



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](#)

고 병 준 교수 지도
석사학위 청구논문

LC-MS/MS를 이용한 저분자 어류
콜라겐 펩타이드의 중요 바이오 마커
식별 분석방법 개발

2024

성신여자대학교 대학원

미래응용과학학과

우 민 영

LC-MS/MS를 이용한 저분자 어류
콜라겐 펩타이드의 중요 바이오 마커
식별 분석방법 개발

고 병 준 교수 지도

이 논문을 석사학위논문으로 제출함

2023년 11월

성신여자대학교 대학원

미래응용과학학과

우 민 영

인 준 서

우민영의 석사학위 논문으로 인준함

2023년 11월

심사위원장.....임동현.....(서명 또는 인)

심사위원고병준.....(서명 또는 인)

심사위원강태현.....(서명 또는 인)

성신여자대학교 대학원

논문개요

콜라겐은 포유류에 가장 큰 비율을 차지하는 단백질군이다. 따라서 시중에서 대부분 판매되는 식용 콜라겐은 주로 포유류인 소나 돼지를 주 원료로 사용하고 있다. 하지만 질병 전파 및 종교적 문제로 인해 최근에는 포유류 콜라겐의 대체제인 해양 콜라겐에 관심이 쏠리고 있다. 해양 콜라겐 중에서도 어류 콜라겐은 비교적 포유류 콜라겐과 생체 적합성이 높아 다른 해양 생물들에 비해 가장 많이 활용되고 있다. 따라서 콜라겐의 품질관리와 제품 인증을 위해 다양한 어종 콜라겐의 특성을 식별할 수 있는 분석 프로토콜이 필요하다. ELISA와 FTIR은 콜라겐 단백질을 분석하는데 기존에 쓰이던 실험 방식들이다. 그러나 콜라겐의 경우, 낮은 반응성, 결과의 불일치, 그리고 위양성/위음성을 띄는 결과 값들에 의해 ELISA 와 FTIR의 분석방법은 다소 신빙성이 떨어진다. 그러므로 본 논문에서는 SEC 분석을 사용해 콜라겐 펩타이드의 분자량 크기를 확인하기고, 콜라겐 어종 분석을 위해 선택적 bottom-up LC-MS analysis 방법을 개발했다. 그 결과 총 5종의 어류 콜라겐에 해당하는 각각의 펩타이드 바이오 마커를 식별해냈다.

목 차

논문개요

I. 서론	1
II 실험 재료 및 방법	7
1. Chemical and Reagents	7
2. HPLC & LC-MS Equipment	9
3. Preparation of QC and calibration standards	9
4. Sample Preparation	10
1) Tryptic digestion sample	10
① 단백질 추출 및 펩타이드화	10
② 정제 과정 (SPE)	11
2) Undigested sample	11
5. HPLC Experiment	12
6. LC-MS/MS Experiment	12
7. Database Search	15
III. 결과 및 고찰	17
1. 1. SEC-HPLC 분자량 분석	17

2. 2. LC-MS/MS 특이서열 분석	19
1) 전처리 방법 연구	19
2) 컬럼 선택	27
3) 이동상 선택	32
4) 어종의 특이서열	36

IV. 결론	43
------------------	----

참고문헌

ABSTRACT

List of Figures

Figure 1. Collagen's most common PTMs	16
Figure 2. SEC molecular distribution of each collagen (220nm)	18
Figure 3. Pep-map of trastuzumab peptide with 25°C(a) and 37°C(b) denaturation temperature	22
Figure 4. Total sequence coverage of trastuzumab with 25°C(a) and 37°C(b) denaturation temperature	23
Figure 5. Total sequence coverage of trastuzumab with 10 ug(a), 25 ug(b), 50 ug(c), and 100 ug(d) digested proteins	25
Figure 6. Table of residue local confidence% for tryptic digested #16 (a), #135(b) collagen and un-digested #16(c), #135(d) collagen	26
Figure 7. TIC comparison between CSH and BEH of #11(a), #23(b), and #180(c) collagen	29
Figure 8. Corresponding RT peaks from each collagen peptide. Each line represent change in the elution time of each peptide. (#11(a), #23(b), #180(c))	30
Figure 9. Pep-map TIC obtained from different mobile phase (0.1%FA, 0.2%FA included water and ACN) of #180(a) and #182(b) collagen	34
Figure 10. Pep-map MS spectrum obtained from different mobile phase (0.1%FA, 0.2%FA included water and ACN) of #180(a) and #182(b) collagen	35
Figure 11. Venn diagram of all 5 types of fish collagen peptide	38

List of Tables

Table 1. Collagen Sample List	8
Table 2. LC–MS/MS gradient conditions	14
Table 3. Total number of de novo peptide identified for #16 and #135 collagen with different pretreatment	26
Table 4. Number of total peptide ID from BEH and CSH column analysis	31
Table 5. List of overlapping peptide sequence	39
Table 6. Specific peptide sequence in each fish collagen (Tilapia, Catfish, Cod, Pollock, and Carp)	40

I. 서론

콜라겐은 피부, 뼈, 인대, 힘줄, 연골 등 신체 내 다양한 결합 조직에 존재하는 가장 풍부한 단백질이다.¹ 또한 콜라겐은 상처 치유 과정의 중요한 구성 요소로, 새로운 조직 성장을 위한 기질 역할을 하며 지혈, 염증, 증식 및 리모델링을 포함한 상처 치유의 모든 단계에서 필수적으로 사용된다.^{1 2} 주로 콜라겐을 구성하는 아미노산은 글리신(33%), 프롤린 및 하이드록시 프롤린(22%)으로 3개의 α 사슬이 삼중 나선 형태를 띠고 있다. 각 알파 사슬의 분자량은 약 100 kDa이며 평균적으로 1,014개의 아미노산으로 구성되어 있다.³ 그리고 왼쪽 나선형 알파 사슬들이 3개가 모여 오른쪽 방향의 슈퍼 나선형태 구조(4차 구조)를 갖게 된다. 이런 콜라겐 구조는 인접한 사슬 내의 글리신 아미노산들 사이에서 수소 결합을 형성해 어떠한 조직보다 안정한 형태를 갖는다.⁴ 하지만 인체의 노화가 지속되면서 기질내 금속단백질 가수분해 효소의 합성량이 증가하고, 그에 따라 콜라겐 분해 속도도 증가하게 된다. 그렇게 인체내에서 분해되는 콜라겐을 보충하기 위해 식품에 함유된 콜라겐(평균 분자량 ~29 kDa)이나 건강 보조식품으로 분류되는 식이콜라겐 보충제에 대한 관심이 늘어나고 있다.

현재 콜라겐 펩타이드 시장은 콜라겐의 인체내 효능에 대한 인식이 높아짐으로써 지속적으로 성장하고 있다. 콜라겐 펩타이드 시장은 연평균 성장률(CAGR)이 6.5%로 예측되며 2028년까지 16억 달러의 매출을 달성할 것으로 예상된다. 콜라겐 펩타이드는 식음료의 건강 보조 식품뿐만 아니라 화장품, 헬스케어 및 의약품의 생리 활성 성분으로 활용도가 많다.⁵ 기존 콜라겐들은 동물성 유래 소재로써 소와 돼지(피부, 뼈, 인대, 힘줄, 연골)로부터 가공되어 이용되었다. 그러나 전염될 수 있는 질병이나, 종교적, 그리고 문화적

신념의 문제로 인해 동물성 콜라겐의 적용과 사용이 제한되어 왔다. ⁶ 반면 어류는 동물성 콜라겐과 달리 낮은 비용과, 더 높은 수율로 화학적 및 생물학적 특성이 포유류와 유사한 콜라겐을 얻을 수 있다. 따라서 포유류의 대체제인 해양성 공급원을 사용해 콜라겐을 추출할 수 있는 방법들이 계속해서 개발되고 있다. 이는 콜라겐을 추출할 수 있는 기본적인 원료수급의 제한과 기존의 포유류 기원 콜라겐이 가지고 있는 단점들을 개선시킨 새로운 대체 자원으로 검토되고 있다.²

이렇게 다양한 활용가치가 있는 콜라겐을 최대한의 효율로 회수와 정제를 위해 다양한 방법이 사용되어 왔다. 일반적인 콜라겐 추출 공정은 비콜라겐 성분을 제거하는 추출 전 단계, 추출 공정 단계, 그리고 사전 농축과 정제 단계(초여과, 침전, 용해화, 투석) 등 여러 단계로 구성된다. ⁷ 추출 전 단계에서는 주로 세척, 동물 조직별 분리, 그리고 다음 전처리를 수월하게 진행하기 위해 동물의 조직들을 작게 자르거나 다져서 크기를 줄이는 작업으로 진행된다. ⁸ 이 작업 이후, 추출의 효율을 높이고 비콜라겐성 물질을 제거하기 위해 가벼운 화학적 전처리가 수행된다. 일반적으로 원료와 추출 방법에 따라 다양한 전처리(알칼리성 또는 산성 처리)방법이 잇따른다. ⁹ 산화 전처리에서는 조각된 원료를 유기산을 이용해 변성시킨다. ¹⁰ 다양한 유기산들 중 아세트산이 가장 추출성이 높으며 제일 흔히 사용되는 유기산 용매이다. ¹¹ 고로, 아세트산을 이용해 용매를 콜라겐 구조 내로 침투시켜 분자 간 및 분자 내 비공유 결합을 유도 분리한다. 또한 추출 단계에 앞서 뼈, 연골, 비늘 등 미네랄 함량이 높은 신체 부위에서 콜라겐 추출을 극대화 하기 위해 원료를 탈염 처리를 한다. 이 단계에서 일반적으로 EDTA 또는 HCl 이 사용된다. ¹² 동물 조직내의 네이티브 콜라겐은 피브릴 형태의 가교결합과 분자내 및 분자간의 강한 산화작용 때문에 거의 어떠한 수용성 용매에도 용해되지 않는 구조적 안정성을 갖는다. ¹³ 따라서

콜라겐 단백질의 가용성을 높이고 안정적인 분리를 달성하기 위해 기본적으로 acid-soluble collagen(ASC)과 pepsin-soluble collagen(PSC) 추출 기술이 사용된다.¹⁴ ASC 에서 사용되는 HCl 과 AcOH 는 콜라겐의 삼중나선을 가수분해하고 단일 사슬을 용액에 용해시켜 큰 분자량의 단백질이 더 짧은 펩타이드(0.3~8 kDa)로 해중합되는 과정을 거친다.¹⁵ 펩신을 이용한 PSC 는 ASC 다음으로 가장 많이 사용되는 방식으로, 삼중 나선의 텔로펩타이드 영역(케토이민 결합의 다수 존재)을 절단하여 용액 내 콜라겐 펩타이드의 침출을 촉진하고 회수율을 높일 수 있다.¹⁶ 이렇게 얻은 콜라겐은 저분자량(low molecular weight, LMW) 펩타이드로, 상대적으로 더 작은 분자량과 친수성이 높은 화합물로 가수분해 될 수 있다.¹⁷ 이를 콜라겐 가수분해물이라고 한다. 단백질 분해 효소를 이용한 가수분해 방식은 식품 산업에서 이미 상용화 되어 있고, 효소는 고온에서의 반응이 필요 없으므로 단백질의 변성없이 가공이 가능하다. 효소를 통해 가수분해된 펩타이드는 아미노산 서열 프로파일을 예측할 수 있다는 장점 또한 존재한다. 이렇게 생산된 저분자량 펩타이드 콜라겐은 혈류내 순환이 더 빨라져 생체 이용률을 더 높이는 것으로 추정되고 있어(체내 흡수율이 최대 1.5 배 더 높음) 더 큰 펩타이드 및 모단백질의 콜라겐보다 선호되는 제형이다.¹⁸

그러나 최근 식품 산업과 콜라겐 시장의 세계화로 인해 더 저렴하거나 불법 원료인 성분이 콜라겐 생산에 사용될 수 있으며, 그 결과 잘못된 라벨링, 허위공정 또는 소비자를 현혹하는 오류된 표기들이 사용될 수 있다.¹⁹ 특히, 수산업계의 허위 라벨링은 전 세계적으로 문제가 되고 있으며, 상업용 수산물의 평균 1/3은 허위 라벨링 또는 근거 없는 품종의 표기로 적발된 바 있다.²⁰ 따라서 식품의 규정 준수와 식품의 품질 및 안전을 보장하기 위해서는 콜라겐의 원료와 품종을 식별하고 구별할 수 있는 분석법이 필수적이다.

현재 각기 다른 콜라겐의 어종들을 추적하는 방법에는 여러 가지 분석방법들이 존재한다. 그 중에서도, 항원-항체 결합 특이성을 이용한 enzyme-linked immunosorbent assay (ELISA)와 같은 면역학적 분석방법이 특정 단백질을 식별할 때 가장 일반적으로 사용된다.²¹ 하지만 ELISA는 각 종의 상동성이 높은 단백질에 대해서는 종별 식별이 어렵다.²² Fourier transform infrared (FTIR)은 콜라겐의 존재와 화학 성분을 평가하고 인식하는 데 사용되는 도구이며, 콜라겐의 유형을 식별하는 데에도 사용된다. 또한 다양한 접근법으로 추출한 콜라겐 성분을 비교하거나 각각의 분리방법이 콜라겐 성분에 미치는 영향을 조사하는 데 적용되었다.²³ 하지만 FTIR 또한 콜라겐 내에 다른 어종의 콜라겐이나 기타 원료가 포함된 혼합물 콜라겐들을 분별하기에는 어려움이 존재했다.

