creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

고 병 준 교수 지도
석사학위 청구논문

질량 분석 기반 생체 분자의 펩타이드 및 글라이칸 특성 연구

- 면역 및 구조 단백질의 특성 규명을 위한 질량 분석 기반 다중 분석 접근 -

2025

성신여자대학교 대학원
미래응용과학학과
천 윤 서

질량 분석 기반 생체 분자의 펩타이드 및 글라이칸 특성 연구

- 면역 및 구조 단백질의 특성 규명을 위한 질량 분석 기반 다중 분석 접근 -

고 병 준 교수 지도

이 논문을 석사학위논문으로 제출함

2025년 5월

성신여자대학교 대학원

미래응용과학학과

천 윤 서

인 준 서

천윤서의 석사학위 논문으로 인준함

2025년 6월

심사위원장 이 재 원 (인)

심 사 위 원 고 병 준 (인)

심 사 위 원 현 정 은 (인)

성신여자대학교 대학원

논문 개요

단백질체학(Proteomics)은 생물학적 시스템 내에 존재하는 단백질과 펩타이드의 다양성과 복잡성을 이해하기 위한 중요 연구 분야이다. 이러한 단백질의 특성 규명을 위해 주로 사용되는 질량 분석법은 높은 정확도로 인해 강력한 분석 도구로 여겨진다. 이에 따라 본 연구에서는 질량 분석을 활용해 정밀 분석 기법을 확립하고, 이를 HLA class Ⅰ 복합체, 치료용 항체, 콜라겐 등 다양한 생체 단백질에 적용함으로써 구조 분석 및 특성 분석을 위한 전략을 제시하고자 하였다.

먼저, HLA는 Human Leukocyte Antigen의 약자로, 인간 백혈구 항원을 의미한다. HLA는 class I, II, III로 분류되며, 이 중 HLA class I 분자는 내인성 유래 펩타이드를 $CD8^+$ T 세포에 제시함으로써 체내 면역을 수행하는 메커니즘에 관여한다. 이 매개 면역 요법은 단순히 펩타이드를 면역 반응에 사용하는 것을 넘어, 암과 같은 질환의 발병 및 진행 과정에서 다양한 생물학적 역할을 수행한다. 따라서 이러한 항종양 치료 방법을 확립하기 위해서는 종양 세포에서 발현되는 HLA class I 결합 펩타이드의 서열 데이터를 평가 및 수집하기 위한 적절한 분석법의 개발이 중요하다.

이를 위한 실험에는 인간 대장암 세포주 HCT-116이 사용되었다. 실험에 앞서, 펩타이드를 효율적으로 추출하기 위한 cell lysis 조건을 최적화하고자 다양한 lysis 시간 조건으로 실험을 진행했다. 그 결과, 4시간 처리군에서 가장 높은 추출 효율을 보였으며, lysis 시간이 경과할수록 HLA class I 펩타이드 추출 효율이 감소하는 경향을 확인할 수 있었다. 이후 Orbitrap 질량분석기를 사용해 추출한 펩타이드의 분석을 수행했고, 데이터 분석은 PEAKS Studio 및 Proteome Discoverer 소프트웨어를 활용했다. 먼저

Peaks Studio에서 De novo 서열 분석을 통해 펩타이드 서열을 예측하고, 이를 FASTA format 파일로 변환해 Proteome Discoverer에서 데이터베이스로 활용해 동정되는 펩타이드 수를 확인하였다. 이 분석을 통해 총 1400개 이상의 HLA 펩타이드 서열을 규명할 수 있었다. 또한, HLA class I 분자의 면역학적 기능에 관여하는 N-글라이칸을 분리해 분석하고, UPLC-FLR 결과와 질량 분석 결과를 비교함으로써 전체적인 글라이칸 구조를 확인했다.

두번째로, 질량 분석을 기반으로 치료용 단일클론 항체의 글라이코실화 특성을 분석하였다. 치료용 단일클론 항체는 다양한 질환에 대한 치료 효과로 현대 의학에서 사용 추세가 꾸준히 증가하고 있다. 이에 따라 생산된 항체의 이질성 평가는 필수적이며, 그중에서도 항체 글라이코실화는 치료 항체의 효능에 직접적인 영향을 미치기 때문에 항체 품질 검증 과정에서 핵심적인 평가 요소이다. 따라서 글라이칸 특성 규명을 위한 분석법의 최적화는 중요한 요소로 간주된다.

본 연구에서는 다양한 오리지널 항체 및 그 바이오시밀러 샘플의 글라이코실화 패턴을 파악하기 위해 항체 자체에 대한 LC-MS 분석과 분리한 N-글라이칸에 대한 LC-MS 분석을 병행했다. 우선, 항체 샘플들에 별도의 효소 처리를 하지 않은 채로 질량 분석을 수행해 분자량을 확인했다. 추가적으로 PNGase F의 처리를 통한 분석으로 mass shift를 확인하여 항체에 부착된 글라이칸 조성을 유추했다. 또한, 분리한 N-글라이칸은 초고성능 액체크로마토그래피-형광 분석(UPLC-FLR)과 LC-MS 분석 기법을 활용해 분석했고, 이를 통해 다양한 글라이칸 구조를 식별할 수 있었다. 주요 글라이칸의 peak 면적 백분율 총합을 계산한 결과, 일부 항체를 제외한 대부분의 시료에서 80% 이상의 백분율 총합이 확인되어 항체 품질 측면에서 양호한 특성을 보임을 확인하였다. 또한, 항체 간에 G0F, G1F, G2F 구조의 분포 비율 차이가 관찰되었으며, 이를 통해 시료 간 미세한 글라이칸 패턴 차이

를 확인할 수 있었다.

마지막으로, 생물의 주요 구성 성분 중 하나인 콜라겐에 대해 질량 분석 기반의 특성 분석을 수행했다. 일반적으로, 효소 처리로 얻은 저분자 어류 콜라겐은 다양한 산업 분야에서 활용도가 높다. 그러나 시중에 유통되는 콜라겐은 어종에 따라 기능과 안정성에 차이가 존재하며, 대부분 가공된 형태로 제공되기 때문에 원료의 품질과 유래를 명확히 파악하기 어렵다. 따라서 콜라겐 성분에 대한 정밀한 분석을 통해 어종별 특성을 평가할 수 있는 기술의 개발이 필요하다. 하지만, 저분자 어류 콜라겐은 다른 단백질에 비해 분자량이 작고 다양한 펩타이드로 구성되어 있어, 기존의 분석 방법만으로는 정확한 서열을 파악하기 어렵다. 이를 극복하기 위해 본 연구에서는 액체 크로마토그래피와 결합한 질량 분석 방식을 활용해 복잡한 펩타이드 혼합물의 서열 내에서 어종별 공통 서열과 특이서열을 동정할 수 있는 분석 전략을 확립했다. 사용한 시료는 명태, 틸라피아. 대구 각 3종류, 총 9개의 어류 콜라겐이며, 이를 Orbitrap 질량 분석기로 분석했다. 데이터 분석에는 PEAKS Studio와 Proteome Discoverer 소프트웨어를 활용했다. 이 역시 동일한 절차에 따라 PEAKS Studio에서 De novo 서열 분석을 수행한 후, 자체 구축한 데이터베이스를 활용하여 Proteome Discoverer에서 어종 간 공통 서열 및 특이서열을 확인하였다.

목 차

논문개요

Ⅰ. 서론 ····· 1

Ⅱ. 재료 및 실험방법 ····· 10

1. 세포주 및 실험 재료 ····· 10

2. HLA Class I 분석 ····· 11

2-1. Cell lysis 시간 설정 ····· 11

1) Cell lysis ····· 11

2) 단백질 정량 ····· 11

3) 웨스턴 블롯(Western blot) ····· 12

2-2. HLA 펩타이드 분리 및 분석 ····· 13

1) 면역 침강(Immunoprecipitation) ····· 13

2) HLA 펩타이드 용출 ····· 14

3) 정제 ····· 14

4) LC-MS/MS 조건 ····· 15

5) 데이터베이스 구축 및 서열 동정 ····· 19

2-3. HLA Class I 분자의 글라이칸 분석 ····· 21

1) 항체 제거 및 시료 준비 ····· 21

2) 글라이칸 분석 전처리 ····· 21

3) 글라이칸 LC-FLR 분석 ····· 25

4) 글라이칸 LC-MS 분석 ····· 25

3. 치료용 항체 분석 ····· 27

3-1. 치료용 항체의 글라이칸 패턴 분석 ····· 27

1) 시료 준비 및 LC-MS 분석 조건 ····· 27

2) Deconvolution을 통한 분자량 분석 ····· 28
3) 질량 분석 기반 글라이칸 동정 ····· 30
4. 어류 저분자 콜라겐 펩타이드 분석 ····· 31
4-1. 콜라겐 펩타이드 서열 동정 ····· 31
1) 시료 준비 및 분석 장비 조건 ····· 31
2) 분석 전 장비 보정 및 설정 ····· 31
3) LC-MS/MS 분석 조건 ····· 32
4) 결과 데이터 분석 ····· 35

Ⅲ. 결과 ····· 36
1. HLA Class I 분석 결과 ····· 36
1-1. 펩타이드 분석 결과 ····· 36
1) 단백질 정량 ····· 36
2) 웨스턴 블롯 ····· 38
3) HLA Class I 펩타이드 LC-MS/MS 결과 ····· 40
1-2. 글라이칸 구조 동정 결과 ····· 42
2. 치료용 항체 분석 결과 ····· 49
2-1. 글라이칸 패턴 분석 결과 ····· 49
1) Deconvolution 분석 ····· 49
2) UPLC-FLR 및 질량 분석 기반 글라이칸 동정 ····· 61
3. 어류 저분자 콜라겐 펩타이드 분석 결과 ····· 77
3-1. LC-MS/MS 결과 ····· 77
1) 펩타이드 서열 간 상동성 및 차이점 분석 ····· 77
Ⅳ. 결론 ····· 86

참고문헌
ABSTRACT

그림 목차

그림 1. HLA class I 펩타이드 제시 경로 ····· 5
그림 2. 주요 글라이칸(G0F, G1F, G2F)의 구조 ····· 7
그림 3. RapiFluor-MS 형광 물질 구조 ····· 23
그림 4. GlycoWorks RapiFluor-MS N-glycan kit workflow ····· 24
그림 5. BCA standard curve (4시간) ····· 37
그림 6. Lysis 시간 변화에 따른 HLA class I 단백질 검출 결과 ····· 39
그림 7. HLA 펩타이드 질량 분석 Total ion chromatogram ····· 41
그림 8. HLA class I 펩타이드의 길이별 분류 ····· 41
그림 9. HLA Class I 글라이칸 LC-MS 크로마토그램 ····· 43
그림 10. RapiFluor-MS labeled N-글라이칸 분자량 ····· 43
그림 11. 7개의 글라이칸 peak에 대한 질량 스펙트럼 확대 이미지 ····· 44
그림 12. UPLC-FLR 및 LC-MS 기반 글라이칸 동정 결과 ····· 58
그림 13. 항체 분자량 deconvolutioin 결과 ····· 51
그림 14. 항체별 N-글라이칸 질량 분석 크로마토그램 ····· 62
그림 15. Humira N-글라이칸 질량 스펙트럼 ····· 67
그림 16. 항체별 N-글라이칸 형광 분석 결과 ····· 71
그림 17. 콜라겐 질량 분석 Total ion chromatogram ····· 79
그림 18. 콜라겐 어종 내 공통 및 특이서열 비교 ····· 83
그림 19. 콜라겐 어종 간 공통 및 특이서열 비교 ····· 84

표 목차

표 1. LC-MS/MS gradient (HLA class I 펩타이드) ····· 17
표 2. Mass spectrometer 분석 조건 (HLA class I 펩타이드) ····· 18
표 3. Proteome Discoverer 분석 파라미터 (HLA class I 펩타이드) ····· 20
표 4. LC-MS 분석 gradient (HLA class I 글라이칸) ····· 26
표 5. Mass spectrometer 분석 조건(HLA class I 글라이칸) ····· 26
표 6. LC-MS 분석 gradient (항체) ····· 29
표 7. Mass spectrometer 분석 조건 (항체) ····· 29
표 8. 실험에 사용한 콜라겐 샘플 목록 ····· 33
표 9. LC-MS/MS gradient (콜라겐) ····· 33
표 10. Mass spectrometer 분석 조건 (콜라겐) ····· 34
표 11. Proteome Discoverer 분석 파라미터 (콜라겐) ····· 35
표 12. BSA standard 농도 및 흡광도 (4시간) ····· 37
표 13. 글라이칸 Peak별 질량값 및 동정 결과 ····· 47
표 14. PNGase F 처리 유무에 따른 항체 시료 분자량 ····· 56
표 15. N-글라이칸 조합별 이론 분자량 계산 ····· 58
표 16. N-글라이칸 조합 및 intensity 측정 결과 ····· 59
표 17. 항체별 주요 글라이칸 구성 백분율(%) ····· 76
표 18. ALC ≥ 75% 조건에서 식별된 펩타이드 수 ····· 82
표 19. 콜라겐 어종 내 공통 서열 목록 ····· 85
표 20. 콜라겐 어종 간 공통 서열 목록 ····· 85

I. 서 론

일반적으로 단백질 또는 펩타이드 분해물의 분석에는 역상 초고성능 액체 크로마토그래피(Reverse Phase Ultra-Performance Liquid Chromatography, RP-UPLC)와 질량 분석기(Mass Spectrometer)를 결합한 LC-MS 방식이 주로 사용된다. RP-UPLC는 높은 분리능과 질량 분석기와의 우수한 호환성으로 인해 선호되며, 이동상으로는 주로 물(Water)와 아세토니트릴(Acetonitrile)과 같은 유기용매의 혼합물이 사용된다. 여기에 Formic acid와 같은 산을 첨가하면 분석 물질을 양성자화시키고 고정상과의 불필요한 이온 상호 작용을 줄여 분석의 정밀도를 향상시킬 수 있다. 또한, 질량 분석은 높은 감도를 바탕으로, 복잡한 단백질 및 펩타이드 혼합물 내에서도 정확한 식별이 가능하다.[1] 본 연구에서는 이러한 질량 분석 기반 방법을 활용해 다양한 생체 단백질 및 펩타이드 시료의 특성 분석을 수행했다.

최근 암 발병률은 전연령대에 걸쳐 지속적으로 증가하는 추세를 보이고 있으며, 전세계적인 주요 사망 원인 중 하나로 자리잡았다. 이에 따라 보다 효과적인 암 치료법 개발에 대한 요구가 높아지고 있다. 이러한 치료 전략 중 하나로, 면역 시스템을 활용한 접근법이 주목받고 있다. 대표적인 항 종양 면역 메커니즘으로는 인간 백혈구 항원(HLA) 복합체가 T세포 수용체에 의해 인식되어 발생하는 세포 독성 면역 기전이 존재한다. 체내 면역 시스템에서 작용하는 HLA는 유전 구조와 기능에 따라 class I, class II, class III로 구분되는데, 이 중 HLA class I은 암세포 내에서 발현된 유전자 유래 펩타이드와 결합한 후 이 항원을 세포 표면에 운반하고 제시하는 역할을 수행한다. 이를 통해 CD8+ T세포는 돌연변이 단백질을 발현하는 암과 같은 병리학적인 세포를 식별해 파괴할 수 있다.

HLA class I의 유전자는 α-사슬과 β2-마이크로글로불린의 이종이합체(heterodimer)로 구성되어 있고, 이 중 α-사슬은 6번 염색체의 단완(p-arm) 부분에 위치하는 유전자에 의해 발현된다. α-사슬은 α1, α2, α3 도메인으로 구성되고, 이중 α1과 α2 도메인은 항원 펩타이드가 결합하는 부위인 peptide-binding groove를 형성하며, α3 도메인은 CD8 분자와 결합하여 T 세포를 활성화하는 역할을 수행한다. β2-마이크로글로불린의 유전자는 15번 염색체에 존재하며 항원 제시에는 관여하지 않지만 α-사슬과 결합하여 안정적인 구조를 형성함으로써 세포 표면 발현에 도움을 제공한다.[2]

일반적으로 HLA class I 분자는 약 8-10개의 아미노산으로 구성된 짧은 펩타이드를 결합한다. 이 펩타이드 항원은 종양 세포에 의해서만 독점적으로 발현되는 종양 특이적 항원(Tumor-Specific Antigen, TSA) 또는 정상 세포에서도 발현되지만 종양 조직에서 과발현되는 종양 관련 항원(Tumor-Associated Antigen, TAA)으로 분류된다.[3]

HLA Class I 펩타이드의 발생 과정은 다음과 같다. 암세포 내에서 발현된 단백질은 프로테아좀에 의해 짧은 펩타이드로 분해되고 항원 가공 수송체(Transporter associated with Antigen Processing, TAP)에 의해 소포체 내강으로 유입된다. 이 펩타이드는 소포체에 위치한 아미노펩티다제인 ERAP 1 및 ERAP2의 추가적인 작용에 의해 HLA class I 분자에 결합하기 적합한 형태인 8-12개의 아미노산 길이로 절단된다. 이후 Tapasin, TAP, calreticulin 및 ERP57로 구성된 펩타이드 로딩 복합체의 도움을 받아 HLA class I 분자에 결합한다. 완전히 조립된 HLA class I 복합체는 소포체를 떠나 세포막으로 전달되고, 제시된 펩타이드는 CD8+ T 세포에 의해 인식되어 세포 독성 면역 반응을 유도하게 한다(그림 1).[4,5]

일반적으로 암세포는 유전적 불안정성을 가지며, 염색체 염기서열의 결실 또는 돌연변이로 인해 유전자 발현에 변화가 생길 수 있다.[6] 따라서 암 치료

의 정밀성을 높이기 위해서는 면역 반응에 관여하는 중요한 요소인 HLA 펩타이드의 유전적 변이와 이에 따른 서열의 변화를 정확히 파악하는 것이 중요하다. 하지만 HLA 펩타이드의 정확한 서열을 기존의 데이터베이스만으로 예측하는 것은 어렵다. 따라서 기존 데이터베이스에 raw data를 매칭해 서열을 확인하는 방식보다는 아미노산 질량 차이를 기반으로 펩타이드 서열을 예측해주는 De novo sequencing 방식을 통해 자체 데이터베이스를 생성하는 것이 효과적인 분석 전략이 될 수 있다. 이는 다양한 암 유래 변이를 수집 및 정리해 하나의 데이터로 정리한 방대한 데이터베이스를 구축할 수 있게 하며, 이는 향후 정밀한 암 치료 전략 수립에 기반이 될 수 있다. 따라서 본 연구에서는 질량 분석을 기반으로 효율적인 데이터베이스 구축을 위한 최적의 분석법 개발을 목표로 실험을 진행했다.

HLA class I의 글라이코실화 패턴 또한 면역 반응에서 중요한 요소라고 할 수 있다. HLA class I 분자의 경우, N-결합 글라이칸은 α 사슬의 86번 위치에서 발견되는 아스파라진(Asparagine, N) 잔기에 부착된다.[7] 일반적으로, HLA class I 분자의 당화는 소포체 막에 위치한 올리고당 전이효소(oligosaccharyltransferase, OST) 복합체에 의해 촉매된다. 이는 초기 HLA class I 폴리펩타이드 사슬의 특정 아스파라긴 잔기에 N-글라이칸을 부착시키며, 특히 단백질이 소포체로 이동한 직후에 $Glc_3Man_9GlcNAc_2$ 구조의 글라이칸을 전달하는 역할을 수행한다.[8] 이후 이 글라이칸은 glucose 잔기 두 개가 제거되어 $Glc_1Man_9GlcNAc_2$의 형태를 가지며, 단백질이 올바르게 접히면 남은 Glc 잔기도 제거되고 단백질은 골지체로 이동할 수 있게 된다. cis-골지체에서는 추가적인 mannose 잔기가 제거되며, medial 및 trans 골지체를 거치면서 GlcNAc, fucose, galactose 등의 당이 부착되며 복합형 글라이칸의 가지가 형성되는 가공이 일어난다. 이러한 당화 성숙을 마친 단백질은 세포 표면으로 수송된다.[9]

HLA class I에 부착된 당 구조는 항원 처리 및 제시 경로의 여러 핵심 구성 요소와의 상호 작용을 조절한다. 예를 들어, HLA class I에 부착된 당이 $Glc_1Man_9GlcNAc_2$일 경우, 이는 샤페론 단백질인 칼넥신(calnexin) 및 칼레티쿨린(calreticulin)과의 상호작용을 촉진해 HLA class I의 접힘 및 조립 과정을 촉진한다. 이후, 칼레티쿨린 및 ERP57과의 상호작용을 통해, 글라이칸은 HLA class I이 타파신에 결합하여 펩타이드 로딩 복합체와 통합될 수 있도록 한다. 또한, HLA class I의 소포체/cis-골지체에서 세포막으로의 이동 역시 부착된 글라이칸에 의해 조절된다. 펩타이드가 적재된 HLA class I 복합체에서 glucosidase I/II에 의해 말단 glucose 잔기가 제거되면, 소포체를 떠나 세포막으로의 이동이 시작된다.[7] 이처럼, HLA 분자의 글라이코실화 여부는 면역 반응의 다양한 단계에 영향을 미치는 중요한 요소로 작용한다. 하지만 암과 같은 특수한 병리적 상황에서는 HLA 글라이칸 생합성 경로에 이상이 생겨 비정상적인 글라이코실화 패턴이 나타날 수 있으며, 이러한 변화는 질병의 바이오마커로 활용될 수 있다. 따라서, HLA class I의 글라이코실화는 면역 반응 조절에 있어 핵심적인 역할을 하므로, 그 구조적 특성을 정확하게 규명하는 것이 중요하다. 이에 본 연구에서는 HLA class I 펩타이드 서열 분석 외에도, N-글라이칸을 분리하여 질량 분석 및 HILIC-UPLC-FLR 검출기를 사용함으로써 HLA class I 분자에 부착되어있는 당의 구조를 분석하고자 하였다.

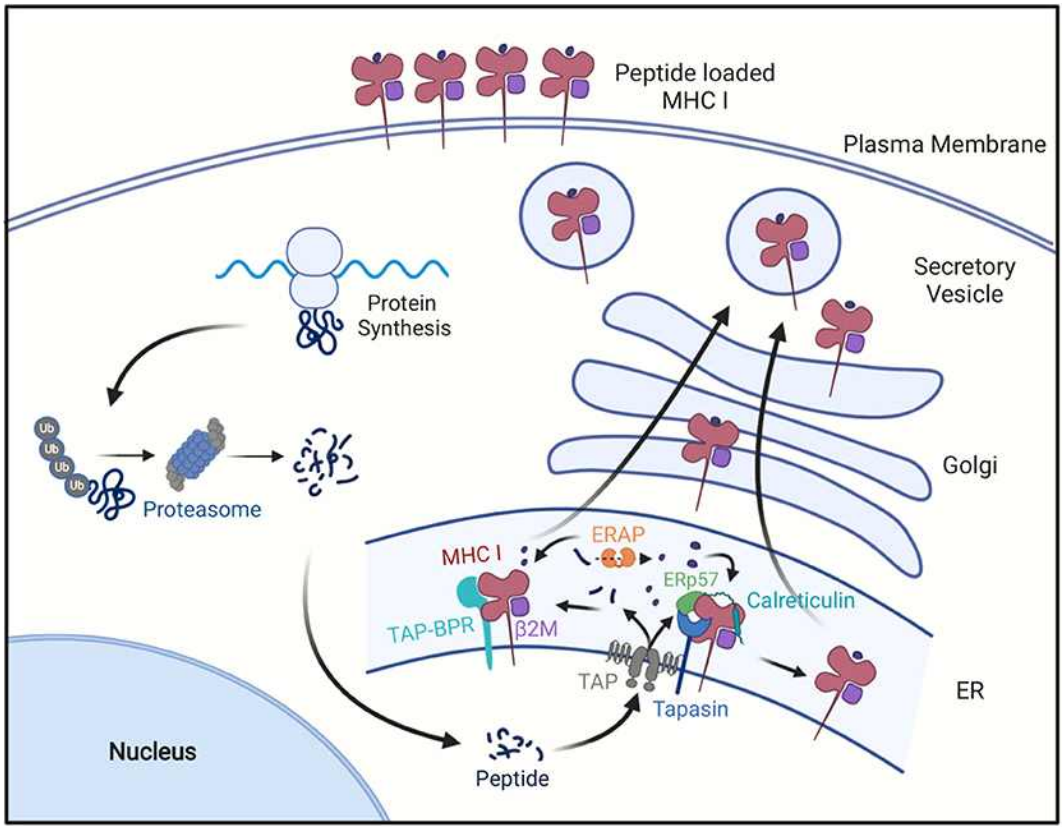


그림 1. HLA class I 펩타이드 제시 경로[5]

프로테아좀에 의해 분해된 올리고펩타이드는 TAP 수송체를 통해 소포체로 운반된다. 이후, ERAP1에 의해 추가로 절단된 펩타이드는 펩타이드 로딩 복합체(Tapasin, TAP, calreticulin, ERP57)의 작용으로 HLA 분자에 결합한다. 조립된 HLA 복합체는 세포 표면으로 운반되어 T 세포에 항원을 제시한다.

단일 클론 항체와 같은 바이오 의약품은 뛰어난 임상적 효능과 표적 특이성으로 인해 제약 개발 산업에서 핵심적인 영역으로 자리잡았다. 이러한 항체는 일반적으로 세포주에서 생산되며, 이 생산 과정에서 전사 후 변형(post-transcriptional modification, PTM) 또는 구조적인 변화가 발생하여 최종 단백질 형태가 결정된다. 이는 항체에 이질성을 초래할 가능성이 있으며, 특히 산화, 탈아미드화, 글라이코실화와 같은 PTM의 조성 변화는 항체의 생물학적 활성을 변화시켜 안정성에 영향을 미칠 수 있다.[10]

단일 클론 항체에서 관찰되는 가장 일반적인 유형의 글라이칸은 Fc 영역의 아스파라긴 위치 297(Asn-297)에 부착되는 N-글라이칸이다.[11] 이 위치에는 $Glc_3Man_9GlcNAc_2$ 올리고당이 소포체 내의 올리고당 전이 효소 복합체에 의해 전달된다. 이후 N-글라이칸은 glycosidase 및 glycosyltransferase 효소들에 의해 연속적인 구조적 가공을 거치며 다양하게 변형된다.[12] 특히, GlcNAc, galactose 그리고 sialic acid가 확률적으로 첨가되는 성질을 가지고 있어 이로 인해 다양한 구조적 형태가 관찰된다.[11]

그 중에서도 G0F, G1F, G2F는 주요 glycoform으로, 항체의 구조적 안정성과 생물학적 기능에 큰 영향을 미치지 않는 정상적인 구조에 해당한다. 이들 간의 차이는 galactose의 부착 여부에 있으며, G0F는 galactose가 없는 형태이고, G1F는 galactose가 하나 추가된 형태, G2F는 galactose가 2개 추가된 구조이다(그림 2)[13]. 일반적으로, 이 세 glycoform의 합산 비율이 80% 이상을 차지하면 항체는 구조적, 품질적인 측면에서 안정적이라고 판단할 수 있다.

