(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。Int. Cl. 7
C12Q 1/68

(11) 공개번호 특2003 - 0007221
(43) 공개일자 2003년01월23일

(21) 출원번호 10 - 2002 - 0041787
(22) 출원일자 2002년07월16일

(30) 우선권주장 1020010042886 2001년07월16일 대한민국(KR)

(71) 출원인
아람 바이오시스템 주식회사
서울특별시 동대문구 장안3동 464 - 1 장안빌딩 405호

(72) 발명자
황인환
경상북도포항시남구지곡동교수APT9동1702호
김대헌
경상남도산청군시천면천평리812번지
이용직
경상북도포항시남구효자동승리APT3동404호
김경파
경상북도포항시효자동포항공대식물단백질이동연구단

(74) 대리인
송호찬
채윤
허은순

심사청구 : 없음

(54) 세포내 과정의 관측 방법 및 그를 위해 사용되는 재조합유전자

요약

본 발명은 세포 내에서 특정 단백질의 이동과 공간적 분포의 형성에 관련된 특이적 특성을 형광 영상을 통하여 선별적으로 확인하는 방법 및 그 방법에 사용되는 재조합 유전자에 관한 것이다. 상기 방법은, (a) 세포내 특정 소기관으로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩(coding)하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자를 준비하는 단계; (b) 상기 재조합 유전자가 세포 내에서 발현될 수 있도록 프로모터(promoter) 부분과 터미네이터(terminator) 부분을 포함하여 구성된 혼성 플라스미드를 준비하는 단계; (c) 상기 (b)단계에서 준비한 혼성 플라스미드 1종 또는 2종 이상을 세포 내에 주입하여 형질전환세포를 준비하는 단계; (d) 상기 형질전환세포에서 신호단백질과 형광단백질이 연결된 리포터 단백질이 발현되도록 하는 단계; 및 (e) 상기 형질전환세포 내에서 발현되는 리포터 단백질의 분포를 발현 중, 이동 중, 또는 그 후에 형광영상 측정을 통하여 관측하는 단계를 포함하여 구성된다.

대표도

도 1

색인어

형질전환, 세포내 소기관, 신호단백질, 형광단백질, 리포터 단백질, 트래피킹

명세서

도면의 간단한 설명

도1은 본 발명에서 그룹 I로서 분류한, 막으로 구성된 세포내 소기관으로 이동하는 리포터 단백질들의 구성 모식도이다.

도2는 상기 그룹 I으로 분류한 리포터 단백질들의 세포 내 발현을 보여주는 신호 영상으로서,

a)는 AtOEP7:GFP가 엽록체 막에 위치하고 있음을 나타내는 사진으로서, 붉은색은 원래 엽록체가 내는 형광이고, 엽록체 막과 상기 단백질이 내는 초록색의 형광이 겹쳐 보이는 부분은 노란색의 형광으로 나타나며,

b)는 a)와 동일한 결과를 필터에 의해 엽록체의 붉은 형광을 제거한 사진이며,

c)는 루비스코 복합단백질이 엽록체의 스트로마(stroma)에 위치하는 것을 나타내는 사진이며,

d)는 Cab:GFP가 엽록체에서 발광하는 것을 나타내는 사진이고,

e)는 GFP에 연결한 루비스코 활성단백질(RA)이 엽록체에 위치하는 것을 나타내는 사진이고,

f)는 GFP에 연결한 F1-H+-ATPase-가 미토콘드리아에 위치하는 것을 나타내는 사진이며,

g)는 GFP에 연결한 퍼록시좀 이동 모티브(SKL)가 퍼록시좀의 막에 위치하는 것을 나타내는 사진이고,

h)는 GFP에 연결한 핵 위치신호(NLS)가 핵에 위치하는 것을 나타내는 사진이다.

도3은 발현된 AtOEP7:GFP에 대한 웨스턴 블랏(Western blot) 사진으로서, T, S, 및 M은 각각 전체 추출물, 용액 부분, 및 막 부분을 나타낸다.

도4는 본 발명에서 그룹 II로서 분류한, 엔도소말 트래피킹(endosomal trafficking)에 의해 세포내 소기관으로 이동하는 리포터 단백질들의 구성 모식도이다.

도5는 상기 그룹 II로 분류한 리포터단 백질들의 세포 내 발현을 보여주는 신호 영상으로서,

a)는 H+-ATPase:GFP가 플라즈마 막에 위치하고 있음을 나타내는 사진이고,

b)는 시알기전환효소(ST)가 골지체에서 형광을 나타낸 사진인데, 이 중 적색 형광은 엽록체의 자연 형광에 의한 것이며,

c)는 리포터 단백질 BiP:RFP가 소포체의 루멘(lumen)을 나타내는 사진이며,

d)는 500:GFP:KKXX가 소포체의 막을 나타내는 사진이며,

e)는 526:GFP가 저장액포를 나타내는 사진이며,

f)는 Chi-n:RFP:Chi-c에 의하여 저장액포가 형광을 나타내는 사진이며,

g)는 신호단백질로 키티나제(chitinase)를 사용하되 카르복시 부분을 연결하지 않은 Chi-n:RFP는 저장액포에 도달하지 못하고 소포체에 점의 형태로 존재함을 나타내는 사진이고,

h)는 SPO:GFP에 의한 형광이 용해성 액포 전체에 고루 분포함을 나타내는 사진이다.

도6은 본 발명에서 그룹 III으로서 분류한, 인지질에 특이적으로 결합하는 리포터 단백질들의 구성 모식도이다.

도7은 본 발명에서 도7은 상기 그룹 III으로 분류한 리포터 단백질들의 세포 내 발현을 보여주는 신호 영상으로서,

a)는 리포터 단백질 GFP:EBD가 각 액포들의 외벽에 형광을 나타내어 이들이 3-인산 포스파티딜이노시톨(phosphatidylinositol 3-phosphate; PI(3)P)을 포함하는 위치로 파악되었으며,

b) 및 c)는 각각, GFP:FAPP1과 GFP:PH가 각각 4-인산 포스파티딜이노시톨 (phosphatidylinositol 4-phosphate ; PI(4)P) 및 4,5-이인산 포스파티딜이노시톨 (phosphatidylinositol 4,5-diphosphate; PI(4,5)P2)이 분포하는 플라즈마 막에 형광을 나타냄을 도시하는 사진이다.

도8은 본 발명에 따라 그룹 I로 분류된 리포터 단백질의 세포내 이동(trafficking)에 대한 약물의 저해 효과를 나타내는 사진으로서, 3-인산화 포스파티딜이노시톨(PI(3)P) 및 4-인산화 포스파티딜이노시톨(PI(4)P)의 저해제인 워트마닌의 RbcS:GFP의 수송에 대한 영향을 나타내는 사진으로서,

a)는 정상 대조군에서는 RbcS:GFP의 녹색 형광이 예상대로 RbcS:GFP가 수송되어야 하는 엽록체 내로 이동된 것이 관찰된 사진이고,

b)와 c)는 모두 워트마닌 존재 시 녹색 형광이 엽록체 내로 수송되지 못하고 작은 반점 혹은 엉킨 형태로 관찰된 것을 나타내는 사진이다.

도9는 본 발명에 따라 그룹 II로 분류된 리포터 단백질의 세포내 이동(trafficking)에 대한 약물의 저해 효과를 나타내는 사진으로서,

a) 및 b)는 정상 대조군에서 500:GFP:KKXX의 녹색 형광이 수많은 네트웍 형태로 관찰되는 것을 나타내는 사진이며,

c) 및 d)는 각각 BafA1 존재 시 녹색 형광이 플라즈마막과 액포막에서 고리 모양으로 나타나는 것이 관찰된 사진이다.

도10은 두 가지 리포터 단백질의 동시 발현 시 특이적인 저해제에 의한 변화를 나타내는 사진으로서,

a) 및 b)는 정상 대조군에서 각각 녹색과 적색의 형광을 나타내는 사진이고,

d)는 브레펠딘 A를 처리하였을 때 BiP에 대한 영향은 없음을 나타내는 사진이며,

e)는 브레펠딘 A를 처리하였을 때 골지체의 파괴로 인하여 ST가 골지체에 존재하지 못하고 소포체로 이동하였음을 나타내는 사진이고,

c) 및 f)는 각각 브레펠딘 A로 처리하지 않은 상기 대조군들과 브레펠딘 A를 처리하였을 때의 결과가 형광 현미경 하에서 뚜렷한 차이를 나타냄을 보여주는 사진들이다.

도11은 소기관의 구조를 파괴하는 brefeldin A의 작용을 시각적으로 비교할 수 있는 사진으로서,

a)는 정상 대조군에서는 BiP:RFP의 적색 형광이 소포체의 구조를 따라 분포하고 있음을 나타내는 사진이고,

c)는 brefeldin A를 첨가하였을 때 소포체의 구조가 파괴되고 있음을 가시적으로 보여주는 사진이며,

b) 및 d)는 상기 a) 및 c)의 시료에 대하여, 리포터 단백질에 의한 형광을 사용하지 않고, 광학 현미경으로 관찰한 결과를 나타내는 사진이다.

도12는, 본 발명에서 그룹 III으로 분류되는 리포터 단백질들에 대한 약물의 저해 작용을 나타내는 도면으로서,

a)의 사진들은 각각, 약물을 처리하지 않은 대조군 세포에서는 EBD가 결합하는 인지질 PI(3)P의 분포를 따라서 각 액포의 막에서 녹색의 형광이 나타나는 것을 볼 수 있으나, 저해제 워트마닌과 LY294002를 처리한 세포에서는 세포질 전체에 형광이 고루 분포함을 나타내는 것이고,

b)의 사진들은 각각, EBD의 1358번째 아미노산에 변이가 생긴 GFP:EBDC1358S를 대상으로 동일한 실험을 실시하였을 때는 세포 내의 형광 분포에 영향을 주지 못하는 것을 나타내는 사진들이다.

도13은 단백질 발현 저해제에 의한 형광 강도의 변화를 나타내는 사진들로서,

a)는, 정상 대조군(b)에 비하여, 시클로헥시미드(cycloheximide)를 처리한 세포에서의 녹색 형광의 강도가 감소한 것을 나타내는 사진이고,

c) 및 d)는 세포 내에 있는 엽록체에 의한 자연 형광으로서, 세포의 성장 및 대사 상태를 상대적으로 비교하기 위하여 관측한 사진이다.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 세포분자유전학 분야의 기술로서 세포 내에서 특정 단백질이 특정 세포내 소기관으로 이동하는 과정을 선별적으로 관찰할 수 있게 하고, 또한 상기 특정 단백질의 발현 과정 또는 세포내 이동 과정에 영향을 미치는 신약 물질을 선별적으로 검색할 수 있도록 하는 기술이며, 특히 형광단백질로 표지된 리포터단백질(reporter protein)이 발현될 수 있도록 준비한 형질전환세포를 이용하는 세포 기반의 검출시스템에 관한 기술이다.

최근의 생명과학분야 기술의 급속한 발전은 생명체의 전체 유전자 서열 또는 전체 단백질의 발현 양상에 대한 일차적 정보를 수집하는 것을 가능하게 하여, 이러한 총괄적 일차 정보들을 대상으로 유용유전자나 유용단백질을 찾아내거나, 이러한 대량의 생명정보들 사이의 상호 연관관계를 밝히고 이를 이용하고자 하는 유전체학(Genomics)과 단백질체학(Proteomics)이라는 새로운 기술개발 분야의 중요성이 부각되고 있다. 막대한 양의 총괄적 생명정보로부터 이러한 유용 정보들을 발굴하고 이용하기 위해서는 본 발명에서 구현한 바와 같은 복잡한 생명현상의 세부적 과정들을 보다 선별적으로 분리하여 빠르게 관측할 수 있는 기술의 개발이 필요불가결하며, 최근의 여러 종래 기술들에서도 이러한 기술개발 경향이 다양하게 나타나고 있다.

유전체학 및 단백질체학 분야의 발전에서 비롯될 방대한 양의 유용 생명정보들의 가장 대표적인 사용처는 신약 개발 분야라고 할 수 있다. 이러한 유용 생명정보들에 근거한 효율적이고도 선별적인 신약 검색 시스템의 구축은 향후의 생명과학기술 분야의 발전에 주도적 역할을 할 것으로 기대되고 있으며, 이에 따라 새로운 신약 검색 시스템의 구축을 위한 기술개발 연구들이 최근 들어 아주 활발하게 추진되고 있다.

신약 검색법의 개발은 기본적으로 (1) 질병의 원인 또는 생명현상 제어와 관련된 특이적인 타겟(target)의 발견 및 설정, (2) 상기 특이적인 목표에 미치는 약물의 영향을 선별적으로 구분하여 판정할 수 있는 분석 방법의 개발 및 최적화로 이루어진다. 따라서, 효율적인 신약 검색법의 개발을 위해서는 대량의 유전 정보와 단백질 정보에 의해 나타나는 복잡한 생명현상 중에서 특정 현상의 원인을 제공하는 구체적인 요인, 즉 유전 정보나 단백질 정보, 또는 이들의 기능과 연관된 세포내(intracellular) 또는 세포간(intercellular) 과정을 선별적으로 구분하여 관측할 수 있는 기술의 개발이 선행되어야 한다. 이러한 선별적 관측을 가능하게 하는 기술은 특정 현상의 발현을 일어나게 하는 구체적 생명 정보 요인, 즉 신약 검색의 세부적 타겟을 발견하는 것을 용이하게 할 뿐만 아니라, 신약 검색 기술을 보다 선별적 수준에서 실시할 수 있는 체계화된 분석 방법을 구축할 수 있는 기반을 제공하게 된다.

상기 특이적 타겟을 대상으로 약물의 영향을 판정하기 위한 분석 방법은 다음과 같은 3가지로 구분할 수 있다. 즉, 질병의 발현 메카니즘에 결정적인 기여를 하는 효소나 수용체들의 활성에 대한 약물의 직접적인 영향을 관찰하여 판정하는 단백질 (또는 분자) 기반의 방법; 생명체 개체의 형태학적 또는 생화학적 특성에 미치는 약물의 영향을 관찰하여 판정하는 개체 기반의 방법; 그리고, 세포나 세포내 소기관의 형태, 세포내 단백질의 발현, 이동 및 대사 등에 미치는 약물의 영향을 관찰하고 판정하는 세포 기반의 방법으로 구분할 수 있다. 단백질 (또는 분자) 기반 방법의 경우, 정확한 목표 단백질을 대상으로 분석을 수행할 수 있다는 장점이 있지만, 시험관 내(in vitro)에서 수행되므로 세포내의 여러 가지 복합적인 요인들에 의한 영향을 고려할 수 없기 때문에 신약후보물질을 생체에 적용하기 위해서는 장시간을 요하는 많은 추가적인 실험들이 필요하게 된다. 그런 반면, 개체 기반 또는 세포 기반의 방법은, 개체 수준 또는 세포 수준에서 일어나는 종합적인 변화를 관찰할 수는 있으나, 약물의 영향이 정확한 목표 유전자 또는 목표 단백질에 대해 미치는 선별적 결과임을 확인하기 위해서는, 단백질 기반의 방법과 마찬가지로 장시간을 요하는 추가실험이 필수적이게 된다.

이러한 종래의 신약 검색법들의 문제점들을 극복하기 위한 방안으로서 분자유전학적 형질전환 기법을 이용한 새로운 기술들이 개발되고 있다. 형질전환에 의하여 특이적으로 나타나는 개체 또는 세포 수준에서의 변화를 관찰함으로써, 특정 유전자의 유전적 기능 또는 이로부터 발현된 단백질의 생물학적 기능을 밝히고, 이러한 형질전환개체 또는 형질전환세포를 대상으로 형질전환에 의해 일어나는 특이 현상에 미치는 약물 후보물질의 영향을 관측함으로써 신약물질을 보다 선별적으로 검색할 수 있게 된다.

개체 또는 세포에서 일어나는 특정 현상을 선별적으로 관측하기 위하여 전통적으로 동위원소 표지법이 많이 사용되어 왔으나, 최근 들어 세포 내에서 발현이 가능한 녹색 형광단백질(Green Fluorescent Protein; GFP)과 적색 형광단백질(Red Fluorescent Protein; RFP)이 개발되고(Morise 등, 1974), 상기 형광단백질들의 여러 변이체들이 개발됨에 따라(Evans, WO98/21355), 세포 내에서의 형광단백질의 발현을 이용하는 다양한 연구결과들이 보고되고 있다. 형광단백질을 사용한 연구 예로서, 유전자의 발현 및 단백질의 세포내 분포에 관한 연구를 비롯하여(Chalfie와 Prasher, US5491084), 세포내 기관의 가시화(Kost 등, 1998), 분비경로에 따른 단백질 수송의 가시화(Kaether와 Gerdes, 1995), 식물세포에서의 단백질의 발현 양상의 가시화(Hu와 Cheng, 1995), 초파리 배아에서의 단백질의 발현 양상의 가시화(Davis와 Viestra, 1998) 등이 있다. 형광단백질 발현을 이용한 신약 검색 방법으로는, 돌연변이 탐색을 위하여 변형된 GFP가 발현되게 형질전환한 개체를 이용하는 방법(Ward와 Chalfie, WO95/21191), 세포외 신호를 세포내로 도입하는 것을 검출하기 위하여, 세포 표면 수용체의 활성에 관여하는 전사조절인자와 형광단백질이 연결된 혼성단백질이 발현되게 형질전환한 세포를 이용하는 방법(Harpold 등, US5401629) 및 단백질 인산화효소의 활성부위를 가지는 단백질과 형광단백질이 연결된 혼성단백질이 발현되게 형질전환한 세포를 이용하여, 인산화효소와 관련된 세포내 과정에 영향을 미치는 생물학적 활성물질을 검색하는 방법(Thastrup 등, WO96/23898) 등이 있다.

형광단백질의 발현을 이용하는 이러한 다양한 연구들 중에서 특정 단백질이 세포내에서 발현되고 이동되는 과정을 형

광영상을 통하여 가시화함으로써, 복잡한 세포내 과정들 중에서 특정 단백질과 관련된 세포내 과정을 선별적으로 용이하게 확인할 수 있는 예들이 제시되고 있는데(Ward와 Chalfie, WO95/21191; Kost 등, 1998; Gillooly 등, 2000; Pih 등, 2000), 이러한 연구결과들은 특정 세포내 과정에 영향을 미치는 약물을 선별적으로 검색할 수 있는 방안을 제공한다. 종래의 다양한 연구들을 통하여 특정 세포내 소기관으로 이동되는 특성을 지닌 신호단백질들이 다수 알려져 있다. 즉, 엽록소 결합 단백질은 엽록체로 이동하며, SV40의 핵 위치신호 부위(NLS domain)는 핵으로 이동하며(Goldfarb 등, 1986), 프록시좀(peroxisome) 신호이동 모티브인 SKL은 퍼록시좀으로 이동하며(Davis와 Viestra, 1998), F1-H+-ATPase-:RFP는 미토콘드리아로 이동하는 (Niwa 등, 1999) 등 대부분의 단백질은 그 기능과 연관된 특정 세포내 소기관으로 이동하게 된다. 따라서, 이와 같이 세포 내에서 특이적 이동 특성을 나타내는 신호단백질에 형광단백질이 연결된 혼성단백질이 세포 내에서 발현되게 함으로써, 특정 신호단백질의 세포내 이동 특성을 가시화하여 관찰하는 것을 용이하게 할 수 있으며(Harpold 등, US5401629; Kost 등, 1998), 상기 특정 세포내 과정에 미치는 약물의 영향을 선별적으로 관찰하고 평가하는 것을 가능하게 할 수 있다(Morinaga 등, 1999). 또 다른 예로서 특정 인지질과 특이적으로 결합하는 것으로 알려진 단백질에 형광단백질이 연결된 혼성단백질이 세포 내에서 발현되게 함으로써 인지질 특이적 세포내 과정의 관찰을 가능하게 할 수 있다. 예를 들어 PLC-Delta PH(Stauffer 등, 1998), AtPH 및 FAPP1(Dowler 등, 2000) 구역들은 PI(4,5)P2, PI(3)P 및 PI(4)P 인지질과 각각 특이적으로 결합하는 것으로 알려져 있는데, PH 구역과 GFP를 포함하는 혼성단백질을 세포 내에서 발현시킴으로써 이 혼성단백질이 플라스마 막으로 이동한다는 것을 밝힌 예가 있다 (Kost 등, 1998).

상기 기술된 내용에서 예시된 바와 같이, 특정 단백질의 세포내에서의 이동 및 분포 특성을 형광단백질로 표지하여 관측함으로써 세포내 과정의 세부적 규명 및 이에 근거한 선별적 신약 검색이 가능하다고 할 수 있다. 그러나, 수천에서 수만 가지 이상에 이르는 단백질들이 관여하는 대단히 복잡한 세포내 과정들을 선별적으로 체계화하여 관측할 수 있는 실질적 방안을 마련하기 위해서는, 다양한 세포내 과정들을 선별적으로 관측할 수 있는 형광단백질로 표지된 다수의 혼성단백질들을 준비하여야 하고, 이러한 혼성단백질들이 세포 내에서 일으키는 특이적 특성들의 규명이 선행되어야 한다. 또한, 이러한 혼성단백질의 발현이 세포내에서 효율적으로 일어나도록 형질전환세포를 준비하는 방법 및 혼성단백질로부터 방출되는 형광신호를 효율적으로 영상화하고 분석할 수 있는 구체적이고도 규격화된 방안이 마련되어야 한다.

발명이 이루고자 하는 기술적 과제

상기와 같은 요구에 따라, 본 발명은 복잡하고도 다양한 세포내 과정들을 선별적으로 관측하는 방법 및 이러한 방법에 근거하여 특정 세포내 과정에 영향을 미치는 약물을 판정하는 선별적 신약 검색 방법을 구체화하여 제공하는 것을 그 목적으로 한다.

