

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(43) 국제공개일
2010년 12월 9일 (09.12.2010)

PCT

(10) 국제공개번호
WO 2010/140774 A2

- (51) 국제특허분류:
G01N 33/68 (2006.01) G01N 27/62 (2006.01)
G06F 19/00 (2006.01)
- (21) 국제출원번호: PCT/KR2010/002745
- (22) 국제출원일: 2010년 4월 30일 (30.04.2010)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2009-0048024 2009년 6월 1일 (01.06.2009) KR
- (71) 출원인 (US 을(를) 제외한 모든 지정국에 대하여): 한국기초과학지원연구원 (KOREA BASIC SCIENCE INSTITUTE) [KR/KR]; 대전광역시 유성구 어은동 52, 305-333 Daejeon (KR).
- (72) 발명자: 권
- (75) 발명자/출원인 (US 에 한하여): 권요셉 (KWON, Joseph) [KR/KR]; 전라북도 전주시 완산구 서신동 동아한빛아파트 109-505, 560-793 Jeollabuk-do (KR). 박치열 (PARK, Chiyoul) [KR/KR]; 광주광역시 북구 오치동 957-5 화이트빌라 203, 500-837 Gwangju (KR). 노국필 (ROH, Gookpil) [KR/KR]; 전라남도 목포시 산정동 1112 신안비치 팔레스아파트 202-1205, 530-350 Jeollanam-do (KR). 이태훈 (LEE, Taehoon) [KR/KR]; 광주광역시 북구 일곡동 쌍용아파트 101-1303, 500-793

Gwangju (KR). 최중순 (CHOI, Jongsoon) [KR/KR]; 대전시 유성구 어은동 99 한빛아파트 119-301, 305-755 Daejeon (KR).

(74) 대리인: 김원준 (KIM, Wonjoon); 대전시 서구 만년동 386 골드벤처타워 205 호, 302-823 Daejeon (KR).

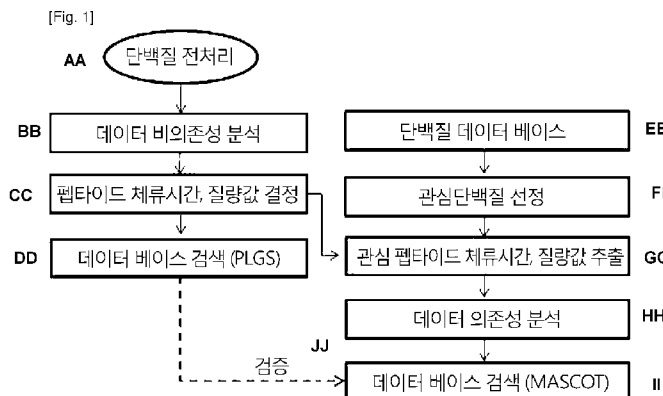
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 유럽 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[다음 쪽 계속]

(54) Title: PROTEIN ANALYSIS METHOD IN WHICH A DATA-INDEPENDENT ANALYSIS METHOD AND A DATA-DEPENDENT ANALYSIS METHOD ARE COMBINED

(54) 발명의 명칭: 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 분석방법



AA ... Protein pretreatment
BB ... Data-independent analysis
CC ... Determination of peptide retention time and mass value
DD ... Database search (PLGS)
EE ... Protein database
FF ... Select a protein of interest
GG ... Extract the retention time and mass value of a peptide of interest
HH ... Data-dependent analysis
II ... Database search (MASCOT)
JJ ... Verification

(57) Abstract: The present invention relates to a method for analyzing proteins using a mass spectrometer, and to a qualitative and quantitative protein analysis method in which a data-independent analysis method and a data-dependent analysis method are combined, wherein said qualitative and quantitative protein analysis method comprises: (A) a step of pretreating a protein mixture; (B) a step of acquiring information on a peptide retention time and on a mass value of the peptide through data-independent analysis; (C) a step of searching a database (PLGS) on the basis of the information on the peptide retention time and on the mass value of the peptide, and performing a first qualitative and quantitative analysis on the protein; (D) a step of extracting information on a predetermined protein from the information on a peptide retention time and on a mass value of the peptide; (E) a step of conducting a data-dependent analysis using the information extracted in step (D); (F) a step of searching a database (MASCOT) on the basis of the information on a peptide retention time and on a mass value of the peptide, acquired in step (E), and performing a second qualitative and quantitative analysis on the protein; and (G) a step of comparing the result of step (C) and the result of step (F) to verify the qualitative and quantitative analysis on the protein.

(57) 요약서:

[다음 쪽 계속]

WO 2010/140774 A2

**공개:**

- 국제조사보고서 없이 공개하며 보고서 접수 후 이를 별도 공개함 (규칙 48.2(g))

본 발명은 질량분석기를 이용하여 단백질을 분석하는 방법에 관한 것으로서, (A) 단백질 혼합물을 전처리하는 단계; (B) 데이터 비의존성 분석을 통해 펩타이드의 체류시간 및 질량값 정보를 획득하는 단계; (C) 상기 체류시간 및 질량값 정보를 바탕으로 데이터베이스(PLGS)를 검색하여 1 차로 단백질을 정량정성하는 단계; (D) 상기 체류시간 및 질량값 정보로부터 소정의 단백질에 관한 정보를 추출하는 단계; (E) 상기 (D)에서 추출된 정보를 활용하여 데이터 의존성 분석을 수행하는 단계; (F) 상기 (E)로부터 얻어진 체류시간 및 질량값 정보를 바탕으로 데이터베이스(MASCOT)를 검색하여 상기 단백질을 2 차로 정량정성하는 단계; 및 (G) 상기 (C)의 결과와 상기 (F)의 결과를 비교분석하여 상기 단백질의 정량정성을 검증하는 단계; 를 포함하는 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법에 관한 것이다.

명세서

발명의 명칭: 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 분석방법

기술분야

- [1] 본 발명은 질량분석기를 이용하여 단백질을 분석하는 방법에 관한 것으로서, 보다 상세하게는 데이터 비의존성 분석법(MSE)으로 분석된 단백질을 데이터 의존성 분석법(DDA)으로 정밀하게 재검증하여 단백질을 분석하는 방법에 관한 것이다.

[2]

배경기술

- [3] 프로테오믹스(proteomics; 단백질체학)는 세포 혹은 조직에서 어떤 단백질들이 발현되며 어떤 변화를 보여주는지 종합적으로 연구하는 분야이다. 프로테오믹스는 90년대 이후 질량분석법의 획기적 발전과 더불어 단백질들의 아미노산 서열에 대한 데이터베이스가 구축되고, 이의 활용이 가능해지면서 시작되었다.
- [4] 개별 단백질을 독립적으로 분석하던 통상의 단백질 생화학과는 달리 프로테오믹스는 연구대상의 볼륨, 속도, 분리수단의 자동화 및 게놈/단백체 DB정보와의 연동성 측면에서 매우 다른 특성을 지닌다. 프로테오믹스는 세포내 전체단백질을 연구하는 대형 스케일의 다단계 고속 분석 기술이므로 연구대상도 단백질의 발현(expression), 기능(function), 구조(structure) 및 생합성후 구조변형(posttranslational modification: PTM)과 관련 단백질간의 결합성(protein-protein interaction)에 초점을 두고 있기 때문에 유전체학보다 복잡하며 데이터양도 방대하다. 프로테오믹스에 의하면 세포의 생리상태변화에 따른 결합성과 기능을 분석하고 파악할 수 있다. 따라서 이를 활용하여 유전자 정보만으로는 알 수 없는 단백질의 이성구조(isoform), 인산화 등의 수식화(post translational modification)와 결합 파트너 등을 분석함으로써 질병의 발생기작을 연구하고, 그의 진단이나 치료에 직접 활용할 수 있게 된다.
- [5]
- [6] 일반적으로 프로테오믹스에서는, 세포 내에서 분리된 단백질 혼합물을 소정의 방법으로 절단하여 펩타이드로 만들고 질량분석기를 이용하여 펩타이드의 질량스펙트럼을 얻고, 이 결과와 기존의 데이터베이스를 비교하는 방법으로 단백질을 정량정성분석하게 된다. 즉, 질량분석기로부터 얻어진 데이터와 데이터 은행(NCBI, EXPASY, ETS 등)에 등록된 각 단백질들의 서열을 이용하여 가상적인 단편화를 통하여 예측된 데이터들을 비교 검토함으로써 시료 내에 존재하는 단백질들을 동정하는 것이다. 이에 의하면 검색 결과로 바로 유전체 및 유전자 서열 데이터 은행으로 연결되어 유전자정보를 얻을 수 있을 뿐만 아니라,

데이터 은행에 등록되는 단백질 정보량이 기하급수적으로 증가하고 있으므로 매우 유용하다.