구조적 안정성과 더불어, Fc 부위의 글라이코실화는 항체의 면역학적 기능에 중요한 영향을 미친다. 특히 세포 수용체와의 결합이나 ADCC(Antibody-dependent cellular cytotoxicity)와 같은 effector 기능을 조절하는 것으로 알려져 있다. 또한 말단 galactose 잔기는 CDC(Complement

dependent cytotoxicity) 활성과 Fc-C1q 상호작용을 향상시킬 수 있으며, α 1,6-연결된 fucose를 부착시키는 core fucosylation은 ADCC 및 Fc-FcγRs 결합을 저해할 수 있다.[12] 이처럼 글라이칸 특성화는 단순한 구조 분석의 의미뿐 아니라 치료제의 효능과 안정성 판단을 위한 필수적인 요소라고 할 수 있다.

본 연구에서는 상용화된 항체 의약품의 특성을 파악하기 위해 총 5종의 오리지널 항체와 6종의 바이오시밀러 항체, 총 11개 샘플의 주요 글라이칸 패턴을 분석했다. 우선, PNGase F 효소 유무에 따른 항체의 분자량을 LC-MS로 확인하여 글라이칸 조성을 확인했다. 이후 더 정확한 동정을 위해 N-글라이칸만을 분리한 후, HILIC-UPLC-FLR 및 LC-MS 분석을 수행했다. LC-MS 분석에서는 각 크로마토그램 peak별 질량값을 계산하여 글라이칸을 동정하였고 이를 UPLC-FLR 결과 크로마토그램과 비교하여 상호 검증을 수행하였다. 결과적으로, 오리지널 및 바이오시밀러 항체 간의 미세한 글라이칸 패턴 차이를 확인할 수 있었으며, 이와 같은 상호 보완적인 접근 방식은 항체의 품질 분석 결과에 대한 신뢰성을 보여준다.

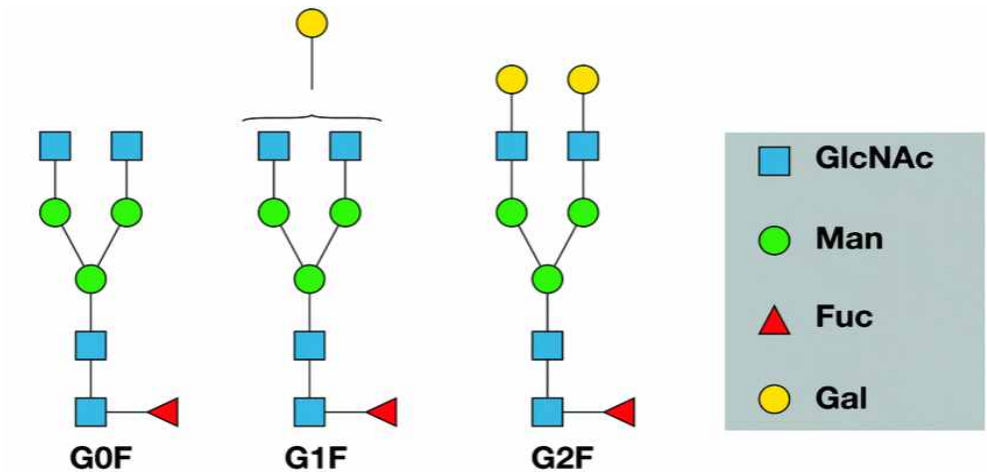


그림 2. 주요 글라이칸(G0F, G1F, G2F)의 구조[13]

콜라겐은 동물에서 가장 풍부한 구조 단백질이며, 인간의 경우 체내 단백질 함량의 약 30%를 차지한다. 콜라겐은 그 구조에 따라 I형, II형, III형과 같은 섬유질 및 IV형, VI형, VII형과 같은 비섬유질 등을 포함하여 총 28가지 유형으로 분류된다. 이중 I형 콜라겐은 가장 일반적인 유형의 콜라겐으로, 동일한 두 개의 α1 사슬과 하나의 α2 사슬로 구성된 삼중 나선 구조를 형성한다. 이 α1 사슬은 일반적으로 Gly-XY 서열의 반복으로 구성되며, 여기서 X와 Y는 각각 프롤린(Proline)과 하이드록시프롤린(Hydroxyproline)이 주로 위치힌다.[14]

I형 콜라겐은 높은 생분해성 및 생물학 활성도를 가지고 있어 화장품 및 제약 산업에서 널리 이용된다. 그러나 동물 유래 콜라겐의 개발은 동물 유래 감염병의 우려와 종교적 제약으로 인해 제한적이다.[15] 이러한 측면에서 해양 생물은 콜라겐의 잠재적인 공급원으로 주목받고 있다. 특히 최근 몇 년간 어류 유래 콜라겐은 콜라겐 수요 증가에 대응하는 시장 제품으로 부상하고 있다. 어류 콜라겐은 유망한 바이오 소재로서 동물성 콜라겐에 비해 체내 흡수율이 1.5배 이상 높으며,[16] 노화 방지 특성과 항균 능력으로 인해 화장품 산업과 기능성 식품 분야에서 활발히 활용되고 있다.[15] 그러나 2021년, 30개국 이상에서 실시한 연구에 따르면 시판되는 해양 생물 유래 제품 중 약 36%에서 원재료 오표기가 발견되었다. 이는 고가의 어종을 상업적 가치가 낮은 유사한 어종으로 대체하는 등의 사기 행위가 발생하고 있는 것을 뜻한다. 이러한 오표기는 알레르기 유발 물질의 존재 가능성을 증가시켜 소비자에게 심각한 건강 질환을 초래할 수 있다.[17] 따라서 제품에 표기된 어종이 실제 사용된 어종과 일치하는지 확인하기 위해서는 엄격한 원료 검증 절차를 거쳐야 하며, 이를 위한 종 특이적인 펩타이드 서열 분석법이 요구된다. 최근에는 ELISA를 이용한 면역학적 접근법과 PCR을 이용하는 DNA 분석법의 개발이 활발히 이루어지고 있다. 하지만 ELISA는 높은 민감도와

선택성을 제공하지만, 콜라겐 간의 아미노산 서열 유사성으로 인해 항체의 교차 반응이 발생할 수 있다는 문제점이 있다. 또한 PCR 기반 DNA 분석은 높은 특이성으로 어류의 기원을 추적할 수 있지만 DNA는 쉽게 분해될 수 있기 때문에 상대적으로 불안정하다는 한계가 있다.[18] 이러한 단점을 보완하기 위한 단백질체학은 어류의 기원 확인을 위한 유망한 접근 방식으로 주목받고 있다. 일반적으로, 단백질은 DNA에 비해 분해와 부패에 강하고, 오염 물질의 영향을 적게 받기 때문에 대량 분석에도 적합하다. 특히 질량 분석 기반 기법을 기반으로 한 펩타이드 분석은 정확한 아미노산 서열을 파악할 수 있어 콜라겐의 서열 동정에 효과적으로 활용된다.[19]

따라서 본 연구에서는 액체 크로마토그래피가 결합된 Orbitrap 질량 분석기를 이용하여 틸라피아, 대구, 명태 총 3종의 어류 유래 저분자 펩타이드 혼합물의 분석을 수행했다. 혼합물은 역상 크로마토그래피를 사용해 분리된 뒤 이온화되어 질량 분석이 수행되었고 이를 통해 각 펩타이드에 대한 질량 스펙트럼을 얻을 수 있었다. 분석 결과, 각각의 시료에 존재하는 펩타이드 아미노산 서열을 기반으로 한 고유한 질량 스펙트럼을 확인함으로써 서로 다른 어종 사이의 서열 상동성과 차이점을 확인할 수 있었다. 이 과정에서 얻어진 특정 m/z 값의 peak는 바이오마커로 활용할 수 있으며, 이를 콜라겐 원료 어종의 명확한 식별에 이용하였다. 결과적으로, LC-MS/MS 분석은 서로 다른 어종 간의 특이적 펩타이드 서열 및 서열 상동성을 확인하는 데 있어 효과적인 정보를 제공하는 방식임을 확인할 수 있었다.

Ⅱ. 재료 및 실험방법

1. 세포주 및 실험 재료

본 연구에서 HLA class Ⅰ 분석에 사용된 인간 대장암 세포주 HCT-116은 오송첨단의료산업진흥재단으로부터 제공받았다. Cell lysis에는 Corning 사에서 제작한 3 mL Tissue grinder(Cat no. 7724-3)를 사용했으며 면역 침강 실험에는 앱클론에서 생쥐 복수(Ascites) 방식을 통해 생산한 W6/32 항-HLA class Ⅰ항체와 Cytiva 사의 Protein A Sepharose fast flow (Cat no. 17127902)를 사용했다. HLA class Ⅰ 펩타이드 용출은 Costar spin-X column(Cat no. 8160)을 구매해 사용했고 정제 과정에서는 Sep-pak tC18 cartridge (Cat no. WAT054960)을 통해 HLA 펩타이드를 분리했다. HLA 및 항체 시료의 N-글라이칸 분석은 Waters 사의 GlycoWorks RapiFluor-MS N-glycan kit(Cat no. 176004154)를 이용하여 수행했다.

콜라겐 실험에 사용된 시료는 총 9개로, 모두 식품 의약품 안전처로부터 제공받은 상용 콜라겐 제품이다. 사용한 어종은 틸라피아(Tilapia), 대구(Cod), 명태(Pollock)의 세 종으로, 어종별로 서로 다른 콜라겐 제품을 3종씩 선정해 분석하였다. 틸라피아 유래 콜라겐은 No. 23, 157, 176, 대구 유래 콜라겐은 No. 16, 28, 153, 명태 유래 콜라겐은 No. 116, 156, 173으로 준비하였다.

치료용 단일클론 항체의 글라이칸 특성화 실험에 사용된 항체 11종(Humira, Yuflyma, Mabthera, Truxima, Avastin, Alymsys, Vegzelma, Remicade, Remsima, Herceptin, Herzuma)은 모두 식품 의약품 안전처로부터 제공받았다.

2. HLA Class Ⅰ 분석

2-1. Cell lysis 시간 설정

(1) cell lysis

-80°C에서 보관하였던 대장암 세포주 HCT-116 1.3×10^8 cell 총 5개를 녹인 후, 각 튜브에 lysis buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 1 mM Na3VO4, 1% Triton X-100, 10% Glycerol, 1/2 정량의 cOmplete™ Mini EDTA free Protease Inhibitor Cocktail) 1 mL를 첨가하였다. 이후 파이펫팅을 통해 세포 pellet과 lysis buffer를 충분히 혼합하였으며, 혼합물은 glass tissue grinder에 옮겨 유리 막대를 이용해 총 35회 왕복하며 파쇄하였다. 이어서, 세포 파쇄물을 2 mL 마이크로 튜브에 옮긴 후, 남은 잔여물을 회수하기 위해 tissue grinder의 벽면에 cell lysis buffer 1 mL를 넣고 파이펫팅하여 잔여 세포를 동일한 튜브에 모았다. 이후 각각의 시료를 4°C에서 4시간, 8시간, 18시간, 24시간, 48시간 동안 반응시켜 lysis를 진행했다.

(2) 단백질 정량

세포 용해가 완료된 후, 시료를 13,000 RPM에서 2분간 원심분리하여 상층액을 분리하였다. 이후 상층액을 얻어 BCA assay를 수행하였다. 단백질 정량을 위한 standard curve를 작성하기 위해 Sigma 사의 BSA(Bovine serum albumin, Cat no. A7030)를 사용했으며, 0.01, 0.025, 0.05, 0.1, 0.5, 0.75, 1, 2 mg/mL의 농도로 희석하여 96-well plate에 10 µL씩 분주하였다.

실험 시료는 각각 1/2, 1/5, 1/10, 1/20, 1/50으로 희석한 후 동일하게 각 well에 10μL씩 분주하였다. 이후 Pierce™ BCA Protein Assay Kit의 Reagent A solution과 Reagent B solution을 각 49:1로 혼합한 reagent를 각 well에 200μL씩 첨가하였다. Plate는 호일로 감싼 뒤 37°C 인큐베이터에서 30분간 반응시켰다. 반응 후, 562 nm에서 흡광도를 microplate Reader로 측정했으며, standard curve를 기반으로 단백질 농도를 계산하였다.

(3) 웨스턴 블롯(Western blot)

lysis 시간에 따른 HLA class Ⅰ의 양을 확인하기 위해 웨스턴 블롯을 수행하였다. 우선 cell lysate 6 μg에 4× Laemmli Sample Buffer (Cat no. 161-0747, Bio-rad) 2.5 μL를 혼합하고, 증류수를 넣어 최종 부피 10 μL의 시료를 준비하였다. 이후 SDS PAGE용 running gel과 stacking gel을 제조한 뒤, gel이 다 굳으면 준비된 시료 10 μL를 각 well에 로딩하고, 단백질 크기 확인을 위해 protein ladder (Cat no. BR1610374, Bio-rad) 5 μL를 함께 로딩하였다. 전기영동은 80 V에서 110 V로 순차적으로 전압을 올리며 1시간 30분간 진행했다. 전기영동이 종료되면, gel에서 membrane으로 단백질을 옮기는 transfer 단계를 진행하기 위해 PVDF membrane (Cat no. IPVH00010, Millipore)을 겔 크기보다 크게 잘라 100% methanol에 4분간 넣고 활성화시킨 뒤, 1× transfer buffer에 담가두었다. Transfer cassette는 아래에서부터 PAD, filter paper, PVDF membrane, gel, filter paper, PAD 순으로 배치해 조립하였다. 마지막 Filter paper를 덮은 후에는 기포 제거를 위해 15ml conical tube를 사용해 상하좌우로 눌러가며 밀어주었다. 조립 과정에서는 membrane이 마르지 않도록 transfer buffer를 지속적으로 적셔주었다. 조립된 cassette는 PowerPacTM HC High-Current Power Supply 장치를 사용하여 100 V에서 2시간 동안 전기영동 시켰다. Transfer 과정 중에는

발열에 대비하여 얼음팩을 넣고 지속적으로 교체해주었다. Transfer가 완료된 membrane은 1× TBST로 세척한 후, skim milk powder로 만든 blocking solution에 담가 4°C shaker에서 overnight을 수행하였다. 다음날 membrane을 1x TBST로 15분간 세척해준 후, blocking solution에 항-HLA class Ⅰ W6/32 항체를 2000:1로 희석한 용액과 상온에서 1시간 동안 반응시켰다. 1차 항체 반응 후에는 1× TBST 용액으로 30분간 세척했고 이어서 blocking solution에 HRP가 부착된 goat 유래 anti-mouse 2차 항체를 20000:1 비율로 희석한 용액과 상온에서 1시간 동안 반응시켰다. 반응 종료 후에는 동일하게 1× TBST로 30분간 세척하였다. 이후 단백질 검출을 위해 ECL solution A와 B를 1:1 비율로 혼합하여 chemiluminescent substrate를 제조한 뒤, 이를 membrane에 골고루 뿌려 반응시켰다. 이후 SHST Multifunctional Imager를 사용해 단백질 밴드를 관찰했다.

2-2. HLA 펩타이드 분리 및 분석

(1) 면역 침강(Immunoprecipitation)

먼저 Protein A beads를 100 μL 준비하여 1× PBS(pH 7.4) 1 mL로 세척하였다. 세척은 탭핑 후 4°C에서 4,000 RPM으로 2분간 원심분리하는 과정을 2회 반복하여 수행했다. 이후 항-HLA class Ⅰ항체(W6/32) 1 mg을 beads와 혼합하고 4°C에서 3시간 동안 rotator에서 반응시켰다. 반응 후에는 동일하게 1× PBS(pH 7.4) 1 mL를 넣고 4°C 4,500 RPM에서 5분간 원심분리하여 상층액을 제거하는 과정을 2회 반복했다. 이후 항체가 결합된 beads 혼합물에 전 단계에서 얻은 용해된 lysate의 상층액을 첨가하여, 4°C rotator에서 overnight 반응을 진행했다.

(2) HLA 펩타이드 용출

Overnight 반응이 완료된 시료는 4°C, 4,500 RPM으로 5분간 원심분리한 뒤, 상층액을 제거하였다. 그런 다음, 항체, beads, 및 HLA 복합체 침전물에 MS grade Water 500 µL를 넣고 잘 섞어주었다. 이후, beads를 효과적으로 분리하기 위해 Costar spin-X column(Cat.no. 8160)을 준비하였다. 먼저 column에 MS grade Water 500 µL를 넣고 4,100 RPM에서 3분간 원심분리하여 column을 활성화했다. 다음으로, 앞서 준비한 시료 혼합물을 column에 옮기고 8,000 RPM에서 20분간 원심분리하였다. 이후, 시료가 들어있던 튜브에 다시 MS grade Water 500 µL를 넣고 남은 잔여물과 잘 섞은 후, 이를 같은 column에 옮겨 10,000 RPM에서 30분간 원심분리했다. 마지막으로 세척을 위해 MS grade Water 500 µL를 한 번 더 column에 넣어 10,000 RPM에서 30분간 원심분리를 수행하였다. 마지막으로 모든 시료가 column 하단으로 충분히 내려간 것을 확인한 후에는 새로운 마이크로 튜브를 column 하부에 받치고 10% acetic acid 200 µL를 column 상단에 넣어 4°C에서 10분간 반응시켰다. 이후 8,000 RPM에서 30분간 원심분리하여 시료 용출을 진행했다.

(3) 정제

용출된 시료 복합체로부터 HLA class Ⅰ 펩타이드를 분리하기 위해 정제 과정을 수행하였다. 우선 Sep-pak tC18 cartridge(Cat no. WAT054960)를 시험관 위에 설치하여 준비하였고, 카트리지를 활성화하기 위해 MS grade acetonitrile 1 mL를 넣고 syringe를 이용해 천천히 통과시켰다. 다음으로, 카트리지 평형화를 위해 solvent A (5% Acetonitrile, 95% water,

0.1% formic acid) 1 mL를 세 번 연속 통과시켰다. 앞서 용출한 시료 200 µL에 solvent A 800 µL를 첨가하여 잘 혼합한 후, 이를 카트리지에 로딩하고, 원래 시료가 담겨있던 마이크로 튜브를 아래에 받쳐 시료가 완전히 통과되도록 하였다. 이후 카트리지를 다시 시험관 위에 올린 뒤, 통과된 시료를 동일 카트리지에 재로딩하여 한 번 더 필터링하였다. 이어서 solvent A를 동일한 방식으로 1 mL씩 세 번 반복하여 통과시키며 세척을 수행하였다. 마지막으로, 새 마이크로 튜브를 카트리지 하단에 설치하고 70% acetonitrile을 250 µL씩 두 번에 걸쳐 통과시켜 HLA 펩타이드를 용출하였다. 최종적으로 용출된 시료에서 70% Acetonitrile을 진공 농축기를 이용해 모두 증발시킨 후, 시료를 - 20°C에서 보관하였다.

(4) LC-MS/MS 조건

앞선 전처리 과정을 통해 정제된 HLA 펩타이드는 Thermo Fisher사의 Vanquish Neo UHPLC 시스템과 Q-Exactive HF Orbitrap 질량 분석기를 이용하여 NanoLC-MS/MS 방식으로 분석하였다 (표 1). 분석 컬럼으로는 EASY-Spray™ PepMap™ Neo(Cat. No. ES75500PN)를 사용하였다. NanoLC는 µL/min 단위의 낮은 유속을 사용하기 때문에 이온화 효율이 크게 향상되어 극소량의 시료로도 고감도의 분석이 가능하다. 추가로 PepMap™ Neo 트랩 카트리지 (Cat. 174502)를 사용해 분석 전 샘플의 정제 및 농축을 수행함으로써 질량 분석의 효율을 높였다. 이동상은 A라인에 0.1% Formic acid in Water, B라인에 0.1% Formic acid in Acetonitrile로 준비하였으며, 세척 용액은 0.1% FA in 75% isopropanol (IPA)로 제조하였다. 분석 전, Pierce사의 LTQ Positive Ion Calibration Solution (Cat. No. 88322)을 사용하여 positive mode에서의 질량 분석기 보정을 수행하였다. 또

한, 컬럼의 안정화를 진행하였으며, 이후 최종 유속은 0.25 µL/min으로 설정하였다. 컬럼 온도는 40°C, 샘플 온도는 4°C로 유지하였다. 이동상 조성은 0.1% FA in ACN 2%에서 시작하여 90분간 42%까지 선형 증가시키는 gradient 조건으로 설정하였으며, 분석 종료 후에는 자동으로 10분간의 컬럼 세척이 이루어졌다(표 1).

질량 분석 조건은 다음과 같다(표 2). NSI source에서 sheath gas flow rate, Aux gas flow rate, sweep gas flow rate는 모두 0으로 설정하였으며, Spray voltage는 1.8kV, Capillary temperature는 275°C로 설정하였다. 펩타이드는 Data Dependent Acquisition (DDA) 방식으로 분석하였으며, full MS scan에서는 m/z 범위 400 - 2000, 해상도 120,000, AGC target $3e^6$으로 설정하였다. 이후 intensity 기준 상위 15개의 이온을 선택하여 MS/MS 분석을 수행하였고, 이때 해상도는 15,000, AGC target $1e^5$, normalized collision energy (NCE)는 27로 설정하였다.

표 1. LC-MS/MS gradient

Time	%A	%B	Flow [μL/min]
initial	98	2	0.25
5	98	2	0.25
65	72	28	0.25
90	58	42	0.25
90		Column Wash	

표 2. Mass spectrometer 분석 조건

Thermo scientific Q-Exactive HF Orbitrap	
Experiments	Full MS/ dd MS^2(Top N)
Runtime	100 min
Polarity	Positive
In-source CID	0.0 eV
Full MS	
Resolution	120000
AGC target	$3e^6$
Maximum IT	100
scan range	400 to 2000 m/z
dd-MS^2	
Resolution	15000
AGC target	$1e^5$
Maximum IT	50 ms
Loop count	15
Top N	15
Isolation window	1.6 m/z
(N)CE	27
Charge state	+2,+3

(5) 데이터베이스 구축 및 서열 동정

일반적으로 암세포는 유전적으로 불안정하여 염색체의 결실 또는 돌연변이로 인해 유전자 발현의 변화가 발생할 수 있다. 이로 인해, 기존의 데이터베이스에 raw data file을 대조해 펩타이드의 서열을 예측하는 것은 한계가 있다. 따라서 아미노산 질량 차이를 기반으로 펩타이드 서열을 예측해주는 De novo sequencing을 사용해서 자체 데이터베이스를 생성하고 이를 적용하는 방식이 보다 효과적인 접근법이라 할 수 있다.

우선 PEAKS Studio 소프트웨어를 이용하여 +1부터 +4까지의 precursor charge state를 기준으로 데이터를 필터링하는 Data Refinement 과정을 수행하였다. 이후 De novo 분석 조건은 precursor ion에 대한 error tolerance는 20 ppm, fragment ion에 대한 error tolerance는 0.2 Da으로 설정하였다. De novo 분석을 통해 얻은 펩타이드 서열들은 통합해 하나의 FASTA 파일로 변환하여 자체 database를 생성하였으며, 해당 파일은 Thermo Fisher사의 Proteome Discoverer 3.2 소프트웨어에서 데이터베이스로 사용하여 추가적인 필터링 및 검증 과정을 거쳤다. Proteome Discover 스프트웨어의 상세 분석 조건은 다음과 같다(표 3).

표 3. Proteome Discoverer 분석 파라미터

Max. Missed Cleavage	2
Min. Peptide Length	8
Max. Peptide Length	12
Precursor Mass Tolerance	20 ppm
Fragment Mass Tolerance	0.2 Da
Enzyme	None

2-3. HLA class I 분자의 글라이칸 분석

(1) 항체 제거 및 시료 준비

Costar Spin-X 컬럼을 통해 얻어진 항체, HLA class I 복합체 혼합물에서 글라이칸이 검출될 수 있는 항체를 제거하기 위해 Amicon Ultra 100K Centrifugal Filters (Cat. No. UFC510024, Millipore)를 사용하였다. 먼저, 필터를 활성화하기 위해 MS-grade water 500 µL를 첨가한 후 4°C, 14,000 RPM에서 10분간 원심분리하였다.

활성화 후, 새로운 마이크로 튜브를 받친 상태에서 용출된 시료 200 µL를 필터에 가하고 동일 조건(4°C, 14,000 RPM, 10분)으로 원심분리를 수행하였다. 필터링을 통해 상부에 남아있는 항체를 제거한 뒤, 진공 농축기를 사용해 하부에 수집된 용출액의 10% acetic acid를 완전히 증발시킴으로써 최종 시료를 확보하였다.

(2) 글라이칸 분석 전처리

건조된 시료에 MS-grade water 20 µL를 첨가한 뒤, 충분히 섞어주었다. 변성 시약은 RapiGestTM SF Surfactant 10 mg에 5× GlycoWorksTM Rapid Buffer 200 µL를 가하여 완전히 용해시켜 준비하였으며, 이 중 6 µL를 시료에 첨가하여 혼합했다. 이후 water 2 µL를 추가하여 최종 부피를 28 µL로 조정한 뒤, 90°C에서 3분간 shaking하여 변성 반응을 수행하였다. 다음으로, 시료를 실온에서 3분동안 방치한 뒤, Glycoworks Rapid PNGaseF (Cat no. P0710S) 2 µL를 첨가하고 파이펫팅하여 충분히 혼합한 다음 55°C에서 10분간 shaking하여 탈당 반응을 진행하였다. 탈당 반응 후에도 다시 시료

를 실온에서 3분간 방치한 후 형광 라벨링을 수행하였다. 라벨링 시약은 GlycoWorksTM RapiFluor-MS Reagent Powder 23 mg에 GlycoWorksTM Reagent Solvent 335 µL를 가하여 용해한 후 준비하였으며, 이 중 12 µL를 시료에 첨가하였다. 라벨링 반응은 상온에서 10분간 shaking하여 수행하였다(그림 3).

