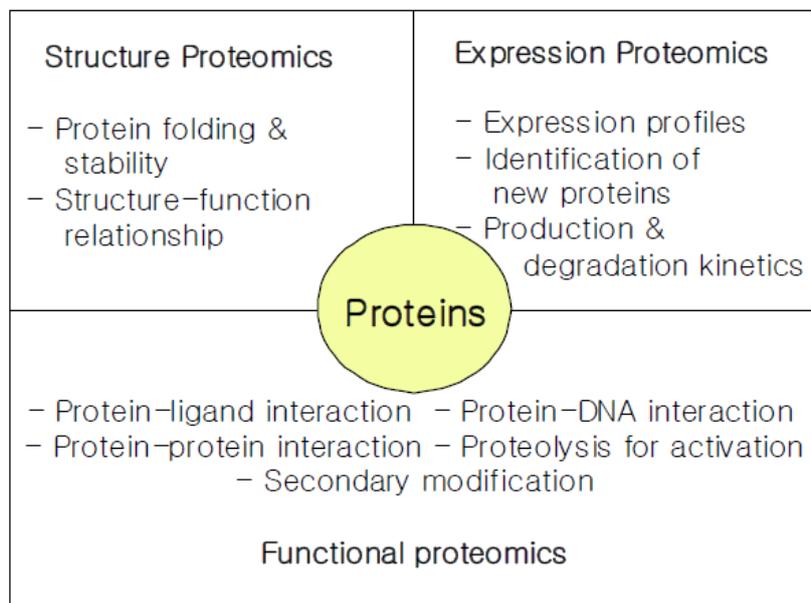


질량분석기 기반의 단백질체학(Proteomics) 소개

현대 바이오 연구의 한 획으로 자리매김한 게놈 프로젝트는 약 2~3만여 개의 인간 유전자에 대한 정보를 밝혔 다. 이 성과는 생명 현상의 이해를 위해 유전자 간 상호 관계에 얽힌 문제를 풀어야 한다는 더 큰 과제를 던져 주었으며 유전체학(genomics)이라는 바이오 분야의 탄생으로 이어졌다. 더불어 세포 내 여러 기능들을 실질적으로 담당하는 단백질들의 집합체인 단백질체(proteomics)에 대한 의미도 새롭게 인식되었고 이런 단백질체를 연구 하는 단백질체학 즉, 프로테오믹스(proteomics)라는 학문 분야가 주목을 받게 되었다.[1,2]

