UNIVERSITY OF CALIFORNIA SAN DIEGO

Detection of proteases in acidified breast milk and evaluation of the effect of pasteurizing on

their catalytic activity

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Stephanie Luedtke

Committee in charge:

      Professor Anthony O'Donoghue, Chair
      Professor Eric Schmelz, Co-Chair
      Professor Brenda Bloodgood

2022

The Thesis of Stephanie Luedtke is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I'd like to dedicate this document to those who have supported me in everything I do.

To my loved ones, friends, and family alike.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

.

VITA

2021   Bachelor of Science in Human Biology, University of California San Diego

2022   Master of Science in Biology, University of California San Diego

PUBLICATIONS

Johansen-Leete, J., Ullrich, S., Fry, S. E., Frkic, R., Bedding, M. J., Aggarwal, A., Ashhurst, A. S., Ekanayake, K. B., Mahawaththa, M. C., Sasi, V. M., Luedtke, S., Ford, D. J., O'Donoghue, A. J., Passioura, T., Larance, M., Otting, G., Turville, S., Jackson, C. J., Nitsche, C., & Payne, R. J. (2022). Antiviral cyclic peptides targeting the main protease of SARS-COV-2. *Chemical Science*, *13*(13), 3826–3836. https://doi.org/10.1039/d1sc06750h

Luskin, K., Mortazavi, D., Bai-Tong, S., Bertrand, K., Chambers, C., Schulkers-Escalante, K., Luedtke, S., ghassemian, M., Odonoghue, A., Geng, B., Leibel, S., & Leibel, S. (2021). Allergen shedding in human milk from mothers of preterm infants: Proteomic and peptidomic feasibility and pilot analysis. *Research Square*. https://doi.org/10.21203/rs.3.rs-1120623/v1

Luedtke, S., Bojo, C., Li, Y., Luna, E., Pomar, B., & Radić, Z. (2021). Backbone conformation shifts in X-ray structures of human acetylcholinesterase upon covalent organophosphate inhibition. *Crystals*, *11*(11), 1270. https://doi.org/10.3390/cryst11111270

Radić, Z., Luedtke, S., Bojo, C., Li, Y., Luna, E., & Pomar, B. (2020). Shifts in Cα backbone conformation in X-ray structures of human acetylcholinesterase covalently inhibited by organophosphorates and organophosphoramidates revealed by Pacct3 Comparative Analysis. *Proceedings of The 2nd International Online Conference on Crystals*. https://doi.org/10.3390/iocc_2020-07339

ABSTRACT OF THE THESIS

Detection of proteases in acidified breast milk and evaluation of the effect of pasteurizing on

their catalytic activity

by

Stephanie Luedtke

Master of Science in Biology

University of California San Diego, 2022

Professor Anthony O'Donoghue, Chair
Professor Eric Schmelz, Co-Chair

Meeting the nutritional demand for high-risk infants is challenging and problems such as necrotizing enterocolitis, growth faltering, and neurodevelopmental impairment persist in neonatal intensive care. The use of pasteurized donor human milk is increasing for high-risk infants particularly when mother's milk is insufficient, or breastfeeding is not possible. Proteins

are a major source of nutrients in milk and are degraded into peptides and subsequently into nutritious amino acids by proteases. Proteomics studies have identified at least 38 different proteases in human milk, but little research has been performed to biochemically characterize these enzymes.

We utilized our mass spectrometry-based peptide digestion assay to detect and characterize the active proteases in human milk. We first acidified the milk to <pH 3.0 to mimic the pH of the infant stomach. We then characterized the protease activity in these acidified samples and discovered two distinct enzymes. One protease removes tripeptides from the N-terminus of our synthetic peptide library and is therefore a triaminopeptidase. We developed an internally quenched fluorogenic substrate that is efficiently cleaved by this enzyme and a tripeptide-chloromethylketone inhibitor that irreversible inactivates the enzyme. In addition, we discovered an endopeptidase that was sensitive to pepstatin and is likely to be the aspartyl protease, cathepsin D.

In the US, milk banks must pasteurize donated milk before distributing it to infants in need, such as those in NICUs. However, the effect of heat pasteurization (62.5°C for 30 min) on enzymatic activity of specific proteases has not been studied. We determined that the two proteases are inactivated by pasteurization. However, we added 10.4 nM of recombinant human cathepsin D into the pasteurized milk to show that the aspartyl protease activity could be recovered to normal activity. We are currently trying to identify the triaminopeptidase and plan to spike this enzyme into pasteurized human milk too in order to recover activity. Our long-term goal is to develop a protease-rich, human milk supplement that is added to pasteurized milk.

CHAPTER 1

INTRODUCTION

      Preterm birth (<37 weeks of gestation) has been one of the leading causes of infant mortality and physiological disabilities [1]. With an increase in preterm deaths in 2016 across all socioeconomic classes [2], the concern for sufficient care is brought to the forefront of nutrition innovation. Sufficient nutrients for preterm infants, who are considered high-risk, are essential to increase body weight for proper development and to reduce mortality [4]. The preterm infant mothers' milk (MoM) is considered the ideal source of nutrition because of the nutritional properties cater to their infants needs [5,6]. Human breast milk has also been shown to reduce the incidence of common physiological issues in neonatal intensive care, such as gastrointestinal tract infections, necrotizing enterocolitis (NEC), sepsis, chronic lung disease, and neurodevelopmental delays [5,7,8]. However, when MoM is unavailable, as is the case when the infant is admitted into an intensive care unit or if the mother is physiologically unable to produce milk, feeding alternatives are necessary [7].

      Donor breast milk (DBM) as an alternative to MoM, is a viable soluble to infants in the NICU. DBM is collected from mothers who have an excess production of milk or who has lost an infant themselves. Mother's breast milk and donor's breast milk are both shown to decrease the risk of necrotizing enterocolitis when compared to primarily formula-fed premature infants [5,7,8]. However, infants who are primarily fed DBM have shown to have the need to be supplements with bovine based in order to make their weight gain comparable to infants who are fed MOM [7]. When looking at how DBM is processed, samples are pooled, treated, tested, and distributed [9]. To be

1

considered safe for consumption, DBM needs to under processed and treated with heat [9]. The Holder Pasteurization method (HoP) is a high-temperature treatment protocol for DBM that is recommended by the Human Milk Banking Association of North America (HMBANA) and European Milk Bank Association (EMBA) for the inactivation of harmful microbial agents such as viruses and bacteria [9]. A significant decrease in the bioactive nutritional content of donor breast milk, including enzymes such as amylase and lipase, is observed when compared to mother's breast milk due to pasteurization [7,10]. There have been several studies concerning the detrimental effects of pasteurization on different macronutrients that make up the composition of the milk, such as bioactive proteins and enzymes previously stated before [11,12]. This decrease in macronutrients in pasteurized donor breast milk has been considered to affect the effectiveness of donor breast milk, which can be supported by the slower growth of DBM-fed infants as mentioned above [7].

Proteases are a class of enzymes that breakdown proteins into peptides and subsequently into amino acids. Many peptides are bioactive and can function as antimicrobials, antioxidants, or immunomodulators [13]. In addition, amino acids are important nutrients for the infant [14,15]. When in the breast, human milk is kept at a physiologically neutral pH (~pH 7.4) [17]. Proteases that have shown to be active in human milk under these conditions are plasmin, carboxypeptidase B2, kallikrein, elastase, thrombin, and cytosol aminopeptidase, and additional exopeptidases and endopeptidases [16-20]. When the infant consumes MOM, the milk is exposed to a stomach pH that ranges from pH 2[19,21]. This allows proteases that were inactive milk in at neutral pH, such as procathepsin D, to become activated to cathepsin D through conformational shifts catalyzed with acid exposure. Cathepsin D is the most prominent protease acid-acting protease in MoM [20]. The loss of this protease activity in milk that has been pasteurized could be detrimental to the

2

gastrointestinal health of an infant [14]. Studies have shown that pasteurization of MoM alters the peptide composition in milk and therefore likely affects the proteases that are outlined above [16-18,22]. However, the effect that pasteurization has on protease kinetics has not yet been explored to the same extent, especially when pertaining to acidified milk that is, physiologically significant as it mimics the pH conditions in the infant stomach. By analyzing the protease content in MoM before and after HoP treatment at pH 3.5, we can begin to discover the effects that treatment has on milk protease activity. Therefore, we can begin to create ways to supplement DBM with proteases that are inactivated due to HoP.

Taking into consideration that activity of enzymes such as lipase and amylase are abolished in human breast milk because of pasteurization, we predict that proteolytic activity is also decreased. We propose that specific proteases, active at pH 3.5 are negatively affected by HoP due to their inherent susceptibility to high temperature. Throughout the course of this project, we aim to identify the proteases that are active in acidified donor human breast milk and determine if they are affected by pasteurization. These findings will be important as it will reveal if unpasteurized MoM milk has superior enzymatic capacity when compared to pasteurized DBM.

In this study, we used an in-house peptide digestion assay to quantify protease activity in milk. This assay is term Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS) and uses a proprietary, physicochemically diverse library of 228 14-mer synthetic peptide substrates for direct monitoring of protease activity through liquid chromatography tandem mass spectrometry (LC-MS/MS) sequencing of peptide cleavage products. The ability of the MSP-MS assay to accurately determine protease substrate specificity for complex mixtures of proteases, has been validated in plasma, cyst fluid, cell extracts and secretory vesicles [23,24,25]. As a result of identifying specific cleavage locations along a peptide chain using MSP-MS methods, we were able to design

that are efficiently cleaved by the target protease [23]. We have detected two acid acting proteases that are inactivated by the HoP process, cathepsin D and a triaminopeptidase. One enzyme is sensitive to pepstatin and therefore we anticipate that this enzyme is, cathepsin D. The second enzyme is a tri-aminopeptidase but we have not identified the enzyme. By adding recombinant cathepsin D into pasteurized milk, we are able to regain the proteolytic activity back to the levels of cathepsin D present in unpasteurized milk. Our long-term goal is to understand the role of acid-acting proteases in milk protein digestion and to develop a protease-rich supplement to aid protein digestion in infants.

METHODS

Human milk samples

Deidentified human breast milk samples were provided by Dr. Maryanne Tigchelaar Perrin from the University of North Carolina, Greensboro. In total, 39 samples of 1.5 mL were shipped to UC San Diego on dry ice. Samples were aliquoted into 50 µL volumes and stored at stored at 80°C. For the studies, samples were thawed no more than two times. The samples before analysis are referred to as whole milk, which contains lipids. To generate skim milk, we first thaw the milk samples and pool them. We centrifuged the pooled samples at 10000xg for 10 minutes at room temperature, then removed the lipid layer that is isolated at the top as a result of centrifugation. We centrifuged the pooled sample again, remove the opaque supernatant and place into a new eppendorf tube for analysis.

For MSP-MS assays, 10 whole milk samples were thawed, pooled at equal volume and then, split into two tubes. One tube was subjected to the Holder Pasteurization protocol as outlined below while the second tube was a control (unpasteurized). Both samples were then centrifuged at 10,000 x g for 10 min at 24°C. After centrifugation, the top thick lipid layer was removed, and the

remaining supernatant was then transferred into a new 0.6 mL Eppendorf tube and centrifuged again at 10,000 x g for another 5 min at 24°C. The remaining lipid layer was removed, and the supernatant was once again transferred into a new 0.6 mL Eppendorf tube. The samples were further aliquoted and stored at -20°C for future use.

Holder Pasteurization Protocol

Whole milk samples were split into 2 tubes of equal volume. One tube was incubated in a preheated water bath (Thermo Fisher Scientific, Waltham, MA, USA) at 62.5°C for 30 minutes while the other tube (control) was incubated at (24°C) for 30 minutes. Both sets of samples were then placed on ice for rapid cooling and moved to (4°C) for 20 minutes.

Peptide Cleavage Site Identification by Multiplex Substrate Profiling (MSP) Mass Spectrometry Assay: Pasteurization and Inhibitor Analysis

The MSP-MS sample preparation protocol was followed as outlined in O'Donoghue et *al*., 2012 [26]. The milk samples were skimmed, and the preparation of the breast milk sample used in the MSP-MS analysis was pooled from multiple milk samples and then diluted (Figure 1.1). A peptide library was added to the pooled, diluted milk samples.

The peptide library consists of an equimolar mixture of 228 synthetic peptides that are each 14 residues total in length [26]. The tetradecapeptide library contains 18 out of the 20 natural amino acids, omitting cysteine and methionine and includes norleucine. It can be used to detect the presence of endopeptidases and exopeptidases. The peptide library was diluted 21.9-fold in either Assay Buffer 1 (0.1 M NaCitrate buffer, pH 3.5) or Assay Buffer 2 (Dulbecco's phosphate buffered saline, pH 7.2) such that the concentration of each peptide was 1 µM. A thawed pooled human milk sample was diluted 25-fold in either assay buffer 1 or assay buffer 2. 12.5 µL diluted milk and 12.5 µL of diluted peptides were combined, such that the final concentrations of the

peptide and the human milk sample are 0.5 µM and 1 in 50, respectively. This reaction was incubated at a 24°C with a 10 µL aliquots removed after 60 and 240 minutes and placed into Eppendorf tubes containing 20 µL of 8 M urea, quenching any proteolytic activity. These samples were labeled T60 and T240. The control sample (T0) consisted of 10 µL of diluted human milk mixed with 20 µL of 8 M urea prior to the addition of 10 µL of the diluted peptide pool. The assay was performed with four replicates for each time point in assay buffer 1 and 2. Samples T0, T60, and T240 were stored at -80°C then desalted with spin columns (Rainin, Oakland, CA, USA) and then injected into the Q-Exactive Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA, USA)[20].

Desalting consisted of washing a spin column with ACN (acetonitrile) and equilibrated with 0.1% TFA. Samples were then acidified to below pH 2 using 2% TFA buffer and loaded into the spin columns. The peptides were considered bound and washed with 0.1% TFA. The peptides were eluted from the spin columns with 50% ACN/0.1% TFA, dried in a SpeedVac and stored at −80 °C.

The assays outlined above were also performed in the presence of several class specific protease inhibitors that included the aspartyl protease inhibitor, pepstatin (2 µM), a serine protease inhibitor, AEBSF (500 µM), an aminopeptidase inhibitor, bestatin (2 and 10 µM), and a metalloprotease inhibitor, 1,10-phenanthroline (500 µM).
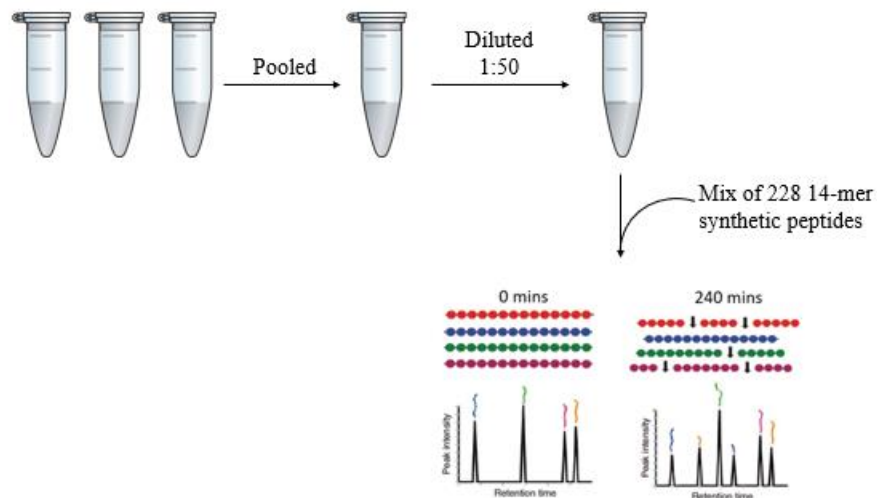
Figure 1.1: Analysis of acidified breast milk using MSP-MS methods. Multiple mother's samples are combined in equal parts and then diluted using acidic assay buffer. The diluted pooled breast milk sample is then combined and incubated with a mix of 228 synthetic peptides that will be used to detect cleavage sites from proteases present in the milk sample over time. The more time the sample is incubated, the more cleavage sites are identified. Analysis of the peptide mixture and the breast milk sample after 240 minutes is used to compare the protease cleavage sites against the control, or breast milk sample after no incubation. Each sample reading is done in quadruplicate (n=4). Datasets are filtered by showing proteolytic cleavages above a fold change (from the control) of 8. All proteolytic cleavages shown are under a q-value of .05.

Data Analysis and Presentation

Peptide quantification of raw MSP-MS data files was processed through PEAKS software (Bioinformatics Solutions Inc.). Label-free quantification was used in order to compare the raw milk and pasteurized datasets. The data was normalized by NormalyzerDE [27]. The median normalized file was utilized to run a second analysis. Processed data reports from PEAKS and Normalyzer were then input into RStudio version 1.4.1717 so that cleavage sites and specific intensity values were to be outputted into an excel sheet format. The data was filtered according to the fold change of intensity values between T240 and T0 for the control (raw) human milk data (calculated using GraphPad Prism software). A fold change greater than or equal to 5 was selected

for both the pasteurization and inhibitor dataset. The data was further filtered by the q-value that

resulted from the above calculation, which was set to less than 0.05.

Information on peptide bond cleavages observed using multiplex substrate profiling was

visualized by graphs made in Excel and Adobe Illustrator.

Fluorogenic Peptide Acquisition

A commercially available internally quenched substrates (IQ) for human cathepsin D was

purchased from CPC Scientific that consists of the sequence MCA-GKPILFFRL-K(DNP)-dArg-

NH2). Three peptides were custom synthesized by GenScript (9 mg) that consisted of the sequence

S-K(MCA)-LFRK-   Lys(DNP)-NH2(SKLFRK),   SW-K(MCA)-FRK-Lys(DNP)-NH2

(SWKFRK), and K(MCA)-WLFRK- Lys(DNP)-NH2 (KWLFRK).

Cathepsin D Activation

Recombinant human Cathepsin D protein was purchased from R&D systems at a stock

concentration of 10 µM. The sample was distributed into 1 µL aliquots and placed into -80 °C for

storage. The protease was thawed and diluted to 476 nM with pH 3.5 activation buffer (50 mM

NaCitrate and Citric Acid, adjusted with HCl). Afterward, the sample was incubated for 30 minutes

in a water bath at 37°C. After 30 minutes, the sample was further diluted to a working

concentration of 1 nM dilution with pH 3.5 assay buffer (50 mM NaCitrate and Citric Acid,

adjusted with HCl).

Fluorogenic Reporter Assays

For all fluorogenic reporter assays, black 384-well flat bottom well plates (Thermo

Scientific, Waltham, MA, USA) were used and the volume contained in each well was 30 µL.

Fluorescence was quantified in a Synergy HTX Multi-Mode Microplate Reader (BioTek,

Winooski, VT, USA) with excitation and emission wavelengths of 320 nm and 400 nm for

internally, respectively. The max velocity was calculated from 10 sequential readings within a 2-hour assay runtime. Raw protease activity was reported as the change in relative fluorescent units per second (RFUs/sec), which was normalized by multiplying by the dilution–of the protease sample.

For fluorogenic inhibition assays, the aspartic acid protease inhibitor pepstatin (MP Biomedicals) and AAF-chloromethylketone (Cayman Chemical) were diluted assay buffer 1 to 8 µM and 40 µM, respectively. An equal volume of diluted inhibitor was combined with a diluted milk sample and incubated at 24°C for 15 minutes. 15 µL aliquots of the protease-inhibitor mixture were then placed into a well of a 384-well assay plate to be mixed with 15 µL of diluted substrate. The final concentration of the protease and substrate in the wells varied while the final concentration of inhibitors in each well was 2 µM (pepstatin) and 10 µM (AAF-CMK).

For pH curve assays, skim milk was diluted 25-fold in 0.1 M Citrate Phosphate and citric acid that ranged from pH 2.2-7.8. The fluorescent substrates were diluted in the same buffer. The diluted milk was combined with 15 µL of 20 µM of SKLFRK or 10 µM of IQ substrate KPILFFRLK. As a result, each well contained 0.6 µL of skim milk and 15 µL of diluted SKLFK (20 µM) or diluted KPILFFRLK (10 µM). The final concentration of SKLFRK was 10 µM and the final concentration of diluted KPILFFRLK was 5 µM.

For large scale proteolytic activity screens using SKLFRK or KPILFFRLK, whole milk samples from 39 different mothers were split into two tubes, raw milk control and treated milk that would be Holder Pasteurized. Each sample was ran in duplicate. Each well had about 0.6 µL of raw or pasteurized whole breast milk and 15 µL of diluted SKLFRK (20 µM) or diluted KPILFFRLK (10 µM). The final concentration of SKLFRK was 10 µM and the final concentration of diluted KPILFFRLK was 5 µM.

Whole milk was pasteurized and a 4.5 µL aliquot was removed and spiked with 0.5 µL of 10 µM recombinant procathepsin D, making a 1 in 10 dilution of recombinant procathepsin D (1 µM). This spiked milk sample was then diluted 1 in 2 by adding all 5 µL into 5 µL of pasteurized milk, making 10 µL of 0.5 µM of procathepsin D-pasteurized solution. A 1 in 2 dilution was repeated until was achieved (5 times). 5 µL from 0.03125 µM solution was then added to 10 µL of pasteurized whole milk, making a 1 in 3 dilution. This is repeated 6 times until reaching 0.0000253125 µM. The procathepsin D-pasteurized milk solution is diluted 1/5 with buffer (0.1 M NaCitrate and Citric Acid, adjusted with HCl, pH 3.5) at the time the diluted substrate, KPILFFRLK (20 µM) is added in order to prevent the activation of cathepsin D prematurely before reading. After combining with the substrate-buffer solution, the final concentration of pasteurized milk was 1 in 10, which was used to trace back to whole milk calculations. The final amount of pasteurized milk-cathepsin D solution is 3 µL and the final concentration of substrate was 10 µM.

Triaminopeptidase Identification

A whole breast milk aliquot of 500 µL was first acidified with 10 µL of 1/5 diluted 6N HCl in order to achieve pH 3.5-4.0. The aliquot was then split into two vials of 250 µL each. In one vial, the control, 0.825 µL of 10 mM AAF-CMK was added and in the other vial, the working sample, 0.825 µL of 10 mM activity-based probe was added (Figure 1.2). The inhibitor concentration in each vial is 33 µM. The vials were left to incubate at 24°C for 1 hour and 20 minutes. 29.5 µL was removed from each vial and assayed with 0.5 µL of 2 mM substrate SKLFRK to check if proteolytic activity was eliminated. The final substrate concentration in the assay is 33 µM. Afterwards, 0.726 µL of 10 mM activity-based probe to the first vial with AAF-CMK and 0.726 µL of 10 mM AAF-CMK was added to the second vial with activity-based probe. The vials were left to incubate at 24°C for another 1 hour and 20 minutes. Another 29.5 µL was

removed from each vial and assayed with 0.5 µL of 2 mM substrate SKLFRK to check if proteolytic activity was consistent with the first round of assays.

After the prepping of the whole milk samples, 190 µL of milk in each vial to continue to work with. The Click-&-Go™ Protein Capture Kit (Click Chemistry Tools, Scottsdale, AZ) protocol was followed. Picolyl azide agarose resin was then mixed with the prepared milk samples. A copper catalyst solution was combined with the agarose resin and milk to complete the click reaction. The resin bound proteins were then reduced and alkylated by adding DTT and heating to 70°C and adding a 40 mM iodoacetamide solution to the beads. The resin was washed with SDS buffer, 8M urea/100 mM Tris pH 8, and 20% acetonitrile solution. The resin then underwent an on-bound trypsin digest and desalted.

Activity Based Alkyne Probe   AAF-Chloromethylketone
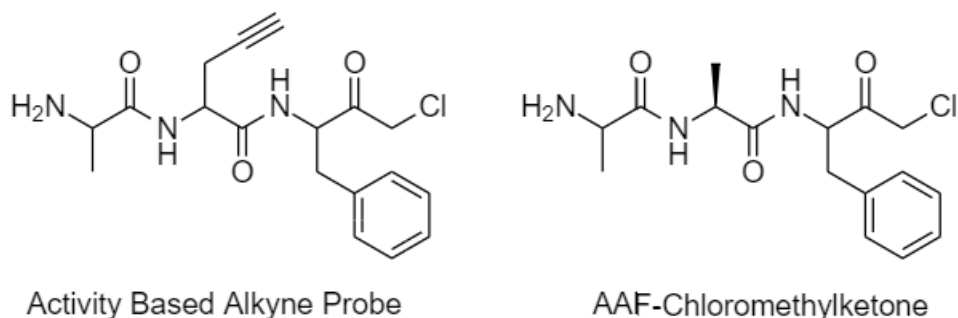
Figure 1.2: Comparison between the activity-based alkyne probe against the serine protease inhibitor, AAF-CMK. An alkyne is added to 2 carbon long chain. This probe is later used for click chemistry reactions in order to identify cleavage products of a serine triaminopeptidase.

Endopeptidase Identification

Isolation of aspartic acid protease was done using a pepstatin pull down assay with pepstatin a-agarose beads (Sigma-Aldrich, Burlington, MA, USA).

The beads were prepared by removing 100 µL of solution from pepstatin a-bead vial and placing it in a clean 2 mL vial. Then 900 µL of assay buffer (.1M NaCitrate with Citric Acid pH 3.5) was added to the beads and centrifuged at 3000 RPM for 1 minute. The supernatant was then removed from the beads (around 900 µL) and discarded. Then 1 mL of assay buffer was added, and the solution was spun at 3000 RPM for 1 minute again. Finally, 950 µL of supernatant was removed and we were left with around 50 µL of beads.