마지막으로, GlycoWork^{TM} HILIC Elution Plate (Cat. No. 186002780, Waters)를 사용하여 글라이칸 정제를 수행하였다. 모든 정제 과정은 Waters Positive Pressure-96 Processor를 통해 일정한 압력을 가하여 수행되었다. 우선 각 사용 예정 well에 MS-grade water 200 µL를 넣어 고정상을 활성화하였다. 이후 85% acetonitrile (ACN) 200 µL를 두 차례 반복하여 통과시켜 well을 평형화시켰다. 준비된 시료 400 µL를 well에 가하고 압력을 가해 통과시킨 후, 세척을 위해 formic acid:water:ACN=1:9:90 (v/v/v) 용액 600 µL를 두 차례 반복하여 적용하였다.

글라이칸 용출은 새로운 96-well 플레이트를 받친 상태에서 200 mM ammonium acetate in 5% ACN 용출 용액을 30 µL씩 세 차례 반복하여 통과시켜 수행하였으며, 수집된 용출액에는 DMF/ACN 희석액 210 µL를 추가하여 희석하였다(그림 4).

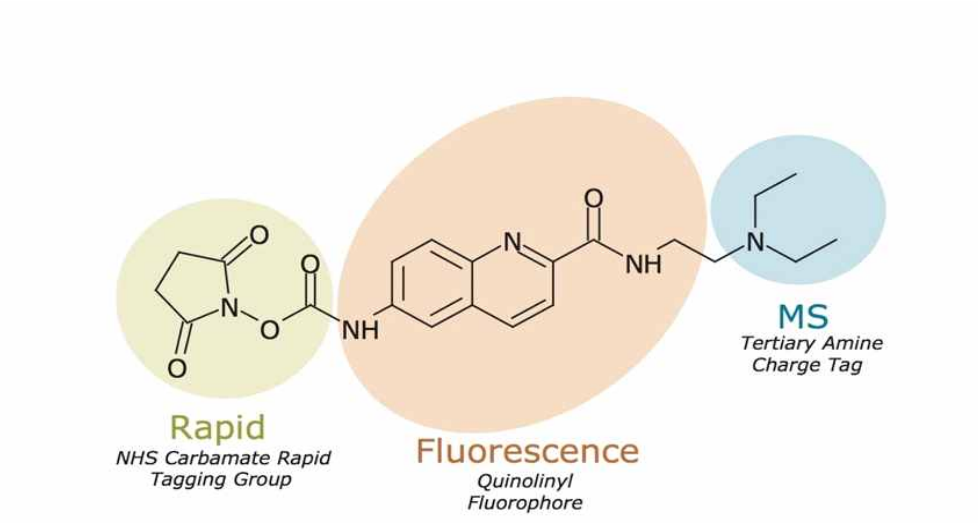


그림 3. RapiFluor-MS 형광 물질 구조[20]

NHS 카바메이트 반응기와 퀴놀린 기반 형광체 그리고 3차 아민 그룹으로 이루어져 있다.

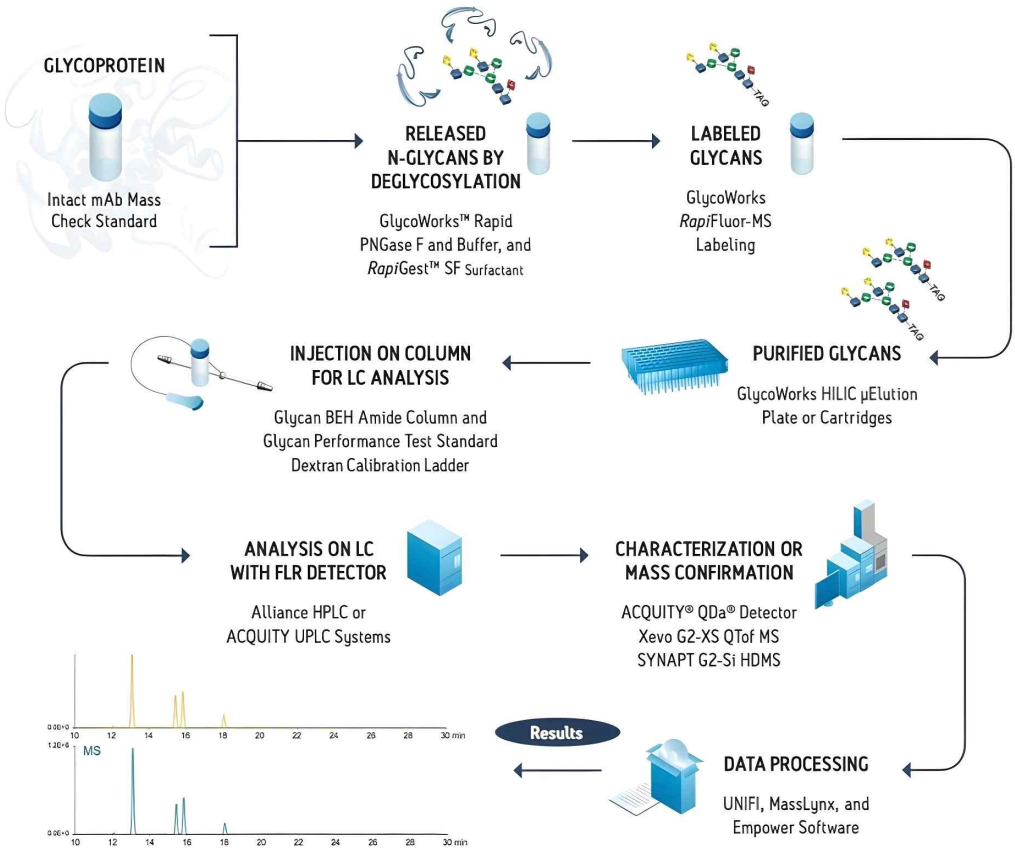


그림 4. GlycoWorks RapiFluor-MS N-glycan kit workflow[20].

HLA class I 시료에 변성 시약, PNGase F, 형광 라벨링 시약을 순차적으로 처리한 후, HILIC plate를 사용해 N-글라이칸을 얻어냈다.

(3) 글라이칸 LC-FLR 분석

전처리를 통해 얻어진 라벨링된 글라이칸은 AQUITY Premier UPLC 시스템(Waters)을 사용해 분석했다. 형광 라벨링된 글라이칸 검출을 위해 FLR(Fluorescence) 검출기(EX 265 nm/EM 425 nm)를 사용했으며, 컬럼은 ACQUITY UPLC Glycan BEH Amide Column (130Å, 1.7 µm, 2.1 × 150 mm)을 사용했다. 이동상은 A 라인 50mM ammonium formate pH 4.4, B 라인 100% ACN으로 구성했다, 분석 전에는 A 라인과 B 라인에 각각 5분간 prime을 실시했고, needle wash는 80% ACN으로 99초간 진행했다. seal wash는 10% ACN으로 실행했다. 또한 sample manager 부분은 15초간 10회 prime을 수행했다. 분석은 컬럼 안정화 후 시작했으며, 유속은 0.4 mL/min으로 설정했다. 분석 중 컬럼 온도는 60°C, 샘플 보관 온도는 4°C로 유지했다. Gradient 조건은 이동상 A의 비율이 25%에서 시작하여 40분간 선형적으로 100%까지 증가하는 방식으로 설정했다(표 4).

(4) 글라이칸 LC-MS 분석

글라이칸의 질량 분석은 ACQUITY UPLC I-Class PLUS(Waters) LC와 연결된 Xevo G2-XS QTOF(Waters)를 이용하여 수행하였다. 컬럼 및 이동상, gradient 조성은 앞선 형광 분석에서 사용한 조건과 동일하게 수행했다. 분석에 앞서 200 pg/µL 농도의 Leucine-enkephalin을 질량 분석기에 연결한 뒤, 3회 purge하여 556.28 m/z 값이 안정적으로 도출되는지 확인하였다. 이어서 sodium iodide를 세 차례 purge하여 scan range인 100 - 2000 m/z 범위 전반에 걸친 mass calibration을 수행하였다. MS 분석 조건은 다음과 같다(표 5).

표 4. LC-MS 분석 gradient

Time (min)	Flow rate (mL/min)	%A	%B	Curve
initial	0.4	25	75	6
35	0.4	46	54	6
36.5	0.2	100	0	6
39.5	0.2	100	0	6
43.1	0.2	25	75	6
47.6	0.4	25	75	6
55	0.4	25	75	6

표 5. Mass spectrometer 분석 조건

Parameter	Condition
Source	
Capillary Voltage(kV)	2.6
Sampling cone	26
Source offset	80
Temperature (°C)	
Source	120
Desolvation	250
Gas Flows	
Cone Gas(L/h)	0
Desolvation Gas(L/h)	800

3. 치료용 항체 분석

3-1. 치료용 항체의 글라이칸 패턴 분석

(1) 시료 준비 및 LC-MS 분석 조건

항체 시료는 총 11종으로, 오리지널 항체 의약품인 Humira, Mabthera, Avastin, Remicade, Herceptin과 이들의 바이오시밀러인 Yuflyma (Humira), Truxima(Mabthera), Alymsys 및 Vegzelma(Avastin), Remsima (Remicade), Herzuma(Herceptin)를 준비했다.

먼저, -20°C 냉동고에 보관되어 있던 항체 시료들을 4°C에서 해동시켜 준비한다. 이후 N-글라이칸이 분리된 항체의 분자량을 구하고자 각 시료의 단백질 양과 PNGase F 용량이 10:1 비율이 되도록 혼합해 충분히 파이펫팅 해준 후 37°C, 300 RPM으로 overnight 반응을 진행한다. 반응 완료 후, 시료는 MS grade water를 사용해 최종 농도가 0.075 mg/mL가 되도록 희석했다. 대조군으로는 PNGase F 처리 없이 동일 농도로 희석한 항체 시료를 준비했다. 이후 준비된 모든 시료들을 Q-TOF로 분석하여 분자량 차이를 확인했다.

질량 분석에는 Waters 사의 UPLC I-CLASS PLUS와 결합된 Waters Xevo G2-XS Q-TOF를 사용했다. 분석에 사용한 컬럼은 BioResolve^{TM} RP mAb Polyphenyl, 450Å, 2.7 µm 2.1 × 50mm (Cat no. 186008944)으로, 주로 항체와 같은 고분자량 물질 분석에 적합하다. 분석 중 컬럼 온도는 80°C, 샘플 보관 온도는 4°C로 설정하였다. 이동상은 A 라인에 0.1% FA in water, B 라인에 0.1% FA in ACN을 사용하였고, 초기 조성은 B라인 5%

에서 시작해 3.7분까지 95%로 선형 증가하는 gradient를 사용하였다(표 6). Seal wash와 Needle wash에는 각각 10% ACN과 80% ACN을 사용하였다. 자세한 질량 분석 조건은 표 7에 정리하였다. 분석 시간은 LC 조건과 동일하게 6.5분으로 설정했고 Full mass scan range는 400~4000 m/z로 설정했다.

(2) Deconvolution을 통한 분자량 분석

질량 분석 후, 스펙트럼 해석을 통해 분자량을 도출할 수 있는 deconvolution 분석을 수행하기 위해 Waters UNIFI 소프트웨어를 사용했다. 사용한 parameter condition은 다음과 같다. MaxEnt1 항목에서는 각 시료의 질량 스펙트럼이 포함되는 1500~4000 Daltons/e 범위를 input m/z range로 입력해 분석 범위를 설정하였고, Iterations 15, Output resolution 0.2 Daltons, Minimum intensities는 left, right 모두 33%로 설정하였다.

표 6. LC-MS 분석 gradient

Time (min)	Flow rate (mL/min)	%A	%B	Curve
initial	0.4	95	5	6
1	0.4	95	5	6
1.01	0.2	95	5	6
3.50	0.2	5	95	6
3.70	0.4	5	95	6
4.00	0.4	95	5	6
4.50	0.4	5	95	6
5.00	0.4	95	5	6
5.50	0.4	95	5	6

표 7. Mass spectrometer 분석 조건

Parameter	Condition
Source	
Capillary Voltage (kV)	3
Sampling cone	150
Source offset	80
Temperature (°C)	
Source	150
Desolvation	500
Gas Flows	
Cone Gas(L/h)	0
Desolvation Gas(L/h)	800

(3) 질량 분석 기반 글라이칸 동정

-20°C에서 냉동 보관한 항체 시료들을 4°C에서 해동시켜 실험에 사용했다. 각 시료는 20 µg씩 분주한 후 변성시약, PNGase F 및 형광 시약을 처리했고 마지막으로 HILIC 정제를 수행했다. 모든 과정은 앞선 HLA class I 글라이칸 획득 실험의 단계와 동일하다(그림 4). 이후 전처리 과정을 통해 얻어진 글라이칸은 AQUITY Premier UPLC 시스템 및 AQUITY I-Class와 연결된 Q-TOF로 분석하였으며 각각의 분석 파라미터 조건 역시 앞선 실험과 동일하게 설정하였다(표 4, 표 5).

4. 어류 저분자 콜라겐 펩타이드 분석

4-1. 콜라겐 펩타이드 서열 동정

(1) 시료 준비 및 분석 장비 조건

틸라피아, 대구, 명태의 총 9종 콜라겐(표 8)을 일부 채취해 MS-grade Water에 녹여 0.5 mg/mL로 분주하였다. 서열 분석에는 Thermo사의 Vanquish Neo UHPLC 시스템과 Q-Exactive HF Orbitrap 질량 분석기가 사용되었다. 또한 75 μm의 미세한 내경을 가진 EASY-Spray™ Pepmap™ Neo 컬럼을 사용하여 혼합물 내 펩타이드를 더 정밀하게 분리하였다. 추가로 PepMap™ Neo 트랩 카트리지(Cat no. 174502)를 함께 사용함으로써 분석 전 단계에서 샘플 내 불순물을 제거했다.

(2) 분석 전 장비 보정 및 설정

분석 전에는 정확한 m/z 측정을 통한 데이터 신뢰성 확보를 위해 Pierce 사의 LTQ Positive Ion Calibration Solution을 사용하여 질량 분석기 보정을 수행하였다. 이후 컬럼 안정화를 진행하였으며, 최종 flow는 0.25 μL/min으로 설정하였다. LC 분석은 역상 크로마토그래피 방식으로 진행했으며 0.1% FA(formic acid) in water 및 0.1% FA in ACN을 각 A 라인과 B 라인 이동상으로 사용했다. 또한, 시료 분석 전후 세척을 위해 0.1% FA in 75% IPA(Isopropanol)을 사용했다.

(3) LC-MS/MS 분석 조건

시료 분석에 사용된 컬럼은 40°C로 유지했고, 샘플 온도는 4°C로 설정하였다. 전체 분석 시간은 95분으로, 이동상 조성은 B 라인 2%에서 시작해 85분간 42%까지 선형 증가시키는 gradient 조건으로 설정하였다. 분석 종료 후에는 자동으로 10분간의 컬럼 세척 및 안정화가 이루어졌다(표 9). 시료 분석 전후에는 컬럼 잔류물 여부 확인과 클리닝을 위해 공시료(blank)를 각각 주입하였다. 분석은 positive mode로 수행되었으며, m/z scan range는 200~1800 m/z으로 설정했다. 또한, full scan 후 상위 intensity를 가진 이온을 선택해 MS/MS 분석을 수행하는 방식인 DDA(Data-Dependent Acquisition)를 사용하였다. Ion source 조건인 sheath gas flow rate, Aux gas flow rate, sweep gas flow rate는 모두 0으로 설정하였으며, Spray voltage는 1.8 kV, Capillary temperature는 275°C로 설정하였다. 자세한 질량 분석 조건은 다음과 같다(표 10). full scan에서의 해상도는 120,000, AGC target은 $3e^{6}$로 설정하였고 MS/MS 분석에서는 해상도 15,000, AGC target $1e^{5,}$ normalized collision energy (NCE)는 27로 설정하였다. 주입된 펩타이드는 전기분무이온화(electrospray ionization, ESI)을 통해 질량분석기 내부로 주입되었고 이에 따른 m/z 스펙트럼을 기반으로 분석을 진행하였다.

표 8. 실험에 사용한 콜라겐 샘플 목록

No	검체명	제형	이름	원산지	Certificate 유무	분자량(Da)	콜라겐/젤라틴 함유량(첨가물 유무)
16	어바들오브네이쳐 노르웨이 피쉬콜라겐 180	Powder	대구	노르웨이 (Seagarden AS)	o	500~2000	98.90%
23	피쉬 콜라겐	Powder	틸라피아	베트남,방글라데시 (한국 젤텍)	o	1000	100%
28	코드 피쉬 콜라겐	Powder	대구	노르웨이 (Seagarden AS)	o	500~2000	99%
116	그린약초 저분자 피쉬 콜라겐	Powder	명태(Alaska Pollock)	러시아	o	300 이하	100%
153	오리진 마린 피쉬 콜라겐 100	Powder	대구	노르웨이	o	1000	99%
156	저스트콜라겐 초저분자 피쉬 콜라겐	Powder	명태(Alaska Pollock)	러시아	o	300	100%
157	준코 프랑스저분자 피쉬 콜라겐	Powder	틸라피아	프랑스 (LESAFFRE INGREDIENTS SERVICES) Naticol BPMG	o	2000 이하	100%
173	The best 저분자 피쉬 콜라겐 펩타이드	Powder	명태(Alaska Pollock)	베트남 (VINH WELLNESS)	o	500	100%
176	좋은아침 저분자 피쉬 콜라겐	Powder	틸라피아	프랑스 (LESAFFRE INGREDIENTS SERVICES)	o	2000	100%

표 9. LC-MS/MS gradient

Time	%A	%B	Flow [μL/min]
initial	98	2	0.25
5	98	2	0.25
60	72	28	0.25
85	58	42	0.25
85	Column Wash		

표 10. Mass spectrometer 분석 조건

Thermo scientific Q-EXactive HF Orbitrap	
Experiments	Full MS/ dd MS^2(Top N)
Runtime	95 min
Polarity	Positive
In-source CID	0.0 eV
Full MS	
Resolution	120000
AGC target	3e6
Maximum IT	100
scan range	200 to 1800 m/z
dd-MS^2	
Resolution	15000
AGC target	1e5
Maximum IT	50 ms
Loop count	15
Top N	15
Isolation window	1.6 m/z
(N)CE	27
Charge state	+1, +2, +3, +4

(4) 결과 데이터 분석

분석 종료 후, raw data file을 PEAKS Studio 소프트웨어의 De novo sequencing 기능을 사용해 분석했고, 이를 통해 대략적인 펩타이드 서열을 예측했다. 우선 +1～+4까지의 precursor charge state를 설정해 데이터를 필터링하는 Data Refinement를 수행했다. 이후, De novo 분석 parameter 설정 시, precursor ion 20 ppm, fragment ion 0.5 Da의 error tolerence를 적용했다. 또한, 콜라겐의 특성상 Proline과 Lysine에 hydroxylation이 빈번하게 일어나기 때문에 이를 fixed PTM으로 설정하여 분석을 진행하였다. 이 결과로 예측된 서열은 FASTA 파일로 변환하여 Proteome Discoverer 소프트웨어에서 데이터베이스로 활용함으로써 교차검증을 진행했다. Proteome Discover 소프트웨어에서 사용한 분석 조건은 표 11에 정리하였다.

표 11. Proteome Discoverer 분석 파라미터

Min. Peptide Length	7
Max. Peptide Length	25
Precursor Mass Tolerance	20 ppm
Fragment Mass Tolerance	0.5 Da
Enzyme	None
Static Modifications	Hydroxylation(P, K)

Ⅲ. 결 과

1. HLA Class I 분석 결과

1-1. 펩타이드 분석 결과

(1) 단백질 정량

인간 대장암 세포(HCT-116) 1.3×10^8 cell 총 5개를 각각 4시간, 8시간, 18시간, 24시간, 48시간으로 나누어 lysis를 진행했다. 이후, 각 lysate 내의 총 단백질 농도를 측정하기 위해 BCA assay를 수행했다. 우선, BSA 용액을 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2 mg/mL의 농도로 희석하여 standard curve를 작성하였다. 각 시간 조건에 대해 측정된 흡광도를 계산하여 그린 standard curve는 모든 R^2 값이 0.99 이상으로 도출되었으며, 표 12와 그림 5에는 대표적으로 4시간 조건의 결과를 제시하였다. 다음으로, 각 시간 조건에서 얻어낸 lysate를 각 1/50, 1/20, 1/10, 1/5, 1/2로 희석해 흡광도를 측정하였고 이를 standard curve에 대입하여 평균값으로 단백질의 농도를 계산했다. 그 결과, 각 시간에 따른 총 단백질 농도는 다음과 같았다: 4시간 15.76mg/ml, 8시간 8.25mg/mL, 18시간 20.35mg/mL, 24시간 21.55mg/mL, 48시간 14.51mg/mL.

표 12. BSA standard 농도 및 흡광도 (4시간)

BSA 농도(mg/mL)	흡광도(562nm)
0.025	0.002
0.05	0.006
0.1	0.035
0.25	0.105
0.5	0.216
0.75	0.329
1	0.417
2	0.784

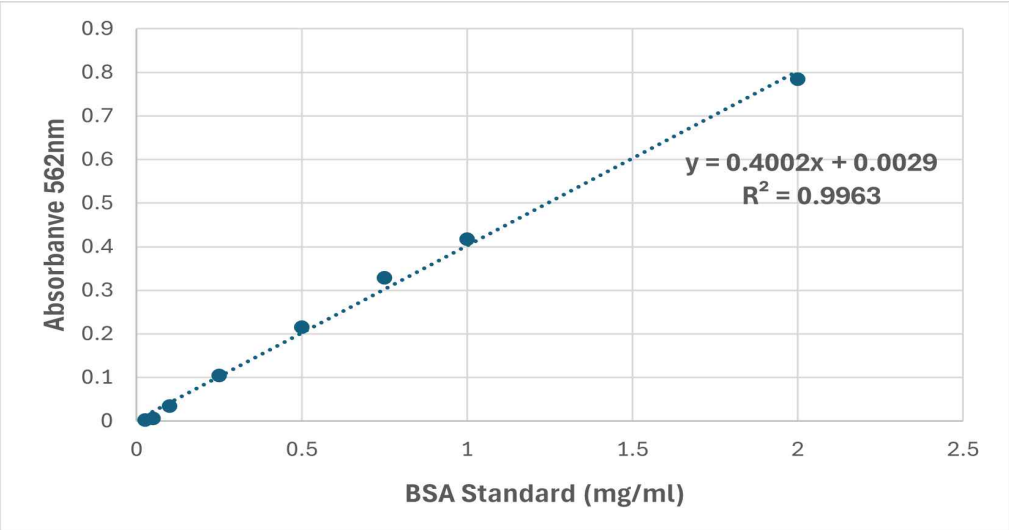


그림 5. BCA standard curve (4시간)

(2) 웨스턴 블롯

각 시간별로 측정된 단백질 농도를 바탕으로, 각 샘플에서 6 μg의 단백질을 취하여 웨스턴 블롯을 진행했다. 실험 결과. 모든 시간 조건(4시간, 8시간, 18시간, 24시간, 48시간)에서 HLA class I 단백질 밴드가 관찰되었고, 해당 밴드들은 37 kDa과 50 kDa 사이 위치에서 확인되었다(그림 6). 또한, 모든 Well에 동일한 단백질 양(6 μg)을 주입했음에도 불구하고, 4시간 lysate에서의 단백질 발현이 가장 높게 관찰되었으며, 시간이 지남에 따라 밴드의 두께가 점차 감소하는 경향을 보였다. 이는 HLA class I 단백질의 발현량이 lysis 시간 경과에 따라 저하됨을 의미한다. 따라서 후속 실험에서는 lysis 시간을 4시간으로 고정하여 실험을 진행했다.

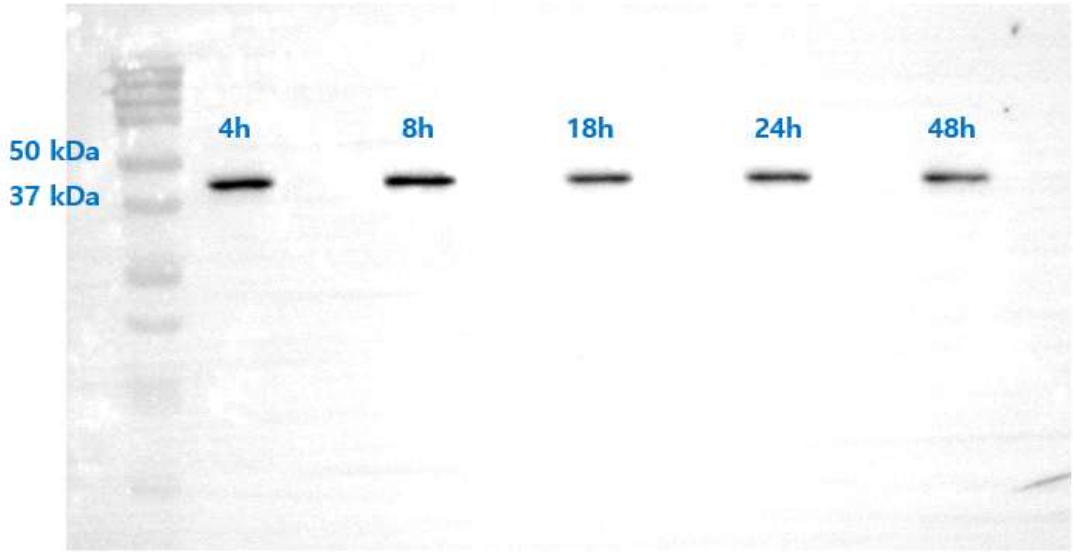


그림 6. Lysis 시간 변화에 따른 HLA class I 단백질 검출 결과

왼쪽부터 순서대로 4h, 8h, 18h, 24h, 48h lysis 조건의 lysate로부터 얻은 HLA class I 단백질의 western blot 결과를 나타내었다.

