box에서는 엄격하게 보존되고 있으나, yeast UBR1에서는 그렇지 않다는 것을 알 수 있다(도 5 참조).

[0129] 또한 hNTAQ1 M1Q 돌연변이체의 경우, 기질 아미노산 Gln* 인식에 관여하는 여러 잔기가 확인되었다(도 3C, 3F 참조). 친수성 Ser28, Tyr80 및 His81 잔기는 기질 Gln N-degron과 수소 결합을 형성하고 hTAQ1의 소수성 Cys26, Trp78 및 Met149 잔기가 상호 작용에 참여하는 것을 알 수 있었으며(도 3F). Gln1* 잔기에 더하여,

Glu2* 및 Asn4* 또한 이전에 공개된 구조와 같이 수소 결합 상호작용을 나타내는 것을 확인할 수 있었다.

[0130] 따라서 이러한 구조적 분석결과로부터, 주어진 단백질의 N-말단에 기질 서열을 융합(부착)시켜 수득한 복합체 형태의 단백질 결정이 펩타이드 침지(peptide soaking) 또는 공결정화(co-crystallization)의 종래 방법을 통해 수득한 결정과 유사하다는 것이 확인함으로써, 본 발명에서 고안한 LC3B 재조합 벡터 시스템으로 발현 및 정제한 단백질 복합체이 결합 포켓에서 기질 서열의 정확한 인식을 통해 안정된 형태의 단백질 결정체를 수득할 수 있음을 알 수 있었다.

[0132] **<실시예 3>**

[0133] **N-데그론 경로 관련 생화학적 연구의 활용 가능성 분석**

[0134] 나아가 본 발명자들은 본 발명의 벡터 시스템이 N-degron 복합체의 구조 결정뿐만 아니라 in vitro에서 N-degron 경로의 생화학적 특성 분석에도 적용할 수 있는지를 확인하기 위한 실험을 수행하였다.

[0135] 보통 번역(translation) 생성물은 N-말단에 Met 잔기를 함유한다. 한편, Arg/N-degron 경로와 관련된 N-degron 기질들은 종종 단백질 분해절단 또는 Arg/N- 데그론 경로의 다른 멤버들에 의한 변형과 같은 번역 후 변형을 통해 다른 N- 데그론 서열을 갖는 것으로 알려져 있다. 따라서 이러한 점에서 종래 in vitro 시스템을 통한 타겟 단백질은 N-말단에 Met 잔기를 가지게 되므로, 실제 세포 내에서 일어나는 정확한 N-degron 경로의 분석이 어려운 문제점이 있었다.

[0136] 따라서 본 발명에서 고안한 LC3B 벡터 시스템은 임의의 단계에서 발견되는 다양한 종류의 아미노산 잔기를 함유하는 N-데그론 기질을 제조할 수 있으며, 이들의 생화학적 활성을 분석할 수 있을 것으로 예상할 수 있었다.

[0137] 이에, 본 발명자들은 p62/SQSTM1에 의해 인식되는 BiP/GRP78의 아르기닐화된 N-말단(이하, R-BiP 이라고 함)을 이용한 분석을 수행하였는데, 이를 위해 본 발명의 LC3B 벡터에 R-BiP가 발현되도록 R-BiP를 암호화하는 유전자 서열을 삽입한 후, in vitro에서 과발현시키고 컬럼 크로마토그래피를 통해 R-BiP 단백질을 정제하였다, 이후 MBP-PB1-p62의 ZZ 도메인을 R-BiP와 혼합한 혼합물을 Superose 6겔 필터컬럼을 통해 단백질 복합체의 형태를 모니터링 하였다.

[0139] 그 결과, 도 6A에 나타낸 바와 같이, 본 발명의 LC3B 벡터 시스템으로 정제된 아르기닐화된 BiP의 N-말단은 MBP-PB1-ZZ와 함께 이동되어 복합체 형태로 존재하는 것으로 확인함으로써 본 발명에서 제조한 R-BiP이 정상적으로 MBP-PB1-ZZ의 기질로 활성을 갖는다는 것을 알 수 있었다.

[0141] 또한, 본 발명자들은 본 발명의 방법으로 분리 정제한 단백질이 N-데그론 경로 연구를 위한 모델 단백질로 사용될 수 있는지를 분석하기 위한 실험을 수행하였는데, 이를 위해 모델 단백질로서 녹색 형광 단백질(GFP)를 사용하였는데, 이는 단백질 수준 및 위치의 시각화와 형광 강도의 감소를 측정함으로써 단백질 분해를 실시간 모니터링하는데 사용될 수 있다. 이를 위해 본 발명에서는 루시닐화된 GFP(L-GFP: leucylated-GFP)를 사용하였다. ClpS는 대장균 N- 말단 규칙 경로에서 중요한 역할을 하는데, N- 데그론 기질들은 ClpP 펩티다제(ClpAP)와 함께 AAA+ 샤페론 ClpA에 의해 분해된다. ClpAP-특이적 어댑터로서, ClpS는 N-말단 불안정화 잔기 (특히 소수성 타입-2N- 데그론 잔기: Leu, Phe, Trp 및 Tyr)에 강하게 결합하고 분해를 위해 기질을 ClpAP에 표적화한다. 본 발명자들은 포유류 세포에서 신규한 2형 타겟단백질의 동정을 위해 ClpS-접합된 컬럼을 준비하였다. 그런 뒤, ClpS 컬럼이 제대로 작동하는지 확인하기 위해 L-GFP를 분리 및 정제한 후, 컬럼에 적용하였다(도 6B 참조). L-GFP가 ClpS 컬럼을 통해 이동하게 되면, 컬럼에 접하된 ClpS와의 결합을 통해 GFP 색, 즉 그린색이 발색되는 것을 확인할 수 있었다. 따라서 본 발명자들은 본 발명의 방법을 세포 추출물로부터 표적 단백질을 분리 및 추출하는 데에도 적용할 수 있음을 알 수 있었다.