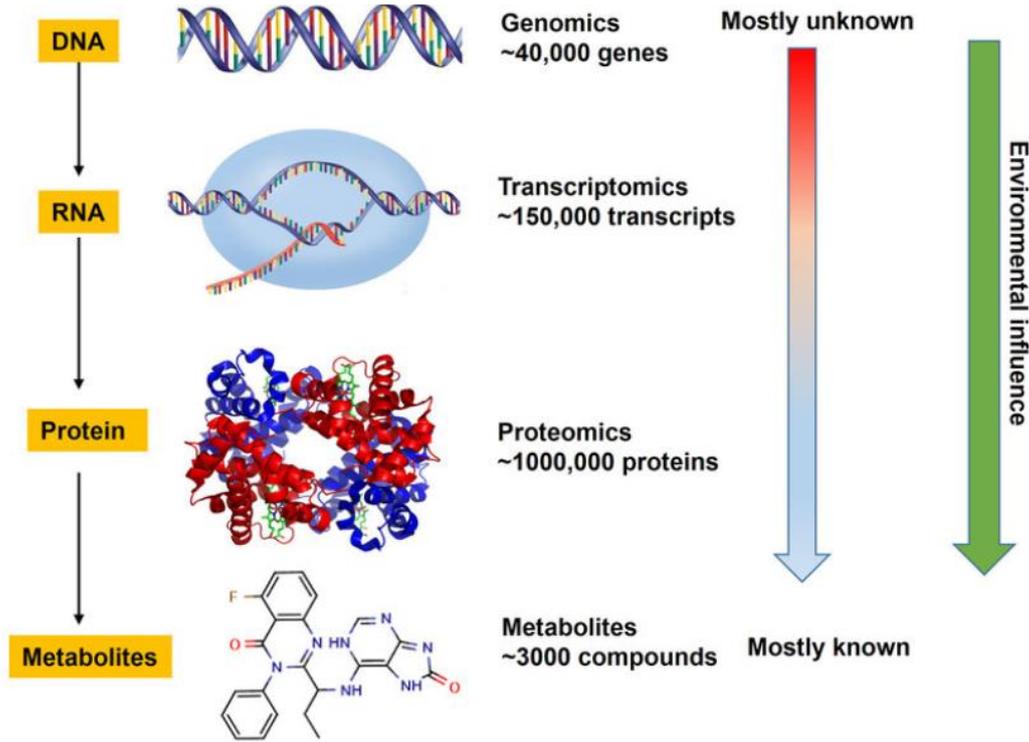
프로테오믹스(proteomics)란 유전자에 암호화되어 발현되는 단백질 전체를 세포기능별로 확인하고자 하는 학문이다. 프로테오믹스(proteome)이란 단백질(protein)과 전체(-ome)의 합성어로서 하나의 생물 종에서 만들어지는 모든 단백질에 대한 분석을 의미할 수 있으나[3], 실제로는 게놈(genome)의 상대어로서 **PROTE**in expressed by a **genOME**의 합성어로 인식되고 있는 용어이다. 프로테오믹스는 프로테오믹스(proteome)의 어미에 ~학, ~론을 의미하는 접미사 -ics가 붙어 프로테오믹스를 연구하는 방법과 기술을 포괄적으로 의미하는 말로, 번역한다면 '단백질체분석학'이라고 할 수 있다. 프로테오믹스는 세포내 전체단백질을 연구하는 대형 스케일의 다단계 고속분석기술이므로 연구대상도 단백질의 발현(expression), 기능(function), 구조(structure) 및 생합성 후 구조변형, 관련 단백질간의 결합성(protein-protein interaction)에 초점을 두고 질병의 진행과정과 연계시켜 총괄적으로 분석하는 기술이므로 유전체학보다 복잡하며 데이터양도 매우 크다.[4]



Proteomics의 최근 연구 기술 동향, 식품기술, 2005, 18(1), 3-15.

단백질 연구는 과거의 생명공학 연구가 진행해 오던 기술 패러다임을 벗어나 방대한 양의 정보처리가 요구되는 거대학문분야로 발전될 수밖에 없는데, 그 이유는 유전자는 그 길이에 무관하게 A(아데닌), T(티민), G(구아닌), C(시토신)라는 4가지의 유전적 코드만 이용하는 반면 단백질의 경우 20가지 아미노산으로 구성된 고분자이기 때문이다. 일부 RNA 촉매 기능을 제외하고는 N개의 서열을 지니는 유전자가 지닌 이들 4개의 물질은 4^N개의 구조

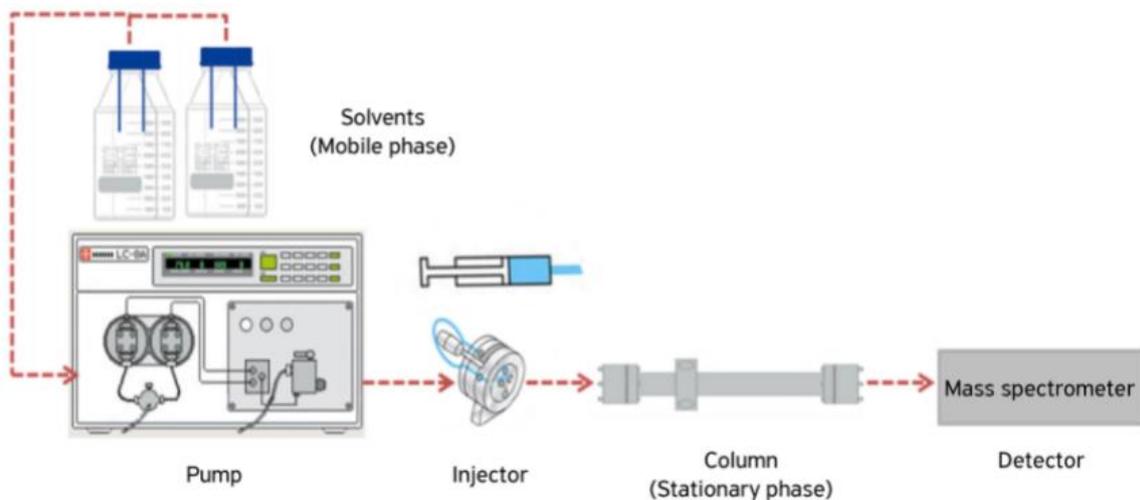
를 지닐 수밖에 없어 생체 내에서 비트에 해당되는 정보저장 역할을 제외하고는 그 이용이 다분히 제한적인 반면 단백질의 경우 20^N 에 해당하는 기하학적 도형체로 만들 수 있기 때문이다.[3]



https://www.researchgate.net/figure/The-advantages-of-metabolomics-over-other-omics_fig1_321045124