(3) HLA class I 펩타이드 LC-MS/MS 결과

HCT-116 세포로부터 분리한 HLA Class I 펩타이드를 Orbitrap으로 분석한 후(그림 7), PEAKS Studio 소프트웨어를 이용하여 De novo sequencing을 수행하였다. 그 결과, 총 4057개의 펩타이드 서열이 검출되었다. 이후, 이 펩타이드 서열들을 하나의 FASTA 파일로 변환하여 Proteome Discoverer 소프트웨어의 참조 데이터베이스로 사용하였다. 추가적으로, minimum peptide length 8, maximum peptide length 12의 조건을 설정하여 필터링을 수행하였으며, 그 결과 총 1445개의 펩타이드 서열 데이터가 도출되었다. 또한, 동정된 펩타이드를 길이별로 분류하여 시각화한 결과(그림 8), 10개의 아미노산으로 구성된 10mer 펩타이드가 387개로 8~12mer 중에 가장 높은 비율로 나타났다. 이와 유사하게 9mer 펩타이드가 높은 동정 비율을 보였고, 그 뒤를 이어 11mer, 12mer, 8mer 순으로 높은 비율을 차지했다.

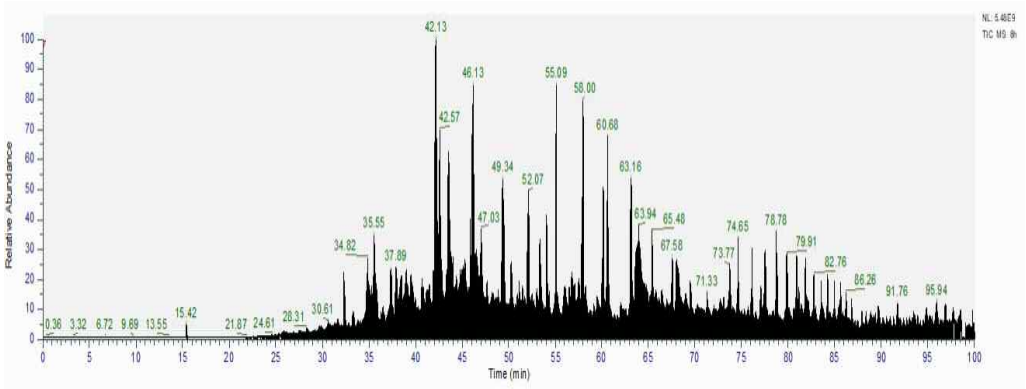
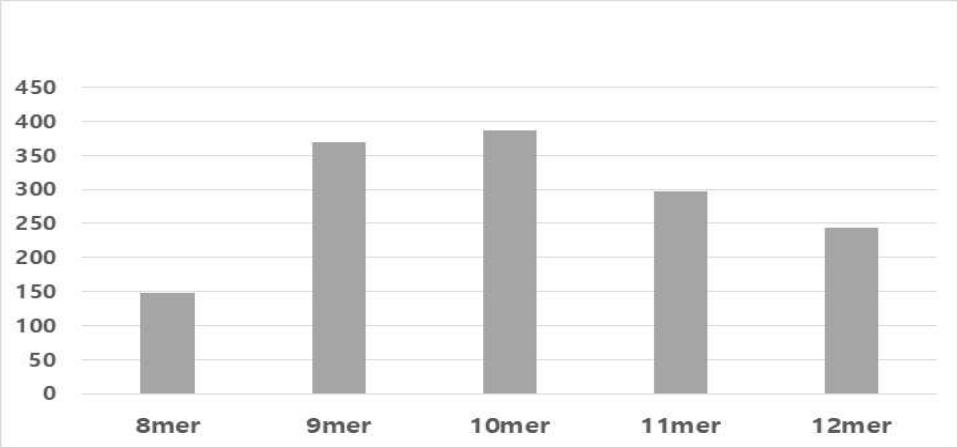


그림 7. HLA 펩타이드 질량 분석 Total ion chromatogram



	8mer	9mer	10mer	11mer	12mer
peptide	147	369	387	297	244

그림 8. HLA class I 펩타이드의 길이별 분류

가로축은 펩타이드 길이, 세로축은 동정된 서열 개수를 나타낸다.

1-2. 글라이칸 구조 동정 결과

HCT-116 세포에서 분리한 HLA class I 분자에 대해 LC-MS를 수행해 얻은 크로마토그램의 총 7개 peak에 대해 글라이칸 동정을 시도하였다 (그림 9). 먼저, 각 peak의 질량 스펙트럼을 분석하여 분자량을 계산하였고, 이를 그림 10에 제시된 형광 표지 기준 분자량과 비교함으로써 해당 peak에 대응하는 글라이칸 구조를 추정하였다.

전체적으로, 각 peak의 글라이칸 질량 스펙트럼은 동위 원소 스펙트럼과 0.5 m/z의 차이를 보였고, 이를 통해 +2 전하 상태임을 확인할 수 있었다. 이를 기반으로 단일 동위 원소 질량을 계산해 peak 1부터 peak 7까지의 글라이칸 구조를 동정할 수 있었다(그림 11, 표 13). 동정된 글라이칸 중에는 G0F, G1F, G2F가 높은 비율로 나타났으며 각각 G1F, G0F, G2F 순으로 분포하였다. 이외에도 G0F-GlcNAc, G1F-GlcNAc, G2F-GlcNAc와 같이 GlcNAc 잔기가 한 개 제거된 형태의 글라이칸도 함께 검출되었다.

이를 통해 총 7개의 peak에서 동정된 글라이칸들을 UPLC-FLR 분석을 통해 얻은 크로마토그램에 정리하였으며, 이는 그림 12에 제시했다. 또한, 각 peak별 글라이칸에 대한 용출 시간(retention time), 면적(area), 면적 백분율(percent area), 피크 높이(peak height) 등의 정보를 함께 표로 정리해 나타내었다.

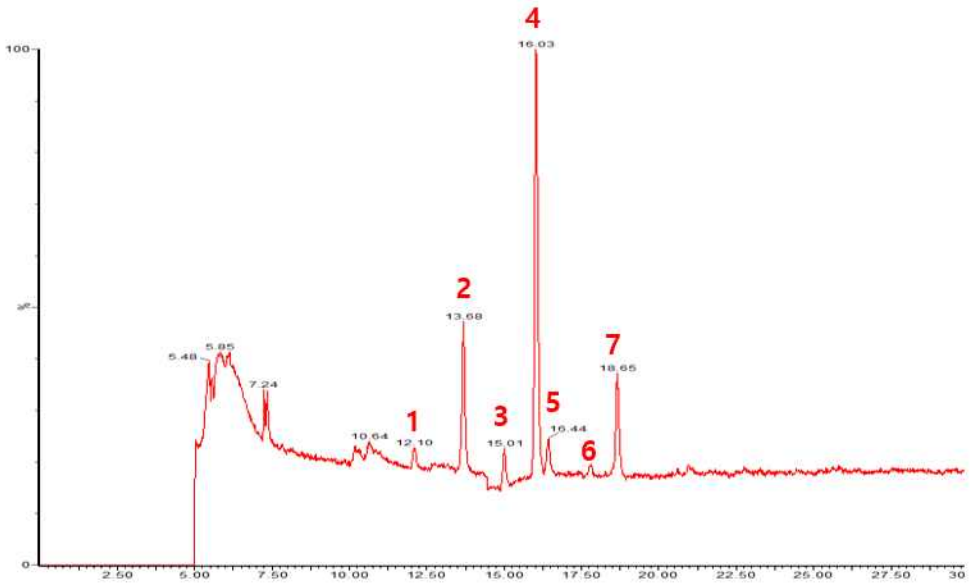
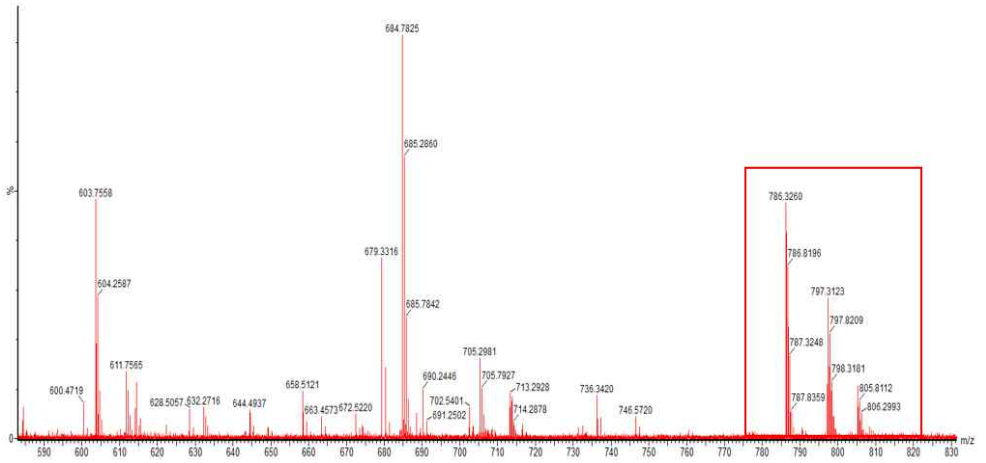


그림 9. HLA Class I 글라이칸 LC-MS 크로마토그램

	RT[1)]	Detected *m/z*	CS[2)]	Detected MW	Theoretical MW	ppm	Glycan
1	14.05	814.8401	2	1627.6645	1627.6611	2.09	G0
2	15.15	887.8705	2	1773.7253	1773.719	3.55	G0F
3	16.33	989.409	2	1976.8023	1976.7984	1.97	G0F+GN
4	16.72	895.8666	2	1789.7175	1789.714	1.96	G1
5	17.1	895.8667	2	1789.7177	1789.714	2.07	G1
6	17.71	968.8968	2	1935.7779	1935.7719	3.10	G1F
7	18.15	968.8964	2	1935.7771	1935.7719	2.69	G1F
8	18.59	1070.4361	2	2138.8565	2138.8512	2.48	G1F+GN
9	18.96	1070.4361	2	2138.8565	2138.8512	2.48	G1F+GN
10	19.64	976.8928	2	1951.7699	1951.7668	1.59	G2
11	20.55	1049.9224	2	2097.8291	2097.8247	2.10	G2F
12	21.03	1151.462	2	2300.9083	2300.9041	1.83	G2F+GN
13	21.7	1114.4439	2	2226.8721	2226.8673	2.16	G1F+NANA
14	22.97	1122.4412	2	2242.8667	2242.8622	2.01	G2+NANA
15	23.75	1195.4704	2	2388.9251	2388.9201	2.09	G2F+NANA
16	24.46	1297.0099	2	2592.0041	2591.9995	1.77	G2F+GN+NANA
17	26.55	1341.0186	2	2680.0215	2680.015	2.43	G2F+2NANA
18	26.92	1442.5595	2	2883.1033	2883.0949	2.91	G2F+GN+2NANA

그림 10. RapiFluor-MS labeled N-글라이칸 분자량

(a)



(b)

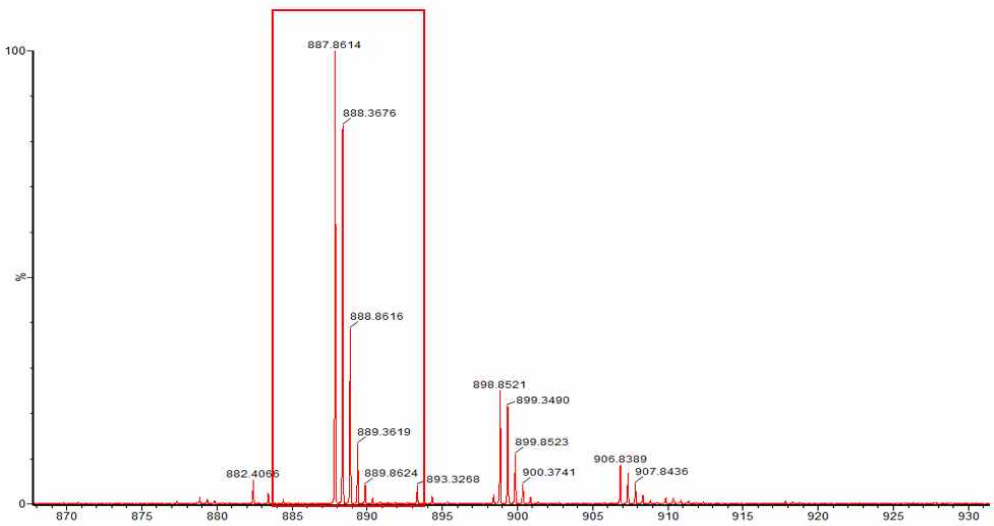
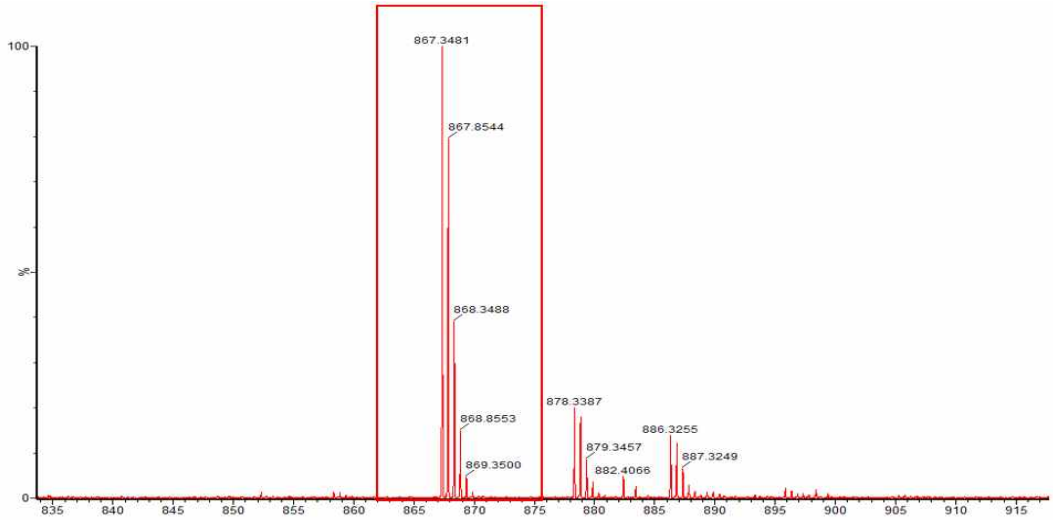


그림 11. 7개의 글라이칸 peak에 대한 질량 스펙트럼 확대 이미지

(a) Peak 1, G0F-GlcNAc, (b) Peak 2, G0F

(a)



(b)

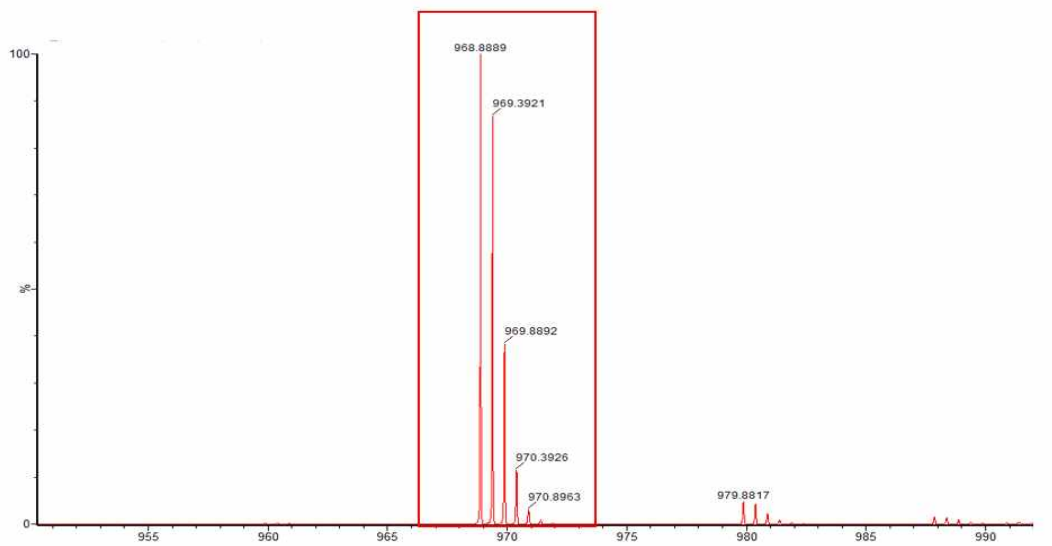
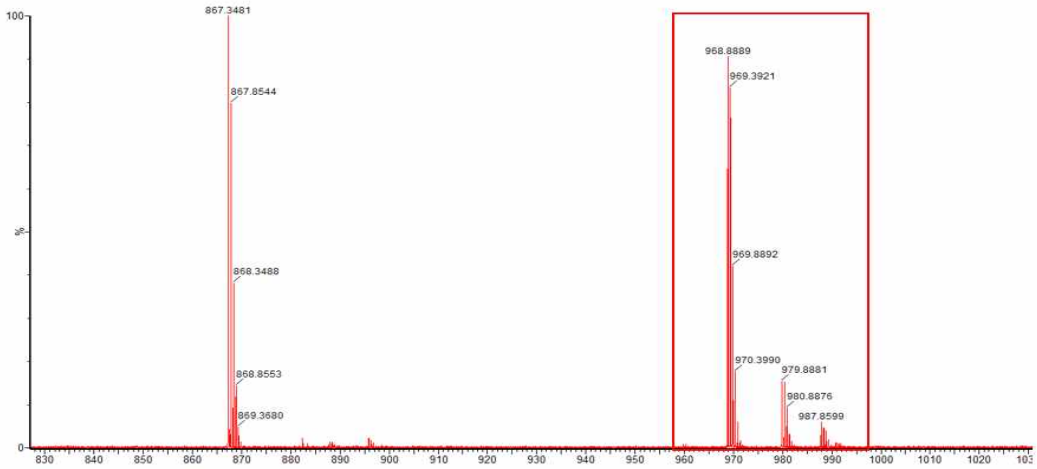


그림 11. Continued

(a) Peak 3, G1F-GlcNAc, (b) Peak 4, G1F

(a)



(b)

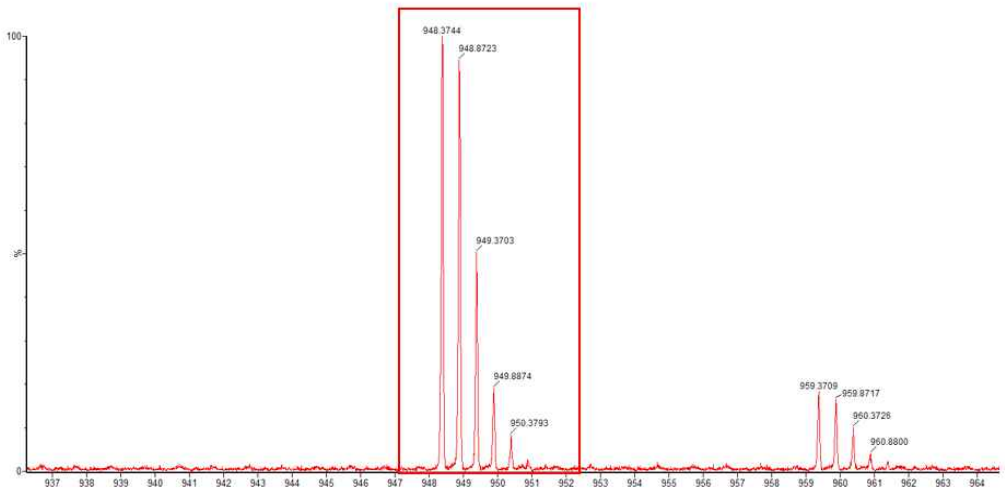


그림 11. Continued

(a) Peak 5, G1F, (b) Peak 6, G2F-GlcNAc

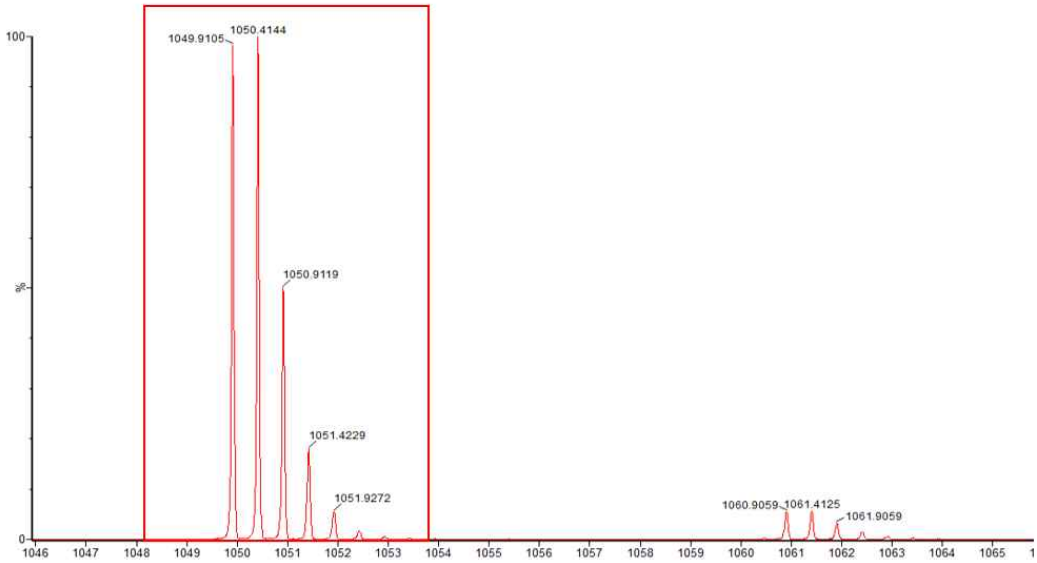
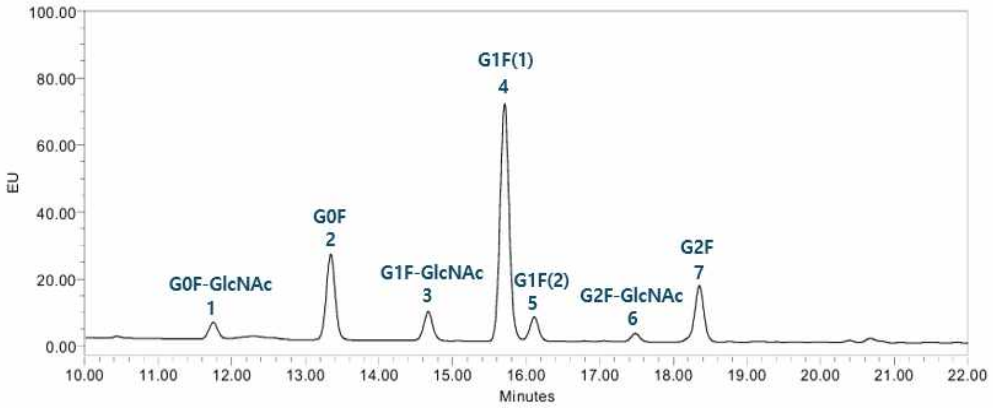


그림 11. Continued. Peak 7, G2F

표 13. 글라이칸 Peak별 질량값 및 동정 결과

	Mass(Dalton)	Glycan
Peak 1	1570.652	G0F-GlcNAc
Peak 2	1773.7228	G0F
Peak 3	1732.6962	G1F-GlcNAc
Peak 4	1935.7778	G1F
Peak 5	1935.7778	G1F
Peak 6	1894.7488	G2F-GlcNAc
Peak 7	2097.821	G2F



	Retention time	Area	% Area	Height
Peak 1	11.749	408847	3.48	49183
Peak 2	13.345	2156793	18.33	255865
Peak 3	14.669	735568	6.25	86580
Peak 4	15.708	6122629	52.05	709849
Peak 5	16.110	621783	5.29	72380
Peak 6	17.484	212041	1.80	25179
Peak 7	18.355	1505949	12.80	169517

그림 12. UPLC-FLR 및 LC-MS 기반 글라이칸 동정 결과

2. 치료용 항체 분석 결과

2-1. 글라이칸 패턴 분석 결과

(1) Deconvolution 분석

치료용 항체의 대략적인 N-글라이칸 패턴을 파악하기 위해 전처리하지 않은 항체 시료와 PNGase F를 처리해 N-글라이칸을 제거한 항체 시료를 동시에 질량 분석했다. 이후 결과 크로마토그램을 기반으로 분자량을 확인하는 deconvolution 분석을 수행하여 모든 시료의 분자량을 확인했다. 정리한 deconvolution 결과는 그림 13에, 각 시료의 분자량은 표 14에 나타냈다.

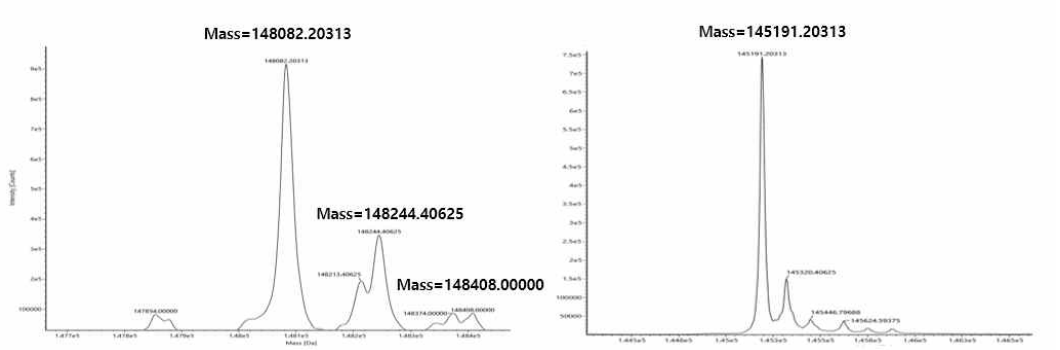
우선 PNGase F를 처리한 항체 시료들의 결과를 살펴보면, 대부분의 시료에서 3~4개의 deconvoluted peak가 관찰되었고, 이 중 가장 intensity가 높은 main peak 외에 128 Da 또는 162 Da의 질량 차이가 나는 peak들이 함께 나타났다. 이는 다른 단백질이 검출된 것이 아닌, 각 항체 중쇄의 C-말단에서 lysine 잔기가 제거되는 lysine clipping이 발생했거나, 제거되지 않은 O-글라이칸의 존재로 인한 것으로 생각할 수 있다. 따라서, 무처리 항체 시료와 PNGase F 처리 시료 간의 분자량 차이는 PNGase F 처리 시료에서 관찰된 가장 intensity가 높은 deconvoluted peak의 분자량 값을 기준으로 사용하여 계산했다.