[7]

[8] 질량분석기는 이온화원과 질량분석관(검출기)에 따라 여러 가지 이름으로 불린다.

[9]

시료 단백질이나 펩타이드를 이온화시키는 방법으로는 전기분무이온화법(electrospray ionization; ESI)과 매트릭스상레이저탈리이온화법(matrix assisted laser desorption ionization; MALDI)이 대표적으로 이용되고 있다. ESI는 액체상의 시료를 이온화시키는 방식으로 액체크로마토그래피 같은 분리 방법과 직접 연결이 용이하다. MALDI는 매트릭스와 시료를 혼합한 후 건조시켜 결정으로 만든 다음 레이저에 의해 이온화가 일어나도록 한다.

[10]

현재 널리 사용되는 질량분석관으로 이온트랩(ion trap), 비행시간형(time of flight; TOF), 사극자형(quadrupole; Q), 그리고 푸리에 변환이온공명형(fourier transform ion cyclotron resonance; FT-ICR) 등이 있는데, 단독으로 또는 두 개 이상을 조합하여 탄뎀(직렬)질량분석기 형태로 사용된다.

[11]

탄뎀질량분석기 중 삼중사극자는 사극자형 분석관 3개(Q_1 , Q_2 , Q_3)를 직렬로 연결하며 만들어진 분석기이다. 가운데 위치하는 Q_2 에서는 주입된 중성가스가 시료이온과 충돌하여 이온들이 단편화된다. 삼중사극자는 스캔모드와 단편화모드 두 가지 방식으로 운전된다. ① 스캔모드에서는 Q_1 분석기만이 작동되어 주입된 모든 m/z 값을 지나는 이온들이 기록되며, 1초 이내에 전체 이온의 질량분석이 가능하다. ② 단편화모드에서는 Q_1 , Q_2 , Q_3 모두 사용된다. Q_1 (질량필터)은 사극자에 가해지는 전압이 어느 특정 m/z 값만을 가지는 이온만 통과되도록 조절(필터링)되고 통과된 이온은 충돌실(Q_2)로 넘어간다. 충돌실로 들어온 이온은 아르곤 가스와 충돌하여 단편화된다. 단편화된 이온들은 Q_3 로 들어가 질량 대 전하비로 분리되어 검출기에 기록된다.

[12]

이러한 삼중사극자 분석기를 활용한 분석론 중의 하나가 데이터의존성분석(Data Dependent Analysis; DDA)법이다. DDA법은, 스캔모드(MS)로 시료 중의 모든 펩타이드 이온에 대한 질량 대 전하량(m/z) 값을 얻고, 이어서 단편화모드(MS/MS)로 펩타이드를 단편화하여 단편화된 이온에 대한 질량 대 전하량(m/z) 값을 얻는 방식이다. 이때 MS와 MS/MS가 교차되어 데이터(스펙트럼)를 생성하게 된다.

[13]

DDA 분석법은 정확한 체류시간(retention time)과 질량값 (m/z)의 정보를 입력하면 주어진 정보에 해당하는 시료중의 물질만 분석할 수 있는 장점이 있다. 그러나 펩타이드의 이온의 관찰된 크기가 큰 물질들이 분석될 확률이 높아 적은 양의 펩타이드의 경우 (절편을 생성하지 못하여) 분석되지 못하는 단점이 있다.

[14]

[15]

한편, 최근 DDA와는 개념이 다른 펩타이드 정보를 얻는 방법론으로 고충돌

에너지와 저충돌에너지를 동시에 적용시키는 데이터 비의존성 분석법(High/low collision energy MS ; MS^E)가 알려졌다. 이 또한 삼중극자 분석기에서 적용되며, DDA가 데이터 의존성 분석법이라면 MS^E는 데이터 비의존성 분석법이라고 할 수 있다. MS^E법은 단위 시간에 들어온 모든 펩타이드를 동시에 충돌기체와 충돌시켜 단편화시키고, 혼합된 펩타이드 절편 정보를 펩타이드의 액상 크로마토그래피의 체류시간과 얻어진 질량값의 패턴을 연동시켜 분석에 활용될 MS/MS 스펙트럼 정보를 생성하는 방식이다.

- [16] 이러한 MS^E법은 DDA법과는 달리, 관찰된 이온의 높이와는 무관하게 펩타이드 절편을 생성하므로 상대적으로 적은 양의 펩타이드 분석에 보다 유리하다. 그러나 MS^E법은 Waters사의 PLGS(Proteinlynx Global Searver)에서만 단백질을 분석할 수 있으며, 연구자들이 가장 많이 이용하는 MASCOT 등에는 적합하지 못한 형식이라는 단점이 있다. 그러나 시료 중의 함량이 미량인 단백질이라도 분석할 수 있다는 강력한 장점을 가지고 있다. 예를 들어, 혈액 단백질은 23종의 단백질이 98%를 차지하며 나머지 2%에 우리가 보고자 하는 바이오마커가 있는 것으로 생각하고 있다. 이러한 미량의 단백질을 분석하기 위해서는 양적으로 많은 단백질을 제거하여 미량의 단백질을 농축하는 과정을 거쳐야 한다. 하지만 혈액샘플은 많이 얻을 수 없어 농축도 한계가 있다. 또한, 막단백질의 경우 대부분 양적으로 풍부한 세포질 내부의 단백질이 오염되어 막단백질 분석을 방해하게 된다. 이는 기존에 주로 활용하던 DDA 방법이 양적으로 우세한 펩타이드를 먼저 분석하기 때문에 겪는 어려움인 것이다. DDA 방법을 사용하여 분석하더라도 농축되어 있으며 컬럼에서 고효율로 분리가 일어난다면 좋은 분석결과를 기대할 수 있다. 그러나 여러 가지 방법들의 개발에도 불구하고 미량 단백질의 분석 및 검증은 우리가 겪는 난관중의 하나이다.

- [17] 본 발명에서 제안된 MS^E-DDA 방법은 개별적으로 MS^E와 DDA 분석법으로 이미 알려져 있는 방법이지만 두 분석법의 장단점을 극복하여 미량 단백질과 미량 단백질의 화학적 변화를 분석, 검증할 수 있을 것이다. 이는 MS^E의 양적인 차별 없는 펩타이드 분석과 DDA의 높은 선택적 분석의 조화로 가능하다. MS^E로부터 최적화된 최소한의 정보를 DDA로 줌으로써 보다 확실한 단백질의 정보를 검증할 수 있는것이다.

[18]

발명의 상세한 설명

기술적 과제

- [19] 단백질의 정확한 동정을 위해서는, 특정단백질에 반응하는 항체(antibody)를 이용하는 것이 일반적이다. 그러나 항체는 고가일 뿐만 아니라, 모든 단백질이 항체가 개발된 것이 아니기 때문에 단백질의 확인동정에 사용되기 부적합하다.
- [20] 본 발명은 질량분석기를 이용하여 수식화된 단백질을 신속하고도 용이하게 검출하고 정성분석할 수 있는 방법을 제공하고자 한다.

[21] 또한 본 발명은 단백질 데이터베이스를 활용한 정보와 질량분석기를 이용하여 샘플내에 미량 존재하는 단백질을 신속하고도 용이하게 검출하고 동정할 수 있는 방법을 제공하고자 한다.