상기 목적을 달성하기 위하여,

첫째, 본 발명에서는 세포내 특정 소기관으로의 위치이동을 제어하는 신호를 가진 특정 신호단백질과 형광을 나타내는 형광단백질이 연결된 리포터 단백질이 발현될 수 있도록 형질전환된 세포를 효율적으로 준비할 수 있는 방안을 제공하고자 한다.

둘째, 본 발명은 신호단백질을 코딩(coding)하는 유전자와 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 다양한 재조합 유전자를 제조할 수 있는 구체적 방법들을 제공하며, 이러한 재조합 유전자가 세포 내에서 효율적으로 발현되도록 하는 혼성 플라스미드의 제조 방법을 제공하고자 한다.

또한, 상기 혼성 플라스미드를 사용하여 형질전환세포를 준비하고, 상기 형질전환세포 내에서 리포터 단백질을 효율적으로 발현시킬 수 있는 체계적인 방법을 제공하고자 한다. 나아가, 상기 리포터 단백질들의 발현 과정 및 이동 과정에서 나타나는 리포터 단백질의 세포 내에서의 세부적 분포를 형광영상을 통하여 상기 세포가 살아있는 상태를 유지하면서 계속하여, 또는 단계별로 관측할 수 있는 구체적 방법을 제공하고자 한다.

이와 같이 형질전환세포를 준비하여 특정 단백질의 세포 내에서의 발현과 이동에 따른 특이적 공간적 분포를 측정할 수 있는 체계화된 효율적인 방법을 제공함에 더하여, 본 발명에서는 이러한 형질전환세포를 이용하여 특정 세포내 과정에 영향을 미치는 약물을 판정할 수 있는 선별적 신약 검색법을 제공하는 것을 그 목적으로 한다.

상기 선별적 세포내 과정의 관측법 및 이에 근거한 선별적 신약 물질 검색법의 활용 효율을 보다 구체화시키고 다양화시키기 위하여, 본 발명에서는 또한 핵, 미토콘드리아, 엽록체, 퍼록시좀(peroxisome), 세포막, 소포체, 골지체, 저장 액포(storage vacuole), 용해성 액포(lytic vacuole), 및 전액포(prevacuolar compartment) 등의 다양한 세포내 소기관으로 특이적으로 이동하는 특성을 지닌 다양한 신호단백질을 대상으로 한 재조합 유전자를 제공하고자 한다. 또한, 상기 재조합 유전자의 제조에서부터 형질전환세포의 준비에 이르는 전반적인 과정들을 최적화하여 제공하는 것을 그 목적으로 한다.

발명의 구성 및 작용

상기와 같은 목적을 달성하기 위하여, 본 발명의 첫번째 측면에서는,

세포 내에서 특정 단백질의 이동과 공간적 분포에 관한 특이적 특성을 확인하는 방법으로서,

(a) 세포내 특정 소기관으로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩(coding)하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자를 준비하는 단계와;

(b) 상기 재조합유전자가 세포 내에서 발현될 수 있도록 프로모터(promoter) 부분과 터미네이터(terminator) 부분을 포함하여 구성된 혼성 플라스미드를 준비하는 단계와;

(c) 상기 (b)단계에서 준비한 혼성 플라스미드 1종 또는 2종 이상을 세포 내에 주입하여 형질전환세포를 준비하는 단계와;

(d) 상기 형질전환세포에서 신호단백질과 형광단백질이 연결된 리포터 단백질이 발현되도록 하는 단계; 및

(e) 상기 형질전환세포 내에서 발현되는 리포터 단백질의 분포를 발현 중, 이동 중, 또는 그 후에 형광영상 측정을 통하여 관측하는 단계를 포함하여 구성되는 방법을 제공한다.

또한, 본 발명의 두번째 측면에서는,

세포 내에서 특정 단백질의 이동과 공간적 분포 형성에 관련된 특이적 특성에 영향을 미치는 화학물질을 검색하는 방법으로서,

(a) 세포내 특정 소기관으로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩(coding)하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자를 준비하는 단계;

(b) 상기 재조합 유전자가 세포 내에서 발현될 수 있도록 프로모터(promoter) 부분과 터미네이터(terminator) 부분을 포함하여 구성된 혼성플라스미드를 준비하는 단계;

(c) 상기 (b)단계에서 준비한 혼성플라스미드 1종 또는 2종 이상을 세포내에 주입하여 형질전환세포를 준비하는 단계;

(d) 상기 형질전환세포에서 신호단백질과 형광단백질이 연결된 리포터단백질이 발현되도록 하되, 상기 리포터단백질이 발현되기 전, 후 또는 동시에, 생물학적 활성을 가질 것으로 추정되는 화학물질로 상기 형질전환세포를 처리하는 단계;

(e) 상기 화학물질로 처리된 형질전환 세포 내에서 발현되는 리포터단백질의 분포를 발현 중, 이동 중, 또는 그 후에 형광영상 측정을 통하여 관측하는 단계; 및

(f) 상기 (e)단계에서 측정한 형광영상을, 화학물질로 처리하지 않은 대조군의 형광영상과 비교하여 화학물질의 영향을 판정하는 단계를 포함하여 구성되는 방법을 제공한다.

본 발명의 세번째 측면에서는, 상기 방법들에서 사용되며 세포내 특정 소기관으로 이동하여 세포내 단백질의 이동 및 세포내 소기관의 공간적 분포를 가시화할 수 있도록, 그것이 주입된 형질전환된 세포 내에서 발현되는 리포터 단백질을 코딩(coding)하는 재조합 유전자를 제공하는데, 상기 재조합유전자는 세포내 특정 소기관으로의 위치 이동을 제어하는 부분을 포함하는 신호단백질을 코딩하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결되어 제조되는 것을 특징으로 한다.

상기 본 발명의 구성을 구체적으로 설명하면, 본원 발명은 핵, 미토콘드리아, 엽록체, 퍼록시좀(peroxisome), 세포막, 소포체, 골지체, 저장액포(storage vacuole), 용해성 액포(lytic vacuole), 및 전액포(prevacuolar compartment) 등의 세포내 소기관과 3종의 인지질로 각각 이동하는 특이적 특성을 가진 NLS(핵 위치 신호 서열), AtOEP7, Cab (엽록소 a/b 결합 단백질), SKL(퍼록시좀 이동 모티브), RbcS(루비스코 스몰 서브유닛), RA (루비스코 액티바제) 및 F1- H+ - ATPase, H+ - ATPase, BiP(샤프론 결합단백질), ST (시알릴트랜스퍼라제), Chi(키티나제), 혼성클론 526, 클론 491, 클론 500, AtVTI1a, SPO(스포라민), EBD, AtPH, FAPP, 및 PH 등의 신호단백질들을 선정하고, 이들을 대상으로 형광단백질로 표지된 리포터 단백질이 발현될 수 있는 형질전환세포를 준비하는 방법을 제공한다. 그리고, 상기 형질전환세포를 이용한 선별적 세포내 과정의 관측법 및 이에 근거한 선별적 신약 검색법을 규격화하여 구현하였다.

보다 구체적으로, 상기 본 발명의 방법은, 세포 내에서 특정 위치로 이동하는 단백질들을 선정하여 그 신호단백질의 전부 또는 위치이동의 정보를 포함하고 있는 일부분에 해당하는 유전자를 합성하고, 이들 신호단백질과 함께 연결되어 위치를 시각화 할 수 있도록 형광을 나타내는 형광단백질의 유전자를 합성하여, 이들이 각각 하나 또는 둘 이상 연결된 재조합 유전자를 제조하는 과정을 포함한다. 신호단백질과 형광단백질이 결합하는 방법에 따라 신호단백질의 기능에 차이가 발생하며, 본 발명에서는 신호단백질이 정확한 위치이동을 할 수 있는 리포터단백질의 구성을 밝혀내고 그 제조법을 확립하였다.

본 발명은 또한, 상기 재조합 유전자가 세포 내에서 발현이 가능하도록 혼성 플라스미드를 제조하는 과정을 포함한다. 이는 프로모터, 터미네이터(terminator), 또는 다른 필요한 인자들(factors)을 포함하는 핵산 운반체에 상기 재조합 유전자를 연결하는 방법으로 구성될 수 있다. 혼성 플라스미드를 주입하여 세포를 형질전환시키는 방법은 당 분야의 통상의 지식을 가진 기술자들에게 잘 알려진 바와 같이, PEG(polyethylene glycol), 인산칼슘, DEAE - 덱스트란 등의 화학물질을 이용하는 방법, 양이온성 지질을 이용하는 라이포펙션(lipofection) 방법, 마이크로인젝션법, 일렉트로포레이션법, 일렉트로퓨전법 등을 사용할 수 있으며, 이들에 제한되지 않는다. 이 때 도입되는 혼성 플라즈미드는 1종을 사용할 수도 있고 2종 이상을 사용하여 두 가지 이상의 리포터단백질이 동시에 발현될 수 있도록 할 수도 있다. 형질전환된 세포에서 신호단백질과 형광단백질이 연결된 리포터 단백질이 효율적으로 발현되도록 조건을 최적화할 필요가 있는데, 이는 당 분야에서 통상의 지식을 가진 자에 의하여 각 신호단백질 및 형광단백질의 종류에 따라 적절히 조절하여 선택할 수 있을 것이며, 본원 발명의 경우, 이후 설명하는 실시예에서 그들의 구체적인 아미노산 서열이나 그에 해당하는 유전자 서열을 나타내는 혼성 플라스미드의 구성을 상세히 설명한다.

본 발명에 따른 리포터 단백질을 구성함에 있어서, 세포 내에서 단백질이 각 소기관으로 이동하는 여러 가지 메카니즘에 대하여 공통적으로 적용될 수 있음을 보이기 위하여, 신호단백질을 단백질 이동신호 메카니즘에 근거하여 세가지 그룹으로 분류하였다. 즉, 신호단백질의 특정 부위가 세포 내에서의 이동을 위한 직접적 인지 신호로 작용하는 체계에 해당하는 세포내 소기관들인 핵, 엽록체, 미토콘드리아 등에 특이적인 단백질들(이하, 본원 명세서에서 편의상 그룹 I로 분류하기로 한다. 도1 참조)과, 신호단백질의 특정 부위가 소포(vesicle)에 포획되는 신호로 작용하여 소포에 둘러싸여 이동하는 엔도소말 이동(endosomal trafficking) 체계에 해당하는 세포내 소기관들인 소포체, 골지체, 용해성 액포, 저장액포, 세포막 등에 특이적인 단백질들(이하, 본원 명세서에서 편의상 그룹 II라고 한다. 도4 참조)로 세분하여, 그들이 각 기관별로 이동하는 과정과 그 공간적 분포를 가시화 할 수 있도록 하는 구체적 방안을 제시하였다. 이와 함께, 세포내 신호전달 과정에 관여하는 인지질에 특이적인 결합을 하는 단백질들(이하, 본원 명세서에서 편의상 그룹 III이라고 한다. 도6 참조)을 선정함으로써 특정 인지질이 존재하는 세포내 부위를 관측할 수 있는 방안도 제시하였다.

이들 신호단백질은 세포 내의 각 소기관으로 이동할 때 그 작용과 관련하여, 세포 내에서의 위치가 각 소기관의 막, 내부, 외부까지 특이적으로 표시될 수 있도록 구성되었으며, 두 가지 이상의 단백질의 위치를 동시에 확인할 수 있도록 두 가지 이상 색깔의 형광단백질을 각각 사용하는 방안도 제시하고 있다.

이후 설명하는 본 발명의 실시예에서는 리포터 단백질을 구성함에 있어서 위치를 가시화하기 위한 형광단백질로 녹색 형광단백질(GFP, Davis와 Viestra, 1998) 및 적색 형광단백질(RFP)을 사용하였으나, 이 외에 다른 형광단백질을 사용하여 동일한 방법으로 본 발명에 따른 리포터 단백질을 구성할 수 있음은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 기술자라면 충분히 이해할 수 있을 것이다.

상기 형질전환된 세포에서 발현되는 리포터 단백질로부터 방출되는 특정 파장의 형광을 형광현미경을 사용하여 상기 세포가 살아있는 상태에서 계속하여 영상화할 수 있게 함으로써, 세포 내에서의 리포터 단백질의 발현과정 및 이동과정을 단계별로 세밀하게 가시화할 수 있었다.

특정 단백질이 발현되어 특정 소기관으로 이동하는 과정과 그 공간적 분포를 확인할 수 있게 됨에 따라, 이를 이용하여 특정 단백질의 세포내 이동을 저해 또는 촉진하는 화학물질을 효율적으로 검색할 수 있는 선별적 신약 검색 시스템을 구축할 수 있었다. 즉, 생물학적 활성을 가지고 있을 것으로 추정되는 화학물질을, 상기 본 발명에 따른 형질전환세포 내에서 상기와 같은 리포터 단백질이 발현되기 전, 후 또는 발현과 동시에 상기 형질전환세포에 처리하고, 이들이 단백질의 이동에 미치는 영향에 의하여 나타나는 세포내 단백질의 분포를, 그러한 화학물질을 처리하지 않은 대조군과 비교함으로써, 처리한 화학물질의 영향을 확인할 수 있음을 보여 주었다.

동일한 방법에 의하여, 리포터단백질의 형광신호의 증감에 따라 발현되는 정도를 비교할 수도 있으므로, 세포 내에서의 단백질의 전사과정(transcription) 또는 번역과정(translation)을 저해 또는 촉진하는 물질의 검색에도 사용할 수 있음을 보여 주었다.

또한, 상기 리포터단백질의 형광신호의 분포 및 형태에 따라, 처리한 화학물질에 의한 세포내 소기관의 변형, 손상, 파괴 등과 같은 형태 변화를 관찰할 수 있으므로, 이와 같은 세포내 기관의 변화를 유발하는 세포독성을 나타내는 물질들을 선별할 수 있음도 보여 주었다.

상기 본 발명에서 제시하는 선별적 신약 검색 시스템이 실질적으로 적용될 수 있음을 보이기 위하여, 이후 실시예에서 상세히 나타낸 바와 같이, 바필로마이신 A1, 워트마닌 및 브레펠딘 A 등의 기지의 세포내 기능 저해제를 사용하여 단백질의 세포내 이동과정과 발현과정이 저해되는 경우 및 세포내 소기관의 변형을 초래하는 세포독성이 나타나는 경우를 실질적으로 관측할 수 있음을 확인하였다. 또한 이후 실시예에서, 단백질의 이동 메카니즘에 따른 차이에 대하여 선별적으로 확인하기 위하여, 상기에서 정의한 그룹 I, II, III에 해당하는 신호단백질에 대하여 각각 실시하여 그 결과를 확인하였다.

본 발명의 구성 및 그 유용성은 첨부된 도면과 함께 본 발명의 바람직한 실시예에서 상세히 설명된다. 그와 같은 실시예는 본 발명을 설명하기 위한 것이며 발명을 한정하기 위한 것이 아니다. 본 발명에 속한 기술 분야에서 통상의 지식을 가진 자는 본원 명세서에 첨부된 도면, 발명의 상세한 설명 및 특허청구범위에 기재된 내용로부터 본 발명의 다른 목적 및 장점을 쉽게 인식할 수 있을 것이다.

(실시예)

< 실시예 1> 막을 통하여 소기관으로 이동되는 단백질들(그룹 I)을 발현시키기 위한 혼성 플라즈미드 제조

F1-H+-ATPase- (접속번호 D88374)의 트랜짓 펩티드(transit peptide)(서열번호 27)는 두 개의 특이 프라이머(CTTTAATCAATGGCAATG(서열번호 1)와 CCATGGCCTGAACTGCTCTAAGCTT(서열번호 2))를 이용하여 람다잽II(λZAPII) cDNA 라이브러리를 대상으로 중합효소 연쇄반응(polymerase chain reaction, PCR)에 의하여 증폭하여 녹색 형광단백질 (GFP)의 아미노 말단과 접합하여 F1-H+-ATPase-:RFP (Niwa et al., 1999)를 제조하였다. 연결된 융합 유전자는 pBluescript에 도입하여 35S 프로모터의 조절유전자의 영향에 들도록 F1-ATPase:RFP 혼성 플라스미드를 제조하였는데, 이후의 혼성 플라스미드를 형질전환하기 위한 과정에서도 이와 같은 방법을 실시하였다.

루비스코(rubisco; ribulose bisphosphate carboxylase) 복합단백질에 대한 리포터단백질을 발현시키기 위하여, 람다잽II cDNA 라이브러리를 대상으로 두개의 특이 프라이머 (CCTCAGTCACACAAAGAG(서열번호 3)와 ACTCGAGGGAATCGGTAAGGTCAG(서열번호 4))를 사용하여 PCR을 실시하여 루비스코의 스몰서브유닛(small subunit)의 트랜짓 펩티드(서열번호 29)를 증폭시켰다. PCR 생성물은 pBluescript에 서브클론하여 GFP 및 적색형광단백질(RFP)의 아미노 말단에 프레임이 맞도록 접합시켜 각각 RbcS:GFP 또는 RbcS:RFP를 제조하였다.

엽록소a/b-결합 단백질을 위하여 람다잽II cDNA 라이브러리를 대상으로 두개의 특이 프라이머(TAGAGAGAAACGATGGCG(서열번호 5) 및 GGATCCCGTTTGGGAGTGGAACTCC(서열번호 6))를 사용하여 PCR 증폭하여 제조한 유전자(서열번호 31)를 도입하여 Cab:GFP 혼성플라스미드를 제조하였다.

루비스코 액티바제(rubisco activase, RA)의 트랜짓 펩타이드(서열번호 33)는 두 개의 특이 프라이머(TCTAGAATGGCCGCCGCAGTTTCC(서열번호 7)와 GGATCCATCTGTCTCCATCGGTTTG(서열번호 8))를 이용하여 람다잽II cDNA 라이브러리를 대상으로 PCR 증폭하여, GFP 또는 RFP의 5' 말단과 접합시켜 RA:GFP 및 RA:RFP를 제조하였다.

완두콩의 OEP14와 유사한 애기장대의 외막 단백질 AtOEP7에 대한 유전자(서열번호 35)를 제조하기 위하여, AtOEP7 코딩 부위를 애기장대의 유전체 DNA를 대상으로 자연적으로 존재하는 종료 신호를 제거하도록 고안된 두 개의 특이 프라이머(OEP7-F: GACGACGACGCAGCGATG(서열번호 9)와 OEP7-R: GGATCCCCAAACCCTCTTTGGATGT(서열번호 10))를 이용하여 PCR 증폭하였다. GFP 또는 RFP의 5' 말단과 접합하여 각각 AtOEP7:GFP 및 AtOEP:RFP를 제조하였다.

핵 위치 신호(nuclear location signal; NLS)(서열번호 37)에 대한 혼성단백질 NLS:GFP는 전에 기술한 방법으로 제조하였다(Pih et al., 2000). NLS:RFP는 NLS:GFP에서 GFP의 코딩부분을 RFP로 치환하여 제조하였다.

프록시좀(peroxisome)에 대한 이동 모티브 SKL(serine, lysine, leucine)에 대한 리포터 단백질인 GFP:SKL을 제조하기 위하여, 326GFP(Davis and Viestra, 1998)를 템플레이트(template)로 두 개의 특이 프라이머(CCGTATGTTACATCACC(서열번호 11)와 TTATAGCTTTGATTTGTATAGTTCATCCAT(서열번호 12))를 사용하여 PCR 증폭하여 혼성 플라스미드를 제조하였다.

상기의 방법으로 제조한 혼성 플라스미드를 발현시킨 각 리포터단백질의 모식도를 도1에 표시하였다.

< 실시예 2> 프로토플래스트의 제조 및 혼성 플라스미드에 의한 세포의 형질전환

a) 프로토플래스트의 제조

온실 내 토양에서 배양된 3 - 4주 된 애기장대의 잎 조직(5 g)을 5 - 10 mm^2 의 크기로 새 면도날로 잘게 자른 후, 이를 효소용액(0.25% 마세로자임(Macerozyme) R - 10, 1.0% 셀룰라제(Cellulase) R - 10, 400 mM 만니톨(mannitol), 8 mM CaCl2, 5 mM Mes - KOH, pH 5.6) 50 ml를 이용하여 22℃에서 천천히 진탕(50 - 75 rpm)하면서 방치하였다. 프로토플래스트 현탁액은 100 μm 메시로 여과한 후 46 xg 에서 5분간 원심분리하여 회수하였다. 프로토플래스트는 W5용액 (154 mM NaCl, 125 mM CaCl2, 5 mM KCl, 5 mM 글루코즈(glucose), 1.5 mM Mes - KOH, pH 5.6) 5 - 10 ml로 재현탁하여, 21% 수크로즈 20 ml 위에 얹어 78 xg 에서 5분간 원심분리하였다. 손상되지 않은 온전한 프로토플래스트를 층간에서부터 회수하여 20ml의 W5용액에 담구었다. 프로토플래스트는 다시 55xg 에서 5분간 원심분리하고, 이를 20 ml의 W5용액에 다시 재현탁하여 얼음 위에 30분간 방치하였다.

b) 혼성 플라스미드 DNA의 분리 및 형질전환

혼성 플라스미드는 퀴아젠 컬럼 (Valencia, CA)을 이용하여 제조사의 추천 방법에 따라 분리하였다. DNA를 프로토플래스트로 삽입하기 위하여 프로토플래스트를 다시 46xg 에서 5분간 원심분리한 후, MaMg용액 (400 mM 만니톨(Mannitol), 15 mM MgCl2, 5 mM Mes - KOH, pH 5.6)에 5x106 프로토플래스트/ml 의 밀도로 재현탁하였다. 혼성 플라스미드 클론은 PEG 매개 형질전환법으로 애기장대의 프로토플래스트에 삽입시켰다(Kim 등, 2001; Jin 등, 2001). 약 20 - 50μg의 플라즈미드 DNA(2 μg/μℓ)를 프로토플래스트 현탁액 300μℓ와 혼합한 후, 325μℓ의 PEG 용액(400 mM 만니톨(Mannitol), 100 mM Ca(NO3)2, 40% PEG 4000)을 가하고 부드럽게 혼합하였다. 혼합액은 실온에서 30분간 방치하고 W5 용액 10 ml로 희석하였다. 프로토플래스트는 50 xg 에서 5분간 원심분리하여 희석하고, W5 용액 3 ml로 재현탁하여 22℃의 암실에 방치하였다.