[0143] 또한, 계층적 Arg/N-데그론 경로(hierarchical Arg/N-degron)의 첫 번째 단계는 N- 말단 디아미다제(deamidase)에 의해 촉매된다. 효모 유래의 Nta1은 아스파라긴(asparagine)과 글루타민(glutamine)을 아스파르테이트(aspartate)와 글루타메이트(glutamate)로 각각 전환되는 것을 촉매하는 것으로 알려져 있다. 또한 포유동물 시스템에서, NTAQ1 및 NTAN1은 각각 Gln 및 Asn을 탈아미드화 한다.

[0144] 이에 본 발명자들은 본 발명의 LC3B 벡터 시스템을 사용하여 기질 서열이 부착된 yNta1 및 hNTAQ1 효소를 제조하였고, 본 발명에서 제조된 yNta1 및 hNTAQ1 효소는 탈아미드화를 위한 추가적인 보조인자가 필요 없음을 확인하였다. 따라서, 본 발명의 방법으로 분리 및 정제한 효소적으로 활성을 갖는 yNta1 및 hTAQ1은 원하는 생성물 서열 자체를 생성하는 것을 확인할 수 있었다(도 6C 및 표 2 참조).

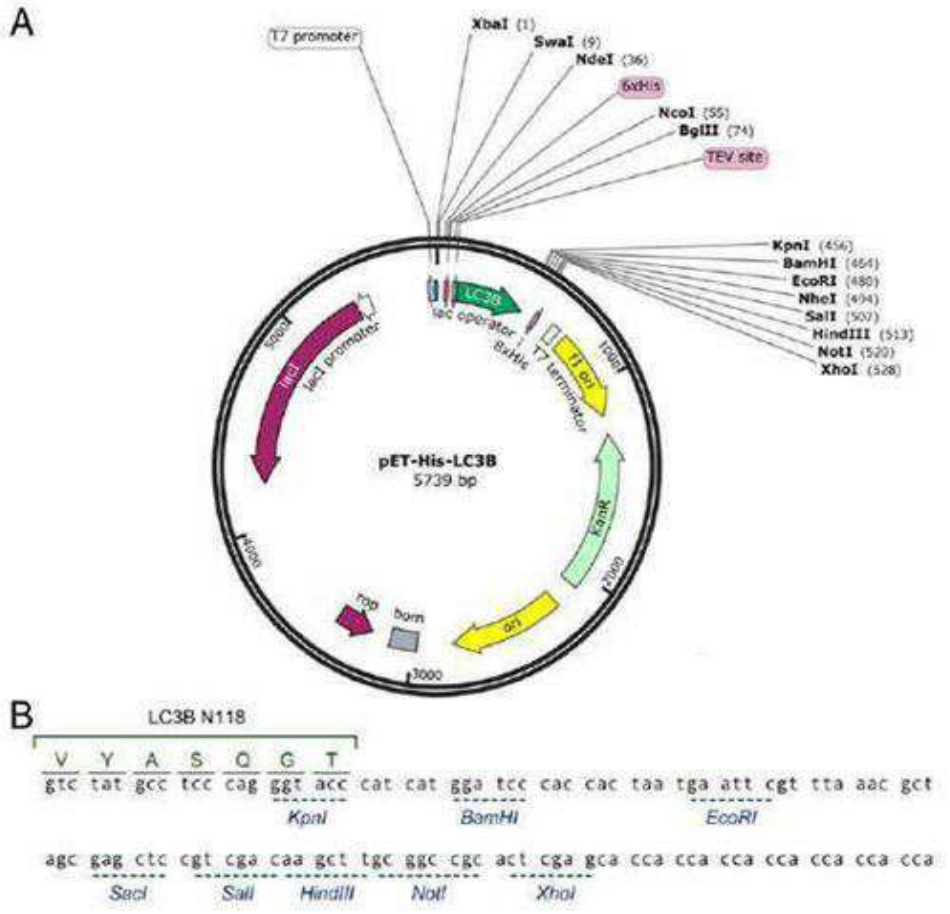
[0145] 따라서 이러한 결과는 hNTAQ1 C28S 돌연변이체가 잔류 효소 활성을 보유하고 있어 단백질의 정제 과정 동안, 기질 서열에서 Gln의 일부가 Glu로 전환되었음을 N- 말단 서열 분석 결과를 통해 확인하였다.