그래서 본 연구는 분석기기 중 극미량의 시료로 고감도의 검출이 가능하며 동시에 각각의 미세한 질량 차이를 통해 아미노산 서열들을 특징지어 분리할 수 있는 liquid chromatography-mass spectrometry (LC-MS/MS)를 사용했다. 본 연구에서 사용된 LC-MS/MS는 시료가 역상분배(RP)-LC의 물리적 이동상을 통해 분리되어 ion source 부분에서 이온화 된 후 mass analyzer (Q-TOF)에서 질량에 따라 이온 조각이 분리된다. 그리고 검출기에 의해 이온 시그널들이 detection되면서 질량 스펙트럼을 얻게 된다.²⁴

분석 기기 외에도 분석된 정보들을 해석하는 도구, 특히 펩타이드 식별 알고리즘을 사용하는 processing software도 매우 중요하다. 펩타이드를 식별한다는 것은 펩타이드 내에 존재하는 아미노산 서열을 결정하는 과정이고, 이렇게 알아낸 펩타이드 서열을 통해 단백질의 종류를 유추할 수 있다. 질량 분석 데이터를 사용하여 펩타이드를 식별하는 방법에는 데이터베이스 검색, 스펙트럼

라이브러리 검색, 그리고 *de novo* 접근법 등 여러 가지의 검색도구가 존재한다.²⁵ 이 중 가장 일반적이고 가장 잘 알려진 펩타이드 식별 알고리즘은 데이터베이스 검색방법이다. 데이터베이스는 계층 시퀀싱에서 파생된 단백질 서열의 모음으로, 일반적으로 기존에 연구되었던 유기체에서 인코딩된 모든 단백질들을 포함한다. 그리고 데이터베이스 검색은 간단한 규칙을 사용하여 데이터베이스 펩타이드 서열에서 이론적 모델 스펙트럼을 생성하고, 이러한 단순화된 모델 스펙트럼을 직접 분석한 스펙트럼과 비교하여 대조율을 매기게 된다.²⁶ 그래서 데이터베이스 검색은 다른 방법들보다 더 정확하며 가장 빈번히 사용되고 있다. 또 다른 방법인 스펙트럼 라이브러리 검색은 관찰된 스펙트럼을 단백질 서열정보가 아닌 기준 물질들의 실제 스펙트럼 라이브러리를 통해 비교된다. 스펙트럼 라이브러리 검색 방법은 데이터베이스 검색보다 프로세싱 과정이 더 빠르고 민감하다. 하지만 스펙트럼 라이브러리는 일반적으로 인간과 생쥐를 포함한 소수의 모델에만 사용할 수 있다.²⁷ 데이터베이스나 스펙트럼 라이브러리 검색방법을 사용할 수 없는 경우, 데이터베이스나 라이브러리가 필요 없는 *de novo* 펩타이드 식별 검색을 이용할 수 있다. *De novo* 알고리즘은 다른 검색도구들과 달리 단편 이온 피크 사이의 간격(즉, 질량 차이)을 계산해서 직접 펩타이드 서열을 결정하고, 다양한 아미노산 조합 중에서 가장 최상의 서열을 선택하여 결과값을 도출해준다.²⁸ 본 논문에서는 데이터베이스 검색과 *de novo* searching 접근법을 모두 적용하여 정밀 질량 분석에 새로운 콜라겐 펩타이드 식별 방법을 개발했다.

따라서 LC-MS/MS를 이용한 분석은 단백질 분자의 질량을 정확하게 측정하여 단백질의 식별, 변형 및 상호 작용을 연구하는 데 필수적이다.²⁹ 그래서 본 연구는 질량 스펙트럼과 이론적 질량 데이터, 그리고 *de novo* sequencing을 이용하여 이미 database에 존재하는 기 확보된 펩타이드 서열들의 정보와

그렇지 못한 서열들을 가지고 특징적인 펩타이드를 검출하여 저분자 어류 콜라겐의 표적 서열들을 식별했다.

II. 실험 재료 및 방법

1. Chemical and Reagents:

실험에 사용된 저분자 콜라겐 시료들은 출처가 확인된 시중에 판매되고 있는 콜라겐 총 21개를 선별하여 실험하였다. 21개의 콜라겐 샘플들은 명태(pollock)에서 유래된 시료 4개 (no.116, 156, 173, 180), 대구(cod)에서 유래된 시료 3개 (no.16, 28, 153), 틸라피아(Tilapia)에서 유래된 시료 3개 (no.23, 157, 176), 메기(Catfish)에서 유래된 시료 3개 (no.11, 135, 174), 틸라피아와 잉어(carp)에서 유래된 시료 5개 (no. 3, 7, 146, 150, 168), 틸라피아와 메기에서 유래된 시료 1개 (no.171), 그리고 연어에서 유래된 시료 2개(no.21, 182) 로 구성이 되어있다. (Table 1.)

이 실험을 위해 사용된 water, LC/MS grade (Cat.no. W6-212)와 acetonitrile, LC/MS grade (Cat.no. A955-1) 용액은 Fisher Chemical사에서 구매하여 사용했으며, Formic Acid, LC-MS Grade (Cat.no. 28905)는 Thermo Scientific™사의 제품을 이용하였다. 샘플 전처리를 위한 1M Tris(hydroxymethyl) aminomethane buffer (Cat.no.IBS-BT019)는 인트론바이오에서 구매했으며, DL-Dithiothreitol (Cat.no.D9779)과 Iodoacetamide(Cat.no.I1149)는 Sigma Aldrich사의 것을 사용했다. Denaturation 용액을 제조하기 위한 Urea (Cat.no. GE3329)는 Glentham Life Science 제품을 사용했다. RapiZyme Trypsin (Cat.no. 186010106)는 Waters사에서 구매했으며 단백질 정제에 사용된 Sep-Pak C18 1 cc Vac Cartridge, 50 mg Sorbent (Cat.no.WAT054960)또한 Waters사의것을 사용했다.

Table 1. Collagen Sample List

No	검체명	판매국	유통/판매인	제조/수입원	제형	어종	원산지	Certificate 유무	저분자 타이틀	분자량 (Da)	콜라겐/젤라틴 함유량 (검각물 유무)
3	저분자 피쉬콜라겐	한국	신비인터내셔널	㈜젤텍2공장	Powder	틸라피아, 잉어	베트남, 방글라데시 (한국 젤텍)	○	저분자	1000 이하	100%
7	다당 저분자 피쉬콜라겐	한국	다당	㈜젤텍2공장	Powder	틸라피아, 잉어	베트남, 방글라데시 (한국 젤텍)	○	저분자	1000 이하	100%
11	피쉬콜라겐 분말	한국	신비인터내셔널	-	Powder	메기	베트남 (VINH HOAN COLLAGEN)	○	저분자	500 이하	100%
16	어바폴오브네이저 노르웨이 피쉬콜라겐 180	한국	볼렉오닉스	비타민나우	Powder	대구 (Gadus morhua)	노르웨이 (Seagarden AS)	○	저분자	500-2000	98.8%
21	블루콜라겐 파우더	캐나다	케이씨아이코리아	KCI Trading Inc. Port Coquitlam B.C	Powder	연어	캐나다	X	저분자	-	79.37%
23	피쉬콜라겐	한국	-	힐링	Powder	틸라피아	베트남, 방글라데시 (한국 젤텍)	○	저분자	1000	100%
28	코드 피쉬 콜라겐	한국	뉴로하스	비타민나우	Powder	대구 (Cod, Gadus morhua)	노르웨이 (Seagarden AS)	○	저분자	500-2000	99%
116	그린약초 저분자 피쉬 콜라겐	한국	그린약초	오벨로	Powder	명태 (Alaska Pollock)	러시아 (HUBEI NUTRATIDE BIOTECH)	○	저분자	300 이하	100%
135	락토테미 저분자 콜라겐 펩타이드	한국	박부자네	디딩인터내셔널	Powder	메기 (Pangasius hypophthalmus)	베트남 (VINH HOAN COLLAGEN)	○	저분자	500	100%
146	푸른돌판 저분자 피쉬콜라겐	한국	푸른산	㈜젤텍2공장	Powder	틸라피아, 잉어 (Carp)	베트남, 방글라데시 (한국 젤텍)	○	저분자	500 이하	100%
150	초저분자 피쉬콜라겐	한국	큐어영	담원인터내셔널	Powder	틸라피아, 잉어 (Katla)	인도 (ATHOS Collagen)	○	저분자	300 이하	100%
153	오리진 마린 피쉬 콜라겐 100	한국	오뉴바이오	비타민나우	Powder	대구	노르웨이	○	저분자	1000	99%
156	저스트콜라겐 초저분자 피쉬콜라겐	한국	콜마 제공	저스트 콜라겐	Powder	명태 (Alaska Pollock)	러시아	○	저분자	300	100%
157	준코 프랑스 저분자 피쉬콜라겐	한국	준코 코리아	트라이콤 무역	Powder	틸라피아	프랑스 (LESAFFRE INGREDIENTS SERVICES) Naticol BPMG	○	저분자	2000 이하	100%
168	저분자 피쉬 콜라겐 펩타이드 비오틴	한국	뉴센트리	채움 바이오	Tablet	틸라피아(나일), 틸라피아, 잉어 (Katla)	인도	○	저분자	300	98.8%
171	웰리스 피쉬콜라겐 펩타이드	한국	다인내추럴	팜두팜2공장	Powder	틸라피아, 메기 (Pangasius hypophthalmus)	이탈리아	○	저분자	3000 이하	100%
173	The best 저분자 피쉬콜라겐 펩타이드	한국	비카인	-	Powder	명태 (Alaska Pollock)	베트남 (VINH WELLNES)	○	저분자	500	100%
174	저분자 피쉬 콜라겐 펩타이드	한국	명성코리아	키오코리아	Powder	메기 (Pangasius hypophthalmus)	베트남 (VINH HOAN COLLAGEN)	○	저분자	500	100%
176	좋은아침 저분자 피쉬 콜라겐	한국	권병푸드	트라이콤 무역	Powder	틸라피아	프랑스 (LESAFFRE INGREDIENTS SERVICES)	○	저분자	2000	100%
180	자연지에 알래스카 저분자 피쉬콜라겐	한국	닥터스랩	아마존트레이딩	Powder	명태 (Alaska Pollock)	미국 (CERTIFIED NUTRACEUTICAL)	○	저분자	300	100%
182	초저분자 콜라겐 펩타이드	한국	세영상사	케이씨아이코리아	Powder	연어	캐나다 (KCI Trading Inc. Port Coquitlam B.C.)	○	저분자	500	79.37%

2. HPLC & LC-MS Equipment:

시중에서 판매되고 있는 저분자 어류 콜라겐 펩타이드의 전체적인 분자량의 측정은 용매 탈기 장치, 바이너리 펌프, 컬럼 히터 및 자동 샘플러가 장착된 SEC-HPLC (Waters Alliance e2695 Separation Module)를 통해 분석하였다. 검출기는 190~800 nm 영역대에서 모든 파장을 볼 수 있는 Waters 2998 PDA Detector를 사용하였다. 평균 분자량을 얻기 위해서는 Empower GPC software를 사용했다.

LC-MS 장비는 ACQUITY UPLC I-Class 시스템에 연결되어 전기분무 이온화(ESI) 인터페이스를 이용하는 Waters Xevo G2-XS QTOF(MA, USA) 질량 분석기기를 사용했다. 검출은 ACQUITY UPLC TUV Thermally Enhanced (Cat.no. 186015031) 검출기를 통해 UPLC분석을 수행했다. 데이터 수집 및 처리는 Empower 3.6.1. 와 MassLynx V4.2 SCN 1018소프트웨어가, 펩타이드 매핑 후 데이터 번역을 통한 동정과정은 PEAKS studio X 소프트웨어가 사용되었다.

3. Preparation of QC and calibration standards

HPLC 표준의 경우 트라스투주맙과 gel filtration standard (Bio Rad, Cat.no.1511901)을 사용하여 LC 분석의 품질을 확인하였다. 150 mg의 트라스투주맙은 75 ml 1X PBS버퍼에 녹여 2 mg/ml의 스톡 솔루션을 만들어 필요한만큼 분주하여 표준물질로 사용되었고, gel filtration standard는 1.8 ml의 MS water로 녹여 10 mg/ml의 stock solution을 만든 후 필요한 만큼 분주하여 사용하였다.

ESI-QTOF-MS는 질량 측의 보정을 위해 sodium iodide (NaI)를 이용하였고 leucine enkephalin으로 실시간 보정을 진행했다. Leucine enkephalin (Cat.no. 186006013)은 Waters사에서 제공된 standard로 사용하였고, sodium iodide는 10 mg sodium hydroxide (NaOH)알갱이(≥99%, Sigma-Aldrich)를 50 mL물에 용해시켜 5mM NaOH 스톡용액을 만들어 주었다. 그 다음, 100 ml 부피 플라스크에 10 mL의 5 mM NaOH용액을 옮겨준 후 90 ml의 이소프로필알코올(IPA)로 희석하여 NaOH standard 용액을 만들었다.

4. Sample Preparation

1) Tryptic digestion sample

① 단백질 추출 및 펩타이드화

Pre-made buffer 1M Tris-HCl(pH 8)을 50 mM로 희석해주어 이 실험의 모든 용질을 녹여주는 solvent스톡을 준비해 준다. 그리고 solvent stock 50 mM Tris-HCl에 8M 요소/5 mM DL-Dithiothreitol(DTT)를 용해하여 변성 버퍼를 준비한다. 50 *ug*의 표적 단백질양을 기준으로 용액의 농도가 1 mg/ml가 되도록 50 *ul*의 변성 버퍼에 녹여준 다음 37°C에서 1시간 동안 배양한다. 알킬화 용액은 Iodoacetamide(IAA)를 MS grade water에 녹여 30 mM 농도로 만들어 준다. 50 *ul*의 단백질 sample 용액에 IAA의 최종 농도가15 mM로 희석되도록 IAA 용액을 50 *ul* 첨가하여 25°C 암실에서 45분 간 배양한다. 단백질 sample 용액 내에 존재하는 요소의 농도를 2M으로 낮추기 위해, 100 *ul*의 50 mM Tris-HCl(pH 8)을 4M의 요소가 존재하는 100 *ul*단백질 sample 용액과 섞어준다. 100 *ug* 라피자임 트립신 1바이알을 MS grade water 200

ul 에 재혼합하여 0.5mg/mL 용액을 만든다. 이렇게 용매화된 트립신은 -20 °C에서 1달동안 보관할 수 있다. 준비한 단백질 샘플에 20 ul의 트립신을 넣어 주어 1:5의 효소: 단백질 비율로 37 °C에서 1시간 동안 분해 반응을 진행해 준다.