[22]

과제 해결 수단

[23] 전술한 과제를 해결하기 위한 본 발명은, 질량분석기를 이용한 단백질 정량정성분석 방법으로서, **(A)** 단백질 혼합물을 전처리하는 단계; **(B)** 데이터 비의존성 분석을 통해 펩타이드의 체류시간 및 질량값 정보를 획득하는 단계; **(C)** 상기 체류시간 및 질량값 정보를 바탕으로 데이터베이스(PLGS)를 검색하여 1차로 단백질을 정량정성하는 단계; **(D)** 상기 체류시간 및 질량값 정보로부터 소정의 단백질에 관한 정보를 추출하는 단계; **(E)** 상기 (D)에서 추출된 정보를 활용하여 데이터 의존성 분석을 수행하는 단계; **(F)** 상기 (E)로부터 얻어진 체류시간 및 질량값 정보를 바탕으로 데이터베이스(MASCOT)를 검색하여 상기 단백질을 2차로 정량정성하는 단계; 및 **(G)** 상기 (C)의 결과와 상기 (F)의 결과를 비교분석하여 상기 단백질의 정량정성을 검증하는 단계;를 포함하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법에 관한 것이다.

[24]

[25] 또한 본 발명은, 질량분석기를 이용한 단백질 정량정성분석 방법으로서, **(A)** 단백질 데이터베이스를 참조하여 관심 단백질을 선정하는 단계; **(B)** 단백질 데이터베이스를 참조하여 상기 관심 단백질 펩타이드의 이론상 체류시간 및 질량값을 추정하는 단계; **(C)** 상기 추정된 정보를 활용하여 데이터 의존성 분석을 수행하는 단계; 및 **(D)** 상기 (C)의 결과를 기초로 상기 단백질 데이터베이스를 검색하여 분석된 단백질이 상기 (A)에서 선정된 관심 단백질인지를 확인하는 단계;를 포함하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법에 관한 것이다.

[26]

[27] 본 발명에 있어서, 상기 질량분석기는 삼중사극자 질량분석기인 것이 바람직하다.

[28] 또한 본 발명에서, 상기 단백질은 세포내에 미량 존재하는 단백질, 예를 들면 막단백질일 수 있다.

[29] 본 발명에서 상기 단백질은 번역후 수식화(PTM)된 단백질, 예를 들면 시스테인이 함유된 단백질일 수 있다.

[30]

[31] 또한 본 발명은, 상기 단백질 정량정성 분석방법을 수행하는 프로그램 자체 및 상기 프로그램이 저장된 저장매체에 관한 것이다.

[32]

[33] 다시 말해 본 발명은, 도 1과 같이, 1차로 데이터 비의존성 분석법(MSE)으로 분석된 단백질을 데이터 의존성 분석법(DDA)으로 정밀하게 재검증하여 단백질을 분석하는 방법에 관한 것이다.

[34] 또한 본 발명은, 1차로 데이터 비의존성 분석법(MSE)으로 분석된 단백질을 데이터 의존성 분석법(DDA)으로 정밀하게 재검증하여 단백질을 분석하는 방법을 수행할 수 있는 프로그램에 관한 것이다.

[35]

발명의 효과

[36] 데이터 비의존성 방법으로부터 생성된 펩타이드 절편은 샘플 중에 상대적으로 적은 양이 존재하더라도 그 정보를 줄 수 있는 방법이다. 본 발명은 기존에 개발된 데이터 비의존성 방법으로 분석된 결과를 미리 계산된 생물학적 정보와 비교하여 분석하고자 하는 펩타이드들의 정보를 얻고자 하는 것이다. 또한, 얻어진 정보를 데이터 의존성 분석 모드에 대입하여 우리가 원하는 펩타이드 절편을 생성시켜 단백질을 분석, 검증하는데 활용할 수 있게 한다.

[37] 이때 사용하는 생물학적 정보는 주로 상대적으로 양이 적은 단백질에 효과적인 것으로 확인 되었다. 단백질이 상대적으로 적게 존재하게 되면 기존에 많이 사용하는 데이터 의존성 방법으로는 분석이 제대로 되지 않을 확률이 크기 때문에 본 발명된 방법을 사용하게 되면 상대적으로 분석 열세에 있는 단백질을 분석할 수 있게 된다.

[38] 따라서, 본 발명에 의하면,

[39] 질량분석기를 이용하여 수식화된 단백질을 신속하고도 용이하게 검출하고 정성분석할 수 있으며,

[40]

[41] 또한 단백질 데이터베이스 정보와 질량분석기를 이용하여 샘플내에 미량 존재하는 단백질을 신속하고도 용이하게 검출하고 정량, 정성 분석할 수 있게 된다.

[42]

[43] 본 발명에 의한 방법은, 산업적, 학문적 중요성을 가진 단백질의 화학변형 즉, 단백질의 PTM(post translational modification) 정보를 빠르고 효과적으로 얻을 수 있게 한다. 이러한 정보는 세포 신호전달 연구, 의약품 개발 등의 중요한 정보로 활용될 수 있다.

[44]

도면의 간단한 설명

[45] 도 1은 본 발명에 의한 단백질 동정 및 검증 과정을 보여주는 흐름도.

[46] 도 2는 본 발명의 방법에 따라 시스테인을 포함한 펩타이드 정보를 추출한 include list의 적용상황을 보여주는 프로그램 화면(A)과 상기 추출한 include list의 분포도를 보여주는 도표(B).

- [47] 도 3은 본 발명에 의한 실시예에서, include list를 DDA 모드에 적용시켜 TIC한 결과를 보여주는 도표.
- [48] 도 4는 본 발명에 의해 검색된 단백질 수와, 종래 방법에 의해 검색된 단백질 수를 비교하여 보여주는 다이어그램.
- [49] 도 5는 본 발명의 실시예에서, 막단백질에 대한 include list의 분포도를 보여주는 도표.
- [50] 도 6은 본 발명의 실시예에 의해 분석된 막단백질의 정보와, 종래 기술에 의해 분석된 막단백질의 정보를 비교해서 보여주는 도표.
- [51] 도 7은 특정 단백질 Slr0906을 검색할 때, 본 발명에 의한 방법(MSE-DDA)과 종래 방법(DDA)에서 활용하는 펩타이드 정보의 차이를 보여주는 도표.

[52]

발명의 실시를 위한 최선의 형태

- [53] 이하 실시예를 들어 본 발명을 보다 상세히 설명한다. 그러나 이러한 실시예는 본 발명의 기술적 사상의 내용과 범위를 쉽게 설명하기 위한 예시일 뿐, 이에 의해 본 발명의 기술적 범위가 한정되거나 변경되는 것은 아니다. 또한 이러한 예시에 기초하여 본 발명의 기술적 사상의 범위 안에서 다양한 변형과 변경이 가능함은 당업자에게는 당연할 것이다.

[54]

- [55] 특정 체류시간과 질량값 정보를 포함한 목록을 "Include list"라 부르며 이 정보는 DDA 분석모드에 활용되어 특정 단백질의 분석을 위해 사용될 것이다.

- [56] 본 발명에서는 MSE로부터 얻어진 단백질 정보로부터 그 단백질을 분석하기 위해 사용되었던 펩타이드의 체류시간과 질량값을 취하여 include list를 편리하게 만들 수 있는 프로그램을 제작하였다. MSE 분석으로부터 얻어진 단백질들에 어떤 의미를 부여하느냐에 따라 재검증하고자 할 때 사용되는 include list는 달라지게 된다.

[57]

- [58] 실시예 : 시스테인에 특정 화학적 변화 분석

[59]

- [60] 목적하는 실험이 아미노산중 시스테인에 특정 화학적 변화를 관찰 하는 실험이라면 MSE로부터 얻어진 단백질 정보 중에서 시스테인을 포함하는 단백질을 선별하게 되고 이를 분석할 때 사용되었던 펩타이드 정보를 모아 include list를 생성시킬 수 있다.