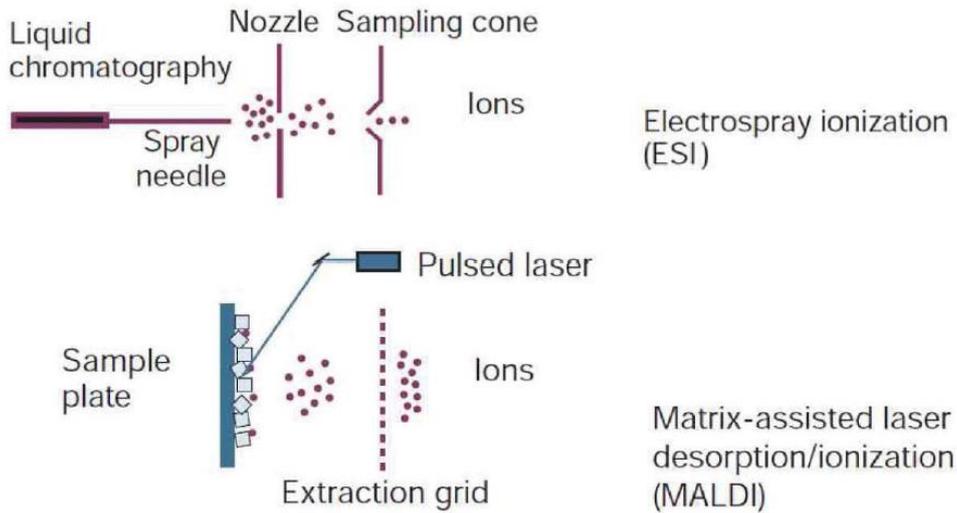
<LC-MS Based Detection of Differential Protein Expression >

프로테오믹스 연구에 여러 분석 기술을 사용할 수 있지만 분리분석기술과 질량분석기술을 결합한 액체 크로마토그래피-질량분석(LC-MS)이 protein 및 peptide 혼합물 분석에 널리 사용되고 있다. 질량분석법의 기본 과정은 아래 그림에서 볼 수 있듯이 샘플이 주입되고 샘플 분자가 이온화 과정을 통해 조각 이온으로 분해되고 최종적으로 분해된 이온이 검출기를 통해 검출된다. 분석 결과는 질량 대 전하비율(m/z)과 상대적 존재비의 함수관계를 나타내는 스펙트럼 형태로 표시되고 측정된 질량과 이미 알려진 질량의 상관관계나 특징적인 조각화 패턴(fragmentation pattern)을 통해 식별이 이루어진다.[5]