② 정제 과정 (SPE)

OASIS HLB 1cc Vac Cartridge(30 mg)를 15 mL falcon tube에 올려 준비한다. 필터 카트리지의 활성화를 위해 1 mL의 100% ACN을 컬럼에 넣고 주사기의 압력을 이용해 resin이 마르지 않게 천천히 용액을 통과시켜 준다. ACN이 통과된 카트리지를 시료와 유사한 화학 환경으로 교체하기 위해 1 mL의 solvent A (5% ACN+95% MS grade water)를 위와 같이 주사기를 이용해 컬럼이 마르지 않게 3번 통과시켜준다. 평형화가 이루어진 컬럼에 펩타이드화된 단백질 샘플을 로딩한 후 방금 단백질 샘플이 담겨있던 마이크로 튜브에 카트리지를 옮겨 sample을 통과시켜 준다. 마이크로 튜브에 담긴 샘플을 다시 한번 카트리지에 로딩 시켜주어 다시한번 필터링을 진행해 준다. 이 다음, 단백질과 결합된 카트리지의 세정을 위해 1 mL의 solvent A를 3번 반복하여 통과시킨다. 카트리지를 새로운 마이크로 튜브에 얹혀 놓고 70%의 ACN을 컬럼에 넣어 마지막에 기포가 생길 때까지 주사기의 압력으로 모든 용매를 추출한다. 이렇게 분석 물질과 흡착제 사이의 결합을 끊어 정제된 단백질만을 용출한 시료는 진공농축기를 이용해 25°C에서 70% ACN을 모두 증발시킨다. 용매가 모두 증발된 샘플은 분석할 때까지 -80°C에서 보관했다.

2) Undigested sample

준비된 가루 콜라겐을 각각 저울을 이용해 2.0 mg을 측정하여 분주한다. 분주

된 콜라겐들을 MS grade water로 녹여 샘플의 농도가 2.0 mg/ml 가 되도록 맞추어 준다. Vortexing을 통해 가루들을 충분히 녹여준 다음 centrifuge로 모든 sample을 다운 시켜준다. 이렇게 준비된 sample은 Advantec사 cellulose acetate membrane filter(Cat.no. 13CP020AS)를 사용하여 육안으로 보이는 이물질들을 걸러내고 단백질 펩타이드들을 정제한다. 정제된 sample들은 분석을 위해 insert vial이 내재된 glass vial에 100 μ l 이상 각각 분주했다.

5. . HPLC Experiment

Size exclusion chromatography (SEC)를 통해 상대적 분자량을 측정하기 위해 size marker standard를 이용해 각 elution time에 해당하는 질량 값의 기준점을 만들어 주었다. Isocratic 이동상은 1X PBS (pH.7.4, HanLAB)를 사용하였고 seal wash와 needle wash는 10% ACN과 50% ACN을 사용했다. 컬럼은 TSKgel G3000SWXL 7.8 mm I.D x 30 cm, 5 μ m를 이용했으며 컬럼온도는 25°C로 유지해주었다. 각 샘플을 전체 런타임은 20분이며 total injection vol.은 각 sample의 단백질 양이 10 μ g이 들어가도록 설정하였다. PDA검출기의 파장은 단백질의 흡광도를 확인할 수 있는 220 nm와 280 nm로 특정하여 콜라겐들을 분석하였다.

6. LC-MS/MS Experiment

LC-MS/MS는 대기압 양이온 이온화 분석기 모드에서 “sensitive” mode 데이터 수집을 선택하여 분석을 행하였다. 정확한 전구체의 질량과 전구체에서 유래된 단편 이온 데이터를 수집하기 위해 m/z 300-2000 질량 범위가 선택되었다. 샘플링 콘 전압, 소스 온도 그리고 용해 온도는 각각 25V, 150°C 및 450°C로 설정해주었다. 용해 가스는 질소가 사용되었고 콘 가스 유속은 50

L/h이었다. 충돌 가스로는 아르곤이 사용되었으며, 고에너지 이온화를 위해 충돌 에너지는 LM CE와 HM CE Ramp Start는 6-65, LM CE 와 HM CE Ramp End는 9-85사이의 전압을 이용하였다. 분석은 전체 스캔 모드에서 수행되었으며 MS스캔 시간은 0.2초, MS/MS 스캔 시간은0.1초로 고정하였다. 최적화된 MS 조건의 질량 정확도와 재현성을 보장하기 위해 200 pg/ml leucine enkephalin(양이온 mode: m/z 556.2771)을 10 *ul*/min의 유속으로 lock spray를 사용하였다. LC gradient 분석조건은 Table 2. 와 같다.

Table 2. LC–MS/MS gradient conditions

Time (min)	Flow rate (mL/min)	% of A	% of B	Curve
0	0.20	100	0	6
2	0.20	100	0	6
92	0.20	55	45	6
94	0.20	55	45	6
95	0.20	10	90	6
97	0.20	10	90	6
98	0.20	100	0	6
120	0.20	100	0	6

7. Database Search

데이터베이스는 UniProt 데이터베이스에서 각 어류의 콜라겐 항목을 선택하여 생성되었다. 데이터 검색은 PEAKS Studio X 를 사용했으며 효소 조건은 모든 염기서열 잔기에 분해가 일어날 수 있다 가정하여 'all amino acid'로 설정하였다. 펩타이드 필터로는 다음 PEAKS search 기준이 사용되었다. 데이터 리파인먼트는 min charge state 1 과 max charge state 3 로, error tolerance 는 전구체 20 ppm 과 fragment ion 0.5 Da 으로 설정했다. Digestion mode 는 unspecific 그리고 maximum missed cleavage per peptide 는 10 으로 세팅하였다. 콜라겐 post-translation modification (PTM)에 가장 많이 일어나는 두 가지 아미노산은 프롤린과 라이신이다. (Figure 1.) 그러므로, variable modification (max PTM per peptide: 3)은 프롤린과 라이신 잔기에 hydroxylation (15.99 Da), hydroxyallysine (14.96 Da)과 allysine (-1.03)이 포함되었다.

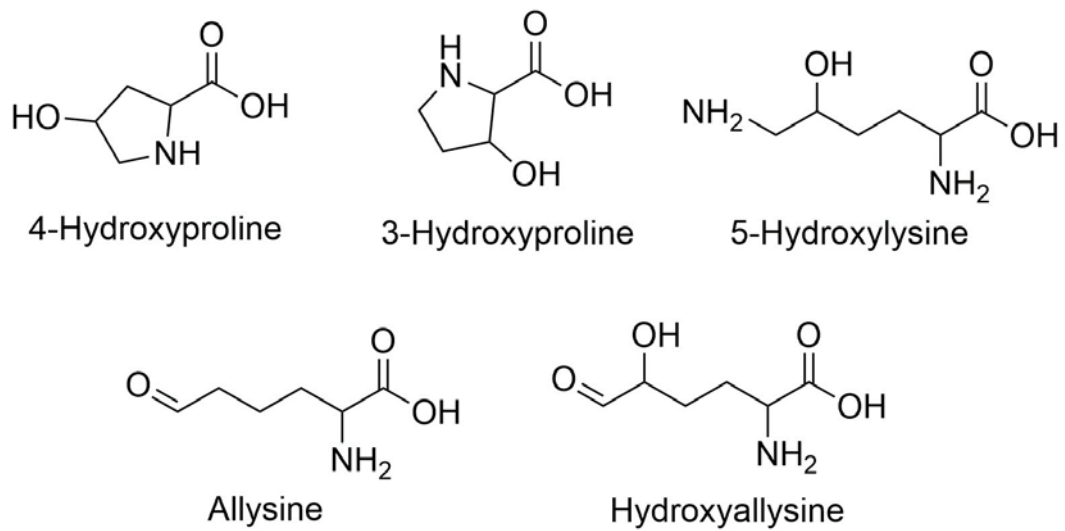


Figure 1. Collagen's most common PTMs ³⁰

Ⅲ. 결과 및 고찰

1. SEC-HPLC 분석

SEC-HPLC로 측정된 21가지 콜라겐 샘플(1-21)의 평균 분자량 분포도는 Figure 2. 나와 있다. Figure 2.에 제시된 분자량 분포는 모두 동일한 농도의 조건에서 측정된 것이다. 샘플이 컬럼에서 머무르는 시간이 길어질수록 분자량이 감소하며, 머무르는 시간은 약 10분에서 20분 사이에 존재했다. 특정 머무름 시간에 신호 강도가 클수록 그 특정 머무름 시간대에 해당되는 단백질 분자량의 함량이 더 많음을 나타낸다. 본 결과 값은 SEC로 220 nm 파장에서 찍은 콜라겐 RT 값들을 나타내었다. 거의 모든 콜라겐 단백질들이 6-13 min 사이에서 모두 용출되었으며 2-3개의 peak로 분리되어 용출되었다.

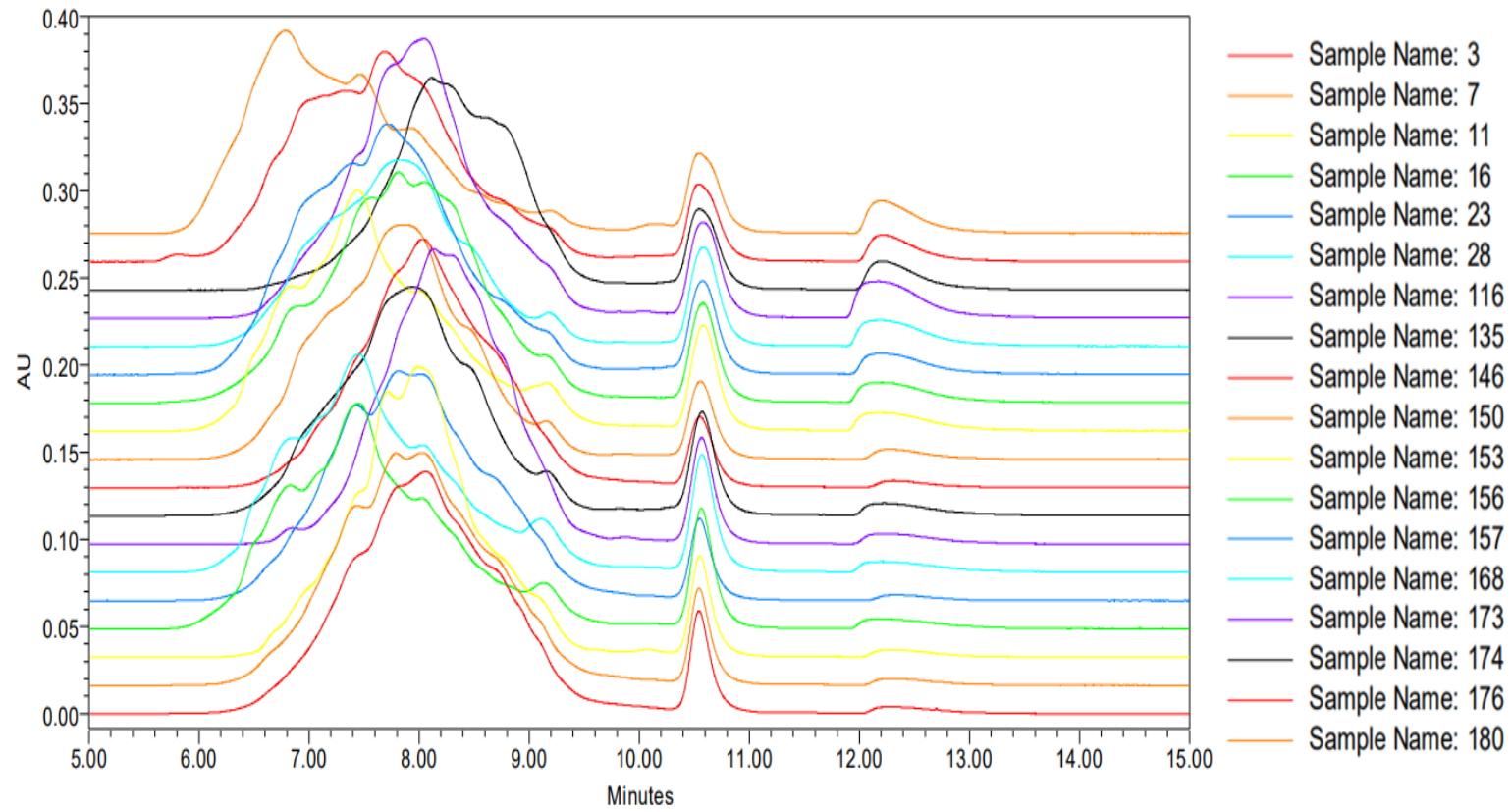


Figure 2. SEC molecular distribution of each collagen (220nm)