[61]

- [62] (1) 단백질 전처리

- [63] 많은 단백질은 S-S 공유결합을 가지고 있다. 이는 시스테인과 시스테인이 연결되는 것이다. 특정 조건 즉, 병적인 조건에서 단백질의 S-S 결합이 깨지게 되는데 이를 확인하기 위하여 두 가지 화학물질로 공유결합 시켜 샘플을

만들었다. 샘플에 Iodoacetamide를 처리하게 되면 시스테인에 +57.02 Da의 질량차가 나며 N-ethyl maleimide(NEM)를 처리하면 +111.03 Da의 질량차가 나게 된다.

- [64] 단백질 샘플에 Iodoacetamide를 처리한 후 S-S 결합을 깨는 물질인 DTT (dithiothreitol)을 처리하였다. 그 후 NEM을 처리하게 되면 처음부터 S-S 결합이 깨져있는 단백질과 그렇지 않은 단백질을 구별할 수 있게 된다.

[65]

[66] (2) 데이터 비의존성 분석 및 데이터베이스 검색

- [67] Nano-UPLC와 Synapt HDMS tandem mass spectrometry(Waters)를 연결한 nano-UPLC-MS^E mode에서 시료를 분석하였다. 분석조건은 다음 표와 같다.

[68]

컬럼	내경 75 μ m×25cm 10min gradient
최종유속 및 압력	350nL/min 및 8000psi
ramping 조건	-low collision energy는 4eV -high collision energy는 15~40 eV
질량값을 보정하기 위하여 200 fmol/ μ L Glu-fibrinopeptide(785.8426 Da [M+2H] ²⁺)를 500 nL/min 속도로 30초 마다 사용	

- [69] 3중실험으로 얻은 기초정보(raw data)는 PLGS에서 처리되어 펩타이드와 fragmentation tolerance는 automatic mode로 sprot database 이용하여 단백질을 검색하였다.

[70]

[71] (3) EMRT 테이블 작성 및 include list 결정

- [72] MS^E 실험을 통하여 생성된 EMRT 정보 중 Cystein을 포함한 펩타이드 서열을 포함한 단백질에 대해 체류시간과 이온차수를 계산하여 include list를 작성하였다(도 2 참조).

[73]

[74] (4) 데이터 의존성 분석

- [75] 상기 include list를 DDA 모드에 적용시켜 도 3과 같은 TIC(Total Ion Chromatography) 결과를 얻었다. DDA 실험에 사용된 LC 전개용매 및 유속은 데이터 비의존성 실험과 동일하게 하였다. 각 샘플은 autosampler를 통하여 5 μ l씩 주입하였으며 C18 trapping column에서 염을 제거하고 농축시켰다. 내부표준물질로 100fmol/ml glu-fibrinopeptide B를 600nL/min의 속도로 주입하여 이온화하였다. 질량분석은 V 모드로 m/z 50~1990 영역을 스캔하고 최대 3개의 전구이온을 단편화시키는 것으로 프로그래밍하였다.

[76] 도 3에서, 첫 번째~세 번째 그래프는 include list에 존재하는 질량값을 골라 단편화한 결과를 보여주는 것이고, 네 번째 그래프는 150분간 처리(크로마토그래피)한 결과를 보여주는 TIC이다.

[77]

[78] (5) 데이터베이스 검색(재검증)

[79] MASCOT v 2.2 프로그램을 이용하여 단백질 데이터베이스는 IPI_mouse_v3.44.fasta를 사용하였다. Carbamidomethylation(C) 와 N_ethylmaleimide를 variable modification 으로 하고 펩타이드 tolerance 100ppm, ms/ms tolerance 0.2Da으로 검색하였다(도 4).

[80] 도 4에서, MS^E 결과는 PLGS로 분석하였으며 MS^E-DDA와 DDA는 MASCOT을 이용하여 검색한 것이다. 도에서 볼 수 있듯이, MS^E방법으로 검색된 단백질 중 Cystein을 포함한 부분만 정보를 추출하여 분석하였을 경우 88개의 단백질을 확인 할 수 있었고, 이는 DDA 모드로만 분석하였을 때와는 상당히 다른 양상을 보인다. 또한, 본 발명에서 목적했던 시스테인에 화학적으로 표지된 N-ethyl maleimide는 이 높은 score로 확인되어 특정 PTM 분석에 충분히 활용될 수 있음을 확인하였다. 이는 MS^E에서 얻은 EMRT table이 신뢰할만하며 이 정보를 바탕으로 include list 자동 생성은 유효하게 작동함을 알 수 있다.

[81] 도 7에서 확인할 수 있듯이, 정확한 정보를 제공받은 데이터 의존성 분석법(MS^E-DDA)을 활용할 경우 DDA 모드에서는 분석되지 못한 40개의 시스테인의 변형정보를 포함한 단백질을 확인할 수 있었으며 이는 DDA모드만을 사용했을 때는 분석할 수 없는 단백질이다.

[82] 데이터 비의존성 분석을 통하여 얻어진 정보를 미량 단백질에 활용하여 검증한 예이며 이 방법을 활용한다면 단백질의 화학적 변형정보를 보다 정확하게 얻을 수 있는 좋은 방법이 될 것이다.

[83]

[84] 실시예 : 막단백질의 분석 및 재검증

[85]

[86] 산업적, 학문적 중요성을 가진 막단백질을 질량분석기를 이용하여 분석하는데 어려움을 겪는 이유는 상대적인 양에 있다. 본 발명에서는 이점에 착안하여 상대적으로 적은 양으로 존재하는 막단백질을 데이터 비의존성 방법으로 분석한 후 막단백질 정보만 추출하여 데이터 의존성 분석을 할 수 있도록 하였으며 이를 통하여 존재하고 있는 보다 높은 신뢰도로 분석하여 검증할 수 있도록 하였다.

[87] 분석하고자 하는 단백질이 막단백질이라면 사용하는 단백질 데이터 베이스 중 막단백질을 미리 예측을 한 후 그 목록과 비교하여 include list 를 생성하는 방법을 사용할 수 있다. 본 실시예를 통하여 본 발명이 양적으로 열세에 있는 단백질 혼합물의 분석에 응용할 수 있음을 알 수 있다.

[88]

[89] (1) 데이터베이스 검색 및 막단백질 예측

[90] Synechocytosis 단백질 데이터베이스에 총 3661개의 단백질 정보가 있으며 이를 TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>)과 Signal P 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>)을 이용하여 총 706 개의 막단백질 정보를 추출하였다.

[91] 추출한 막단백질 정보는 text file 형태로 저장하여 사용하였다.

[92]

slr1450, slr1456, slr1708, sl11942, slr2013, slr2010sl10002, sl11021, ssr2422, sl11158 sl11155, slr0533, slr1895, slr0678, slr2003, sm10005, slr2016, ssr0550, ssl6077, slr1187 sl10498, ssr3307.....이하생략

[93]

[94] (2) 데이터 비의존성 분석 및 데이터베이스 검색

[95] Nano-UPLC와 Synapt HDMS tandem mass spectrometry(Waters)를 연결한 nano-UPLC-MS^E mode에서 시료를 분석하였다. 분석조건은 다음 표와 같다.

[96]

컬럼	내경 75 μ m×25cm 10min gradient
최종유속 및 압력	350nL/min 및 8000psi
ramping 조건	-low collision energy는 4eV -high collision energy는 15~40 eV
질량값을 보정하기 위하여 200 fmol/ μ L Glu-fibrinopeptide(785.8426 Da [M+2H] ²⁺)를 500 nL/min 속도로 30초 마다 사용	

[97] 세 번 반복 실험으로 얻은 펩타이드의 질량 정보가 포함된 기초정보(raw data)는 PLGS에서 처리되어 sprout database에 대하여 단백질을 검색하였다. 단백질 검색 조건은 fragment tolerance 100ppm, MS/MS tolerance 0.1Da, Enzyme trypsin, Missed cleavages 1, Fixed modification Carbamidomethylation (C), Variable modification Oxidation (M) 이다.