< 실시예 3> 그룹 I 리포터 단백질들의 세포 내 발현 및 위치 관찰

실시예 1에서 제조한 혼성 플라스미드를 실시예 2의 방법으로 세포에 형질전환 후 시간 변화에 따른 리포터 단백질들의 발현을 형광현미경(독일Zeiss, Axioplan fluorescence microscope)과 냉각형 CCD 카메라를 사용하여 영상을 잡아 관찰하였다. 사용된 필터는 XF116(exciter: 474AF20, dichroic: 500DRLP, emitter: 510AF23), XF33/E(exciter: 535DF35, dichroic: 570DRLP, emitter: 605DF50), 및 XF137 (exciter: 540AF30, dichroic: 570DRLP, emitter: 585ALP)(Omega, Inc., Brattleboro, VT)이며, 각각 GFP, RFP 및 엽록소의 자체 형광을 관찰하기 위하여 사용되었다. 결과는 Adobe (Mountain View, CA) Photoshop 소프트웨어를 사용하여 분석하여 위색상으로 나타내었다.

a) 엽록체로 이동하는 리포터 단백질들의 위치

리포터단백질 AtOEP7:GFP의 녹색 형광은 엽록체 막의 외부에서 관찰되었다. 도 2의 a)에서 적색 형광은 엽록체 본래의 형광이므로 이를 필터로 제거한 결과가 도2의 b)에 나타나 있다. 본 결과는 이러한 엽록체 막으로의 이동 신호를 가진 신호단백질과 형광의 표지 단백질을 포함한 혼성단백질이 모두 엽록체 막으로 수송되었음을 나타낸다.

리포터단백질 RbcS:GFP, Cab:GFP 및 RA:GFP의 혼성 단백질의 녹색 형광이 나타나는 위치를 분석한 결과는 각각 도2의 c), d), 및 e)와 같았다. 상기 도면에 나타난 바와 같이, 루비스코 복합 단백질은 엽록체의 스트로마(stroma)에 위치하며, Cab:GFP는 엽록체에서 발광하였다. 이러한 결과는 RbcS, Cab 및 RA가 가지는 이동 신호와 형광의 표지 단백질을 포함한 혼성단백질이 모두 엽록체 내로 이동하였음을 나타낸다.

b) F1-H+-ATPase-:RFP 의 미토콘드리아로의 이동

리포터단백질 F1-H+-ATPase-:RFP에 의한 적색 형광이 미토콘드리아 내에서 관찰되었다(도2의 f)). 이 결과는 미토콘드리아 내로의 이동 신호를 가진 신호단백질과 형광단백질을 포함한 단백질이 모두 미토콘드리아 내로 이동하였음을 나타낸다.

c) GFP:SKL의 퍼록시좀으로의 수송

GFP:SKL 혼성단백질의 녹색형광이 나타나는 위치를 분석한 결과, 도2의 g)와 같이 GFP:SKL의 녹색 형광이 퍼록시좀으로 이동함이 관찰되었다. 본 결과는 이러한 퍼록시좀으로의 이동 신호 SKL(serine, lysine, leucine)와 형광의 표지 단백질을 포함한 혼성단백질이 모두 퍼록시좀으로 이동하였음을 나타낸다.

d) NLS:GFP의 핵으로의 수송

혼성단백질의 녹색형광이 나타나는 위치를 분석한 결과, 도2의 h)와 같이 NLS:GFP의 녹색 형광이 핵 내로 이동함이 관찰되었다. 본 결과는 이러한 핵 내로의 이동 신호와 형광의 표지 단백질을 포함한 혼성단백질이 핵 내로 수송되었음을 나타낸다.

< 실시예 4> 웨스턴 블랏 분석법으로 AtOEP7:GFP의 엽록체 막으로의 수송 확인

AtOEP7:GFP의 혼성 플라즈미드는 실시예 1과 같이 제조하였다. 제조된 혼성 플라즈미드를 실시예 2에서의 방법으로 형질전환하고, 24시간 방치하였다. 세포액 5 ml를 원심분리한 후 4℃의 마쇄용액(10mM EDTA, 50 mm HEPES-KOH, 0.33 M 소르비톨(sorbitol), 0.5 g/l BSA, 5 mM 소디움 아스코로베이트(sodium ascorbate)) 5 ml에 현탁시켜서, 호모게나이저(homogenizer)로 3초 간격으로 20분간 처리하여 전체 단백질 용액을 제조하였다. 전체 단백질 용액을 100,000xg로 원심분리하여 액상 부분과 막 부분을 각각 분리하였다. 각각의 시료를 SDS를 포함하는 7.5% 폴리아크릴아마이드 젤에서 전기영동하였으며, 전개된 젤을 PVDF 나일론 막에 붙여서 단백질을 이동시킨 후, 이를 폴리클로날 앤티-(polyclonal anti-)GFP 항체로 결합시켰다.

그 결과, 세포 내에서 발현된 신호단백질은 세포질에 존재하지 않고 모두 엽록체 엔벨롭 막으로 이동하였음을 보여주고 있다(도3). 이는 기존의 웨스턴 블랏 방법으로 확인하는 단백질의 분포를 본 발명에서 제시하는 방법으로 동일하게 확인할 수 있음을 나타낸다.

< 실시예 5> 세포질 내에서 엔도소말 트래피킹(endosomal trafficking) 과정을 통하여 소기관으로 이동되는 단백질들(그룹 II)을 발현시키기 위한 혼성 플라스미드의 제조

H+-ATPase의 코딩 부위의 전범위(애기장대의 AHA2)(서열번호 39)를 두개의 특이 프라이머 (GAGATGTCGAGTCTCGAA(서열번호 13) 및 CTCGAGCACAGTGTAGTGACTGG(서열번호 14))를 사용하여 위와 같은 방법으로 증폭하여(서열번호 33), GFP의 5' 말단과 연결하였다. 연결된 융합 유전자는 pUC 벡터에 도입하여 35S 프로모터의 조절유전자의 영향에 들도록 혼성 플라스미드를 제조하였는데, 이는 이후의 실시예에서도 같은 방법으로 실시하였다.

애기장대의 전체 cDNA에서 두개의 특이적인 프라이머 BIP5(5'-TACGCAAAGTTTCC-GAT-3'(서열번호 15))와 BIP3(5'-CTAGAGCTCATCGTGAGA-3'(서열번호 16))을 사용하여 샤프론 결합단백질(BiP) 유전자(접속번호 D82817)를 증폭시켰다. 이 유전자의 아미노 말단 부위(44개 아미노 잔기)(서열번호 41)와 카복시 말단 부위(44 아미노 잔기)(서열번호 43)를 각각 GFP와 RFP의 아미노 말단과 카복시 말단에 연결하여 BiP:GFP와 BiP:RFP 융합유전자를 제조하였다.

시알기 전달효소(sialtransferase, ST)의 cDNA(서열번호 45)는 람다잽II cDNA 라이브러리를 대상으로 두 개의 특이 프라이머(ATGATTCATACCAACTTGAAG(서열번호 17)와 GGATCCACAACGAATGTTCCGGAA(서열번호 18))를 이용하여 증폭하였다. GFP 또는 RFP를 ST의 카복시 말단에 프레임에 맞게 접합시켜 ST:GFP 및 ST:RFP를 제조하였다.

키티나제(Chitinase)에 대한 혼성단백질 Chi-n:RFP:Chi-c를 제조하기 위하여 종료 신호를 제외한 RFP 코딩부위를 포함하는 DNA 조각(서열번호 47 및 49)을 콩의 키티나제 cDNA(접속번호 M13968)의 Sma I 및 EcoRV 사이트에 접합하였다(서열번호 37 및 38).

액포 분류 수용체(vacuolar sorting receptor)로 작용하는 단백질 BP-80의 세포질 부위(cytoplasmic tail)를 제거한 클론 500을 발현시키기 위하여, 500:GFP의 혼성 플라즈미드는 종료 신호가 제거된 GFP코딩부위를 먼저 클론 500의 EcoRI 위치에 삽입시켜서 제조하였다(Kim 등, 2001).

BP-80의 세포질 부위가 TIP(tonoplast intrinsic protein)의 세포질 부위로 치환된 혼성클론 526을 위하여, 혼성클론 526의 EcoRI 위치에 GFP의 코딩부위를 삽입하여 526:GFP의 혼성플라즈미드를 제조하였다.

BP-80 단백질인 클론 491에 대한 리포터단백질을 발현시키기 위하여, 491:GFP 및 491:RFP는 각각 종료 신호가 제거된 GFP및 RFP의 5 말단에 클론 491을 삽입하여 혼성 플라즈미드를 구성하여 제조하였다.

500:GFP:KKXX는 다음과 같이 제조하였다. 종료 신호가 제거된 GFP코딩 부위를 먼저 클론500의 EcoRI 위치에 삽입시킨 후(Jiang and Rogers, 1998), 두개의 특이 프라이머 (GGATCCTCTAGAGGATCGATCCGG(서열번호 19)와 TTAGATGAGTTTCTTTTTCTCAAAGAAAGTTTTCAAAAGGAATCCCCCTCC(서열번호 20))를 이용하여 PCR증폭함으로써 KKXX를 500:GFP의 말단에 접합시켰다.

트랜스 골지 네트웍에서 저장전액포로 이동하는 애기장대의 t-SNARE의 일종인 AtVTI1a(Zheng 등, 1999)를 발현시키기 위하여, RFP의 카복시 말단에 AtVTI1a의 코딩 유전자(서열번호 51)를 결합시켜 RFP:AtVTl1a를 제조하고, AtVTl1a의 카르복시 말단에 GFP의 코딩 부위를 접합시켜 혼성 플라즈미드 AtVTI1a:GFP의 재조합 유전자를 구성하여 혼성 플라즈미드를 제조하였다.

스포라민의 발현을 위하여 스포라민 B의 유전자(서열번호 53) 카복시 말단에 GFP를 연결하여 SPO:GFP 혼성단백질에 대한 혼성 플라즈미드를 구성하였다.

상기의 방법으로 제조한 혼성 플라스미드를 발현시킨 각 리포터 단백질의 모식도는 도4에 표시하였다.

< 실시예 6> 그룹 II 리포터 단백질들의 세포내 발현 및 위치 관찰

H+-ATPase:GFP, ST:GFP, BiP:RFP, 526:GFP, Chi-n:RFP:Chi-c, 500:GFP:KKXX의 혼성 플라즈미드의 제조는 실시예 4에서와 같이 실시하였으며, 이들을 실시예 2의 방법으로 세포에 형질전환 후 시간 변화에 따른 리포터 단백질들의 발현을 실시예 3에 설명한 바와 같이 형광 현미경으로 관찰한 결과들 중 일부는 다음과 같다.

리포터단백질 H+-ATPase:GFP는 플라즈마 막에 형광이 나타났으며(도5의 a)), ST:GFP는 골지체에 형광을 나타내었는데(도5의 b)), 이들 결과 중 적색 형광은 엽록체의 자연 형광에 의한 것이다. 리포터단백질 BiP:RFP와 500:GFP:KKXX에 의하여 나타나는 형광의 위치는 소포체였는데, 이들은 각각 소포체 루멘(lumen)과 막을 표시하였다(도5의 c) 및 d)). 526:GFP에 의해서는 저장액포의 막이(도5의 e)), Chi-n:RFP:Chi-c에 의하여 저장액포(도5의 f))가 각각 형광을 나타내었다. 신호단백질로 키티나제(chitinase)를 사용할 때, 카르복시 부분을 연결하지 않은 Chi-n:RFP는 저장액포에 도달하지 못하고 소포체에 점의 형태로 존재하였다(도5의 g)). SPO:GFP에 의한 형광은 용해성 액포 전체에 고루 분포하고 있었다(도5의 h)).

< 실시예 7> 인지질에 특이적인 단백질들(그룹 III)을 발현시키기 위한 혼성플라즈미드 제조

GFP:EBD 혼성유전자를 제조하기 위하여 사람의 초기 엔도소말 항원(early endosome antigen) 1 (EEA1)의 C-말단에 해당하는 코딩부위(아미노산 잔기 1257에서 1411) 유전자(서열번호 55)를 두 개의 프라이머 5'-GAATTCGTGGCAATCTAGTCAACGG-3'(서열번호 21) 및 5'-CTAATGTTAGTGTAATATTAC-3'(서열번호 22)와 함께 중합효소 연쇄반응(PCR)으로 증폭하였으며, 이를 종료신호가 제거된 GFP 코딩 유전자의 C-말단에 연결하였다. 연결된 융합유전자는 pUC 벡터에 도입하여 35S 프로모터의 조절유전자의 영향에 들도록 혼성 플라스미드를 제조하였는데, 이는 이후의 실시예에서도 같은 방법으로 실시하였다.

EBD 변이체의 혼성 플라즈미드 구성을 위하여 GFP:EBDC1358S는 1358 위치에 아미노산이 세린(serine)으로 치환되도록 프라이머를 구성하여 제조하였다.

애기장대의 PH 구역(Pleckstrin homology domain)(서열번호 57)에 대한 혼성단백질 GFP:AtPH는 두개의 프라이머 5'-CCCGGGAAATGGAGAGTATGTGGCGA-3'(서열번호 23)와 5'-TAATCACCGCCTGTGATCATA-3'(서열번호 24)를 사용하였고, PH 구역을 포함하는 FAPP(서열번호 59)의 혼성단백질 GFP:FAPP는 두개의 프라이머 5'-CTCGAGATGGAGGGGGTTCTGTACAAG-3'(서열번호 25)와 5'-TCACGCTTTGGAGCTCCCAAGGGC-3'(서열번호 26)을 사용하여 PCR 증폭하였으며, PH:GFP는 Kost B 등의 방법(1998)으로 제조하였다. PH 유전자(서열번호 61).

상기의 방법으로 제조한 혼성 플라스미드를 발현시킨 각 리포터 단백질의 모식도는 도6에 표시하였다.

< 실시예 8> 그룹 III 리포터 단백질들의 세포내 발현 및 위치 관찰

실시예 7의 방법으로 제조한 혼성 플라스미드 GFP:EBD, GFP:AtPH, GFP:FAPP, GFP:PH를 실시예 2의 방법으로 세포에 형질전환 후, 시간에 따른 리포터 단백질들의 발현을 실시예 3에서 설명한 바와 같이 형광현미경으로 관찰한 결과들 중 일부는 다음과 같다.

리포터단백질 GFP:EBD와 GFP:AtPH는 각 액포들의 외벽에 형광을 나타내어 이들이 3-인산 포스파티딜이노시톨(phosphatidylinositol 3-phosphate; PI(3)P)를 포함하는 위치로 파악되었으며, GFP:FAPP와 GFP:PH는 각각 4-인산 포스파티딜이노시톨 (phosphatidylinositol 4-phosphate; PI4P)과 4,5-이인산 포스파티딜이노시톨 (phosphatidylinositol 4,5-diphosphate; PI(4,5)P2)이 분포하는 플라즈마 막에 형광을 나타내었다. 이는 인지질이 분포하는 곳으로 단백질을 이동시키기 위하여, 인지질에 특이적인 단백질의 전체 또는 일부를 신호단백질로 사용하는 것이 가능함을 보여주고 있다(도 7 참조).

< 실시예 9> 워트마닌(wortmannin)이 RbcS:GFP의 수송에 미치는 영향

a) 혼성 플라스미드 제조, 형질전환 및 혼성 단백질의 발현

RbcS:GFP의 혼성 플라즈미드는 실시예 1과 같이 제조하였다. 혼성 플라즈미드의 분리, 프로토플래스트의 제조, 및 형질전환은 실시예 2와 같이 실시하였다. 프로토플래스트 현탁액에 워트마닌을 5 μg/ml 농도로 처리한 후 형질전환하고 암실에서 배양하였다. 혼성단백질의 발현 관찰은 실시예 3과 같이 실시하였다.

b) 화학물질 워트마닌이 혼성단백질의 세포내 이동에 미치는 영향

3-인산화 포스파티딜이노시톨(PI3P)과 4-인산화 포스파티딜이노시톨(PI4P)에 대하여 특이적으로 저해하는 워트마닌(Ui 등, 1995)이 RbcS:GFP의 수송에 미치는 영향을 검사하였다. 혼성단백질의 녹색형광이 나타나는 위치를, 화학물질을 처리하지 않은 정상 대조군과 워트마닌을 처리하였을 경우와 비교 분석하였다. 도8에 나타난 바와 같이, 정상 대조군에서는 RbcS:GFP의 녹색 형광이 예상대로 RbcS:GFP가 수송되어야 하는 엽록체 내로 이동된 것이 관찰되는 반면에(도8의 a)), 워트마닌 존재 시에는 녹색 형광이 엽록체 내로 수송되지 못하고 작은 반점 혹은 엉킨 형태로 관찰되었다(도8의 b) 및 c)). 본 결과는 워트마닌이 엽록체로 이동되어야 하는 단백질의 세포질에서 엽록체로의 이동을 저해함을 나타낸다.

< 실시예 10> 액포형 H+-ATPase의 저해제인 바필로마이신 A1(bafilomycin A1, BafA1)의 500:GFP:KKXX의 역행성의 트래피킹(retrograde trafficking)에 미치는 영향

a) 혼성 플라스미드 제조, 형질전환 및 혼성단백질의 발현

500:GFP:KKXX의 혼성 플라즈미드는 실시예 5와 같이 제조하였고, 혼성 플라즈미드의 분리, 프로토플래스트의 제조, 및 형질전환은 실시예 2와 같이 실시하였다. 프로토플래스트 현탁액에 바필로마이신 A1를 5 μg/ml 농도로 처리한 후 형질전환하고 암실에서 배양하였으며, 혼성단백질의 발현관찰은 실시예 3와 같이 실시하였다.

b) 화학물질 바필로마이신 A1(BafA1)이 혼성단백질의 세포내 이동에 미치는 영향

혼성단백질의 녹색형광이 나타나는 위치를 화학물질을 처리하지 않은 정상 대조군과 BafA1를 처리하였을 경우와 비교 분석하였다. 도9에서 나타낸 바와 같이 정상 대조군에서는 500:GFP:KKXX 의 녹색 형광이 수많은 네트웍 형태로 관찰되는 반면(도9의 a)와 b)), BafA1 존재 시에는 녹색 형광이 플라즈마 막(도9의 c))과 액포막(도9의 d))에서 고리 모양으로 나타났다. 본 결과는 BafA1이 애기장대의 프로토플래스트에서 500:GFP:KKXX의 역행성의 트래피킹이 저해되고, 표지 형광단백질이 플라즈마 막 또는 액포막으로 이동됨을 나타낸다.

< 실시예 11> 두 가지 리포터단백질의 동시 발현시 특이적인 저해제에 의한 변화

a) 혼성 플라스미드 제조, 형질전환 및 혼성단백질의 발현

각각 소포체와 골지체에 특이적인 리포터단백질 BiP:GFP와 ST:RFP의 혼성 플라즈미드는 실시예 5와 같이 제조하였다. 혼성 플라즈미드의 분리, 프로토플래스트의 제조, 및 형질전환은 실시예 2와 같이 실시하였다. 프로토플래스트 현탁액에 브레펠딘 A를 5 μg/ml 농도로 처리한 후, 형질전환하고 암실에서 배양하였다. 혼성단백질의 발현관찰은 실시예 3과 같이 실시하였다.

b) 두 가지 리포터 단백질이 내는 형광의 분포

브레펠딘 A(BFA)는 동물 세포에서 ADP-리보실레이션 인자(ribosylation factors)(Arfs)의 저해제로 알려져 있는 화학물질이다(Morinaga 등, 1999). 정상 대조군에서는 BiP:GFP에 의한 녹색 형광이 소포체에(도10의 a)), 그리고 ST:RFP에 의한 적색 형광이 골지체에(도10의 b)) 각각 나타났으나, BFA를 처리하였을 때, 골지체의 파괴로 인하여 ST가 골지체에 존재하지 못하고 소포체로 이동하였음을 관찰할 수 있었다(도10의 e)). 따라서 대조군과 화학물질을 처리한 결과는 형광 현미경 하에서 뚜렷한 차이를 나타내고(도10의 c)와 f)), 이는 두 가지 이상의 단백질 이동에 대한 결과를 두 가지 이상의 형광을 사용함으로써 동시에 측정할 수 있음을 보이고 있다. ST:RFP의 결과만으로 보면, 이 실험 결과는 골지체 구조 파괴 과정을 보여주는 결과이기도 하다.