2. LC-MS/MS 특이서열 분석

1) 전처리 방법 연구

시료 준비 중에 생기는 많은 번역 후 변형 (post-translational modification, PTM)은 온도, 배양 시간, 효소의 종류 등 다양한 변수들의 반응에 따라 달라진다. 따라서 본 연구에서는 In-sol 효소분해 반응의 온도차와 분해되는 단백질 총량의 변화를 통해 적절한 전처리 방법을 평가했다. 효소가 단백질의 모든 영역에 접근하여 완전한 소화가 이루어지기 위해서는 먼저 단백질이 충분히 변성되어 펼쳐지고, 환원되는 것이 중요하다. 따라서, 이런 변성반응의 가속화를 위해 보통은 특정온도에서 변성반응을 수행한다. 본 실험은 In-sol 효소분해 반응에 쓰이는 Urea+DTT 변성 시약의 최적 온도를 찾기 위해 가장 많이 사용되는 프로토콜들을 기반으로 1시간동안 25℃와 37℃ 온도에서 변성반응을 비교했다. 모든 분해반응은 같은 100 μ g의 trastuzumab을 사용하였고 트립신의 총량은 5 μ g으로 고정해 16시간 가수분해했다. 각 25℃와 37℃에서 반응시켜 분석한 trastuzumab의 peptide mapping spectrum은 Figure 3. 과 같다. 25℃에서 변성반응을 거친 단백질은 대부분 분해가 되지 않아 TIC spectrum에서 분해된 peak들이 관찰되지 않았다. 25-30 머무름 시간(RT) 사이에 존재하는 peak들은 분해되지 않은 trastuzumab의 intact form으로 관찰되었다. 반면, 37℃ 항온기에서 변성반응을 진행한 trastuzumab은 다양한 펩타이드 형태로 분해가 되어 6-33 RT 사에서 다수의 peak들이 관찰되었다. 이 data를 기반으로 UNIFI 소프트웨어 내에 존재하는 trastuzumab 서열 library와 비교하여 매칭률을 확인해 보았다. (Figure 4.) 25℃의 샘플은 경쇄 78%와 중쇄 70% 커버리지로 총 73%의 항체서열 커버리지를 보였다. 다음 37℃ 샘플은 경쇄 100% 커버리지 그리고 중쇄 90% 커버리지를 보였으며 총

평균적으로 93%의 순수한 매칭률을 보여 주었다. 이를 통해 37°C에서의 변성 화원반응이 단백질의 분해에 더 적합한 온도였음을 확인할 수 있었다.

다음은 LC-MS/MS 분석을 위해 필요한 digestion 단백질 양의 최소값을 찾기 위해 10 *ug*, 25 *ug*, 50 *ug*, 그리고 100 *ug*의 trastuzumab을 가지고 트립신 효소반응을 진행하였다. Denaturation과 reduction반응은 37°C에서 진행하였고, 단백질과 트립신의 비율은 1:20으로 넣어 16시간 반응시켜 주었다. 마지막 질량분석기에 들어가는 펩타이드의 총량은 모두 5 *ug*으로 고정했다. Figure 5.는 각각 가수분해된 trastuzumab의 coverage를 나타내었다. 예상대로 단백질의 양이 많아질수록 실험 도중 단백질 펩타이드의 손실이 줄어들어 서열 검출량이 커졌다. 결론적으로, 90%이상의 coverage를 갖는 단백질 분해 실험을 재현하기 위해선 단백질이 100 *ug* 이상 필요하다.

가수분해된 콜라겐은 분자량 분포가 넓은 펩타이드로 구성되어 있다. 그리고 이런 펩타이드를 가지고 트립신 분해효소 반응을 진행하면 분자량이 더 작은 펩타이드로 분해된다. 이렇게 가수분해를 거친 콜라겐 혼합물과 추가로 효소반응을 거치지 않은 두가지 시료를 가지고 펩타이드 맵핑 분석을 했다. Figure 6.에 나와 있는 차트들은 비교분석에 이용된 콜라겐들의 *de novo* processing 데이터를 local confidence% 수치대로 나열한 결과값들이다. PEAKS 프로그램은 *de novo*서열의 각 아미노산에 대한 신뢰도 점수를 할당하는데, 점수의 범위는 0%에서 99%까지 표시된다. 그리고 *de novo* 전체 펩타이드 서열은 ALC(Average of Local Confidence) 점수로 평가된다. ALC는 서열에 포함된 모든 아미노산의 local confidence 점수의 평균값이다. 따라서, 전체적으로 tryptic digested 콜라겐 #16과 #135은 digestion을 하지 않은 콜라겐보다 식별된 아미노산의 개수도 적었지만 90%이상이 되는 local

confidence의 비율도 낮았다. Table 3. 는 *de novo* 분석된 펩타이드 서열 개수를 정리해 놓았다. 예상과 달리 추가적인 가수분해를 거친 저분자 콜라겐들은 전처리를 거치지 않은 콜라겐들보다 분석효율이 현저히 저해됐다. 이미 저분자의 형태를 띄고 있던 어류 콜라겐들이 다시 한번 가수분해가 되면서 훨씬 더 작은 크기의 분자량을 갖게 되었고, 트립신 처리를 한 후 SPE 정제를 거쳤을 때 펩타이드들과 고정상 레진과의 결합력이 저해되면서 더 많은 손실이 생겼을 것으로 예상된다. 고로, 본 연구에서 저분자 콜라겐의 분석은 부가적인 전처리를 거치지 않고 cellulose acetate(CA) syringe filter를 통해 간단한 정제과정을 거친 후 바로 분석되었다.

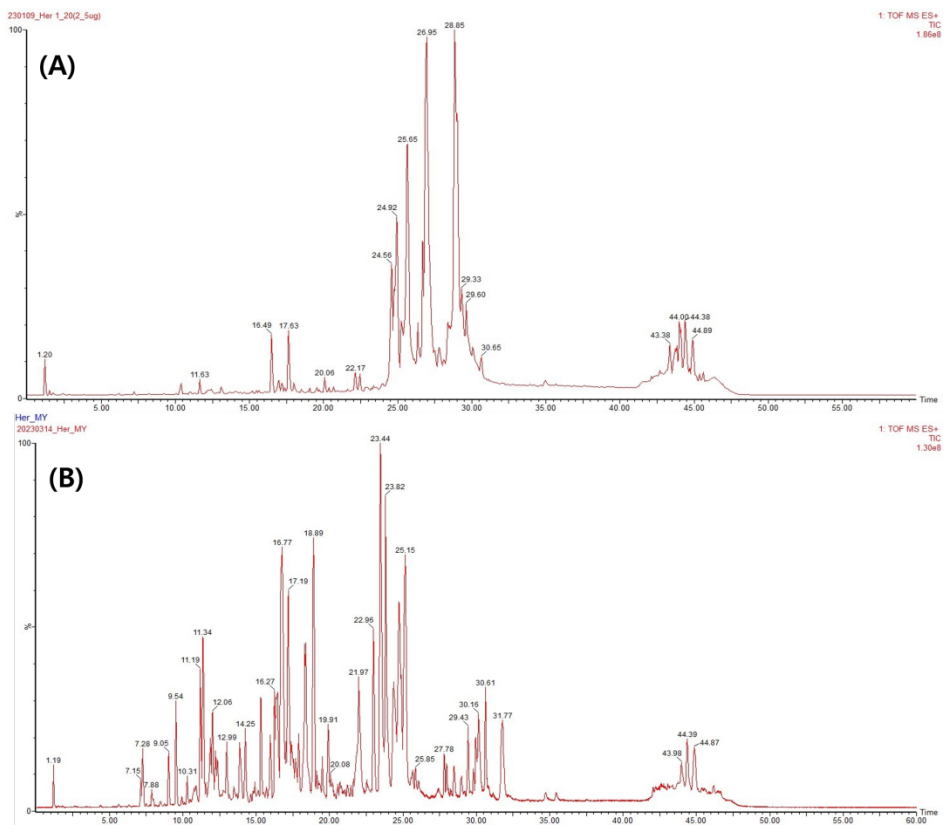


Figure 3. Pep-map of trastuzumab peptide with 25°C (a) and 37°C (b) denaturation temperature.



Figure 4. Total sequence coverage of trastuzumab with 25 °C (a) and 37 °C (b) denaturation temperature.

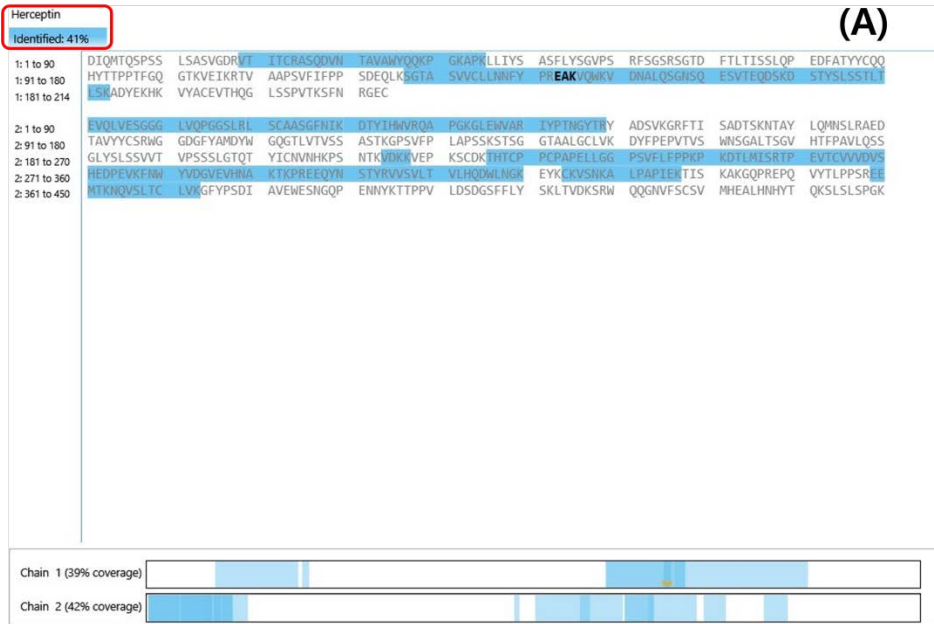




Figure 5. Total sequence coverage of trastuzumab with 10 μ g (a), 25 μ g (b), 50 μ g (c), and 100 μ g (d) digested proteins.

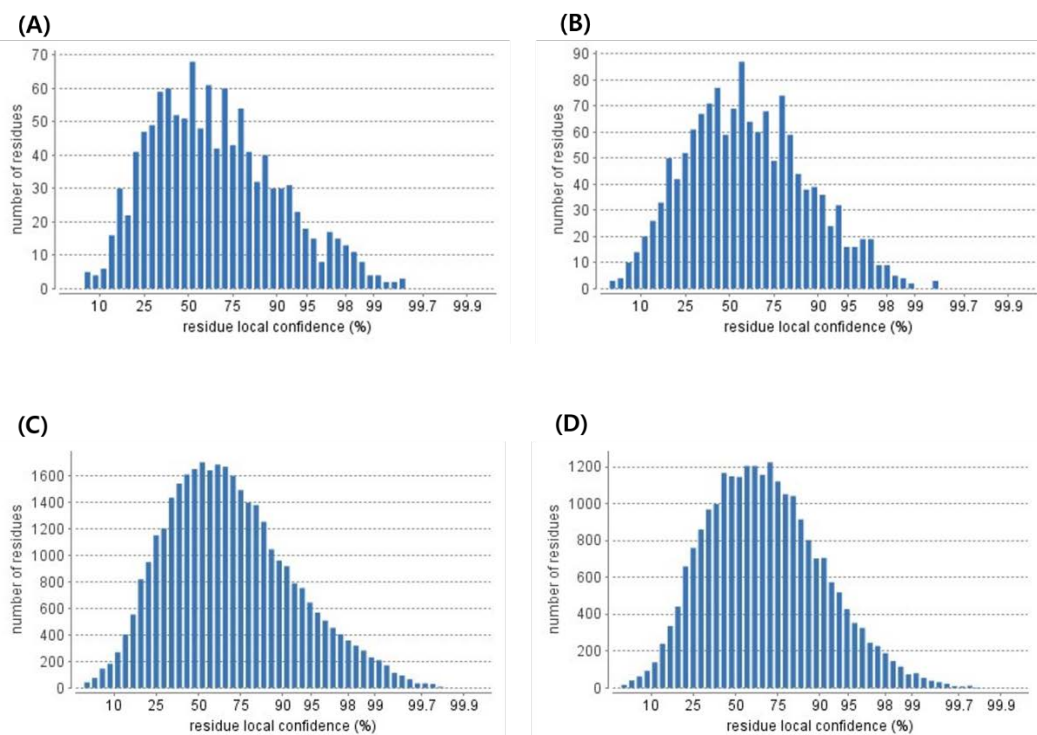


Figure 6. Table of residue local confidence% for tryptic digested #16 (a), #135(b) collagen and un-digested #16(c), #135(d) collagen.

Table 3. Total number of *de novo* peptide identified for #16 and #135 collagen with different pretreatment.