The breast milk was skimmed as mentioned previously. We added 200 µL of skim milk to 200 µL of assay buffer in order to acidify the milk before the addition of the agarose beads. At this point, 40 µL of the skim milk-acid buffer solution was reserved for fluorescent assay verification of existing proteolytic activity. The rest of the diluted skim milk (around 360 µL) was added to the 50 µL of beads. The skim milk and beads incubated at 22°C using a rotisserie for 15 minutes. After 15 minutes, centrifuge the mixture at 3000 RPM for 1 minute. Collect 40 µL of the supernatant for fluorescent assay verification of the absence of proteolytic activity since the proteases should be bound to the beads and no longer in the solution. The supernatant was removed, making sure to not disturb the beads. The beads were washed by adding 1 mL of assay buffer, spun down at 3000 RPM for 1 minute, and then the supernatant was removed. This process was done twice. The beads were then stored in -80°C until ready to perform a trypsin digest.

Proteomic studies

The proteases underwent a trypsin digest in order to isolate the proteases from the beads and be read through mass spectroscopy.

Proteins bound to Click-&-Go™ Protein Capture beads were digested with trypsin for downstream proteomic studies. Digestion buffer (100 mM Tris, 2 mM CaCl2, 10% acetonitrile, pH 8) was used to resuspend beads and to wash them. 10 µL of 0.1 µg/µL of LC/MS grade trypsin

in 200 µL of digestion buffer and the beads with the bound protein were incubated at 37°C for 6 hours. After incubation, the resin was pelleted, and the supernatant was collected and processing for mass spectroscopy.

Proteins bound to pepstatin-agarose were not wash and immediately subjected to 10 µL of 0.1 µg/µL of LC/MS grade trypsin in 200 µL of digestion buffer and the beads with the bound protein were incubated at 37°C for 6 hours. After incubation, the resin was pelleted, and the supernatant was collected and processing for mass spectroscopy.

Statistical Analysis

For the MSP-MS data, cleavage products with fold-change >8 between T240 and T0 and q-value < 0.05 were for further analysis. Each time point had quadruplicates in a specific pH. For both raw human milk analysis and pasteurization milk analysis. To calculate percent activities of the pasteurized data, the average of the raw milk samples was taken at timepoint T240 and normalized to 100%. The statistical analyses of pasteurized data are where then presented as the mean (n=4) ± SE.

For fluorescent assays, all statistical analyses in the fluorescent assay figures are presented as the mean (n=3) ± SE. Calculation of max velocity is always calculated from 10 points throughout a runtime (2 hours), reporting in RFUs/sec.

RESULTS

Protease Activity Profiling of Inhibitor Treated Human Breast Milk Samples in Acidic Conditions by MSP-MS

The inhibitor portion of the MSP-MS assay was to assess the commonalities between itself and the pasteurization MSP-MS dataset. Pepstatin is an aspartic peptidase inhibitor that is known to inhibit acid protease cathepsin D, a protease that was previously shown be present in human

breast milk [18,19,20,26]. The fold change was calculated in the same manner as outlined above for the pasteurization analysis. After filtering the overall data set to only include fold change to be equal to or greater than 8 and q value lower than 0.05, the averages of the dataset where raw milk was incubated for 240 minutes (n=4) and the dataset where pepstatin inhibited milk was incubated for 240 minutes (n=4) were then compared and graphed. Aminopeptide cleavage was inhibited by pasteurization (2 µM) (Figure 1.2). Aminopeptide cleavages that were inhibited by pasteurization (including triaminopeptides KWL*I, VMF*L, GLY*F, and HWA*F) maintained their proteolytic activity in the presence of pepstatin. This indicates that this aminopeptidase is not of the aspartyl peptidase family. However, endopeptide cleavage activity is significantly lowered by pepstatin. This is especially true for peptides containing phenylalanine in the P1 position and arginine in the P2′ position. This matches the cleavage profile of cathepsin D, as cathepsin D has been shown to preferentially cleave residues L, F, A, Y, E and D in the P1 position [28]. The endoproteases that is both inhibited by pasteurization and pepstatin is proposed to cathepsin D.

Protease Activity Profiling of Pasteurized Human Breast Milk Samples in Acidic Conditions by MSP-MS

The skimming of the donor milk sample pool is outlined in the sample processing protocol because it was essential for the operation of the Q-Exactive Mass Spectrometer.

After 240 mins incubation, 12 peptides were cleaved by proteases in raw milk. The peptides and cleavage site locations are listed in Figure 1.3. PHWQREVIF*FRLNTP, GPKLTYDF*WIQNLP, GDQPVSRGLY*FITH, KWLIHPTFSY*LRWP, LDQIYKLSNW*FFEQ, YTRLNGEAVLF*LSK, LIVQHRLFTYF*RAW, and LDGYWDHKFEL*LTW are all present in the raw control, however, are eliminated in both the pepstatin and pasteurization datasets. This indicates that these endopeptidases that have similar

14

active sites, with phenylalanine or lysine being in the P1 or P1′ position, is cleaved by a single aspartic protease and it is affected by pasteurization. VMF*LREKQYDPVSL, HWA*FRSRYHGPLAH, KWL*IHPTFSYMRWP and GLY*FRYEIPQADQW are all present in the raw control and pepstatin datasets, however, are eliminated in the pasteurization dataset. Considering that all of these peptides are cleaved in the third position, the results indicate that a tri-aminopeptidase that is not inhibited by pepstatin, making it a different protease from the aspartic protease mentioned earlier, is negatively affected by pasteurization.

When the same milk samples were pasteurized and incubated with the peptide library, many of the cleaved products found in the raw milk assay was not detected. These data indicate that pasteurization inactivate the milk protease.

4 detected triaminopeptides, peptides that are cleaved in the third cleavage position, were lowered to below 5% of original activity when detected in the pasteurized data set. As for the endopeptidase cleavages, 8 endopeptidases in the raw milk dataset were lowered to 5% of original activity in the pasteurized milk dataset. These results indicate that there are two classes of proteases, aminopeptidase and endopeptidase, that were inactivated by pasteurization.

Aminopeptidase cleavages that are dramatically reduced as stated previously are characterized by the amino acids located in the P1, P2, P3, and P1′ cleavage sites. The following amino acid sequences were targeted for cleavage by aminopeptidases in the raw milk dataset and were not targeted for cleavage in the pasteurized milk dataset (cleavage location is indicated by a "*"): KWL*I, VMF*L, GLY*F, and HWA*F. Out of the beforementioned peptides, phenylalanine P1 position, leucine in the P1′ position, and tryptophan in the P2 position.

As for the endopeptides cleavages, the most abundant amino acid in the P1 and P1′ is phenylalanine. Arginine often is neighboring phenylalanine in the P2′ position.
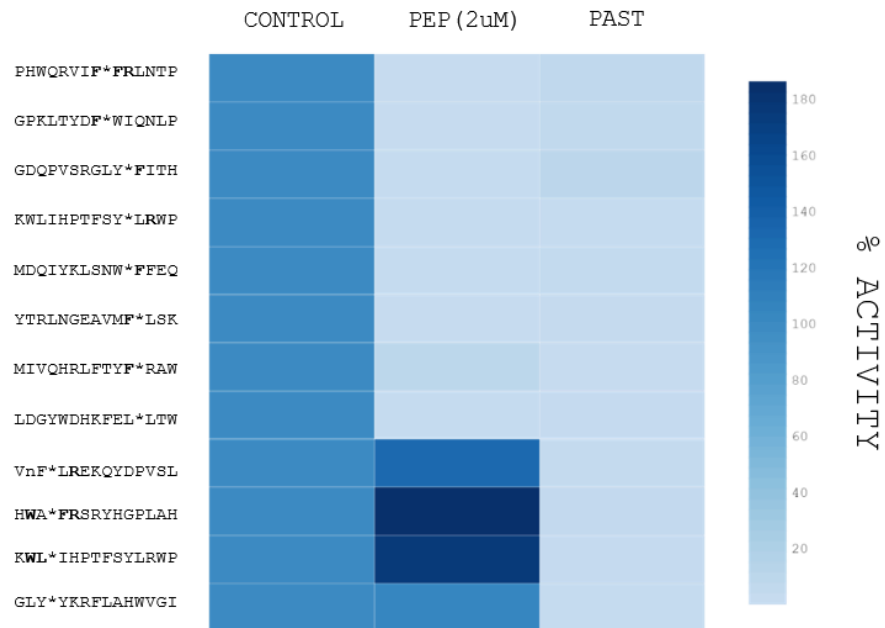
Figure 1.3: Analysis of peptides that were present in raw acidified breast milk that were absent in the Holder pasteurized and/or pepstatin (2 µM) treated breast milk MSP-MS data sets. The control in this figure is a global control that combines two controls from two different datasets.

Identification of Endoproteases Active in Acidic Conditions Affected by Pasteurization

There was a significant loss of endopeptidase activity at pH 3.5 following pasteurization

or pepstatin treatments. There is evidence for the protease being detected is cathepsin D as many

of the peptides cleaved by human milk are also cleaved by human recombinant cathepsin D [29].

To verify the findings that were seen in the MSP-MS analysis of endoproteases activity in acidic

human breast milk, fluorescent assays were completed with a reporter substrate that is designed

for cathepsin D, KPILFFRLK. In this study, we were working with a reporter substrate that was

an internally quenched peptide chain (IQ), with a fluorophore towards the N-terminus and a

quencher for the fluorophore on the carboxyterminal. When an endopeptidase such as cathepsin D

cleaves the middle of the peptide chain, the quencher moves away from the fluorophore and a light

16

emittance can be read. There was an abundance of activity, or fluorescence, that resulted from the addition of the reporter substrate to human breast milk, that is eliminated by the addition of pepstatin (Figure 1.4). This indicates that we are looking at an endopeptidase that preferentially cleaves in the middle of the peptide, since the fluorophore was blocking the N-terminus which makes it difficult for aminopeptidases to recognize and cleave a peptide.
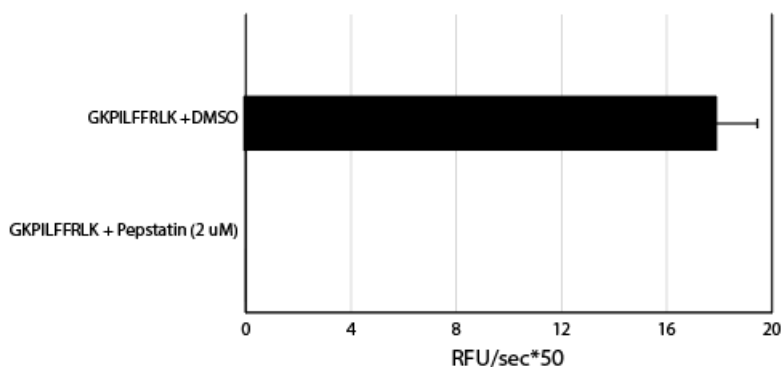
Figure 1.4: Verification of proteolytic activity in the presence of cathepsin D reporter substrate, KPILFFRLK. The negative control was acidic buffer mixed with each substrate (not shown). Quantification of protease activity is measured as max velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. The proteolytic activity is normalized according to the concentration of sample used, in this case, a final dilution of 50-fold was utilized. Data is presented as mean of triplicates (n=3) ± SD. There is abundant proteolytic activity in breast milk in response KPILFFRLK.

Identification of Triaminopeptidases in Acidic Conditions Affected by Pasteurization

MSP-MS analysis had shown that pasteurization, however not pepstatin, had a significant impact on the activity of the protease that cleaves four triaminopeptides in human breast milk in pH 3.5 conditions (Figure 1.3). Fluorescent assays were used for the verification of this triaminopeptidase in human breast milk at pH 3.5. Because we were working with an IQ fluorescent substrate for a triaminopeptidase, we wanted to make sure that the fluorescent lysine that resides at the non-prime side doesn't interfere with the recognition of the active site. If the

lysine was in a place that hindered recognition, the substrate would fail to be cleaved. The lysine also couldn't be on the N-terminus because triaminopeptidases need a free amine on the N-terminus to recognize the substate. Thus, we developed a set of custom IQ peptides that shared major amino acid residues of the four triaminopeptides found in the mass spectrometry assay, as well as a fluorescent lysine that interchanged between the P1, P2, P3 position of the peptide. Three IQ substrates were developed with serine being in the P3 position, tryptophan being in the P2 position, and leucine being in the P1 position with the lysine fluorophore placed in different positions on the non-prime side of the scissile bond [30]. The IQ that had the most activity of the three IQs developed was the IQ with the lysine fluorophore in the P2 position, SKLFRK (Figure 1.5). This peptide is also cleaved in the presence of pepstatin, so we know that this peptide is not cleaved by an aspartyl protease, and it follows the patterns that we saw in the MSP-MS data. This fluorescent substrate was used for the rest of the triaminopeptidase fluorescent assays.

When the triaminopeptidase and human breast milk are combined with SKLFRK, cleavage activity follows the same patterns in both the control (DMSO) and pepstatin (2 µM), which means that we looked towards other inhibitors. Furthermore, when human milk is treated with a chloromethylketone inhibitor, AAF-CMK (20 µM), all activity is eliminated while the pepstatin-treated milk maintains complete activity (Figure 1.6). Furthermore, it is important to note that cathepsin D is not affected by AAF-CMK. These results give strong evidence that we are looking at two different proteases that have two different functions at an optimum pH of 3.4 (Figure 1.7). Since this triaminopeptidase is inhibited by a chloromethylketone, we can assume that the protease is a serine protease.
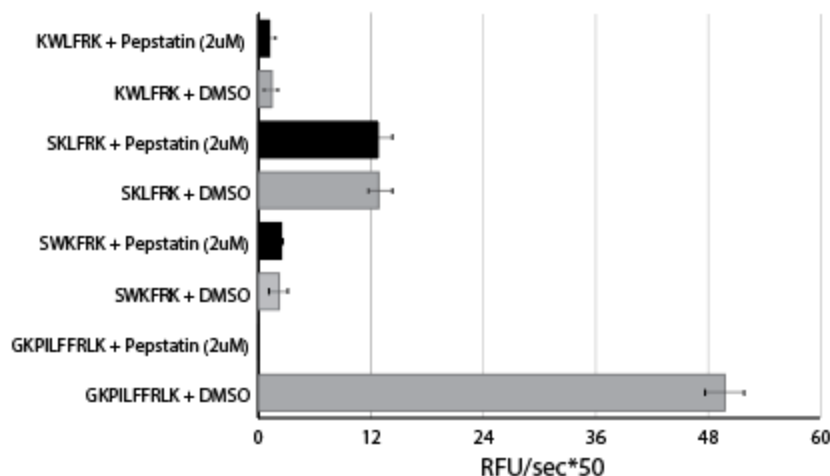
Figure 1.5: Fluorescent substrate assay of proteolytic activity in raw human breast milk and pepstatin (2 µM) treated milk at pH 3.5. IQ substrates KWLFRK, SKLFRK, and SWKFRK are flanked with a lysine MCA fluorophore either on the P1, P2, or P3 peptide position and a DNP quencher on the C terminus. KPILFFRLK serves as a positive control. The negative control was acidic buffer mixed with each substrate (not shown). Quantification of protease activity is measured as max velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. The proteolytic activity is normalized according to the concentration of sample used, in this case, a final dilution of 50-fold was utilized. Data is presented as mean of triplicates (n=3) ± SD. SKLFRK is shown to be the most rapidly cleaved substrate with the lysine fluorophore in the P2 position.
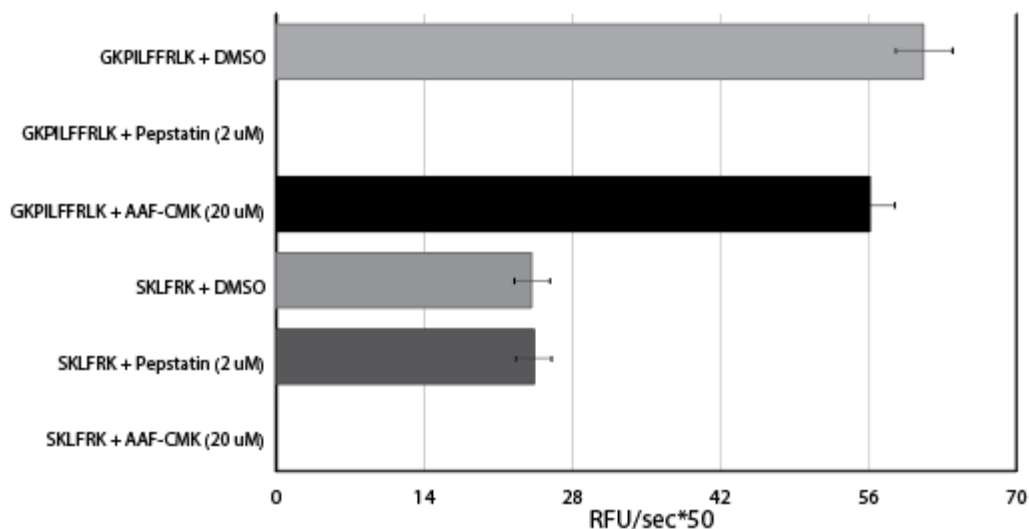
Figure 1.6: Fluorescent substrate assay of proteolytic activity with SKLFRK and KPILFFRLK in raw human breast milk, pepstatin (2 µM) treated milk, AAF-CMK (10 µM) treated milk at pH 3.5. IQ substrate SKLFRK is flanked with a lysine MCA fluorophore on the P2 peptide position and a DNP quencher on the C terminus. KPILFFRLK serves as a positive control. The negative control was acidic buffer mixed with each substrate (not shown). Quantification of protease activity is measured as maximum of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. Data is presented as mean of triplicates ($n=3$) $\pm$ SD. The proteolytic activity is normalized according to the concentration of sample used, in this case, a final dilution of 50-fold was utilized. Human milk protease cleaved KPILFFRLK in the presence of AAF-CMK inhibitor. However, human milk protease was not able to cleave SKLFRK in the presence of AAF-CMK inhibitor and SKLFRK is still cleaved when milk proteases are exposed to pepstatin.
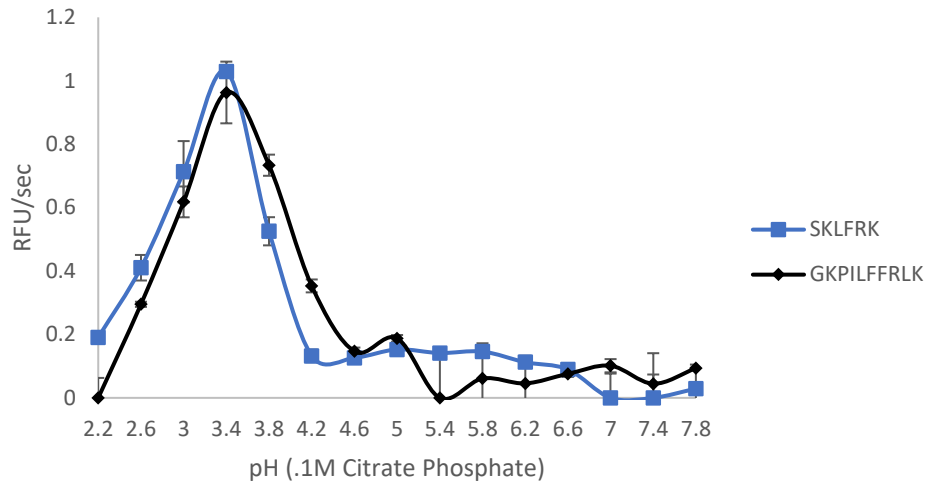
Figure 1.7: Analyzing the pH optimum of KPILFFRLK and SKLFRK through pH curves. Data is presented as mean of triplicates (n=3) ± SE. Quantification of protease activity is measured as maximum velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. The pH optimum for both KPILFFRLK and SKLFRK resides at pH 3.4.

Large Scale Fluorescent Assays

Using the fluorescent data from pooled samples, we wanted to test the substrates KPILFFRLK and SKLFRK against a panel of human breast milk samples to analyze the effects that pasteurization on individual milk samples. The screen was done with whole milk in order to replicate what would be done in a milk bank setting. To verify that activity is consistent between whole milk and skim milk, or raw milk and milk supernatant, KPILFFRLK and SKLFRK was assayed in both conditions (Figure 1.8). There is no change of activity between the two substrates in either condition. This allowed the large-scale assay to occur with unprocessed milk.

Out of the 39 human breast milk samples, all KPILFFRLK proteolytic activity was eliminated, indicating that the endoprotease that cleaves this peptide was inactivated with the Holder pasteurization protocol (Figure 1.9). Despite the varying activity of each sample, the activity is eliminated completely in every sample, indicating that pasteurization has a substantial,

global effect on different milk samples from different mothers. When the assay is repeated with

SKLFRK, activity that was once present in the milk sample is eliminated (Figure 1.10). These

results together not only support what is observed in the MSP-MS assay, they also indicate there

is a large negative impact of pasteurization on proteases that could be physiologically important

in infant digestion.

Reintroduction of Proteolytic Activity

In order to reestablish activity that is lost from pasteurization, the reintroduction of

recombinant cathepsin D into the pasteurized milk sample was tested. The concentration of

cathepsin D that resembles the activity seen in the large-scale assay with KPILFFRLK and the

sample of raw milk assayed before pasteurization is 10.4 nM (Figure 1.11)**.**
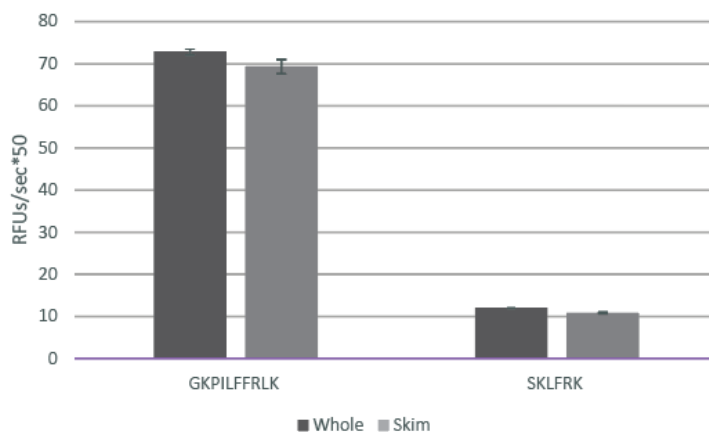
Figure 1.8: Comparison of whole and skim breast milk proteolytic activity resulting from the addition of KPILFFRLK and SKLFRK. Data is presented as mean of triplicates (n=3) ± SE. Quantification of protease activity is measured as maximum velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. The proteolytic activity is normalized according to the concentration of sample used, in this case, a final dilution of 50-fold was utilized. The difference in proteolytic activity between the two conditions, whole and skim, are negligible. It can be assumed that interchanging between whole or skim is in terms of reading fluorescence would give similar results.
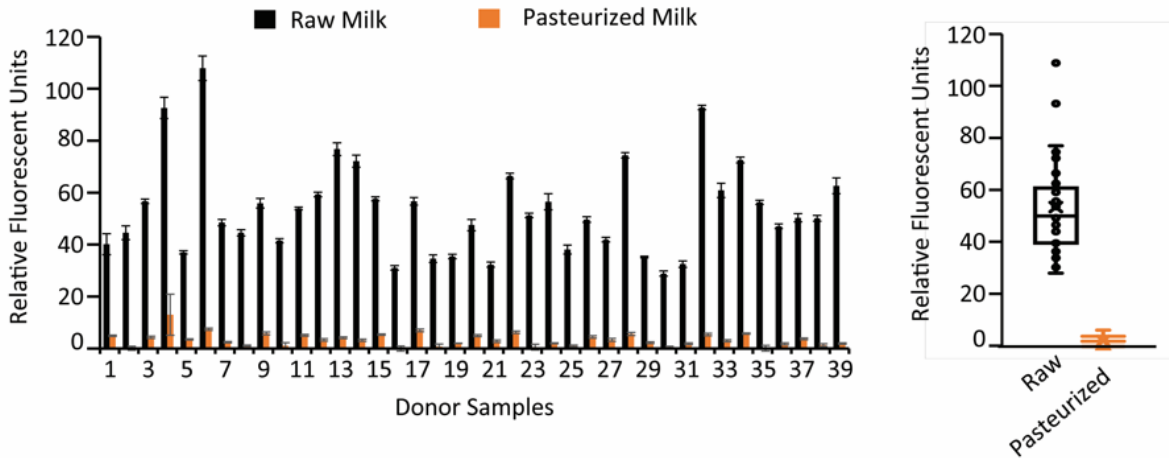
Figure 1.9: Analysis of acidic breast milk proteolytic activity in the presence of KPILFFRLK across 39 different breast milk samples before and after pasteurization. Data is presented as mean of triplicates (n=2) ± SE. Quantification of protease activity is measured as maximum velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. Although the intensity of proteolytic activity varies between each sample, the proteolytic activity in all samples is eliminated with the Holder Pasteurization treatment.
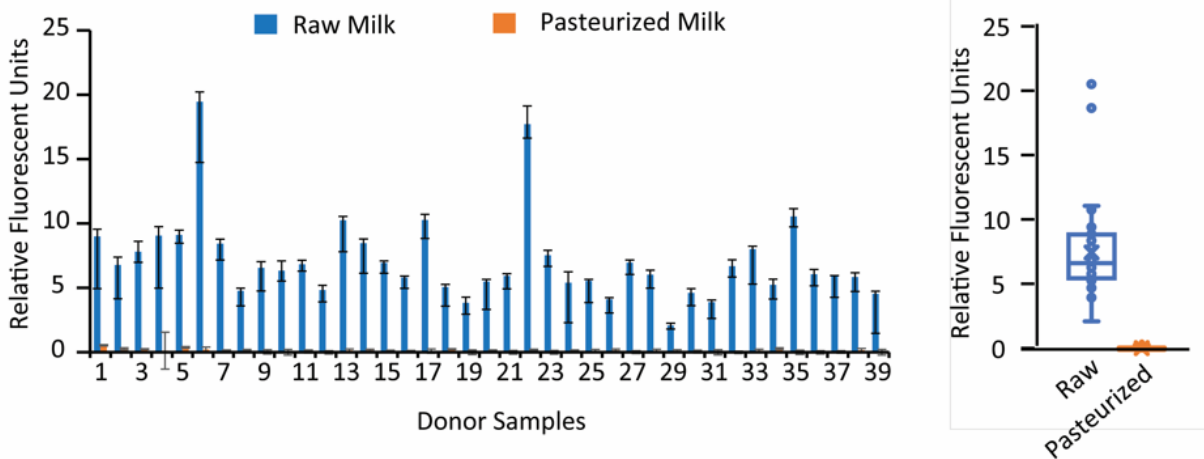
Figure 1.10: Analysis of acidic breast milk proteolytic activity in the presence of SKLFRK across 39 different breast milk samples before and after pasteurization. Data is presented as mean of triplicates (n=2) ± SE. Quantification of protease activity is measured as maximum velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. Although the intensity of proteolytic activity varies between each sample, the proteolytic activity in all samples is eliminated with the Holder Pasteurization treatment.
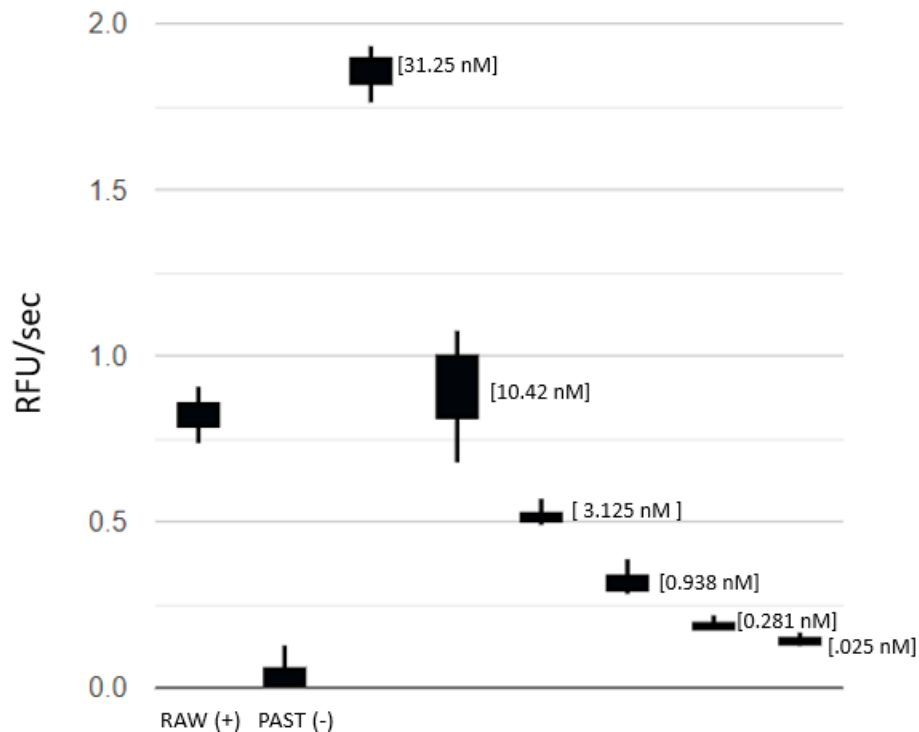
Figure 1.11: Reintroduction of recombinant human cathepsin D in pasteurized human breast milk samples. Data is presented as mean of triplicates (n=3) ± SE. Quantification of protease activity is measured as maximum velocity of ten points of the course of a two-hour reading period, labelled as relative fluorescent units per second. The "RAW" and "PAST" dataset is from a large-scale proteolytic activity analysis using 39 acidified breast milk samples. This dataset is included in this figure to compare the level of activity that is conserved across experiments. Pasteurized breast milk is tested to verify that no proteolytic activity when introduced to KPILFFRLK is present (not shown). When comparing the large-scale assay and positive control to the dilutions of recombinant cathepsin D, 10.42 nM of cathepsin D resembles that of the nontreated raw human breast milk samples.