[0147] 이상의 결과를 통해 본 발명자들은 본 발명에서 고안한 LC3B 벡터 시스템을 사용할 경우, N-데그론 경로에 관여하는 N-recognin, 기질 또는 이들의 복합체를 용이하게 분리 및 정제할 수 있을 뿐만 아니라 분석하고자 하는 목적 단백질의 N 말단에 목적하는 임의의 아미노산 잔기가 형성되도록 디자인할 수 있어, in vitro 시스템을 통해서도 N-데그론 경로에 관여된 단백질들의 활성 및 구조 연구를 정확하고 편리하게 수행할 수 있음을 알 수 있었다.

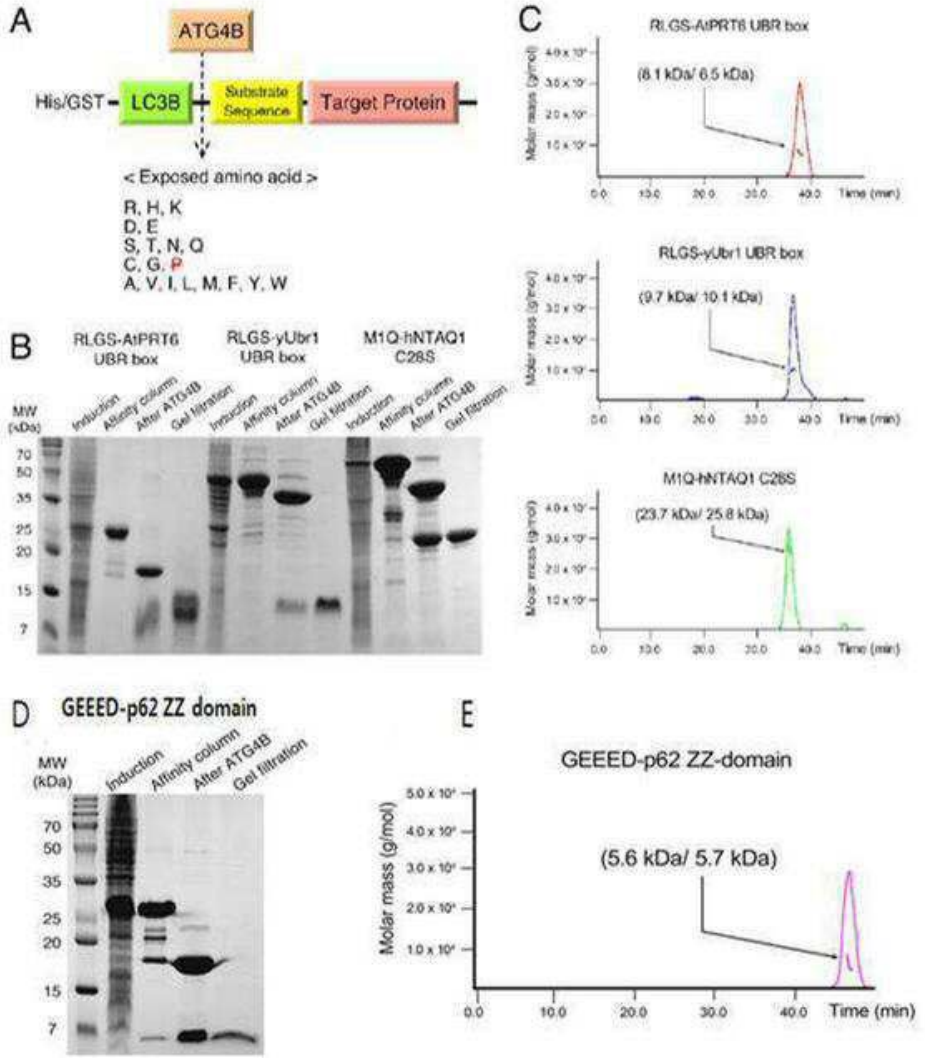
[0148] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