< 실시예 12> 화학물질 브레펠딘 A가 소포체의 형성 및 구조 유지에 미치는 영향

a) 혼성 플라스미드 제조, 형질전환 및 혼성단백질의 발현

BiP:RFP의 혼성 플라즈미드는 실시예 5와 같이 제조하였다. 혼성 플라스미드의 분리, 프로토플래스트의 제조, 및 형질전환은 실시예 2와 같이 실시하였다. 프로토플래스트 현탁액에 브레펠딘 A(BFA)를 5 μg/ml 농도로 처리한 후, 형질전환하고 암실에서 배양하였다. 혼성단백질의 발현 관찰은 실시예 3과 같이 실시하였다.

b) 화학물질의 혼성단백질의 세포내 이동에 미치는 영향

혼성단백질의 적색형광이 나타나는 위치를 화학물질을 처리하지 않은 정상 대조군과 BFA를 처리하였을 경우와 비교분석하였다. 도11의 a)에서와 같이, 정상 대조군에서는 BiP:RFP의 적색 형광이 소포체의 구조를 따라 분포하고 있으나, BFA를 첨가하였을 때, 소포체의 구조가 파괴되고 있음을 가시적으로 보여준다(도11의 c)). 상기 각각의 동일한 시료에 대한 결과를, 리포터단백질에 의한 형광을 사용하지 않고 광학 현미경으로 관찰한 결과(도 11의 b)와 d))와 비교하면, 각 소기관에 특이적인 결합에 의한 리포터 단백질에 의한 관찰의 효과를 알 수 있다.

< 실시예 13> 인지질에 특이적인 단백질의 이동을 저해하는 실험

a) 혼성 플라스미드 제조, 형질전환 및 혼성단백질의 발현

리포터단백질 GFP:EBD와 GFP:EBDC1358S는 실시예 7과 같은 방법으로 구성한 혼성 플라스미드로부터 제조하여 사용하였다. 프로트플라스트에 구축한 재조합 DNA 운반체를 실시예2의 방법으로 도입하고, 워트마닌은 1.0 μg/ml, 포스파티딜이노시톨 3-키나제에 특이적 저해제인 2-(4-모폴리닐)-8-페닐-4H-1-벤조피란-4-온{2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-on}(LY294002, Vlahos 등, 1994)는 10 μg/ml의 농도로 각각 처리한 후, 22℃에서 배양하면서 시간별로 형광을 측정하였다.

b) 화학물질의 혼성단백질의 세포내 분포에 미치는 영향

약물을 처리하지 않은 세포에서는 EBD가 결합하는 인지질 PI(3)P의 분포를 따라서 각 엔도솜(endosome)에서 녹색의 형광이 나타나지만, 저해제 워트마닌과 LY294002를 처리한 세포에서는 세포질 전체에 형광이 고루 분포함을 볼 수 있었다(도12의 a)). 한편, EBD의 1358번째 아미노산에 변이가 생긴 GFP:EBDC1358S를 대상으로 동일한 실험을 실시하였을 때는, 세포 내의 형광 분포에 영향을 주지 못하였다.

< 실시예 14> 단백질의 발현 저해제 시클로헥시미드(cycloheximide)에 의한 발현의 변화

a) 혼성 플라스미드 제조, 형질전환 및 혼성단백질의 발현

RA:GFP의 혼성 플라즈미드는 실시예 1과 같이 제조하였다. 혼성 플라스미드의 분리, 프로토플래스트의 제조, 및 형질전환은 실시예 2와 같이 실시하였다. 프로토플래스트 현탁액에 시클로헥시미드(cycloheximide)를 5 μg/ml 농도로 처리한 후 형질전환하고, 22℃에서 암실에서 배양하였다. 혼성단백질의 발현 관찰은 실시예 3과 같이 실시하였다.

b) 화학물질의 혼성단백질의 발현에 미치는 영향

정상 대조군(도13의 b))에 비하여 시클로헥시미드를 처리한 세포에서의 녹색 형광의 강도가 감소하였다(도13의 a)). 이는 리포터 단백질의 발현이 저해되어 나타난 현상으로, 화학물질이 단백질의 전이 및 번역 과정에 미치는 영향을 가시화할 수 있음을 나타낸다. 도13의 c)와 d)의 적색 형광은 세포 내에 있는 엽록체에 의한 자연 형광으로서, 세포의 성장 및 대사 상태를 상대적으로 비교하기 위하여 측정하였다.

이상 설명한 본 발명은 전술한 실시예 및 첨부된 도면에 의해 한정되는 것이 아니다. 본 발명은 본원 명세서에 기재된 기술적 사상 및 본원 청구범위에 기재된 발명의 기본 구성을 벗어나지 않는 범위 내에서 여러 가지 치환, 변형 및 수정이 가능하며, 그와 같은 치환, 변경 및 수정은 본 발명의 범위에 속한다.

발명의 효과

본 발명을 통하여 세포 내에서의 특정 단백질의 이동과 공간적 분포의 형성에 관련된 특이적 특성을 선별적으로 확인하는 방법 및 이러한 특정 단백질의 세포 내에서의 특이적 특성에 영향을 미치는 화학물질을 검색하는 선별적 신약 검색 방법을 체계화한 검출시스템을 구현하였다.

다양한 세포내 소기관으로 이동되는 특성을 가진 다수의 리포터단백질을 각각 발현할 수 있는 형질전환세포를 준비하는 구체화된 방법들을 제공함으로써 그 적용 대상을 확장하였으며, 이에 따라 상기 선별적 세포내 과정 관측법 및 이에 근거한 선별적 신약 검색법을 보다 체계적으로 종합하여 활용할 수 있는 기반을 마련하였다.

또한, 몇 가지 기지 약물들을 사용하여 이러한 약물들이 세포내 과정에 미치는 다양한 영향들의 구체적 예를 시험함으로써, 약물의 세부적 영향을 구분하여 판정할 수 있는 방안을 마련하였다. 보다 구체적으로, (1) 특정 단백질이 발현되어 특정 소기관으로 이동하는 과정과 그 공간적 분포에 미치는 약물의 영향을 관측함으로써, 특정 단백질의 세포내 이동을 저해 또는 촉진하는 화학물질을 선별적으로 검색할 수 있는 방안을 제시하였으며; (2) 리포터단백질의 발현에 미치는 약물의 영향을 형광신호의 세기 변화로부터 관측함으로써, 세포내에서의 단백질의 전사과정(transcription) 또는 번역과정(translation)을 저해 또는 촉진하는 물질을 선별적으로 검색할 수 있는 방안을 제시하였고; (3) 세포내 소기관의 변형, 손상, 파괴 등을 유발하는 약물의 영향을 관측함으로써, 세포독성을 유발하는 물질을 선별적으로 검색할 수 있는 방안을 마련하였다.

< 참고 문헌>

(특 허)

Harpold, M. M., and Brust, P.US 5,401,629. ASSAY METHOD AND SOMPOSITIONS USEFUL FOR MEASURING THE TRANSDUCTION OF AN INTRACELLULAR SIGNAL.

Thastrup, O., Tullin, S., Poulsen, L. K., and Bjorn, S. P.WO96/23898. A METHOD OF DETECTING BIOLOGICALLY ACTIVE SUBSTANCES

Ward, W., and Chalfie, M.WO95/21191. BIOLUMINESCENT INDICATOR BASED UPON THE EXPRESSION OF A GENE FOR A MODIFIED GREEN-FLUORESCENT PROTEIN.

Chalfie, M. and Prasher, D. US 5,491,084.USES OF GREEN-FLUORESCENT PROTEIN

Evans, K.WO98/21355. MUTANTS OF GREEN FLUORESCENT PROTEIN.

(논 문)

Davis, S. J., and Viestra, R. D.(1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant Mol. Biol.36, 521-528.

Dowler, S., Currie, R. A., Campbell, D. G., Deak, M., Kular, G., Downes, C. P., and Alessi, D. R.(2000). Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. Biochem J.351(Pt 1), 19-31.

Gillooly, D. J., Morrow, I. C., Lindsay, M., Gould, R., Bryant, N. J., Gaullier, J - M., Parton, R. G., and Stenmark, H.(2000). Localization of phosphatidylinositol 3 - phosphate in yeast and mammalian cells. EMBO J. 19, 4577 - 4588.

Goldfarb, D. S., Gariepy, J., Schoolnik, G., and Konberg, R. D.(1986). Synthetic peptides as nuclear localization signals. Nature322, 641 - 644.

Hu and Cheng(1995). Expression of aequorea green fluorescent protein in plant cells. FEBS Lett.369, 331 - 334.

Jiang. L., and Rogers, J. C.(1998). Integral membrane protein sorting to vacuoles in plant cells: evidence for two pathways. J. Cell Biol.143, 1183 - 1199.

Jin, J. B., Kim, T. A., Kim, S. J., Lee, S. H., Kim, D. H., Cheong, G. - W., and Hwang, I.(2001). A new Dynamin - like protein, ADL6, is involved in trafficking from the trans - Golgi network to the central vacuole in arabidopsis. Plant Cell (in press)

Kaether and Gerdes(1995) Visualization of protein transport along the secretory pathway using green fluorescent protein. FEBS Lett.369, 267 - 271.

Kim, D. H., Eu, Y. J., Yoo, C. M., Kim, Y. W., Pih, K. T., Jin, J. B., Kim, S. J., Stenmark, H., and Hwang, I.(2001). Trafficking of Phosphatidylinositol 3 - Phosphate from the trans - Golgi Network to the Lumen of the Central Vacuole in Plant Cells. Plant Cell13, 287 - 301.

Kost, B., Spielhofer, P., and Chua, N. H.(1998). A GFP - mouse talin fusion protein labels plant actin filaments in vivo and visualizes the actin cytoskeleton in growing pollen tubes. Plant J.16, 393 - 401.

Morinaga, N., Adamik, R., Moss, J., and Vaughan, M. (1999). Brefeldin A inhibited activity of the sec7 domain of p200, a mammalian guanine nucleotide - exchange protein for ADP - ribosylation factors. J. Biol. Chem.274, 17417 - 17423.

Morise et al.(1974). Intermolecular energy transfer in the bioluminescent system of aequorea. Biochem.13, 2656 - 2662.

Niwa, Y., Hirano, T., Yoshimoto, K., Shimizu, M., and Kobayashi, H.(1999). Non - invasive quantitative detection and applications of non - toxic, S65T - type green fluorescent protein in living plants. Plant J.18, 455 - 463

Pih, K. T., Yi, M. J., Liang, Y. S., Shin, B. J., Cho, M. J., Hwang, I., and Son, D.(2000). Molecular clonging and targeting of a fibrillarin homolog from Arabidopsis. Plant Physiol.123, 51 - 58.

Stauffer, T. P., Ahn, S. and Meyer, T.(1988). Receptor - induced transient reduction in plasma membrane PtdIns(4,5)P2 concentration monitored in living cells. Curr. Biol.8, 343 - 346.

Ui, M., Okada, T., Hazeki, K., and Haxeki, O.(1995). Wormannin as a unique probe for an intracellular signaling protein, phosphoinositide 3 - kinase. TIBS20, 303 - 307.

Vlahos, C., Matter, W. F., Jui, K., and Brown, R. F.(1994). A specific inhibitor of phosphatidylinositol 3 - kinase, 2 - (4 - morpholinyl) - 8 - phenyl - 4H - 1 - benzopyran - 4 - one (LY294002). J. Biol. Chem.269, 5241 - 5248.

Weiss, E. L., Bishop, A. C., Shokat, K. M., and Drubin, D. G.(2000). Chemical genetic analysis of the budding-yeast p21-activated kinase cla4p. Nat. Cell Biol.2, 677-685.

Zheng, H., von Mollard, G. F., Kovaleva, V., Stevens, T. H., and Raikhel, N. V.(1999). The plant vesicle-associated SNARE AtVTI1a likely mediates vesicle transport from the trans-Golgi network t the prevacuolar compartment. Mol. Biol. Cell10, 2251-2264.

(57) 청구의 범위

청구항 1.

세포 내에서 특정 단백질의 이동과 공간적 분포에 관한 특이적 특성을 확인하는 방법으로서, 상기 방법은,

(a) 세포내 특정 소기관으로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩(coding)하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자를 준비하는 단계;

(b) 상기 재조합 유전자가 세포 내에서 발현될 수 있도록 프로모터(promoter) 부분과 터미네이터(terminator) 부분을 포함하여 구성된 혼성 플라스미드를 준비하는 단계;

(c) 상기 (b)단계에서 준비한 혼성 플라스미드 1종 또는 2종 이상을 세포 내에 주입하여 형질전환세포를 준비하는 단계;

(d) 상기 형질전환세포에서 신호단백질과 형광단백질이 연결된 리포터 단백질이 발현되도록 하는 단계; 및

(e) 상기 형질전환세포 내에서 발현되는 리포터 단백질의 분포를 발현 중, 이동 중, 또는 그 후에 형광영상 측정을 통하여 관측하는 단계를 포함하여 구성되는 방법.

청구항 2.

식물 세포 내에서 특정 단백질의 이동과 공간적 분포에 관한 특이적 특성을 확인하는 방법으로서, 상기 방법은,

(a) 식물 세포내 특정 소기관으로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자를 준비하는 단계;

(b) 상기 재조합 유전자가 세포 내에서 발현될 수 있도록 프로모터(promoter) 부분과 터미네이터(terminator) 부분을 포함하여 구성된 혼성 플라스미드를 준비하는 단계;

(c) 상기 (b)단계에서 준비한 혼성 플라스미드 1종 또는 2종 이상을 세포벽이 제거된 식물세포의 프로토플래스트(protoplast) 내에 주입하여 형질전환세포를 준비하는 단계;

(d) 상기 형질전환세포에서 신호단백질과 형광단백질이 연결된 리포터 단백질이 발현되도록 하는 단계; 및

(e) 상기 형질전환세포 내에서 발현되는 리포터 단백질의 분포를 발현 중, 이동 중, 또는 그 후에 형광영상 측정을 통하여 관측하는 단계를 포함하여 구성되는 방법.

청구항 3.

제1항에 있어서, 상기 세포가 진핵 세포인 것을 특징으로 하는 방법.

청구항 4.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 형광단백질이 녹색 형광단백질(Green Fluorescent Protein; GFP), 적색 형광단백질(Red Fluorescent Protein; RFP), 이들의 변형체 및 이들의 유도체로 구성되는 군으로부터 선택되는 것임을 특징으로 하는 방법.

청구항 5.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 리포터단백질이 이동하는 세포내 특정 소기관이 핵, 미토콘드리아, 및 퍼록시좀(peroxisome)으로 구성되는 군으로부터 선택되는 소기관인 것을 특징으로 하는 방법.

청구항 6.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 재조합 유전자가 포함하고 있는 신호단백질 코딩부분이 NLS(핵위치 신호 서열), SKL(퍼록시좀 이동 모티브), 및 F1- H+-ATPase로 구성되는 군으로부터 선택되는 단백질의 전체 또는 위치이동 신호부분을 포함하는 그 단백질의 일부를 코딩하고 있는 것을 특징으로 하는 방법.

청구항 7.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 리포터단백질이 이동하는 세포내 특정 소기관이 세포막, 소포체, 골지체, 리보좀, 리소좀, 세포골격 (cytoskeleton), 중심체, 용해성 액포(lytic vacuole), 및 전액포(prevacuolar compartment)로 구성되는 군으로부터 선택되는 소기관인 것을 특징으로 하는 방법.

청구항 8.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 재조합 유전자가 포함하고 있는 신호단백질 코딩부분이 H+-ATPase, BiP(샤프론 결합단백질), ST(시알릴트랜스퍼라제), 클론 491, AtVTI1a, 및 SPO(스포라민)으로 구성되는 군으로부터 선택되는 단백질의 전체 또는 위치이동 신호부분을 포함하는 그 단백질의 일부를 코딩하고 있는 것을 특징으로 하는 방법.

청구항 9.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 재조합 유전자가 포함하고 있는 신호단백질 코딩부분이 인지질에 특이적으로 결합하는 단백질을 코딩하고 있는 것을 특징으로 하는 방법.

청구항 10.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 재조합 유전자가 포함하고 있는 신호단백질 코딩부분이 인지질에 특이적으로 결합하는 특성을 가진 EBD 및 PH로 구성되는 군으로부터 선택되는 단백질의 전체 또는 위치이동 신호부분을 포함하는 그 단백질의 일부를 코딩하고 있는 것을 특징으로 하는 방법.

청구항 11.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 형광영상을 측정하기 위하여 특정 형광 파장을 선별적으로 통과시키는 광학 필터가 장착된 형광현미경을 사용하여 1개 또는 2개 이상의 형광 파장에 대한 형광영상을 각각 얻는 것을 특징으로 하는 방법.

청구항 12.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 혼성 플라즈미드를 세포 내에 주입하는 방법이, PEG(polyethylene glycol), 인산칼슘, 및 DEAE-덱스트란(DEAE-Dextran)으로 구성된 군으로부터 선택되는 화학물질을 이용하는 방법, 양이온성 지질을 이용한 라이포펙션(lipofection) 방법, 마이크로인젝션 방법, 일렉트로포레이션 방법 및 일렉트로퓨전 방법으로 구성되는 군으로부터 선택되는 방법인 것을 특징으로 하는 방법.

청구항 13.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 혼성 플라즈미드를 세포 내에 주입하기 위하여, 5% 내지 40%의 PEG를 사용하는 것을 특징으로 하는 방법.

청구항 14.

제1항 내지 제3항 중 어느 하나의 항에 있어서, 상기 리포터 단백질의 발현 과정, 이동 과정 또는 이동 결과의 관측을 5분 내지 80시간의 범위 내에서 배양 후 또는 배양 중에 실시하는 것을 특징으로 하는 방법.

청구항 15.

세포내 특정 소기관으로의 위치 이동을 제어하는 부분을 포함하는 신호단백질을 코딩(coding)하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자로서, 상기 신호단백질을 코딩하는 유전자가 H+-ATPase, BiP(샤프론 결합단백질), 클론 491, AtVTI1a, SPO(스포라민), 및 ST(시알기 전달효소)로 구성되는 군으로부터 선택되는 단백질의 전체 또는 위치이동 신호부분을 포함하는 그 단백질의 일부를 코딩하고 있는 것을 특징으로 하는 재조합 유전자.

청구항 16.

세포내 인지질로의 위치이동을 제어하는 부분을 포함하는 신호단백질을 코딩하는 유전자와 형광을 나타내는 형광단백질을 코딩하는 유전자가 각각 하나 또는 둘 이상 연결된 재조합 유전자로서, 상기 신호단백질을 코딩하는 유전자가 EBD를 코딩하는 것을 특징으로 하는 재조합 유전자.

청구항 17.

제15항 또는 제16항에 있어서, 상기 형광단백질을 코딩하는 유전자가 녹색 형광단백질(Green Fluorescent Protein; GFP), 적색 형광단백질(Red Fluorescent Protein; RFP), 이들의 변형체, 및 이들의 유도체로 구성되는 군으로부터 선택되는 하나 또는 둘 이상을 코딩하고 있는 것을 특징으로 하는 재조합 유전자.

청구항 18.

제15항에 기술된 재조합 유전자로부터 발현된 리포터단백질이 이동하는 세포내 특정 소기관이 세포막, 소포체, 골지체, 용해성 액포(lytic vacuole), 및 전액포(prevacuolar compartment)로 구성되는 군으로부터 선택되는 세포내 소기관인 것을 특징으로 하는 재조합 유전자.

청구항 19.

제15항 내지 제17항 중 어느 한 항에 기술된 재조합 유전자에 프로모터 부분과 터미네이터 부분을 연결하여 세포 내에서 리포터단백질의 발현이 가능하게 만든 혼성 플라즈미드.

청구항 20.

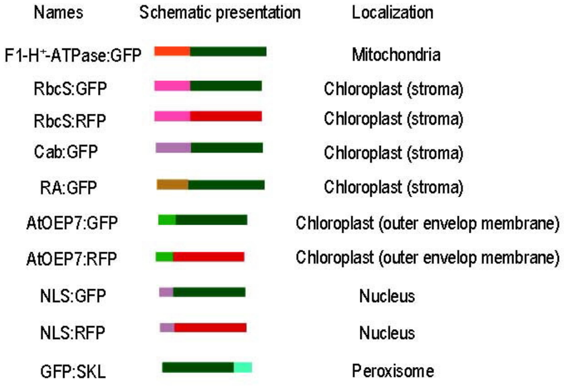
제15항 내지 제17항 중 어느 한 항에 기술된 재조합 유전자로부터 발현된 혼성단백질.

청구항 21.

제19항에 있어서, 상기 프로모터는 35S 프로모터인 것을 특징으로 하는 혼성 플라스미드.

