



Peptidomic characterization of protease activity and biomarker discovery of COPD pathogenesis associated with HIV in clinical BALF samples



Monica E. Kruk¹, Danielle Weise¹, Sarah Samorodnitsky², Eric Lock², Subina Mehta³, Pratik Jagtap³, Christine Wendt^{1,4}, and Timothy J. Griffin³

¹Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Minnesota, Minneapolis, MN, USA
²Division of Statistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA
³Department of Biochemistry, Mol. Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA
⁴Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

Introduction

- Chronic obstructive pulmonary disease (COPD) and HIV can both significantly affect an individual's quality of life. Additionally, there is increased risk of COPD for people living with HIV (PWH) with 1 in 10 diagnosed.
- COPD is known to have increased inflammation and proteolytic activity within the lungs, which can result in tissue destruction and health decline.
- To uncover mechanisms behind COPD progression in PWH, we have quantified endogenous peptides using MS-based peptidomic analysis of bronchoalveolar lavage fluid (BALF) samples from PWH with or without COPD, and correlating findings to lung function test results (FEV1 percent predicted).
- Based on detected peptide sequences, we used customized software tools to assess potential protease activity enriched in these clinical samples.
- Our analysis has revealed promising new results on endogenous peptides within BALF samples and enriched protease activities that indicate possible mechanisms underlying COPD pathogenesis in PWH.

Methods

- Acquire intact and degraded protein information and quantitation with Fraggpipe.
 - Group 1 Proteins**
 - All proteins ≥ 3 kDa
 - Detected by aptamers
 - Group 2 Proteins**
 - Ultracentrifuged
 - Depleted
 - Detected by MS
 - Group 3 Proteins**
 - Ultracentrifuged
 - ≥ 3 kDa
 - Depleted
 - Detected by MS
 - Degraded Proteins**
 - Ultracentrifuged
 - ≤ 3 kDa
 - Detected by MS
- Detected intact proteins were filtered to functionally active proteases/peptidases according to MEROPS, a literature-based online database
- Added MEROPS information for each protease to Excel file.

Amino acid	P4	P3	P2	P1	P1'	P2'	P3'	P4'
Gly	28	46	19	5	29	99	59	31
Pro	44	4	67	4	0	2	17	11
Ala	57	58	39	11	174	69	36	49
Val	59	40	15	4	22	11	15	25
Leu	39	13	51	13	9	88	40	51
Ile	47	19	4	0	11	20	26	35
Met	8	19	9	0	4	12	7	3
Phe	3	3	19	2	1	11	23	4
Tyr	5	15	33	2	8	6	2	12
Trp	1	0	4	0	0	0	0	0
Ser	20	65	64	0	99	27	43	19
Thr	38	20	28	15	36	10	17	14
Cys	2	4	19	1	3	3	2	2
Asn	5	10	15	0	12	4	10	18
Gln	5	23	7	0	18	26	33	24
Asp	7	26	27	1	12	11	14	14
Glu	21	32	12	1	10	33	32	45
Lys	21	25	11	129	52	7	53	57
Arg	35	18	13	253	0	0	2	0
His	9	2	4	0	12	0	4	0

example of protease summary from MEROPS
- Used z-scores and custom Excel functions to assign proteases to cleavage sites surrounding detected cleaved proteins' peptide(s).

$z = \frac{\text{observed value} - \text{mean average}}{\text{standard deviation}}$
- Performed statistical analyses to determine correlation to FEV1pp.

Results

Proteases

Protease Family	Protein Count
Matrix Metalloproteinase	13
Cathepsin	11
Kallikrein	11
Aminopeptidase	10
Caspase	8
Carboxypeptidase	7
Disintegrin-Metalloproteinase	9
Dipeptidyl Peptidase	4
Granzyme	4
Ubiquitin	4
Calpain	3
Elastase	3
Coagulation factor	3
Proteasome	3
Transmembrane Serine Protease	3

Figure 1. Proteins identified as proteases according to the MEROPS database. **A)** Top 15 protease families with highest counts of members detected. **B)** Counts of proteases detected in different parts of a bronchoalveolar lavage fluid sample.

Figure 2. Pathways and connection between proteases. **i)** Top pathways of proteases using Reactome.org. **ii)** Network connectivity of all proteases detected and mapped using the STRING-database.

Substrates

Protein Description	Peptide Count
Glyceraldehyde-3-phosphate dehydrogenase	140
Alpha-enolase	136
Vimentin	114
Tubulin beta-4B chain	91
Pyruvate kinase PKM	90
Myosin-9	87
Tubulin alpha-1A chain	69
Aldehyde dehydrogenase	61
Annexin A1	53
Ceruloplasmin	52
Keratin, type II cytoskeletal 8	51
Histone H1.5	51
Histone H2A type 2-C	50
Fructose-bisphosphate aldolase A	48
Napsin-A	47

Figure 3. Detection of peptide fragments and cleavage sites in proteins found to be cleaved with significant correlation to lung function (FEV1pp). **A)** Top 15 proteins degraded from proteolytic activity. **B)** Illustration of analyzed parts of protein. **C)** Example chromatograms of detected peptides with similar sequences.

Figure 4. Pathways and connection between protein substrates significantly correlated to FEV1pp. **i)** Top pathways identified for substrates using Reactome.org.