[98]

[99] (3) EMRT 테이블 작성 및 include list 결정

[100] 막단백질로 예측된 gene index와 데이터 비의존성 데이터 (EMRT 테이블)를 비교하여 해당되는 막단백질을 분석하기 위해 사용된 펩타이드의 체류시간과 차수 등을 추출하여 데이터 의존성 분석을 할 수 있도록 include list를 생성하였다.

[101] include list의 자동생성을 위한 프로그램을 아래에 예시하였다.

[102] <include list 생성 프로그램의 예>

[103]

```
#!/usr/bin/perl
use strict;
use warnings;
use Getopt::Long;
my $usage = "$0 -i inputfile -o outputfile -e parameter[0|1]\n";
my ($inputfile, $listfile, $outfile, $param);
GetOptions('i=s'=>\$inputfile, 'l=s'=>\$listfile, 'o=s'=>\$outfile);
die "\n$usage" if ( not defined $inputfile or not defined $outfile or not
defined $listfile);
die "\nCheck files\n" if ( not -e $inputfile or not -e $listfile);
my @list;
open FILE, $listfile;
for (<FILE>) {
    chomp;
    push @list, $_;
}
close FILE;
open OFILE, ">$outfile";
print OFILE "Mass,RT(second),charge\n";
open FILE, $inputfile;
my $tmp = <FILE>;
for ( <FILE> ) {
    chomp;
    my @a = split /\./, $_;
    foreach my $pro ( @list ) {
        if ( $a[5] =~ /$pro/ ) {
            print OFILE "$a[2],", $a[3]*60, ",1\n";
            print OFILE ($a[2]+1.0078)/2, ", ", $a[3]*60, ",2\n";
            print OFILE ($a[2]+2*1.0078)/3, ", ", $a[3]*60, ",3\n";
            last;
        }
    }
}
close FILE;
close OFILE;
```

[104] 위에서 작성된 프로그램을 이용하여 도 5와 같은 분포의 include list를 추출하게 되었다. 도 5는 추출된 분포를 보여주는 것이고 실제로는 텍스트 파일 형태로 활용하게 된다.

[105] 도 5에서, x축은 분석컬럼에서 머무름 시간이며 y축은 질량값이다. 그래프 상에서 각 점은 막단백질에서 유래된 해당 펩타이드의 체류시간과 질량값을 나타낸다.

[106]

[107] (4) 데이터 의존성 분석

[108] 분석조건은 다음 표와 같다.

[109]

컬럼	내경 75 μ m×25cm 10min gradient
최종유속 및 압력	350nL/min 및 8000psi
ramping 조건	-low collision energy는 4eV -high collision energy는 15~40 eV
질량값을 보정하기 위하여 200 fmol/ μ L Glu-fibrino펩타이드(785.8426 Da [M+2H] ²⁺)를 500 nL/min 속도로 30 초 마다 사용	

[110] 데이터 의존성 실험에 사용된 LC 전개용매 및 유속은 데이터 비의존성 실험과 동일하게 하였다. 각 샘플은 autosampler를 통하여 5 μ l씩 주입하였으며 C18 trapping column에서 염을 제거하고 농축시켰다. 내부표준물질로 100fmol/ml glu-fibrino 펩타이드 B를 600nL/min의 속도로 주입하여 이온화하였다. 질량분석은 V 모드로 m/z 50~1990 영역을 스캔하고 최대 3개의 전구이온을 단편화시키는 것으로 프로그래밍하였다.

[111]

[112] (5) 데이터베이스 검색(재검증)

[113] 본 발명에 의한 방법(MSE-DDA 분석법) 및 종래 기술에 의한 방법(MSE, DDA 분석법)에 따라 막단백질을 분석하였다(도 6).

[114] 도 6에서, 데이터 비의존성 분석법에 의해 분석된 막단백질정보를 x 축으로 볼 때 검정색 막대그래프는 Include list 정보를 가지고 분석한 데이터 의존성 분석이며 붉은색 막대 그래프는 Include list 정보를 가지고 있지 않은 상태의 데이터 의존성 분석이다.

[115] MSE-DDA는 그래프에서와 같이 보다 높은 데이터베이스 스코어를 나타내어 두배 이상의 향상이 있다. 이는 보다 정확한 시간에 펩타이드 정보를 주기 때문에 양적인 손실 없이 실험이 진행되기 때문이다. 뿐만 아니라 분석된 단백질 개수 또한 MSE-DDA 가 DDA 보다 높은 것을 알 수 있다.

[116] MSE로 분석(X 축)은 되었지만 MSE-DDA에서 분석되지 않은 단백질들은 양적으로 적게 분포되어 있는 것을 확인하였으며 정확한 펩타이드 분석 정보 없어 그 신뢰도가 낮다고 할 수 있다.

[117]

[118] 구체적인 예로서, 단백질 Slr0906(Galactose mutarotase and related enzymes)을 찾기 위해 MSE-DDA방법과 DDA방법에서 펩타이드 정보를 어떻게 인식하고 있는지 확인할 수 있다(도 7).

[119]

[120] 도 7에서 A는, MS^E-DDA를 통한 분석으로서 8개의 펩타이드 정보를 활용한 것으로 해당되는 펩타이드의 SIC(Selected Ion Chromatography) 결과이다. 도 7에서 B는, 같은 단백질을 분석하기 위해 DDA방법으로부터 나온 데이터의 일부로서 4 개의 펩타이드 정보가 사용된 것을 알 수 있다. 이렇게 같은 단백질을 분석하기 위해 사용된 펩타이드 개수가 다른 이유는 데이터 비의존성 분석결과를 토대로 한 것인지 아닌 것인지의 차이이다.

[121]

[122] MS^E-DDA가 보다 정확한 펩타이드 정보를 활용함으로써 특정 단백질을 분석하기 위해 보다 많은 펩타이드 정보가 활용된다. 따라서 단백질의 스코어가 증가되는 현상을 볼 수 있다. 보다 높은 스코어는 분석된 단백질의 신뢰도를 향상시킬 수 있는 지표가 되므로 MS^E-DDA 방법을 통한 단백질 검증은 유용하게 활용될 수 있을 것이다.

[123]

산업상 이용가능성

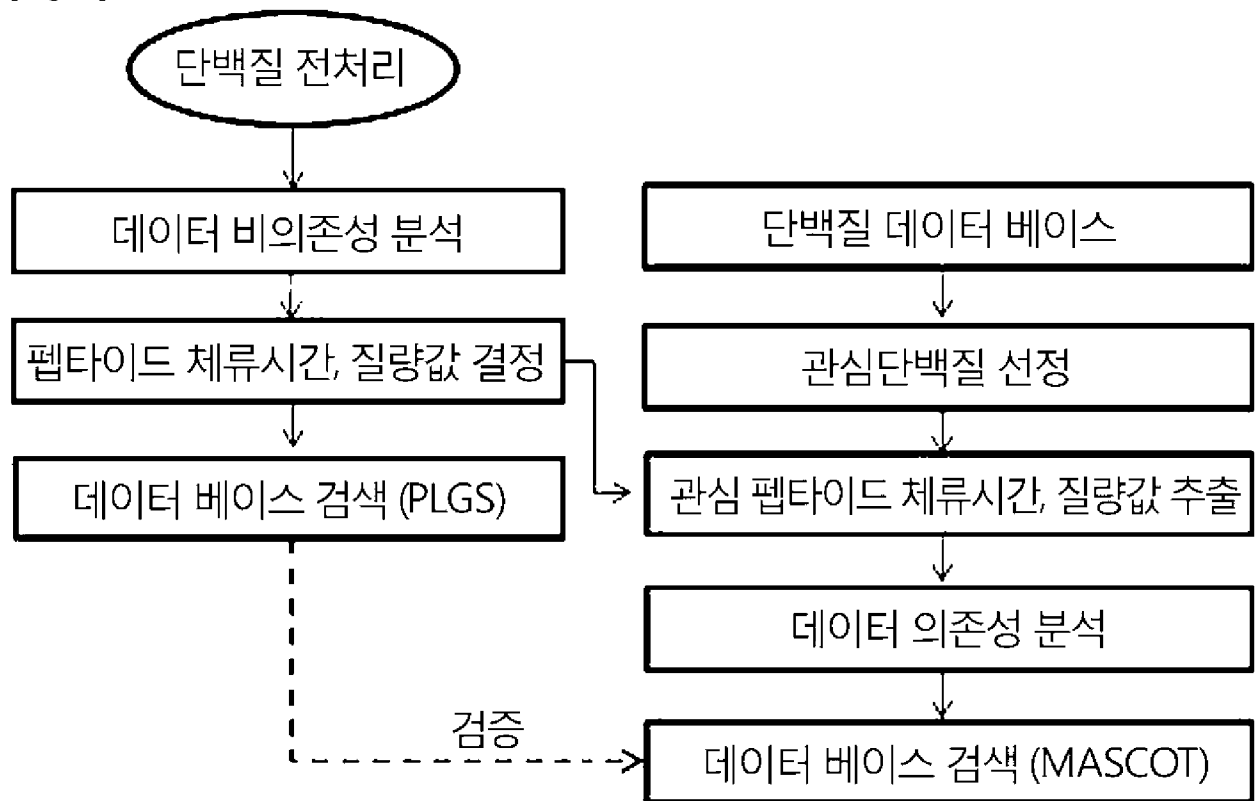
[124] 본 발명에 의하면, 수식화된 단백질과 샘플내에 미량 존재하는 단백질을 용이하게 검출하고 정량, 정성 분석할 수 있어 세포 신호전달 연구, 의약품 개발에 매우 유용하다.