DISCUSSION

At pH 3.5, an endopeptidase and a triaminopeptidase were identified to be major proteolytic contributors to the cleavages identified in raw human breast milk. The two proteases are also entirely inhibited by pasteurization treatment. The endopeptidase is greatly affected by pepstatin and pasteurization, while the aminopeptidase is greatly affected by AAF-CMK and

pasteurization. These are two fundamentally different proteases that share the same fate under the Holder pasteurization protocol.

Endopeptidase identification at pH 3.5: Cathepsin D

The abundance of peptides cleaved by endopeptidases that were not cleaved in the pasteurized dataset gave rise to the analysis of commonly known proteases that are active in acidic pH in human milk samples. Cathepsin D's amino acid preferences matched the residues that were not cleaved, leading to the analysis of human breast milk and the cathepsin D reporter substrate KPILFFRLK. The quality of KPILFFRLK as a fluorescent reporter substrate gives clear indication that cathepsin D-like proteases are present in milk and is inhibited in breast milk that is pasteurized.

Procathepsin D is a lysosomal protease precursor that has been detected in human breast milk at neutral pH which then undergoes a conformational shift when exposed to acidic condition to become activated. The roles of cathepsin D span many human biological systems and plays a huge part in bodily development. Active cathepsin D is involved in protein degradation of lysosomes, polypeptide hormones, chemokines, and growth factors as well as the activation of enzymatic precursors and brain antigen processing [31]. The lack of such a versatile protease alone, especially at a young and vulnerable state as that of a preterm infant, would greatly contribute to the complications in later adult life, which preterm infants are known to be at higher risk for. Cathepsin D being completely inactivated after pasteurization could provide some insight into additional reasons why infants are not able to protect themselves against NEC or grow as efficiently with donor breast milk compared to those provided with MOM [5].

Triaminopeptidase identification at pH 3.5

After preforming a click reaction with an alkyne activity-based probe and an azide agarose resin bead, the identity of the protease that exhibits triaminopeptidase activity is inconclusive. The

products of two proteases were identified in the acidified milk with the activity-based probe, which were cytosol aminopeptidase and CNDP2. Cytosol aminopeptidase has been seen in breast milk previously, but active at pH 7.4 [16-20]. The pH optimum of the protease that cleaves SKLFRK is pH 3.5 and exhibits no activity at neutral pH. As for CNDP2, or cytosolic non-specific dipeptidase, it has been found in the exosomes derived from breast milk [32]. The pH optimum of CNDP2, like cytosol aminopeptidase, is at neutral pH [33]. Not only is the pH optimum off, but the enzyme was only found to cleave dipeptides in the presence of Mn^{2+}.

There is the potential that we found latent form or function for these proteases in breast milk specially that differs from other parts of the body. Without the isolation of the protease itself to test, we can only speculate that possibility.

FUTURE DIRECTIONS

The direction of this study would be to study active proteases at pH 7.4. A large part of our analysis of the effects of pasteurization included analyzing proteases active at neutral pH with there being MSP-MS. There is also more preexisting knowledge about the proteases that exist at neutral pH. We also want to continue to find a conclusive identity of the triaminopeptidase present throughout our studies

The overarching goal is large scale breast milk analysis that is user friendly. An easy-to-use plate preloaded with fluorescent substrates would be used to assess the proteolytic content quality of a donor milk sample, so that appropriate supplementation can be added if needed. These tests would be accessible by human donor banks so that testing can occur in a quick manner. Proteases that are found to be inactive early on in testing could play a factor in preventing detrimental physical effects to a preterminal later on in life, as proper supplementation is essential for such high-risk children [5]. Supplementation development for donor breast milk and formula

would go hand in hand with this goal as banks and formula production companies would need a protease product to safely supplement their product.

Our study would also branch out into a functionality endeavor, where we would try to find the exact purpose of all proteases found in breast milk when ingested by the infant. This would be in the form of a clinical type of study where different modes of feeding are used and we would analyze the effects on the child. Assuming we have complete knowledge of the proteomics of breast milk via the before-mentioned assays, we would then be able to start to generalize.

CONCLUSION

Pasteurization has a larger effect on proteolytic activity than we previously hypothesized. Implications of such results would ultimately call for one of two of the following actions: the development of proper milk fortification or the altering of the disinfection protocol of the donated milk. With the second option being too out of scope for the time being, referencing the recent pandemic, proper milk fortification with additional proteases that are lost to pasteurization, such as cathepsin D, would be more achievable and proposed to be affective in lowering risks for serious medical complications for already high-risk children.

Acknowledgements

REFERENCES

[1] Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J. E., Cousens, S., Mathers, C., & Black, R. E. (2016). Global, regional, and national causes of under-5 mortality in 2000-15: anupdated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England), 388(10063), 3027–3035. https://doi.org/10.1016/S0140-6736(16)31593-8

[2] Singh, G. K., & Yu, S. M. (2019). Infant Mortality in the United States, 1915-2017: Large Social Inequalities have Persisted for Over a Century. International journal of MCH and AIDS,8(1), 19–31. https://doi.org/10.21106/ijma.271

[3] Vogel, J. P., Chawanpaiboon, S., Moller, A.-B., Watananirun, K., Bonet, M., &Lumbiganon, P. (2018). The global epidemiology of preterm birth. Best Practice & Research Clinical Obstetrics & Gynaecology, 52, 3–12. https://doi.org/10.1016/j.bpobgyn.2018.04.003

[4] Agostoni, C., Buonocore, G., Carnielli, V. P., De Curtis, M., Darmaun, D., Decsi, T., Domellöf, M., Embleton, N. D., Fusch, C., Genzel-Boroviczeny, O., Goulet, O., Kalhan, S. C., Kolacek, S., Koletzko, B., Lapillonne, A., Mihatsch, W., Moreno, L., Neu, J., Poindexter, B., Puntis, J., … ESPGHAN Committee on Nutrition (2010). Enteral nutrient supply for preterm infants: commentary from the European Society of Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition. Journal of pediatric gastroenterology and nutrition, 50(1), 85–91. https://doi.org/10.1097/MPG.0b013e3181adaee-++-0

[5] Harding, J. E., Cormack, B. E., Alexander, T., Alsweiler, J. M., & Bloomfield, F. H. (2017). Advances in nutrition of the newborn infant. The Lancet, 389(10079), 1660–1668. https://doi.org/10.1016/s0140-6736(17)30552-4

[6] Wallenborn, J. T., Levine, G. A., Carreira dos Santos, A., Grisi, S., Brentani, A., & Fink, G. (2021). Breastfeeding, Physical Growth, and Cognitive Development. Pediatrics, 147(5). https://doi.org/10.1542/peds.2020-008029

[7] Meier, P., Patel, A., & Esquerra-Zwiers, A. (2017). Donor Human Milk Update: Evidence, Mechanisms, and Priorities for Research and Practice. The Journal of pediatrics, 180, 15–21. https://doi.org/10.1016/j.jpeds.2016.09.027

[8] Radmacher, P. G., & Adamkin, D. H. (2017). Fortification of human milk for preterm infants. Seminars in Fetal and Neonatal Medicine, 22(1), 30–35. https://doi.org/10.1016/j.siny.2016.08.004

[9]Weaver, G., Bertino, E., Gebauer, C., Grovslien, A., Mileusnic-Milenovic, R., Arslanoglu, S., Barnett, D., Boquien, C. Y., Buffin, R., Gaya, A., Moro, G. E., Wesolowska, A., & Picaud, J. C. (2019). Recommendations for the Establishment and Operation of Human Milk Banks in Europe: A Consensus Statement From the European Milk Bank Association (EMBA). Frontiers in pediatrics, 7, 53. https://doi.org/10.3389/fped.2019.00053

[10] Patel, A. L., & Kim, J. H. (2018). Human milk and necrotizing enterocolitis. Seminars in pediatric surgery, 27(1), 34–38. https://doi.org/10.1053/j.sempedsurg.2017.11.007

[11] Binte Abu Bakar, S. Y., Salim, M., Clulow, A. J., Nicholas, K. R., & Boyd, B. J. (2021). Human milk composition and the effects of pasteurisation on the activity of its components. Trends in Food Science & Technology, 111, 166–174. https://doi.org/10.1016/j.tifs.2021.02.055

[12] Peila, C., Moro, G. E., Bertino, E., Cavallarin, L., Giribaldi, M., Giuliani, F., Cresi, F., & Coscia, A. (2016). The Effect of Holder Pasteurization on Nutrients and Biologically-Active Components in Donor Human Milk: A Review. Nutrients, 8(8), 477. https://doi.org/10.3390/nu8080477

[13] Wada, Y., & Lönnerdal, B. (2020). Bioactive peptides derived from human milk proteins: an update. Current opinion in clinical nutrition and metabolic care, 23(3), 217–222. https://doi.org/10.1097/MCO.0000000000000642

[14] Antalis, T. M., Shea-Donohue, T., Vogel, S. N., Sears, C., & Fasano, A. (2007). Mechanisms of disease: protease functions in intestinal mucosal pathobiology. Nature clinical practice. Gastroenterology & hepatology, 4(7), 393–402. https://doi.org/10.1038/ncpgasthep0846
[14] Dallas, D. C., & German, J. B. (2017). Enzymes in Human Milk. Nestle Nutrition Institute workshop series, 88, 129–136. https://doi.org/10.1159/000455250

[16] Elmlinger, M. W., Grund, R., Buck, M., Wollmann, H. A., Feist, N., Weber, M. M., Speer, C. P., & Ranke, M. B. (1999). Limited Proteolysis of the IGF Binding Protein-2 (IGFBP-2) by a Specific Serine Protease Activity in Early Breast Milk. Pediatric Research, 46(1), 76–81. https://doi.org/10.1203/00006450-199907000-00013

[17] Demers-Mathieu, V., Nielsen, S. D., Underwood, M. A., Borghese, R., & Dallas, D. C. (2018). Changes in Proteases, Antiproteases, and Bioactive Proteins From Mother's Breast Milk to the Premature Infant Stomach. Journal of pediatric gastroenterology and nutrition, 66(2), 318–324. https://doi.org/10.1097/MPG.0000000000001719

[18] Nielsen, S. D., Beverly, R. L., & Dallas, D. C. (2017). Milk Proteins Are Predigested Within the Human Mammary Gland. Journal of Mammary Gland Biology and Neoplasia, 22(4), 251–261. https://doi.org/10.1007/s10911-018-9388-0

[19] Kelly, E. J., Newell, S. J., Brownlee, K. G., Primrose, J. N., & Dear, P. R. F. (1993). Gastric acid secretion in preterm infants. Early Human Development, 35(3), 215–220. https://doi.org/10.1016/0378-3782(93)90108-7

[20] Větvicka, V., Vágner, J., Baudys, M., Tang, J., Foundling, S. I., & Fusek, M. (1993). Human breast milk contains procathepsin D--detection by specific antibodies. Biochemistry and molecular biology international, 30(5), 921–928.

[21] Gan, J., Zheng, J., Krishnakumar, N., Goonatilleke, E., Lebrilla, C. B., Barile, D., & German, J. B. (2019). Selective Proteolysis of α-Lactalbumin by Endogenous Enzymes of Human Milk at Acidic pH. Molecular Nutrition & Food Research, 63(18), 1900259. https://doi.org/10.1002/mnfr.201900259

[22] Deglaire, A., Oliveira, S. D., Jardin, J., Briard-Bion, V., Kroell, F., Emily, M., Ménard, O., Bourlieu, C., & Dupont, D. (2019). Impact of human milk pasteurization on the kinetics of peptide release during in vitro dynamic digestion at the preterm newborn stage. Food Chemistry,281, 294–303. https://doi.org/10.1016/j.foodchem.2018.12.086

[23] Maffioli, E., Jiang, Z., Nonnis, S., Negri, A., Romeo, V., Lietz, C. B., Hook, V., Ristagno, G., Baselli, G., Kistler, E. B., Aletti, F., O'Donoghue, A. J., & Tedeschi, G. (2020). High-Resolution Mass Spectrometry-Based Approaches for the Detection and Quantification of Peptidase Activity in Plasma. Molecules (Basel, Switzerland), 25(18), 4071. https://doi.org/10.3390/molecules25184071

[24] Ivry, S. L., Sharib, J. M., Dominguez, D. A., Roy, N., Hatcher, S. E., Yip-Schneider, M. T., Schmidt, C. M., Brand, R. E., Park, W. G., Hebrok, M., Kim, G. E., O'Donoghue, A. J., Kirkwood, K. S., & Craik, C. S. (2017). Global Protease Activity Profiling Provides Differential Diagnosis of Pancreatic Cysts. Clinical cancer research : an official journal of the American Association for Cancer Research, 23(16), 4865–4874. https://doi.org/10.1158/1078-0432.CCR-16-2987

[25] O'Donoghue, A. J., Jin, Y., Knudsen, G. M., Perera, N. C., Jenne, D. E., Murphy, J. E., Craik, C. S., & Hermiston, T. W. (2013). Global substrate profiling of proteases in human neutrophil extracellular traps reveals consensus motif predominantly contributed by elastase. PloS One, 8(9), e75141. https://doi.org/10.1371/journal.pone.0075141

[26] O'Donoghue, A. J., Eroy-Reveles, A. A., Knudsen, G. M., Ingram, J., Zhou, M., Statnekov, J. B., Greninger, A. L., Hostetter, D. R., Qu, G., Maltby, D. A., Anderson, M. O., Derisi, J. L., McKerrow, J. H., Burlingame, A. L., & Craik, C. S. (2012). Global identification of peptidase specificity by multiplex substrate profiling. Nature methods, 9(11), 1095–1100. https://doi.org/10.1038/nmeth.2182

[27] Sun, H., Lou, X., Shan, Q., Zhang, J., Zhu, X., Zhang, J., Wang, Y., Xie, Y., Xu, N., & Liu, S. (2013). Proteolytic Characteristics of Cathepsin D Related to the Recognition and Cleavage of Its Target Proteins. PLoS ONE, 8(6). https://doi.org/10.1371/journal.pone.0065733

[28] National Center for Biotechnology Information. PubChem Bioassay Record for AID 668998, Source: ChEMBL. https://pubchem.ncbi.nlm.nih.gov/bioassay/668998. Accessed June 3, 2021.

[29] Jiang, Z., Lietz, C. B., Podvin, S., Yoon, M. C., Toneff, T., Hook, V., & O'Donoghue, A. J. (2021). Differential Neuropeptidomes of Dense Core Secretory Vesicles (DCSV) Produced at Intravesicular and Extracellular pH Conditions by Proteolytic Processing. ACS chemical neuroscience, 12(13), 2385–2398. https://doi.org/10.1021/acschemneuro.1c00133

[30] Ivry, S. L., Knudsen, G. M., Caiazza, F., Sharib, J. M., Jaradeh, K., Ravalin, M., O'Donoghue, A. J., Kirkwood, K. S., & Craik, C. S. (2019). The lysosomal aminopeptidase tripeptidyl peptidase 1 displays increased activity in malignant pancreatic cysts. Biological Chemistry, 400(12), 1629–1638. https://doi.org/10.1515/hsz-2019-0103

[31] Benes, P., Vetvicka, V., & Fusek, M. (2008). Cathepsin D--many functions of one aspartic protease. Critical reviews in oncology/hematology, 68(1), 12–28. https://doi.org/10.1016/j.critrevonc.2008.02.008

[32] Katz, M. L., Johnson, G. C., Leach, S. B., Williamson, B. G., Coates, J. R., Whiting, R. E., Vansteenkiste, D. P., & Whitney, M. S. (2017). Extraneuronal pathology in a canine model

of CLN2 neuronal ceroid lipofuscinosis after intracerebroventricular gene therapy that delays neurological disease progression. Gene Therapy, 24(4), 215–223. https://doi.org/10.1038/gt.2017.4

[33] Admyre, C., Johansson, S. M., Qazi, K. R., Filén, J. J., Lahesmaa, R., Norman, M., Neve, E. P., Scheynius, A., & Gabrielsson, S. (2007). Exosomes with immune modulatory features are present in human breast milk. Journal of immunology (Baltimore, Md. : 1950), 179(3), 1969–1978. https://doi.org/10.4049/jimmunol.179.3.1969

[34] Bellia, F., Vecchio, G., & Rizzarelli, E. (2014). Carnosinases, their substrates and diseases. Molecules (Basel, Switzerland), 19(2), 2299–2329. https://doi.org/10.3390/molecules19022299

CHAPTER 2

<u>INTRODUCTION</u>

There has been increasing recognition of the protective role of human milk and early food exposure in the development of atopic conditions such as eczema, food allergy, and asthma [1–4]. Asthma is the most prevalent chronic disease in children, affecting over 300 million people worldwide [5] and disproportionately affects preterm infants. In one study, infants born before 37 weeks (moderate to late preterm) were 50% more likely to develop asthma, and infants born before 32 weeks (extremely to very preterm) were three times as likely to develop asthma [6]. While mechanical ventilation and other causes of direct damage to the lung may account for some of this increased risk for the development of asthma in the preterm population, other factors may also play a role. Lower rates of atopy in term infants were associated with exclusive breastfeeding for the first four months of life [7], lack of antibiotic exposure (either maternal intrapartum or early in infancy) [8], vaginal delivery [9] and furry pets in the home [10]. Preterm infants spend this critical period in the neonatal intensive care unit, and some are exposed to antibiotics and c-section delivery but most importantly, almost all experience most of their nutrition via a nasogastric tube or a bottle containing previously frozen maternal or donor breast milk, instead of fresh breastmilk via breastfeeding.

Recommendations regarding the introduction of allergenic foods to infants has shifted from avoiding allergenic foods until 1 year of age to early introduction prior to 6 months of age, as the latter has been shown to be associated with a decreased risk of food allergy during a critical window of the infant's immune development [4]. This approach is thought to reduce the induction of Type 2 inflammation that is responsible for allergic conditions. Given that early exposure to

food proteins can alter the development of food allergies later in life, early nutrition has become an area of interest in studying the pathogenesis of atopic disease [11]. Over 70% of peanut reactions occur on the first known exposure, indicating prior sensitization, perhaps transcutaneously, via inhalation, or via human milk [12,13]. Since human milk is often the primary source of nutrition in infancy, it has been postulated that it could be a source for allergen introduction [14]. Although IgE mediated reactions to human milk are rare, they do occur, demonstrating its immunogenic nature [15,16].

Airborne allergens from house dust mite have also been found in human milk at similar quantities to food allergens [17]. One dust mite protein, Der p1 has demonstrated both Toll-like receptor agonist and protease activities, which could potentially initiate allergic immune responses [18]. In addition to food and environmental allergens, human milk also contains many other bioactive substances, including endogenous proteases and protease inhibitors, immunoglobulins, soluble receptors, cytokines, human milk oligosaccharides (HMOs), fatty acids, and microbes [11,19 20–23].

Human milk proteomics studies have utilized different methods for protein identification, including Western blotting, ELISA, and mass spectrometry, which may account for heterogeneity between experiment results [24–31]. Mass spectrometry has been utilized to determine the protein content in human milk, however, there are few studies that are primarily focused on allergenic proteins [11]. Studies utilizing mass spectrometry, which allows for a broader untargeted search for proteins, have identified 1200-1600 total proteins in human milk, the vast majority of non-human proteins derived from cow's milk, with one study also identifying dog, horse, cat, chicken and rice proteins [15,32]. The degree of protein alteration prior to its appearance in human milk is unknown, and studies are conflicting. Some studies have found that peanut Ara h 1 and 2, 14 ovalbumin [32],

and gliadin [25] are not degraded in human milk. This contrasts with other studies that identified fragments of ß-Lactoglobin [15,33] and α-S1 casein [53].

Proteases have been identified in human milk and appear to play a significant role in infant digestion, but the interaction between endogenous human milk proteases and non-human proteins such as house dust mite proteases and human milk fortifiers (HMFs) proteases has not been extensively studied [34].

The development of tolerance versus sensitization to allergens is complex and depends on the interaction and often multi-directional relationship between many different factors, such as maternal history, milk composition, gut immunology, and microbiome and external environment[21]. Allergen shedding in human milk may be a way to educate the infant's immune system and modulate allergy risk in the infant [35]. Given advancements in medicine that have led to increased survival of preterm infants, we aim to examine longer term outcomes such as allergic sensitization or tolerance in this population by examining human milk protein composition and exposure. Due to newer techniques in proteomic and peptidomic analysis, and the paucity of data regarding the ability of HM to induce tolerance or sensitization to allergens, we developed a feasibility pilot study on a subset of milk samples to investigate how nutrition and environmental exposures may impact allergen shedding in human milk in preterm infants.

METHODS

Sample Collection

Four human milk samples (2 from mothers of term infants and 2 from mothers of preterm infants) were analyzed from the Mommy's Milk Human Milk Biorepository (HMB) to evaluate the feasibility of analyzing human milk samples by untargeted mass spectrometry (Table 2.1.1). The Mommy's Milk HMB was founded at the University of California, San Diego in 2014 with

the goal of building a constant but rotating inventory of 3,000 human milk samples available for future research [36]. Following informed consent, women provide 50 mL up to a full pump of expressed breast milk (convenience sample). Participants are interviewed about their sociodemographic characteristics, pregnancy history, dietary intake using a standard questionnaire (https://www.nutritionquest.com), medication exposure, lifestyle habits, maternal stress, anxiety and depression, breastfeeding behaviors, and signs and symptoms of potential adverse reactions in the offspring. Data on growth of the infant/toddler are captured from medical records, and neurodevelopmental assessments are conducted longitudinally. Sample collections occur at UC San Diego, community sites, or the participant's home. Human milk samples are stored and shipped on ice within 24 hours of collection to the Mommy's Milk lab where the sample is aliquoted and stored at −80°C until requested for study analysis.