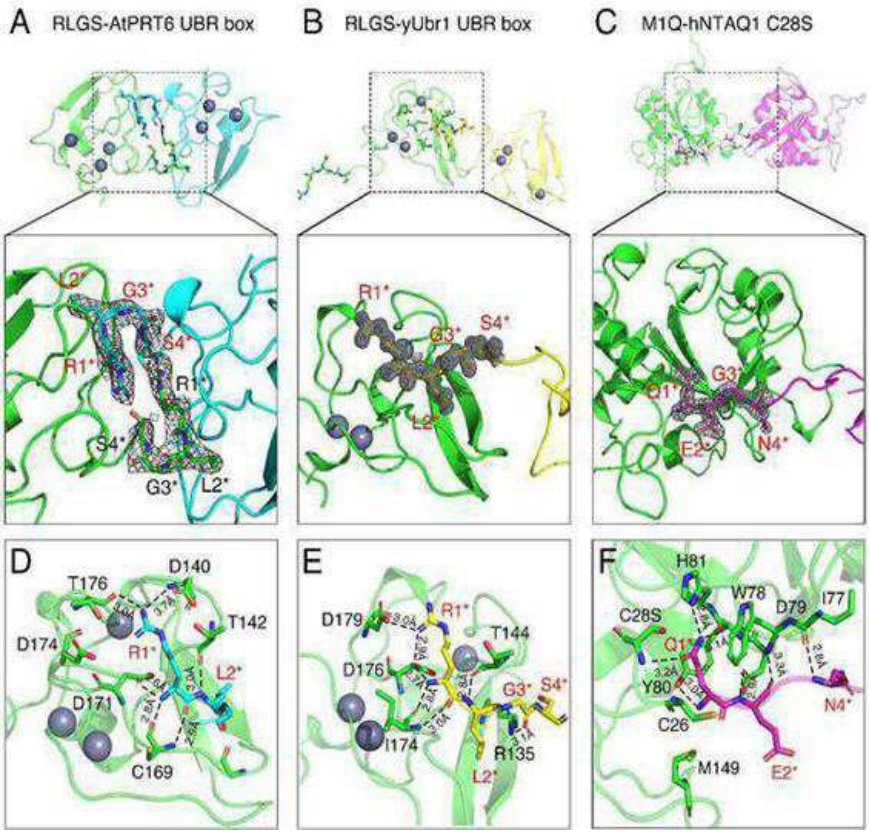
도면1



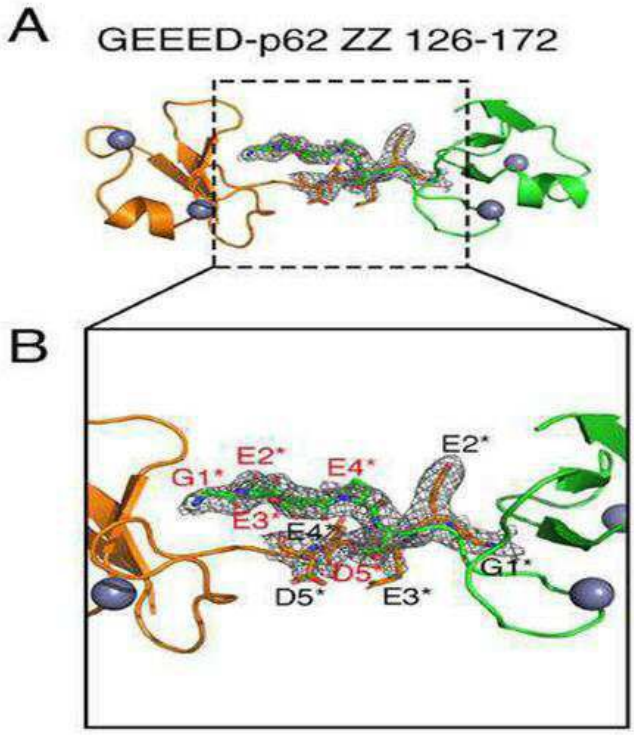
도면2


A
ATG4B
His/GST
LC3B
Substrate Sequence
Target Protein
< Exposed amino acid >
R, H, K
D, E
S, T, N, Q
C, G, P
A, V, I, L, M, F, Y, W
B
RLGS-AtPRT6 UBR box
RLGS-yUbr1 UBR box
M1Q-hNTAQ1 C28S
MW (kDa)
Induction
Affinity column
After ATG4B
Gel filtration
70
50
35
25
20
15
7
C
RLGS-AtPRT6 UBR box
(8.1 kDa/ 6.5 kDa)
RLGS-yUbr1 UBR box
(9.7 kDa/ 10.1 kDa)
M1Q-hNTAQ1 C28S
(23.7 kDa/ 25.8 kDa)
Molar mass (g/mol)
Time (min)
D
GEEED-p62 ZZ domain
MW (kDa)
Induction
Affinity column
After ATG4B
Gel filtration
70
50
35
25
20
15
7
E
GEEED-p62 ZZ-domain
(5.6 kDa/ 5.7 kDa)
Molar mass (g/mol)
5.0 x 10⁴
4.0 x 10⁴
3.0 x 10⁴
2.0 x 10⁴
1.0 x 10⁴
0.0
10.0
20.0
30.0
40.0
Time (min)