또한, 무처리 항체 시료의 deconvolution 결과를 보면, PNGase F 처리군에 비해 더 많은 deconvolution peak가 나오는 것을 알 수 있다. 이는 항체의 N-글라이코실화 위치에 존재하는 다양한 글라이칸인 G0F, G1F, G2F 글라이칸의 조합으로 인한 것으로 보인다. 글라이칸 조합 확인은 무처리 항체 시료의 각 deconvolution peak와 PNGase F를 처리한 항체 시료의

deconvolution peak의 차이를 계산하는 방식으로 진행했다. 이후 이 차이는 표 15에 정리해놓은 각 조합별 글라이칸 질량 값과 비교하였다. 표에 정리한 각 글라이칸의 분자량은 N-글라이칸이 항체의 Asn 297에 부착될 때 발생하는 탈수 축합 반응(condensation reaction)에 의해 물 분자(H_2O, 18 Da)가 제거된다는 사실을 반영해 작성되었다. 단, G1F+G1F, G0F+G2F 조합은 동일한 분자량을 가지므로, 질량 정보만으로는 정확한 구조를 구분하기 어렵다. 따라서 결과 정리 시 두 가지 경우의 수를 모두 제시했다. 또한, Deconvolution 과정에서 미세한 질량 오차(± 10 Da)가 발생할 수 있으므로 이를 고려해 글라이칸을 동정하였다. 이러한 분석을 통해 성공적으로 각 항체 시료의 글라이칸 조합과 상대적 intensity를 확인할 수 있었다(표 16).

결과적으로, 본 실험에서는 G0F+G0F, G0F+G1F, G0F+G2F, G1F+G1F, G1F+G2F 총 5가지 조합의 존재 비율을 확인했다. 모든 항체 시료에서 G0F+G0F, G0F+G1F, G1F+G1F 또는 G0F+G2F 의 조합을 확인했으며, 추가적으로 Mabthera와 그 바이오시밀러인 Truxima, Herceptin과 그 바이오시밀러인 Herzuma에서만 G1F+G2F의 조합이 관찰되었다. 이는 해당 항체 시료들에서 상대적으로 G2F의 비율이 높아 나타난 결과인 것으로 해석된다.

(a)



(b)

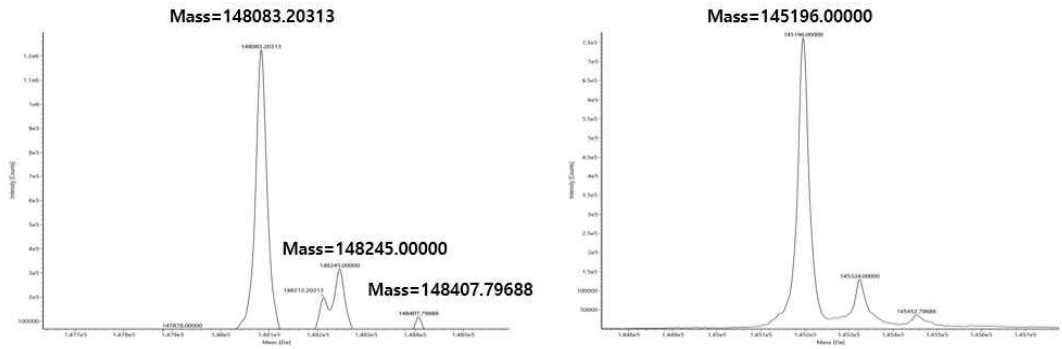
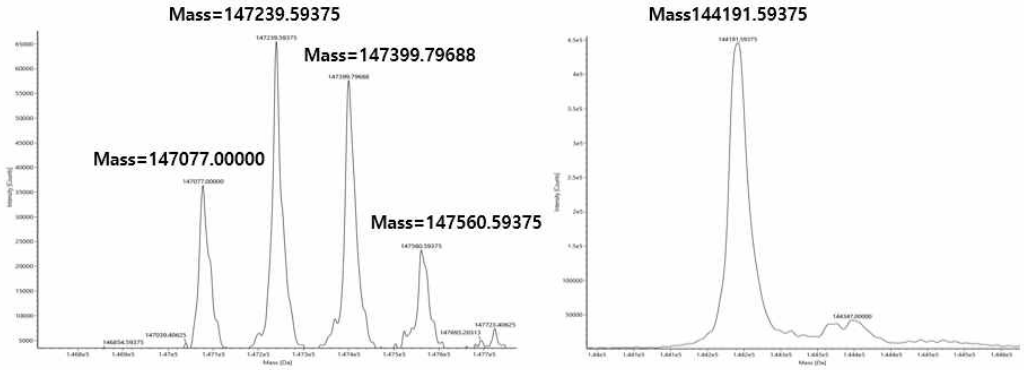


그림 13. 항체 분자량 deconvolutioin 결과

(a)Humira, (b)Yuflyma. 좌측은 전처리하지 않은 항체의 분자량, 우측은 PNGase F를 처리한 항체의 분자량을 나타낸다.

(a)



(b)

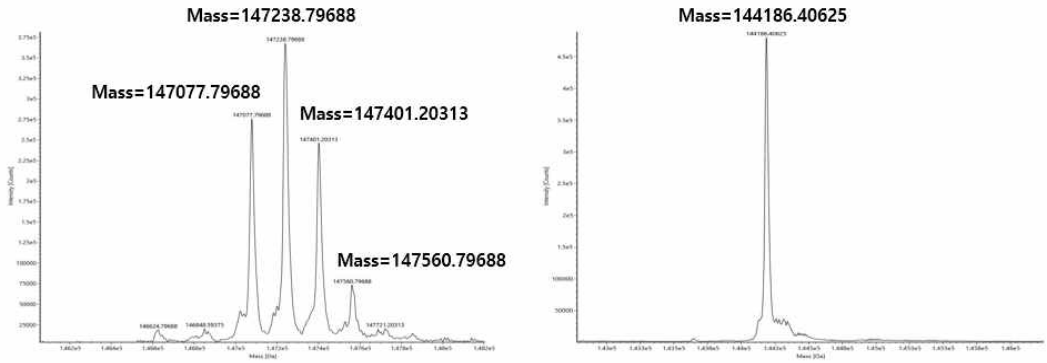
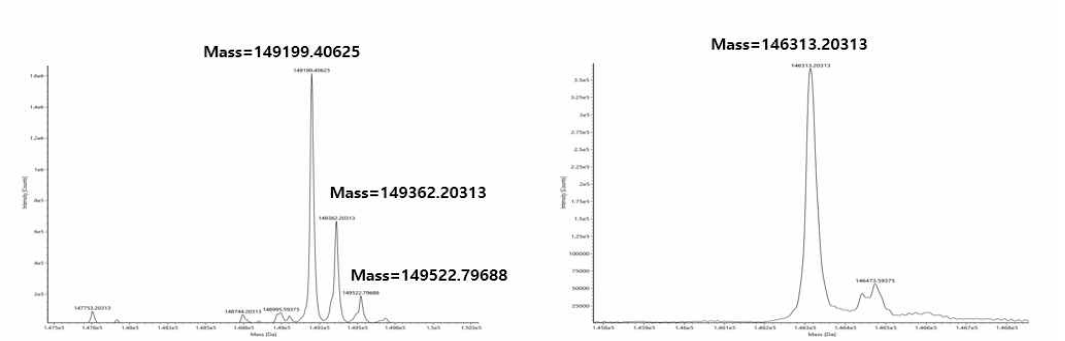


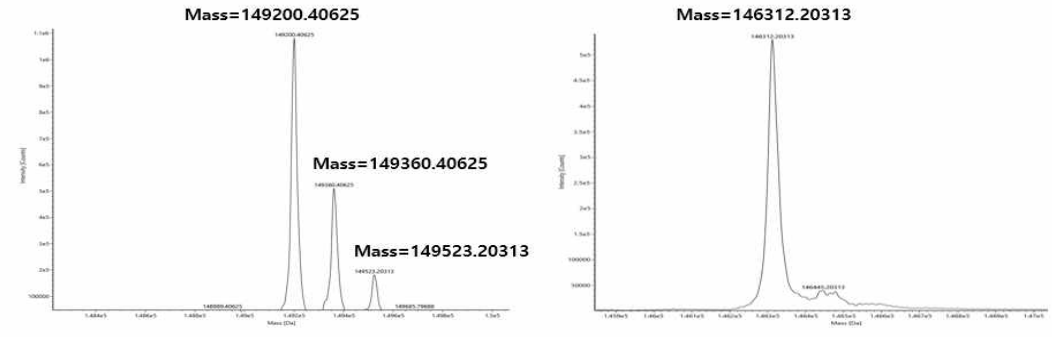
그림 13. Continued.

(a)Mabthera, (b)Truxima. 좌측은 전처리하지 않은 항체의 분자량, 우측은 PNGase F를 처리한 항체의 분자량을 나타낸다.

(a)



(b)



(c)

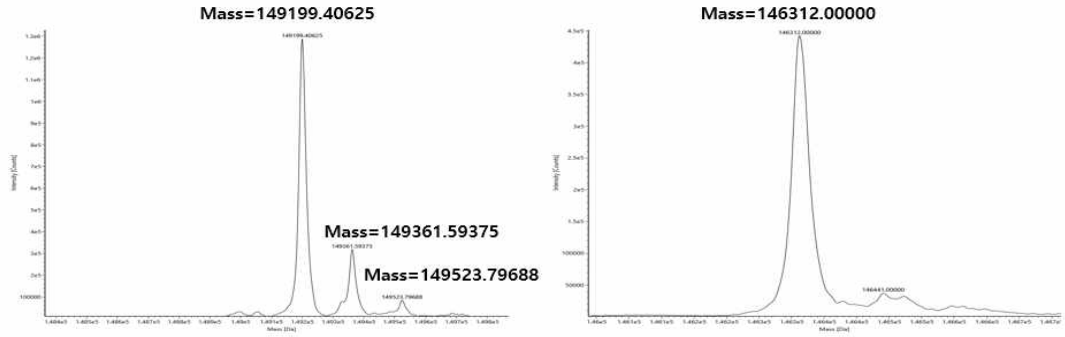
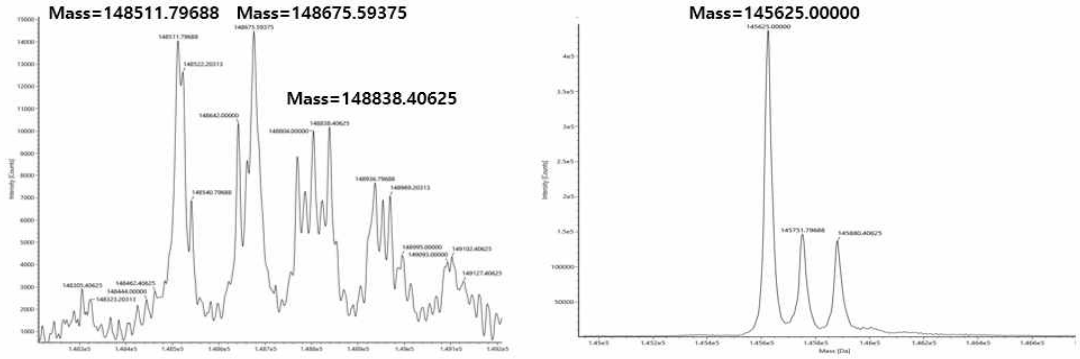


그림 13. Continued.

(a)Avastin, (b)Alymsys, (c)Vegzelma. 좌측은 전처리하지 않은 항체의 분자량, 우측은 PNGase F를 처리한 항체의 분자량을 나타낸다.

(a)



(b)

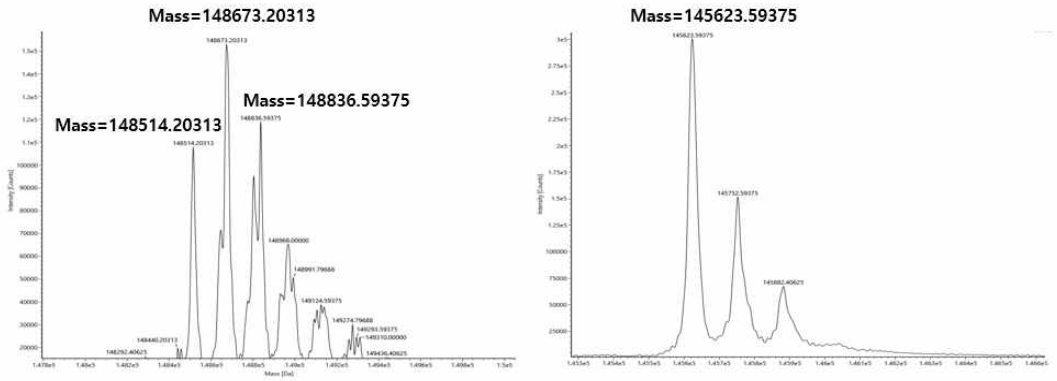
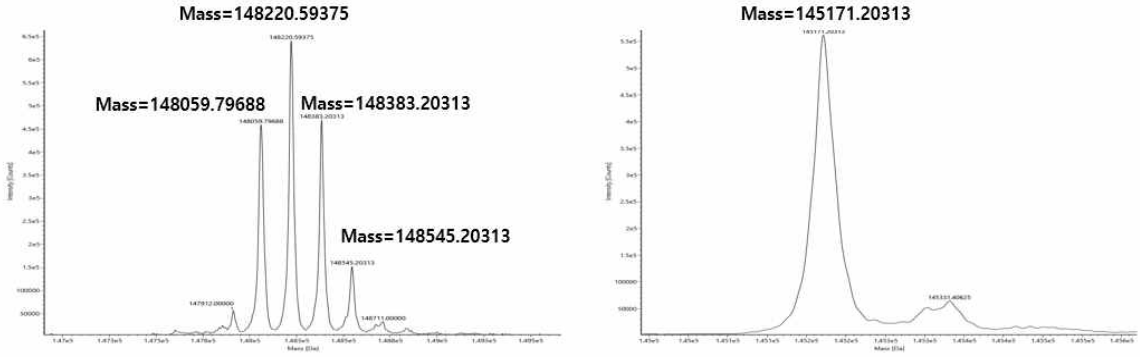


그림 13. Continued.

(a)Remicade, (b)Remsima. 좌측은 전처리하지 않은 항체의 분자량, 우측은 PNGase F를 처리한 항체의 분자량을 나타낸다.

(a)



(b)

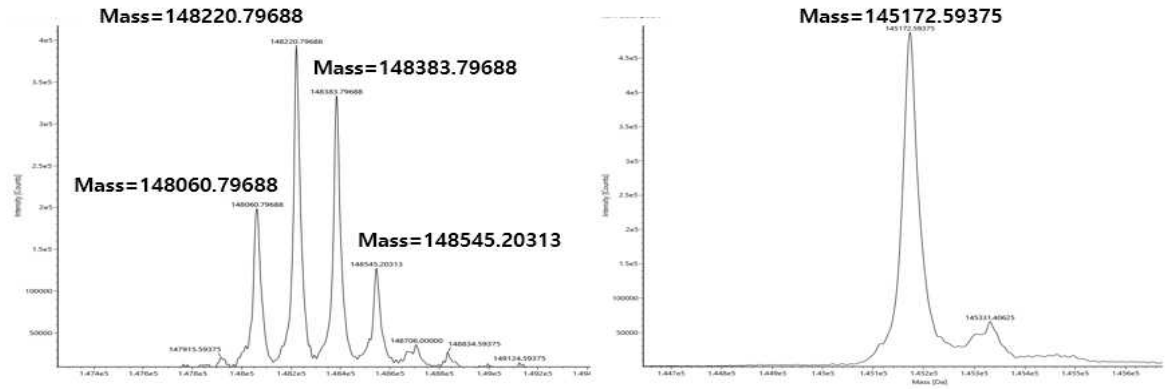


그림 13. Continued.

(a)Herceptin, (b)Herzuma. 좌측은 전처리하지 않은 항체의 분자량, 우측은 PNGase F를 처리한 항체의 분자량을 나타낸다.

표 14. PNGase F 처리 유무에 따른 항체 시료 분자량

		무처리 Molecular Weight (Da)	PNGase F 처리 Molecular Weight (Da)
Adalimumab	Humira	148082.20313	145191.20313
		148244.40625	
		148408.00000	
	Yuflyma	148083.20313	145196.00000
		148245.00000	
		148407.79688	
Rituximab	Mabthera	147077.00000	144191.59375
		147239.59375	
		147399.79688	
		147560.59375	
	Truxima	147077.79688	144186.40625
		147238.79688	
		147401.20313	
		147560.79688	
Bevacizumab	Avastin	149199.40625	146313.20313
		149362.20313	
		149522.79688	
	Alymsys	149200.40625	146312.20313
		149360.40625	
		149523.20313	
	Vegzelma	149199.40625	146312.00000
		149361.59375	
		149523.79688	

표 14. Continued

		무처리 Molecular Weight (Da)	PNGase F 처리 Molecular Weight (Da)
Infliximab	Remicade	148511.79688	145625.00000
		148675.59375	
		148838.40625	
	Remsima	148514.20313	145623.59375
		148673.20313	
		148836.59375	
Trastuzumab	Herceptin	148059.79688	145171.20313
		148220.59375	
		148383.20313	
		148545.20313	
	Herzuma	148060.79688	145172.59375
		148220.79688	
		148383.79688	
		148545.20313	

표 15. N-글라이칸 조합별 이론 분자량 계산

Glycan 조합	Molecular Weight (Da)
G0F + G0F	2889.088866
G0F + G1F	3051.14169
G0F + G2F	3213.194513
G1F + G1F	3213.194514
G1F + G2F	3375.247337
G2F + G2F	3537.30016

표 16. N-글라이칸 조합 및 intensity 측정 결과

		Δ Molecular Weight (Da)	Composition	Relative intensity(%)
Adalimumab	Humira	2891.000	G0F+G0F	67.76
		3053.203	G0F+G1F	25.94
		3216.797	G0F+G2F/G1F+G1F	6.30
	Yuflyma	2887.203	G0F+G0F	73.80
		3049	G0F+G1F	19.16
		3211.797	G0F+G2F/G1F+G1F	7.04
Rituximab	Mabthera	2885.406	G0F+G0F	19.81
		3048	G0F+G1F	35.79
		3208.203	G0F+G2F/G1F+G1F	31.73
		3369	G1F+G2F	12.67
	Truxima	2891.391	G0F+G0F	28.64
		3052.391	G0F+G1F	38.23
		3214.797	G0F+G2F/G1F+G1F	25.48
		3374.391	G1F+G2F	7.65
Bevacizumab	Avastin	2886.203	G0F+G0F	65.27
		3049	G0F+G1F	27.03
		3209.594	G0F+G2F/ G1F+G1F	7.70
	Alymsys	2888.203	G0F+G0F	60.90
		3048.203	G0F+G1F	28.89
		3211	G0F+G2F/G1F+G1F	10.21
	Vegzelma	2887.406	G0F+G0F	76.04
		3049.594	G0F+G1F	18.80
		3211.797	G0F+G2F/G1F+G1F	5.16

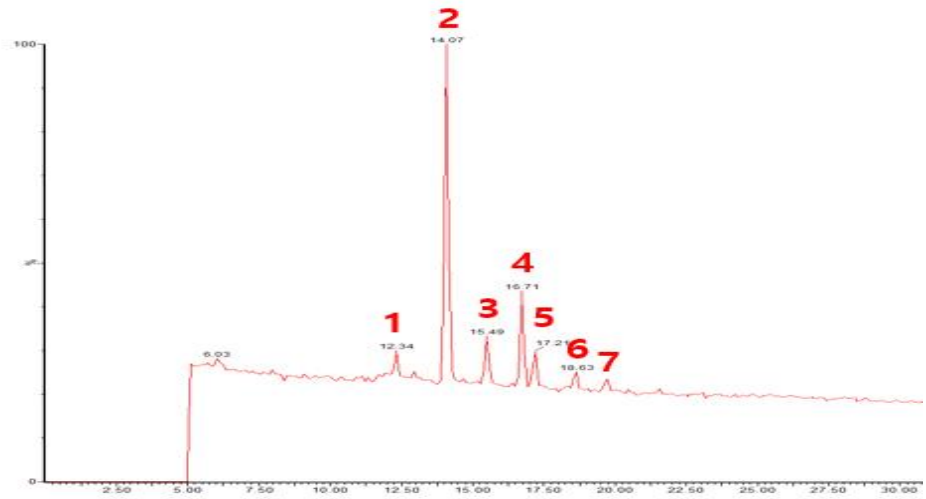
표 16. Continued.

		Δ Molecular Weight (Da)	Glycan Composition	Relative intensity(%)
Infliximab	Remicade	2886.797	G0F+G0F	37.10
		3050.594	G0F+G1F	38.79
		3213.406	G0F+G2F/G1F+G1F	24.11
	Remsima	2890.609	G0F+G0F	28.65
		3049.609	G0F+G1F	40.47
		3213	G0F+G2F/G1F+G1F	30.88
Trastuzumab	Herceptin	2888.594	G0F+G0F	26.58
		3049.391	G0F+G1F	37.48
		3212	G0F+G2F/G1F+G1F	27.23
		3374	G1F+G2F	8.71
	Herzuma	2888.203	G0F+G0F	19.03
		3048.203	G0F+G1F	37.48
		3211.203	G0F+G2F/G1F+G1F	31.43
		3372.609	G1F+G2F	12.06

(2) UPLC-FLR 및 질량 분석 기반 글라이칸 동정

각 항체 시료의 N-글라이칸을 분리 후 질량 분석을 수행한 결과, 그림 14와 그림 15와 같은 크로마토그램 및 질량 스펙트럼이 확인되었다. 그림 15의 질량 스펙트럼은 대표적으로 Humira 항체 시료의 결과를 나타낸 것이다. 동정된 글라이칸들은 +2의 전하 상태를 띠고 있었으며 이를 바탕으로 계산한 질량값은 그림 10과 비교하였다. 전체적으로, 모든 항체 시료에서 주요 글라이칸인 G0F, G1F, G2F가 공통적으로 검출되었으며, 각각 G0F, G1F, G2F 순으로 높은 비율을 보였다. 이외에 일부 시료에서는 G0, Mannose 5, Mannose 6, G0F-GlcNAc, G0-GlcNAc 등의 글라이칸들도 함께 확인되었다. 이 중 고만노스(high-mannose) 계열인 Mannose 5, Mannose 6 글라이칸은 항체의 반감기를 줄여 제거율을 상승시키는 특성이 있어 치료용 항체에서는 바람직하지 않은 글라이칸으로 분류된다[12]. 그러나 본 분석에서는 해당 글라이칸이 매우 낮은 비율로 검출되어, 품질 기준 내 수용 가능한 범위로 판단하였다. 이후 UPLC-FLR 분석을 통해 얻은 결과 크로마토그램에 각 peak별로 동정된 글라이칸을 표시하였다(그림 16). 주요 글라이칸들의 백분율 총합은 표 17에 제시하였다. 모든 항체 시료들은 주요 글라이칸 총합 비율이 80% 이상인 수치를 보였으며 이는 품질적으로 양호한 수준임을 의미한다. 또한 PNGase F 처리 유무를 통한 글라이칸 조합 확인 실험 결과에서 나타났듯이, Mabthera, Truxima, Herceptin, Herzuma는 다른 항체 시료들에 비해 높은 G2F 비율을 보여주었다.