Table 2.1.1: Maternal Dietary History: Mommy's Milk Human Milk Repository (HMB) Samples

Sample ID number	Infant's gestational age	Infant's gender	Cow milk	Egg	Wheat	Nuts	Fish / Shellfish
R1	36 weeks	Female	no	no	yes	yes	yes
R2	33 weeks	Female	yes	no	yes	yes	yes
R3	39 weeks	Male	yes	yes	yes	yes	yes
R4	40 weeks	Female	no	yes	yes	yes	yes

Table 2.1.2: Maternal Dietary History: Microbiome, Atopy, and Prematurity (MAP) Samples

Sample ID number	Cow's milk	Egg	Soy	Wheat	Peanut	Tree Nuts	Fish	Shrimp	Shellfish
J7	rarely	daily	never	daily	daily	daily	daily	never	never
J8	weekly	weekly	rarely	weekly	never	never	weekly	weekly	never
J9	daily	daily	rarely	weekly	daily	weekly	rarely	rarely	rarely
J10	weekly	weekly	rarely	weekly	weekly	monthly	monthly	never	never
J11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
J12	daily	daily	rarely	daily	weekly	weekly	monthly	monthly	never

We also analyzed 5 human milk samples and 1 formula sample (J11) from mothers and preterm infants recruited into the Microbiome, Atopy and Prematurity (MAP) pilot study (under

review) (Table 2.1.2). The MAP study population recruited 48 preterm infants, ≤34 weeks, and their mothers, from Jacobs Medical Center at UCSD in the Newborn Intensive Care Unit (NICU) and Scripps Memorial La Jolla NICU. At birth, prenatal (maternal antibiotics and diet, pregnancy morbidities, smoking, pet ownership, family history of asthma and other social and demographic information) and perinatal (method of delivery, need for resuscitation) factors/exposures were documented at the time of enrollment. Parents were given a History and Allergy Questionnaire at enrollment which asked about family history of asthma, smoking, allergies, medication, dietary history during pregnancy and postnatally (Supplemental table 1). Milk, stool and saliva samples from the study participants were collected weekly and stored immediately at 4°C and transferred to -80°C within 36 hours post collection. For this pilot study, first week milk samples were analyzed from 6 infant/parent couplets. Since the majority of preterm infants require fortification to support their growth and development, half of the milk samples (3/6) analyzed from the MAP study contained human milk or formula-based fortifiers. To control for those additives, an additional 2 samples of regular and hydrolyzed HMF (Enfamil human milk fortifier acidified liquid (Mead Johnson)) and one formula sample (Enfamil premature (Mead Johnson)) were analyzed. This research was performed in accordance with the ethical principles for medical research involving human subjects outlined in the Declaration of Helsinki. The study protocol was approved by the University of California, San Diego's Human Research Protections Program IRB# 181711.

Sample preparation for proteomic analysis

Milk samples were thawed on ice prior to preparation for proteomic analysis. Guanidine-HCl was added to 2 µL of milk sample to achieve a final concentration of 6 M. The samples were boiled for 10 minutes followed by 5 minutes cooling at room temperature. The boiling and cooling cycle was repeated 3 times. The proteins were precipitated with the addition of methanol to a final

volume of 90% followed by vortex and centrifugation at 14,000 rpm on a benchtop microfuge for 10 minutes at 25°C. The soluble fraction was removed by flipping the tube onto an absorbent surface and tapping to remove any liquid. The pellet was suspended in 200 µL of 8 M Urea made in 100 mM Tris pH 8.0. Tris (2-carboxyethyl) phosphine and chloroacetamide were added to final concentrations of 10 mM and 40 mM, respectively and the mixture was vortexed for 5 minutes. 3 volumes of 50 mM Tris pH 8.0 was added to the sample to reduce the final urea concentration to 2 M. Trypsin was added in a 1:50 protein ratio and incubated at 37°C for 12 hours. The solution was then acidified using TFA (0.5% TFA final concentration) and mixed. Samples were desalted using 100 mg C18-StageTips as described by the manufacturer protocol. The peptide concentration of the samples was measured using BCA after resuspension in sample loading buffer and a total of 0.5 µg was injected for each label free quantification run.

Sample preparation for peptidomic analysis

To remove high molecular weight milk proteins, 100 µL of human milk was mixed with 900 µL of methanol and vortexed for 5 seconds. The samples were kept at 25°C for 30 minutes followed by centrifugation at 12,000 rpm for 10 minutes at 25°C. 500 µL of supernatant was transferred to a fresh tube dried in a vacuum centrifuge. The samples were hydrated in 0.5 mL of 0.5% formic acid and 5% acetonitrile (ACN) solution and desalted using a Sep-PAK C18 1 cc Vac (Waters Corporation, Milford MA) according to the manufacturer's protocol with the exception that 40% ACN was used to elute peptides. The eluents were dried in speed-vac in preparation for mass spectrometry analysis

Liquid Chromatography with Tandem Mass Spectrometry (LCMS-MS)

Trypsin-digested peptides were analyzed by ultra-high pressure liquid chromatography (UPLC) coupled with tandem mass spectroscopy (LC-MS/MS) using nano-spray ionization. The

nanospray ionization experiments were performed using a Orbitrap fusion Lumos hybrid mass spectrometer (Thermo) interfaced with nano-scale reversed-phase UPLC (Thermo Dionex UltiMate™ 3000 RSLC nano System) using a 25 cm, 75-micron ID glass capillary packed with 1.7-µm C18 (130) BEH TM beads (Waters corporation). Peptides were eluted from the C18 column into the mass spectrometer using a linear gradient (5–80%) of ACN (Acetonitrile) at a flow rate of 375 µL/min for 2h. The buffers used to create the ACN gradient were: Buffer A (98% H2O, 2% ACN, 0.1% formic acid) and Buffer B (100% ACN, 0.1% formic acid). Mass spectrometer parameters are as follows; an MS1 survey scan using the orbitrap detector (mass range (m/z): 400-1500 (using quadrupole isolation), 120,000 resolution setting, spray voltage of 2,200 V, Ion transfer tube temperature of 275°C, AGC target of 400,000, and maximum injection time of 50 ms) was followed by data dependent scans (top speed for most intense ions, with charge state set to only include +2-5 ions, and 5 second exclusion time, while selecting ions with minimal intensities of 50,000 at which the collision event was carried out in the high energy collision cell (HCD Collision Energy of 30%), and the fragment masses were analyzed in the ion trap mass analyzer (With ion trap scan rate of turbo, first mass m/z was 100, AGC Target 5000 and maximum injection time of 35ms). Protein identification and label free quantification was carried out using Peaks Studio 8.5 (Bioinformatics solutions Inc.)

Analysis was performed in two separate runs – the first included the samples from the breast milk repository, and the second included the samples from the MAP study, formula, and fortifier. Database searches were carried out against a reference database that included FASTA protein sequences of known protein allergens from the University of Nebraska (http://www.allergenonline.org/ (used version 19, published 02/10/19)) that was combined with human proteome UniProt sequences using Peaks 8.5 (Bioinformatics Solutions) search engine.

This database included known allergenic peptides and all predicted human proteome. The hits were filtered at 1% false discovery rate (FDR) before being considered for further analysis. All positive peptide sequence results were verified by blasting original sequences against UniProt and NCBI (https://www.uniprot.org/blast/) to confirm accuracy. Results where at least one of the two of the databases did not have > 70% identity to the labeled species-specific protein were discarded.

Protease Activity

Samples were diluted 50-fold in 50 mM Tris-HCl, pH 9.0, 150 mM NaCl and assayed with 25 µM RR-AMC (Santa Cruz Biotechnology, sc-281540) in triplicate wells on a black 384-well plate. The final volume in each well was 30 µL and the assay was performed at 25°C. Activity was monitored for 2 hours on a BioTek HTX plate reader with excitation of 360 nm and emission of 460 nm. Activity was reported as the change in relative change in fluorescence units per sec (RFU/sec). HMB samples did not contain fortifier and did not undergo protease analysis. (Summary of milk feed content and analysis performed- Table 2.2).

Table 2.2: Milk Sample Composition and type of analysis

Sample ID	Milk Sample Composition	Proteomics	Peptidomics
J7	HM + fortifier	Yes	Yes
J8	HM	Yes	Yes
J9	HM	Yes	Yes
J10	HM + fortifier	Yes	Yes
J11	preterm formula	Yes	Yes
J12	HM	Yes	Yes
R1	HM	Yes	No
R2	HM	Yes	No
R3	HM	Yes	No
R4	HM	Yes	No
HM: human milk; fortifier: Enfamil Human milk fortifier (non-hydrolyzed)			

RESULTS

Proteomic Feasibility Study from the Mommy's Milk Human Milk Biorepository samples

To determine the feasibility of detecting allergenic peptides/proteins in human milk, we performed mass spectrometry on four (2 term and 2 preterm) trypsin-digested breast milk samples (R1, R2, R3, R4) and utilized the University of Nebraska FASTA and UniProt databases for a total of 2211 sequences for comparison. Each sample had between 806 and 1007 peptides with 28 to 38 non-human proteins per sample encompassing 23 different plant and animal species (Table 2.3). We detected peptides from various food, venom/salivary, and airborne sources. There were no appreciable differences between term and preterm samples in terms of total protein content. One sample accounted for over 50% of the non-human peptide variability (R4).

Table 2.3: List of pertinent bovine peptides and all NH non-bovine peptides stratified by sample proportion in Fortifiers, HBM, and MAP samples

Species	Sample	Proteins	Digested vs Undigested
Bos taurus (cow)	14/14	β-Lg (**Bos d 5**)	Digested and non-digested, in formula samples only
	13/14	α-S1-casein (**Bos d 9**), β-casein (**Bos d 11**), serum albumin (**Bos d 6**)	-Bos d 9 – 3 isoforms; digested only in the human samples, digested and non-digested in formula samples
			-Bos d 11 – 5 isoforms; both digested and non-digested in human and formula samples
			-Bos d 6 – 5 isoforms; all digested in human and formula samples
	12/14	α-lactalbumin (**Bos d 4**)	4 isoforms were digested in human and formula samples; one isoform digested and non-digested in formula only
	9/14	α-lactalbumin precursor	Digested in human and human samples
	8/14	C3, β-casein A3, ceruloplasmin	-C3 – 2 isoforms; digested in human and formula samples
			-β-casein A3 – digested and non-digested in human and formula samples
			-ceruloplasmin – 4 isoforms; digested in human and formula samples
	7/14	Histone H4, antithrombin III	-histone H4 – 19 isoforms, digested in human and formula samples
			-antithrombin III – digested in human and formula samples
	6/14	Lactadherin, κ-casein (**Bos d 12**)	-lactadherin – digested in human and formula samples
	5/14	Monocyte CD14	2 isoforms, digested in human and formula samples
	4/14	C4, lactoperoxidase, α-S2-casein (**Bos d 10**), Ig J	-C4 – digested and non-digested in human samples only
			-Ig J – digested in human and formula samples
	3/14	C7, PGH2 isomerase, CD 59, LPS binding protein	-C7 – 2 isoforms, digested in human and formula samples

Table 2.3: List of pertinent bovine peptides and all NH non-bovine peptides stratified by sample proportion in Fortifiers, HBM, and MAP samples

	2/14	C9, CD 109, C5a	-C9 – 2 isoforms; digested in human and formula samples
			-C5 – digested in human and formula samples
	1/14	Lipocalin (**Bos d 2**), CD 5, CD 81, CD 166, T cell surface CD3	-Bos d 2 – digested in formula samples only
Bos mutus (yak)	14/14	*β*-casein isoform X2	Digested and non-digested in human and formula samples
Felis catus (cat)	13/14	Serum albumin (**Fel d 2**)	All digested in human and formula samples
Equus caballus (horse)	12/14	Serum albumin (**Equ c 3**)	All digested in human and formula samples
Canis lupus familiaris (dog)	11/14	Serum albumin (**Can f 3**)	All digested in human and formula samples
	8/14	Albumin precursor	All digested in human and formula samples
	2/14	Albumin partial, CE1	-CE1: digested in formula samples only
Sus scrofa (pig)	10/14	Serum albumin (**Sus s 1**), albumin partial	Sus s 1 and albumin partial – all digested in human and formula samples
Dermatophagoides farinae (dust mite)	9/14	Heat shock protein (**Der f 28**)	All digested in human and formula samples
	5/14	Triosephosphate isomerase (**Der f 25**)	2 isoforms; digested in human samples only
	1/14	Tubulin α chain	
Aedes aegypti (mosquito)	10/14	Heat shock cognate 70 (**Aed a 8**)	All digested in human and formula samples
Salmo salar (salmon)	5/14	β-enolase (**Sal s 2**)	Digested in human samples only
	4/14	Enolase, aldolase A	-Enolase – digested in human samples only
	3/14	Collagen α (**Sal s 6.0102**)	Digested and non-digested, in formula samples only

Table 2.3: List of pertinent bovine peptides and all NH non-bovine peptides stratified by sample proportion in Fortifiers, HBM, and MAP samples

	2/14	Aldolase A fructose bisphosphate	
Hevea brasiliensis (latex)	8/14	Enolase 1 (**Hev b 9**)	Digested, in human samples only
	2/14	Enolase 2 (**Hev b 2**)	
Procambarus clarkii (crayfish)	7/14	Triosephosphate isomerase (**Pro c 8**)	Digested, in human samples only
Penicillium citrinum (fungus)	4/14	Heat shock 70 kDa protein (**Pen c 19**)	Digested, in human samples only
Blattella germanica (cockroach)	3/14	α-amylase (**Bla g 11**)	Digested, in human samples only
Gallus gallus (chicken/egg)	5/14	β-enolase	Digested, in human samples only
Scylla paramomosain (crab)	4/14	Triosephosphate isomerase	
Prunus dulcis (almond)	3/14	Pru du 6, Pru du 6.0101	
Oreochromis mossambicus (tilapia)	3/14	Tropomyosin	
Thunnus albacares (tuna)	3/14	α-amylase (**Thu a 2.0101**)	Digested, in human samples only
Triticum aestivum (wheat)	2/14	HMW glutenin subunit 1By9, Tri a TPIS, triosephophat-isomerase (**Tri a 31**)	-HMW glutenin subunit 1By9 – digested, in human samples only -Tri a TPIS – digested, in human samples only -Tri a 31 – digested, in human samples only
Curvularia lunata (mold)	1/14	Enolase	
Daucus carota (carrot)	1/14	Peptidyl-prolyl cis trans isomerase	
Catharanthus roseus (periwinkle)	1/14	Peptidyl-prolyl cis trans isomerase	
Ambrosia artemisiifolia (ragweed)	1/14	Putative pectate lyase precursor (**Amb a 1**)	Digested, in human samples only

Proteomic Pilot Study from the MAP samples

After the protocol was validated with the samples from the HMB repository, a subsequent mass spectroscopy run was performed on 5 human milk samples, 1 formula sample, 2 regular milk fortifier samples, and 2 hydrolyzed milk fortifier samples. All the milk samples except for one (J9) were a mix of maternal expressed breast milk or donor breast milk. J9 contained maternal expressed breast milk only.

We performed proteomic and peptideomic analysis on the MAP samples and HMFs. We identified a total of 784 digested non-human proteins, 754 that were bovine in origin and 30 non-bovine in origin. Hydrolyzed HMF samples had significantly fewer proteins (average of 48) compared to non-hydrolyzed HMF (average of 264), human milk (average of 256) and formula (average of 236) (Figure 2.1).
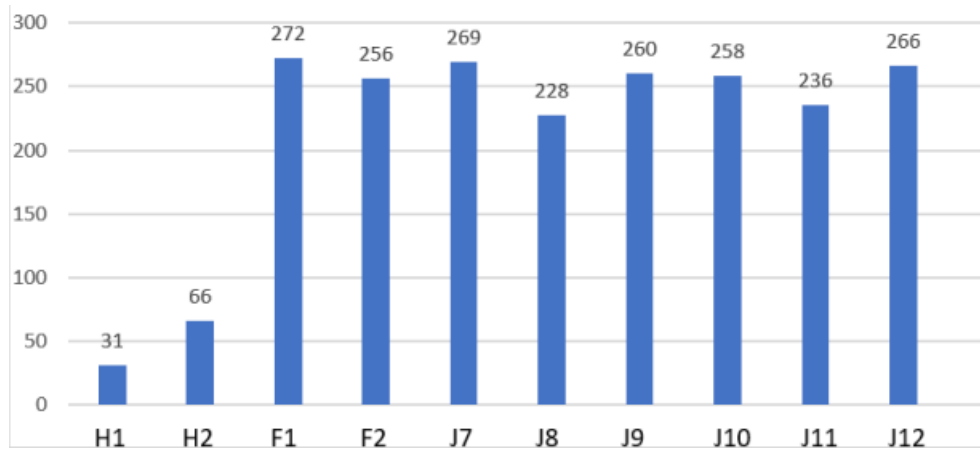
Figure 2.1: Total Non-Human Proteins in MAP Samples. H- Hydrolyzed Human Milk Fortifier, F- Regular Human Milk Fortifier, J11 Enfamil Preterm Formula.

In total, we identified proteins from 23 different species, including aeroallergens, food and contact allergens (Table 2.3). For quality control purposes peptide alignment maps were made for two common allergens, betalactoglobulin and cat albumin. The maps demonstrated consistent

45

overlap, thereby supporting that these proteins were indeed identified and digested similarly between samples. (Supplemental Figure 1)

Bovine peptides were the most numerous of the non-human peptides detected in human milk samples without HMF. Specific allergenic bovine peptides (β-lactoglobulin, α- and β-casein, α-lactalbumin) were in the highest relative quantification in regular fortifier and formula, intermediate in hydrolyzed fortifier and lowest in human milk samples (Supplemental Figure 2).

Peptidomics studies of the MAP samples identified peptides that were generally less than 40 amino acids in length (Figure 2.2). Most non-bovine peptides were digested (28/33) and found in samples with human milk and fortifier which contrasts with only a few (2/28) being found in the formula sample.
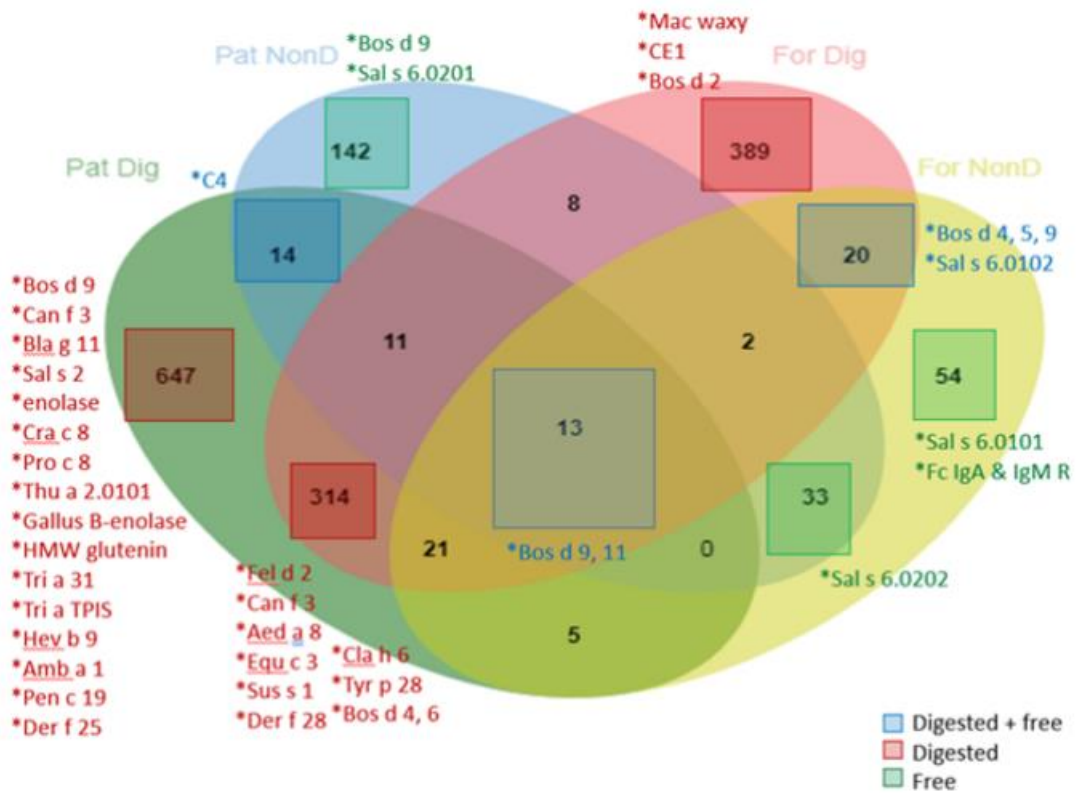
Figure 2.2: Distribution of Free and Digested Peptides in MAP Samples. Pat Dig= Patient samples Digested, Pat NonD= Patient samples Non-digested. For Dig= Formula sample Digested, For NonD= Formula sample Non-digested

Protease activity

Upon analysis of the amino acids at the N- and C- termini of peptides, we discovered that there was a high frequency of proline (P) and glutamine (Q) residues in the MAP samples. This was not surprising because human casein is the major protein in these samples and 17% of residues in this protein are proline while 11% are glutamine. Samples J8-J11 are notable as they have peptides that were cleaved after lysine (K) and arginine (R) at the N-terminal side of peptides. This is not seen for J7 and J12 (Figure 2.3). To quantify protease activity in these human milk samples, we assayed the samples with the fluorogenic substrates, ArgArg-AMC. We found that the most

abundant protease activity was in human milk sample J9, while the formula only sample J11, had

the lowest activity (Figure 2.4).
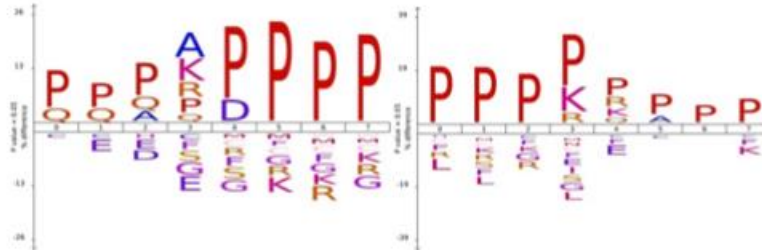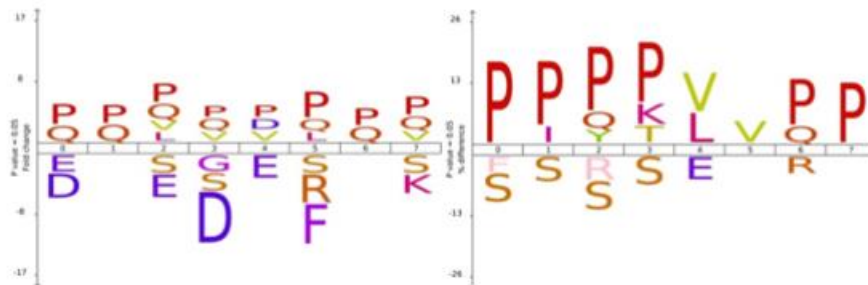
Figure 2.3: Protease Activity in MAP Samples Frequency of amino acids at the 4 positions or either side of the cleaved bond (Positions 0 to 7 where cleavage occurs between 3 and 4). The left image on the tab corresponds to the amino acids found on the Nterminal side of the peptides while the right image corresponds to the amino acids found on the C-terminal side of the peptides. Sample J7 is fortified HBM while J9 is unfortified HBM. Amino Acid Codes: Proline-P, Glutamine-Q, Lysine (K), Arginine (R).
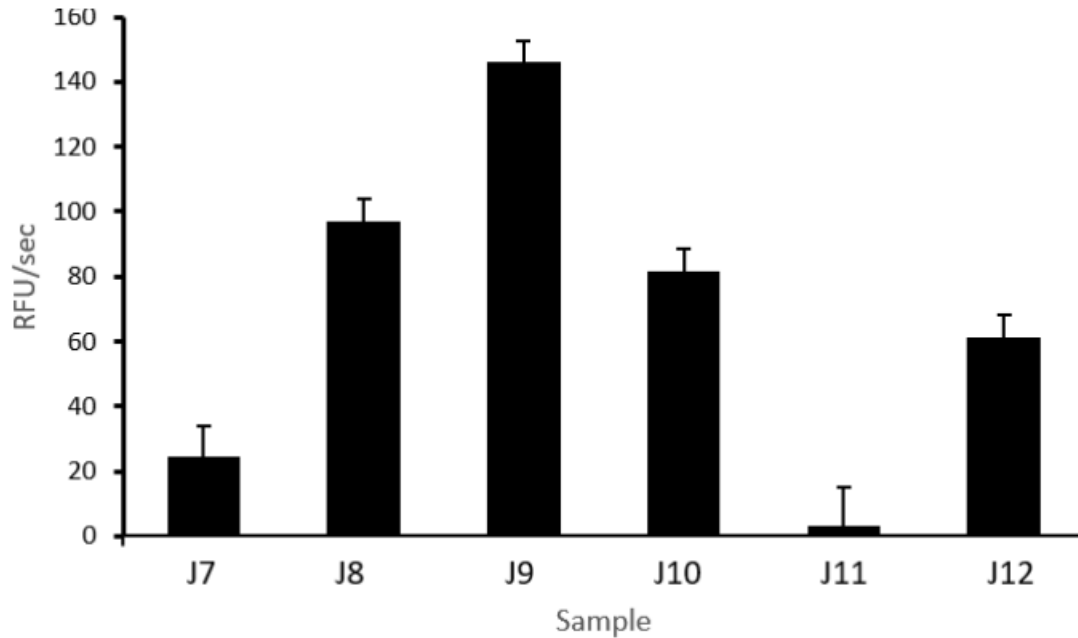
Figure 2.4: Protease activity in MAP Samples. Proteolytic activity in milk samples using the fluorogenic dipeptide substrates Arg-Arg-AMC. J9 Maternal Expressed Breast Milk Only. J11 Fortified Formula

DISCUSSION

To our knowledge, this is the first combined human milk mass spectrometry and protease analysis using clinically relevant NICU preterm milk samples. While the detection of food peptides in the milk samples is interesting and somewhat expected, the size, breadth, and variety of food and aeroallergens as well as other non-food peptides is fascinating, especially when comparing human milk to cow milk formula. Furthermore, we found differential protease activity between the samples with the highest being in maternal expressed breast milk alone, without fortifier (J9) and the lowest in formula (J11).