Collagen #	Pretreatment	# of <i>De novo</i> peptide
16	Trypsin	97
	-	3317
135	Trypsin	125
	-	2523

2) 컬럼 선택

펩타이드 분리를 위한 C18 고정상은 peak capacity (Pc) 을 개선하고 강력한 이온 페어링 에이전트에 대한 의존도를 낮춰 ESI-MS 와 더 잘 호환되는 것이 이상적이다. Pc 는 머무름 시간(RT) 내에서 분석물질들이 분리될 수 있는 피크의 최대 개수를 의미한다.³¹ 그래서 본 실험은 저분자 콜라겐 분석에 최적화된 컬럼을 찾기 위해 펩타이드 분리를 위한 charged-surface hybrid(CSH) C18 컬럼과 ethylene-bridged hybrid(BEH) C18 컬럼의 성능을 비교해 보았다. BEH C18 은 알킬 엔드 캡핑의 실리카 하이브리드 고정상으로 만들어져 있다. CSH C18 은 BEH C18 과 동일한 실리카 입자층 기반으로 만들어졌지만, 작은 전하량의 염기 작용기가 표면에 결합되어 있다.^{32,33} CSH 에 존재하는 표면 전하는 대략 5 pKa 로 보고된다. 따라서, CSH 입자는 이동상의 조성과 유형에 따라 산성 pH(<5)에서 양전하의 형태를 띈다. 그러므로 CSH C18 은 BEH C18 보다 향상된 로딩성능과 더 큰 Pc 를 보이는 것으로 알려져 있다. 추가적으로 CSH 는 FA 이온 페어링에 대한 의존도가 현저히 낮아 고감도가 요구되는 MS 애플리케이션에 이상적인 컬럼으로 보고되었다. 본 실험에선 각 컬럼의 peak capacity 를 극대화하기 위해 0.2% FA 를 포함한 이동상을 이용하였고, 2 시간의 gradient 로 #11, #23, #180 콜라겐을 비교 분석했다. 추가적으로, 펩타이드의 회수율과 피크 모양을 개선하기 위해 60°C 로 컬럼온도를 설정해주었다. Figure 7.와 Figure 8.를 통해 CSH C18 이 BEH C18 보다 우수한 피크 모양을 생성한 것을 확인할 수 있었다. 이는 CSH 의 peak 들이 용출되는 시간이 더 앞당겨져 peak 의 폭이 좁아졌으며, 더 대칭적인 모양의 peak 이 형성됐다. 그리고 Figure 8 과 같이 같은 질량의 펩타이드 peak 들을 비교해 보았을 때 각 peak 의 분리도도 CSH 가 더 우수하다는 것을 관측할 수 있었다. 마지막으로 Table 4.에 보이는 총

프로파일링 펩타이드 서열은 PEAKS software 를 통해 database searching 과 *De novo* searching 을 거친 결과값들이다. 이 모든 데이터 합쳤을 때 식별된 총 펩타이드 서열의 수는 BEH C18 분석물들이 더 많았으나, 데이터 품질의 척도인 *de novo* ALC% 값을 기준으로 상위 20%의 데이터들만 수집한 결과 CSH C18 로 분석한 샘플서열의 개수가 더 많았다. #11 은 126 개, #23 은 26 개, 그리고 #180 은 73 개의 서열 개수 차이로 CSH C18 컬럼의 분석효율이 더 뛰어나다는 것을 증명할 수 있었다.

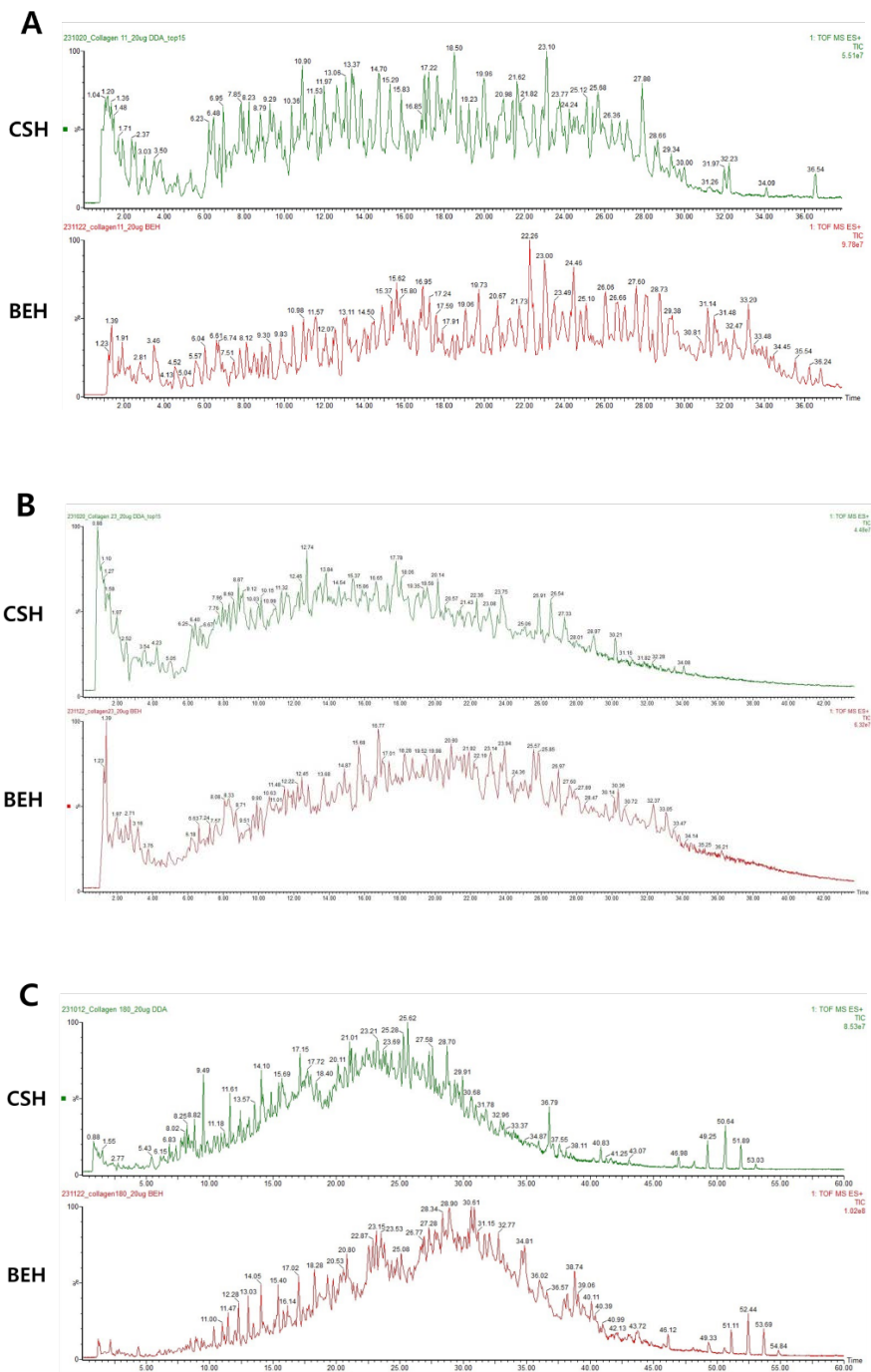


Figure 7. TIC comparison between CSH and BEH of #11 (a), #23(b), and #180(c) collagen.

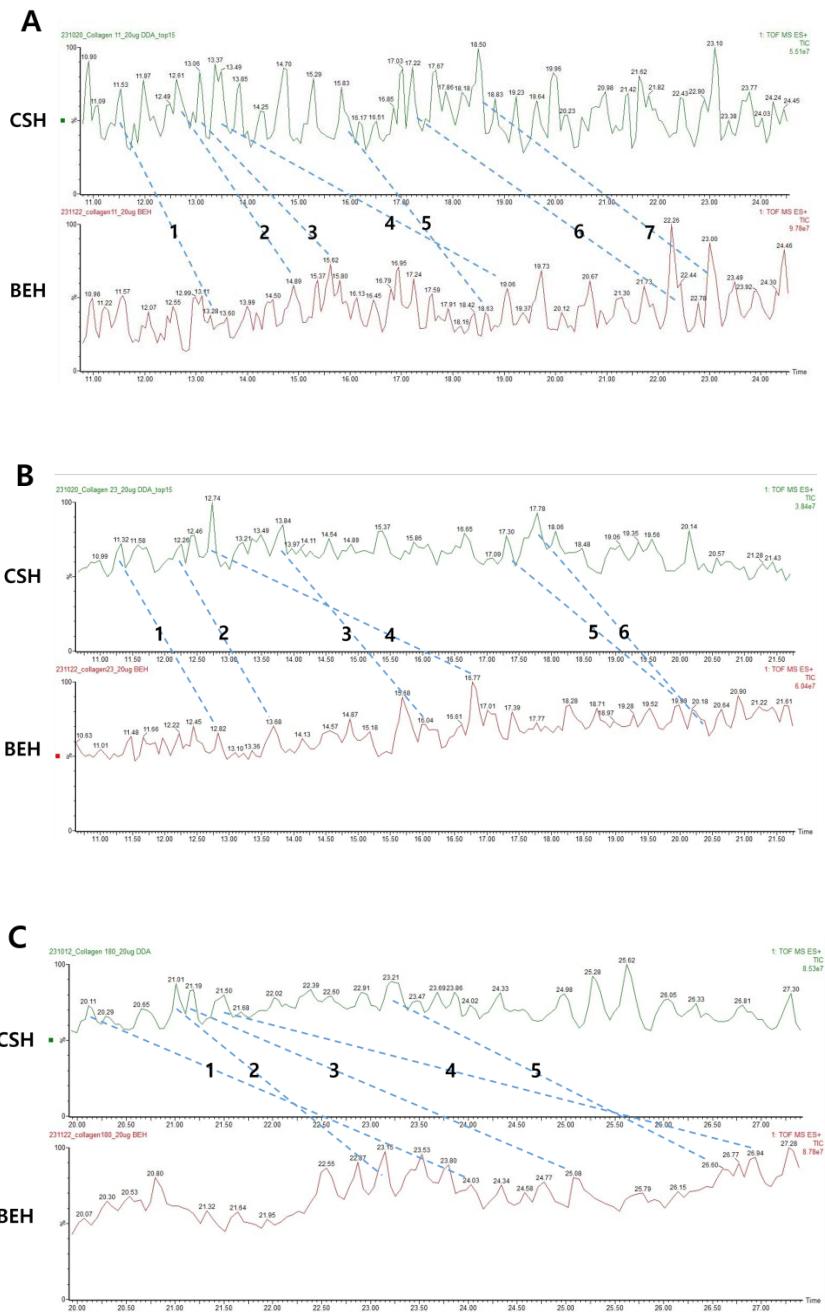


Figure 8. Corresponding RT peaks from each collagen peptide. Each line represent change in the elution time of each peptide. (#11(a), #23(b), #180(c))

Tabel 4. Number of total peptide ID from BEH and CSH column analysis.

Collagen #	Type of column	Total # of peptide ID	Total # of credible peptide ID	Total Denovo search (ALC % \geq 50)	Credible Denovo search (ALC % \geq 80)	Total PEAKS search
11	BEH	2344	232	2203	91	141
	CSH	2290	358	2227	295	63
23	BEH	3107	239	3000	132	107
	CSH	2432	265	2380	213	52
180	BEH	2377	288	2169	80	208
	CSH	2315	361	2153	199	162

3) 이동상 선택

최근 실리카겔 고정상과 함께 쓰이는 formic acid(pKa = 3.75)와 같은 약한 산화제는 ESI를 크게 억제하지 않으면서 효율적인 분리를 제공하기 때문에 펩타이드 분석을 위한 이동상 산성제의 대안으로 부상하고 있다.³⁴ Jadeja et al.³⁵에 따르면 최근 2022, 1~3월 사이에 작성된 논문들 중 99%의 논문이 이동상 산성제로 formic acid(FA)를 사용했으며, 이 중 91%는 0.1%의 FA를 최적의 농도로 간주했다. 이는 포름산이 낮은 농도에서도 펩티드와 강한 이온 쌍을 형성하는 능력이 있어 더 효율적인 이온화를 유도한다. 그리고 이런 이온화의 효율이 중요한 이유는 단일 전하를 띄는 전구체는 대부분 MS/MS 분석 과정에서 식별이 되지 않기 때문이다.³⁶ 만약 단일이온 전구체가 단편화가 된다면 fragment들의 다수는 전하가 없는 상태의 펩타이드로 조각화 되어, 이들은 2차 전자 복사 효과로 시그널이 증폭되는 electron multiplier detector에 검출되지 않게 된다. 그러므로, 전하 상태가 높은 전구체는 더 큰 단편화 효율을 나타낸다고 볼 수 있다. ESI에서 트립틱 펩타이드는 일반적으로 두 개 이상의 전하를 지니고 있어서, fragmentation이된 펩타이드 파편들은 최소 하나 이상의 양성자를 전달받을 수 있게 된다. 트립신은 라이신과 아르기닌 잔기의 C-말단 쪽에서 펩타이드 결합을 절단한다. 즉, 거의 모든 트립틱 펩타이드들의 단편은 C-말단에 아르기닌(R) 또는 라이신(K)이 존재하므로 염기성 특징을 지닌 아르기닌과 라이신의 결사슬이 다중 전하 이온화의 표적이 된다.³⁷ 하지만 저분자 콜라겐 가수분해물은 트립신으로 분해된 펩타이드의 (700-1500 Da) 크기보다 훨씬 더 작은 크기의 펩타이드로 존재할 수 있다.³⁸ 또한, 시중에 판매되고 있는 가수분해된 다수의 저분자 콜라겐 펩타이드들은 다양한 방법을 통해 가수분해를 진행하므로 이들의 서열내에 이온화가 진행될 수 있는 염기성 아미노산의 존재를 알 수 없다. 그러므로, 본 실험에서 식별된 대부분의 콜라겐

서열들은 단일전하를 띄고 있는 전구체 이온으로부터 유래되었다. 따라서 이런 단일이온의 전하량을 향상시키기 위해 본 실험은 0.1% FA의 기본 용매조건과 0.2% FA의 산성조건에서 비교 분석을 진행했다. 콜라겐 분석은 추가적인 전처리를 거치지 않은 상태로 CSH C18컬럼을 사용해 분석했다. 결과적으로, Figure 9.을 통해 0.2% FA를 이용한 분석물의 전체적인 스펙트럼의 intensity가 더 우수하다는 것을 확인 할 수 있었다. #180 콜라겐은 0.1%FA 조건에서 3.13×10^7 , 0.2%FA 조건에서 5.15×10^7 으로 약 1.65배의 intensity 차이를 보였고, #182 콜라겐도 0.1%FA 조건에서 3.11×10^7 , 0.2%FA 조건에서 5.23×10^7 intensity로 약 1.68배 가량 검출 강도가 상승했다. 이 외에, 0.1% FA 이동상에서 단일전하를 띄고 있었던 펩타이드 전구체들이 0.2% FA 이동상 조건에서 1+가 하나 더 붙은 2가 전하의 peak들로 관찰되었다. (Figure 10.) #180 콜라겐에서는 618.3066 m/z 값을 가지고 있던 단일전하 전구체가 0.2% FA 조건에서는 2가 전하의 309.6622 m/z peak으로 분리되어 관찰되었다. 그리고 #182 콜라겐은 879.4804 m/z의 단일전하 전구체가 440.2461 m/z로 분석되었다. 결과적으로, 이온화의 효율을 올려주는 0.2% FA 산화제를 이용하여 단일이온 콜라겐 전구체들의 형성을 저해하였다.