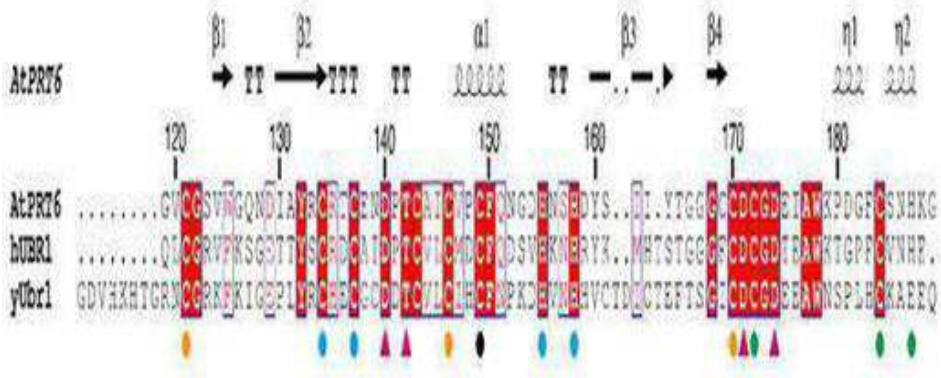
도면3


A RLGS-AtPRT6 UBR box
B RLGS-yUbr1 UBR box
C M1Q-hNTAQ1 C28S
D
E
F

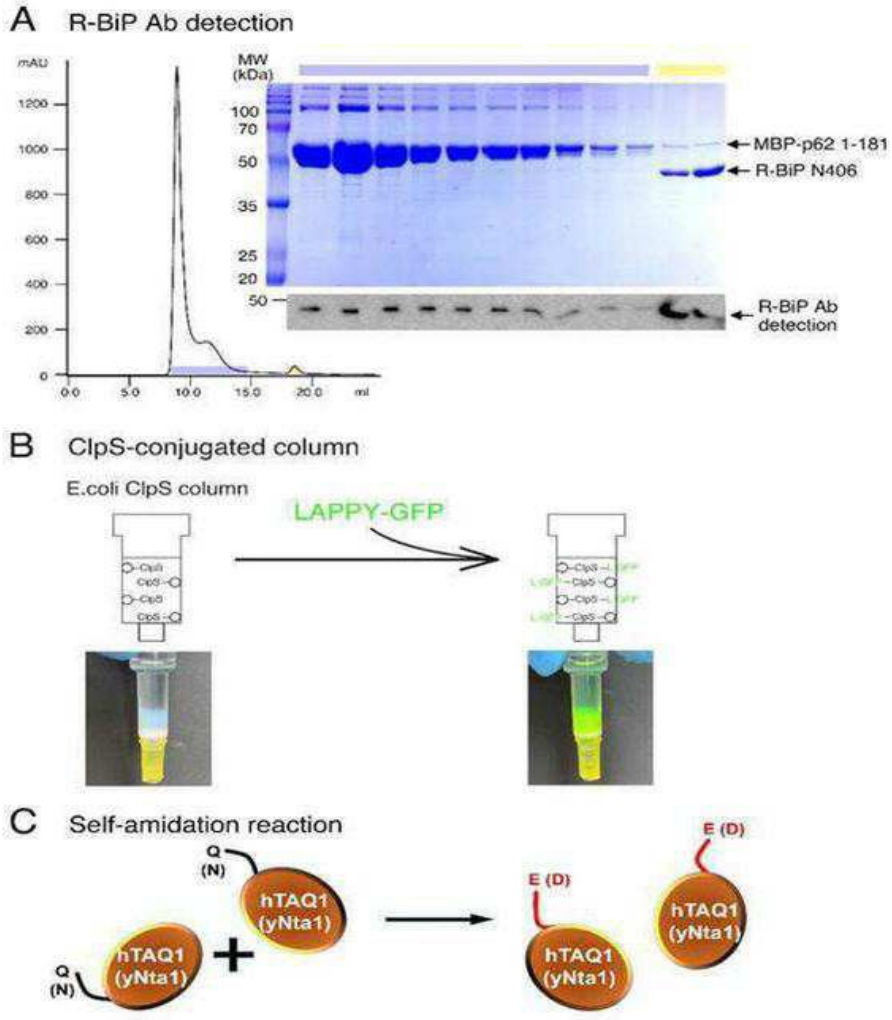
도면4


A GEEED-p62 ZZ 126-172
B

도면5

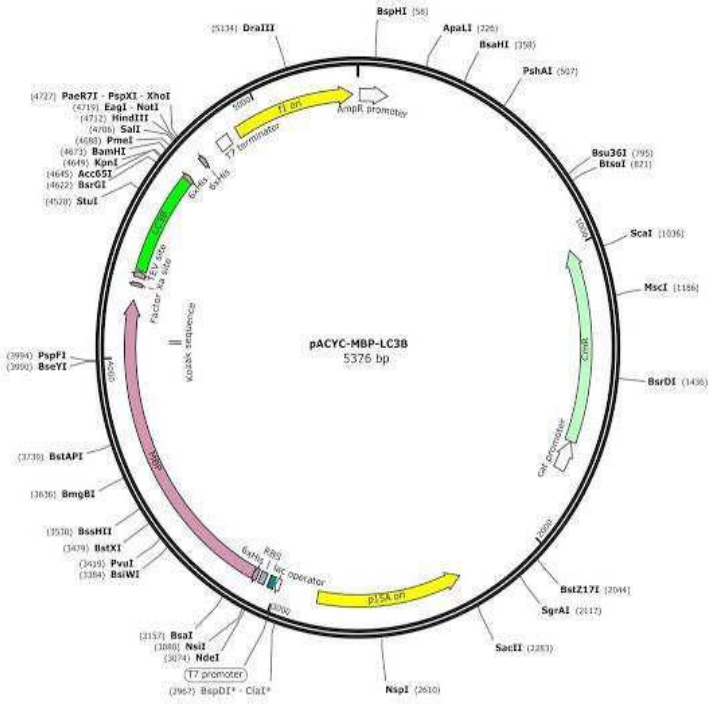

AtPRT6
hUBR1
yUbr1

도면6


A
R-BiP Ab detection
MW (kDa)
MBP-p62 1-181
R-BiP N406
R-BiP Ab detection
B
ClpS-conjugated column
E.coli ClpS column
LAPPY-GFP
C
Self-amidation reaction
hTAQ1 (yNta1)
E (D)
Q (N)