The presence of allergen peptides in human milk does not appear to be accidental and may be linked to the development of allergy. In one study, there was an increase in atopy in children who were breastfed by atopic mothers and found to have high HM dust mite (Der p 1) levels; this was not noted in the offspring of mothers without atopic history regardless of Der p 1 level in

49

human milk [18]. In food allergy, maternal cow's milk avoidance was associated with increased cow's milk allergy in offspring, mediated by a lower cow's milk specific IgA and possibly the lack of cow's milk protein exposure [23]. In our analysis, bovine peptides were the most numerous of the non-human peptides detected in the human milk only samples (without formula or fortifier) and specific allergenic bovine peptides (β-lactoglobulin, α- and β- caseins, α-lactalbumin) were found in the highest relative quantification in regular fortifier and formula samples and lowest in the human milk samples.

Multiple other common food allergens have been identified in human breastmilk (HBM) studies. Ovalbumin has been detected in HBM in 8.3%-76% of subjects [24,30,31,37], while ovomucoid was identified in 78% of subjects in one study. 30 There appeared to be a dose-response phenomenon between maternal egg intake and infant serology, whereby for each additional egg ingested, HBM ovalbumin concentration increased by 25% and infant egg-specific IgG4 increased 22%. 37 Egg protein (β-enolase) was found in our analysis although specific ovalbumin and ovomucoid peptides were not identified. Regarding peanut protein, one study of 23 lactating females found that after a 50 g oral peanut load, 48% of female subjects' HM samples contained peanut. 14Another small study demonstrated peanut allergen (Ara h 6) in HM that was functional and IgE-reactive as evidenced by in vitro assays and the observation that administration to mice lead to partial oral tolerance [38]. Peanut protein was not identified in our analysis. With wheat protein, gliadin was detected in 67.5% – 100% breast milk samples in two studies [25,39]. Multiple different wheat peptides were identified in our samples, but not gliadin.

In the food diaries associated with the maternal milk samples that were not augmented with fortifier or formula, some foods that were reported as consumed did not show up in the samples. Conversely, in other cases, foods that were not reportedly consumed, did show up in the samples.

While recall bias and ensuing inaccuracy may partially account for these discrepancies, there is also the issue of timing of food consumption with respect to appearance in human milk. Moreover, the capability of excreting specific proteins may vary between mothers and further impact the presence of allergenic proteins, which further complicates attempts at correlating dietary ingestion and breast milk peptides.

We know that antigen-presenting cells introduce processed allergens to T-helper lymphocytes and proceed down a TH2 pathway in allergic conditions. 40 How the allergen is processed, the role of proteases, and the exact conformation of different allergenic proteins in human milk is not known, although the size of the original protein was better elucidated in our study. We demonstrated that many bovine peptides are found digested (original protein size> 40 amino acids) and free (original protein< 40 amino acids), indicating that there are a variety of different parent proteins. These proteins were mostly found shared between the human milk plus fortifier or formula samples. Conversely, human milk samples without fortifier had relatively few digested bovine peptides, supporting that most cow's milk-derived peptides originated from smaller proteins. Interestingly, cow's milk allergy is one of the first to appear in infants' and the majority of those are sensitized to caseins, which may be able to cross the GI border relatively intact as they coagulate in acidic conditions and may be less susceptible to proteolysis [32,41]. A variety of caseins of different sizes were identified in our formula, fortifier, and human milk samples, although the allergenicity of these specific caseins are not definitively established in this current analysis. Additionally, the exact origin of these proteins, although presumably diet-derived, is unknown. As opposed to various sizes, the majority of nonbovine peptides in our human milk samples were digested, thereby originating from peptides over 40 amino acids in length.

Assuming that most of these peptides were generated by proteases, we looked at the amino acid sequence in the protein that ended up being the substate for cleavage. Sample J9 (pure maternal expressed breast milk) from the MAP study showed the highest protease activity. Previous proteomic studies have shown differences in the presence of proteases and protease inhibitors in HM between allergenic and non-allergenic mothers [42]. There is evidence that an imbalance between protease and protease inhibitors in HBM could allow for easier penetration of allergens[43,44]. Specifically, reduced cystatin, a protein inhibitor that has been detected in HM, secreted by epithelial cells has been linked to easier penetration of Der p1 through skin [45]. Furthermore, protease inhibitors have been detected in the stool of infants who have received HBM indicating that these protease inhibitors may be active in the gastrointestinal tract [46]. It is thought this complex interplay between allergens, proteases, and protease inhibitors is important in the pathogenesis of atopy, and protease inhibitors are being evaluated as a potential therapeutic agent to treat asthma and other atopic conditions.

There are several limitations to this study. We started with a small batch of samples to assess feasibility in this pilot trial. There is inconsistency between dietary documentation between the two sample groups. We will have a more consistent and larger sample size in our future analysis. Other sample-based limitations include the lack of multiple "pure" samples that contain only maternal expressed breastmilk without fortifier or the use of pooled donor human milk. Theoretically, a subtractive analysis could be considered, with inference of protein content of breast milk via exclusion of proteins found in fortifier, however, this is limited due to the overlap of proteins between fortifier and human milk and the variability between the samples, including differences in protein content between the two samples of the same fortifier. Moreover, since a large proportion of preterm infants receive supplementation, donor milk, or formula, our results

reflect the real world setting in the NICU. Our subject dietary history did not include the temporal relationship of specific food ingestion and sample collection. Thus, secretion kinetics cannot be concluded, and contamination/inadvertent consumption is an issue with the self-reported dietary histories. Closer analysis of maternal diet and timing of consumption may help to determine the kinetics of human milk peptides and the degree of contamination (dietary or via mass spectrometry) that could account for the detection of proteins that are not found in the diet. Database limitations are also possible. We did not manually blast all proteins against NCBI and Uniprot databases, only those which were positively identified, so it is possible that there were false negatives and proteins were not identified due to inaccurate database sequences.

FUTURE DIRECTIONS AND CONCLUSIONS

We have taken a large step forward in identifying what a preterm infant immune system may encounter in their milk feeds, however, it was beyond the scope of this study to determine the origin of the human milk peptides identified. This is an area we plan to investigate in the future. Peptides may be secreted by lactocytes or enter via the bloodstream. It is also not known where proteolytic cleavage occurs, whether it is locally in the breast or in the GI tract/blood, which could be further investigated by paired blood samples in future studies.

The interaction between allergen, protease, and protease inhibitors also warrants further investigation. Identifying which proteases and protease inhibitors are present in our MAP samples would be of great interest, particularly if their presence or absence augments the development of atopic conditions in infants who have been in the NICU. We do plan to follow subjects out to 5 years and look for the development of allergic outcomes in our MAP cohort. The use of formula, fortifiers, and donor milk are important in optimizing the growth and development of preterm

infants. However, their use may have unintended longterm consequences, that need further investigation.

In conclusion, the detection of various allergenic peptides and protease activity in our milk samples raises more questions about how modifying feeds in the NICU may impact the development of atopy in preterm infants. Ultimately, whether human milk can serve to induce allergic sensitization or tolerance in an infant is an area of research that needs much further exploration.

Acknowledgements

REFERENCES

[1] Björkstén, B., Aït-Khaled, N., Innes Asher, M., Clayton, T. O., Robertson, C., & ISAAC Phase Three Study Group (2011). Global analysis of breast feeding and risk of symptoms of asthma, rhinoconjunctivitis and eczema in 6-7 year old children: ISAAC Phase Three. Allergologia et immunopathologia, 39(6), 318–325. https://doi.org/10.1016/j.aller.2011.02.005

[2] Kull, I., Melen, E., Alm, J., Hallberg, J., Svartengren, M., van Hage, M., Pershagen, G., Wickman, M., & Bergström, A. (2010). Breast-feeding in relation to asthma, lung function, and sensitization in young schoolchildren. *The Journal of allergy and clinical immunology*, *125*(5), 1013–1019. https://doi.org/10.1016/j.jaci.2010.01.051

[3] Gdalevich, M., Mimouni, D., & Mimouni, M. (2001). Breast-feeding and the risk of bronchial asthma in childhood: a systematic review with meta-analysis of prospective studies. *The Journal of pediatrics*, *139*(2), 261–266. https://doi.org/10.1067/mpd.2001.117006

[4] Du Toit, G., Roberts, G., Sayre, P. H., Bahnson, H. T., Radulovic, S., Santos, A. F., Brough, H. A., Phippard, D., Basting, M., Feeney, M., Turcanu, V., Sever, M. L., Gomez Lorenzo, M., Plaut, M., & Lack, G. (2015). Randomized trial of peanut consumption in infants at risk for peanut allergy. *New England Journal of Medicine*, *372*(9), 803–813. https://doi.org/10.1056/nejmoa1414850

[5] Braman S. S. (2006). The global burden of asthma. *Chest*, *130*(1 Suppl), 4S–12S. https://doi.org/10.1378/chest.130.1_suppl.4S

[6] Been, J. V., Lugtenberg, M. J., Smets, E., van Schayck, C. P., Kramer, B. W., Mommers, M., & Sheikh, A. (2014). Preterm birth and childhood wheezing disorders: a systematic review and meta-analysis. *PLoS medicine*, *11*(1), e1001596. https://doi.org/10.1371/journal.pmed.1001596

[7] Silvers, K. M., Frampton, C. M., Wickens, K., Pattemore, P. K., Ingham, T., Fishwick, D., Crane, J., Town, G. I., Epton, M. J., & New Zealand Asthma and Allergy Cohort Study Group (2012). Breastfeeding protects against current asthma up to 6 years of age. *The Journal of pediatrics*, *160*(6), 991–6.e1. https://doi.org/10.1016/j.jpeds.2011.11.055

[8] Raciborski, F., Tomaszewska, A., Komorowski, J., Samel-Kowalik, P., Białoszewski, A. Z., Walkiewicz, A., Lusawa, A., Szymański, J., Opoczyńska, D., Drużba, M., Borowicz, J., Lipiec, A., Kapalczynski, W. J., & Samoliński, B. (2012). The relationship between antibiotic therapy in

early childhood and the symptoms of allergy in children aged 6-8 years - the questionnaire study results. *International journal of occupational medicine and environmental health*, *25*(4), 470–480. https://doi.org/10.2478/S13382-012-0056-0

[9.]Roduit, C., Scholtens, S., de Jongste, J. C., Wijga, A. H., Gerritsen, J., Postma, D. S., Brunekreef, B., Hoekstra, M. O., Aalberse, R., & Smit, H. A. (2009). Asthma at 8 years of age in children born by caesarean section. *Thorax*, *64*(2), 107–113. https://doi.org/10.1136/thx.2008.100875

[10]Fujimura, K. E., Johnson, C. C., Ownby, D. R., Cox, M. J., Brodie, E. L., Havstad, S. L., Zoratti, E. M., Woodcroft, K. J., Bobbitt, K. R., Wegienka, G., Boushey, H. A., & Lynch, S. V. (2010). Man's best friend? The effect of pet ownership on house dust microbial communities. *The Journal of allergy and clinical immunology*, *126*(2), 410–412.e4123. https://doi.org/10.1016/j.jaci.2010.05.042

[11] Munblit, D., Peroni, D. G., Boix-Amorós, A., Hsu, P. S., Van't Land, B., Gay, M., Kolotilina, A., Skevaki, C., Boyle, R. J., Collado, M. C., Garssen, J., Geddes, D. T., Nanan, R., Slupsky, C., Wegienka, G., Kozyrskyj, A. L., & Warner, J. O. (2017). Human Milk and Allergic Diseases: An Unsolved Puzzle. *Nutrients*, *9*(8), 894. https://doi.org/10.3390/nu9080894

[12]Ewan P. W. (1996). Clinical study of peanut and nut allergy in 62 consecutive patients: new features and associations. *BMJ (Clinical research ed.)*, *312*(7038), 1074–1078. https://doi.org/10.1136/bmj.312.7038.1074

[13] Al-Muhsen, S., Clarke, A. E., & Kagan, R. S. (2003). Peanut allergy: an overview. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, *168*(10), 1279–1285.

[14] Vadas, P., Wai, Y., Burks, W., & Perelman, B. (2001). Detection of peanut allergens in breast milk of lactating women. *JAMA*, *285*(13), 1746–1748. https://doi.org/10.1001/jama.285.13.1746

[15] Picariello, G., De Cicco, M., Nocerino, R., Paparo, L., Mamone, G., Addeo, F., & Berni Canani, R. (2019). Excretion of Dietary Cow's Milk Derived Peptides Into Breast Milk. *Frontiers in nutrition*, *6*, 25. https://doi.org/10.3389/fnut.2019.00025

[17] Rekima, A., Bonnart, C., Macchiaverni, P., Metcalfe, J., Tulic, M. K., Halloin, N., Rekima, S., Genuneit, J., Zanelli, S., Medeiros, S., Palmer, D. J., Prescott, S., & Verhasselt, V. (2020). A role for early oral exposure to house dust mite allergens through breast milk in IgE-mediated food allergy susceptibility. *The Journal of allergy and clinical immunology*, *145*(5), 1416–1429.e11. https://doi.org/10.1016/j.jaci.2019.12.912

[18] Baïz, N., Macchiaverni, P., Tulic, M. K., Rekima, A., Annesi-Maesano, I., Verhasselt, V., & EDEN Mother-Child Cohort Study Group (2017). Early oral exposure to house dust mite allergen through breast milk: A potential risk factor for allergic sensitization and respiratory allergies in children. *The Journal of allergy and clinical immunology*, *139*(1), 369–372.e10. https://doi.org/10.1016/j.jaci.2016.07.021

[19] Rodríguez J. M. (2014). The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation?. *Advances in nutrition (Bethesda, Md.)*, *5*(6), 779–784. https://doi.org/10.3945/an.114.007229

[20] Kleinman, R. E., & Walker, W. A. (1979). The enteromammary immune system: an important new concept in breast milk host defense. *Digestive diseases and sciences*, *24*(11), 876–882. https://doi.org/10.1007/BF01324906

[21] Rajani, P. S., Seppo, A. E., & Järvinen, K. M. (2018). Immunologically Active Components in Human Milk and Development of Atopic Disease, With Emphasis on Food Allergy, in the Pediatric Population. *Frontiers in pediatrics*, *6*, 218. https://doi.org/10.3389/fped.2018.00218

[22] Peroni, D. G., Pescollderungg, L., Piacentini, G. L., Rigotti, E., Maselli, M., Watschinger, K., Piazza, M., Pigozzi, R., & Boner, A. L. (2010). Immune regulatory cytokines in the milk of lactating women from farming and urban environments. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, *21*(6), 977–982. https://doi.org/10.1111/j.1399-3038.2010.00995.x

[23].Seppo, A. E., Savilahti, E. M., Berin, M. C., Sampson, H. A., & Järvinen, K. M. (2017). Breast milk IgA to foods has different epitope specificity than serum IgA-Evidence for enteromammary link for food-specific IgA?. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, *47*(10), 1275–1284. https://doi.org/10.1111/cea.12945

[24] Fukushima, Y., Kawata, Y., Onda, T., & Kitagawa, M. (1997). Consumption of cow milk and egg by lactating women and the presence of beta-lactoglobulin and ovalbumin in breast milk. *The American journal of clinical nutrition*, *65*(1), 30–35. https://doi.org/10.1093/ajcn/65.1.30

[25] Chirdo, F. G., Rumbo, M., Añón, M. C., & Fossati, C. A. (1998). Presence of high levels of non-degraded gliadin in breast milk from healthy mothers. *Scandinavian journal of gastroenterology*, *33*(11), 1186–1192. https://doi.org/10.1080/00365529850172557

[26] Axelsson, I., Jakobsson, I., Lindberg, T., & Benediktsson, B. (1986). Bovine beta-lactoglobulin in the human milk. A longitudinal study during the whole lactation period. *Acta paediatrica Scandinavica*, *75*(5), 702–707. https://doi.org/10.1111/j.1651-2227.1986.tb10277.x

[27] van Odijk, J., Kull, I., Borres, M. P., Brandtzaeg, P., Edberg, U., Hanson, L. A., Høst, A., Kuitunen, M., Olsen, S. F., Skerfving, S., Sundell, J., & Wille, S. (2003). Breastfeeding and allergic disease: a multidisciplinary review of the literature (1966-2001) on the mode of early feeding in infancy and its impact on later atopic manifestations. *Allergy*, *58*(9), 833–843. https://doi.org/10.1034/j.1398-9995.2003.00264.x

[28] Järvinen, K. M., Martin, H., & Oyoshi, M. K. (2019). Immunomodulatory effects of breast milk on food allergy. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology*, *123*(2), 133–143. https://doi.org/10.1016/j.anai.2019.04.022

[29] Coscia, A., Orrù, S., Di Nicola, P., Giuliani, F., Varalda, A., Peila, C., Fabris, C., Conti, A., & Bertino, E. (2012). Detection of cow's milk proteins and minor components in human milk using proteomics techniques. *The journal of maternal-fetal & neonatal medicine : the official journal*

of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians, *25 Suppl 4*, 54–56. https://doi.org/10.3109/14767058.2012.715015

[30] Kilshaw, P. J., & Cant, A. J. (1984). The passage of maternal dietary proteins into human breast milk. *International archives of allergy and applied immunology*, *75*(1), 8–15. https://doi.org/10.1159/000233582

[31] Cant, A., Marsden, R. A., & Kilshaw, P. J. (1985). Egg and cows' milk hypersensitivity in exclusively breast fed infants with eczema, and detection of egg protein in breast milk. *British medical journal (Clinical research ed.)*, *291*(6500), 932–935. https://doi.org/10.1136/bmj.291.6500.932

[32] Zhu, J., Garrigues, L., Van den Toorn, H., Stahl, B., & Heck, A. (2019). Discovery and Quantification of Nonhuman Proteins in Human Milk. *Journal of proteome research*, *18*(1), 225–238. https://doi.org/10.1021/acs.jproteome.8b00550

[33] Picariello, G., Addeo, F., Ferranti, P., Nocerino, R., Paparo, L., Passariello, A., Dallas, D. C., Robinson, R. C., Barile, D., & Canani, R. B. (2016). Antibody-independent identification of bovine milk-derived peptides in breast-milk. *Food & function*, *7*(8), 3402–3409. https://doi.org/10.1039/c6fo00731g

[35] Macchiaverni, P., Rekima, A., van den Elsen, L., Renz, H., & Verhasselt, V. (2021). Allergen shedding in Human milk: Could it be key for immune system education and allergy prevention? *Journal of Allergy and Clinical Immunology*, *148*(3), 679–688. https://doi.org/10.1016/j.jaci.2021.07.012

[36] Bandoli, G., Bertrand, K., Saoor, M., & Chambers, C. D. (2020). The Design and Mechanics of an Accessible Human Milk Research Biorepository. *Breastfeeding medicine : the official journal of the Academy of Breastfeeding Medicine*, *15*(3), 155–162. https://doi.org/10.1089/bfm.2019.0277

[37] Palmer, D. J., Gold, M. S., & Makrides, M. (2005). Effect of cooked and raw egg consumption on ovalbumin content of human milk: a randomized, double-blind, cross-over trial. *Clinical and*

experimental allergy : journal of the British Society for Allergy and Clinical Immunology, *35*(2), 173–178. https://doi.org/10.1111/j.1365-2222.2005.02170.x

[38] Bernard, H., Ah-Leung, S., Drumare, M. F., Feraudet-Tarisse, C., Verhasselt, V., Wal, J. M., Créminon, C., & Adel-Patient, K. (2014). Peanut allergens are rapidly transferred in human breast milk and can prevent sensitization in mice. *Allergy*, *69*(7), 888–897. https://doi.org/10.1111/all.12411

[39] Troncone, R., Scarcella, A., Donatiello, A., Cannataro, P., Tarabuso, A., & Auricchio, S. (1987). Passage of gliadin into human breast milk. *Acta paediatrica Scandinavica*, *76*(3), 453–456. https://doi.org/10.1111/j.1651-2227.1987.tb10498.x

[40] Ozdemir, C., Akdis, M., & Akdis, C. A. (2010). T-cell response to allergens. *Chemical immunology and allergy*, *95*, 22–44. https://doi.org/10.1159/000315936

[41] Zhu, J., Garrigues, L., Van den Toorn, H., Stahl, B., & Heck, A. (2019). Discovery and Quantification of Nonhuman Proteins in Human Milk. *Journal of proteome research*, *18*(1), 225–238. https://doi.org/10.1021/acs.jproteome.8b00550.

[42] Hettinga KA, Reina FM, Boeren S, Zhang L, Koppelman GH, Postma DS, et al. (2015) Difference in the Breast Milk Proteome between Allergic and Non-Allergic Mothers. PLoS ONE 10(3): e0122234. https://doi.org/10.1371/journal.pone.0122234

[43] Smith, P. K., & Harper, J. I. (2006). Serine proteases, their inhibitors and allergy. *Allergy*, *61*(12), 1441–1447. https://doi.org/10.1111/j.1398-9995.2006.01233.x

[44] Gregory, L. G., & Lloyd, C. M. (2011). Orchestrating house dust mite-associated allergy in the lung. *Trends in immunology*, *32*(9), 402–411. https://doi.org/10.1016/j.it.2011.06.006

[45] Cork, M. J., Danby, S. G., Vasilopoulos, Y., Hadgraft, J., Lane, M. E., Moustafa, M., Guy, R. H., Macgowan, A. L., Tazi-Ahnini, R., & Ward, S. J. (2009). Epidermal barrier dysfunction in

atopic dermatitis. *The Journal of investigative dermatology*, *129*(8), 1892–1908. https://doi.org/10.1038/jid.2009.133

[46] Davidson, L. A., & Lönnerdal, B. (1990). Fecal alpha 1-antitrypsin in breast-fed infants is derived from human milk and is not indicative of enteric protein loss. *Acta paediatrica Scandinavica*, *79*(2), 137–141. https://doi.org/10.1111/j.1651-2227.1990.tb11429.x

CHAPTER 3

ABSTRACT

Antivirals that specifically target SARS-CoV-2 are needed to control the COVID-19 pandemic. The main protease (M^{pro}) is essential for SARS-CoV-2 replication and is an attractive target for antiviral development. Here we report the use of the Random nonstandard Peptide Integrated Discovery (RaPID) mRNA display on a chemically cross-linked SARS-CoV-2 M^{pro} dimer, which yielded several high-affinity thioether-linked cyclic peptide inhibitors of the protease. Structural analysis of M^{pro} complexed with a selenoether analogue of the highest-affinity peptide revealed key binding interactions, including glutamine and leucine residues in sites S_1 and S_2, respectively, and a binding epitope straddling both protein chains in the physiological dimer. Several of these M^{pro} peptide inhibitors possessed antiviral activity against SARS-CoV-2 *in vitro* with EC_{50} values in the low micromolar range. These cyclic peptides serve as a foundation for the development of much needed antivirals that specifically target SARS-CoV-2.

INTRODUCTION

The COVID-19 pandemic, caused by infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused widespread morbidity and mortality as well as devastation to the global economy since the disease was first reported in late 2019 in Wuhan, China[1]. At the time of writing there has been more than 200 million confirmed cases and 4.3 million deaths worldwide as a result of COVID-19 [2]. There has been significant effort from the global research community to develop effective vaccines for COVID-19; this has been enormously successful, with adenoviral vectored vaccines, protein vaccines and mRNA vaccines now in

widespread use across the world. Whilst vaccines will enable protective immunity in most, there will be populations where vaccine-based immunity may fail, and these individuals will be vulnerable to SARS-CoV-2 infection in the future. It is furthermore unclear what changes will appear in the virus in contemporary SARS-CoV-2 viral variants (highlighted by the recent emergence of the delta variant B.1.617.2)[3], and how those variants will navigate both convalescent and vaccine immune responses. Given that this is the third coronavirus that has crossed via zoonoses, antiviral development against SARS-CoV-2, and future coronaviruses with pandemic potential, are desperately needed in addition to prophylactic vaccines.

While there have been significant efforts toward the discovery of effective antivirals for SARS-CoV-2, the vast majority of molecules that have completed (or are currently being assessed in) clinical trials were originally developed for other infectious and inflammatory disease indications and are being repurposed for COVID-19. For example, remdesivir, originally trialed for Ebola, is currently the only antiviral drug to be approved by the U.S. Food and Drug Administration (FDA) for the treatment of COVID-19. While the molecule has been shown to possess some activity during early infection, it has shown limited to no efficacy in a number of trials[4,5] as well as in infections in patients hospitalized with COVID-19[6]. Other repurposed antiviral drugs that have entered trials include the HIV combination therapy lopinavir-ritonavir[7,8], type I interferon treatments [9,10], and the antimalarial hydroxychloroquine [11–13]; however, these have not demonstrated improvement in disease progression over standard care. In fact, it has recently been suggested that many repurposing efforts may be compromised by experimental artefacts reflecting the physicochemical properties of certain drugs rather than specific target-based activities [14]. To date, the most effective therapeutic intervention for improving COVID-19 patient outcomes in a hospital setting is the use of the corticosteroid dexamethasone, which reduces

inflammation-mediated lung injury associated with SARS-CoV-2 infection in patients with elevated levels of C-reactive protein [15,16]. Based on the above, there is an urgent need for the discovery of effective antivirals for COVID-19, ideally with mechanisms of action that specifically target proteins critical in the SARS-CoV-2 lifecycle.