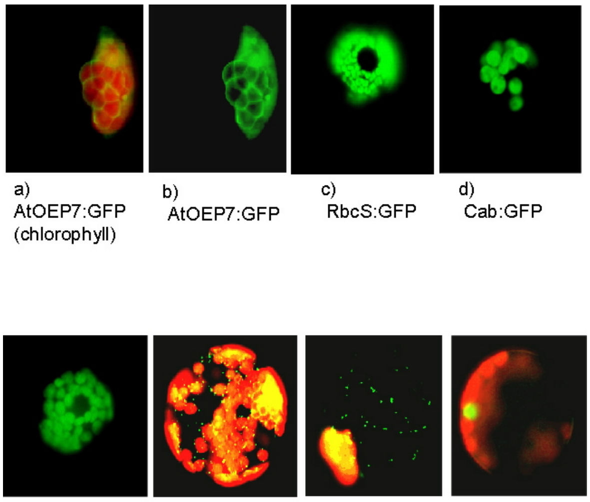
도면

도면 1



Names	Schematic presentation	Localization
F1-H$^+$-ATPase:GFP		Mitochondria
RbcS:GFP		Chloroplast (stroma)
RbcS:RFP		Chloroplast (stroma)
Cab:GFP		Chloroplast (stroma)
RA:GFP		Chloroplast (stroma)
AtOEP7:GFP		Chloroplast (outer envelop membrane)
AtOEP7:RFP		Chloroplast (outer envelop membrane)
NLS:GFP		Nucleus
NLS:RFP		Nucleus
GFP:SKL		Peroxisome

도면 2



a) AtOEP7:GFP (chlorophyll)

b) AtOEP7:GFP

c) RbcS:GFP

d) Cab:GFP

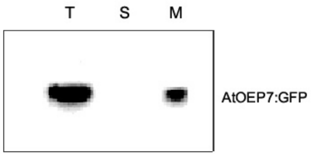
e) RA:GFP

f) F1-H$^+$-ATPase:GFP

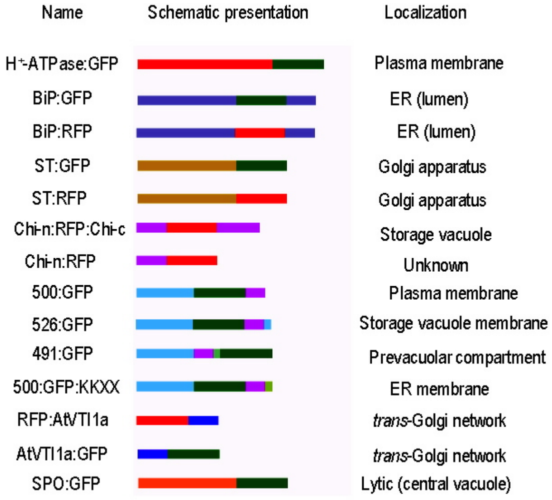
g) GFP:SKL

h) NLS:GFP

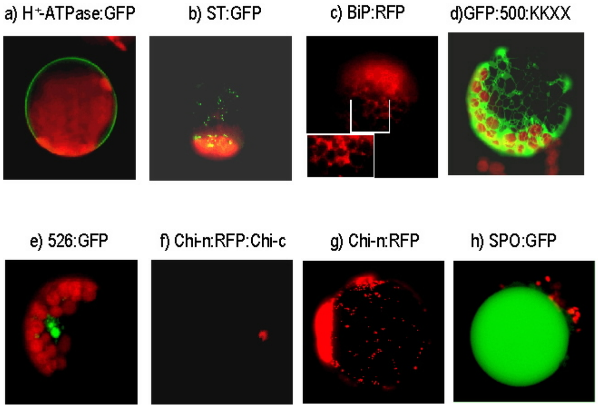
도면 3


T
S
M
AtOEP7:GFP

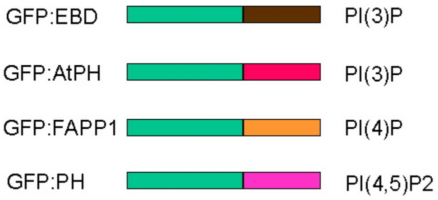
도면 4


Name
Schematic presentation
Localization
H+-ATPase:GFP
Plasma membrane
BiP:GFP
ER (lumen)
BiP:RFP
ER (lumen)
ST:GFP
Golgi apparatus
ST:RFP
Golgi apparatus
Chi-n:RFP:Chi-c
Storage vacuole
Chi-n:RFP
Unknown
500:GFP
Plasma membrane
526:GFP
Storage vacuole membrane
491:GFP
Prevacuolar compartment
500:GFP:KKXX
ER membrane
RFP:AtVTI1a
trans-Golgi network
AtVTI1a:GFP
trans-Golgi network
SPO:GFP
Lytic (central vacuole)

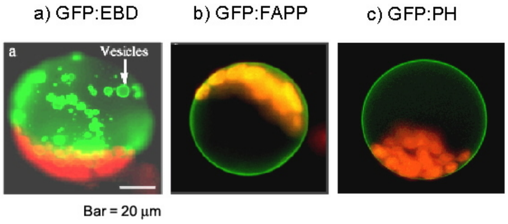
도면 5


a) H+-ATPase:GFP
b) ST:GFP
c) BiP:RFP
d)GFP:500:KKXX
e) 526:GFP
f) Chi-n:RFP:Chi-c
g) Chi-n:RFP
h) SPO:GFP

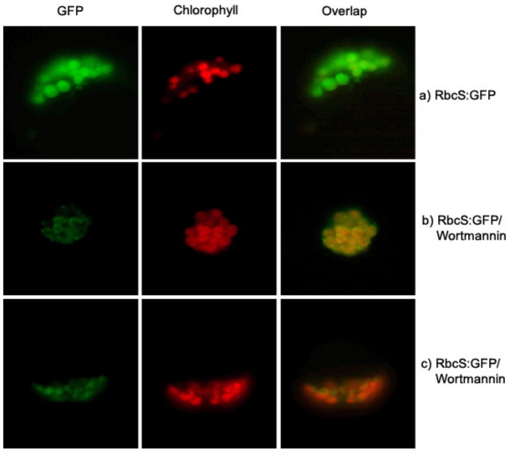
도면 6


GFP:EBD
PI(3)P
GFP:AtPH
PI(3)P
GFP:FAPP1
PI(4)P
GFP:PH
PI(4,5)P2

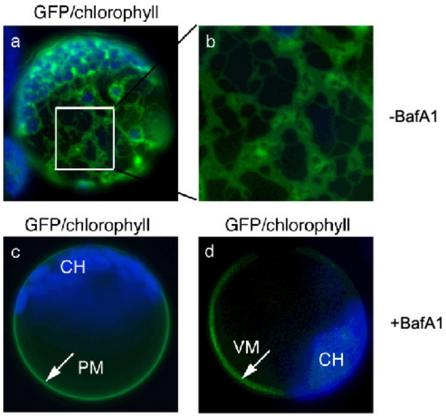
도면 7


a) GFP:EBD
b) GFP:FAPP
c) GFP:PH
a
Vesicles
Bar = 20 µm

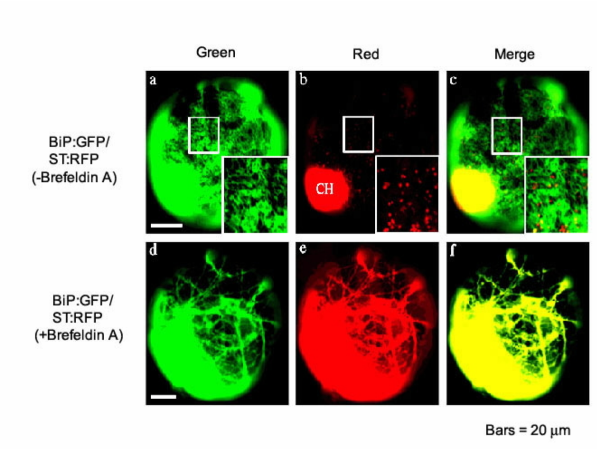
도면 8


GFP
Chlorophyll
Overlap
a) RbcS:GFP
b) RbcS:GFP/ Wortmannin
c) RbcS:GFP/ Wortmannin

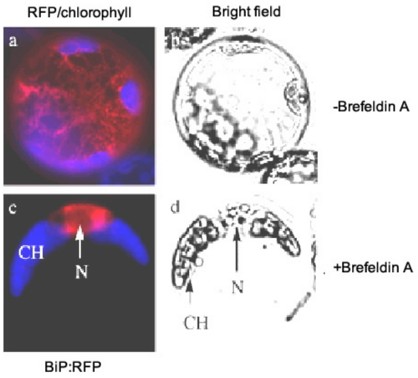
도면 9


GFP/chlorophyll
a
b
-BafA1
GFP/chlorophyll
GFP/chlorophyll
c
CH
PM
d
VM
CH
+BafA1