도면7

A



B

LC3B N118

V	Y	A	S	Q	G	T
gtc	tat	gcc	tcc	cag	ggt	acc

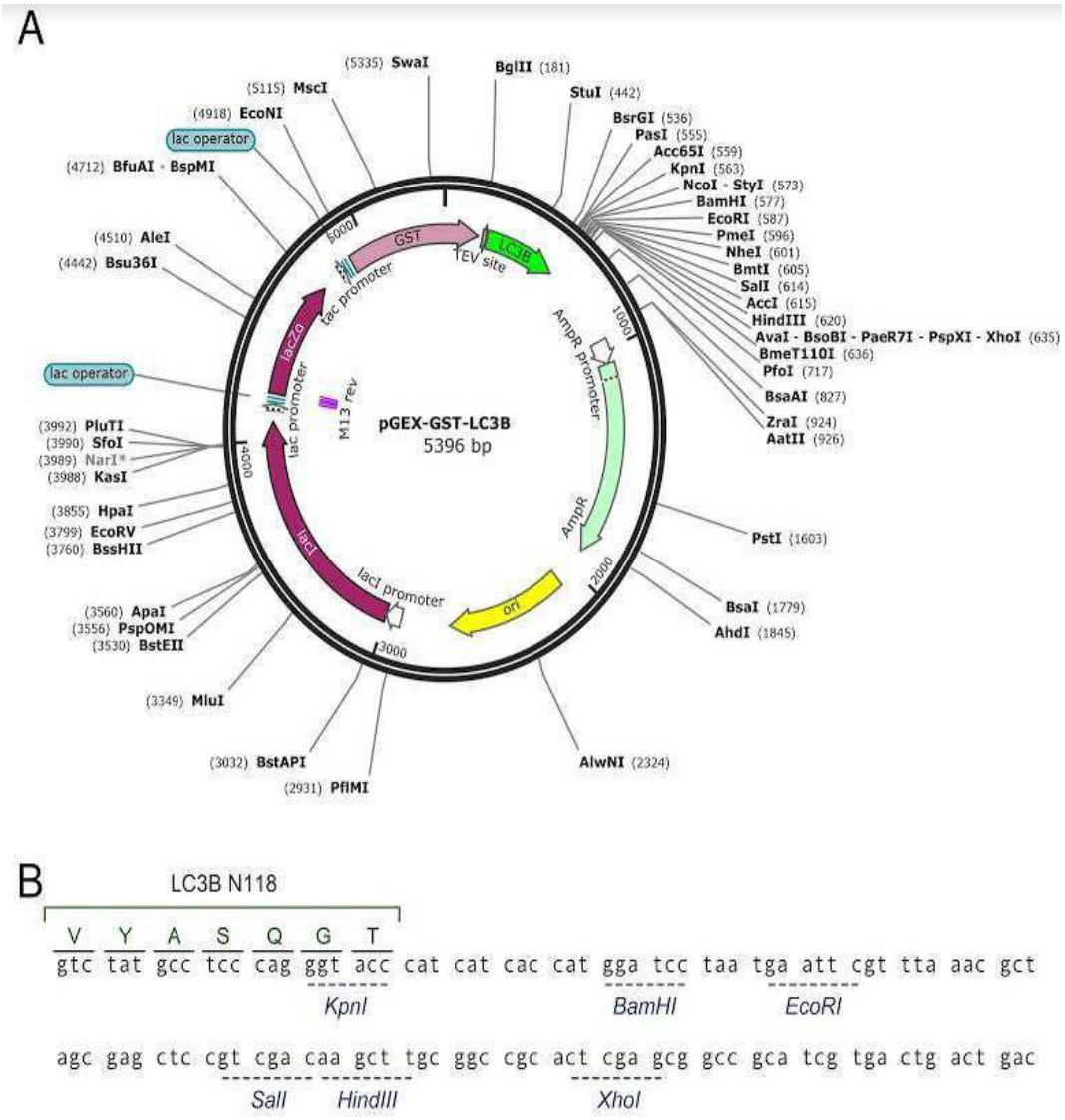
cat cat cac cat cac cac taa tgg atc cga att cgt tta