Infection of human cells by SARS-CoV-2 is initiated by interaction between the receptor binding domain of the trimeric viral spike protein (S) with the host cell-surface receptor angiotensin converting enzyme 2 (ACE2) (Figure 3.1). Following receptor binding of the virus, the spike protein is activated by cleavage between the S1 and S2 domains leading to host cell entry via two distinct pathways: 1) an endocytic pathway through endosomal–lysosomal compartments with spike cleavage facilitated by lysosomal cathepsins, or 2) a cell surface pathway following activation by a serine protease such as TMPRSS2 [17–19]. Following proteolysis, the N-terminus of the cleaved S2 domain is embedded into the cell membrane and leads to fusion of the membranes of the virus and the host cell, followed by transfer of viral RNA into the cytoplasm [20].
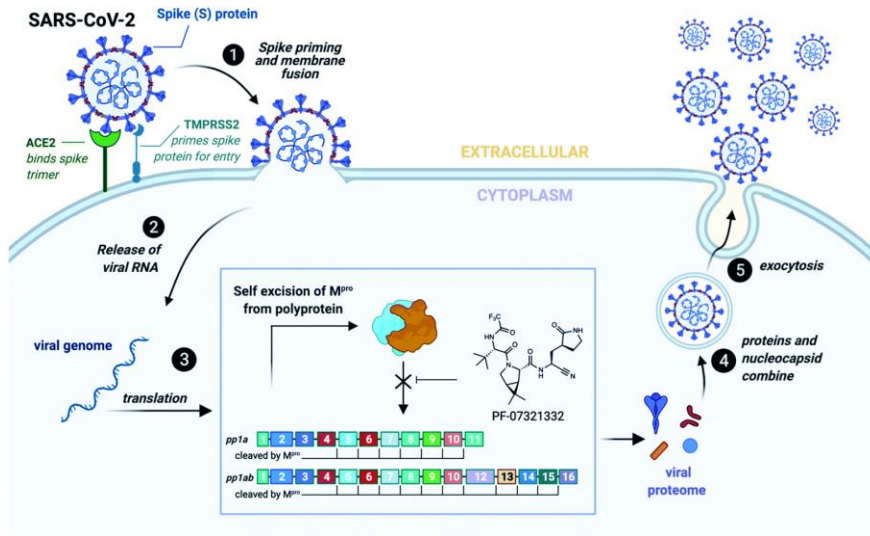
Figure 3.1: Mechanism of SARS-CoV-2 entry into host cells and replication (cell-surface entry pathway mediated by TMPRSS2 shown). The proteolytic activity of SARS-CoV-2 Mpro on the two polyproteins pp1a and pp1ab is shown in the box, including the structure of the covalent peptidomimetic Mpro inhibitor PF-07321332 developed by Pfizer.

Viral gene expression within the host cell results in the translation of two overlapping polyproteins, pp1a and pp1ab. Embedded within these polyproteins are sixteen non-structural proteins critical for viral replication, the majority of which form the viral replication and transcription complex (RTC) including the viral RNA-dependent RNA polymerase (RdRp, nsp12) and helicase (nsp13)[20]. These proteins become functional only after proteolytic release by two viral proteases. The first of these is a domain of nsp3 called the papain-like protease (PL^pro) which cleaves the pp1a and pp1ab at three sites, releasing nsp1, nsp2 and nsp3[20,21]. The second is the SARS-CoV-2 main protease (M^pro), also called nsp5 or the chymotrypsin-like protease (3CL^pro), which cleaves pp1a and pp1ab at a minimum of 11 distinct cleavage sites to release nsp4-16 (Figure 3.1)[20]. Interestingly, M^pro has also been found to aid in immune evasion by inhibiting type I IFN production, contributing to the impaired type I IFN response that has become a hallmark of severe SARS-CoV-2 infection, with persistent viral load and poor patient outcomes [22–25]. M^pro forms a catalytically active homodimer which cleaves with high specificity at Leu-Gln↓Xaa

(where ↓ represents the cleavage site and Xaa can be Ser, Ala or Asn)[26–28]. Such sequence specificity has not been observed for any human proteases and therefore peptide or peptidomimetic based inhibitors are predicted to inhibit SARS-CoV-2 M^{pro} with high selectivity and with minimal off-target effects in humans [20,29].

The key role of M^{pro} for the replication and viability of SARS-CoV-2 has naturally led to the search for novel inhibitors of the protease. Perhaps the most promising of these are the peptidomimetic compounds developed by Pfizer, inspired by PF-00835231 (IC_{50} = 4-8 nM against SARS-CoV-2 M^{pro}) that was originally developed against SARS-CoV-1 M^{pro}, which possesses high homology to the SARS-CoV-2 protease [30–32]. Specifically, a phosphate prodrug of this inhibitor (PF07304814) has recently completed a phase 1b trial (clinical trials identifier NCT04535167). A second-generation orally available peptidomimetic M^{pro} inhibitor developed by Pfizer (PF-07321332, Figure 3.1) has also recently entered phase 2 clinical trials for treatment of COVID-19 (clinical trials identifier NCT04960202) [33]. Both molecules possess a γ-lactam mimic of the glutamine (Gln) residue found at the P1 position in physiological cleavage sites and also inhibit SARS-CoV-2 M^{pro} via a covalent mechanism through electrophilic warheads embedded within the inhibitors [30]. Specifically, PF-00835231 possesses a hydroxymethyl ketone moiety, while PF-07321332 contains a nitrile warhead, both of which react with the catalytic cysteine (Cys145) to inactivate the protease. In addition to these clinical candidates, a number of other peptidomimetic inhibitors of SARS-CoV-2 M^{pro} are currently under preclinical investigation [34–37], including repurposed drugs such as boceprevir, a serine protease inhibitor approved in 2011 for the treatment of hepatitis C [35,38,39], and the feline anticoronaviral drug GC376 [35,38].

Macrocyclic peptides are attractive chemotypes for medicinal chemistry efforts due to their ability to bind targets with high affinity and selectivity, whilst exhibiting greater proteolytic

stability and membrane permeability than their linear counterparts [40–42]. In this work we describe several potent cyclic peptide inhibitors of the SARS-CoV-2 M^{pro}, identified through the use of the Random nonstandard Peptide Integrated Discovery (RaPID) technology, which couples mRNA display with flexizyme-mediated genetic code reprogramming [42–44]. Importantly, we also report a crystal structure of the SARS-CoV-2 M^{pro} dimer bound to our most potent cyclic peptide inhibitor that highlights the residues important for binding both at the catalytic site and across the dimer interface. Finally, we demonstrate that three of the cyclic peptides identified exhibited antiviral activity against SARS-CoV-2 *in vitro*, with an additional peptide gaining antiviral activity upon conjugation to a cell penetrating peptide.

METHODS

Plasmid construction

The gene of the SARS-CoV-2 main protease (M^{pro}) [27] was cloned in between the *Nde*I and *Xho*I sites of the T7 vector pET-47b (+). The construct contains the M^{pro} self-cleavage-site (SAVLQ↓SGFRK; arrow indicating the cleavage site) at the N-terminus. At the C-terminus, the construct contains a modified PreScission cleavage site (SGVTFQ↓GP) connected to a His$_6$-tag. All plasmid constructions and mutagenesis were conducted with cloning and QuikChange protocols using mutant T4 DNA polymerase.

Protein expression

Wild-type M^{pro} was expressed in BL21 DE3 cells transformed with the desired plasmid. Protein expression was conducted in a Labfors 5 bioreactor (INFORS HT, Switzerland). After induction, the culture was grown at 18°C overnight for protein expression. Cells were harvested by centrifugation at 5,000 x g for 15 minutes and lysed by passing twice through a Emulsiflex-C5 homogenizer (Avestin, Canada). The lysate was centrifuged at 13,000 x g for 60 minutes and the

filtered supernatant was loaded onto a 5 mL Ni-NTA column (GE Healthcare, USA) equilibrated with binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol). The protein was eluted with elution buffer (binding buffer containing, in addition, 300 mM imidazole) and the fractions were analyzed by 12% SDS-PAGE. PreScission cleavage and TEV cleavage were conducted in binding buffer in the presence of 1 mM DTT with a protein-to-protease ratio of 100:1. Following cleavage of the His6-tag, the buffer was exchanged to 20 mM HEPES-KOH pH 7.0, 150 mM NaCl, 1 mM DTT, 1 mM EDTA. All samples were analyzed by mass spectrometry using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific, USA) coupled with an UltiMate S4 3000 UHPLC (Thermo Scientific, USA).

Protein cross-linking

C-terminally His-tagged SARS-CoV-2 M^{pro} (25 µM) was incubated with disuccinimidyl glutarate (H6/D6, Creative Molecules Inc.) (250 µM) in aqueous buffer (20 mM HEPES pH 7.6, 100 mM NaCl) at 37°C for 1 hour before the reaction was quenched by dilution of the protein to 10 µM (monomer concentration) by addition of aqueous buffer containing 20 mM Tris-HCl pH 7.6, 100 mM NaCl. Cross-linking efficiency was analyzed by 12% SDS-PAGE and stained with Sypro™ Ruby.

Proteomics to determine crosslink sites on SARS-CoV-2 M^{pro}

Crosslinked proteins were digested with trypsin and peptides desalted as described previously [55]. Peptide mixtures were analyzed by LC-MS/MS using data dependent acquisition with HCD fragmentation on a Thermo Orbitrap Eclipse mass spectrometer. Crosslinked peptides were identified with a 2% false-discovery rate using the Byonic search engine (Protein Metrics) and a custom database containing the 3CL protein sequence and common proteomics contaminants (e.g. trypsin, albumin, keratins, etc.). DSG crosslinks and hydrolysis products were allowed for

the 3CL protein only. Carbamidomethylation was specified as a fixed modification on C, while oxidation of methionine, deamidation of N/Q and pyro-Glu for N-terminal Q/E were variable modifications. Plots of crosslinked residues were generated using UCSF Chimera 1.15rc with the 3CL dimer structure (PDB: 6Y2E).

RaPID mRNA display selection

Library preparation and display selection were performed as described previously with slight modifications [44,56–58]. Briefly, an mRNA library comprising an AUG start codon, 4–15 NNS (N = A, C, G or T; S = C or G) codons, a TGC cysteine codon and a 3' region encoding a linking sequence (Gly-Asn-Leu-Ile) with mRNA lengths pooled proportional to theoretical diversity was prepared as described previously. The mRNA library was then ligated to a puromycin-linked oligonucleotide using T4 ligase. The puromycin-linked mRNA was then translated using the PURExpress ΔRF kit (New England Biolabs) with the addition of release factors 2 and 3 per the manufacturer's instructions along with a "Solution A" containing 19 amino acids (-Met) (prepared as previously described) and either $tRNA^{fMet}_{Ini}$-N-chloroacetyl-L-Tyr or $tRNA^{fMet}_{Ini}$-N-chloroacetyl-D-Tyr (25 μM) to initiate translation and facilitate peptide macrocyclization. Translation was performed at 100 μL scale for the first round and 2.5 μL scale for subsequent rounds. Following translation, ribosomal denaturation was performed by addition of EDTA before reverse transcription with RNase H- reverse transcriptase. After exchange into selection buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Tween-20) through G-25 resin the peptide-mRNA:cDNA libraries were incubated with cross-linked linked SARS-CoV-2 M^{pro} (200 nM) immobilized on Dynabeads™ His-tag Isolation & Pulldown (Life Technologies) at 4°C for 30 min after which the beads were washed with ice-cold selection buffer (3 x 100 μL). The beads were then suspended in aqueous 0.1 vol.% Triton™ X-100 (100 μL) and heated to 95°C for 5 min to

recover binding peptides. The recovered cDNA was amplified by PCR, purified by ethanol precipitation, and transcribed with T7 RNA polymerase to yield an enriched mRNA library to enter the following round of selection. Subsequent rounds were performed in the same manner with the addition of six negative selection steps against beads alone to remove bead binding sequences. Sequencing of recovered cDNA from rounds seven through nine was conducted using an iSeq high-throughput sequencer (Illumina).

Cyclic peptide synthesis

For materials and methods, full synthetic details and characterization (HPLC, ESI-MS) of all peptides, see Supplementary Methods and Supplementary Figures 14-19.

SARS-CoV-2 M^{pro} activity assay

The M^{pro} inhibition assay protocol was adapted and modified from [27,47] and was carried out with a reaction volume of 100 µL in 96-well-plates (black, polypropylene, U-bottom; Greiner Bio-One, Austria) using an aqueous buffer composed of 20 mM Tris-HCl pH 7.3, 100 mM NaCl, 1 mM EDTA, 1 mM DTT. All measurements were performed in triplicate. The compounds were incubated with recombinant SARS-CoV-2 M^{pro} for 10 min at 37°C. The enzymatic reaction was initiated by addition of the FRET substrate (DABCYL)-KTSAVLQ↓SGFRKM-E(EDANS)-NH$_2$ (Mimotopes, Australia). The final concentrations for IC$_{50}$ determination amounted to 25 nM M^{pro} and 25 µM substrate, with inhibitor concentrations ranging from nanomolar to micromolar. The final concentrations for K_i determination amounted to 12.5 nM enzyme and 10, 20, 35 and 50 µM substrate, with a control and three inhibitor concentrations ranging from nanomolar to micromolar. The fluorescence signal was monitored by a fluorophotometer (Infinite 200 PRO M Plex; Tecan, Switzerland) for 5 min at 490 nm, using an excitation wavelength of 340 nm. Initial velocities were derived from the linear range of the enzymatic reaction. For IC$_{50}$ determination,

100% enzymatic activity was defined as the initial velocity of control triplicates containing no inhibitor and the percentage of inhibition was calculated in relation to 100% enzymatic activity. An EDANS standard curve generated as described by Ma *et al.* [35] was used to convert fluorescence intensity to the amount of cleaved substrate (calibration curve). Data sets were analyzed with Prism 9.2 (Graph Pad Software, USA) to generate dose-response curves and calculate IC_{50} values, as well as to generate Michaelis-Menten curves and calculate K_i values assuming the appropriate inhibition mode.

Crystallography

To obtain a crystallographic complex of SARS-CoV-2 M^{pro} and **1**, the peptide was dissolved in DMSO to a stock concentration of 50 mM. Protein, prepared as described above, was diluted to 4 mg/mL in TBS (50 mM Tris-HCl pH 7.5, 300 mM NaCl) and incubated with 2.5-fold molar excess (~300 μM) **1** for 2 hours on ice to saturate the binding sites of the protease. The complex was briefly buffer exchanged to remove unbound peptide using an Amicon Ultra 0.5 mL centrifugal filter with a 10 kDa MWCO to a final protein concentration of 4 mg/mL. This complex was used for high-throughput crystallography trials at 18 °C using drop sizes of 0.5 μL protein and 0.5 μL reservoir, which yielded a hit in the Index sparse matrix screen (Hampton Research). This hit was optimized but did not yield crystals of suitable diffraction quality; however, the crystals were able to be crushed and used to seed crystals with better morphology. Serial seeding was performed using a protein concentration of 1 mg/mL for several rounds until crystals of suitable diffracting quality were obtained, which were formed at 18 °C in a drop size of 1 μL reservoir, 1.5 μL protein, and 0.5 μL seed stock against a reservoir solution of 25% PEG 3350, 0.1 M Bis-Tris pH 6.5, 0.3 M NaCl. Crystals were flash frozen without cryoprotecting and diffraction data were collected at 100 K using the MX2 beamline at the Australian Synchrotron. As the diffraction was

71

not adequate to unambiguously model the inhibitor into the crystal structure, we solved the co-crystal structure of the SARS-CoV-2 M^{pro}-**Se-1** complex. The complex was prepared the same as with **1** but with overnight incubation at 4°C. The complex was diluted to a protein concentration of 1 mg/mL and crystallized at 18°C in a drop size of 1 µL reservoir, 1.5 µL protein and 0.5 µL seed stock against a reservoir solution of 22% PEG 3350, 0.1 M Bis-Tris pH 6.0, 0.3 M NaCl. Crystals were seeded using crystal seeds of the SARS-CoV-2 M^{pro}-**1** crystals. Crystals formed as thin plates after ~4–5 days and were flash frozen without cryoprotection.

Diffraction data were collected at 100 K using the MX2 beamline at the Australian Synchrotron [59]. Reflections collected were indexed and integrated using XDS [60] and scaled in Aimless (CCP4) [61]. The phase problem was overcome by molecular replacement in Phaser MR (CCP4) [62], using PDB ID 7JKV as the search model. The structure was refined by iterative rounds of rebuilding in Coot [63], and twin refinement in Refmac [64,65]. The crystal was twinned P21 (-h, -k, (k+l)); initially autoindexing as C221). Data collection and refinement statistics are given in Supplementary Table 3. The final structure was deposited in the Protein Data Bank (PDB: 7RNW).

SARS-CoV-2 Infectivity Assays

HEK293-ACE2-TMPRSS2 cells stably expressing human ACE2 and TMPRSS2 were generated as previously described [66]. A high content fluorescence microscopy approach was used to assess the ability of the cyclic peptide M^{pro} inhibitors to protect cells from SARS-CoV-2 induced cytopathic effects in permissive cells. In brief, the engineered cell line succumbs to viral cytopathic effects after 6 to 18 hours post infection. Cytopathic effects can be enumerated, as cells and their nuclei collapse into large syncytia after 18 hours of viral culture. The remaining cells outside of

the syncytia increase in a dose dependent manner the lower the viral titers are and/or if a viral inhibitor is introduced within the culture.

For testing the cyclic peptides, compounds were initially diluted in cell culture medium (DMEM-5% FCS) to make 4x working stock solutions and then serially diluted further in the above media to achieve a 2-fold dilution series. On the day of the assay, HEK293-ACE2-TMPRSS2 cells were trypsinized, stained with Nucblue in suspension and then seeded at 16,000 cells in a volume of 40 µL of DMEM-5% FCS per well in a 384-well plate (Corning #CLS3985). Diluted compounds (20 µL) were added to the cells and the plates containing cells and compounds incubated for 1 hour at 37°C, 5% CO_2. 20 µL of virus solution at 8×10^3 TCID50/mL [67] was then added to the wells and plates were incubated for a further 24 hours at 37°C, 5% CO_2. Stained cells were then imaged using the InCell 2500 (Cytiva) high throughput microscope, with a 10×0.45 NA CFI Plan Apo Lambda air objective. Acquired nuclei were counted using InCarta high-content image analysis software (Cytiva) to give a quantitative measure of CPE. Virus inhibition/neutralization was calculated as %N= (D-(1-Q))x100/D, where; "Q" is the value of nuclei in test well divided by the average number of nuclei in untreated uninfected controls, and "D"=1-Q for the wells infected with virus but untreated with inhibitors. Thus, the average nuclear counts for the infected and uninfected cell controls get defined as 0% and 100% neutralization respectively. To account for cell death due to drug toxicity, cells treated with a given compound alone and without virus were included in each assay. The % neutralization for each compound concentration in infected wells was normalized to % neutralization in wells with equivalent amount of compound but without the virus to yield the final neutralization values for each condition. Inhibition curves and 50% (EC_{50}) effective concentrations were determined by non-linear regression analysis using GraphPad Prism software (version 9.1.2, GraphPad software, USA).

Cytotoxicity and targeted proteomics on HEK293-ACE2-TMPRSS2 cell line

HEK293-ACE2-TMPRSS2 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 4.5 g/L D-Glucose, L-Glutamine and 110 mg/L sodium pyruvate (Gibco), supplemented with 10% foetal calf serum (Hyclone). Cells were sub-cultured between 70-90% confluency and tested regularly to ensure free of mycoplasma contamination. Cytotoxicity of **1**, **pen-1**, **6** and **pen-6**, was determined by an Alamar Blue HS (Invitrogen) cell viability assay as per manufacturer's instructions. Briefly, 5×10^4 cells were seeded into wells of a 96-well flat bottom culture plate (Corning) and once adhered, compounds or vehicle control (DMSO) were added at varying concentrations and incubated for 24 hours (37°C, 5% CO_2). Alamar Blue HS cell viability reagent (Invitrogen) was added and the cells further incubated for 2-3 hours. Relative fluorescent units (RFU) were determined per well at ex/em 560/590 nm (Tecan Infinite M1000 pro plate reader). Increasing RFU is proportional to cell viability.

For targeted proteomics, cells in 6-well plates at 70% confluency were treated with vehicle control (DMSO) or inhibitors at 10 μM. At various time points (performed in triplicate), the cells were washed 3 times with 2 mL DPBS (Gibco), scraped and lysed in 4% SDC buffer (4% sodium deoxycholate, 0.1 M Tris-HCl pH 8.0) then immediately heated to 95°C for 10 mins before freezing at −30°C prior to processing for targeted proteomics [68]. Cell lysates were digested to peptides as described previously. Peptides were analyzed by LC-MS/MS using a data-independent acquisition method as described previously [69]. Extracted ion chromatograms for fragment ions derived from tryptic peptides were plotted using Xcalibur Qual Browser (Thermo Scientific). The area under the curve for each peak was used for quantification.

RESULTS

Selection against chemically cross-linked SARS-CoV-2 M^pro homodimer

In order to identify cyclic peptide inhibitors of SARS-CoV-2 M[pro], we sought to utilize RaPID, which allows the screening of $>10^{12}$ cyclic peptides for affinity against a protein target of interest immobilized on magnetic beads. However, functional SARS-CoV-2 M[pro] is a homodimer with relatively weak affinity between the monomers ($K_D = 2.5$ μM), and we hypothesized that the protein may exist in a predominantly monomeric (i.e. inactive) form when immobilized on magnetic beads [27]. Consistent with this, we found that the catalytic activity (measured by fluorescent substrate cleavage) of C-terminally His6-tagged M[pro] was significantly diminished (ca. 30%) when immobilized on His Pull-down Dynabeads™ compared to that of the wild-type protein in solution (Supplementary Fig. 1a); this indicated that a significant proportion of the immobilized M[pro] was unable to form an active homodimer. We therefore investigated the use of a chemical cross-linking strategy to covalently lock SARS-CoV-2 M[pro] in the homodimeric state to ensure that RaPID resulted in selection of ligands to the catalytically active form of the protease. The presence of proximal lysine residues near the dimer interface in the crystal structure of SARS-CoV-2 M[pro] (PDB: 6Y2E) prompted us to investigate a number of amine-reactive bifunctional cross-linkers of differing lengths for this purpose [27]. Efficient cross-linking was obtained by disuccinimidyl glutarate (DSG) (Figure 3.2.1 and 3.2.2), with subsequent mapping by LC-MS/MS revealing intermolecular cross-links between lysine residues on the different monomer chains (K97-K97*, K97-K90*). We also observed cross-links between K12 and K97, which are positioned in close proximity on different monomers of the dimer, as well as within the same monomer unit; however, in this case we could not differentiate between intermolecular or intramolecular cross-links by mass spectrometry (Supplementary Fig. 1b). Importantly, cross-linked M[pro] exhibited catalytic activity comparable to that of wild-type M[pro] (in solution), following immobilization on magnetic beads (Supplementary Fig. 1a). Based on these data, we

moved on to RaPID selections against the cross-linked SARS-CoV-2 M^{pro} with a view to discovering novel cyclic peptide inhibitors.
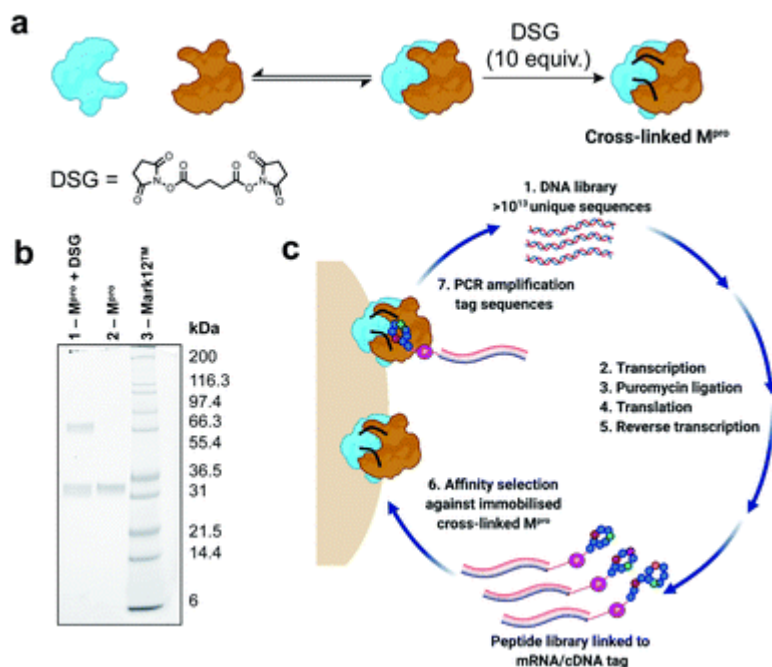
Figure 3.2: (1) Scheme for cross-linking of SARS-CoV-2 Mpro with disuccinimidyl glutarate (DSG). (2) SDS-PAGE gel of cross-linking reaction of Mpro. Lane 1: Reaction of SARS-CoV-2 Mpro (25 µM) with DSG (10 equivalents relative to Mpro monomer) for 1 h at 37°C in aqueous buffer (20 mM HEPES, 100 mM NaCl, pH 7.6). Lane 2: SARS-CoV-2 Mpro. Lane 3: Mark12™ ladder. (3) Workflow for the discovery of macrocyclic peptide ligands of cross-linked SARS-CoV-2 Mpro using RaPID mRNA display.