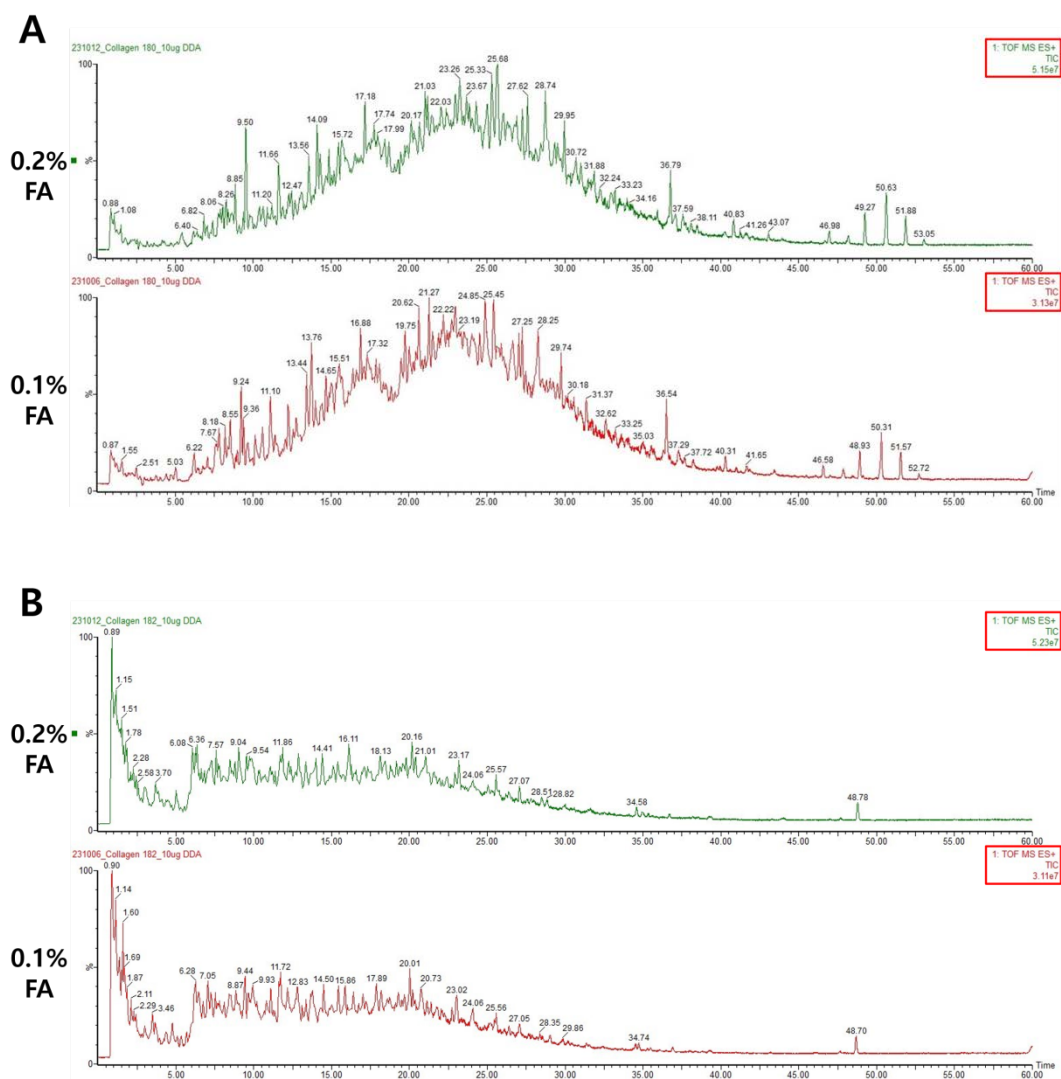


Figure 9. Pep-map TIC obtained from different mobile phase (0.1%FA, 0.2%FA included water and ACN) of #180(a) and #182(b) collagen.

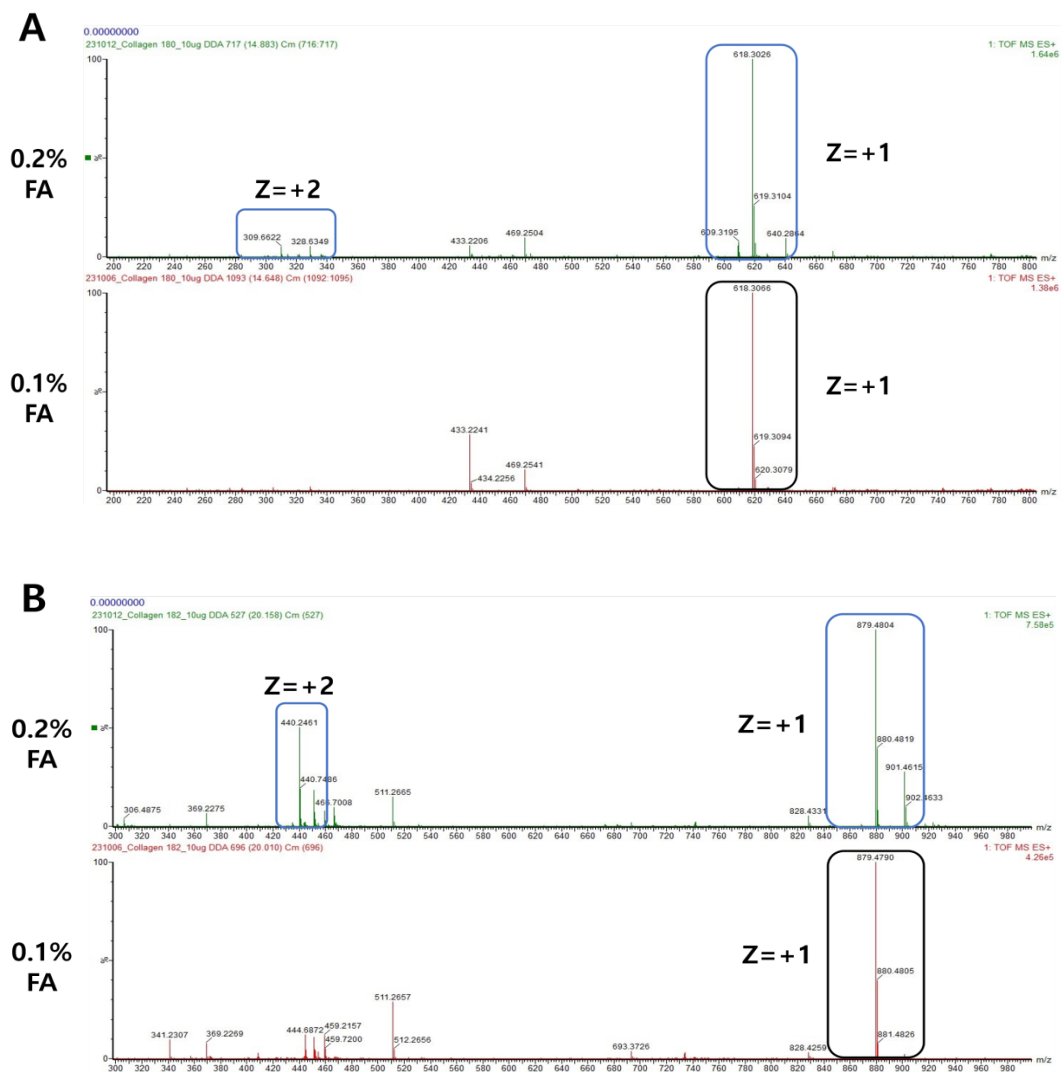


Figure 10. Pep-map MS spectrum obtained from different mobile phase (0.1%FA, 0.2%FA included water and ACN) of #180(a) and #182(b) collagen.

4) 어종의 특이서열

펩타이드 바이오마커는 한 종에 고유한 펩타이드 서열을 의미하며 다른 종과 구체적으로 구별할 수 있는 서열이어야 한다.³⁹ 펩타이드 마커의 선별은 크게 네 가지 단계로 나뉜다. 먼저, 고분해능 LC/MS를 사용하여 상위 15개 스펙트럼의 DDA 스캔을 수행했다. 이렇게 얻은 1단계 질량 분석 데이터를 PEAKS Studio로 가져왔다. 그리고 수집된 데이터에 대해 단백질 식별은 UniProt 데이터베이스와 *de novo* 시퀀싱을 사용하여 특성 분석을 진행했다. 이렇게 얻은 다양한 펩타이드들 중 5가지의 어종들 사이에서 서로 겹치는 비특이적인 펩타이드들은 제거하였다. MS/MS를 통해 분석된 수많은 결과값들 중 Database searching을 통해 완벽하게 식별된 펩타이드 서열들과 *de novo* sequencing을 통해 얻어낸 average local confidence score (ALC%)가 80 이상인 결과값들을 모아 이전보다 높은 정확도를 나타내는 펩타이드 서열 데이터들을 도출하였다. 그렇게 킬라피아는 총 400개, 대구는 141개, 잉어는 275개, 명태는 48개, 그리고 메기는 21개의 각각 신빙성이 높은 펩타이드 서열들을 검출하였다. 그리고 이들 중 각 어종의 서열들 사이에서 중복된 펩타이드 서열들을 제외하기 위해 각각의 데이터 값들을 비교해보았을 때 Table 5.과 같이 중복된 값들이 나타났다. 이때 잉어와 킬라피아 사이에서 다수의 중복서열들이 식별되었는데, 이는 킬라피아와 잉어가 같이 혼합된 콜라겐 샘플들이 존재하여 나타난 이상치 결과로 추정된다. 결론적으로, 각 어종들 사이에서 중복되는 서열들을 제외한 특이적 펩타이드 마커는 킬라피아 148개, 대구 136개, 명태 41개, 잉어 28개, 그리고 메기 15개가 식별되었다. (Figure 11.) Table 6.에서 보이는 서열들은 *de novo* search에서 90%이상 ALC값을 가지고 있는 데이터와 database search에서 $-10\log P$ 값이 20 이상인 ($P < 0.01$) 데이터들만 모아 정리한 것이다. $-10\log P$ 는 phred quality

score를 의미하며, 단백질 서열 분석에서 식별된 아미노산 단량체 데이터의 품질을 측정하는데 사용된다. Phred quality score가 20을 넘어가게 된다면 그 특정 아미노산 서열의 정확도는 99% 이상이다. 그렇게 해서 추려진 총 펩타이드 서열의 개수는 툴라피아 69개, 메기 7개, 대구 60개, 명태 17개, 그리고 잉어 9개이다.

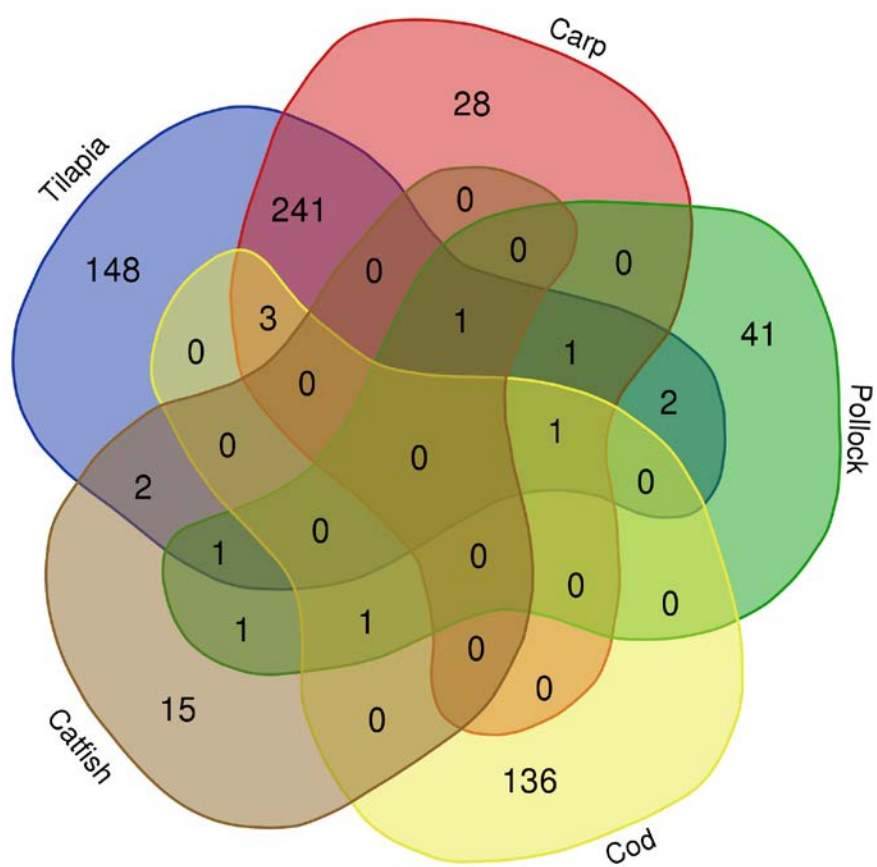


Figure 11. Venn diagram of all 5 types of fish collagen peptide

Table 5. List of overlapping peptide sequence

Fish Origin	Total overlapping peptide	Peptide sequence
Carp, Cod, Pollock, Tilapia	1	HPGL
Carp, Catfish, Pollock, Tilapia	1	RPP(+15.99)GPA
Carp, Pollock, Tilapia	1	ARPK
Carp, Cod, Tilapia	3	WGGAQ
		RGAGEK
		GGAWK
Catfish, Pollock, Tilapia	1	GPRL
Catfish, Cod, Pollock	1	RSGPA
Pollock, Tilapia	2	GDSRK
		KP(+15.99)GPAGPR
Catfish, Tilapia	2	LGPAGPR
		RPAGPA
Catfish, Pollock	1	WPGPK
Carp, Tilapia	241	VGAAAR
		RPAGPAGP
		AGPLGPTG
		LGAK(-1.03)R
		AGPSGPAGPR
		LGPAGPSGPAGPK
		Ect.