Discussion and Conclusion

- The PWH with COPD had a higher average amount of peptide and number of unique peptides detected compared to PWH without COPD. These results are consistent with increased protease activity in COPD.
- 8,334 proteins were detected from the 3 methods, and of those, 138 were proteases based on MEROPS characterization.
- MEROPS protease summaries were used to calculate z-scores of cleavage sites using a sequence window of at least 5 amino acids. A higher average z-score indicates a higher probability recognition as a substrate by a specific protease.
- 11,063 unique peptide sequences were detected, and 2,799 had a significant z-score calculated and protease associations assigned. 31 proteins were found significantly correlated to FEV1pp.
- Matrix metalloproteases, cathepsins, kallikreins, caspases, and exopeptidases were concluded to play major roles in specific cleaving of immune-related proteins and structural tissues.

Future Directions

- Complete analysis of protease-substrate correlation.
- Investigate other protease - substrate interaction predictor algorithms and compare.
- Determine candidate quantifiable peptides for activity of proteases.

Acknowledgements

We would like to thank:

- The National Institute of Health (NIH) for funding for this work (grant R01HL140971)
- LeeAnn Higgins and Kevin Murray at the Center for Metabolomics and Proteomics at the University of Minnesota for acquiring the mass spectrometry data
- The SomaLogic company in Colorado for performing the SomaScan aptamer-based protein detection
- Neil D. Rawlings and the EMBL-European Bioinformatics Institute for the MEROPS database
- The Nesvizhskii Lab's Fraggpipe and Thermo Fisher Scientific's Proteome Discoverer software for PSM matching and protein identification
- SIB, GPR, and EMBL for the STRING-db.org database
- The Global Core Biodata Resource for Reactome.org
- Tim Hulslen and DeepVenn.com for Venn diagrams
- Flicon.com artists (KP Arts, Icongeek26, vectorsmarket15, and Freepik) for images

ii) Top Pathways

Pathway	Proteins
Activation of Matrix Metalloproteinases	21 / 35
Degradation of the extracellular matrix	42 / 148
Collagen degradation	21 / 69
Extracellular matrix organization	43 / 328
Metabolism of Angiotensinogen to Angiotensins	11 / 27
Innate Immune System	40 / 1,340
Neutrophil degranulation	24 / 478
Peptide hormone metabolism	14 / 131
Assembly of collagen fibrils and other multimeric structures	11 / 67
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	12 / 127
Collagen formation	11 / 104
Intrinsic Pathway of Fibrin Clot Formation	7 / 26
Formation of Fibrin Clot (Clotting Cascade)	8 / 43
Trafficking and processing of endosomal TLR	6 / 16
Metabolism of proteins	47 / 2,214

Predicted Protease-Substrate Interactions

Protein Description	Peptide Sequence	Cleave	Protease	z-score
Collagen alpha-2(I) chain	QSLQEETVRKGPAG	GPAG DRGP	MMP13	2.77
Tubulin beta chain	AILVDLEPGTMDSVR	YVPR AILV	GZMK	2.67
Histone H1.4	AAAGYDVEK	KKAL AAAG	GZMM	2.20
Histone H2A type 2-C	AGLQFPVGR	RSSR AGLQ	GZMA	2.06
Ezrin	LRAKEELERQAV	RMAA LRAK	MMP3	1.95
Actin, cytoplasmic 1	APEEHPVLL	PVLL TEAP	CTS5	1.80
Kinogen-1	SSRIGEIKEETT	SPFR SSRI	KLK1	1.81
Phosphoglycerate kinase 1	SLEPVAVELK	VELK SLLG	ELANE	1.74
Vimentin	NLRETNLD	TNLD SLPL	CASP6	1.61
Macrophage-capping protein	QVLGPKPALK	PALK EGNP	CTS5	1.57
Ubiquitin-60S ribosomal protein L40	IIEPSLR	LRGG IIEP	USP4	3.84
Calreticulin	KSGTIFDNF	FDNF LITN	CTSA	2.99
Collagen alpha-2(I) chain	QSLQEETVRKGPAG	GPAG DRGP	MMP13	2.77
Prothrombin	ESYIDGR	IDGR IVEG	F10	2.58
Bactericidal permeability-increasing protein	NVVLQPHQNF	HQNF LLFG	CTSA	2.58
Alpha-1-antitrypsin	LSGVTEEAPLK	APLK LSKA	CPA4	2.55
Deleted in malignant brain tumors 1 protein (12)	DVGSYQEKVDV	RSKR DVGS	FURIN	2.51
Apolipoprotein A-IV	EKESQDKTL	STFK EKES	ATG4B	2.50
Natural resistance-associated macrophage protein 1	MTGDGKQPQR	GPQR LSGS	MMP8	2.30
Putative neutrophil cytosol factor 1B	GDTFIRH	---M GDTF	METAP1D	2.25

Table 1. Examples of detected peptides with assigned protease-substrate prediction z-scores calculated in Excel using custom functions and published substrate sequences found in the MEROPS online database.

ii) Top Pathways

Pathway	Proteins
Aggregophagy	4 / 47
Chaperone Mediated Autophagy	3 / 23
Autophagy	5 / 166
Selective autophagy	4 / 89
Amlyoid fiber formation	4 / 89
Neutrophil degranulation	7 / 478
RHO GTPase Effectors	6 / 325
HCMV Early Events	5 / 215
Glycolysis	4 / 112
Apoptotic execution phase	3 / 54
Glucose metabolism	4 / 143
HCMV Infection	5 / 270
Macroautophagy	4 / 150
Caspase-mediated cleavage of cytoskeletal proteins	2 / 12
HDACs deacetylate histones	3 / 63
Innate Immune System	10 / 1,340
DNA Damage/Telomere Stress Induced Senescence	3 / 70
Immune System	14 / 2,624