To select for peptide binders of SARS-CoV-2 M^{pro} we performed parallel display using the RaPID mRNA display technology, employing peptide libraries initiated with either *N*-chloroacetyl-L-tyrosine or *N*-chloroacetyl-D-tyrosine to induce thioether macrocyclization through reaction with the side chain of a downstream cysteine residue. Tyr was chosen as the initiating amino acid due to its high translation efficiency in place of *N*-formyl-Met [45,46], while the use of D-Tyr in addition to L-Tyr expanded the accessible chemical space of the library. For each selection, a semi-randomized DNA library of $>10^{13}$ unique sequences were transcribed into mRNA followed by ligation to a puromycin linker (Figure 3.2.3). Translation of the mRNA-

puromycin hybrids *in vitro*, followed by reverse transcription, yielded a library of peptide-mRNA:cDNA conjugates that were counter-selected against His Pull-down Dynabeads™ (to remove bead-binding peptides) before an affinity selection step, in which the libraries were panned against bead-immobilized cross-linked SARS-CoV-2 M^{pro}. PCR of the bead-bound fraction yielded an enriched DNA library that was used as the starting point for the subsequent rounds of selection (Figure 3.2.3). Following nine iterative rounds of this process, next-generation sequencing of the final DNA library identified several peptide sequences predicted to bind to SARS-CoV-2 M^{pro} based on high relative abundance (Supplementary Fig. 2).

Table 3.1: In vitro inhibition activity of peptides 1 and 6 against other coronaviral proteases. Data are represented as IC50 values ± SEM

Peptide	SARS-CoV-1 M^{pro} IC$_{50}$/μM	MERS-CoV M^{pro} IC$_{50}$/μM	SARS-CoV-2 PLpro IC$_{50}$/μM
1	0.078 ± 0.002	>10	>10
6	>10	>10	>10

Table 3.2: In vitro inhibition activity of peptides 1 and 6 against a selection of human proteases. Data are represented as IC50 values ± SEM

Peptide	TMPRSS2 IC$_{50}$/μM	Furin IC$_{50}$/μM	Cathepsin B IC$_{50}$/μM	Cathepsin L IC$_{50}$/μM	Cathepsin E IC$_{50}$/μM
1	>10	>10	>10	10.9 ± 1.12	>10
6	>10	>10	>10	5.7 ± 0.45	>10

Synthesis and *in vitro* evaluation of cyclic peptide inhibitors of SARS-CoV-2 M^{pro}

We selected eight peptides from the enriched libraries for synthesis (five L-Tyr initiated and three D-Tyr initiated) based on their abundance in the final DNA library and diversity in sequence (Figure 3.3.1 and Figure 3.3.2). Peptides **1**–**8** were synthesized *via* Fmoc-strategy solid-phase peptide synthesis on Rink amide resin with the N-terminal L- or D-Tyr residue derivatized

with chloroacetic acid to facilitate thioether cyclization with the thiol of a downstream cysteine residue. Deprotection and cleavage from resin followed by cyclization provided the target thioether-linked cyclic peptides **1**–**8** following purification by reverse-phase HPLC. The synthetic cyclic peptides were next evaluated for inhibitory activity against SARS-CoV-2 M^{pro} using a previously reported fluorescence resonance energy transfer (FRET)-based assay [27,47]. Gratifyingly, we observed potent inhibition of proteolytic activity for six of the eight peptides in the series, with IC50 values for peptides **1**–**6** ranging from 0.070 – 12.7 μM (Figure 3.3.3). Interestingly, despite being enriched in the selection, lariat peptide **7** and head to tail cyclic peptide **8** did not show appreciable inhibition of the protease (IC$_{50}$ > 50 μM). We next assessed the selectivity of the two most potent inhibitors, 1 and 6, against a panel of coronaviral and human proteases (Tables 3.1 and 3.2). Peptides 1 and 6 were also assessed for stability in human plasma; both exhibited high stability with 85% of 1 and 80% of 6 remaining after 24 hours (see ESI, Fig. S5†).
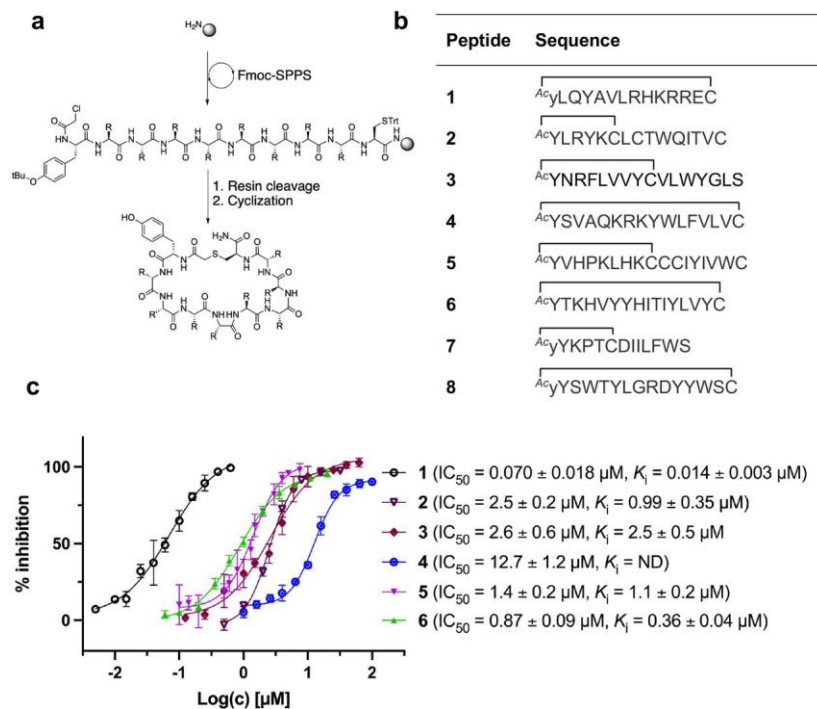
Figure 3.3: (1) Synthesis of cyclic peptides 1–8 via Fmoc-SPPS (see Supplementary methods for full synthetic details). (2) Sequences of peptides 1-8 in one letter amino acid code, thioether cyclization is represented as a black line. (3) In vitro inhibition data of SARS-CoV-2 Mpro for peptides 1–6 with associated IC50 and Ki values ±SEM. NB: cyclic peptides 7 and 8 showed no inhibition of SARS-CoV-2 Mpro at 50 μM, ND = not determined.

The most potent SARS-CoV-2 M^{pro} inhibitor in the series was cyclic peptide **1**, which exhibited an IC_{50} of 70 ± 18 nM and a K_i of 14 ± 3 nM against the protease. Given the superior inhibitory potency of this molecule over the other cyclic peptides that emerged from the RaPID screen (>15-fold), this molecule was selected for further structure-activity investigations.

Probing the binding interaction of lead cyclic peptide 1

We next performed experiments to probe the key interactions of lead cyclic peptide inhibitor **1** with SARS-CoV-2 M^{pro}. Given the critical role dimerization plays in M^{pro} catalytic activity, we first investigated the effect of **1** on the monomer-dimer equilibrium of M^{pro} by multi-angle laser light scattering with size exclusion chromatography (SEC-MALLS). Interestingly, the addition of two molar equivalents of peptide **1** to a solution of wild-type M^{pro} (present in an

79

approximately 1:1 ratio of monomer:dimer) resulted in a shift in the dimer-monomer equilibrium to afford a solution of exclusively homodimeric protease (Figure 3.4.1). This data indicates that cyclic peptide **1** stabilizes the homodimeric form of SARS-CoV-2 M^{pro} and suggests that peptide **1** may bind to the active site present in the dimeric, but not monomeric forms of the protein. To corroborate this data, we used 3D NMR spectroscopy (specifically TROSY-HNCO spectra) to analyze SARS-CoV-2 M^{pro} after titration with peptide **1** (Figure 3.4.2). This revealed shifts of backbone NMR resonances of residues near the active site, consistent with binding of peptide **1** at this location. However, shifts were also observed for residues located far from the active site (Supplementary Fig. 3), suggesting that the protein responds to the binding of **1** with global allosteric changes. In contrast, titration of the obligate monomeric mutant R298A of SARS-CoV-2 M^{pro} [48] with peptide **1** led to no shifts in NMR resonances (Supplementary Fig. 3, see Supplementary Tables 1 and 2 and Supplementary Fig. 4 for full assignment of monomeric M^{pro} mutant). Notably, R298, which is mutated in the obligate monomer, is buried within the dimer interface in the structure of the wild-type protease and therefore the lack of binding of **1** to the monomeric protein is unlikely to be owing to a loss of interactions with this amino acid residue. Taken together, these SEC-MALLS and NMR data therefore suggest that peptide **1** selectively binds to and inhibits the active homodimer of SARS-CoV-2 M^{pro} but does not bind to the inactive monomer.

a

3CLpro WT (1 uM)
3CLpro WT (1 uM) + 1 (2 uM)

Refractive Index (au)

Molecular Weight (kDa)

M^pro peptide 1 dimer
67 kDa

M^pro monomer
36 kDa

volume (mL)

b

c

[M+H]⁺ 1874.46

No M^pro

^{Ac}yLQYAVLRHKRREC

[M+H]⁺ 1874.69

25 nM M^pro

[M+H]⁺ 1874.46

2.5 µM M^pro

[M+H]⁺ 1892.45

H-YAVLRHKRREC ^{Ac}yLQ-OH

m/z

d

Peptide	Sequence	K_i (µM)
1	^{Ac}yLQYAVLRHKRREC	0.014 ± 0.003
1-y1a	^{Ac}aLQYAVLRHKRREC	0.48 ± 0.13
1-L2A	^{Ac}yAQYAVLRHKRREC	>20
1-Q3A	^{Ac}yLAYAVLRHKRREC	>20
1-Y4A	^{Ac}yLQAAVLRHKRREC	0.87 ± 0.09
1-R8A	^{Ac}yLQYAVLAHKRREC	0.094 ± 0.024
1-H9A	^{Ac}yLQYAVLRAKRREC	0.048 ± 0.013
1-K10A	^{Ac}yLQYAVLRHARREC	0.054 ± 0.013
1-R11A	^{Ac}yLQYAVLRHKAREC	1.2 ± 0.3
1-R12A	^{Ac}yLQYAVLRHKRAEC	0.021 ± 0.005
1-E13A	^{Ac}yLQYAVLRHKRRAC	0.19 ± 0.06

Figure 3.4: (1) SEC-MALLS of SARS-CoV-2 Mpro with and without peptide 1. SARS-CoV-2 Mpro (red line) exists in equilibrium between monomeric and homodimeric forms giving rise to two peaks (in ca. 1:1.25 ratio of monomer to dimer) in the size-exclusion chromatogram at a concentration of 1 µM in aqueous buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl). MALLS analysis indicates an approximate molecular weight of 36 kDa for the monomer (calculated molecular weight = 33.8 kDa). After addition of two molar equivalents of 1, Mpro converges predominantly to a homodimer (blue line) with a MALLS reading of 66 kDa (calculated MW of SARS-CoV-2 Mpro dimer = 67.6 kDa, MW of peptide 1= 1874 Da) indicative of formation of the SARS-CoV-2 Mpro homodimer upon binding 1. (2) Cyclic peptide inhibitor 1 binds to the dimeric form of SARS-CoV-2 Mpro. Overlay of projections onto the 15N-1H plane of 3D TROSY-HNCO spectra of 0.3 mM solutions of 15N/13C/2H-labelled wild-type Mpro. Blue and red contour lines show the spectra recorded in the absence and presence of equimolar inhibitor 1, respectively. Assignments are shown for peaks that shift or disappear in response to the inhibitor. (3) Monitoring of the cleavage of cyclic peptide inhibitor 1 with SARS-CoV-2 Mpro. MALDI-TOF mass spectrum of cyclic peptide 1 (top spectrum). Negligible cleavage of 1 was observed following incubation with SARS-CoV-2 Mpro under standard assay conditions (25 nM SARS-CoV-2 Mpro, 5 µM 1, 20 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM DTT, 1 mM EDTA) for 1 h at 37 °C (middle spectrum). Slow cleavage of 1 was observed (ca. ~30% after 1 h) in the presence of a high concentration of SARS-CoV-2 Mpro to 2.5 µM (bottom spectrum). (4) Inhibitory activity of

alanine mutants of lead cyclic peptide 1. Sequences and associated inhibitory constants for peptide 1 analogues, whereby all polar residues within the randomized region of 1 were each systematically mutated to alanine and their inhibitory activity against SARS-CoV-2 Mpro assessed (Cys14 was not mutated as this residue is required for cyclization, Ala substitutions are shown in bold).

Cyclic peptide **1** includes the cleavage recognition motif of M^{pro} found in the natural viral protein substrates, namely a Gln in the P_1 position and a Leu in the P_2 position [27]. It was therefore hypothesized that this Leu-Gln motif embedded within **1** mimics the natural substrate and binds to the catalytic site of the protease. Importantly, this proposal is supported by the purely competitive inhibition mode observed for **1** (Supplementary Fig. 5). However, this also raises the possibility that the protease may be able to cleave **1** next to the Gln-Leu recognition sequence. To test this, we incubated **1** with wild type SARS-CoV-2 M^{pro} and assessed cleavage by MALDI-TOF-MS (Figure 3.4.3). Peptide **1** exhibited notable resistance to proteolysis by M^{pro} with negligible cleavage of peptide **1** observed under standard assay conditions (25 nM M^{pro}) after 1 h. However, incubation of **1** with a high concentration of M^{pro} (2.5 μM) resulted in slow cleavage of peptide **1**, with 30% peptide cleavage observed after 1 h. Analysis of the cleavage reaction by LC-MS/MS confirmed that proteolysis had indeed occurred between Gln3 and Tyr4 (Supplementary Fig. 6a and 6b). We also synthesized an authentic standard of the resulting cleavage product, which was verified to be identical to M^{pro}-cleaved **1** by LC-MS/MS (Supplementary Fig. 6b). Finally, we assessed whether the linear peptide product resulting from M^{pro} cleavage of **1** possessed inhibitory activity against SARS-CoV-2 M^{pro}. Interestingly, the peptide exhibited an IC$_{50}$ of 23.2 ± 5 μM, ~330-fold higher than the IC$_{50}$ of 70 nM for **1** (Supplementary Fig. 6c). This two-orders of magnitude loss in activity upon linearization suggests that the conformation of cyclic peptide **1** is pre-organized for optimal interaction with the protease.

In order to assess the importance of each residue in **1** for M^{pro} inhibitory activity, we systematically replaced all residues in the peptide with alanine (except alkyl side chain-containing amino acids Ala5, Val6 and Leu7) and determined the inhibitory activity of the resulting mutants against M^{pro} (Figure 3.4.4, Supplementary Fig. 7). Consistent with the known recognition sequence for SARS-CoV-2 M^{pro} and supported by the mass spectrometry results described above, mutation of either Leu2 or Gln3 to Ala (that would be predicted to bind in the S2 and S1 recognition sites, respectively) led to more than two orders of magnitude reduction in inhibitory activity. Remarkably, mutation of Tyr4 (which would be predicted at P_1') also led to a significant loss in inhibitor activity (IC_{50} = 1.9 µM); while this is consistent with the established importance of aromaticity in P_1', small residues such as alanine are often found in substrates at this position, and the dramatic reduction in inhibitory potency was therefore unexpected [27,28]. Interestingly, mutation of Arg11, which is distal from the most prominent recognition residues of the cyclic peptide, also led to a significant reduction in inhibitory activity (IC_{50} = 3.4 µM) suggesting that this residue makes important interactions with the protease and/or serves a crucial role in the adoption of the active conformation of the cyclic peptide. In contrast, mutation of Arg8, His9, Lys10, Arg12 or Glu13 resulted in equipotent activity or only a modest reduction in inhibitory activity (IC_{50} values = 90–390 nM).

Co-crystallization of peptide 1 and a selenoether analogue with SARS-CoV-2 M^{pro}

In order to further interrogate the binding mode of inhibitor **1**, we used X-ray diffraction to solve the co-crystal structure of the SARS-CoV-2 M^{pro}-**1** complex to 3.4 Å. The limited resolution of the structure hindered the interpretation of the electron density within the active site. We therefore also solved the structure of a synthetic selenoether analogue of **1** (**Se-1**) in complex with SARS-CoV-2 M^{pro}, in which the cysteine residue involved in thioether peptide

83

macrocyclization was replaced with selenocysteine [49]; we rationalized that the greater electron density of Se in this analogue would aid its placement in the active site. Importantly, **Se-1** displayed identical inhibitory activity as **1** (K_i of **Se-1** = 17 nM vs 14 nM for 1), suggesting its interactions with the protein are also very similar to parent inhibitor **1** (Supplementary Fig. 8). A crystal of this SARS-CoV-2 M^{pro}-**Se-1** complex diffracted to 2.19 Å resolution in the same space group and with similar cell dimensions as the SARS-CoV-2 M^{pro}-**1** complex, thus providing a detailed view of the interactions between the inhibitor and the protease (Supplementary Table 3). The crystal dimensions and packing are unique and not observed in any of the hundreds of deposited SARS-CoV-2 M^{pro} structures, with a notably long *b* axis. There were four protein subunits within the asymmetric unit comprising two physiological dimers (Figure 3.5.1).
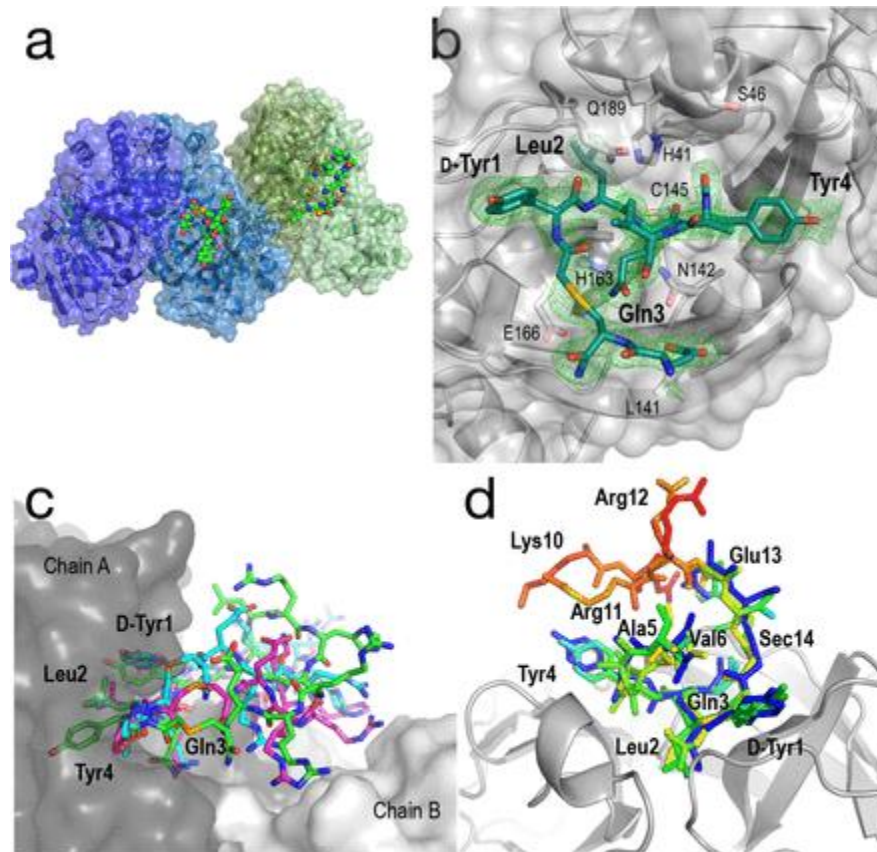
Figure 3.5: Structural analysis of the SARS-CoV-2 Mpro-Se-1 complex (PDB ID: 7RNW). (1) The asymmetric unit of the SARS-CoV-2 Mpro-Se-1 crystal structure, containing two physiological dimers shown as blue and green. The Se-1 peptide is shown in a sphere representation. (2) Zoom of the active site with electron density (Polder omit map, contoured to 3.5σ) showing bound peptide (residues 7–12 disordered). Key active site residues of SARS-CoV-2 Mpro are shown as sticks, and key positions of the Se-1 peptide (residues 1-4) are labelled in bold text. (3) Representative conformations (300 ns) from each of the triplicate simulations suggest that Tyr1, Leu2, Gln3 and Tyr4 are stably bound in the active site of SARS-CoV-2 Mpro, while the remainder of the peptide is more mobile and makes transient interactions across the dimer interface. (4) Se-1 peptide bound to chains A, C and D coloured by B-factor, showing the residues 1-4 are stable and bound within the S3, S2, S1 and S-1 subsites, respectively, whereas the polar half of the peptide (Arg8-Glu13) is either too disordered to accurately model, or is modelled with high B-factors.

Although the peptide is not fully resolved in the structure, 9-13 residues were observable in various chains with good electron density; the selenoether linkage and the first 5 residues (D-Tyr1, Leu2, Gln3, Tyr4, Ala5) were very stable, with the remainder of the peptide (7-12) appearing

somewhat disordered (Figure 3.5.2). The peptide displayed a consensus binding pose across three subunits of the crystal structure. The fourth subunit displayed an alternative binding pose in which Leu2 and Gln3 were present in identical positions, but the flanking D-Tyr1 and Tyr4 residues adopted alternative conformations (Supplementary Fig. 9a and 9b). Closer inspection revealed a non-Pro *cis*-peptide bond between Tyr4 and Ala3 in chains A, C, D, which results in a tight kink in the helix, while in chain B this bond remains in the *trans*-configuration (Supplementary Fig. 9c and 9d). The Gln3 of the peptide occupies the canonical S_1 subsite of the protease (following the numbering by Lee *et al*.) [50] in every peptide:protein complex, facilitated by interactions with His163, Glu166, and Asn142 (Figure 3.5.2). Likewise, Leu2 is always bound in the S_2 subsite comprised of His41, His164 and Gln189. D-Tyr1 is in the solvent-exposed S_3 site bordered by Gln189, Ala191 and Pro16, while the peptide twists at the selenoether/D-Tyr1 linkage, turning away from the canonical S_4 site to be positioned adjacent to Glu166. Tyr4 is therefore positioned in the S_1' (P_1') position in ¾ chains, fully consistent with the slow proteolysis between Gln3 and Tyr4 observed by mass spectrometry (Figure 3.4.3). In chain B, the peptide backbone and Tyr4 have swapped positions (Supplementary Fig. 9a and 9b).

To obtain a better understanding of the interaction of the full peptide with the protein in the absence of crystal contacts, we modelled the missing residues as accurately as possible with the available density and used this structure as the starting point for triplicate 333 ns molecular dynamics (MD) simulations [51] (~1 μs total simulation time; Figure 3.5.3, Supplementary Fig. 10). The results were consistent with the crystal structure, i.e. residues 1-4 were all relatively stable within their respective binding pockets within the substrate cleft of SARS-CoV-2 M^{pro}, with the mobility of the peptide increasing on either side of these residues (Figure 3.5.4). Notably, the

simulations showed regular transient interactions between the peptide and the second chain of the physiological dimer.

`The structure and MD simulations also provide a molecular explanation for the inhibitory activity of the alanine mutants. Specifically, the structure shows that the central interactions are formed by Leu2 and Gln3, consistent with the large reductions in activity when these positions are mutated to Ala, while D-Tyr1 and Tyr4 also form significant interactions with the protease on either side of these residues. Other positions (8, 9, 10, 12 and 13) that were observed to have little influence on inhibitory activity are either disordered or solvent-exposed. MD simulations suggest that Arg11, the mutation of which to Ala had a significant effect on inhibition (Figure 3.4.4), makes contacts across the dimer interface (Supplementary Fig. 9e), interacting with the side chain and main chain carbonyl of Q256*. This is consistent with electron density showing Arg11 within hydrogen bonding distance of the neighboring chain B in one dimer (Supplementary Fig. 9f). These observations are consistent with the SEC-MALLS and NMR data that shows the peptide binds the dimeric form of SARS-CoV-2 M^{pro} exclusively for its mode of inhibition (Figure 3.4.1 and Figure 3.4.2).

Antiviral activity of cyclic peptide SARS-CoV-2 M^{pro} inhibitors

Given the promising inhibitory activity of cyclic peptides **1–6** against SARS-CoV-2 M^{pro}, we next assessed the antiviral activity of the four most potent peptides against SARS-CoV-2 *in vitro*. Specifically, we used ACE-2 and TMPRSS2 overexpressing HEK293T cells infected with SARS-CoV-2 to assess each peptide. The latter cell line is hyper-permissive to SARS-CoV-2 infection with extensive viral syncytia forming after 18 hours post viral infection. Dose-dependent sigmoidal inhibition curves can be generated through enumeration of remaining single cell nuclei using a live nuclear dye and a standard high content fluorescent microscope. In this setting, we

observed a dose-dependent protection for three of the four peptides, with EC_{50} values ranging from 11.8 – 33.1 μM. However, peptide **1**, the most active compound in the enzyme inhibition assay, was inactive up to a concentration of 50 μM, most likely due to low cell permeability owing to the large number of polar and charged residues in this molecule. To facilitate cell entry of **1**, we therefore conjugated its C-terminus to penetratin, a 16 amino acid cell-penetrating peptide (CPP) derived from the *Drosophila* Antennapedia homeodomain [52] (to afford **pen-1**). It should be noted that during the selection process the C-terminus of the peptides is conjugated to a large mRNA:cDNA tag, and therefore addition of a CPP to the C-terminus of the peptide (to afford **pen-1**) was deemed unlikely to affect binding and inhibition of M^{pro}. Pleasingly, **pen-1** had equivalent *in vitro* inhibitory activity against SARS-CoV-2 M^{pro} to peptide **1** (K_i = 9 nM for **pen-1** vs 14 nM for **1**, Supplementary Fig. 11) indicating that addition of a CPP to the C-terminus of the peptide did not affect binding and inhibition of the protease. We also prepared a penetratin conjugate of cyclic peptide **6**, the second most active peptide in the enzyme inhibition assay (K_i = 360 nM). While parent cyclic peptide **6** showed antiviral activity in the cell-based assay (EC_{50} of 33.1 μM), we anticipated that the conjugation to penetratin might further improve potency. We also assessed the cytotoxicity of **1**, **pen-1**, **6** and **pen-6** on HEK293-ACE2-TMPRSS2 cells, with none of the compounds affecting cell viability at a concentration of 50 μM (Supplementary Fig. 12). Pleasingly, we found that **pen-1** and **pen-6** both exhibited significantly improved antiviral activity, with EC_{50} values of 15.2 μM and 6.6 μM, respectively (Figure 3.6). As expected, the marked improvement in antiviral activity of peptide **1** by conjugation to the penetratin CPP correlated with enhanced cellular uptake, whereby LC-MS/MS analysis showed a 5.5-fold increase in levels of **pen-1** compared to **1** in cell lysates (Supplementary Fig. 13).
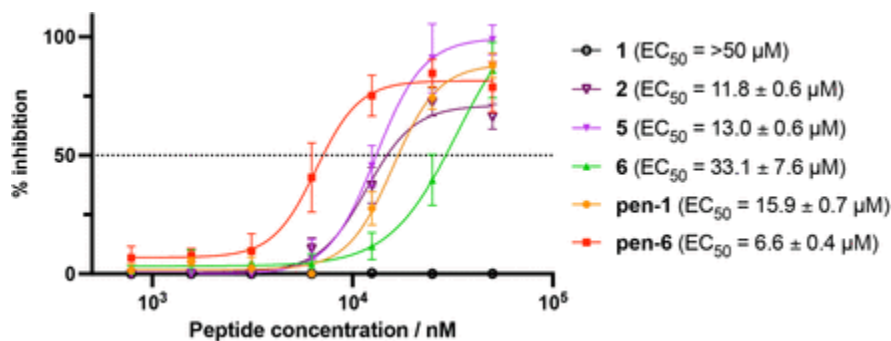
Figure 3.6: Antiviral activity of peptides 1, 2, 5, 6 and penetratin conjugates of 1 and 6 (pen-1 and pen-6). HEK293-ACE2-TMPRSS2 cells were incubated with varying concentrations of cyclic peptide Mpro inhibitors and infected with SARS-CoV-2. Inhibition curves and 50% effective concentrations (EC50) were determined by non-linear regression analysis using GraphPad Prism. Data are the means ± SD of experiments performed in quadruplicate.