KpnI *BamHI*

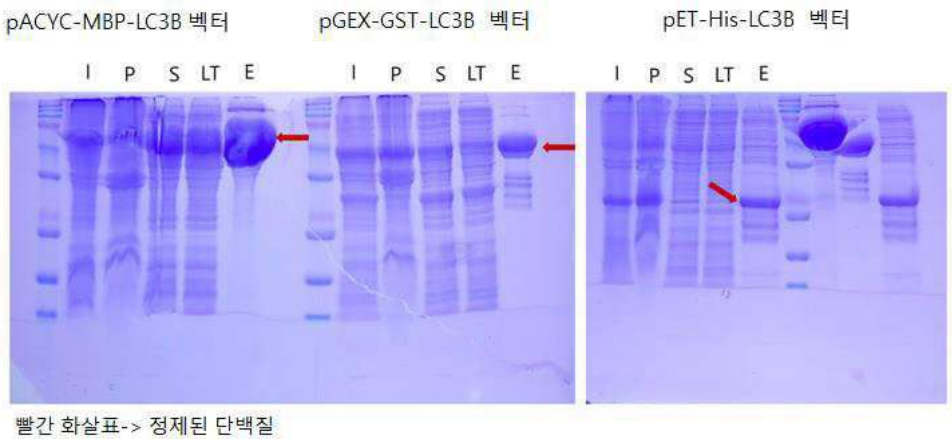
aac gct agc gag ctc cgt cga caa gct tgc ggc cgc act cga gca cca cca cca cca cca

SalI *HindIII* *NotI* *XhoI*

도면8



도면9



빨간 화살표-> 정제된 단백질

서 열 목 록

<110> Industry Academic Cooperation Foundation of Korea University

<120> Recombinant vector for mass production of proteins in the N end rule pathway and use thereof