(a)



(b)

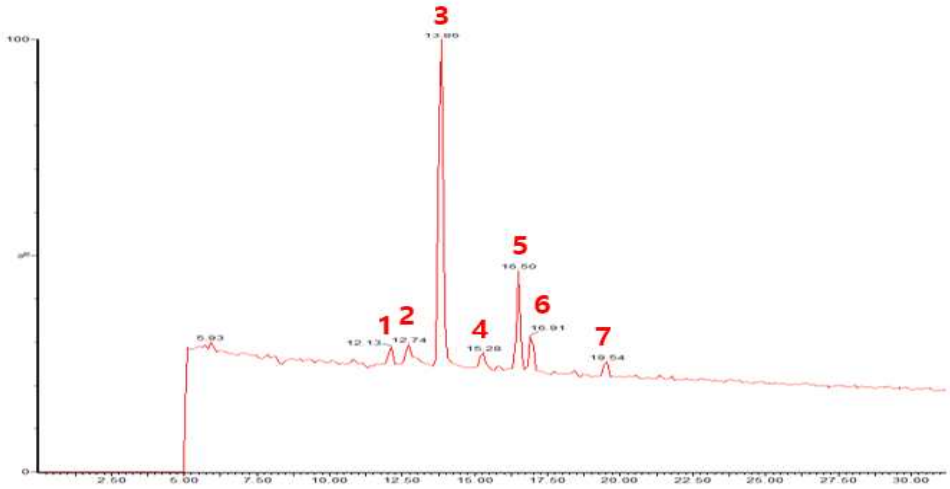
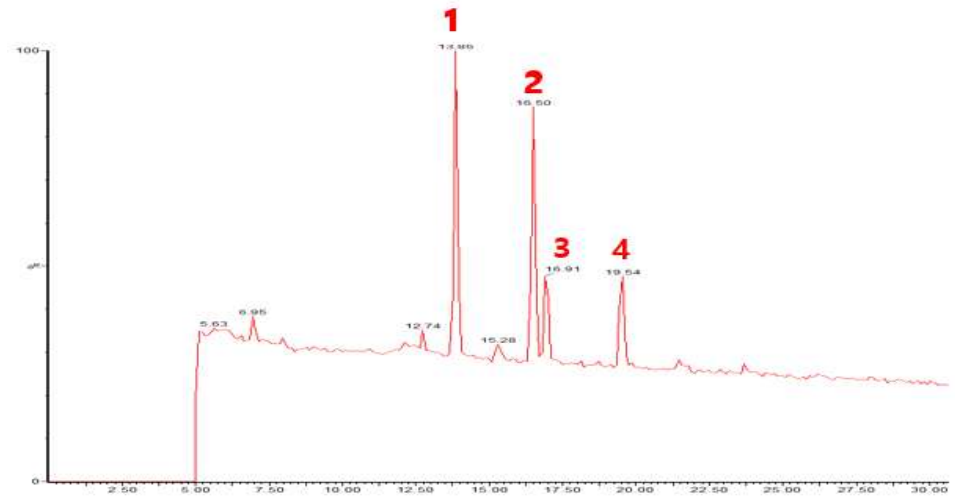


그림 14. 항체별 N-글라이칸 질량 분석 크로마토그램

(a) Humira, (b) Yuflyma

(a)



(b)

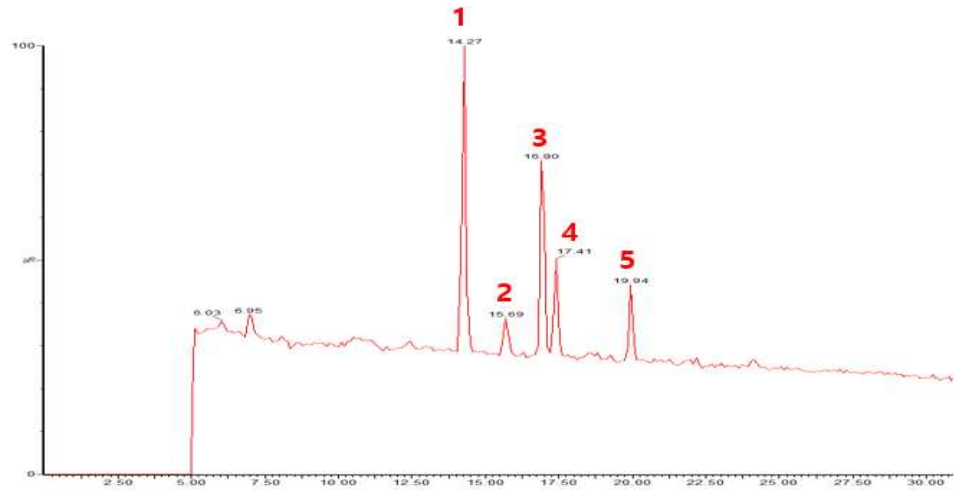
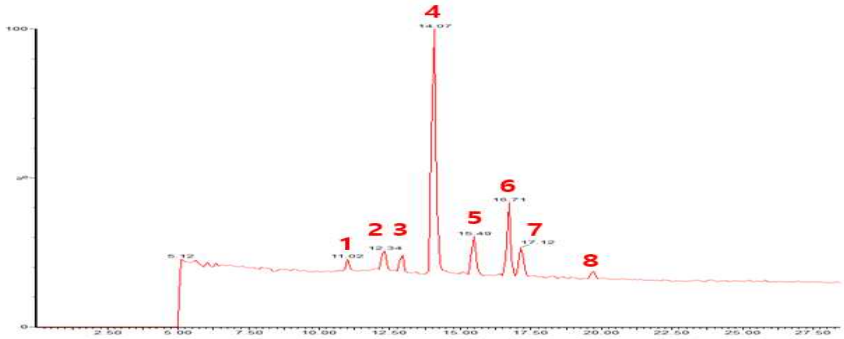


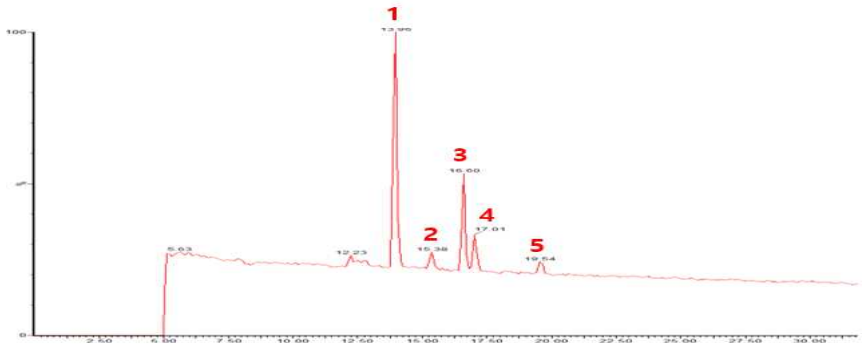
그림 14. Continued

(a) Mabthera, (b) Truxima

(a)



(b)



(c)

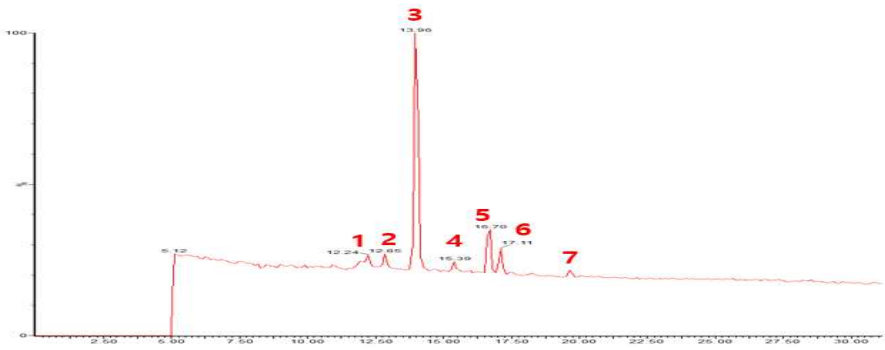
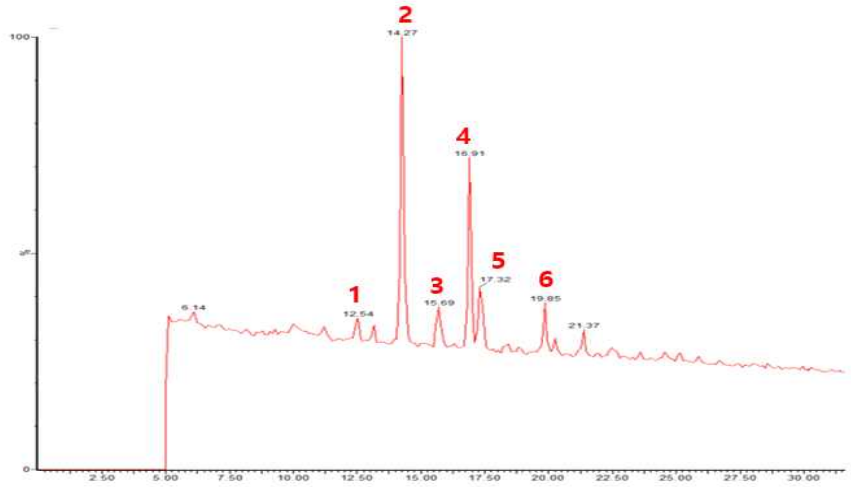


그림 14. Continued

(a) Avastin, (b) Alymsys, (c) Vegzelma

(a)



(b)

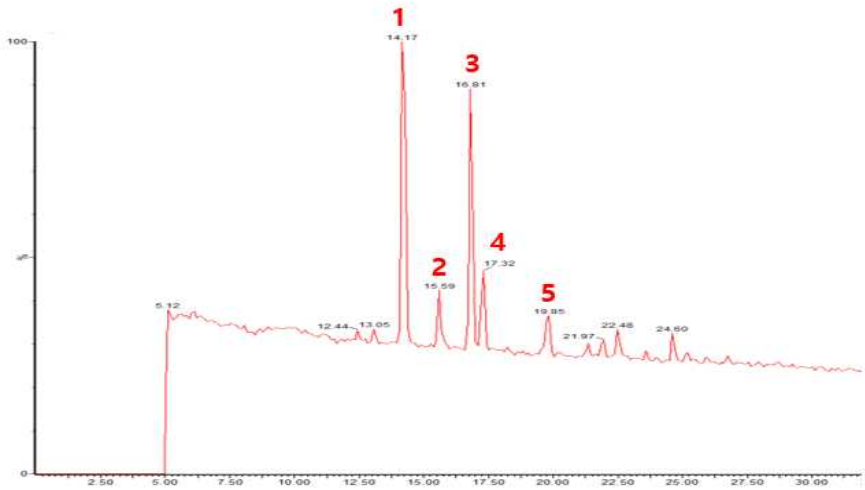
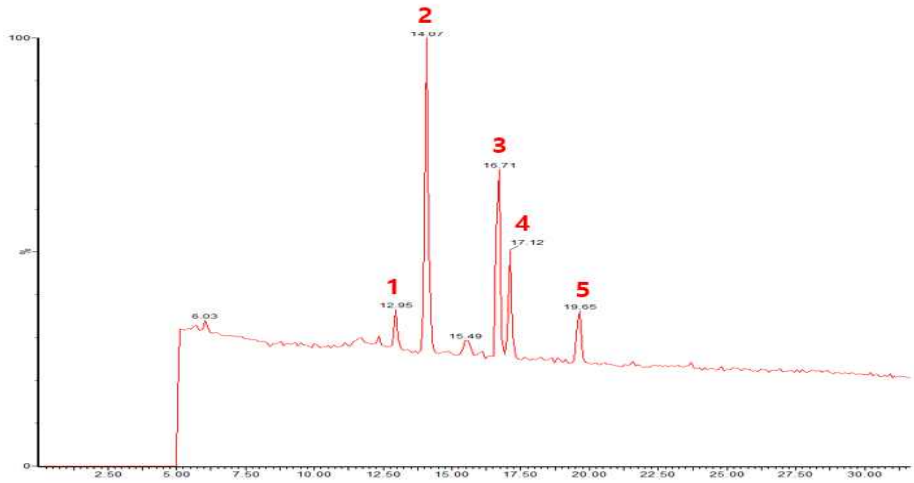


그림 14. Continued

(a) Remicade, (b) Remsima

(a)



(b)

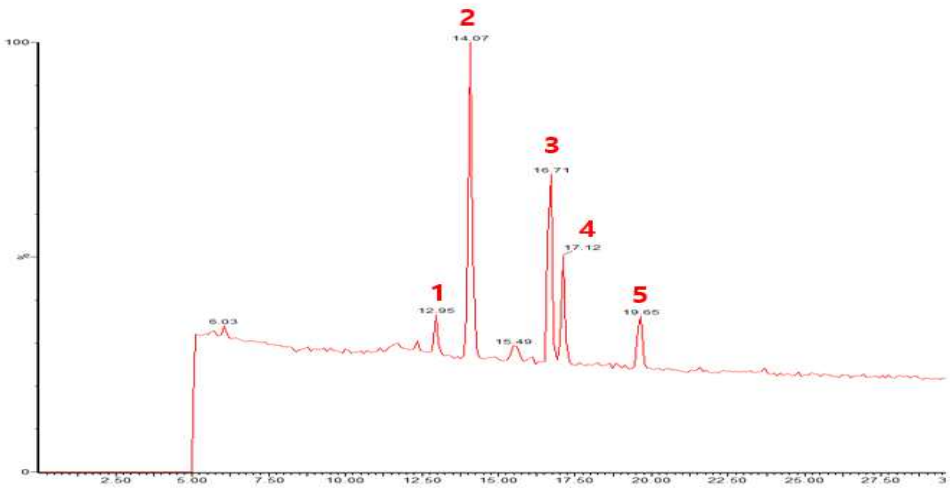
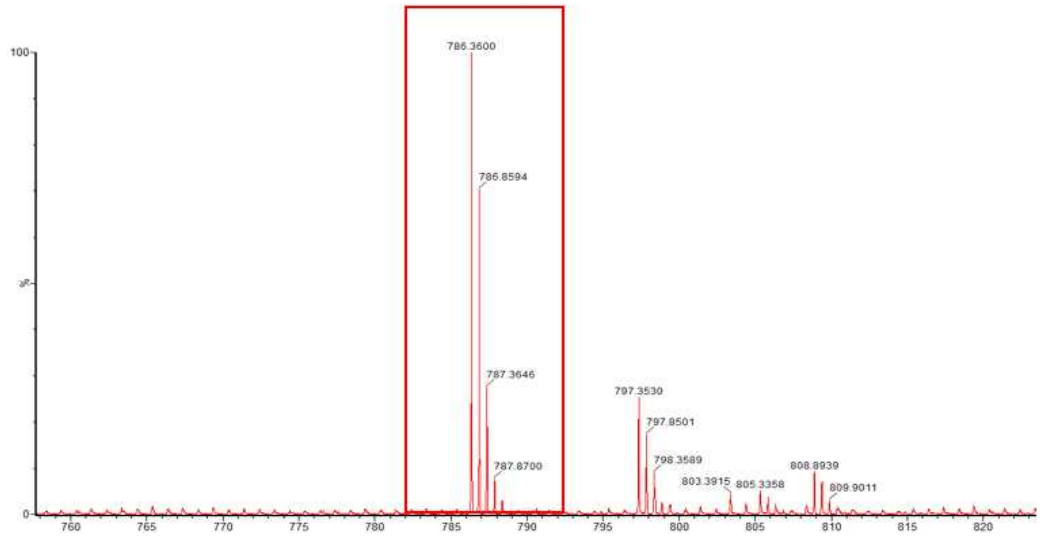


그림 14. Continued

(a) Herceptin, (b) Herzuma

(a)



(b)

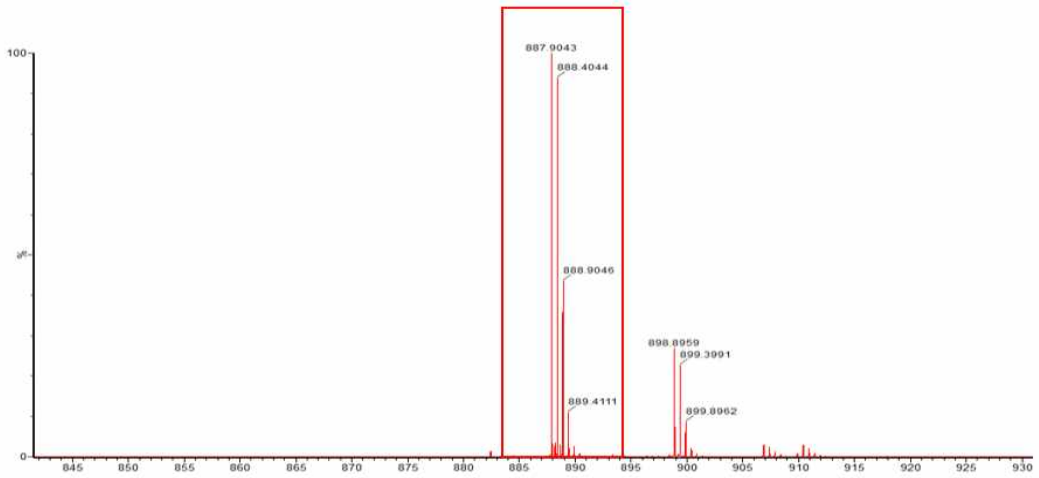
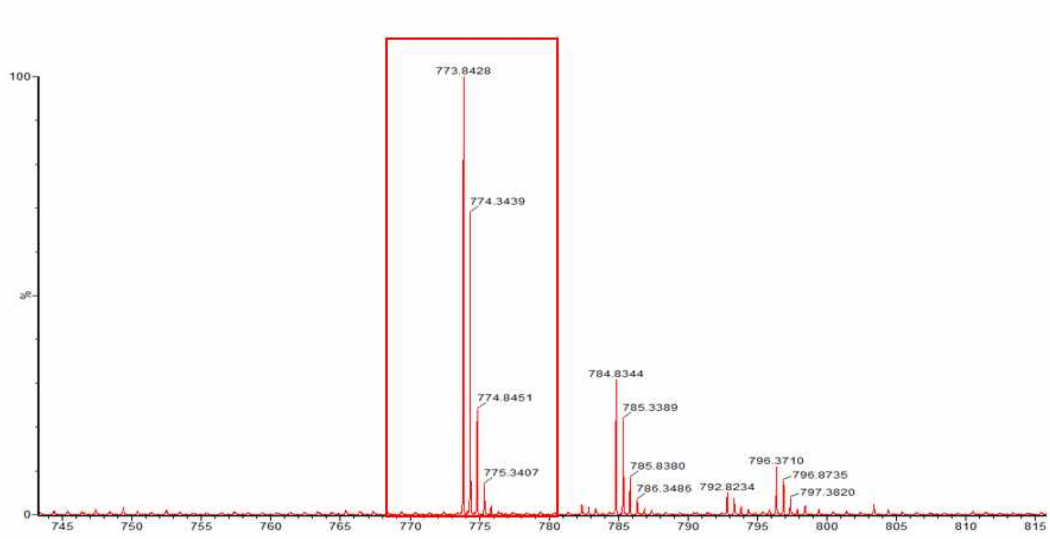


그림 15. Humira N-글라이칸 질량 스펙트럼

(a) G0F-GlcNAc, (b) G0F

(a)



(b)

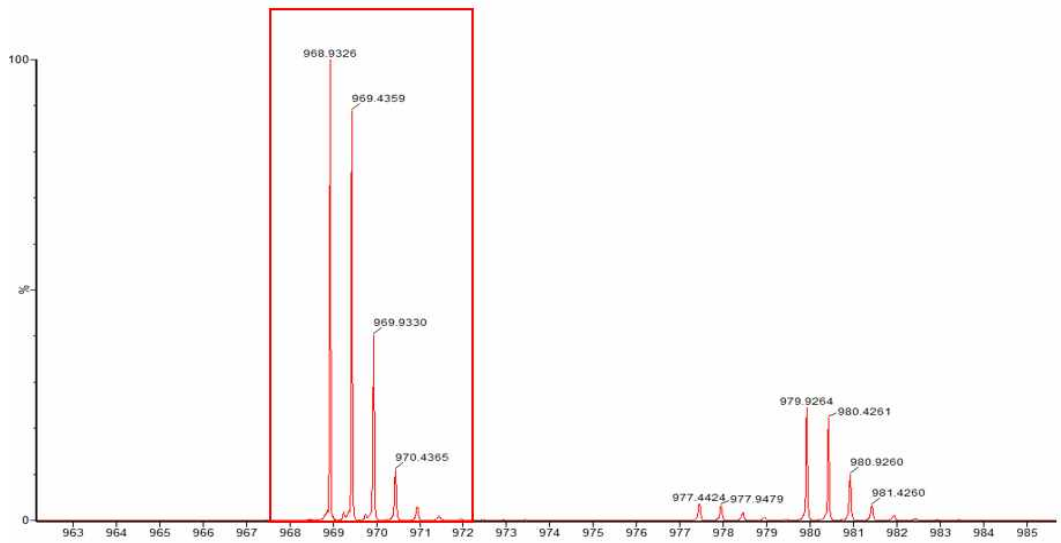
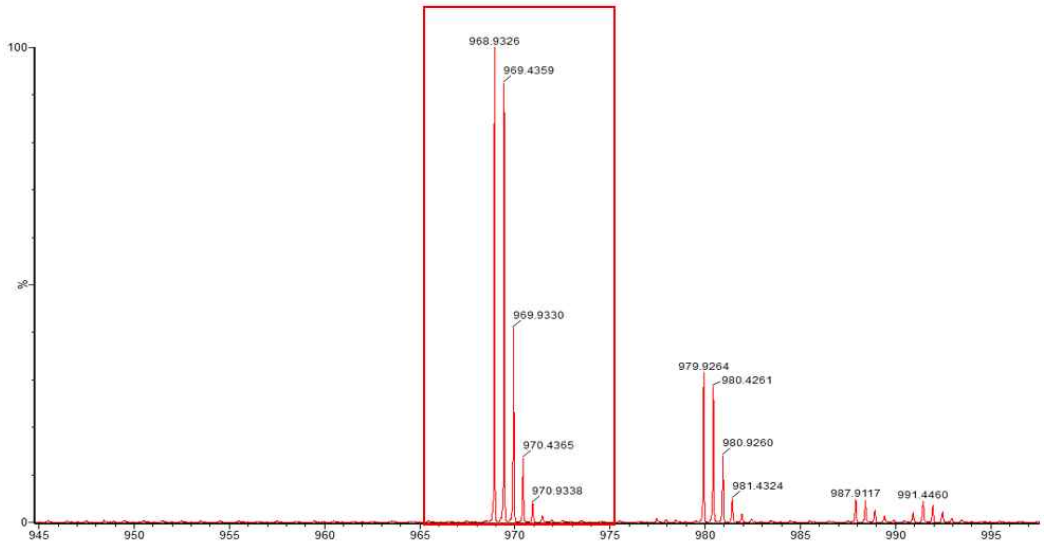


그림 15. Continued

(a) Man 5, (b) G1F(1)

(a)



(b)

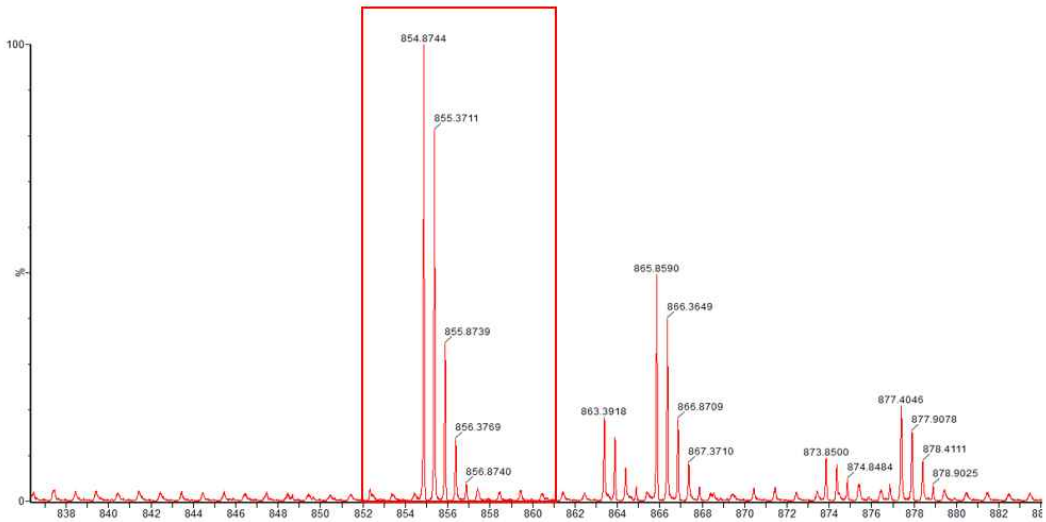


그림 15. Continued

(a) G1F(2), (b) Man 6

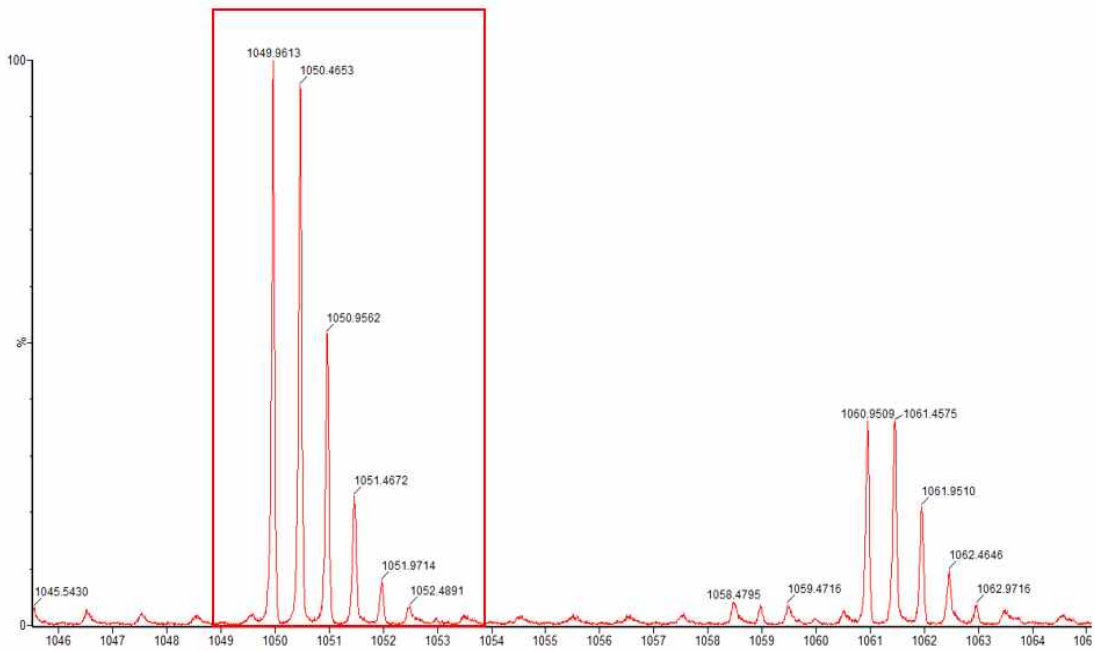
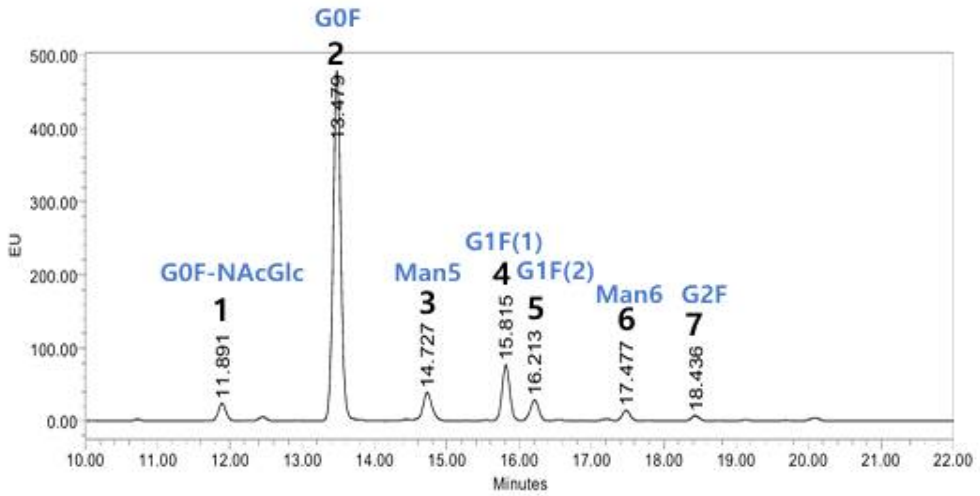


그림 15. Continued. G2F

(a)



(b)