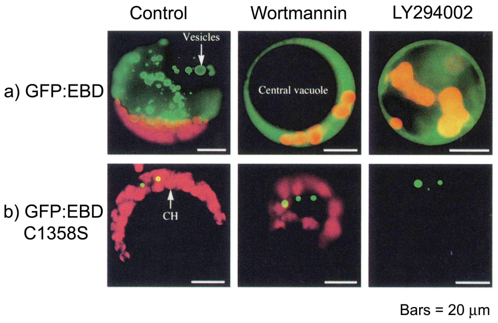
도면 10



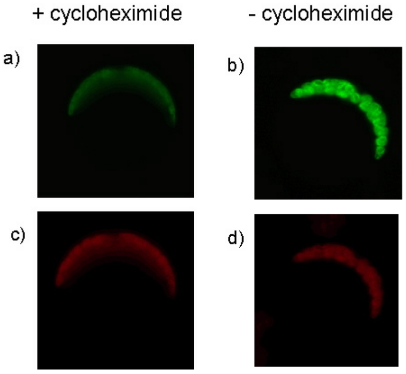
도면 11



도면 12



도면 13



```
<110>    Ahram Biosystems Inc.
<120>    A METHOD FOR INTRACELLUAR PROCESS MONITORING AND RECOMBINANT GENE
         USED THEREFOR
<130>    PN02-535
<150>    KR10-2001-0042886
<151>    2001-07-16
<160>    62
<170>    KopatentIn 1.71
<210>    1
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for partial F1-ATPase coding sequence
<400>    1
ctttaatcaa tggcaatg                                                     18
<210>    2
<211>    25
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(25)
<223>    3' PCR primer for partial F1-ATPase coding sequence
<400>    2
ccatggcctg aactgctcta agctt                                             25
<210>    3
<211>    18
```

```
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for partial rubisco small subunit coding sequence
<400>    3
cctcagtcac acaaagag                                                          18
<210>    4
<211>    24
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(24)
<223>    3' PCR primer for partial rubisco small subunit coding sequence
<400>    4
actcgaggga atcggtaagg tcag                                                   24
<210>    5
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for partial chlorophyll a/b binding protein coding
         sequence
<400>    5
tagagagaaa cgatggcg                                                          18
<210>    6
<211>    25
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(25)
<223>    3' PCR primer for partial chlorophyll a/b binding protein coding
         sequence
<400>    6
ggatcccgtt tgggagtgga actcc                                                  25
<210>    7
<211>    24
```

```
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(24)
<223>    5' PCR primer for partial rubisco acitvase coding sequence
<400>    7
tctagaatgg ccgccgcagt ttcc                                              24
<210>    8
<211>    25
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(25)
<223>    3' PCR primer for partial rubisco acitvase coding sequence
<400>    8
ggatccatct gtctccatcg gtttg                                             25
<210>    9
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for partial AtOEP7 coding sequence
<400>    9
gacgacgacg cagcgatg                                                     18
<210>    10
<211>    25
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(25)
<223>    3' PCR primer for partial AtOEP7 coding sequence
<400>    10
ggatccccaa accctctttg gatgt                                             25
<210>    11
<211>    17
<212>    DNA
<213>    Artificial Sequence
```

```
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(17)
<223>    5' PCR primer for peroxisome targeting sequence
<400>    11
ccgtatgtta catcacc                                                          17
<210>    12
<211>    30
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(30)
<223>    3' PCR primer for peroxisome targeting sequence
<400>    12
ttatagcttt gatttgtata gttcatccat                                           30
<210>    13
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for H+-ATPase coding sequence
<400>    13
gagatgtcga gtctcgaa                                                         18
<210>    14
<211>    23
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(23)
<223>    3' PCR primer for H+-ATPase coding sequence
<400>    14
ctcgagcaca gtgtagtgac tgg                                                   23
<210>    15
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
```

```
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    5' PCR primer for luminal binding protein (BiP) coding sequence
<400>    15
tacgcaaaag tttccgat                                                    18
<210>    16
<211>    18
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(18)
<223>    3' PCR primer for luminal binding protein (BiP) coding sequence
<400>    16
ctagagctca tcgtgaga                                                    18
<210>    17
<211>    21
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(21)
<223>    5' PCR primer for beta-galactoside alpha 2,6-sialtransferase codi
         ng sequence
<400>    17
atgattcata ccaacttgaa g                                                21
<210>    18
<211>    24
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(24)
<223>    3' PCR primer for beta-galactoside alpha 2,6-sialtransferase codi
         ng sequence
<400>    18
ggatccacaa cgaatgttcc ggaa                                             24
<210>    19
<211>    24
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
```

```
<220>
<221>    misc_feature
<222>    (1)..(24)
<223>    5' PCR primer for 500:GFP:KKXX coding sequence comprising an endo
         plamic reticulum targeting sequence
<400>    19
ggatcctcta gaggatcgat ccgg                                          24
<210>    20
<211>    51
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(51)
<223>    3' PCR primer for 500:GFP:KKXX coding sequence comprising an endo
         plamic reticulum targeting sequence
<400>    20
ttagatgagt ttctttttct caaagaaagt tttcaaaagg aatccccctc c            51
<210>    21
<211>    25
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(25)
<223>    5' PCR primer for partial EBD coding sequence
<400>    21
gaattcgtgg caatctagtc aacgg                                         25
<210>    22
<211>    21
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(21)
<223>    3' PCR primer for partial EBD coding sequence
<400>    22
ctaatgttag tgtaatatta c                                             21
<210>    23
<211>    26
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
```

```
<220>
<221>    misc_feature
<222>    (1)..(26)
<223>    5' PCR primer for coding sequence of pleckstrin homology domain o
         f Arabidopsis thaliana (AtPH)
<400>    23
cccgggaaat ggagagtatg tggcga                                              26
<210>    24
<211>    21
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(21)
<223>    3' PCR primer for coding sequence of pleckstrin homology domain o
         f Arabidopsis thaliana (AtPH)
<400>    24
taatcaccgc ctgtgatcat a                                                   21
<210>    25
<211>    27
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(27)
<223>    5' PCR primer for coding sequence of partial pleckstrin homology
         domain of FAPP
<400>    25
ctcgagatgg agggggttct gtacaag                                             27
<210>    26
<211>    24
<212>    DNA
<213>    Artificial Sequence
<220>
<223>    Synthetic Sequence
<220>
<221>    misc_feature
<222>    (1)..(24)
<223>    3' PCR primer for coding sequence of partial pleckstrin homology
         domain of FAPP
<400>    26
tcacgctttg gagctcccaa gggc                                                24
<210>    27
<211>    234
<212>    DNA
<213>    Arabidopsis thaliana
```

```
<220>
<221>    CDS
<222>    (1)..(234)
<223>    Coding sequence for partial gamma subunit of mitochindrial F1-ATP
         ase comprising a mitochondria targeting sequence
<400>    27
atg gca atg gct gtt ttc cgt cgc gaa ggg agg cgt ctc ctc cct tca           48
Met Ala Met Ala Val Phe Arg Arg Glu Gly Arg Arg Leu Leu Pro Ser
  1               5                  10                  15
atc gcc gct cgc cca atc gct gct atc cga tct cct ctc tct tct gac           96
Ile Ala Ala Arg Pro Ile Ala Ala Ile Arg Ser Pro Leu Ser Ser Asp
             20                  25                  30
cag gag gaa gga ctt ctt gga gtt cga tct atc tca act caa gtg gtg          144
Gln Glu Glu Gly Leu Leu Gly Val Arg Ser Ile Ser Thr Gln Val Val
         35                  40                  45
cgt aac cgc atg aag agt gtt aag aac atc caa aag atc aca aag gca          192
Arg Asn Arg Met Lys Ser Val Lys Asn Ile Gln Lys Ile Thr Lys Ala
     50                  55                  60
atg aag atg gtt gct gct tcc aag ctt aga gca gtt cag gcc                  234
Met Lys Met Val Ala Ala Ser Lys Leu Arg Ala Val Gln Ala
 65                  70                  75
<210>  28
<211>  78
<212>  PRT
<213>  Arabidopsis thaliana
<400>  28
Met Ala Met Ala Val Phe Arg Arg Glu Gly Arg Arg Leu Leu Pro Ser
  1               5                  10                  15
Ile Ala Ala Arg Pro Ile Ala Ala Ile Arg Ser Pro Leu Ser Ser Asp
             20                  25                  30
Gln Glu Glu Gly Leu Leu Gly Val Arg Ser Ile Ser Thr Gln Val Val
         35                  40                  45
Arg Asn Arg Met Lys Ser Val Lys Asn Ile Gln Lys Ile Thr Lys Ala
     50                  55                  60
Met Lys Met Val Ala Ala Ser Lys Leu Arg Ala Val Gln Ala
 65                  70                  75
<210>    29
<211>    237
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    CDS
<222>    (1)..(237)
<223>    Coding sequence for partial rubisco small subunit comprising a ch
         loroplast (stroma) targeting sequence
<400>    29
atg gct tcc tct atg ctc tct tcc gct act atg gtt gcc tct ccg gct           48
Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
  1               5                  10                  15
cag gcc act atg gtc gct cct ttc aac gga ctt aag tcc tcc gct gcc           96
```

```
Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
             20                  25                  30
ttc cca gcc acc cgc aag gct aac aac gac att act tcc atc aca agc          144
Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
         35                  40                  45
aac ggc gga aga gtt aac tgc atg cag gtg tgg cct ccg att gga aag          192
Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Gly Lys
     50                  55                  60
aag aag ttt gag act ctc tct tac ctt cct gac ctt acc gat tcc              237
Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser
 65                  70                  75
<210>  30
<211>  79
<212>  PRT
<213>  Arabidopsis thaliana
<400>  30
Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala
  1               5                  10                  15
Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala
             20                  25                  30
Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser
         35                  40                  45
Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Gly Lys
     50                  55                  60
Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Asp Leu Thr Asp Ser
 65                  70                  75
<210>    31
<211>    105
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    CDS
<222>    (1)..(105)
<223>    Coding sequence for partial chlorophyll a/b binding protein compr
         ising a chloroplast (stroma) targeting sequence
<400>    31
atg gcg tcg aac tcg ctt atg agc tgt ggc ata gcc gcc gtg tac cct           48
Met Ala Ser Asn Ser Leu Met Ser Cys Gly Ile Ala Ala Val Tyr Pro
  1               5                  10                  15
tcg ctt ctc tct tct tcc aag tct aaa ttc gta tcc gcc gga gtt cca           96
Ser Leu Leu Ser Ser Ser Lys Ser Lys Phe Val Ser Ala Gly Val Pro
             20                  25                  30
ctc cca aac                                                              105
Leu Pro Asn
         35
<210>  32
<211>  35
<212>  PRT
<213>  Arabidopsis thaliana
<400>  32
```

```
Met Ala Ser Asn Ser Leu Met Ser Cys Gly Ile Ala Ala Val Tyr Pro
  1               5                  10                  15
Ser Leu Leu Ser Ser Ser Lys Ser Lys Phe Val Ser Ala Gly Val Pro
             20                  25                  30
Leu Pro Asn
         35
<210>    33
<211>    204
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    CDS
<222>    (1)..(204)
<223>    Coding sequence for partial rubisco activase comprising a chlorop
         last (stroma) targeting sequence
<400>    33
atg gcc gcc gca gtt tcc acc gtc ggt gcc atc aac aga gct ccg ttg           48
Met Ala Ala Ala Val Ser Thr Val Gly Ala Ile Asn Arg Ala Pro Leu
  1               5                  10                  15
agc ttg aac ggg tca gga tca gga gct gta tca gcc cca gct tca acc           96
Ser Leu Asn Gly Ser Gly Ser Gly Ala Val Ser Ala Pro Ala Ser Thr
             20                  25                  30
ttc ttg gga aag aaa gtt gta act gtg tcg aga ttc gca cag agc aac          144
Phe Leu Gly Lys Lys Val Val Thr Val Ser Arg Phe Ala Gln Ser Asn
         35                  40                  45
aag aag agc aac gga tca ttc aag gtg ttg gct gtg aaa gaa gac aaa          192
Lys Lys Ser Asn Gly Ser Phe Lys Val Leu Ala Val Lys Glu Asp Lys
     50                  55                  60
caa acc gat gga                                                          204
Gln Thr Asp Gly
 65
<210>  34
<211>  68
<212>  PRT
<213>  Arabidopsis thaliana
<400>  34
Met Ala Ala Ala Val Ser Thr Val Gly Ala Ile Asn Arg Ala Pro Leu
  1               5                  10                  15
Ser Leu Asn Gly Ser Gly Ser Gly Ala Val Ser Ala Pro Ala Ser Thr
             20                  25                  30
Phe Leu Gly Lys Lys Val Val Thr Val Ser Arg Phe Ala Gln Ser Asn
         35                  40                  45
Lys Lys Ser Asn Gly Ser Phe Lys Val Leu Ala Val Lys Glu Asp Lys
     50                  55                  60
Gln Thr Asp Gly
 65
<210>    35
<211>    192
<212>    DNA
<213>    Arabidopsis thaliana
```

```
<220>
<221>    CDS
<222>    (1)..(192)
<223>    Coding sequence for partial AtOEP7, a homologue of pea OEP14, com
         prising a chloroplast (outer envelop membrane) targeting sequence
<400>    35
atg gga aaa act tcg gga gcg aaa cag gcg act gtg gtg gtc gca gcg           48
Met Gly Lys Thr Ser Gly Ala Lys Gln Ala Thr Val Val Val Ala Ala
  1               5                  10                  15
atg gcg tta gga tgg tta gcc ata gag atc gct ttc aag cct ttc ctc           96
Met Ala Leu Gly Trp Leu Ala Ile Glu Ile Ala Phe Lys Pro Phe Leu
             20                  25                  30
gat aaa ttc cgc tcc tca atc gac aaa tct gac cca acc aaa gac ccc          144
Asp Lys Phe Arg Ser Ser Ile Asp Lys Ser Asp Pro Thr Lys Asp Pro
         35                  40                  45
gat gac ttc gac acc gcc gct act gca acc aca tcc aaa gag ggt ttg          192
Asp Asp Phe Asp Thr Ala Ala Thr Ala Thr Thr Ser Lys Glu Gly Leu
     50                  55                  60
<210>  36
<211>  64
<212>  PRT
<213>  Arabidopsis thaliana
<400>  36
Met Gly Lys Thr Ser Gly Ala Lys Gln Ala Thr Val Val Val Ala Ala
  1               5                  10                  15
Met Ala Leu Gly Trp Leu Ala Ile Glu Ile Ala Phe Lys Pro Phe Leu
             20                  25                  30
Asp Lys Phe Arg Ser Ser Ile Asp Lys Ser Asp Pro Thr Lys Asp Pro
         35                  40                  45
Asp Asp Phe Asp Thr Ala Ala Thr Ala Thr Thr Ser Lys Glu Gly Leu
     50                  55                  60
<210>    37
<211>    54
<212>    DNA
<213>    SV40
<220>
<221>    CDS
<222>    (1)..(54)
<223>    Coding sequence for SV40 nuclear localization signal (NLS)
<400>    37
atg ggt gct cct cca aaa aag aag aga aag gta gct agt aaa gga gaa           48
Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Ala Ser Lys Gly Glu
  1               5                  10                  15
gaa ctt                                                                   54
Glu Leu
<210>  38
<211>  18
<212>  PRT
<213>  SV40
<400>  38
```

```
Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Ala Ser Lys Gly Glu
  1               5                  10                  15
Glu Leu
<210>    39
<211>    2847
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    gene
<222>    (1)..(2847)
<223>    Gene for H+-ATPase comprising a plasma membrane targeting sequenc
         e
<220>
<221>    CDS
<222>    (1)..(2844)
<223>    H+-ATPase comprising a plasma membrane targeting sequence
<400>    39
atg tcg agt ctc gaa gat atc aag aac gag act gtt gat ctg gaa aaa           48
Met Ser Ser Leu Glu Asp Ile Lys Asn Glu Thr Val Asp Leu Glu Lys
  1               5                  10                  15
att ccg att gag gaa gtt ttc cag cag cta aaa tgt tca agg gaa gga           96
Ile Pro Ile Glu Glu Val Phe Gln Gln Leu Lys Cys Ser Arg Glu Gly
             20                  25                  30
ttg aca acg cag gaa ggg gag gac agg att cag atc ttt ggc ccc aac          144
Leu Thr Thr Gln Glu Gly Glu Asp Arg Ile Gln Ile Phe Gly Pro Asn
         35                  40                  45
aag ctc gaa gag aaa aag gaa agc aaa ctt ctg aag ttt ttg ggg ttt          192
Lys Leu Glu Glu Lys Lys Glu Ser Lys Leu Leu Lys Phe Leu Gly Phe
     50                  55                  60
atg tgg aat cca ctt tca tgg gtc atg gaa atg gct gca atc atg gcc          240
Met Trp Asn Pro Leu Ser Trp Val Met Glu Met Ala Ala Ile Met Ala
 65                  70                  75                  80
att gct ttg gcc aac ggt gat ggt agg cct ccg gat tgg cag gat ttt          288
Ile Ala Leu Ala Asn Gly Asp Gly Arg Pro Pro Asp Trp Gln Asp Phe
                 85                  90                  95
gtt ggt att atc tgt ctg ttg gtt atc aac tct acc atc agt ttt atc          336
Val Gly Ile Ile Cys Leu Leu Val Ile Asn Ser Thr Ile Ser Phe Ile
            100                 105                 110
gaa gaa aac aat gct ggt aat gct gct gct gct ctt atg gct ggt ctt          384
Glu Glu Asn Asn Ala Gly Asn Ala Ala Ala Ala Leu Met Ala Gly Leu
        115                 120                 125
gct cct aaa acc aag gtt ctt agg gat gga aag tgg agt gaa caa gaa          432
Ala Pro Lys Thr Lys Val Leu Arg Asp Gly Lys Trp Ser Glu Gln Glu
    130                 135                 140
gct gct att ctt c cca gga gat att gtt agc att aaa tta gga gac          480
Ala Ala Ile Leu Val Pro Gly Asp Ile Val Ser Ile Lys Leu Gly Asp
145                 150                 155                 160
att atc cca gct gat gcc cgt cta ctt gaa ggt gat cct tta aag gtt          528
Ile Ile Pro Ala Asp Ala Arg Leu Leu Glu Gly Asp Pro Leu Lys Val
                165                 170                 175
```

```
gac caa tct gct cta act gga gag tcc ctt cct gta acc aag cac ccg          576
Asp Gln Ser Ala Leu Thr Gly Glu Ser Leu Pro Val Thr Lys His Pro
            180                 185                 190
ggt caa gaa gtt ttc tct ggt tca acc tgc aaa caa gga gaa atc gag          624
Gly Gln Glu Val Phe Ser Gly Ser Thr Cys Lys Gln Gly Glu Ile Glu
        195                 200                 205
gcg gtt gtt att gcc act ggg gtt cat acc ttc ttc ggt aaa gct gct          672
Ala Val Val Ile Ala Thr Gly Val His Thr Phe Phe Gly Lys Ala Ala
    210                 215                 220
cac ctt gtg gac agc act aac caa gtt gga cat ttc cag aag gtt ctt          720
His Leu Val Asp Ser Thr Asn Gln Val Gly His Phe Gln Lys Val Leu
225                 230                 235                 240
aca gcc att ggg aac ttc tgt atc tgt tcc att gct atc ggt atg gtg          768
Thr Ala Ile Gly Asn Phe Cys Ile Cys Ser Ile Ala Ile Gly Met Val
                245                 250                 255
att gag atc atc gtc atg tat ccg atc caa cgc cga aag tac aga gat          816
Ile Glu Ile Ile Val Met Tyr Pro Ile Gln Arg Arg Lys Tyr Arg Asp
            260                 265                 270
gga att gac aac ctt ttg gtc ctc ttg atc ggt ggt atc ccc att gct          864
Gly Ile Asp Asn Leu Leu Val Leu Leu Ile Gly Gly Ile Pro Ile Ala
        275                 280                 285
atg cct aca gtc ttg tcc gtg acc atg gct att ggg tct cac agg ttg          912
Met Pro Thr Val Leu Ser Val Thr Met Ala Ile Gly Ser His Arg Leu
    290                 295                 300
tct cag caa ggt gcc atc acc aag cgt atg act gcc att gaa gag atg          960
Ser Gln Gln Gly Ala Ile Thr Lys Arg Met Thr Ala Ile Glu Glu Met
305                 310                 315                 320
gca gga atg gat gtc ctg tgc agt gac aaa acc ggg aca cta acc ctc         1008
Ala Gly Met Asp Val Leu Cys Ser Asp Lys Thr Gly Thr Leu Thr Leu
                325                 330                 335
aac aaa ttg agt gtg gac aaa aac ttg gtc gag gtt ttc tgc aag ggt         1056
Asn Lys Leu Ser Val Asp Lys Asn Leu Val Glu Val Phe Cys Lys Gly
            340                 345                 350
gtg gag aaa gat caa gtc cta tta ttt gca gct atg gct tcc agg gtt         1104
Val Glu Lys Asp Gln Val Leu Leu Phe Ala Ala Met Ala Ser Arg Val
        355                 360                 365
gag aac cag gat gcc att gat gca gcc atg gtt ggg atg ctt gct gat         1152
Glu Asn Gln Asp Ala Ile Asp Ala Ala Met Val Gly Met Leu Ala Asp
    370                 375                 380
cca aag gag gct aga gct gga atc agg gaa gtt cac ttc ctt cca ttc         1200
Pro Lys Glu Ala Arg Ala Gly Ile Arg Glu Val His Phe Leu Pro Phe
385                 390                 395                 400
aac cct gtg gat aag aga act gct ttg act tac att gac ggc agt ggt         1248
Asn Pro Val Asp Lys Arg Thr Ala Leu Thr Tyr Ile Asp Gly Ser Gly
                405                 410                 415
aac tgg cac aga gtc agt aaa ggt gct cct gag cag atc ctc gaa ctt         1296
Asn Trp His Arg Val Ser Lys Gly Ala Pro Glu Gln Ile Leu Glu Leu
            420                 425                 430
gcc aaa gcc agc aat gat ctt agc aag aag gtg ctc tcc att att gac         1344
Ala Lys Ala Ser Asn Asp Leu Ser Lys Lys Val Leu Ser Ile Ile Asp
```

```
        435                 440                 445
aag tat gct gag cgt ggt ctt agg tcg ttg gct gtt gct cgc cag gtg         1392
Lys Tyr Ala Glu Arg Gly Leu Arg Ser Leu Ala Val Ala Arg Gln Val
    450                 455                 460
gtg cca gag aaa aca aag gaa agc cca ggt gcg cca tgg gaa ttt gtt         1440
Val Pro Glu Lys Thr Lys Glu Ser Pro Gly Ala Pro Trp Glu Phe Val
465                 470                 475                 480
ggc ttg ttg cca ctt ttt gat ccc cca aga cat gac agt gct gaa aca         1488
Gly Leu Leu Pro Leu Phe Asp Pro Pro Arg His Asp Ser Ala Glu Thr
                485                 490                 495
att cga cgg gct ttg aat ctt ggt gtt aac gtc aag atg atc act ggt         1536
Ile Arg Arg Ala Leu Asn Leu Gly Val Asn Val Lys Met Ile Thr Gly
            500                 505                 510
gac caa ctt gct att ggt aag gaa act ggt cgc aga ctt gga atg gga         1584
Asp Gln Leu Ala Ile Gly Lys Glu Thr Gly Arg Arg Leu Gly Met Gly
        515                 520                 525
aca aac atg tat cca tct tcg gct ctt ctt ggt aca cac aaa gac gca         1632
Thr Asn Met Tyr Pro Ser Ser Ala Leu Leu Gly Thr His Lys Asp Ala
    530                 535                 540
aac ctc gca tcc att cct gtt gag gag ttg att gaa aag gct gat gga         1680
Asn Leu Ala Ser Ile Pro Val Glu Glu Leu Ile Glu Lys Ala Asp Gly
545                 550                 555                 560
ttt gcc gga gtc ttc cca gag cac aaa tac gaa att gtg aaa aag ttg         1728
Phe Ala Gly Val Phe Pro Glu His Lys Tyr Glu Ile Val Lys Lys Leu
                565                 570                 575
cag gag agg aag cat att gtt gga atg act ggt gat ggt gtc aat gat         1776
Gln Glu Arg Lys His Ile Val Gly Met Thr Gly Asp Gly Val Asn Asp
            580                 585                 590
gcc cct gct cta aag aaa gct gat atc ggt att gct gtt gct gat gct         1824
Ala Pro Ala Leu Lys Lys Ala Asp Ile Gly Ile Ala Val Ala Asp Ala
        595                 600                 605
aca gat gct gct cgt ggt gct tca gat atc gtg ctc act gag cct gga         1872
Thr Asp Ala Ala Arg Gly Ala Ser Asp Ile Val Leu Thr Glu Pro Gly
    610                 615                 620
ctc agc gtt att atc agt gct gtt ctc acc agc aga gct att ttc cag         1920
Leu Ser Val Ile Ile Ser Ala Val Leu Thr Ser Arg Ala Ile Phe Gln
625                 630                 635                 640
