Supporting Information for

Chemoproteomic profiling of gut microbiota-associated bile salt hydrolase activity

Bibudha Parasar[1], Hao Zhou[2], Xieyue Xiao[3], Qiaojuan Shi[3], Ilana L. Brito[3,4,5], and Pamela V. Chang[4,5,6]*

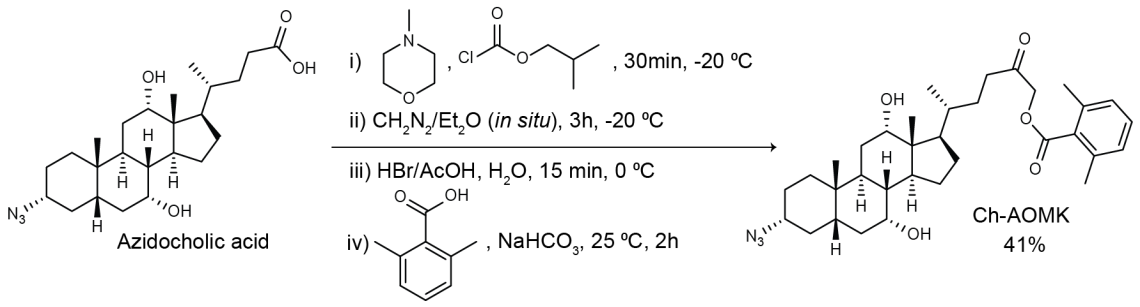
[1]Department of Chemistry and Chemical Biology, [2]Department of Microbiology, [3]Meinig School of Biomedical Engineering, [4]Center for Infection and Pathobiology, [5]Cornell Institute of Host-Microbe Interactions & Disease, [6]Department of Microbiology and Immunology, Cornell University, Ithaca, NY 14853

Contents

General materials and methods: All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. Organic extracts were dried over Na_2SO_4, and solvents were removed with a rotary evaporator at reduced pressure (20 torr), unless otherwise noted. Flash chromatography was performed using Silicycle Siliaflash P60 40-63Å 230-400 mesh silica gel. Analytical thin layer chromatography (TLC) was performed on glass-backed TLC 60 Å silica gel plates, and compounds were visualized by staining with ceric ammonium molybdate and the absorbance of UV light (λ = 254 nm or 365 nm). Reverse phase HPLC was performed using a Shimadzu system equipped with a CBM-20A controller, SPD20AV UV-Vis detector, LC-20AR liquid chromatograph unit, FRC-10A fraction collector, and an Epic Polar 5µm 120Å C18 analytical column (4.6 x 250 mm) at a flow rate of 1 mL/min or a semipreparative column (10 x 250 mm) at a flow rate of 4 mL/min. HPLC samples were filtered with a Millex-LH syringe filter equipped with a 0.45 µm PTFE membrane prior to injection. The water:acetonitrile gradient varied from 100:0 to 55:45 over 50 min with a curve value of -4, after which the gradient increased to 0:100 over 5 min at a linear rate and remained at 0:100 for an additional 10 min. Dichloromethane (DCM), methanol (MeOH), dimethylsulfoxide (DMSO), dimethylformamide (DMF), chloroform, and HPLC-grade water and acetonitrile were used from commercial sources without further purification. UV absorbance readings were measured on a Bio-Tek PowerWave XS microplate spectrophotometer. Anaerobic chamber (Coy Lab, model AC16-113) was maintained with the gas composition (3% hydrogen, 20% carbon dioxide, and 77% nitrogen). Coy Lab (model 2000) incubator was used inside the anaerobic chamber for the growth of anaerobic bacteria.