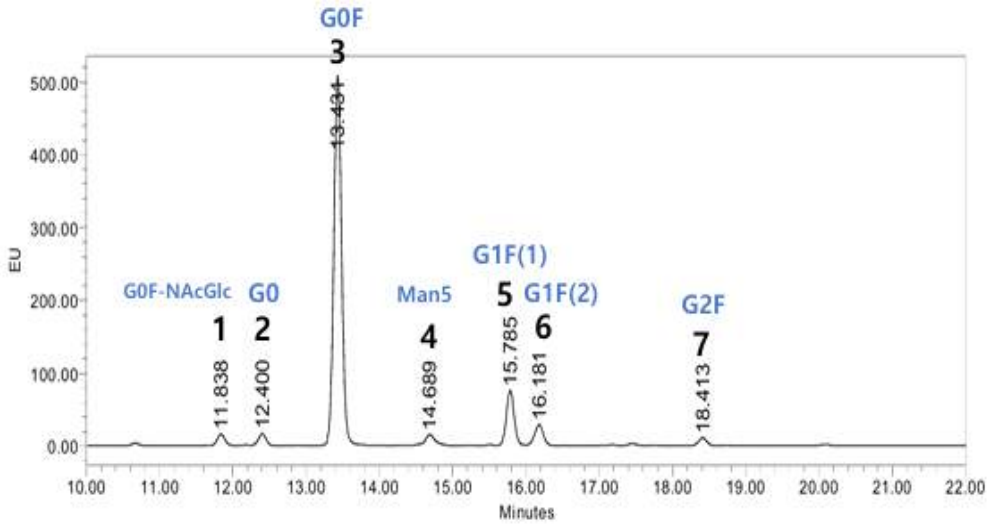
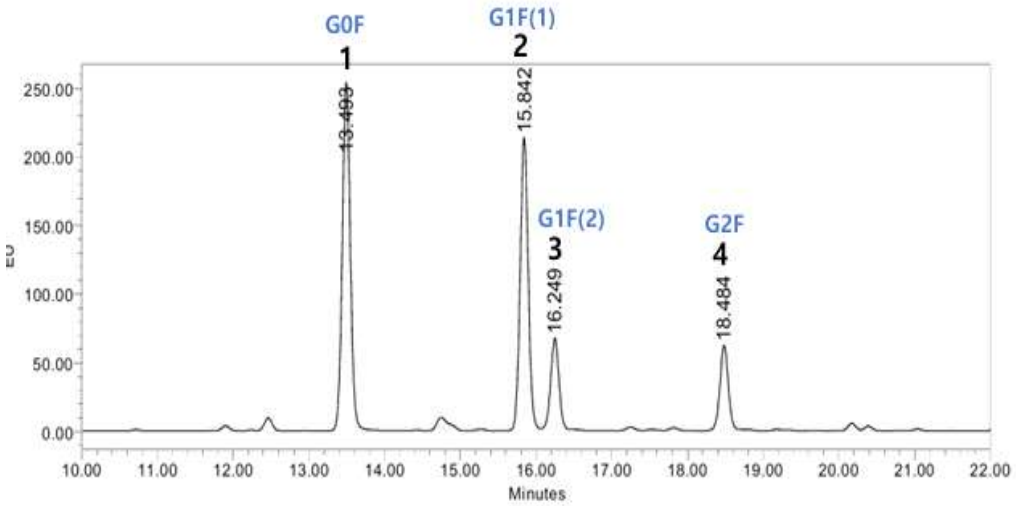


그림 16. 항체별 N-글라이칸 형광 분석 결과

(a) Humira, (b) Yuflyma

(a)



(b)

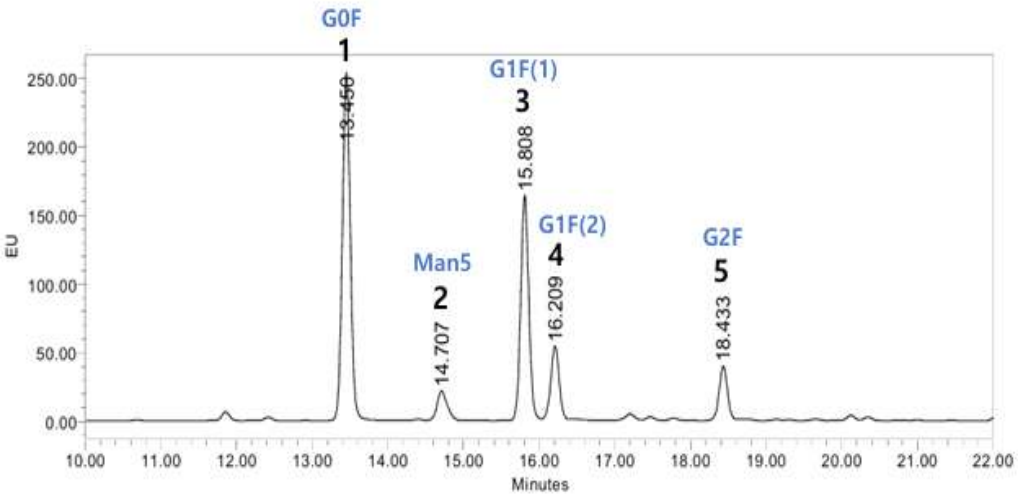
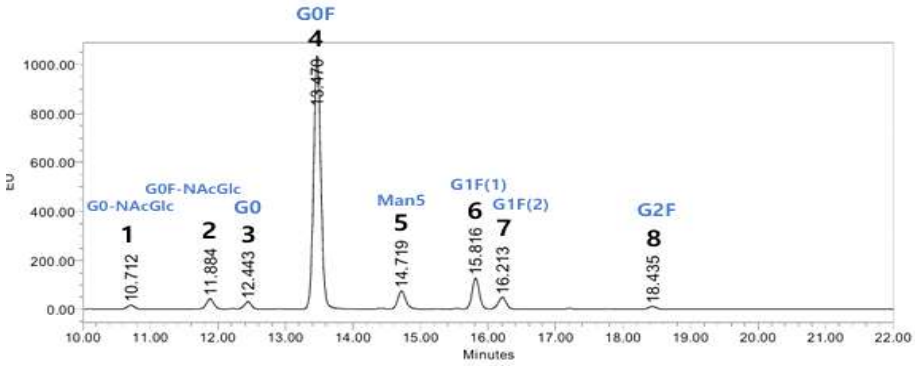


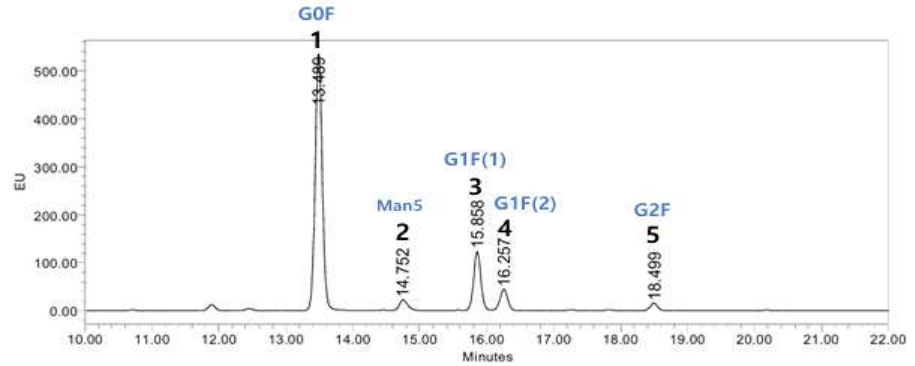
그림 16. Continued

(a) Mabthera, (b) Truxima

(a)



(b)



(c)

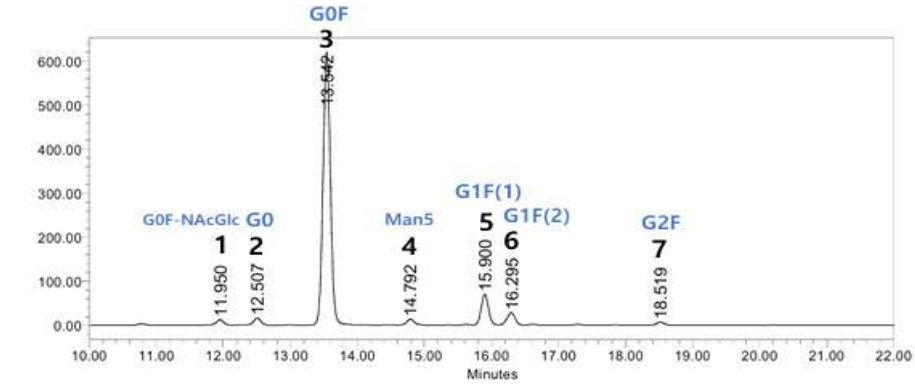
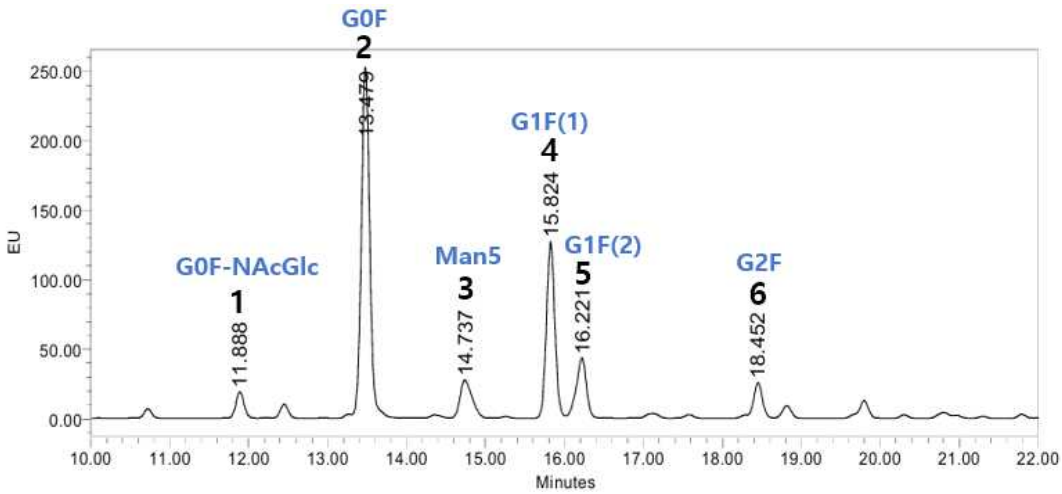


그림 16. Continued

(a) Avastin, (b) Alymsys, (c) Vegzelma

(a)



(b)

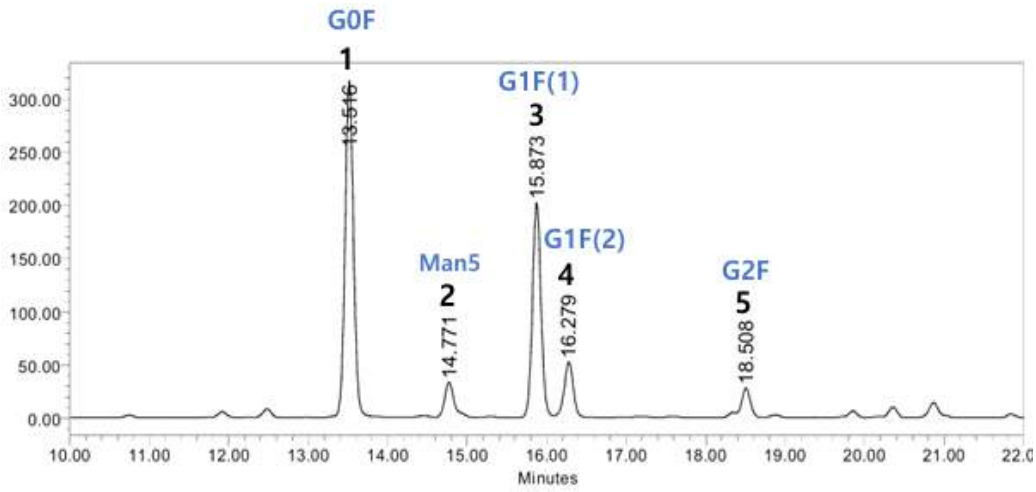
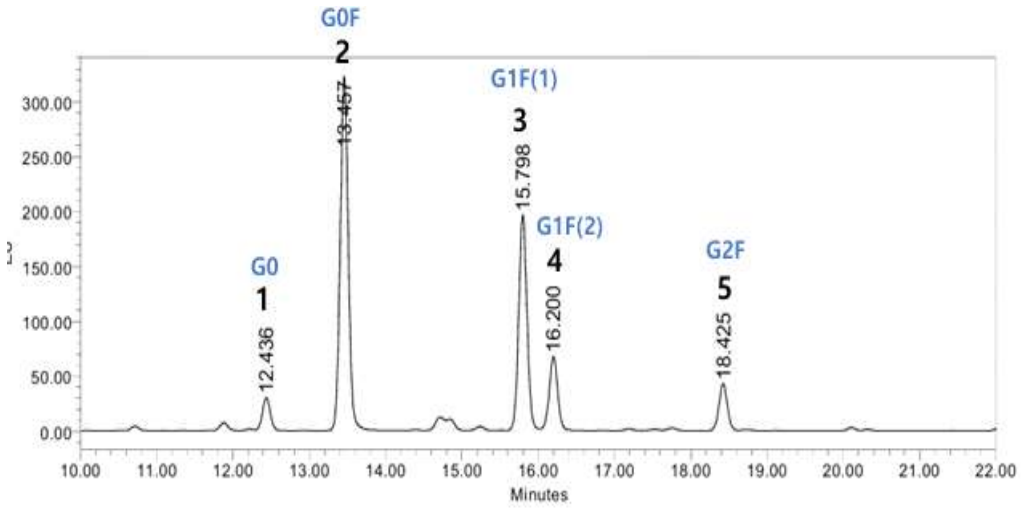


그림 16. Continued

(a) Remicade, (b) Remsima

(a)



(b)

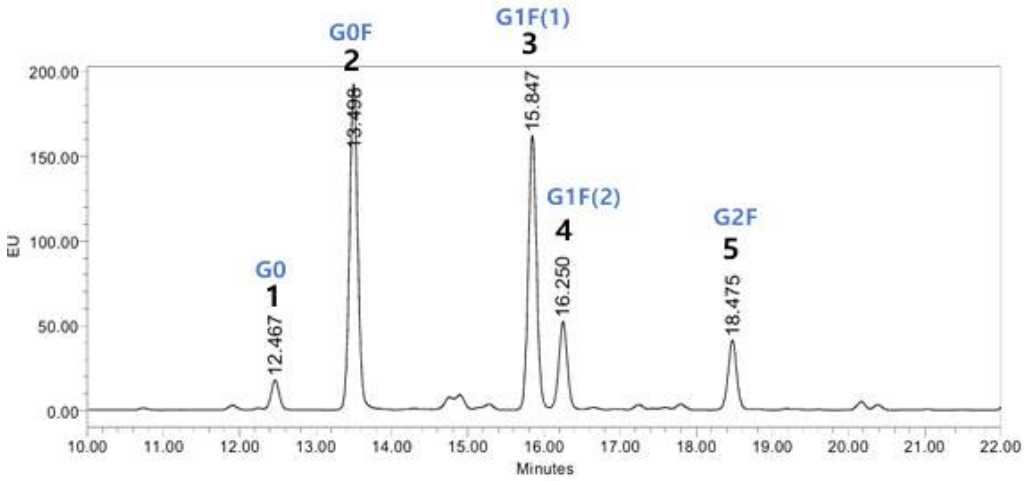


그림 16. Continued

(a) Herceptin, (b) Herzuma

표 17. 항체별 주요 글라이칸 구성 백분율(%)

		G0F	G1F	G1F′	G2F	Major 비율 (%)
Adalimumab	Humira	70.413	11.340	4.680	1.173	87.607
	Yuflyma	74.630	11.320	4.843	1.740	92.533
Rituximab	Mabthera	41.943	35.767	11.460	10.827	99.997
	Truxima	46.457	30.607	10.277	7.593	94.933
Bevacizumab	Avastin	73.867	9.043	3.800	0.913	87.623
	Alymsys	70.833	16.557	6.360	2.227	95.977
	Vegzelma	80.493	9.073	4.097	1.003	94.667
Infliximab	Remicade	47.207	23.927	9.730	5.493	86.357
	Remsima	48.487	31.800	8.707	5.213	94.207
Trastuzumab	Herceptin	48.127	30.240	10.653	6.720	95.740
	Herzuma	40.867	34.913	11.497	9.167	96.443

3. 어류 저분자 콜라겐 펩타이드 분석 결과

3-1. LC-MS/MS 결과

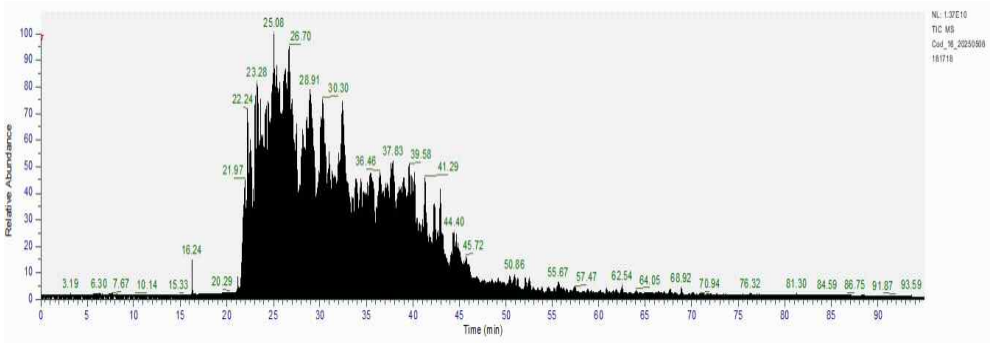
(1) 펩타이드 서열 간 상동성 및 차이점 분석

어종 간 콜라겐 펩타이드 서열 동정을 위해 Orbitrap 장비를 이용하여 LC-MS/MS 분석을 수행했다. 질량 분석 결과 크로마토그램은 그림17에 나타내었다. 이후, Peaks Studio를 활용해 De novo sequencing을 진행했고, 방대한 데이터를 효율적으로 정리하기 위해 ALC(Average Local Confidence)를 고려하여 FASTA 파일로 변환했다. ALC는 De novo sequencing으로 예측한 아미노산 서열 각각에 대해 계산된 local confidence score의 평균값을 의미한다. 즉, local confidence score가 클수록 각 아미노산이 그 위치에 있을 확률이 높아져 서열 예측의 정확성이 증가하기 때문에 이는 신뢰성을 평가하는 주요 지표로 활용할 수 있다. 이 기준에 따라 ALC 수치가 75% 이상인 서열들을 분류했으며, 각 시료별로 식별된 펩타이드 서열 수는 다음과 같다. 대구의 경우, 대구＃16은 296개, 대구＃28은 325개 대구＃153은 329개의 펩타이드 서열이 식별되었다. 명태의 경우, 명태＃116는 1117개, 명태＃156은 874개, 명태＃173은 894개의 펩타이드 서열이 식별되었다. 틸라피아는 틸라피아＃23에서 537개, 틸라피아＃157에서 439개, 틸라피아＃176에서 278개의 서열이 식별되었다(표 18). 전체적으로 명태 유래 콜라겐 펩타이드에서 동정된 펩타이드 서열 수가 가장 많았으며, 이는 명태 시료에 다양한 서열 펩타이드 서열이 존재하고 높은 신뢰도(ALC)를 가진 펩타이드가 상당수 포함되어 있었기 때문으로 판단된다. 이후 펩타이드 서열들을 통합해 FASTA 파일로 만들어 자체 데이터베이스를 생성하였고, 이를 Proteome

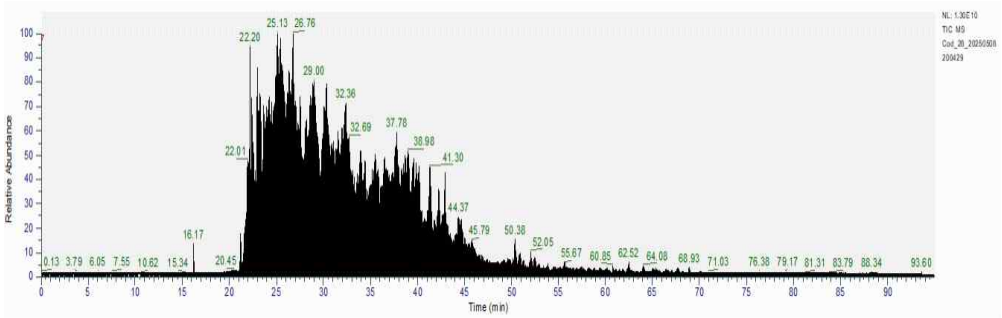
Discoverer 소프트웨어에 삽입하여 참조 데이터베이스로 활용하였다. 이를 통해 대구＃16은 98개, 대구＃28은 91개, 대구＃153은 152개, 명태＃116은 192개, 명태＃156은 203개, 명태＃173은 333개, 틸라피아＃23은 106개, 틸라피아＃157은 159개, 틸라피아＃176은 92개의 펩타이드 서열이 동정되었다(표 18). 결과적으로 봤을 때, 추가적인 조건으로 인한 필터링으로 De novo sequencing을 통해 얻은 결과보다 펩타이드 동정 개수가 감소한 것을 확인하였고, 보다 신뢰성 있는 서열 결과 데이터를 얻을 수 있었다.

다음 단계에서는 확보한 서열 데이터를 활용해 콜라겐 간의 공통 서열과 특이서열을 확인해보았다(그림 18, 그림 19). 우선, 각 어종 내에서 벤다이어그램을 통해 서열 데이터를 비교해보았다. 그 결과, 대구(Cod)에서 16개, 명태(Pollock)에서 6개, 틸라피아(Tilapia)에서 1개의 공통 서열이 각각 확인되었다(표 19). 이어서 각 어종의 서열 데이터를 하나로 모아 어종 간 서열 유사성을 비교해본 결과, 명태와 틸라피아 사이에서 3개의 공통 서열이 검출되었다(표 20). 이러한 분석은 어종 간 콜라겐 서열 유사성 및 특이성을 규명하는 데 기여할 수 있으며 향후 생물 종 특이적 바이오 마커로 활용하기 위한 기반 자료로 활용될 수 있다.

(a)



(b)



(c)

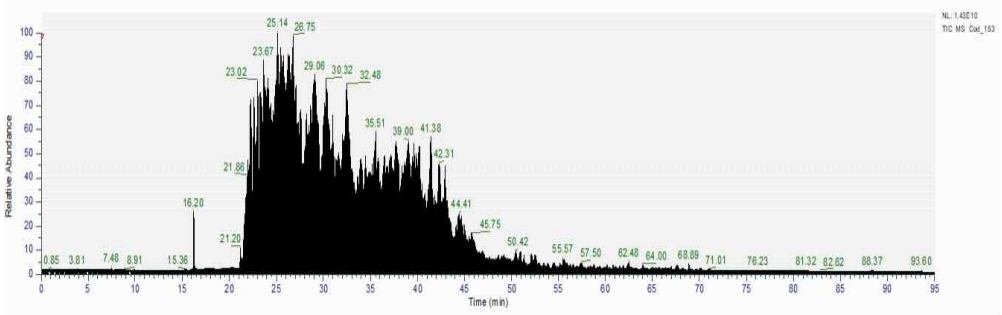
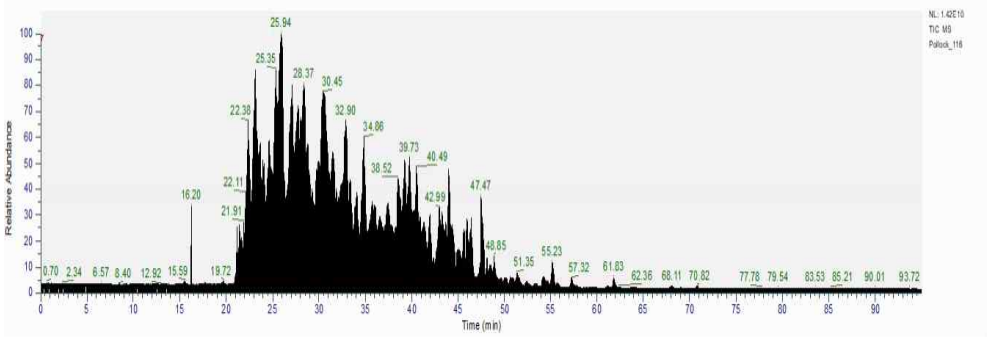


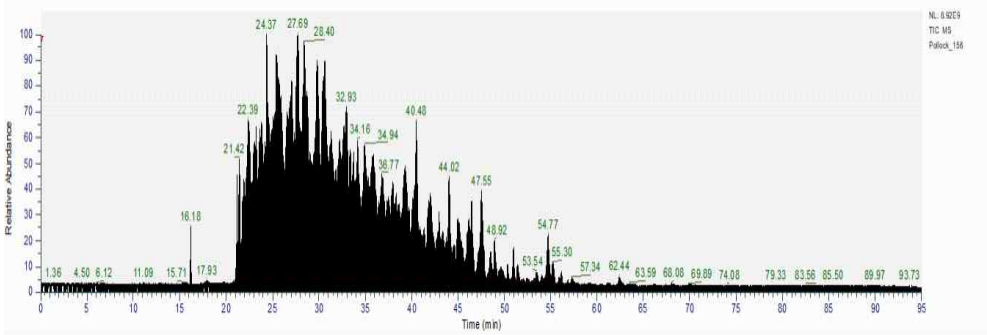
그림 17. 콜라겐 질량 분석 Total ion chromatogram

(a) Cod＃16, (b) Cod＃28, (c) Cod＃153

(a)



(b)



(c)