청구범위

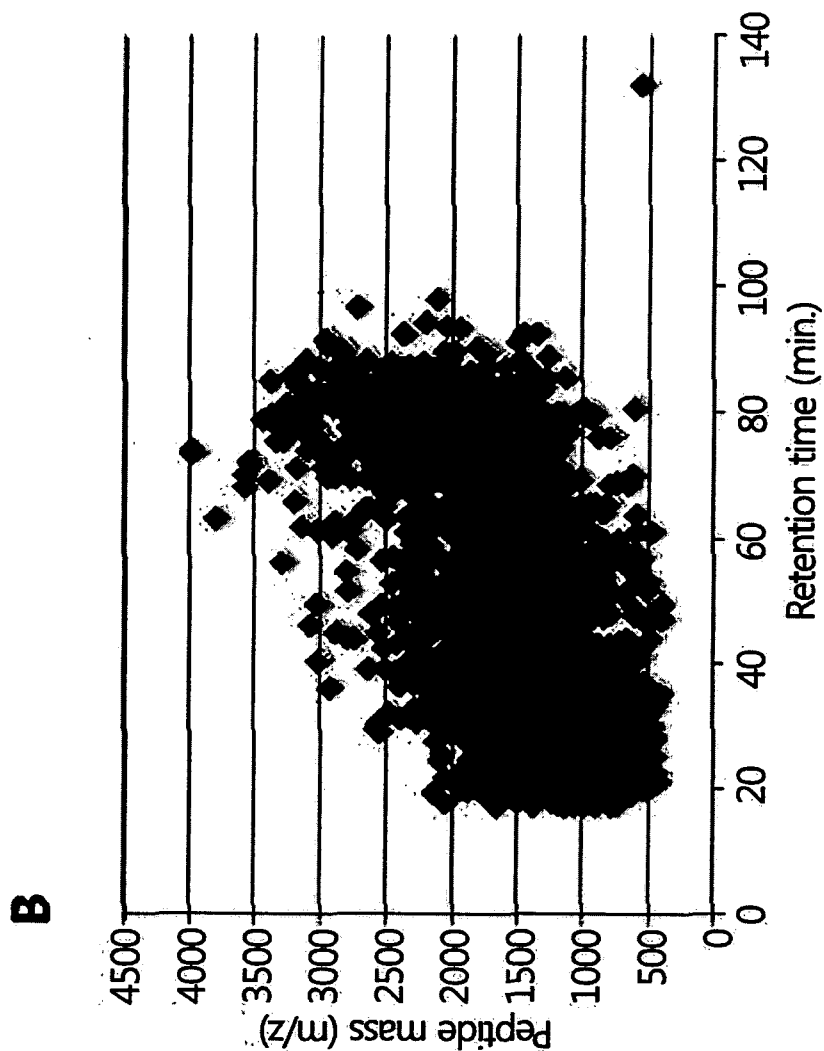
- [청구항 1] 질량분석기를 이용한 단백질 정량정성분석 방법으로서,
 (A) 단백질 혼합물을 전처리하는 단계;
 (B) 데이터 비의존성 분석을 통해 펩타이드의 체류시간 및 질량값 정보를 획득하는 단계;
 (C) 상기 체류시간 및 질량값 정보를 바탕으로 데이터베이스(PLGS)를 검색하여 1차로 단백질을 정량정성하는 단계;
 (D) 상기 체류시간 및 질량값 정보로부터 소정의 단백질에 관한 정보를 추출하는 단계;
 (E) 상기 (D)에서 추출된 정보를 활용하여 데이터 의존성 분석을 수행하는 단계;
 (F) 상기 (E)로부터 얻어진 체류시간 및 질량값 정보를 바탕으로 데이터베이스(MASCOT)를 검색하여 상기 단백질을 2차로 정량정성하는 단계;
 (G) 상기 (C)의 결과와 상기 (F)의 결과를 비교분석하여 상기 단백질의 정량정성을 검증하는 단계;
 를 포함하는 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 2] 제 1 항에 있어서,
 상기 질량분석기는 삼중사극자 질량분석기인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 3] 제 1 항에 있어서,
 상기 단백질은 세포내에 미량 존재하는 단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 4] 제 3 항에 있어서,
 상기 단백질은 막단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 5] 제 1 항에 있어서,
 상기 단백질은 번역후 수식화(PTM)된 단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 6] 제 5 항에 있어서,
 상기 단백질은 시스테인이 함유된 단백질인 것을 특징으로 하는

- 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 7] 제 1 항에 의한 단백질 정량정성 분석방법을 수행하는 프로그램이 저장된 저장매체.
- [청구항 8] 질량분석기를 이용한 단백질 정량정성분석 방법으로서,
 (A) 단백질 데이터베이스를 참조하여 관심 단백질을 선정하는 단계;
 (B) 단백질 데이터베이스를 참조하여 상기 관심 단백질 펩타이드의 이론상 체류시간 및 질량값을 추정하는 단계;
 (C) 상기 추정된 정보를 활용하여 데이터 의존성 분석을 수행하는 단계;
 (D) 상기 (C)의 결과를 기초로 상기 단백질 데이터베이스를 검색하여 분석된 단백질이 상기 (A)에서 선정된 관심 단백질인지를 확인하는 단계;
 를 포함하는 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 9] 제 8 항에 있어서,
 상기 질량분석기는 삼중사극자 질량분석기인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 10] 제 8 항에 있어서,
 상기 단백질은 세포내에 미량 존재하는 단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 11] 제 10 항에 있어서,
 상기 단백질은 막단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 12] 제 8 항에 있어서,
 상기 단백질은 번역후 수식화(PTM)된 단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 13] 제 8 항에 있어서,
 상기 단백질은 시스테인이 함유된 단백질인 것을 특징으로 하는 데이터 비의존성 분석법과 데이터 의존성 분석법을 복합화한 단백질 정량정성 분석방법.
- [청구항 14] 제 8 항에 의한 단백질 정량정성 분석방법을 수행하는 프로그램이 저장된 저장매체.