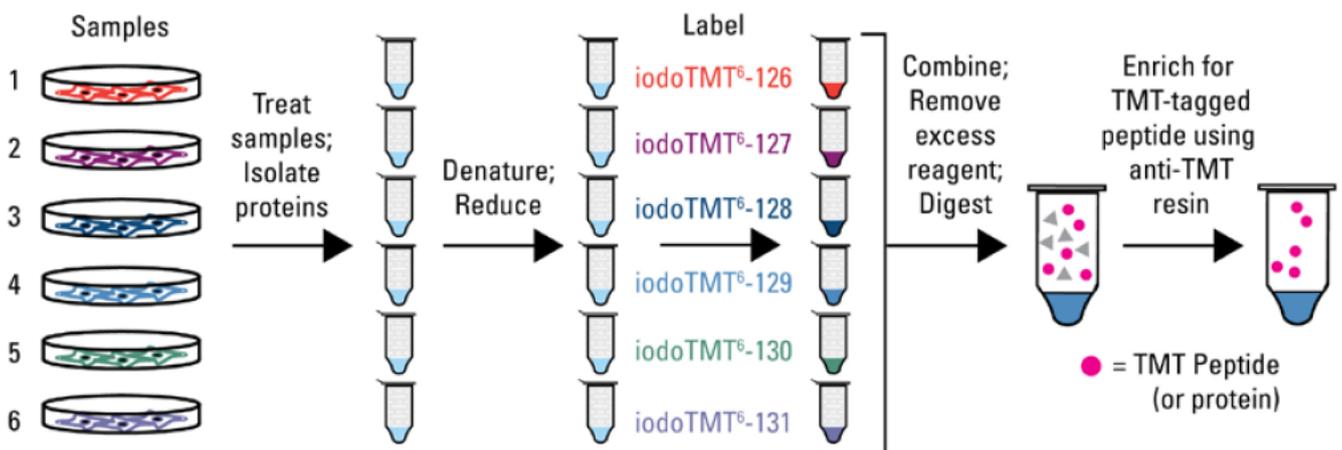
<https://www.scielo.br/j/aib/a/yzsPfxB9P8DXC7vsRTn7S3p/?lang=en#>

질량분석기를 이용한 단백질 연구는 Electrospray ionization(ESI)과 Matrix-assisted laser desorption/ionization (MALDI)와 같은 이온화 방법들이 개발됨으로써 본격화되었다고 할 수 있다. 단백질 분석 LC-MS 시스템에서는 분석민감도를 극대화하기 위하여 NanoSpray Ionization(NSI)을 주로 사용한다. NSI 시스템에서는 분무된 샘플용액 방울(droplet)들의 크기가 작고, 표면적/부피 비율이 높기 때문에 용매 분자들을 증발시키는 것이 용이하다. 이러한 특징은 이온화 효율을 증가시킴으로써 질량분석기의 검출한계를 낮추는 효과를 가져오게 되어 분석민감도가 극대화된다.[6]



프로테오믹스, 어떻게 활용할 수 있을까?, 분자세포생물학뉴스레터, 2013, 1-12.

단백질 정량 분석 방법으로는 크게 Labeled protein quantification 과 Label free protein quantification 방법으로 나뉘어지며 Labeled protein quantification은 Tandem mass tag(TMT) labeling과 Stable isotope label by amino acids in cell culture(SILAC) 방식 등이 있다. 최근 프로테오믹스 분야에서는 SILAC Labeling 방법보다 TMT-Labeling 방법이 널리 사용되고 있으며, 현재 이바이오젠에서는 Label free protein quantification 방법과 함께 TMT Labeling 방법을 이용한 정량분석 서비스를 제공하고 있다. TMT Labeling 방법은 세포 또는 조직으로부터 추출한 단백질을 펩타이드화 한 후, 각 샘플별로 서로 다른 동위원소를 labeling하여 분석하는 방법이다. TMT Labeling 방법의 전반적인 과정은 먼저 비교군별로 약물 처리 후 단백질을 추출하고 펩타이드로 만든 후 TMT labeling을 하게 된다. TMT labeling된 펩타이드를 pooling한 다음, fractionation을 진행하고 fractionation된 샘플을 LC-MS로 분석하고 최종적으로 LC-MS 분석 결과의 스펙트럼을 데이터베이스에 매칭하여 최종 결과를 얻게 된다.



<https://www.thermofisher.com/order/catalog/product/90101#/90101>.

<Application of Proteomics by LC-MS>

LC-MS 질량분석법을 통해 Cell[7,8], Tissue[9], Serum/Plasma[10,11], Culture media 등 다양한 샘플[12]에서 단백질의 동정 및 발현을 확인할 수 있다. 아래에 LC-MS 질량분석법을 통해 프로테오믹스 연구한 사례들을 소개하고자 한다.