aga atg aag aac tat act atc tat gca gtc tca atc acc atc cgt att         1968
Arg Met Lys Asn Tyr Thr Ile Tyr Ala Val Ser Ile Thr Ile Arg Ile
                645                 650                 655
gtg ttt ggt ttc atg ctt att gct ttg ata tgg gaa ttt gac ttc tca         2016
Val Phe Gly Phe Met Leu Ile Ala Leu Ile Trp Glu Phe Asp Phe Ser
            660                 665                 670
gcc ttc atg gtt ctg atc att gcc att ctt aac gac ggt acc atc atg         2064
Ala Phe Met Val Leu Ile Ile Ala Ile Leu Asn Asp Gly Thr Ile Met
        675                 680                 685
aca atc tca aag gac aga gtt aag cca tct ccc aca cct gat agc tgg         2112
Thr Ile Ser Lys Asp Arg Val Lys Pro Ser Pro Thr Pro Asp Ser Trp
    690                 695                 700
aaa ctt aaa gaa att ttt gct act gga gtc gtt cta gga ggc tac cag         2160
```

```
Lys Leu Lys Glu Ile Phe Ala Thr Gly Val Val Leu Gly Gly Tyr Gln
705                 710                 715                 720
gcc atc atg act gtt att ttc ttc tgg gcg gcg cac aag act gac ttt         2208
Ala Ile Met Thr Val Ile Phe Phe Trp Ala Ala His Lys Thr Asp Phe
                725                 730                 735
ttc tcg gac aca ttc ggt gtg agg tcc att agg gac aat aac cac gag         2256
Phe Ser Asp Thr Phe Gly Val Arg Ser Ile Arg Asp Asn Asn His Glu
            740                 745                 750
cta atg ggt gcg gtg tac tta caa gtt agt atc att agt caa gct ctg         2304
Leu Met Gly Ala Val Tyr Leu Gln Val Ser Ile Ile Ser Gln Ala Leu
        755                 760                 765
atc ttc gtc aca aga tca agg agt tgg tct ttt gtt gaa cgt cct gga         2352
Ile Phe Val Thr Arg Ser Arg Ser Trp Ser Phe Val Glu Arg Pro Gly
    770                 775                 780
gca ttg ctg atg att gct ttc ctc att gca caa ctg att gct act ttg         2400
Ala Leu Leu Met Ile Ala Phe Leu Ile Ala Gln Leu Ile Ala Thr Leu
785                 790                 795                 800
att gcg gtt tac gcc aac tgg gaa ttt gca aag att agg ggt att gga         2448
Ile Ala Val Tyr Ala Asn Trp Glu Phe Ala Lys Ile Arg Gly Ile Gly
                805                 810                 815
tgg gga tgg gct ggt gtg atc tgg cta tac agt att gtc aca tac ttc         2496
Trp Gly Trp Ala Gly Val Ile Trp Leu Tyr Ser Ile Val Thr Tyr Phe
            820                 825                 830
cca ttg gac gtt ttc aag ttt gcc att cga tac atc ttg agc gga aag         2544
Pro Leu Asp Val Phe Lys Phe Ala Ile Arg Tyr Ile Leu Ser Gly Lys
        835                 840                 845
gcg tgg ctc aac ttg ttt gag aac aag acg gct ttc acg atg aag aaa         2592
Ala Trp Leu Asn Leu Phe Glu Asn Lys Thr Ala Phe Thr Met Lys Lys
    850                 855                 860
gat tac gga aaa gaa gag aga gag gct caa tgg gca ctt gct caa agg         2640
Asp Tyr Gly Lys Glu Glu Arg Glu Ala Gln Trp Ala Leu Ala Gln Arg
865                 870                 875                 880
aca ctt cac ggt tta cag cca aaa gaa gct gtt aac atc ttc cct gag         2688
Thr Leu His Gly Leu Gln Pro Lys Glu Ala Val Asn Ile Phe Pro Glu
                885                 890                 895
aaa gga agt tac aga gaa ttg tct gag atc gct gag caa gct aag aga         2736
Lys Gly Ser Tyr Arg Glu Leu Ser Glu Ile Ala Glu Gln Ala Lys Arg
            900                 905                 910
aga gct gag atc gct agg ctt agg gag ctg cac aca ctc aag gga cat         2784
Arg Ala Glu Ile Ala Arg Leu Arg Glu Leu His Thr Leu Lys Gly His
        915                 920                 925
gtg gaa tca gtc gtg aag cta aag ggc ttg gac att gaa act ccc agt         2832
Val Glu Ser Val Val Lys Leu Lys Gly Leu Asp Ile Glu Thr Pro Ser
    930                 935                 940
cac tac act gtg           tag                                           2847
His Tyr Thr Val
945
<210>  40
<211>  948
<212>  PRT
```

```
<213>  Arabidopsis thaliana
<400>  40
Met Ser Ser Leu Glu Asp Ile Lys Asn Glu Thr Val Asp Leu Glu Lys
  1               5                  10                  15
Ile Pro Ile Glu Glu Val Phe Gln Gln Leu Lys Cys Ser Arg Glu Gly
             20                  25                  30
Leu Thr Thr Gln Glu Gly Glu Asp Arg Ile Gln Ile Phe Gly Pro Asn
         35                  40                  45
Lys Leu Glu Glu Lys Lys Glu Ser Lys Leu Leu Lys Phe Leu Gly Phe
     50                  55                  60
Met Trp Asn Pro Leu Ser Trp Val Met Glu Met Ala Ala Ile Met Ala
 65                  70                  75                  80
Ile Ala Leu Ala Asn Gly Asp Gly Arg Pro Pro Asp Trp Gln Asp Phe
                 85                  90                  95
Val Gly Ile Ile Cys Leu Leu Val Ile Asn Ser Thr Ile Ser Phe Ile
            100                 105                 110
Glu Glu Asn Asn Ala Gly Asn Ala Ala Ala Ala Leu Met Ala Gly Leu
        115                 120                 125
Ala Pro Lys Thr Lys Val Leu Arg Asp Gly Lys Trp Ser Glu Gln Glu
    130                 135                 140
Ala Ala Ile Leu Val Pro Gly Asp Ile Val Ser Ile Lys Leu Gly Asp
145                 150                 155                 160
Ile Ile Pro Ala Asp Ala Arg Leu Leu Glu Gly Asp Pro Leu Lys Val
                165                 170                 175
Asp Gln Ser Ala Leu Thr Gly Glu Ser Leu Pro Val Thr Lys His Pro
            180                 185                 190
Gly Gln Glu Val Phe Ser Gly Ser Thr Cys Lys Gln Gly Glu Ile Glu
        195                 200                 205
Ala Val Val Ile Ala Thr Gly Val His Thr Phe Phe Gly Lys Ala Ala
    210                 215                 220
His Leu Val Asp Ser Thr Asn Gln Val Gly His Phe Gln Lys Val Leu
225                 230                 235                 240
Thr Ala Ile Gly Asn Phe Cys Ile Cys Ser Ile Ala Ile Gly Met Val
                245                 250                 255
Ile Glu Ile Ile Val Met Tyr Pro Ile Gln Arg Arg Lys Tyr Arg Asp
            260                 265                 270
Gly Ile Asp Asn Leu Leu Val Leu Leu Ile Gly Gly Ile Pro Ile Ala
        275                 280                 285
Met Pro Thr Val Leu Ser Val Thr Met Ala Ile Gly Ser His Arg Leu
    290                 295                 300
Ser Gln Gln Gly Ala Ile Thr Lys Arg Met Thr Ala Ile Glu Glu Met
305                 310                 315                 320
Ala Gly Met Asp Val Leu Cys Ser Asp Lys Thr Gly Thr Leu Thr Leu
                325                 330                 335
Asn Lys Leu Ser Val Asp Lys Asn Leu Val Glu Val Phe Cys Lys Gly
            340                 345                 350
Val Glu Lys Asp Gln Val Leu Leu Phe Ala Ala Met Ala Ser Arg Val
        355                 360                 365
Glu Asn Gln Asp Ala Ile Asp Ala Ala Met Val Gly Met Leu Ala Asp
    370                 375                 380
```

```
Pro Lys Glu Ala Arg Ala Gly Ile Arg Glu Val His Phe Leu Pro Phe
385                 390                 395                 400
Asn Pro Val Asp Lys Arg Thr Ala Leu Thr Tyr Ile Asp Gly Ser Gly
                405                 410                 415
Asn Trp His Arg Val Ser Lys Gly Ala Pro Glu Gln Ile Leu Glu Leu
            420                 425                 430
Ala Lys Ala Ser Asn Asp Leu Ser Lys Lys Val Leu Ser Ile Ile Asp
        435                 440                 445
Lys Tyr Ala Glu Arg Gly Leu Arg Ser Leu Ala Val Ala Arg Gln Val
    450                 455                 460
Val Pro Glu Lys Thr Lys Glu Ser Pro Gly Ala Pro Trp Glu Phe Val
465                 470                 475                 480
Gly Leu Leu Pro Leu Phe Asp Pro Pro Arg His Asp Ser Ala Glu Thr
                485                 490                 495
Ile Arg Arg Ala Leu Asn Leu Gly Val Asn Val Lys Met Ile Thr Gly
            500                 505                 510
Asp Gln Leu Ala Ile Gly Lys Glu Thr Gly Arg Arg Leu Gly Met Gly
        515                 520                 525
Thr Asn Met Tyr Pro Ser Ser Ala Leu Leu Gly Thr His Lys Asp Ala
    530                 535                 540
Asn Leu Ala Ser Ile Pro Val Glu Glu Leu Ile Glu Lys Ala Asp Gly
545                 550                 555                 560
Phe Ala Gly Val Phe Pro Glu His Lys Tyr Glu Ile Val Lys Lys Leu
                565                 570                 575
Gln Glu Arg Lys His Ile Val Gly Met Thr Gly Asp Gly Val Asn Asp
            580                 585                 590
Ala Pro Ala Leu Lys Lys Ala Asp Ile Gly Ile Ala Val Ala Asp Ala
        595                 600                 605
Thr Asp Ala Ala Arg Gly Ala Ser Asp Ile Val Leu Thr Glu Pro Gly
    610                 615                 620
Leu Ser Val Ile Ile Ser Ala Val Leu Thr Ser Arg Ala Ile Phe Gln
625                 630                 635                 640
Arg Met Lys Asn Tyr Thr Ile Tyr Ala Val Ser Ile Thr Ile Arg Ile
                645                 650                 655
Val Phe Gly Phe Met Leu Ile Ala Leu Ile Trp Glu Phe Asp Phe Ser
            660                 665                 670
Ala Phe Met Val Leu Ile Ile Ala Ile Leu Asn Asp Gly Thr Ile Met
        675                 680                 685
Thr Ile Ser Lys Asp Arg Val Lys Pro Ser Pro Thr Pro Asp Ser Trp
    690                 695                 700
Lys Leu Lys Glu Ile Phe Ala Thr Gly Val Val Leu Gly Gly Tyr Gln
705                 710                 715                 720
Ala Ile Met Thr Val Ile Phe Phe Trp Ala Ala His Lys Thr Asp Phe
                725                 730                 735
Phe Ser Asp Thr Phe Gly Val Arg Ser Ile Arg Asp Asn Asn His Glu
            740                 745                 750
Leu Met Gly Ala Val Tyr Leu Gln Val Ser Ile Ile Ser Gln Ala Leu
        755                 760                 765
Ile Phe Val Thr Arg Ser Arg Ser Trp Ser Phe Val Glu Arg Pro Gly
    770                 775                 780
```

```
Ala Leu Leu Met Ile Ala Phe Leu Ile Ala Gln Leu Ile Ala Thr Leu
785                 790                 795                 800
Ile Ala Val Tyr Ala Asn Trp Glu Phe Ala Lys Ile Arg Gly Ile Gly
                805                 810                 815
Trp Gly Trp Ala Gly Val Ile Trp Leu Tyr Ser Ile Val Thr Tyr Phe
            820                 825                 830
Pro Leu Asp Val Phe Lys Phe Ala Ile Arg Tyr Ile Leu Ser Gly Lys
        835                 840                 845
Ala Trp Leu Asn Leu Phe Glu Asn Lys Thr Ala Phe Thr Met Lys Lys
    850                 855                 860
Asp Tyr Gly Lys Glu Glu Arg Glu Ala Gln Trp Ala Leu Ala Gln Arg
865                 870                 875                 880
Thr Leu His Gly Leu Gln Pro Lys Glu Ala Val Asn Ile Phe Pro Glu
                885                 890                 895
Lys Gly Ser Tyr Arg Glu Leu Ser Glu Ile Ala Glu Gln Ala Lys Arg
            900                 905                 910
Arg Ala Glu Ile Ala Arg Leu Arg Glu Leu His Thr Leu Lys Gly His
        915                 920                 925
Val Glu Ser Val Val Lys Leu Lys Gly Leu Asp Ile Glu Thr Pro Ser
    930                 935                 940
His Tyr Thr Val
945
<210>    41
<211>    132
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    CDS
<222>    (1)..(132)
<223>    Coding sequence for amino terminal region of luminal binding prot
         ein (BiP) comprising an endoplasmic reticulum (lumen) targeting s
         equence or its portion
<400>    41
atg gct cgc tcg ttt gga gca aac agc act gtt gtg ctt gcg atc atc           48
Met Ala Arg Ser Phe Gly Ala Asn Ser Thr Val Val Leu Ala Ile Ile
  1               5                  10                  15
ttc ttc gga tgt tta ttt gcg ttt tcc act gca aaa gaa gag gct acg           96
Phe Phe Gly Cys Leu Phe Ala Phe Ser Thr Ala Lys Glu Glu Ala Thr
             20                  25                  30
aag tta gga tca gtt att ggg ata gat ctt ggt acc                          132
Lys Leu Gly Ser Val Ile Gly Ile Asp Leu Gly Thr
         35                  40
<210>  42
<211>  44
<212>  PRT
<213>  Arabidopsis thaliana
<400>  42
Met Ala Arg Ser Phe Gly Ala Asn Ser Thr Val Val Leu Ala Ile Ile
  1               5                  10                  15
Phe Phe Gly Cys Leu Phe Ala Phe Ser Thr Ala Lys Glu Glu Ala Thr
```

```
             20                  25                  30
Lys Leu Gly Ser Val Ile Gly Ile Asp Leu Gly Thr
         35                  40
<210>    43
<211>    243
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    gene
<222>    (1)..(243)
<223>    Gene for carboxyl terminal region of luminal binding protein (BiP
         ) comprising an endoplasmic reticulum (lumen) targeting sequence
         or its portion
<220>
<221>    CDS
<222>    (1)..(240)
<223>    Carboxyl terminal region of luminal binding protein (BiP) compris
         ing an endoplasmic reticulum (lumen) targeting sequence or its po
         rtion
<400>    43
aag ctt gca gac aaa ttg gaa gga gat gag aag gag aag ata gaa gca           48
Lys Leu Ala Asp Lys Leu Glu Gly Asp Glu Lys Glu Lys Ile Glu Ala
  1               5                 1                  15
gcg acg aaa gag gcc ttg gag tgg ctc gac gag aac caa aac tca gag           96
Ala Thr Lys Glu Ala Leu Glu Trp Leu Asp Glu Asn Gln Asn Ser Glu
             20                  25                  30
aaa gaa gag tac gac gag aag ctc aag gag gta gag gca gtg tgt aac          144
Lys Glu Glu Tyr Asp Glu Lys Leu Lys Glu Val Glu Ala Val Cys Asn
         35                  40                  45
cca atc atc aca gca gtt tac cag aga tca ggc gga gca cca ggt gca          192
Pro Ile Ile Thr Ala Val Tyr Gln Arg Ser Gly Gly Ala Pro Gly Ala
     50                  55                  60
gga gga gaa tca tcg act gag gag gaa gat gag tct cac gat gag ctc          240
Gly Gly Glu Ser Ser Thr Glu Glu Glu Asp Glu Ser His Asp Glu Leu
 65                  70                  75                  80
tag                                                                      243
<210>  44
<211>  80
<212>  PRT
<213>  Arabidopsis thaliana
<400>  44
Lys Leu Ala Asp Lys Leu Glu Gly Asp Glu Lys Glu Lys Ile Glu Ala
  1               5                  10                  15
Ala Thr Lys Glu Ala Leu Glu Trp Leu Asp Glu Asn Gln Asn Ser Glu
             20                  25                  30
Lys Glu Glu Tyr Asp Glu Lys Leu Lys Glu Val Glu Ala Val Cys Asn
         35                  40                  45
Pro Ile Ile Thr Ala Val Tyr Gln Arg Ser Gly Gly Ala Pro Gly Ala
     50                  55                  60
Gly Gly Glu Ser Ser Thr Glu Glu Glu Asp Glu Ser His Asp Glu Leu
```

```
 65                  70                  75                  80
<210>    45
<211>    1212
<212>    DNA
<213>    Rattus norvegicus
<220>
<221>    gene
<222>    (1)..(1212)
<223>    Gene for beta-galactoside alpha 2,6-sialyltransferase comprising
         a Golgi apparatus targeting sequence
<220>
<221>    CDS
<222>    (1)..(1209)
<223>    Beta-galactoside alpha 2,6-sialyltransferase comprising a Golgi a
         pparatus targeting sequence
<400>    45
atg att cat acc aac ttg aag aaa aag ttc agc ctc ttc atc ctg gtc           48
Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Leu Phe Ile Leu Val
  1               5                  10                  15
ttt ctc ctg ttc gca gtc atc tgt gtt tgg aag aaa ggg agc gac tat           96
Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Lys Gly Ser Asp Tyr
             20                  25                  30
gag gcc ctt aca ctg caa gcc aag gaa ttc cag atg ccc aag agc cag          144
Glu Ala Leu Thr Leu Gln Ala Lys Glu Phe Gln Met Pro Lys Ser Gln
         35                  40                  45
gag aaa gtg gcc atg ggg tct gct tcc cag gtt gtg ttc tca aac agc          192
Glu Lys Val Ala Met Gly Ser Ala Ser Gln Val Val Phe Ser Asn Ser
     50                  55                  60
aag caa gac cct aag gaa gac att cca atc ctc agt tac cac agg gtc          240
Lys Gln Asp Pro Lys Glu Asp Ile Pro Ile Leu Ser Tyr His Arg Val
 65                  70                  75                  80
aca gcc aag gtc aaa cca cag cct tcc ttc cag gtg tgg gac aag gac          288
Thr Ala Lys Val Lys Pro Gln Pro Ser Phe Gln Val Trp Asp Lys Asp
                 85                  90                  95
tcc aca tac tca aaa ctt aac ccc agg ctg ctg aag atc tgg aga aac          336
Ser Thr Tyr Ser Lys Leu Asn Pro Arg Leu Leu Lys Ile Trp Arg Asn
            100                 105                 110
tat ctg aac atg aac aaa tat aaa gta tcc tac aag gga ccg ggg cca          384
Tyr Leu Asn Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro
        115                 120                 125
gga gtc aag ttc agc gta gaa gca ctg cgt tgc cac ctt cga gac cat          432
Gly Val Lys Phe Ser Val Glu Ala Leu Arg Cys His Leu Arg Asp His
    130                 135                 140
gtg aac gtg tct atg ata gag gcc aca gat ttt ccc ttc aac acc act          480
Val Asn Val Ser Met Ile Glu Ala Thr Asp Phe Pro Phe Asn Thr Thr
145                 150                 155                 160
gag tgg gag ggt tac ctg ccc aag gag aac ttt aga acc aag gtt ggg          528
Glu Trp Glu Gly Tyr Leu Pro Lys Glu Asn Phe Arg Thr Lys Val Gly
                165                 170                 175
cct tgg caa agg tgt gcc gtc gtc tct tct gca gga tct ctg aaa aac          576
```

```
Pro Trp Gln Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Asn
            180                 185                 190
tcc cag ctt ggt cga gag att gat aat cat gat gca gtt ctg agg ttt          624
Ser Gln Leu Gly Arg Glu Ile Asp Asn His Asp Ala Val Leu Arg Phe
        195                 200                 205
aat ggg gcc cct acc gac aac ttc caa cag gat gtg ggc tca aaa act          672
Asn Gly Ala Pro Thr Asp Asn Phe Gln Gln Asp Val Gly Ser Lys Thr
    210                 215                 220
acc att cgc cta atg aac tct cag tta gtc acc aca gaa aag cgc ttc          720
Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe
225                 230                 235                 240
ctc aag gac agt ttg tac acc gaa gga atc cta att gta tgg gac cca          768
Leu Lys Asp Ser Leu Tyr Thr Glu Gly Ile Leu Ile Val Trp Asp Pro
                245                 250                 255
tcc gtg tat cat gca gat atc cca aag tgg tat cag aaa cca gac tac          816
Ser Val Tyr His Ala Asp Ile Pro Lys Trp Tyr Gln Lys Pro Asp Tyr
            260                 265                 270
aat ttc ttc gaa acc tat aag agt tac cga agg ctg aac ccc agc cag          864
Asn Phe Phe Glu Thr Tyr Lys Ser Tyr Arg Arg Leu Asn Pro Ser Gln
        275                 280                 285
cca ttt tat atc ctc aag ccc cag atg cca tgg gaa ctg tgg gac atc          912
Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile
    290                 295                 300
att cag gaa atc tct gca gat ctg att cag cca aat ccc cca tcc tcc          960
Ile Gln Glu Ile Ser Ala Asp Leu Ile Gln Pro Asn Pro Pro Ser Ser
305                 310                 315                 320
ggc atg ctg ggt atc atc atc atg atg acg ctg tgt gac cag gta gat         1008
Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp
                325                 330                 335
att tac gag ttc ctc cca tcc aag cgc aag acg gac gtg tgc tat tat         1056
Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr
            340                 345                 350
cac caa aag ttc ttt gac agc gct tgc acg atg ggt gcc tac gac ccg         1104
His Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr Asp Pro
        355                 360                 365
ctc ctc ttc gag aag aat atg gtg aag cat ctc aat gag gga aca gat         1152
Leu Leu Phe Glu Lys Asn Met Val Lys His Leu Asn Glu Gly Thr Asp
    370                 375                 380
gaa gac att tat ttg ttt ggg aaa gcc acc ctt tct ggc ttc cgg aac         1200
Glu Asp Ile Tyr Leu Phe Gly Lys Ala Thr Leu Ser Gly Phe Arg Asn
385                 390                 395                 400
att cgt tgt         t ga                                                1212
Ile Arg Cys
<210>  46
<211>  403
<212>  PRT
<213>  Rattus norvegicus
<400>  46
Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Leu Phe Ile Leu Val
  1               5                  10                  15
```

```
Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Lys Gly Ser Asp Tyr
             20                  25                  30
Glu Ala Leu Thr Leu Gln Ala Lys Glu Phe Gln Met Pro Lys Ser Gln
         35                  40                  45
Glu Lys Val Ala Met Gly Ser Ala Ser Gln Val Val Phe Ser Asn Ser
     50                  55                  60
Lys Gln Asp Pro Lys Glu Asp Ile Pro Ile Leu Ser Tyr His Arg Val
 65                  70                  75                  80
Thr Ala Lys Val Lys Pro Gln Pro Ser Phe Gln Val Trp Asp Lys Asp
                 85                  90                  95
Ser Thr Tyr Ser Lys Leu Asn Pro Arg Leu Leu Lys Ile Trp Arg Asn
            100                 105                 110
Tyr Leu Asn Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro
        115                 120                 125
Gly Val Lys Phe Ser Val Glu Ala Leu Arg Cys His Leu Arg Asp His
    130                 135                 140
Val Asn Val Ser Met Ile Glu Ala Thr Asp Phe Pro Phe Asn Thr Thr
145                 150                 155                 160
Glu Trp Glu Gly Tyr Leu Pro Lys Glu Asn Phe Arg Thr Lys Val Gly
                165                 170                 175
Pro Trp Gln Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Asn
            180                 185                 190
Ser Gln Leu Gly Arg Glu Ile Asp Asn His Asp Ala Val Leu Arg Phe
        195                 200                 205
Asn Gly Ala Pro Thr Asp Asn Phe Gln Gln Asp Val Gly Ser Lys Thr
    210                 215                 220
Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe
225                 230                 235                 240
Leu Lys Asp Ser Leu Tyr Thr Glu Gly Ile Leu Ile Val Trp Asp Pro
                245                 250                 255
Ser Val Tyr His Ala Asp Ile Pro Lys Trp Tyr Gln Lys Pro Asp Tyr
            260                 265                 270
Asn Phe Phe Glu Thr Tyr Lys Ser Tyr Arg Arg Leu Asn Pro Ser Gln
        275                 280                 285
Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile
    290                 295                 300
Ile Gln Glu Ile Ser Ala Asp Leu Ile Gln Pro Asn Pro Pro Ser Ser
305                 310                 315                 320
Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp
                325                 330                 335
Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr
            340                 345                 350
His Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr Asp Pro
        355                 360                 365
Leu Leu Phe Glu Lys Asn Met Val Lys His Leu Asn Glu Gly Thr Asp
    370                 375                 380
Glu Asp Ile Tyr Leu Phe Gly Lys Ala Thr Leu Ser Gly Phe Arg Asn
385                 390                 395                 400
Ile Arg Cys
<210>    47
```

```
<211>    180
<212>    DNA
<213>    Phaseolus vulgaris
<220>
<221>    CDS
<222>    (1)..(180)
<223>    Coding sequence for amino terminal region of chitinase comprising
         a part of a storage vacuole targeting sequence
<400>    47
atg aag aag aat agg atg atg atg atg ata tgg agc gta gga gtg gtg           48
Met Lys Lys Asn Arg Met Met Met Met Ile Trp Ser Val Gly Val Val
  1               5                  10                  15
tgg atg ctg ttg ttg gtt gga gga agc tac gga gag cag tgt gga agg           96
Trp Met Leu Leu Leu Val Gly Gly Ser Tyr Gly Glu Gln Cys Gly Arg
             20                  25                  30
caa gca gga ggt gca ctc tgc cca ggg ggc aac tgt tgc agc cag ttc          144