Table 6. Specific peptide sequence in each fish collagen (Tilapia, Catfish, Cod, Pollock, and Carp)

Fish Origin	Peptide	Mass	m/z	Length	PTM
Tilapia	AKGPLGPR	794.4763	398.2393	8	
	ARPAGPA	638.3500	639.3593	7	
	ASGPKPR	711.4027	356.7021	7	
	ATGETGKAGDK	1033.5040	517.7576	11	
	CP(+15.99)LLGL	743.4251	372.7133	7	Hydroxylation(PK)
	GPAGPRKPGPA	1003.5560	502.7786	11	
	GPKWPGPK	865.4810	433.7486	8	
	GWGAQ	517.2285	518.2427	5	
	HSK(+15.99)GY	606.2761	607.2915	5	Hydroxylation(PK)
	KGEGPAGLA	798.4235	400.2132	9	
	KVVGAPGP	721.4122	361.7067	8	
	KWAPGPA	725.3860	726.3973	7	
	LTGPR	542.3176	543.3305	5	
	NKGP	414.2227	415.2310	4	
	PGPPGAP	591.3016	592.3195	7	
	PGVVGAP	593.3173	594.3310	7	
	QALGPAGPT	810.4235	811.4150	9	
	RAGA	373.2073	374.2162	4	
	RAPAGPA	638.3500	639.3638	7	
	RGPAF	546.2914	547.3055	5	
	RGPP(+15.99)GAGP	723.3663	724.3816	8	Hydroxylation(PK)
	RGPP(+15.99)GPAG	723.3663	362.6932	8	Hydroxylation(PK)
	RLP(+15.99)GF	604.3333	605.3447	5	Hydroxylation(PK)
	RP(+15.99)GGR	557.3033	558.2998	5	Hydroxylation(PK)
	RP(+15.99)GGVGSAGPQ	997.4941	499.7631	11	Hydroxylation(PK)
	RPAGF	546.2914	547.3030	5	
	RPAGPAL	680.3969	681.4082	7	
	RPSGPAG	640.3292	641.3423	7	
	RSP(+15.99)GAT	603.2976	604.3113	6	Hydroxylation(PK)
	GPVGGPP(+15.99)GNLQN	1064.5250	1065.5100	11	Hydroxylation(PK)
	GEPGNVGPK	853.4293	427.7173	9	
	RGAPGPAGP	778.4086	779.4167	9	
	GVAGPQGPAG	809.4031	810.3864	10	
	TGPIGLPGPA	878.4861	440.2440	10	
	TGPKGETGTPG	1000.4830	501.2502	11	
	PAGSPGDKG	784.3715	785.3621	9	
	AP(+15.99)GIPGPAGPK	976.5341	489.2658	11	Hydroxylation(PK)
	AGPAGPAGAP	764.3817	765.3726	10	
	RGPAAGAPG	778.4086	779.4163	9	
	IGLPGPAGPQ	905.4970	453.7463	10	
	AVGPIGAPG	737.4072	738.3906	9	
	AAGAPGPMG	727.3323	728.3452	9	
	VGAPGPKGNNG	966.4883	967.4803	11	
	GAPGRPRP	863.4725	864.4824	9	
	GPRGPPPLPG	903.4926	904.4777	10	
	GPK(+15.99)GAIGPAGS	926.4821	464.2441	11	Hydroxylation(PK)
	GPK(+15.99)GAIGPAGS	926.4821	464.2441	11	Hydroxylation(PK)
	K(+15.99)JAGPAGAPG	740.3817	741.3668	9	Hydroxylation(PK)
	APGIP(+15.99)GPAGPK	976.5341	977.5204	11	Hydroxylation(PK)
	SWPGPVG	698.3387	699.3493	7	
	RGPAAGAPG	681.3558	682.3687	8	
	VGAIIGGPV	668.3857	335.1965	8	
	KGEKGEQGAP	999.4985	500.7474	10	
	RGAPGPAG	681.3558	682.3690	8	
	IGAPGPVGGPPG	917.4970	459.7596	11	
	GEPGKKGSK	886.4508	887.4384	9	
	GPRGLPGPPG	903.4926	904.4778	10	
<hr/>					
Fish Origin	Peptide	Mass	m/z	Length	PTM
Catfish	GALGPAGPR	794.4398	398.2188	9	
	IGPAGPR	666.3813	334.1887	7	
	LGPP(+15.99)GPP(+15.99)GPA	890.4497	891.4396	10	Hydroxylation(PK)
	P(+15.99)GPVRK	668.3969	335.2103	6	Hydroxylation(PK)
	QLGPSGPP(+15.99)GAP	992.4926	497.2498	11	Hydroxylation(PK)
	WLGAA	516.2697	517.2745	5	
WSGPQ	573.2547	574.2618	5		

Fish Origin	Peptide	Mass	m/z	Length	PTM
	ADGLL	487.2642	488.2657	5	
	AGLLGVGDE	829.4181	830.4210	9	
	AVALGERQA	913.4981	457.7464	9	
	CPGLGKGF	777.3843	778.3922	8	
	DGTL	404.1907	405.1926	4	
	EP(+15.99)SY	510.1962	511.1976	4	Hydroxylation (PK)
	ERGPAGPA	753.3769	754.3833	8	
	FLRMF	712.3730	357.1872	5	
	FSGLDKK	793.4333	397.7193	7	
	GAPGFGAAA	717.3445	718.3507	9	
	GFLGPKG	674.3751	338.1858	7	
	GGP(+15.99)GPQGPKGAAGAR	1292.6580	647.3347	15	Hydroxylation (PK)
	GLSW	461.2274	462.2331	4	
	GLTGLAGPA	755.4177	756.4073	9	
	GPAASPGPM	783.3585	784.3711	9	
	GPAGLRPGA	794.4398	398.2181	9	
	GPLKSGSP	711.3915	356.6962	8	
	GPSGPAGAAGA	811.3824	812.3915	11	
	GPVGKWQ	770.4075	386.2156	7	
	GTLWR	631.3442	316.6823	5	
	HPGLDGAK	793.4082	397.7155	8	
	K(-1.03)QGEHGAAGAK	1051.5050	526.7525	11	All lysine
	KP(+15.99)GFLQR	860.4868	431.2446	7	Hydroxylation (PK)
	LGPKPSVS	840.4705	421.2350	9	
	LGSDDAAGPK	871.4399	872.4307	10	
	LGTDDGAGGK	831.4086	416.7036	10	
	LGTDNAGGK	831.4086	416.7036	9	
	LKPLPGPM	851.4938	426.7478	8	
Cod	LPGAKGDSGPAGPK	1250.6620	626.3374	14	
	MKLL	518.3250	260.1699	4	
	MPGMK	562.2607	563.2631	5	
	NLGALGPK	768.4493	385.2252	8	
	NLGESGRVGVV	1085.583	543.7943	11	
	NYDEYVK	929.4130	465.7123	7	
	P(+15.99)AGGRGGPGS	827.3885	828.3854	10	Hydroxylation (PK)
	P(+15.99)ESYTDH	863.3297	432.6766	7	Hydroxylation (PK)
	PGKPLGSS	741.4021	371.7010	8	
	PSPGM PGDL	869.3953	870.4064	9	
	PVGPPLK	737.4799	369.7401	7	
	QCGPKG	588.2690	589.2744	6	
	QK(-1.03)GEHGAAGAK	1051.5050	526.7527	11	All lysine
	QLGFLGAR	860.4868	431.2423	8	
	RMGPN	573.2693	574.2789	5	
	RQGL	472.2758	473.2871	4	
	RS GAP(+15.99)GF	706.3398	707.3540	7	Hydroxylation (PK)
	RS GAP(+15.99)GFLQR	1103.5840	552.8091	10	Hydroxylation (PK)
	SGPKGSLGAR	928.5090	465.2548	10	
	TGPGK(-1.03)T	558.2650	559.2678	6	All lysine
	TLWR	574.3227	288.1694	4	
	TSIGPAG	601.3071	602.3076	7	
	TSMGM	525.1927	526.1978	5	
	VLGPVASSAAG	927.5025	928.4890	11	
	VSAGDDGAKGE	1004.4410	503.2246	11	
	VSAGDNGAKGE	1003.4570	502.7293	11	
	VSGALGFL	762.4276	763.4139	8	
	VSSGVK(+15.99)GE	777.3868	389.6943	8	Hydroxylation (PK)
	WGGAA	460.2070	461.2146	5	
	WPGVK	585.3275	586.3372	5	
	WSGEAGPKGD	1002.4406	502.2309	10	
	WSGVK	575.3067	576.3147	5	

Fish Origin	Peptide	Mass	m/z	Length	PTM
Pollock	ADINGAI	672.3442	673.3331	7	
	ADKICYL	824.4102	825.3905	7	
	AGNLAHAGA	780.3878	781.3777	9	
	GFIGIQQ	761.4072	381.7032	7	
	GPAW	429.2012	430.2067	4	
	IGIGAAV	599.3643	300.6840	7	
	K(+15.99)HGGGPA	638.3136	639.3263	7	Hydroxylation(PK)
	K(+15.99)PAATGA	630.3336	316.1707	7	Hydroxylation(PK)
	KPAATGA	614.3387	615.3363	7	
	LMIGGFGN	807.3949	808.3873	8	
	RVGGA	458.2601	459.2671	5	
	SSSSSLK	694.3497	695.3511	7	
	STDHQLK	827.4137	828.4041	7	
	VK(+15.99)IGAPI	712.4483	357.2252	7	Hydroxylation(PK)
	VK(+15.99)IGAPIN	826.4912	414.2469	8	Hydroxylation(PK)
	WRANAVNQP	1054.5310	528.2600	9	
	WVG PQ	585.2911	293.6542	5	
	ADINGAI	672.3442	673.3331	7	
	ADKICYL	824.4102	825.3905	7	
	AGNLAHAGA	780.3878	781.3777	9	
	GFIGIQQ	761.4072	381.7032	7	
	GPAW	429.2012	430.2067	4	
IGIGAAV	599.3643	300.6840	7		
K(+15.99)HGGGPA	638.3136	639.3263	7	Hydroxylation(PK)	
K(+15.99)PAATGA	630.3336	316.1707	7	Hydroxylation(PK)	
KPAATGA	614.3387	615.3363	7		
LMIGGFGN	807.3949	808.3873	8		
RVGGA	458.2601	459.2671	5		
SSSSSLK	694.3497	695.3511	7		
STDHQLK	827.4137	828.4041	7		
VK(+15.99)IGAPI	712.4483	357.2252	7	Hydroxylation(PK)	
VK(+15.99)IGAPIN	826.4912	414.2469	8	Hydroxylation(PK)	
WRANAVNQP	1054.5310	528.2600	9		
WVG PQ	585.2911	293.6542	5		

Fish Origin	Peptide	Mass	m/z	Length	PTM
Carp	AAGPAGNN	670.3034	671.3239	8	
	AAGPAVSVHI	920.5079	461.2584	10	
	DAPGNVGP	853.4293	427.7173	9	
	GALGPQGPVG	851.4501	426.7239	10	
	GHRGPIGQ	820.4304	411.2130	8	
	LGPP(+15.99)GPAGPK	962.5184	482.2581	11	Hydroxylation(PK)
	LGPP(+15.99)GPAGPK	905.4970	453.7480	10	Hydroxylation(PK)
	PGAAGPAGPK	821.4395	822.4301	10	
	TGDPGPVG	698.3235	699.3392	8	
	AAGPAGNN	670.3034	671.3239	8	
	AAGPAVSVHI	920.5079	461.2584	10	
	DAPGNVGP	853.4293	427.7173	9	
	GALGPQGPVG	851.4501	426.7239	10	
	GHRGPIGQ	820.4304	411.2130	8	
	LGPP(+15.99)GPAGPK	962.5184	482.2581	11	Hydroxylation(PK)
	LGPP(+15.99)GPAGPK	905.4970	453.7480	10	Hydroxylation(PK)
	PGAAGPAGPK	821.4395	822.4301	10	
TGDPGPVG	698.3235	699.3392	8		

IV. 결론

본 연구에서는 총 5 종의 저분자 어류 콜라겐을 다양한 전처리 과정과 분석 조건을 사용하여 HPLC 로 저분자 콜라겐의 평균 분자량과 LC-MS/MS 로 각 어종에 해당하는 펩타이드 바이오 마커를 분석하여 이에 따라 다음과 같은 결론을 얻었다.