[Fig. 1]



[Fig. 2]



A

Function : 1 Survey

Trap CE Control Transfer CE Control Cone Voltage Lock Mass
Acquisition MS Survey MS/MS Peak Detection Include Exclude Adducts

Include Mass

☒ Enable MS/MS peak selection using an include list
☒ Only select peaks that are on the include list
☐ Preferentially select peaks that are on the include list

Include peaks within +/- mDa of entries on the include list
 Retention Time window +/- seconds

Include File List

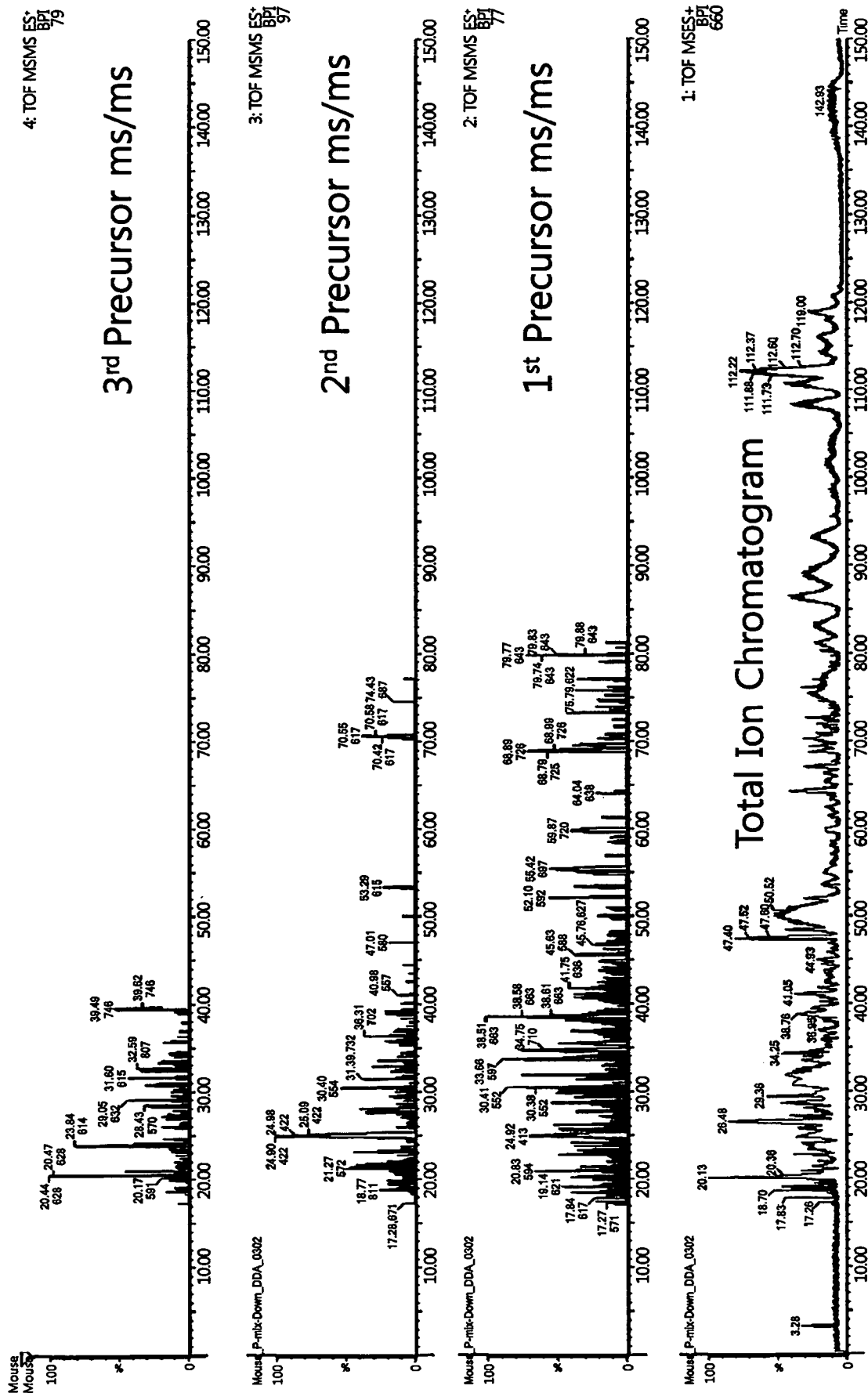
☐ Include from range
☒ Include from file

Mass	Retention Time	Collision Energy	Cone Voltage	Charge State	Scan Time	Int.
500...	33.960	0.00	40.00	2	0.50	
500...	65.460	0.00	40.00	2	0.50	
500...	39.010	0.00	40.00	2	0.50	
500...	28.770	0.00	40.00	2	0.50	
500...	62.620	0.00	40.00	2	0.50	
500...	60.590	0.00	40.00	2	0.50	
500...	63.160	0.00	40.00	3	0.50	
500...	37.780	0.00	40.00	3	0.50	
501...	37.800	0.00	40.00	3	0.50	
501...	30.860	0.00	40.00	2	0.50	
1001...	80.420	0.00	40.00	2	0.50	
1001...	48.920	0.00	40.00	2	0.50	
1001...	33.650	0.00	40.00	2	0.50	
501...	47.640	0.00	40.00	3	0.50	
501...	73.170	0.00	40.00	3	0.50	

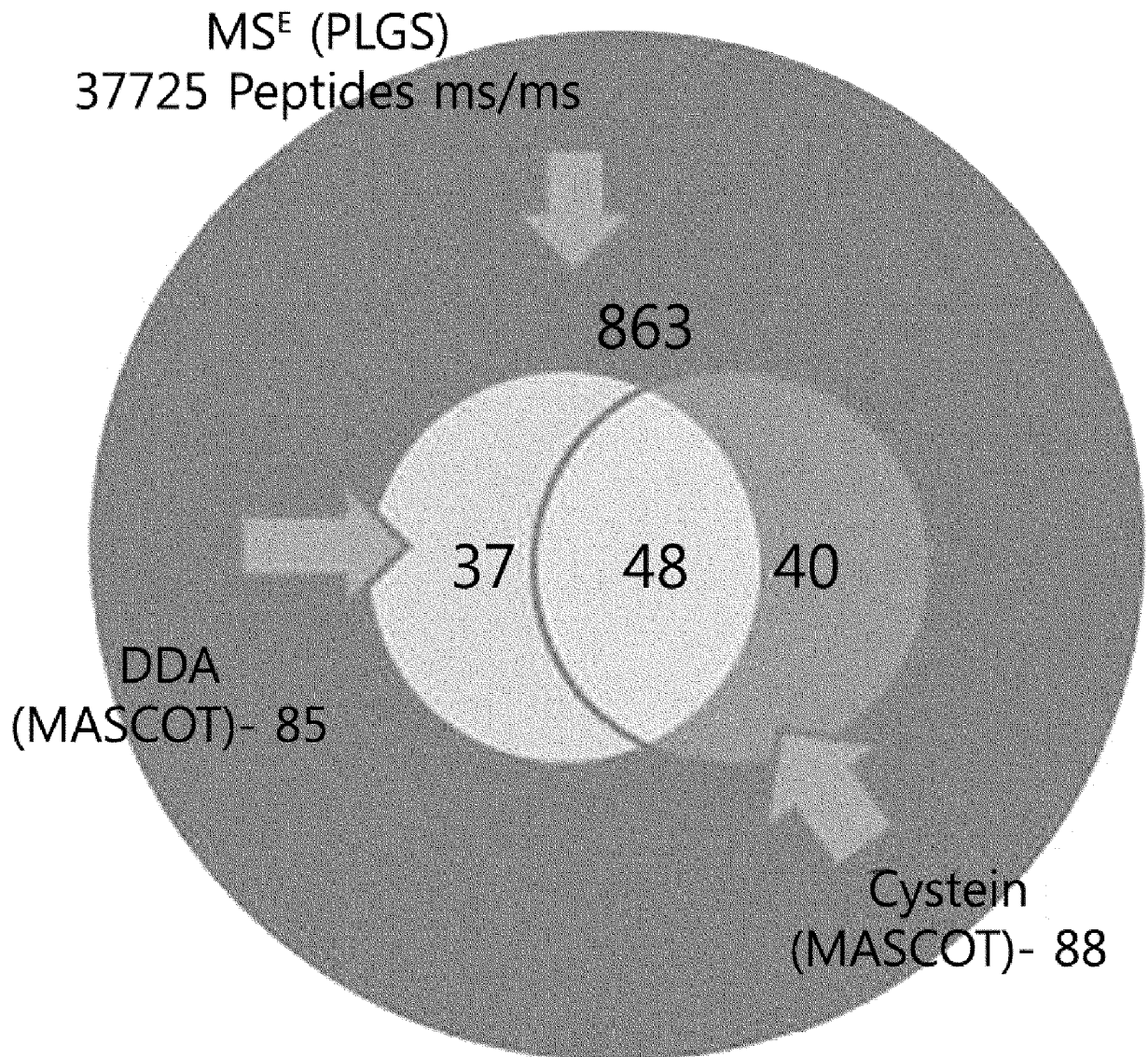
New Add... Delete Save Save As... OK Cancel Apply