<130> NPDC81740

<160> 10

<170> KoPatentIn 3.0

<210> 1

<211> 118

<212> PRT

<213> Artificial Sequence

<220><223> LC3B amino acid sequence

<400> 1

Met Pro Ser Glu Lys Thr Phe Lys Gln Arg Arg Thr Phe Glu Gln Arg
1 5 10 15

Val Glu Asp Val Arg Leu Ile Arg Glu Gln His Pro Thr Lys Ile Pro
20 25 30

Val Ile Ile Glu Arg Tyr Lys Gly Glu Lys Gln Leu Pro Val Leu Asp
35 40 45

Lys Thr Lys Phe Leu Val Pro Asp His Val Asn Met Ser Glu Leu Ile
50 55 60

Lys Ile Ile Arg Arg Arg Leu Gln Leu Asn Ala Asn Gln Ala Phe Phe
65 70 75 80

Leu Leu Val Asn Gly His Ser Met Val Ser Val Ser Thr Pro Ile Ser
85 90 95

Glu Val Tyr Glu Ser Glu Lys Asp Glu Asp Gly Phe Leu Tyr Met Val
100 105 110

Tyr Ala Ser Gln Gly Thr
115

<210> 2

<211> 354

<212> DNA

<213> Artificial Sequence

<220><223> LC3B DNA sequence

<400> 2

atgccgtcgg agaagacctt caagcagcgc cgcaccttcg aacaaagagt agaagatgtc 60

cgacttattc gagagcagca tccaaccaaa atcccggtga taatagaacg atacaagggt 120

gagaagcagc ttcctgttct ggataaaaca aagttccttg tacctgacca tgtcaacatg 180

agtgagctca tcaagataat tagaaggcgc ttacagctca atgctaatca ggccttcttc 240

ctgttggtga acggacacag catggtcagc gtctccacac caatctcaga ggtgtatgag 300

agtgagaaag atgaagatgg attcctgtac atggtctatg cctcccaggg tacc 354

<210> 3

<211> 75

<212> PRT

<213> Artificial Sequence

<220><223> RLGS-AtPRT6 UBR box polypeptide sequence

<400> 3

Arg Leu Gly Ser Gly Val Cys Gly Ser Val Trp Gly Gln Asn Asp Ile
1 5 10 15

Ala Tyr Arg Cys Arg Thr Cys Glu Asn Asp Pro Thr Cys Ala Ile Cys
20 25 30

Val Pro Cys Phe Gln Asn Gly Asp His Asn Ser His Asp Tyr Ser Ile
35 40 45

Ile Tyr Thr Gly Gly Gly Cys Cys Asp Cys Gly Asp Glu Thr Ala Trp
50 55 60

Lys Pro Asp Gly Phe Cys Ser Asn His Lys Gly
65 70 75

<210> 4

<211> 225

<212> DNA

<213> Artificial Sequence

<220><223> RLGS-AtPRT6 UBR box polynucleotide sequence

<400> 4

cgccttggca gcggagtatg tgggtccgtg tggggacaga atgatattgc ttaccgttgc 60

cgcacttgcg aaaatgatcc aacttgcgct atctgcgtgc cctgctttca gaatggagac 120

cataattcgc atgattacag cattatttac acgggaggtg gatgttgtga ttgcggtgat 180

gaaacggctt ggaaaccaga cggattttgt agcaaccaca aagga 225

<210> 5

<211> 86

<212> PRT

<213> Artificial Sequence

<220><223> RLGS-yUbr1 UBR box polypeptide sequence

<400> 5

Arg Leu Gly Ser Gly Asp Val His Lys His Thr Gly Arg Asn Cys Gly

1 5 10 15

Arg Lys Phe Lys Ile Gly Glu Pro Leu Tyr Arg Cys His Glu Cys Gly

20 25 30

Cys Asp Asp Thr Cys Val Leu Cys Ile His Cys Phe Asn Pro Lys Asp

35 40 45

His Val Asn His His Val Cys Thr Asp Ile Cys Thr Glu Phe Thr Ser

50 55 60

Gly Ile Cys Asp Cys Gly Asp Glu Glu Ala Trp Asn Ser Pro Leu His

65 70 75 80

Cys Lys Ala Glu Glu Gln

85

<210> 6

<211> 258

<212> DNA

<213> Artificial Sequence

<220><223> RLGS-yUbr1 UBR box polynucleotide sequence

<400> 6

cgtctgggta gcggagacgt ccataaacat accggaagaa attgtgggag gaaattcaaa 60

ataggggaac ccttgtatag gtgtcatgag tgtggttgcg atgatacttg tgtgctttgt 120

attcattgtt ttaatccaaa agatcatgtg aatcatcatg tttgtaccga tatatgtact 180

gaattcacta gtggaatttg tgactgtgga gatgaagaag cttggaattc tccacttcat 240

tgcaaagctg aagaacaa 258

<210> 7

<211> 205

<212> PRT

<213> Artificial Sequence

<220><223> M1Q-hNTAQ1 C28S polypeptide sequence

<400> 7

Gln Glu Gly Asn Gly Pro Ala Ala Val His Tyr Gln Pro Ala Ser Pro
1 5 10 15

Pro Arg Asp Ala Cys Val Tyr Ser Ser Cys Tyr Ser Glu Glu Asn Ile
20 25 30

Trp Lys Leu Cys Glu Tyr Ile Lys Asn His Asp Gln Tyr Pro Leu Glu
35 40 45

Glu Cys Tyr Ala Val Phe Ile Ser Asn Glu Arg Lys Met Ile Pro Ile
50 55 60

Trp Lys Gln Gln Ala Arg Pro Gly Asp Gly Pro Val Ile Trp Asp Tyr
65 70 75 80

His Val Val Leu Leu His Val Ser Ser Gly Gly Gln Asn Phe Ile Tyr
85 90 95

Asp Leu Asp Thr Val Leu Pro Phe Pro Cys Leu Phe Asp Thr Tyr Val
100 105 110

Glu Asp Ala Phe Lys Ser Asp Asp Asp Ile His Pro Gln Phe Arg Arg
115 120 125

Lys Phe Arg Val Ile Arg Ala Asp Ser Tyr Leu Lys Asn Phe Ala Ser
130 135 140

Asp Arg Ser His Met Lys Asp Ser Ser Gly Asn Trp Arg Glu Pro Pro
145 150 155 160

Pro Pro Tyr Pro Cys Ile Glu Thr Gly Asp Ser Lys Met Asn Leu Asn
165 170 175

Asp Phe Ile Ser Met Asp Pro Lys Val Gly Trp Gly Ala Val Tyr Thr
180 185 190

Leu Ser Glu Phe Thr His Arg Phe Gly Ser Lys Asn Cys
195 200 205

<210> 8

<211> 618

<212> DNA

<213> Artificial Sequence

<220><223> M1Q-hNTAQ1 C28S polynucleotide sequence

<400> 8

caggaaggta atggccccgc tgctgtccac taccagccgg ccagcccccc gcgggacgcc 60

tgcgtctaca gcagctgcta cagcgaagaa aatatttgga agctctgtga atacatcaaa 120

aaccatgacc agtatccttt agaagaatgt tatgctgtct tcatatctaa tgagaggaag 180

atgataccta tctggaaaca acaggcgaga cctggagatg gacctgtgat ctgggattac 240

catgttgttt tgcttcatgt ttcaagtgga ggacagaact tcatttatga tctcgatact 300

gtcttgccat ttccctgcct ctttgacact tatgtagaag atgcctttaa gtctgatgat 360

gacattcacc cacagtttag gaggaaattt agagtgatcc gtgcagattc atatttgaag 420

aactttgctt ctgaccgatc tcacatgaaa gactccagtg ggaattggag agagcctccg 480

ccgccatatc cctgcattga gactggagat tccaaaatga acctgaacga tttcatcagt 540

atggatccca aggtaggatg gggcgccgtc tacacactat ccgaatttac acatcggttt 600

ggcagtaaaa actgctga 618

<210> 9

<211> 52

<212> PRT

<213> Artificial Sequence

<220><223> GEEED-p62 ZZ domain polypeptide sequence

<400> 9

Gly Glu Glu Glu Asp Val Ile Cys Asp Gly Cys Asn Gly Pro Val Val
1 5 10 15

Gly Thr Arg Tyr Lys Cys Ser Val Cys Pro Asp Tyr Asp Leu Cys Ser
20 25 30

Val Cys Glu Gly Lys Gly Leu His Arg Gly His Thr Lys Leu Ala Phe
35 40 45

Pro Ser Pro Phe
50

<210> 10

<211> 156

<212> DNA

<213> Artificial Sequence

<220><223> GEEED-p62 ZZ domain polynucleotide sequence

<400> 10

ggcgaagaag aagatgtgat ctgcgatggc tgcaatgggc ctgtggtagg aacccgctac 60

aagtgcagcg tctgcccaga ctacgacttg tgtagcgtct gcgagggaaa gggcttgcac 120

cgggggcaca ccaagctcgc attccccagc cccttc 156