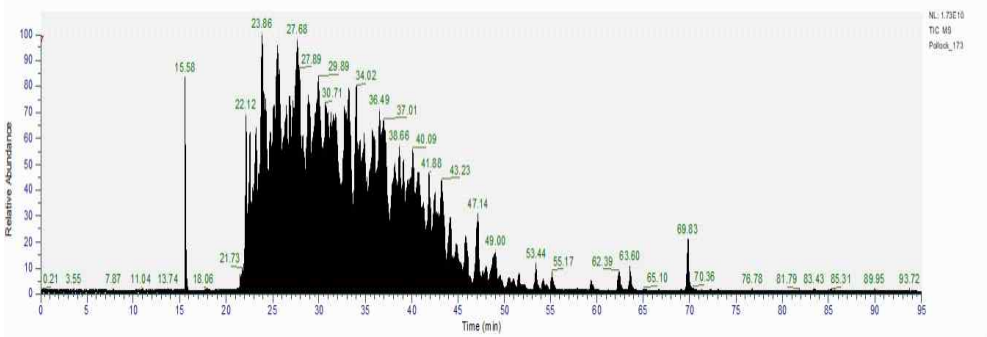
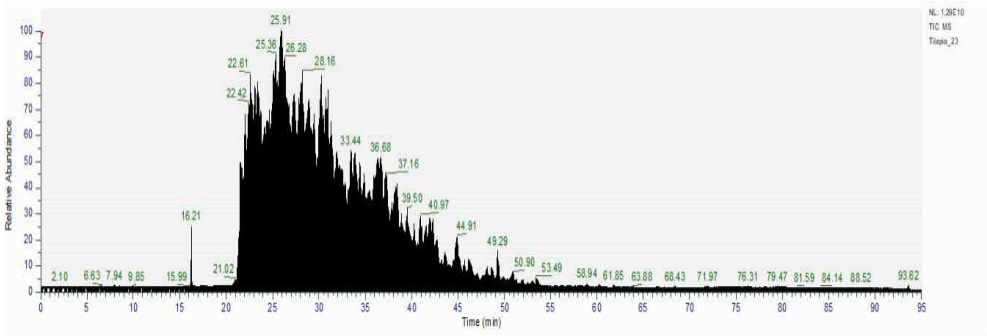


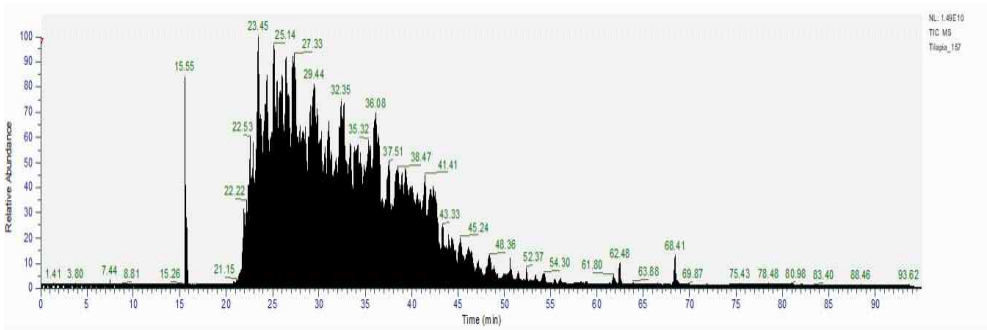
그림 17. Continued

(a) Pollock #116, (b) Pollock #156, (c) Pollock #173

(a)



(b)



(c)

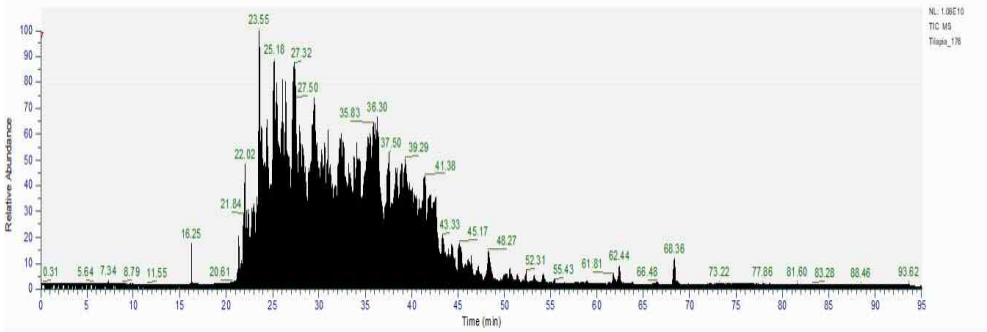


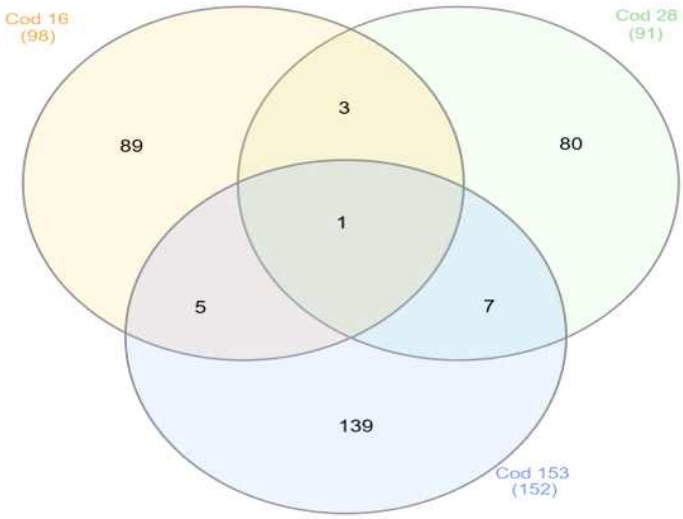
그림 17. Continued

(a) Tilapia #23, (b) Tilapia #157, (c) Tilapia #176

표 18. ALC ≥ 75% 조건에서 식별된 펩타이드 수

Collagen	De novo 기반 펩타이드 수 (Peaks Studio)	DB 검색 기반 펩타이드 수 (Proteome Discoverer)
Cod 16	296	98
Cod 28	325	91
Cod 153	329	152
Pollock 116	1117	192
Pollock 156	874	203
Pollock 173	894	333
Tilapia 23	537	106
Tilapia 157	439	159
Tilapia 176	278	92

(a)



(b)

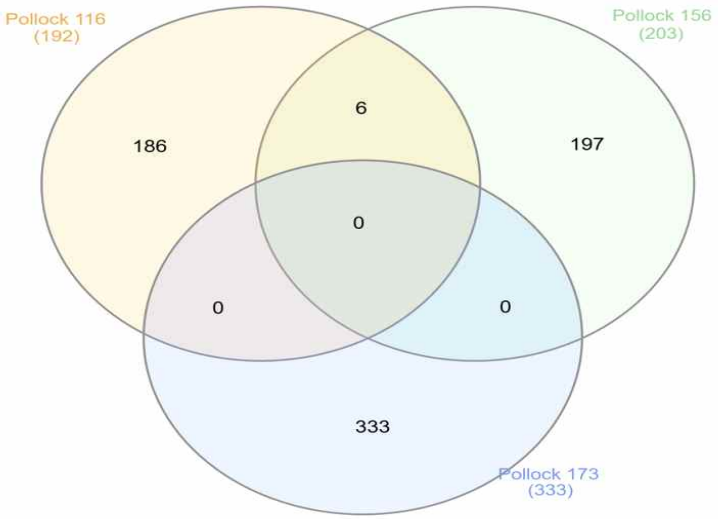


그림 18. 콜라겐 어종 내 공통 및 특이서열 비교

(a) Cod(대구), (b) Pollock(명태). 대구의 경우, 16개의 공통 서열이, 명태의 경우 6개의 공통 서열이 화인되었다.

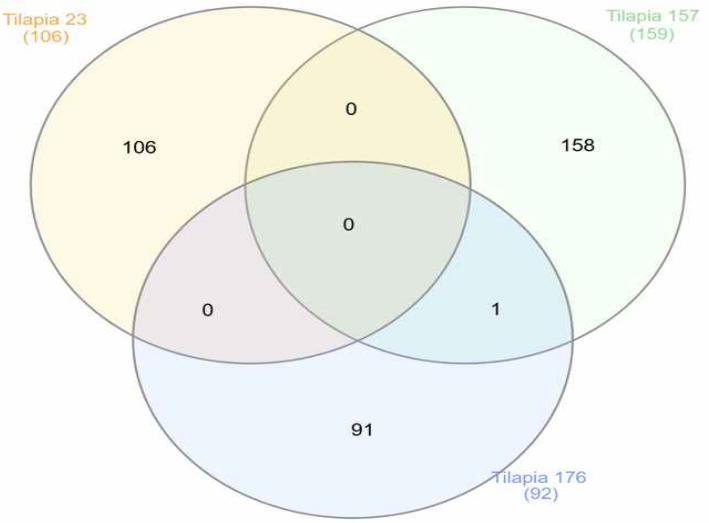


그림 18. Continued

Tilapia(틸라피아), 1개의 공통 서열이 확인되었다.

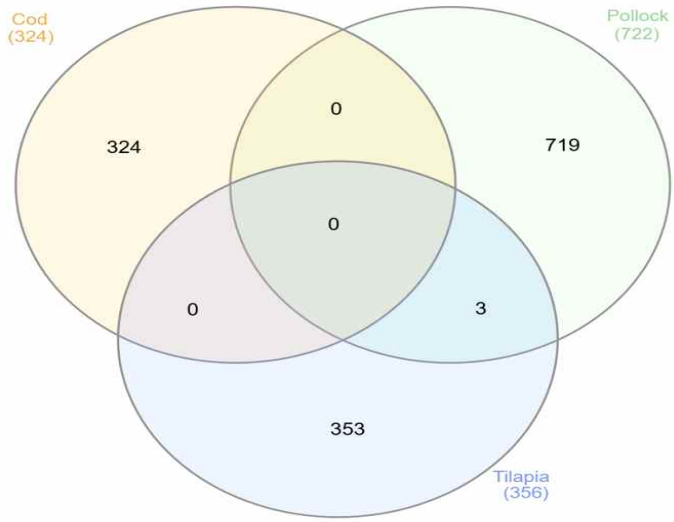


그림 19. 콜라겐 어종 간 공통 및 특이 서열 비교

Pollock(명태)와 Tilapia(틸라피아) 사이에서 3개의 공통 서열이 확인되었다.

표 19. 콜라겐 어종 내 공통 서열 목록

Collagen	Identical sequence
Cod 16 & Cod 28	EGGGGAGQDGYTHK KVGLAHE AVSGPPG
Cod 16 & Cod 153	KGLGESGAHGSSGSEATAQR PGACALASGLLGAQ RVPPALGEYAVGYT EPSGGYEDH APGAGAPGTSGAPGQLGNA
Cod 28 & Cod 153	PGAGLGAAGEGGQLDARG RAGNPGSDGSQVQGN VAAGASGEVGRE LGEGVVNAGAPGAAGAR SVGAGTHAGAGQAH SVGAGHTAGALYT AGGPGPF
Cod 16 & Cod 28 & Cod 153	AGHAGVN
Pollock 116 & Pollock 156	QPVENWPGPPGH TPQGFPGEGR WGAPGAPGSQ AGAPGAPGSQ ALAGVGAE TGPPGAP
Tilapia 157 & Tilapia 176	QDGSLGQDAALLQ

표 20. 콜라겐 어종 간 공통 서열 목록

Collagen	Identical sequence
Pollock & Tilapia	WGGGGVGVVGQHQ ADGLLYT TGELGAA

Ⅳ. 결론

액체 크로마토그래피-질량 분석법은 높은 분리능와 정밀성으로 인해 단백질 특성 확인에 보편적으로 사용되는 기술이다. 이에 따라 본 연구에서는 다양한 전처리 방식을 통해 시료를 준비하고, LC-MS 및 LC-MS/MS를 활용하여 그에 맞는 분석법을 적용해 단백질의 특성을 규명하고자 하였다. 분석에 사용된 시료는 HLA class I 복합체, 치료용 단일 클론 항체, 그리고 콜라겐의 세 가지로 구성되었다.

세포 독성 면역에 관여하는 HLA class I 분자의 펩타이드 서열을 규명하고 글라이칸 패턴을 파악하는 것은 그 분자적 특성을 이해하는 데 있어 필수적인 요소이다. 이를 위해 HLA class I 복합체를 효과적으로 추출할 수 있는 최적의 lysis 조건을 설정하고, 분석을 수행하였다. 우선, lysis 시간 조건을 설정하기 위해 4시간, 8시간, 18시간, 24시간, 48시간으로 실험을 진행했으며, 웨스턴 블롯을 통해 단백질의 발현을 확인한 결과 4시간 조건에서 가장 높은 수준의 발현률이 확인되었다. 이에 따라 이후의 실험은 모두 4시간으로 고정해 진행했다. 다음으로, 면역 침강법을 사용해 HLA class I 복합체만을 특이적으로 분리했고 이를 이용해 HLA class I 펩타이드 서열 분석 및 글라이칸 분석 실험을 각각 수행하였다. 먼저, 펩타이드 서열 동정을 위해 정제된 HLA class I 복합체에 10% acetic acid를 처리해 펩타이드만을 분리하고, 이를 LC-MS/MS로 분석했다. 그 결과, 1400개 이상의 펩타이드 서열이 동정되었으며, 서열들을 분류한 결과 10개의 아미노산 길이(10 mer) 길이의 펩타이드가 가장 높은 비율로 검출되었다. 또한, HLA class I 분자의 글라이칸 동정 실험 결과, G0F, G1F, G2F 형태의 글라이칸이 높은 비율로

관찰되었으며, 그 외에도 GlcNAc가 제거된 형태의 글라이칸들이 추가로 확인되었다. 일반적으로 항체와 같은 단백질 의약품에서는 G0F가 가장 높은 비율로 검출되지만, HLA class I의 경우에는 G1F가 가장 많이 나타났고, 이를 통해 단백질마다 고유한 글라이코실화 패턴을 가지고 있음을 확인할 수 있었다.

단백질 의약품인 치료용 항체의 글라이코실화는 생물학적 활성 및 안정성에 영향을 미치므로, 항체의 글라이칸 패턴을 파악하는 것은 그 특성을 이해하는 데 필수적인 요소로 여겨진다. 이를 위해 총 5종의 오리지널 항체와 6종의 바이오시밀러 항체를 대상으로 PNGase F 처리 여부에 따른 LC/MS 결과를 비교함으로써 글라이칸 조합을 대략적으로 확인했다. 이후 글라이칸만을 분리해 질량 분석 및 LC-FLR 분석 실험을 수행함으로써 글라이칸을 동정하였다. 우선 PNGase F 처리에 따른 항체의 분자량 변화를 분석한 결과, 모든 항체 시료에서 G0F+G0F, G0F+G1F, G1F+G1F 또는 G0F+G2F 조합의 글라이칸이 확인되었다. 이중 Herceptin, Herzuma, Mabthera, Truxima에서는 추가로 G1F+G2F의 조합이 확인되었고, 이는 G2F의 상대적으로 높은 비율에 기인한 것으로 판단된다. 더욱 정밀한 글라이칸 동정을 위해 질량 분석을 수행한 결과, G0F, G1F, G2F의 주요 글라이칸이 순차적으로 높은 비율로 검출되었다. 또한, 일부 항체 시료를 제외하고 대부분에서 주요 글라이칸의 총합이 80% 이상으로 확인되었으며, 이를 통해 분석한 모든 항체 시료가 양호한 품질 특성을 지님을 확인할 수 있었다. 이 외에도 G0, Mannose-5, Mannose-6 및 G0F 또는 G0에서 GlcNAc이 한 개 제거된 형태 등의 글라이칸도 함께 낮은 비율로 검출되었다. 이러한 실험 결과는 항체의 글라이칸 조성이 품질 특성 평가에 있어 유의미한 지표가 될 수 있으며, 항체 간의 비교 평가에 효과적으로 활용될 수 있음을 보여준다.

대구, 명태, 틸라피아 유래의 콜라겐 펩타이드를 LC-MS/MS로 분석하여 공통 서열 및 특이서열을 규명하였다. 우선, 대구 유래 콜라겐 3종의 서열을 분석한 결과, 16개의 공통 서열 및 308개의 특이서열이 확인되었으며, 명태 유래 콜라겐에서는 6개의 공통 서열 및 716개의 특이서열이 확인되었다, 틸라피아 유래 콜라겐에서는 1개의 공통 서열 및 355개의 특이서열이 확인되었다. 이어서, 각 어종의 서열을 통합하여 대구, 명태. 틸라피아 간 비교를 수행한 결과, 명태와 틸라피아 사이에서 3개의 공통 서열이 검출되었고, 대구 324개, 명태 719개, 틸라피아 353개의 특이서열이 검출되었다. 이러한 펩타이드 서열 동정 결과는 공통 서열과 특이서열이 어종별로 구분됨을 보여주며, 이는 식품 원료 추적 및 품질 관리 등 다양한 분야에서 응용 가능한 어류 유래 바이오마커로서의 가능성을 제시한다.

참고 문헌

1. Shi, Y., Xiang, R., Horváth, C., & Wilkins, J. A. (2004). The role of liquid chromatography in proteomics. Journal of Chromatography A, 1053(1-2), 27-36.

2. Cruz-Tapias P, Castiblanco J, Anaya JM. Major histocompatibility complex: Antigen processing and presentation. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, et al., editors. Autoimmunity: From Bench to Bedside [Internet]. Bogota (Colombia): El Rosario University Press; 2013 Jul 18. Chapter 10.Available from: https://www.ncbi.nlm.nih.gov/books/NBK459467/

3. Boyne, C., Lennox, D., Beech, O., Powis, S. J., & Kumar, P. (2021). What is the role of HLA-I on cancer derived extracellular vesicles? Defining the challenges in characterisation and potential uses of this ligandome. International Journal of Molecular Sciences, 22(24), 13554.

4. Roche & Cresswell, 2017)Roche, P. A., & Cresswell, P. (2017). Antigen processing and presentation mechanisms in myeloid cells. Myeloid Cells in Health and Disease: A Synthesis, 209-223.

5. Dhatchinamoorthy, K., Colbert, J. D., & Rock, K. L. (2021). Cancer immune evasion through loss of MHC class I antigen presentation. Frontiers in immunology, 12, 636568.

6. Lee, J.-K., Choi, Y.-L., Kwon, M., & Park, P. J. (2016). Mechanisms and consequences of cancer genome instability: lessons from genome sequencing studies. Annual Review of Pathology: Mechanisms of Disease, 11(1), 283-312.

7. Ilca, F. T., & Boyle, L. H. (2021). The glycosylation status of MHC class I molecules impacts their interactions with TAPBPR. Molecular Immunology, 139, 168-176.

8. Hu, M., Zhang, R., Yang, J., Zhao, C., Liu, W., Huang, Y., Lyu, H., Xiao, S., Guo, D., & Zhou, C. (2023). The role of N-glycosylation modification in the pathogenesis of liver cancer. Cell Death & Disease, 14(3), 222.

9. Hoek, M., Demmers, L. C., Wu, W., & Heck, A. J. (2021). Allotype-specific glycosylation and cellular localization of human leukocyte antigen class I proteins. Journal of proteome research, 20(9), 4518-4528.

10. Mesonzhnik, N., Belushenko, A., Novikova, P., Kukharenko, A., & Afonin, M. (2024). Enhanced N-Glycan Profiling of Therapeutic Monoclonal Antibodies through the Application of Upper-Hinge Middle-Up Level LC-HRMS Analysis. Antibodies, 13(3), 66.

11. Singh, S. K., & Lee, K. H. (2022). Characterization of monoclonal antibody glycan heterogeneity using hydrophilic interaction liquid chromatography-mass spectrometry. Frontiers in Bioengineering and Biotechnology, 9, 805788.

12. Boune et al., 2020; Boyne et al., 2021; Dhatchinamoorthy et al., 2021; Dierickx et al., 2023; Hoek et al., 2021; Hu et al., 2023; Ilca & Boyle, 2021; Kumazawa et al., 2018; Lee et al., 2016; Mesonzhnik et al., 2024; Salvatore et al., 2020; Shi et al., 2004; Singh & Lee, 2022; Tiwari et al., 2023; Varunjikar et al., 2024; Wang et al., 2025; Yalovenko et al., 2020)Boune, S., Hu, P., Epstein, A. L., & Khawli, L. A. (2020). Principles of N-linked glycosylation variations of IgG-based therapeutics: pharmacokinetic and functional considerations. Antibodies, 9(2), 22.

13. Yalovenko, N., Yatsyna, V., Bansal, P., AbiKhodr, A. H., & Rizzo, T. R. (2020). Analyzing glycans cleaved from a biotherapeutic protein using ultrahigh-resolution ion mobility spectrometry together with cryogenic ion spectroscopy. Analyst, 145(20), 6493-6499.

14. Salvatore, L., Gallo, N., Natali, M. L., Campa, L., Lunetti, P., Madaghiele, M., Blasi, F. S., Corallo, A., Capobianco, L., & Sannino, A. (2020). Marine collagen and its derivatives: Versatile and sustainable bio-resources for healthcare. Materials Science and Engineering: C, 113, 110963.

15. Wang, M., Liu, Y., Pallarés, N., El Marsni, Z., Kousoulaki, K., & Barba, F. J. (2025). Determination of collagen types and mineral contents in fish skin and collagen-containing skin-derived protein hydrolysates before and after in vitro simulated digestion. Food & Function, 16(3), 1032-1040.

16. Tiwari, R., Mishra, J., Devhare, L. D., & Tiwari, G. (2023). An updated review on recent developments and applications of fish collagen. Pharma Times, 55(6), 28-36.

17. Varunjikar, M. S., Pineda-Pampliega, J., Belghit, I., Palmblad, M., Grøsvik, B. E., Meier, S., Olsvik, P. A., Lie, K. K., & Rasinger, J. D. (2024). Fish species authentication in commercial fish products using mass spectrometry and spectral library matching approach. Food Research International, 192, 114785.

18. Kumazawa, Y., Taga, Y., Takashima, M., & Hattori, S. (2018). A novel LC - MS method using collagen marker peptides for species identification of glue applicable to samples with multiple animal origins. Heritage Science, 6, 1-9.

19. Dierickx, K., Presslee, S., & Harvey, V. L. (2023). Rapid collagen peptide mass fingerprinting as a tool to authenticate Pleuronectiformes in the food industry. Food Control, 148, 109680.

20. (Lauber et al., 2015)Lauber, M. A., Brousmiche, D. W., Hua, Z., Koza, S. M., Guthrie, E., Magnelli, P., Taron, C. H., & Fountain, K. J. (2015). Rapid preparation of released N-glycans for HILIC analysis using a novel fluorescence and MS-active labeling reagent. Waters Application Note 720005275EN.

ABSTRACT

Mass Spectrometry Based Characterization of Peptides and Glycans in Biomolecules

- Mass Spectrometry Based Multi-Analytical Approach for the Characterization of Immune and Structural Proteins -

Youn Seo Chun
Department of Next Generation Applied Science
Graduate School of
Sungshin Women's University

Proteomics is a vital field of research aimed at understanding the diversity and complexity of proteins and peptides within biological systems. Among the various analytical techniques, mass spectrometry (MS) is widely regarded as a powerful tool due to its high precision in characterizing proteins. In this study, we aimed to establish precise analytical methodologies using mass spectrometry and to apply these approaches to the structural and functional characterization of various biologically relevant proteins, including HLA class I complexes, therapeutic antibodies, and collagen.

The Human Leukocyte Antigen (HLA) system plays a central role in

immune regulation and is categorized into classes I, II, and III. Among them, HLA class I molecules are responsible for presenting endogenously derived peptides to $CD8^+$ T cells, thereby playing a crucial role in the immune response. This antigen presentation not only contributes to immune surveillance but also has significant implications in the development and progression of diseases such as cancer. Therefore, identifying and characterizing the peptide sequences bound to HLA class I molecules expressed in tumor cells is essential for advancing antigen-based cancer immunotherapies.

For this purpose, the human colorectal cancer cell line HCT-116 was used. Prior to peptide isolation, we optimized cell lysis conditions to improve peptide extraction efficiency by varying the lysis duration. We observed the highest extraction efficiency at 4 hours, while longer lysis times resulted in a decline in peptide yield. Peptides were then analyzed using an Orbitrap mass spectrometer, and the resulting data were processed using PEAKS Studio and Proteome Discoverer software. Initially, de novo sequencing was conducted using PEAKS Studio to predict peptide sequences, which were then converted into a FASTA format file. This file was used as a custom database in Proteome Discoverer to identify and quantify the number of peptides matched. Through this approach, we successfully identified over 1,400 HLA class I peptides.

In addition, N-glycans, which are functionally significant in the immunological roles of HLA class I molecules, were analyzed. We compared the UPLC-FLR and mass spectrometry results to confirm the

overall glycan structures associated with HLA class I complexes.

Secondly, we investigated the glycosylation characteristics of therapeutic monoclonal antibodies using mass spectrometry-based approaches. The use of monoclonal antibodies has been steadily increasing in modern medicine due to their therapeutic efficacy against a variety of diseases. Accordingly, evaluating the heterogeneity of these antibodies is essential, especially since antibody glycosylation directly affects their therapeutic efficacy. Therefore, optimizing analytical methods for glycan profiling is considered a crucial step in the quality assessment of therapeutic antibodies.

In this study, we performed a comprehensive analysis of the glycosylation patterns of various originator antibodies and their biosimilar counterparts using a combination of LC-MS analysis of the intact antibodies and LC-MS analysis of the released N-glycans. First, the intact antibody samples were analyzed by mass spectrometry without enzymatic treatment to determine their molecular weights. Subsequently, PNGase F digestion was performed to identify mass shifts indicative of glycan removal, enabling inference of the glycan compositions attached to the antibodies. The released N-glycans were further analyzed using Ultra-Performance Liquid Chromatography with Fluorescence Detection (UPLC-FLR) and LC-MS, which enabled the identification of various glycan structures. Based on the calculated percentage of peak areas for the major glycans, it was observed that most antibody samples exhibited over 80% of total peak area, indicating favorable glycosylation characteristics in terms of quality. In addition, differences in the

distribution ratios of G0F, G1F, and G2F were observed among the antibodies, revealing subtle variations in glycan patterns.

Lastly, we conducted a proteomic analysis of collagen, a key structural protein in biological tissues, using a mass spectrometry-based approach. Low-molecular-weight fish collagen, typically obtained via enzymatic hydrolysis, is widely used in various industries. However, the functional and stability properties of collagen can differ depending on the species of origin, and commercial collagen products are often provided in processed forms, making it difficult to trace their source and quality. Therefore, there is a growing need for analytical techniques capable of accurately evaluating species-specific characteristics of collagen. Due to the low molecular weight and peptide complexity of hydrolyzed fish collagen, conventional methods often fall short in providing accurate sequence identification.

To overcome this limitation, we established a strategy that combines liquid chromatography with mass spectrometry (LC-MS) to analyze complex peptide mixtures and identify both common and species-specific peptide sequences. We analyzed nine collagen samples derived from three fish species—Alaska pollock, tilapia, and cod—using an Orbitrap mass spectrometer. Data analysis was performed using PEAKS Studio and Proteome Discoverer. As with previous experiments, de novo sequencing was conducted using PEAKS Studio, followed by the construction of a custom database that enabled the identification of shared and unique peptide sequences among different fish species in Proteome Discoverer.