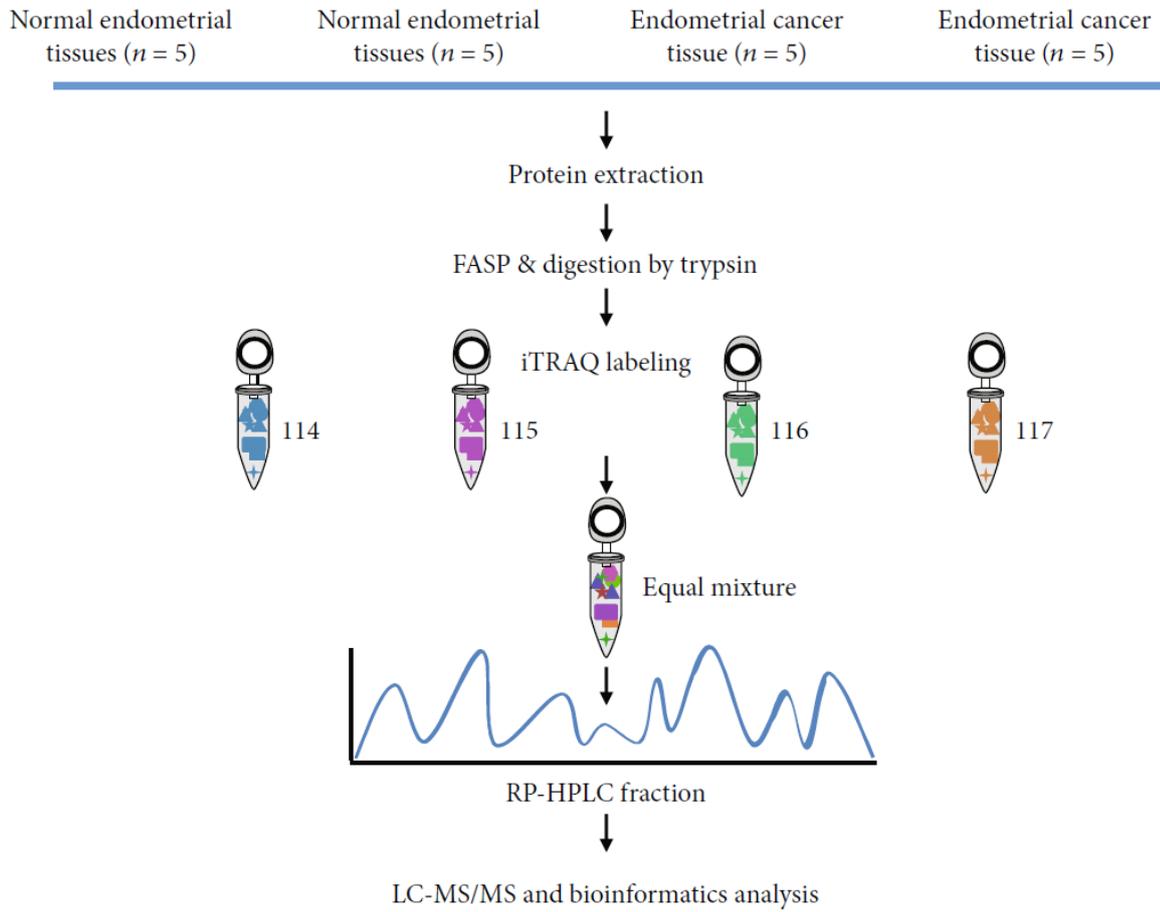
현재 전세계적으로 이슈인 COVID-19 발생의 원인인 코로나 바이러스(SARS-CoV-2) 세포를 분석한 논문이 2020년에 발표가 되었다. 현시점에서 코로나 바이러스(SARS-CoV-2) 세포의 분석은 현 상황을 대응하는데 매우 중요하다. 이에 본 연구에서는 SARS-CoV-2에 감염된 세포의 전체 단백질체를 분석하기 위해 LC-MS 질량분석법을 활용했다.

Source	# identified proteins	# identified peptides
Total	6,512	61,394
SARS2	9	279
<i>Chlorocebus</i>	6,503	61,115

Accession	Protein name	Organism	Description	Coverage [%]	# Peptides	# PSMs
PODTC9	NCAP	SARS2	Nucleoprotein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=N PE=3 SV=1	91	61	676
PODTC1	R1AB	SARS2	Replicase polyprotein 1ab OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=rep PE=1 SV=1	22	128	237
PODTC2	SPIKE	SARS2	Spike glycoprotein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=S PE=1 SV=1	39	53	172
PODTC2	ORF9B	SARS2	Protein 9b OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 PE=3 SV=1	96	12	49
PODTC5	VME1	SARS2	Membrane protein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 PE=3 SV=1	36	13	89
PODTC3	AP3A	SARS2	Protein 3a OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=3a PE=3 SV=1	28	8	30
PODTC7	NS7A	SARS2	Protein 7a OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=7a PE=3 SV=1	23	1	2
PODTC6	NS6	SARS2	Non-structural protein 6 OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=6 PE=3 SV=1	31	2	2
PODTC4	VEMP	SARS2	Envelope small membrane protein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=E PE=3 SV=1	16	1	2