Gln Ala Gly Gly Ala Leu Cys Pro Gly Gly Asn Cys Cys Ser Gln Phe
         35                  40                  45
ggg tgg tgc ggc tcc acc acc gac tac tgc ggc ccg                          180
Gly Trp Cys Gly Ser Thr Thr Asp Tyr Cys Gly Pro
     50                  55                  60
<210>  48
<211>  60
<212>  PRT
<213>  Phaseolus vulgaris
<400>  48
Met Lys Lys Asn Arg Met Met Met Met Ile Trp Ser Val Gly Val Val
  1               5                  10                  15
Trp Met Leu Leu Leu Val Gly Gly Ser Tyr Gly Glu Gln Cys Gly Arg
             20                  25                  30
Gln Ala Gly Gly Ala Leu Cys Pro Gly Gly Asn Cys Cys Ser Gln Phe
         35                  40                  45
Gly Trp Cys Gly Ser Thr Thr Asp Tyr Cys Gly Pro
     50                  55                  60
<210>    49
<211>    411
<212>    DNA
<213>    Phaseolus vulgaris
<220>
<221>    CDS
<222>    (1)..(411)
<223>    Coding sequence for carboxyl terminal region of chitinase compris
         ing a part of a storage vacuole targeting sequence
<400>    49
ata tcc tgg aac tac aac tac ggt cag tgc gga agg gcc att ggg gtt           48
Ile Ser Trp Asn Tyr Asn Tyr Gly Gln Cys Gly Arg Ala Ile Gly Val
  1               5                  10                  15
gac ttg ctc aac aaa cct gat cta gtc gcc act gac tct gtc atc tcc           96
Asp Leu Leu Asn Lys Pro Asp Leu Val Ala Thr Asp Ser Val Ile Ser
             20                  25                  30
```

```
ttc aag tcc gcc ctc tgg ttc tgg atg acc gca cag tcc ccc aag cct          144
Phe Lys Ser Ala Leu Trp Phe Trp Met Thr Ala Gln Ser Pro Lys Pro
         35                  40                  45
tcc tcc cac gac gtc atc acc tct cga tgg acc ccc tcc tct gcc gac          192
Ser Ser His Asp Val Ile Thr Ser Arg Trp Thr Pro Ser Ser Ala Asp
     50                  55                  60
gtc gcc gcc cgc cgg ctt ccc ggc tac ggc act gtg acg aac atc atc          240
Val Ala Ala Arg Arg Leu Pro Gly Tyr Gly Thr Val Thr Asn Ile Ile
 65                  70                  75                  80
aac gga ggc ctg gag tgc ggg cga gga cag gac agc agg gtt caa gac          288
Asn Gly Gly Leu Glu Cys Gly Arg Gly Gln Asp Ser Arg Val Gln Asp
                 85                  90                  95
cgc atc gga ttc ttc aag aga tac tgt gat ctg ctt gga gtc ggt tat          336
Arg Ile Gly Phe Phe Lys Arg Tyr Cys Asp Leu Leu Gly Val Gly Tyr
            100                 105                 110
ggc aac aac ctt gac tgc tac tct cag act cca ttt gga aat tca ctc          384
Gly Asn Asn Leu Asp Cys Tyr Ser Gln Thr Pro Phe Gly Asn Ser Leu
        115                 120                 125
tta ctc tct gac ctt gtc acc tct cag                                      411
Leu Leu Ser Asp Leu Val Thr Ser Gln
    130                 135
<210>  50
<211>  137
<212>  PRT
<213>  Phaseolus vulgaris
<400>  50
Ile Ser Trp Asn Tyr Asn Tyr Gly Gln Cys Gly Arg Ala Ile Gly Val
  1               5                  10                  15
Asp Leu Leu Asn Lys Pro Asp Leu Val Ala Thr Asp Ser Val Ile Ser
             20                  25                  30
Phe Lys Ser Ala Leu Trp Phe Trp Met Thr Ala Gln Ser Pro Lys Pro
         35                  40                  45
Ser Ser His Asp Val Ile Thr Ser Arg Trp Thr Pro Ser Ser Ala Asp
     50                  55                  60
Val Ala Ala Arg Arg Leu Pro Gly Tyr Gly Thr Val Thr Asn Ile Ile
 65                  70                  75                  80
Asn Gly Gly Leu Glu Cys Gly Arg Gly Gln Asp Ser Arg Val Gln Asp
                 85                  90                  95
Arg Ile Gly Phe Phe Lys Arg Tyr Cys Asp Leu Leu Gly Val Gly Tyr
            100                 105                 110
Gly Asn Asn Leu Asp Cys Tyr Ser Gln Thr Pro Phe Gly Asn Ser Leu
        115                 120                 125
Leu Leu Ser Asp Leu Val Thr Ser Gln
    130                 135
<210>    51
<211>    666
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    gene
```

```
<222>    (1)..(666)
<223>    VTI1a gene encoding v-SNARE AtVTI1a protein comprising a trans-Go
         lgi network targeting sequence
<220>
<221>    CDS
<222>    (1)..(663)
<223>    v-SNARE AtVTI1a comprising a trans-Golgi network targeting sequen
         ce
<400>    51
atg agt gac gtg ttt gat gga tat gag cga caa tac tgt gag ctc tcg            48
Met Ser Asp Val Phe Asp Gly Tyr Glu Arg Gln Tyr Cys Glu Leu Ser
  1               5                  10                  15
gcg agt ctt tca aag aag tgt tcc tct gct att tcg ctt gat gga gaa            96
Ala Ser Leu Ser Lys Lys Cys Ser Ser Ala Ile Ser Leu Asp Gly Glu
             20                  25                  30
caa aag aag cag aag ttg tct gaa ata aaa tcc ggg ctg gaa aat gca           144
Gln Lys Lys Gln Lys Leu Ser Glu Ile Lys Ser Gly Leu Glu Asn Ala
         35                  40                  45
gaa gta ttg att agg aaa atg gat ctc gag gca agg acc ctt cca cct           192
Glu Val Leu Ile Arg Lys Met Asp Leu Glu Ala Arg Thr Leu Pro Pro
     50                  55                  60
aat ctt aag tcc agc ctc ctt gtc aaa tta aga gaa ttc aag tct gat           240
Asn Leu Lys Ser Ser Leu Leu Val Lys Leu Arg Glu Phe Lys Ser Asp
 65                  70                  75                  80
ctg aac aat ttc aaa act gaa gtg aag aga atc aca tct gga cag ttg           288
Leu Asn Asn Phe Lys Thr Glu Val Lys Arg Ile Thr Ser Gly Gln Leu
                 85                  90                  95
aac gct gct gct cgt gat gag ttg cta gaa gct ggt atg gct gac aca           336
Asn Ala Ala Ala Arg Asp Glu Leu Leu Glu Ala Gly Met Ala Asp Thr
            100                 105                 110
aag acg gct tca gct gat cag aga gca aga ttg atg atg tca act gag           384
Lys Thr Ala Ser Ala Asp Gln Arg Ala Arg Leu Met Met Ser Thr Glu
        115                 120                 125
cgt ctg ggc cga acc act gac aga gtc aag gac agt cgt aga aca atg           432
Arg Leu Gly Arg Thr Thr Asp Arg Val Lys Asp Ser Arg Arg Thr Met
    130                 135                 140
atg gag act gaa gaa atc ggt gtt tca atc ctc caa gac ttg cac ggt           480
Met Glu Thr Glu Glu Ile Gly Val Ser Ile Leu Gln Asp Leu His Gly
145                 150                 155                 160
cag aga cag tct ctc tta cgt gcc cac gaa acg ctt cat gga gta gac           528
Gln Arg Gln Ser Leu Leu Arg Ala His Glu Thr Leu His Gly Val Asp
                165                 170                 175
gat aac att gga aag agc aag aag atc ttg aca gac atg acg agg agg           576
Asp Asn Ile Gly Lys Ser Lys Lys Ile Leu Thr Asp Met Thr Arg Arg
            180                 185                 190
atg aac aag aac aaa tgg acc att gga gcc atc atc atc gca ttg att           624
Met Asn Lys Asn Lys Trp Thr Ile Gly Ala Ile Ile Ile Ala Leu Ile
        195                 200                 205
gcc gcc att ttc atc atc ttg tac ttc aaa ctc acc aag       taa            666
Ala Ala Ile Phe Ile Ile Leu Tyr Phe Lys Leu Thr Lys
```

```
    210                 215                 220
<210>  52
<211>  221
<212>  PRT
<213>  Arabidopsis thaliana
<400>  52
Met Ser Asp Val Phe Asp Gly Tyr Glu Arg Gln Tyr Cys Glu Leu Ser
  1               5                  10                  15
Ala Ser Leu Ser Lys Lys Cys Ser Ser Ala Ile Ser Leu Asp Gly Glu
             20                  25                  30
Gln Lys Lys Gln Lys Leu Ser Glu Ile Lys Ser Gly Leu Glu Asn Ala
         35                  40                  45
Glu Val Leu Ile Arg Lys Met Asp Leu Glu Ala Arg Thr Leu Pro Pro
     50                  55                  60
Asn Leu Lys Ser Ser Leu Leu Val Lys Leu Arg Glu Phe Lys Ser Asp
 65                  70                  75                  80
Leu Asn Asn Phe Lys Thr Glu Val Lys Arg Ile Thr Ser Gly Gln Leu
                 85                  90                  95
Asn Ala Ala Ala Arg Asp Glu Leu Leu Glu Ala Gly Met Ala Asp Thr
            100                 105                 110
Lys Thr Ala Ser Ala Asp Gln Arg Ala Arg Leu Met Met Ser Thr Glu
        115                 120                 125
Arg Leu Gly Arg Thr Thr Asp Arg Val Lys Asp Ser Arg Arg Thr Met
    130                 135                 140
Met Glu Thr Glu Glu Ile Gly Val Ser Ile Leu Gln Asp Leu His Gly
145                 150                 155               160
Gln Arg Gln Ser Leu Leu Arg Ala His Glu Thr Leu His Gly Val Asp
                165                 170                 175
Asp Asn Ile Gly Lys Ser Lys Lys Ile Leu Thr Asp Met Thr Arg Arg
            180                 185                 190
Met Asn Lys Asn Lys Trp Thr Ile Gly Ala Ile Ile Ile Ala Leu Ile
        195                 200                 205
Ala Ala Ile Phe Ile Ile Leu Tyr Phe Lys Leu Thr Lys
    210                 215                 220
<210>    53
<211>    207
<212>    DNA
<213>    Ipomoea batatas
<220>
<221>    CDS
<222>    (1)..(207)
<223>    Coding sequence for partial sporamin comprising a lytic vacuole t
         argeting sequence
<400>    53
atg aaa gcc ctc gca ctc ttc ttc cta ctt tcc ctc tat ctc ctc ccc           48
Met Lys Ala Leu Ala Leu Phe Phe Leu Leu Ser Leu Tyr Leu Leu Pro
  1               5                  10                  15
aac cca gct cat tcc aag ttc aat ccc atc cgc ctc cgc ccc gcc cac           96
Asn Pro Ala His Ser Lys Phe Asn Pro Ile Arg Leu Arg Pro Ala His
             20                  25                  30
```

```
gaa acg gcg tcg tcc gaa act ccg gtg ctc gac atc aac ggc gac gaa         144
Glu Thr Ala Ser Ser Glu Thr Pro Val Leu Asp Ile Asn Gly Asp Glu
         35                  40                  45
gtc cgg gcc ggc gaa aat tac tac att gtc tcc gcc ata tgg ggc gcc         192
Val Arg Ala Gly Glu Asn Tyr Tyr Ile Val Ser Ala Ile Trp Gly Ala
     50                  55                  60
ggc gga gga ggc ctg                                                     207
Gly Gly Gly Gly Leu
 65
<210>  54
<211>  69
<212>  PRT
<213>  Ipomoea batatas
<400>  54
Met Lys Ala Leu Ala Leu Phe Phe Leu Leu Ser Leu Tyr Leu Leu Pro
  1               5                  10                  15
Asn Pro Ala His Ser Lys Phe Asn Pro Ile Arg Leu Arg Pro Ala His
             20                  25                  30
Glu Thr Ala Ser Ser Glu Thr Pro Val Leu Asp Ile Asn Gly Asp Glu
         35                  40                  45
Val Arg Ala Gly Glu Asn Tyr Tyr Ile Val Ser Ala Ile Trp Gly Ala
     50                  55                  60
Gly Gly Gly Gly Leu
 65
<210>    55
<211>    465
<212>    DNA
<213>    Homo sapiens
<220>
<221>    gene
<222>    (1)..(465)
<223>    Gene for partial EBD (C-terminal domain of human EEA1) comprising
         a phosphatidylinositol 3-phosphate targeting sequence
<220>
<221>    CDS
<222>    (1)..(462)
<223>    Partial EBD (C-terminal domain of human EEA1) comprising a phosph
         atidylinositol 3-phosphate targeting sequence
<400>    55
tgg caa tct agt caa cgg aga gtt agt gag ctt gag aaa caa acg gat          48
Trp Gln Ser Ser Gln Arg Arg Val Ser Glu Leu Glu Lys Gln Thr Asp
  1               5                  10                  15
gac tta cgg ggt gaa att gca gta tta gaa gca acg gtt cag aat aat          96
Asp Leu Arg Gly Glu Ile Ala Val Leu Glu Ala Thr Val Gln Asn Asn
             20                  25                  30
caa gat gaa agg aga gca cta ctg gaa aga tgt ctt aaa gga gaa ggt         144
Gln Asp Glu Arg Arg Ala Leu Leu Glu Arg Cys Leu Lys Gly Glu Gly
         35                  40                  45
gaa ata gaa aag ctt caa acc aaa gta tta gaa ttg caa aga aag ctg         192
Glu Ile Glu Lys Leu Gln Thr Lys Val Leu Glu Leu Gln Arg Lys Leu
```

```
     50                  55                  60
gat aat aca act gca gtg cag gag ctg ggc aga gaa aac caa tca ctt          240
Asp Asn Thr Thr Ala Val Gln Glu Leu Gly Arg Glu Asn Gln Ser Leu
 65                  70                  75                  80
cag atc aaa cat aca caa gcg ttg aat aga aag tgg gcc gaa gac aat          288
Gln Ile Lys His Thr Gln Ala Leu Asn Arg Lys Trp Ala Glu Asp Asn
                 85                  90                  95
gaa gta caa aac tgt atg gcc tgt ggg aaa ggc ttt tca gta aca gtg          336
Glu Val Gln Asn Cys Met Ala Cys Gly Lys Gly Phe Ser Val Thr Val
            100                 105                 110
aga cgg cat cac tgc cga cag tgt gga aat atc ttc tgt gct gaa tgt          384
Arg Arg His His Cys Arg Gln Cys Gly Asn Ile Phe Cys Ala Glu Cys
        115                 120                 125
tca gcc aaa aat gcc tta act cct tcc tcc aag aag cct gtt cgt gtc          432
Ser Ala Lys Asn Ala Leu Thr Pro Ser Ser Lys Lys Pro Val Arg Val
    130                 135                 140
tgt gat gca tgt ttc aat gac ttg caa gga       taa                        465
Cys Asp Ala Cys Phe Asn Asp Leu Gln Gly
145                 150
<210>  56
<211>  154
<212>  PRT
<213>  Homo sapiens
<400>  56
Trp Gln Ser Ser Gln Arg Arg Val Ser Glu Leu Glu Lys Gln Thr Asp
  1               5                  10                  15
Asp Leu Arg Gly Glu Ile Ala Val Leu Glu Ala Thr Val Gln Asn Asn
             20                  25                  30
Gln Asp Glu Arg Arg Ala Leu Leu Glu Arg Cys Leu Lys Gly Glu Gly
         35                  40                  45
Glu Ile Glu Lys Leu Gln Thr Lys Val Leu Glu Leu Gln Arg Lys Leu
     50                  55                  60
Asp Asn Thr Thr Ala Val Gln Glu Leu Gly Arg Glu Asn Gln Ser Leu
 65                  70                  75                  80
Gln Ile Lys His Thr Gln Ala Leu Asn Arg Lys Trp Ala Glu Asp Asn
                 85                  90                  95
Glu Val Gln Asn Cys Met Ala Cys Gly Lys Gly Phe Ser Val Thr Val
            100                 105                 110
Arg Arg His His Cys Arg Gln Cys Gly Asn Ile Phe Cys Ala Glu Cys
        115                 120                 125
Ser Ala Lys Asn Ala Leu Thr Pro Ser Ser Lys Lys Pro Val Arg Val
    130                 135                 140
Cys Asp Ala Cys Phe Asn Asp Leu Gln Gly
145                 150
<210>    57
<211>    438
<212>    DNA
<213>    Arabidopsis thaliana
<220>
<221>    gene
```

```
<222>    (1)..(438)
<223>    Gene for pleckstrin homology domain of Arabidopsis thaliana (AtPH
         ), a putative homologue of human pleckstrin domain, comprising a
         phosphatidylinositol 3-phosphate targeting sequence
<220>
<221>    CDS
<222>    (1)..(435)
<223>    pleckstrin homology domain of Arabidopsis thaliana (AtPH), a puta
         tive homologue of human pleckstrin domain, comprising a phosphati
         dylinositol 3-phosphate targeting sequence
<400>    57
atg gag agt atg tgg cga atc gcg acg gga caa gat ccg agc cgt gaa           48
Met Glu Ser Met Trp Arg Ile Ala Thr Gly Gln Asp Pro Ser Arg Glu
  1               5                  10                  15
gat tac gaa ggg atc gag ttc tgg tca aac cct gag cgt tct ggt tgg           96
Asp Tyr Glu Gly Ile Glu Phe Trp Ser Asn Pro Glu Arg Ser Gly Trp
             20                  25                  30
ctc aca aag caa ggc gat tac atc aaa acc tgg cgt cgt cgt tgg ttc          144
Leu Thr Lys Gln Gly Asp Tyr Ile Lys Thr Trp Arg Arg Arg Trp Phe
         35                  40                  45
gtt ctc aaa cga ggg aag ctt ctc tgg ttc aaa gat caa gcc gct gct          192
Val Leu Lys Arg Gly Lys Leu Leu Trp Phe Lys Asp Gln Ala Ala Ala
     50                  55                  60
gga att cgt gga tct acg ccg cgt ggt gtg atc tcc gtt ggt gat tgt          240
Gly Ile Arg Gly Ser Thr Pro Arg Gly Val Ile Ser Val Gly Asp Cys
 65                  70                  75                  80
ctc acc gtg aaa gga gct gag gat gtt gtg aat aag cct ttt gct ttt          288
Leu Thr Val Lys Gly Ala Glu Asp Val Val Asn Lys Pro Phe Ala Phe
                 85                  90                  95
gag cta tct agt ggt agc tat acc atg ttc ttc att gct gat aat gag          336
Glu Leu Ser Ser Gly Ser Tyr Thr Met Phe Phe Ile Ala Asp Asn Glu
            100                 105                 110
aag gag aaa gaa gag tgg att aat tcg att gga aga tcg att gtg caa          384
Lys Glu Lys Glu Glu Trp Ile Asn Ser Ile Gly Arg Ser Ile Val Gln
        115                 120                 125
cac tcg agg tct gtc acg gat tct gag gtt ctc gat tat gat cac agg          432
His Ser Arg Ser Val Thr Asp Ser Glu Val Leu Asp Tyr Asp His Arg
    130                 135                 140
cgg  tga                                                                 438
Arg
145
<210>  58
<211>  145
<212>  PRT
<213>  Arabidopsis thaliana
<400>  58
Met Glu Ser Met Trp Arg Ile Ala Thr Gly Gln Asp Pro Ser Arg Glu
  1               5                  10                  15
Asp Tyr Glu Gly Ile Glu Phe Trp Ser Asn Pro Glu Arg Ser Gly Trp
             20                  25                  30
```

```
Leu Thr Lys Gln Gly Asp Tyr Ile Lys Thr Trp Arg Arg Arg Trp Phe
         35                  40                  45
Val Leu Lys Arg Gly Lys Leu Leu Trp Phe Lys Asp Gln Ala Ala Ala
     50                  55                  60
Gly Ile Arg Gly Ser Thr Pro Arg Gly Val Ile Ser Val Gly Asp Cys
 65                  70                  75                  80
Leu Thr Val Lys Gly Ala Glu Asp Val Val Asn Lys Pro Phe Ala Phe
                 85                  90                  95
Glu Leu Ser Ser Gly Ser Tyr Thr Met Phe Phe Ile Ala Asp Asn Glu
            100                 105                 110
Lys Glu Lys Glu Glu Trp Ile Asn Ser Ile Gly Arg Ser Ile Val Gln
        115                 120                 125
His Ser Arg Ser Val Thr Asp Ser Glu Val Leu Asp Tyr Asp His Arg
    130                 135                 140
Arg
145
<210>    59
<211>    282
<212>    DNA
<213>    Mus musculus
<220>
<221>    CDS
<222>    (1)..(282)
<223>    Coding sequence for partial pleckstrin homology domain of FAPP (f
         amily A (phosphoinositide binding specific) member 3), comprising
         a phosphatidylinositol 4-phosphate targeting sequence
<400>    59
atg gag ggg gtt ctg tac aag tgg acc aac tat ctc aca ggt tgg cag           48
Met Glu Gly Val Leu Tyr Lys Trp Thr Asn Tyr Leu Thr Gly Trp Gln
  1               5                  10                  15
cct cga tgg ttt gtt ctg gat aat gga atc ctg tcc tac tat gac tca           96
Pro Arg Trp Phe Val Leu Asp Asn Gly Ile Leu Ser Tyr Tyr Asp Ser
             20                  25                  30
cag gat gat gtc tgc aaa ggg agc aaa ggg agt ata aag atg gcg gtc          144
Gln Asp Asp Val Cys Lys Gly Ser Lys Gly Ser Ile Lys Met Ala Val
         35                  40                  45
tgt gag att aaa gtc cat ccc gca gac aac aca aga atg gag tta atc          192
Cys Glu Ile Lys Val His Pro Ala Asp Asn Thr Arg Met Glu Leu Ile
     50                  55                  60
att cca gga gag cag cat ttc tac atg aag gca gta aat gcc gcc gag          240
Ile Pro Gly Glu Gln His Phe Tyr Met Lys Ala Val Asn Ala Ala Glu
 65                  70                  75                  80
aga cag agg tgg ctg gtt gcc ctt ggg agc tcc aaa gcg tgc                  282
Arg Gln Arg Trp Leu Val Ala Leu Gly Ser Ser Lys Ala Cys
                 85                  90
<210>  60
<211>  94
<212>  PRT
<213>  Mus musculus
<400>  60
```

```
Met Glu Gly Val Leu Tyr Lys Trp Thr Asn Tyr Leu Thr Gly Trp Gln
  1               5                  10                  15
Pro Arg Trp Phe Val Leu Asp Asn Gly Ile Leu Ser Tyr Tyr Asp Ser
             20                  25                  30
Gln Asp Asp Val Cys Lys Gly Ser Lys Gly Ser Ile Lys Met Ala Val
         35                  40                  45
Cys Glu Ile Lys Val His Pro Ala Asp Asn Thr Arg Met Glu Leu Ile
     50                  55                  60
Ile Pro Gly Glu Gln His Phe Tyr Met Lys Ala Val Asn Ala Ala Glu
 65                  70                  75                  80
Arg Gln Arg Trp Leu Val Ala Leu Gly Ser Ser Lys Ala Cys
                 85                  90
<210>    61
<211>    525
<212>    DNA
<213>    Rattus norvegicus
<220>
<221>    gene
<222>    (1)..(525)
<223>    Gene for partial pleckstrin homology domain of phospholipase C-de
         lta 1 comprising phosphatidylinositol 4,5-diphosphate targeting s
         equence
<220>
<221>    CDS
<222>    (1)..(522)
<223>    Partial pleckstrin homology domain of phospholipase C-delta 1 com
         prising phosphatidylinositol 4,5-diphosphate targeting sequence
<400>    61
gac tcg ggt agg gac ttc ctg acc ctg cac ggg ctc cag gat gac ccg           48
Asp Ser Gly Arg Asp Phe Leu Thr Leu His Gly Leu Gln Asp Asp Pro
  1               5                  10                  15
gac ctt cag gcc ctt ctg aag ggc agc cag ctt ctg aag gtg aag tcc           96
Asp Leu Gln Ala Leu Leu Lys Gly Ser Gln Leu Leu Lys Val Lys Ser
             20                  25                  30
agc tcg tgg cgt agg gaa cgc ttc tac aag cta cag gag gac tgc aag          144
Ser Ser Trp Arg Arg Glu Arg Phe Tyr Lys Leu Gln Glu Asp Cys Lys
         35                  40                  45
acc atc tgg cag gaa tct cga aag gtc atg agg tcc ccg gag tcg cag          192
Thr Ile Trp Gln Glu Ser Arg Lys Val Met Arg Ser Pro Glu Ser Gln
     50                  55                  60
ctg ttc tcc atc gag gac att cag gag gta cgg atg gga cac cgc aca          240
Leu Phe Ser Ile Glu Asp Ile Gln Glu Val Arg Met Gly His Arg Thr
 65                  70                  75                  80
gaa ggc ctg gag aag ttt gcc cga gac atc ccc gag gat cga tgc ttc          288
Glu Gly Leu Glu Lys Phe Ala Arg Asp Ile Pro Glu Asp Arg Cys Phe
                 85                  90                  95
tcc att gtc ttc aag gac cag cgc aac acc cta gac ctc att gcc cca          336
Ser Ile Val Phe Lys Asp Gln Arg Asn Thr Leu Asp Leu Ile Ala Pro
            100                 105                 110
tca cca gct gac gct cag cac tgg gtg cag ggc ctg cgc aag atc atc          384
```

```
Ser Pro Ala Asp Ala Gln His Trp Val Gln Gly Leu Arg Lys Ile Ile
        115                 120                 125
cac cac tcc ggc tcc atg gac cag cgg cag aag ctg cag cac tgg att          432
His His Ser Gly Ser Met Asp Gln Arg Gln Lys Leu Gln His Trp Ile
    130                 135                 140
cac tcc tgc ttg cga aag gct gat aaa aac aag gac aac aag atg aac          480
His Ser Cys Leu Arg Lys Ala Asp Lys Asn Lys Asp Asn Lys Met Asn
145                 150                 155                 160
ttc aag gag ctg aag gac ttc ctg aag gag ctc aac atc cag  tga             525
Phe Lys Glu Leu Lys Asp Phe Leu Lys Glu Leu Asn Ile Gln
                165                 170
<210>  62
<211>  174
<212>  PRT
<213>  Rattus norvegicus
<400>  62
Asp Ser Gly Arg Asp Phe Leu Thr Leu His Gly Leu Gln Asp Asp Pro
  1               5                  10                    15
Asp Leu Gln Ala Leu Leu Lys Gly Ser Gln Leu Leu Lys Val Lys Ser
             20                  25                  30
Ser Ser Trp Arg Arg Glu Arg Phe Tyr Lys Leu Gln Glu Asp Cys Lys
         35                  40                  45
Thr Ile Trp Gln Glu Ser Arg Lys Val Met Arg Ser Pro Glu Ser Gln
     50                  55                  60
Leu Phe Ser Ile Glu Asp Ile Gln Glu Val Arg Met Gly His Arg Thr
 65                  70                  75                  80
Glu Gly Leu Glu Lys Phe Ala Arg Asp Ile Pro Glu Asp Arg Cys Phe
                 85                  90                  95
Ser Ile Val Phe Lys Asp Gln Arg Asn Thr Leu Asp Leu Ile Ala Pro
            100                 105                 110
Ser Pro Ala Asp Ala Gln His Trp Val Gln Gly Leu Arg Lys Ile Ile
        115                 120                 125
His His Ser Gly Ser Met Asp Gln Arg Gln Lys Leu Gln His Trp Ile
    130                 135                 140
His Ser Cys Leu Arg Lys Ala Asp Lys Asn Lys Asp Asn Lys Met Asn
145                 150                 155                 160
Phe Lys Glu Leu Lys Asp Phe Leu Lys Glu Leu Asn Ile Gln
                165                 170
```