DISCUSSION

In summary, we discovered macrocyclic peptide inhibitors of SARS-CoV-2 M^{pro} using RaPID mRNA display technology. A protein cross-linking strategy was employed to generate a covalently locked catalytically active dimer, facilitating display selection against solid-supported SARS-CoV-2 M^{pro}. This enabled the discovery of potent inhibitors of M^{pro}, including the 14-residue cyclic peptide 1 with a K_i of 14 nM that represents one of the most potent inhibitors reported against the protease to date. SEC-MALLS and NMR spectroscopic studies revealed that this molecule inhibits SARS-CoV-2 M^{pro} by exclusively binding to the catalytic protease dimer, highlighting the importance of controlling the nature of the immobilized protein construct for display selections; in this case by covalently cross-linking the M^{pro} homodimer. Co-crystal structures of M^{pro} and **1** and the selenocysteine derivative **Sec-1** solved by X-ray crystallography revealed the canonical interaction between a Gln residue and subsite S_1, flanked by hydrophobic residues. This half of the peptide was relatively stable, while the hydrophilic face (8–12) was relatively mobile, interacting with solvent and across the dimer interface to facilitate an exclusive

dimer binding mode for the inhibitor. The strain imposed by cyclization prevents the peptide from adopting a fully relaxed conformation and likely explains the very slow turnover, despite the inhibitor possessing canonical residues that bind to the key recognition subsites of the protease. Thus, cyclization of the peptide has essentially converted a peptide substrate into a highly potent SARS-CoV-2 M^{pro} inhibitor, highlighting a key benefit of the cyclic peptides that emerge from RaPID screens in medicinal chemistry studies. Several of the cyclic peptides also exhibited antiviral activity against SARS-CoV-2 *in vitro* and enhancing cellular uptake of cyclic peptides by conjugation to the CPP penetratin led to dramatic improvements in antiviral activity. This was especially pronounced for lead molecule **1**, which was ineffective without the covalent CPP tag but exhibited an EC$_{50}$ of 15.9 μM following fusion to penetratin. Taken together, these compounds now serve as bona fide starting points for the rational design of peptide- or peptidomimetic-based antivirals for COVID-19 that target SARS-CoV-2 M^{pro}. Moreover, the crosslinked protease selection strategy developed as part of this work is expected to inform medicinal chemistry campaigns targeting other dimeric coronaviral main proteases. Future work in our laboratories will involve lead optimization of the discovered peptides for improved activity and cell permeability both by rational design and high-throughput affinity maturation [53,54] with a view to developing molecules with potent antiviral activity *in vivo*.

Otting, Gottfried; Turville, Stuart; Jackson, Colin; Nitsche, Christoph; Payne, Richard. The thesis author was an author of this paper

REFERENCES

[1] Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., … Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

[2] Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious diseases*, *20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

[3] Wall, E. C., Wu, M., Harvey, R., Kelly, G., Warchal, S., Sawyer, C., Daniels, R., Hobson, P., Hatipoglu, E., Ngai, Y., Hussain, S., Nicod, J., Goldstone, R., Ambrose, K., Hindmarsh, S., Beale, R., Riddell, A., Gamblin, S., Howell, M., … Bauer, D. L. V. (2021). Neutralising antibody activity against SARS-COV-2 vocs B.1.617.2 and B.1.351 by BNT162B2 vaccination. *The Lancet*, *397*(10292), 2331–2333. https://doi.org/10.1016/s0140-6736(21)01290-3

[4] Wang, Y., Zhang, D., Du, G., Du, R., Zhao, J., Jin, Y., Fu, S., Gao, L., Cheng, Z., Lu, Q., Hu, Y., Luo, G., Wang, K., Lu, Y., Li, H., Wang, S., Ruan, S., Yang, C., Mei, C., … Wang, C. (2020). Remdesivir in adults with severe COVID-19: A randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*, *395*(10236), 1569–1578. https://doi.org/10.1016/s0140-6736(20)31022-9

[5] Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., Hohmann, E., Chu, H. Y., Luetkemeyer, A., Kline, S., Lopez de Castilla, D., Finberg, R. W., Dierberg, K., Tapson, V., Hsieh, L., Patterson, T. F., Paredes, R., Sweeney, D. A., Short, W. R., Touloumi, G., … ACTT-1 Study Group Members (2020). Remdesivir for the Treatment of Covid-19 - Final Report. *The New England journal of medicine*, *383*(19), 1813–1826. https://doi.org/10.1056/NEJMoa2007764

[6] Dyer O. (2020). Covid-19: Remdesivir has little or no impact on survival, WHO trial shows. *BMJ (Clinical research ed.)*, *371*, m4057. https://doi.org/10.1136/bmj.m4057

[7]Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M., Li, X., Xia, J., Chen, N., Xiang, J., Yu, T., Bai, T., Xie, X., Zhang, L., Li, C., Yuan, Y., … Wang, C. (2020). A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *The New England journal of medicine*, *382*(19), 1787–1799. https://doi.org/10.1056/NEJMoa2001282

[8] Horby, P. W., Mafham, M., Bell, J. L., Linsell, L., Staplin, N., Emberson, J., Palfreeman, A., Raw, J., Elmahi, E., Prudon, B., Green, C., Carley, S., Chadwick, D., Davies, M., Wise, M. P., Baillie, J. K., Chappell, L. C., Faust, S. N., Jaki, T., … Landray, M. J. (2020). Lopinavir–ritonavir in patients admitted to hospital with covid-19 (recovery): A randomised, controlled, open-label, Platform Trial. The Lancet, 396(10259), 1345–1352. https://doi.org/10.1016/s0140-6736(20)32013-4

[9]Schreiber G. (2020). The Role of Type I Interferons in the Pathogenesis and Treatment of COVID-19. *Frontiers in immunology*, *11*, 595739. https://doi.org/10.3389/fimmu.2020.595739

[10]Monk, P. D., Marsden, R. J., Tear, V. J., Brookes, J., Batten, T. N., Mankowski, M., Gabbay, F. J., Davies, D. E., Holgate, S. T., Ho, L. P., Clark, T., Djukanovic, R., Wilkinson, T., & Inhaled Interferon Beta COVID-19 Study Group (2021). Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet. Respiratory medicine*, *9*(2), 196–206. https://doi.org/10.1016/S2213-2600(20)30511-7

[11]Boulware, D. R., Pullen, M. F., Bangdiwala, A. S., Pastick, K. A., Lofgren, S. M., Okafor, E. C., Skipper, C. P., Nascene, A. A., Nicol, M. R., Abassi, M., Engen, N. W., Cheng, M. P., LaBar, D., Lother, S. A., MacKenzie, L. J., Drobot, G., Marten, N., Zarychanski, R., Kelly, L. E., Schwartz, I. S., … Hullsiek, K. H. (2020). A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19. *The New England journal of medicine*, *383*(6), 517–525. https://doi.org/10.1056/NEJMoa2016638

[12]Skipper, C. P., Pastick, K. A., Engen, N. W., Bangdiwala, A. S., Abassi, M., Lofgren, S. M., Williams, D. A., Okafor, E. C., Pullen, M. F., Nicol, M. R., Nascene, A. A., Hullsiek, K. H., Cheng, M. P., Luke, D., Lother, S. A., MacKenzie, L. J., Drobot, G., Kelly, L. E., Schwartz, I. S., Zarychanski, R., … Boulware, D. R. (2020). Hydroxychloroquine in Nonhospitalized Adults

With Early COVID-19 : A Randomized Trial. *Annals of internal medicine*, *173*(8), 623–631. https://doi.org/10.7326/M20-4207

[13]Cavalcanti, A. B., Zampieri, F. G., Rosa, R. G., Azevedo, L., Veiga, V. C., Avezum, A., Damiani, L. P., Marcadenti, A., Kawano-Dourado, L., Lisboa, T., Junqueira, D., de Barros E Silva, P., Tramujas, L., Abreu-Silva, E. O., Laranjeira, L. N., Soares, A. T., Echenique, L. S., Pereira, A. J., Freitas, F., Gebara, O., … Coalition Covid-19 Brazil I Investigators (2020). Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *The New England journal of medicine*, *383*(21), 2041–2052. https://doi.org/10.1056/NEJMoa2019014

[14]Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O'Meara, M. J., Monel, B., Vallet, T., White, K. M., Zhang, Z., Alon, A., Schadt, H., O'Donnell, H. R., Lyu, J., Rosales, R., McGovern, B. L., Rathnasinghe, R., Jangra, S., Schotsaert, M., Galarneau, J. R., Krogan, N. J., … Shoichet, B. K. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science (New York, N.Y.)*, *373*(6554), 541–547. https://doi.org/10.1126/science.abi4708

[15]RECOVERY Collaborative Group, Horby, P., Lim, W. S., Emberson, J. R., Mafham, M., Bell, J. L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L. C., Faust, S. N., Jaki, T., … Landray, M. J. (2021). Dexamethasone in Hospitalized Patients with Covid-19. *The New England journal of medicine*, *384*(8), 693–704. https://doi.org/10.1056/NEJMoa2021436

[16]Keller, M. J., Kitsis, E. A., Arora, S., Chen, J. T., Agarwal, S., Ross, M. J., Tomer, Y., & Southern, W. (2020). Effect of Systemic Glucocorticoids on Mortality or Mechanical Ventilation in Patients With COVID-19. *Journal of hospital medicine*, *15*(8), 489–493. https://doi.org/10.12788/jhm.3497

[17]Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

[18]Bayati, A., Kumar, R., Francis, V., & McPherson, P. S. (2021). SARS-CoV-2 infects cells after viral entry via clathrin-mediated endocytosis. *The Journal of biological chemistry*, *296*, 100306. https://doi.org/10.1016/j.jbc.2021.100306

[19]Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(21), 11727–11734. https://doi.org/10.1073/pnas.2003138117

[20]V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature reviews. Microbiology*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

[21]Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A. R., Tascher, G., Geurink, P. P., Wilhelm, A., van der Heden van Noort, G. J., Ovaa, H., Müller, S., Knobeloch, K. P., Rajalingam, K., Schulman, B. A., Cinatl, J., Hummer, G., … Dikic, I. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, *587*(7835), 657–662. https://doi.org/10.1038/s41586-020-2601-5

[22]Wu, Y., Ma, L., Zhuang, Z., Cai, S., Zhao, Z., Zhou, L., Zhang, J., Wang, P. H., Zhao, J., & Cui, J. (2020). Main protease of SARS-CoV-2 serves as a bifunctional molecule in restricting type I interferon antiviral signaling. *Signal transduction and targeted therapy*, *5*(1), 221. https://doi.org/10.1038/s41392-020-00332-2

[23]Fung, S. Y., Siu, K. L., Lin, H., Yeung, M. L., & Jin, D. Y. (2021). SARS-CoV-2 main protease suppresses type I interferon production by preventing nuclear translocation of phosphorylated IRF3. *International journal of biological sciences*, *17*(6), 1547–1554. https://doi.org/10.7150/ijbs.59943

[24]Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., Jordan, T. X., Oishi, K., Panis, M., Sachs, D., Wang, T. T., Schwartz, R. E., Lim, J. K., Albrecht, R. A., & tenOever, B. R. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, *181*(5), 1036–1045.e9. https://doi.org/10.1016/j.cell.2020.04.026

[25]Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., Breillat, P., Carlier, N., Gauzit, R., Morbieu, C., Pène, F., Marin, N., Roche, N., Szwebel, T. A., Merkling, S. H., Treluyer, J. M., … Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science (New York, N.Y.)*, *369*(6504), 718–724. https://doi.org/10.1126/science.abc6027

[26]Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., & Hilgenfeld, R. (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO journal*, *21*(13), 3213–3224. https://doi.org/10.1093/emboj/cdf327

[27]Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science (New York, N.Y.)*, *368*(6489), 409–412. https://doi.org/10.1126/science.abb3405

[28]Ullrich, S., & Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorganic & medicinal chemistry letters*, *30*(17), 127377. https://doi.org/10.1016/j.bmcl.2020.127377

[29]Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science (New York, N.Y.)*, *300*(5626), 1763–1767. https://doi.org/10.1126/science.1085658

[30]Boras, B., Jones, R. M., Anson, B. J., Arenson, D., Aschenbrenner, L., Bakowski, M. A., Beutler, N., Binder, J., Chen, E., Eng, H., Hammond, H., Hammond, J., Haupt, R. E., Hoffman, R., Kadar, E. P., Kania, R., Kimoto, E., Kirkpatrick, M. G., Lanyon, L., Lendy, E. K., … Allerton, C. (2021). Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID19. *Nature communications*, *12*(1), 6055. https://doi.org/10.1038/s41467-021-26239-2

[31]Hoffman, R. L., Kania, R. S., Brothers, M. A., Davies, J. F., Ferre, R. A., Gajiwala, K. S., He, M., Hogan, R. J., Kozminski, K., Li, L. Y., Lockner, J. W., Lou, J., Marra, M. T., Mitchell, L. J., Jr, Murray, B. W., Nieman, J. A., Noell, S., Planken, S. P., Rowe, T., Ryan, K., … Taggart, B. (2020). Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the

Potential Therapeutic Treatment of COVID-19. *Journal of medicinal chemistry*, *63*(21), 12725–12747. https://doi.org/10.1021/acs.jmedchem.0c01063

[32]Vankadara, S., Wong, Y. X., Liu, B., See, Y. Y., Tan, L. H., Tan, Q. W., Wang, G., Karuna, R., Guo, X., Tan, S. T., Fong, J. Y., Joy, J., & Chia, C. (2021). A head-to-head comparison of the inhibitory activities of 15 peptidomimetic SARS-CoV-2 3CLpro inhibitors. *Bioorganic & medicinal chemistry letters*, *48*, 128263. https://doi.org/10.1016/j.bmcl.2021.128263

[33]Owen, D. R., Allerton, C., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., Boras, B., Cardin, R. D., Carlo, A., Coffman, K. J., Dantonio, A., Di, L., Eng, H., Ferre, R., Gajiwala, K. S., Gibson, S. A., Greasley, S. E., Hurst, B. L., Kadar, E. P., Kalgutkar, A. S., … Zhu, Y. (2021). An oral SARS-CoV-2 M[pro] inhibitor clinical candidate for the treatment of COVID-19. *Science (New York, N.Y.)*, *374*(6575), 1586–1593. https://doi.org/10.1126/science.abl4784

[34]Vandyck, K., & Deval, J. (2021). Considerations for the discovery and development of 3-chymotrypsin-like cysteine protease inhibitors targeting SARS-CoV-2 infection. *Current opinion in virology*, *49*, 36–40. https://doi.org/10.1016/j.coviro.2021.04.006

[35]Dai, W., Zhang, B., Jiang, X. M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., … Liu, H. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science (New York, N.Y.)*, *368*(6497), 1331–1335. https://doi.org/10.1126/science.abb4489

[36]Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y., & Wang, J. (2020). Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell research*, *30*(8), 678–692. https://doi.org/10.1038/s41422-020-0356-z

[37]Yang, K. S., Ma, X. R., Ma, Y., Alugubelli, Y. R., Scott, D. A., Vatansever, E. C., Drelich, A. K., Sankaran, B., Geng, Z. Z., Blankenship, L. R., Ward, H. E., Sheng, Y. J., Hsu, J. C., Kratch, K. C., Zhao, B., Hayatshahi, H. S., Liu, J., Li, P., Fierke, C. A., Tseng, C. K., … Liu, W. R. (2021). A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors*. *ChemMedChem*, *16*(6), 942–948. https://doi.org/10.1002/cmdc.202000924

[38]Kitamura, N., Sacco, M. D., Ma, C., Hu, Y., Townsend, J. A., Meng, X., Zhang, F., Zhang, X., Ba, M., Szeto, T., Kukuljac, A., Marty, M. T., Schultz, D., Cherry, S., Xiang, Y., Chen, Y., & Wang, J. (2022). Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *Journal of medicinal chemistry*, *65*(4), 2848–2865. https://doi.org/10.1021/acs.jmedchem.1c00509

[39]Fu, L., Ye, F., Feng, Y., Yu, F., Wang, Q., Wu, Y., Zhao, C., Sun, H., Huang, B., Niu, P., Song, H., Shi, Y., Li, X., Tan, W., Qi, J., & Gao, G. F. (2020). Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nature communications*, *11*(1), 4417. https://doi.org/10.1038/s41467-020-18233-x

[40]Ghahremanpour, M. M., Tirado-Rives, J., Deshmukh, M., Ippolito, J. A., Zhang, C. H., Cabeza de Vaca, I., Liosi, M. E., Anderson, K. S., & Jorgensen, W. L. (2020). Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS medicinal chemistry letters*, *11*(12), 2526–2533. https://doi.org/10.1021/acsmedchemlett.0c00521

[41]Vinogradov, A. A., Yin, Y., & Suga, H. (2019). Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *Journal of the American Chemical Society*, *141*(10), 4167–4181. https://doi.org/10.1021/jacs.8b13178

[42]Zorzi, A., Deyle, K., & Heinis, C. (2017). Cyclic peptide therapeutics: past, present and future. *Current opinion in chemical biology*, *38*, 24–29. https://doi.org/10.1016/j.cbpa.2017.02.006

[43] Passioura, T., & Suga, H. (2017). A rapid way to discover nonstandard macrocyclic peptide modulators of drug targets. Chemical Communications, 53(12), 1931–1940. https://doi.org/10.1039/c6cc06951g

[44]Peacock, H., & Suga, H. (2021). Discovery of De Novo Macrocyclic Peptides by Messenger RNA Display. *Trends in pharmacological sciences*, *42*(5), 385–397. https://doi.org/10.1016/j.tips.2021.02.004

[38]Kitamura, N., Sacco, M. D., Ma, C., Hu, Y., Townsend, J. A., Meng, X., Zhang, F., Zhang, X., Ba, M., Szeto, T., Kukuljac, A., Marty, M. T., Schultz, D., Cherry, S., Xiang, Y., Chen, Y., & Wang, J. (2022). Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *Journal of medicinal chemistry*, *65*(4), 2848–2865. https://doi.org/10.1021/acs.jmedchem.1c00509

[39]Fu, L., Ye, F., Feng, Y., Yu, F., Wang, Q., Wu, Y., Zhao, C., Sun, H., Huang, B., Niu, P., Song, H., Shi, Y., Li, X., Tan, W., Qi, J., & Gao, G. F. (2020). Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nature communications*, *11*(1), 4417. https://doi.org/10.1038/s41467-020-18233-x

[40]Ghahremanpour, M. M., Tirado-Rives, J., Deshmukh, M., Ippolito, J. A., Zhang, C. H., Cabeza de Vaca, I., Liosi, M. E., Anderson, K. S., & Jorgensen, W. L. (2020). Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS medicinal chemistry letters*, *11*(12), 2526–2533. https://doi.org/10.1021/acsmedchemlett.0c00521

[41]Vinogradov, A. A., Yin, Y., & Suga, H. (2019). Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *Journal of the American Chemical Society*, *141*(10), 4167–4181. https://doi.org/10.1021/jacs.8b13178

[42]Zorzi, A., Deyle, K., & Heinis, C. (2017). Cyclic peptide therapeutics: past, present and future. *Current opinion in chemical biology*, *38*, 24–29. https://doi.org/10.1016/j.cbpa.2017.02.006

[43] Passioura, T., & Suga, H. (2017). A rapid way to discover nonstandard macrocyclic peptide modulators of drug targets. Chemical Communications, 53(12), 1931–1940. https://doi.org/10.1039/c6cc06951g

[44]Peacock, H., & Suga, H. (2021). Discovery of De Novo Macrocyclic Peptides by Messenger RNA Display. *Trends in pharmacological sciences*, *42*(5), 385–397. https://doi.org/10.1016/j.tips.2021.02.004

[45]Goto, Y., Katoh, T., & Suga, H. (2011). Flexizymes for genetic code reprogramming. *Nature protocols*, *6*(6), 779–790. https://doi.org/10.1038/nprot.2011.331

[46]Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., & Suga, H. (2008). Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS chemical biology*, *3*(2), 120–129. https://doi.org/10.1021/cb700233t

[47] Fujino, T., Goto, Y., Suga, H., & Murakami, H. (2013). Reevaluation of the D-amino acid compatibility with the elongation event in translation. *Journal of the American Chemical Society*, *135*(5), 1830–1837. https://doi.org/10.1021/ja309570x

[48]Zhu, L., George, S., Schmidt, M. F., Al-Gharabli, S. I., Rademann, J., & Hilgenfeld, R. (2011). Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral research*, *92*(2), 204–212. https://doi.org/10.1016/j.antiviral.2011.08.001

[49]Shi, J., & Song, J. (2006). The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *The FEBS journal*, *273*(5), 1035–1045. https://doi.org/10.1111/j.1742-4658.2006.05130.x

[50]Hendrickson W. A. (2000). Synchrotron crystallography. *Trends in biochemical sciences*, *25*(12), 637–643. https://doi.org/10.1016/s0968-0004(00)01721-7

[51]Lee, J., Worrall, L. J., Vuckovic, M., Rosell, F. I., Gentile, F., Ton, A. T., Caveney, N. A., Ban, F., Cherkasov, A., Paetzel, M., & Strynadka, N. (2020). Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nature communications*, *11*(1), 5877. https://doi.org/10.1038/s41467-020-19662-4

[52] Meng, B. (2019). A new application of schrödinger-type identity to singular boundary value problem for the schrödinger equation. Boundary Value Problems, 2019(1). https://doi.org/10.1186/s13661-019-1279-9

[53]Kauffman, W. B., Fuselier, T., He, J., & Wimley, W. C. (2015). Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends in biochemical sciences*, *40*(12), 749–764. https://doi.org/10.1016/j.tibs.2015.10.004

[54] Derossi, D., Joliot, A. H., Chassaing, G., & Prochiantz, A. (1994). The third helix of the Antennapedia homeodomain translocates through biological membranes. Journal of Biological Chemistry, 269(14), 10444–10450. https://doi.org/10.1016/s0021-9258(17)34080-2

[55]Nielsen, E. J., Yoshida, S., Kamei, N., Iwamae, R., Khafagy, e., Olsen, J., Rahbek, U. L., Pedersen, B. L., Takayama, K., & Takeda-Morishita, M. (2014). In vivo proof of concept of oral insulin delivery based on a co-administration strategy with the cell-penetrating peptide penetratin. *Journal of controlled release : official journal of the Controlled Release Society*, *189*, 19–24. https://doi.org/10.1016/j.jconrel.2014.06.022

[56]Muto, K., Kamei, N., Yoshida, M., Takayama, K., & Takeda-Morishita, M. (2016). Cell-Penetrating Peptide Penetratin as a Potential Tool for Developing Effective Nasal Vaccination Systems. *Journal of pharmaceutical sciences*, *105*(6), 2014–2017. https://doi.org/10.1016/j.xphs.2016.03.026

[57]Qian, Z., LaRochelle, J. R., Jiang, B., Lian, W., Hard, R. L., Selner, N. G., Luechapanichkul, R., Barrios, A. M., & Pei, D. (2014). Early endosomal escape of a cyclic cell-penetrating peptide allows effective cytosolic cargo delivery. *Biochemistry*, *53*(24), 4034–4046. https://doi.org/10.1021/bi5004102

[58]Bennink, L. L., Smith, D. J., Foss, C. A., Pomper, M. G., Li, Y., & Yu, S. M. (2017). High Serum Stability of Collagen Hybridizing Peptides and Their Fluorophore Conjugates. *Molecular pharmaceutics*, *14*(6), 1906–1915. https://doi.org/10.1021/acs.molpharmaceut.7b00009

[59]Rogers, J. M., Passioura, T., & Suga, H. (2018). Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 10959–10964. https://doi.org/10.1073/pnas.1809901115

[60] Bashiruddin, N. K., Hayashi, M., Nagano, M., Wu, Y., Matsunaga, Y., Takagi, J., Nakashima, T., & Suga, H. (2020). Development of cyclic peptides with potent in vivo osteogenic activity through RaPID-based affinity maturation. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(49), 31070–31077. https://doi.org/10.1073/pnas.2012266117.