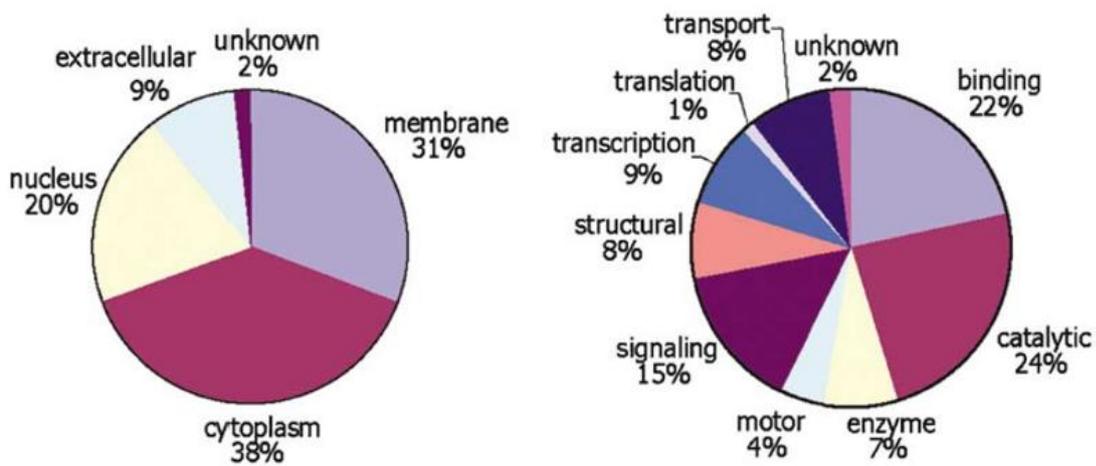
<https://www.biorxiv.org/content/10.1101/2020.04.23.057810v3>.

대표적인 부인과 악성종양의 하나인 자궁내막암(endometrial cancer, EC) 단백질체학 연구에서도 LC-MS를 활용한 질량분석법이 널리 사용되고 있다. 본 연구에서는 Normal endometrial tissue(NE) 샘플과 Endometrial cancer tissue(EC) 샘플에서 protein을 추출하고 trypsin을 사용하여 digestion 후 labeling 방법을 통해 정량 분석을 진행했다.



<https://doi.org/10.1155/2020/5273969>

2005년 미국의 Utah 대학에서는 Cell 샘플과 FFPE 샘플로 LC-MS를 활용하여 분석을 진행했고 각각의 샘플들의 protein list 결과를 확인했다. 더 나아가 cell 샘플과 FFPE 샘플의 분석 결과로 얻은 protein list를 활용하여 Gene ontology 분석을 진행했다.



Laboratory Investigation (2005) 85, 1405–1415.

현재 임상화학검사실에서도 액체 크로마토그래피 질량분석법을 도입하여 운영하고 있으며 2000년대 초반부터 질량분석기를 이용하는 검사의 개발 및 검사 건수가 점점 증가하고 있다. 질량분석법을 이용해서 측정 가능한 물질은 약물, 대사물질 등과 같은 저분자량 물질 이외에도 단백질, 핵산 등과 같은 고분자량 물질이 있다. 질량분석법이 계속적으로 발전하리라 예상하는 이유는 오믹스 분야처럼 단백질, 지질, 대사물질 [13] 등 유사한 물질을 동시에 측정이 가능한 검사가 질량분석법에서는 가능하여 질량분석법에 대해 계속적으로 연구 및 개발이 되고 있기 때문이다.[14]

단백질 기능해석 및 새로운 단백질의 발견은 신약의 개발, 고부가가치 생물소재의 생산으로 연결되어 생물산업의 발전에 크게 기여할 것으로 보이며 대량 정보처리에 기반한 단백질 기능해석이 가능케 하는 단백질 정보학과 프로테오믹스의 융합적인 발전을 꾀할 수 있다면, 향후 현재의 기술역량을 선진국 수준으로 끌어올릴 수 있을 것이다.[3]