생성된 Include list 적용

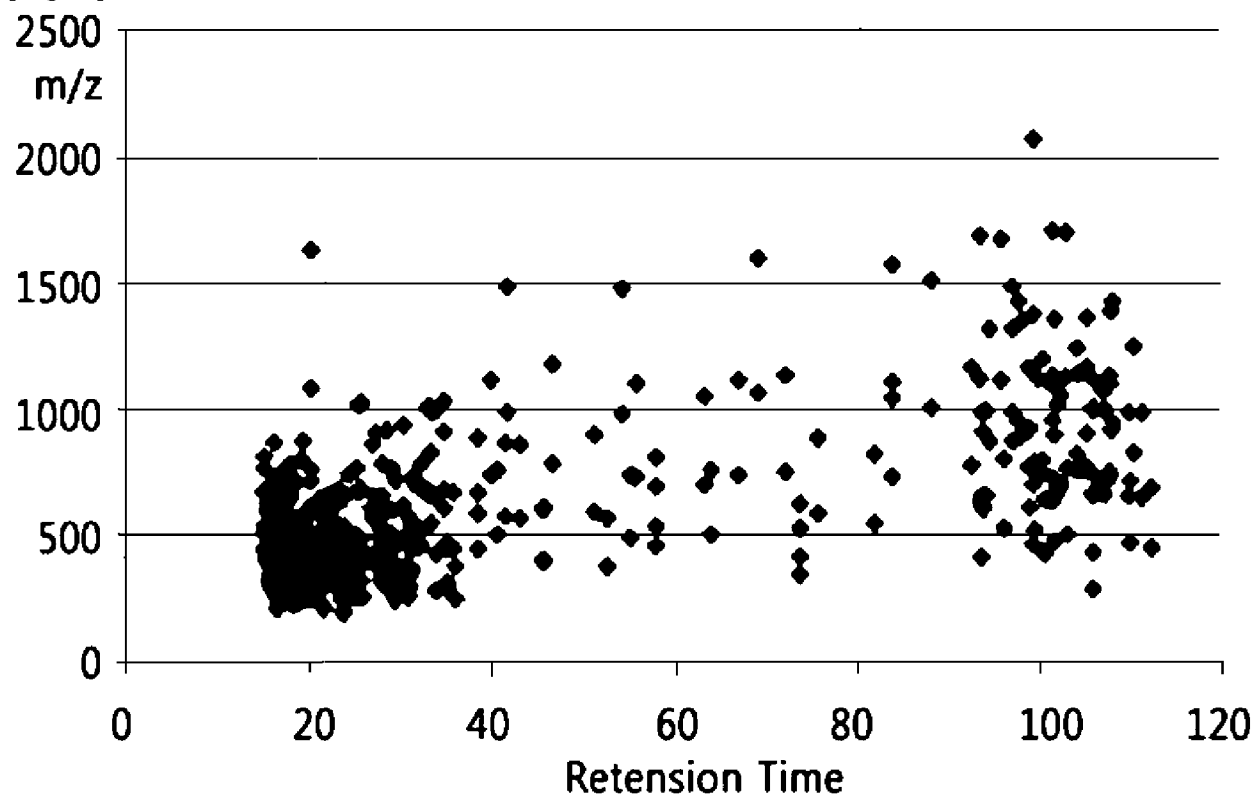
[Fig. 3]



[Fig. 4]

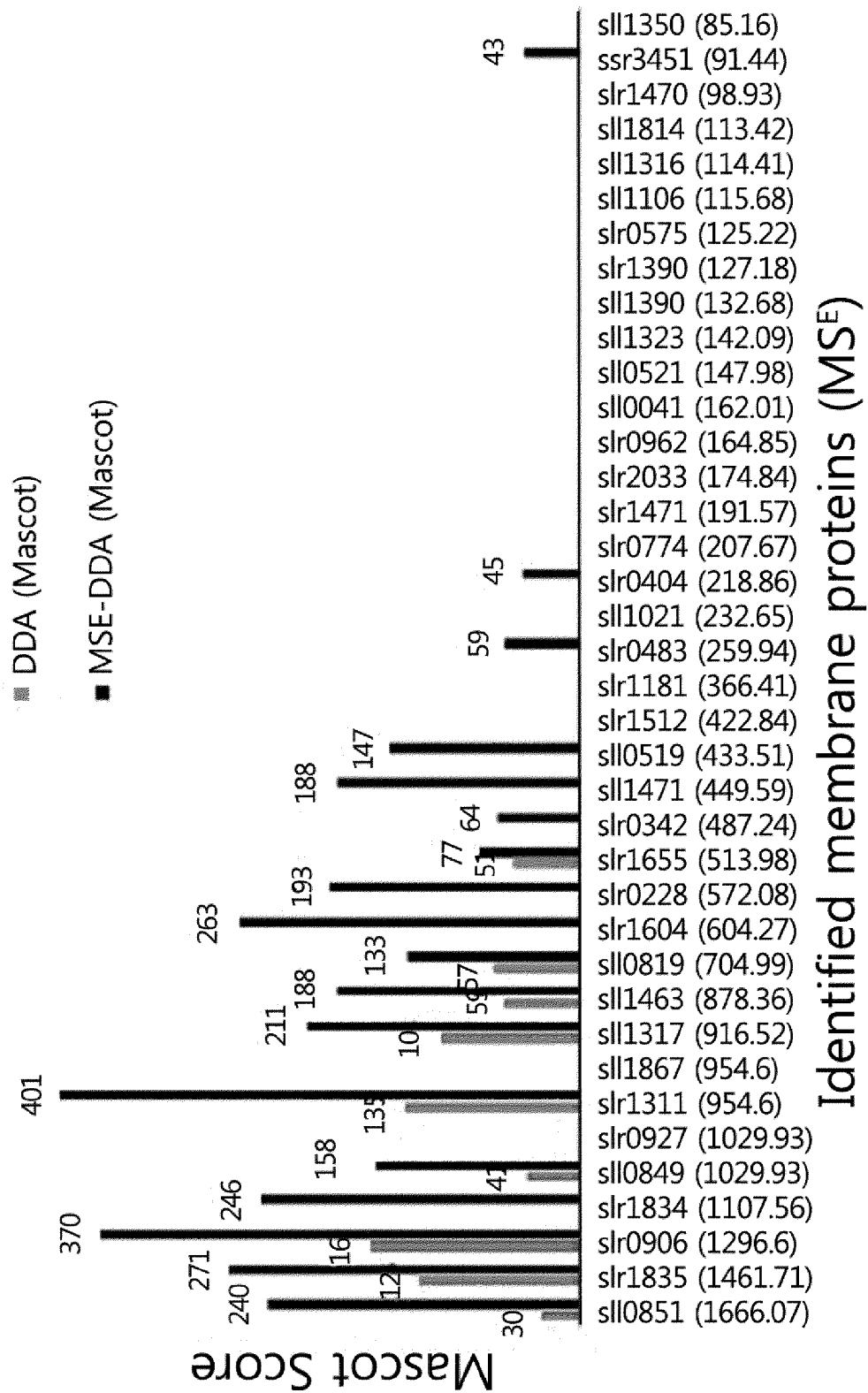


[Fig. 5]



[Fig. 6]

The map of Identified membrane proteins



[Fig. 7]