1) 트립신 효소분해 전처리는 denaturation/reduction 단계에서 25℃의 조건보다 37℃에서 반응을 진행했을 때 촉진된 변성반응에 의해 전체적인 분해효율이 커졌다. 저온에서 가수분해된 trastuzumab 은 펩타이드 맵핑되어 total coverage 를 측정했을 때 73%의 매칭률이 관찰되었다. 반면 고온에서 반응한 trastuzumab 은 total coverage 가 93%로 전체적인 단백질 서열의 매칭률이 90%이상을 넘었다. 트립신 전처리 온도 외에 가수분해되는 단백질의 양에 따라 분석되는 결과값의 차이를 비교했다. 10 ug, 25 ug, 50 ug, 100 ug 의 trastuzumab 을 가지고 모두 같은 조건에서 효소분해반응을 실험하였다. 10 ug의 항체는 41%, 25 ug은 57%, 50 ug은 88%, 100 ug은 99%로, 단백질의 양이 많아 질수록 coverage 가 순차적으로 커졌다. 따라서 펩타이드 맵핑 전처리는 100 ug 이상의 단백질을 가지고 37℃에서 변성 반응을 진행하였다.

2) 저분자 콜라겐을 분석할 때 추가적인 가수분해 전처리를 거친 샘플들은 이미 500-2000 Da 이하의 질량을 갖고 있는 펩타이드들을 더 작게 분해하여 정제과정에서 단백질의 손실을 증가시켰다. 전체적인 회수율이 적어진 펩타이드 샘플로 분석한 결과 전처리를 거치지 않은 샘플들과 비교했을 때 약 20-34%가량의 데이터량이 차이 났다.

3) LC-MS/MS 분석시 CSH C18 컬럼은 BEH C18 컬럼보다 분리율이

뛰어났고 전체적인 저분자/저전하량 펩타이드에 대한 선택성이 BEH 컬럼보다 우수했다. BEH C18에서 분석된 펩타이드 서열의 개수는 CSH C18보다 더 많았지만 실제 데이터 품질을 비교해보았을 때 CSH C18의 데이터들이 더 높은 ALC%로 측정됐다.

4) 저분자 콜라겐의 이온화 효율을 극대화하기 위해 0.2% formic acid를 이동상 내에 존재하는 산화제로 사용해 주었다. 기존 0.1% FA를 사용했을 때보다 단일 전하로 존재하던 전구체들이 2+이상 향상된 이온화 전구체로 관찰되었다. 추가적으로 이온화 효율이 상승하면서 전체적인 스펙트럼의 intensity(검출량)도 1.65-1.68배 커졌다.

5) 위에 보이는 최적의 분석조건들을 적용하여 저분자 어류 콜라겐들을 분석한 결과, 90%이상의 ALC와 $10\log P$ 값이 20 이상인 ($P < 0.01$) 펩타이드 서열들로 추려봤을 때 틸라피아는 69개, 메기는 7개, 대구는 60개, 명태는 17개, 그리고 잉어는 9개의 특이서열이 존재했다.

참고문헌

1. Kadler KE, Baldock C, Bella J, Boot-Handford RP. Collagens at a glance. *J Cell Sci.* 2007;120(12):1955–1958.
2. Felician FF, Xia C, Qi W, Xu H. Collagen from marine biological sources and medical applications. *Chemistry & biodiversity.* 2018;15(5):e1700557.
3. Domene C, Jorgensen C, Abbasi SW. A perspective on structural and computational work on collagen. *Physical Chemistry Chemical Physics.* 2016;18(36):24802–24811.
4. Shoulders MD, Raines RT. Collagen structure and stability. *Annu Rev Biochem.* 2009;78:929–958. Accessed Oct 18, 2023. doi: 10.1146/annurev.biochem.77.032207.120833.
5. Hong H, Fan H, Chalamaiyah M, Wu J. Preparation of low-molecular-weight, collagen hydrolysates (peptides): Current progress, challenges, and future perspectives. *Food Chem.* 2019;301:125222. <https://www.sciencedirect.com/science/article/pii/S0308814619313305>. doi: 10.1016/j.foodchem.2019.125222.

6. Nugraha R, Rahayu F, Nurilmala M. Identification of peptide biomarkers for halal gelatin using bioinformatics techniques. *IOP Conf Ser : Earth Environ Sci.* 2022;1033(1):012063.
<https://dx.doi.org/10.1088/1755-1315/1033/1/012063>. Accessed Oct 12, 2023. doi: 10.1088/1755-1315/1033/1/012063.
7. Schmidt MM, Dornelles R, Mello RO, et al. Collagen extraction process. *International Food Research Journal.* 2016;23(3):913.
8. Jongjareonrak A, Benjakul S, Visessanguan W, Nagai T, Tanaka M. Isolation and characterisation of acid and pepsin-solubilised collagens from the skin of brownstripe red snapper (*lutjanus vitta*). *Food Chem.* 2005;93(3):475-484.
<https://www.sciencedirect.com/science/article/pii/S0308814604007666>.
doi: 10.1016/j.foodchem.2004.10.026.
9. Mahboob S. Isolation and characterization of collagen from fish waste material-skin, scales and fins of catla catla and cirrhinus mrigala. *Journal of food science and technology.* 2015;52:4296-4305.
10. Wang L, Yang B, Wang R, Du X. Extraction of pepsin-soluble collagen from grass carp (*ctenopharyngodon idella*) skin using an artificial neural network. *Food Chem.* 2008;111(3):683-686.

<https://www.sciencedirect.com/science/article/pii/S0308814608004901>.

doi: 10.1016/j.foodchem.2008.04.037.

11. Delgado LM, Shologu N, Fuller K, Zeugolis DI. Acetic acid and pepsin result in high yield, high purity and low macrophage response collagen for biomedical applications. *Biomedical Materials*. 2017;12(6):065009.

12. Jafari H, Lista A, Siekapen MM, et al. Fish collagen: Extraction, characterization, and applications for biomaterials engineering. *Polymers*. 2020;12(10):2230. <https://search.proquest.com/docview/2550255774>.
doi: 10.3390/polym12102230.

13. Nalinanon S, Benjakul S, Visessanguan W, Kishimura H. Use of pepsin for collagen extraction from the skin of bigeye snapper (*priacanthus tayenus*). *Food Chem*. 2007;104(2):593–601.
<https://www.sciencedirect.com/science/article/pii/S0308814606009770>.
doi: 10.1016/j.foodchem.2006.12.035.

14. Matmaroh K, Benjakul S, Prodpran T, Encarnacion AB, Kishimura H. Characteristics of acid soluble collagen and pepsin soluble collagen from scale of spotted golden goatfish (*parupeneus heptacanthus*). *Food Chem*. 2011;129(3):1179–1186.

15. Bai C, Wei Q, Ren X. Selective extraction of collagen peptides with high purity from cod skins by deep eutectic solvents. *ACS Sustainable Chem Eng.* 2017;5(8):7220–7227.
<https://doi.org/10.1021/acssuschemeng.7b01439>. Accessed Oct 13, 2023.
doi: 10.1021/acssuschemeng.7b01439.
16. Hong H, Chaplot S, Chalamaiah M, Roy BC, Bruce HL, Wu J. Removing cross-linked telopeptides enhances the production of low-molecular-weight collagen peptides from spent hens. *J Agric Food Chem.* 2017;65(34):7491–7499.
17. Marcet I, Álvarez C, Paredes B, Díaz M. The use of sub-critical water hydrolysis for the recovery of peptides and free amino acids from food processing wastes. review of sources and main parameters. *Waste Manage.* 2016;49:364–371.
<https://www.sciencedirect.com/science/article/pii/S0956053X16300095>.
doi: 10.1016/j.wasman.2016.01.009.
18. Yamamoto S, Deguchi K, Onuma M, Numata N, Sakai Y. Absorption and urinary excretion of peptides after collagen tripeptide ingestion in humans. *Biological & pharmaceutical bulletin.* 2016;39(3):428–434.

https://www.jstage.jst.go.jp/article/bpb/39/3/39_b15-00624/_article/-char/en. doi: 10.1248/bpb.b15-00624.

19. Karim AA, Bhat R. Gelatin alternatives for the food industry: Recent developments, challenges and prospects. *Trends Food Sci Technol*. 2008;19(12):644-656.

<https://www.sciencedirect.com/science/article/pii/S0924224408002252>. doi: 10.1016/j.tifs.2008.08.001.

20. Brooks C, Parr L, Smith JM, Buchanan D, Snioch D, Hebishy E. A review of food fraud and food authenticity across the food supply chain, with an examination of the impact of the COVID-19 pandemic and brexit on food industry. *Food Control*. 2021;130:108171.

<https://www.sciencedirect.com/science/article/pii/S0956713521003091>. doi: 10.1016/j.foodcont.2021.108171.

21. Tukiran NA, Ismail A, Mustafa S, Hamid M. Development of anti-peptide enzyme-linked immunosorbent assay for determination of gelatin in confectionery products. *Int J Food Sci Tech*. 2016;51(1):54-60.

22. Deng G, Guo S, Zaman F, Li T, Huang Y. Recent advances in animal origin identification of gelatin-based products using liquid

chromatography–mass spectrometry methods: A mini review. *Reviews in analytical chemistry*. 2020;39(1):260–271.

<http://www.degruyter.com/doi/10.1515/revac-2020-0121>. doi:

10.1515/revac-2020-0121.

23. Chuaychan S, Benjakul S, Kishimura H. Characteristics of acid– and pepsin–soluble collagens from scale of seabass (*Lates calcarifer*). *LWT– Food Science and Technology*. 2015;63(1):71–76.

24. Huang EC, Henion JD. LC/MS and LC/MS/MS determination of protein tryptic digests. *J Am Soc Mass Spectrom*. 1990;1(2):158–165.

25. O'Bryon I, Jenson SC, Merkley ED. Flying blind, or just flying under the radar? the underappreciated power of de novo methods of mass spectrometric peptide identification. *Protein Science*. 2020;29(9):1864–1878.

26. Elias JE, Gygi SP. Target–decoy search strategy for increased confidence in large–scale protein identifications by mass spectrometry. *Nature methods*. 2007;4(3):207–214.

27. Griss J. Spectral library searching in proteomics. *Proteomics*. 2016;16(5):729–740.

28. Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA. De novo peptide sequencing and identification with precision mass spectrometry. *Journal of proteome research*. 2007;6(1):114–123.
29. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem*. 2006;78(16):5678–5684.
30. van Huizen NA, Ijzermans JNM, Burgers PC, Luijckx TM. Collagen analysis with mass spectrometry. *Mass Spec Rev*. 2020;39(4):309–335. <https://doi.org/10.1002/mas.21600>. doi: 10.1002/mas.21600.
31. Grushka E. Chromatographic peak capacity and the factors influencing it. *Anal Chem*. 1970;42(11):1142–1147.
32. Wyndham KD, O'Gara JE, Walter TH, et al. Characterization and evaluation of C18 HPLC stationary phases based on ethyl-bridged hybrid organic/inorganic particles. *Anal Chem*. 2003;75(24):6781–6788.
33. Gritti F, Guiochon G. Effect of the pH and the ionic strength on overloaded band profiles of weak bases onto neutral and charged surface hybrid stationary phases in reversed-phase liquid chromatography. *Journal of Chromatography A*. 2013;1282:113–126.

<https://www.sciencedirect.com/science/article/pii/S0021967313001519>.

doi: 10.1016/j.chroma.2013.01.063.

34. Campuzano ID, Lippens JL. Ion mobility in the pharmaceutical industry: An established biophysical technique or still niche? *Curr Opin Chem Biol.* 2018;42:147–159.

<https://www.sciencedirect.com/science/article/pii/S1367593117301680>.

doi: 10.1016/j.cbpa.2017.11.008.

35. Jadeja S, Kupcik R, Fabrik I, Sklenářová H, Lenčo J. A stationary phase with a positively charged surface allows for minimizing formic acid concentration in the mobile phase, enhancing electrospray ionization in LC–MS proteomic experiments. *Analyst.* 2023.

36. Nesatyy VJ, Groh K, Nestler H, Suter MJ. On the acquisition of 1 charge states during high–throughput proteomics: Implications on reproducibility, number and confidence of protein identifications. *Journal of proteomics.* 2009;72(5):761–770.

37. Binossek ML, Schilling O. Enhanced identification of peptides lacking basic residues by LC- ESI- MS/MS analysis of singly charged peptides. *Proteomics.* 2012;12(9):1303–1309.

38. Hustoft HK, Malerod H, Wilson SR, Reubsaet L, Lundanes E, Greibrokk T. A critical review of trypsin digestion for LC–MS based proteomics. *Integrative Proteomics*. 2012;73.
39. Afzaal M, Saeed F, Hussain M, Shahid F, Siddeeg A, Al–Farga A. Proteomics as a promising biomarker in food authentication, quality and safety: A review. *Food Sci Nutr*. 2022;10(7):2333–2346.
<https://doi.org/10.1002/fsn3.2842>. doi: 10.1002/fsn3.2842.

ABSTRACT

Identifying important biomarkers in low-molecular weight fish collagen peptides through LC-MS/MS to differentiate the origins of collagen.

Minyoung Woo

Department of Next generation Applied Science

Graduate School of

Sungshin University

Collagen is the one of the most abundant proteins in mammals. Therefore, edible collagen available in the market is mainly produced from the skin of cattle or pigs. However, recently there has been growing interest in marine collagen that does not contain mammalian components due to concerns about disease transmission and religious issues. Among marine collagens, fish collagen has the highest biocompatibility with mammals, and therefore has been widely implemented in use compared to other marine organisms. Accordingly, in the collagen industry, there is a need to identify various species of collagen for product certification. Reliable species identification of collagen samples is crucial for detecting fraudulent mislabeling. To achieve this goal, ELISA and FTIR can be used for peptide authentication. However, for collagen, issues related to

false positive and negative results, inconsistencies, and low reactivity have been observed with ELISA and FTIR methods. Therefore, we developed a selective LC-MS methodology for collagen species detection and used SEC analysis to confirm the actual size of each collagen peptide. As a result, we identified peptide biomarkers for each of the five species of fish collagen.