이바이오젠에서는 이러한 연구 흐름에 맞춰 실험 방법에 따라서 높은 정확도의 단백질 동정(identification)과 발현(expression)을 최소 수백개에서 수 천개까지 확인 가능한 High Resolution/High Accuracy Mass Spectrometry 분석 기반의 Proteomics 분석서비스를 도입했다. 더 나아가 Proteomics에서만 확인할 수 있는 단백질의 특정 아미노산의 다양한 modification(Phosphorylation, Glycosylation, Methylation, Acetylation, Glycation)[15]을 정확한 질량값을 이용해서 분석할 수 있다.

LC-MS Proteomics service에 대해서는 자사 홈페이지[16]를 통해 확인 가능하며, Protein Expression Profiling Service에 대해서는 아래 내용을 통해 확인할 수 있다.

<LC-MS Proteomics service >

LC-MS Proteomics service는 단백질부터 multi-omics까지 한 번에 분석이 가능한 질량분석기 기반의 단백질체 분석 서비스입니다. 분석 상담 후 다양한 샘플(cell pellet, tissue, gel piece, conditioned media, serum/plasma, others)에서 protein identification 및 protein expression을 profiling 할 수 있으며, ExDEGA Report, GO, Pathway, Clustering 등의 다양한 분석을 지원해 드립니다.

Service Name	Protein Expression Profiling Service
Service Code	LM1001
Sample Requirement	> 150 ug protein or other samples
Sample type	Cell, Tissue, Serum/Plasma, Cell media, Gel piece
Analysis Instrument (HPLC & UPLC)	Thermo Dionex-UltiMate 3000 SystemAgilent 1260 & 1290 System
Analysis Instrument (Mass Spectrometry)	Thermo Q-Exactive MS
LC-MS/MS Method	UPLC & Q-Exactive
Data Format	Protein List, Quantified Protein List (DEPs)

<참고문헌>

1. 프로테오믹스, 어떻게 활용할 수 있을까?, 분자세포생물학뉴스레터, 2013, 1-12.
2. https://www.ibric.org/myboard/read.php?Board=bio_response&id=10
3. 프로테오믹스 연구의 최신동향과 활용, NEWS & INFORMATION FOR CHEMICAL ENGINEERS, 2004, 22(6), 676-683.
4. Proteomics의 최근 연구 기술 동향, 식품기술, 2005, 18(1), 3-15.
5. https://www.agilent.com/cs/library/slidepresentation/Public/5991-5857_Agilent_MS_Theory_KO.pdf.
6. 칩 기반 미세관 HPLC를 이용한 단백질 분석, ANALYTICAL SCIENCE & TECHNOLOGY, 2011, 24(6), 407-413.
7. Targeted proteomics as a tool to detect SARS-CoV-2 proteins in clinical specimens, bioRxiv, 2021.
8. A Quantitative Proteomic Analysis to Reveal Effects of N-acetylcysteine on H₂O₂-induced Cytotoxicity, Current Proteomics, 2019, 16, 1-11.
9. Proteomic Analysis of Human Endometrial Tissues Reveals the Roles of PI3K/AKT/mTOR Pathway and Tumor Angiogenesis Molecules in the Pathogenesis of Endometrial Cancer, BioMed Research International, 2020, 10.
10. A Serum Protein Profile Predictive of the Resistance to Neoadjuvant Chemotherapy in Advanced Breast Cancers, Molecular & Cellular Proteomics 10.10, 2011, 1-13.
11. Identification of novel blood biomarkers of treatment response in cystic fibrosis pulmonary exacerbations by label-free quantitative proteomics, SCIENTIFIC REPORTS, 2019, 9(17126), 1-9.
12. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS, Laboratory Investigation, 2005, 85, 1405-1415.
13. A non-targeted LC-MS metabolic profiling of pregnancy: longitudinal evidence from healthy and pre-eclamptic pregnancies, Metabolomics, 2021, 17(20), 1-12.
14. 임상검사실에서의 액체크로마토그래피-질량분석법에 대한 제언: 도입과 운영, Lab Med Online, 2020, 10(1), 1-9.
15. Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks, SCIENTIFIC REPORTS, 2015, 5, 1-12.
16. <https://www.e-biogen.com/service4.php>.