Anhydrous sodium sulfate, ethylenediaminetetraacetic acid disodium salt, sodium acetate, sodium carbonate, and Nycodenz were purchased from VWR. 2,6-dimethylbenzoic acid was purchased from Accela Chemicals. Sodium thiosulfate pentahydrate was purchased from Alfa Aesar. Silver nitrate was obtained from Macron Fine Chemicals. 4-methylmorpholine, hydrogen bromide solution (33 wt% in acetic acid), and iodoacetamide were purchased from Sigma Aldrich. Isobutylchloroformate was purchased from Lancaster Synthesis. Sodium bicarbonate, trichloroacetic acid (TCA), sodium dodecyl sulfate (SDS), copper(II) sulfate pentahydrate, and sodium ascorbate were purchased from Fisher Scientific. Choloylglycine hydrolase from *Clostridium perfringens*, isopropyl β-D-1-thiogalactopyranoside (IPTG), sodium glycocholate hydrate, dithiothreitol (DTT), phenylmethanesulfonyl fluoride (PMSF), 2-mercaptoethanol, sodium cholate hydrate, deoxycholic acid, sodium chenodeoxycholate, and lithocholic acid were purchased from Sigma Aldrich. Sodium taurocholate hydrate was purchased from Alfa Aesar. Sodium taurodeoxycholate was purchased from Santa Cruz Biotechnology. Sodium taurolithocholate was purchased from Cambridge Isotopes. Sodium taurochenodeoxycholate, tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), and tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were purchased from Cayman Chemical. Rhodamine110-based Fluor 488-alkyne was purchased from Sigma Aldrich. Biotin-PEG4-alkyne (biotin-alkyne) was purchased from Click Chemistry Tools. Zirconia/silica beads (0.1 mm diameter) were obtained from BioSpec Products. High capacity streptavidin agarose beads were purchased from Thermo Fisher Scientific. Precision Plus Protein™ All Blue Standards ladder and Mark12™ Unstained Protein Standard was purchased from Bio-Rad and Thermo Fisher Scientific, respectively. Horse radish peroxidase-conjugated streptavidin (streptavidin-HRP) was obtained from Genetex (GTX85912). Sequencing Grade Modified Trypsin was purchased from Promega. Bovine serum albumin (BSA) was obtained from Amresco. Rabbit anti-FLAG polyclonal antibody was obtained from Sigma Aldrich (F7425), and goat anti-rabbit conjugated HRP antibody was purchased from Jackson

Immunoresearch (111-035-144). Bifidobacterium broth was purchased from HIMedia (M1395-500G). Dextran sodium sulfate was obtained from Chem-Impex International.



Scheme S1: Synthesis of Ch-AOMK.

Synthesis of the Ch-AOMK: Azidocholic acid[1] (300 mg, 0.7 mmol) was dissolved in anhydrous THF (3.64 mL) and stirred in a dry ice/acetone bath at -15 °C for 5 min. N-methylmorpholine (98.1 µL, 0.866 mmol, 1.25 eq.) and isobutyl chloroformate (104 µL, 0.8 mmol, 1.15 eq.) were sequentially added to this solution, and the mixture was stirred at -15 °C for an additional 30 min, during which a white precipitate formed. The reaction was brought to 0 °C. Ethereal diazomethane was generated in situ according to the procedure reported in the Sigma Aldrich technical bulletin (AL-180). A flame polished glass pipette was used to add diazomethane (3 mmol, 3.75 eq.) dropwise to the reaction mixture at 0 °C, and the reaction was slowly warmed to room temperature over 4 h. To generate the corresponding bromomethyl ketone, the reaction mixture was cooled to 0 °C. Hydrogen bromide (33 w% in acetic acid, 5 mL, 75 mmol, 107 eq.) was mixed with 10 mL of water and added to the reaction mixture, dropwise until the evolution of nitrogen gas stopped. The mixture was diluted with ethyl acetate and transferred to a separatory funnel. The organic layer was washed sequentially with water, brine, and $NaHCO_3$, then dried over anhydrous Na_2SO_4. The organic layers were combined and rotovapped to yield a sticky yellow solid.

The crude was dissolved in dry DMF (0.7 mL) and was stirred at room temperature under nitrogen, and 2,6-dimethylbenzoic acid (30 mg, 0.2 mmol, 0.28 eq.) and sodium bicarbonate (17 mg, 0.2 mmol, 0.28 eq.) were added. After 2 h, the reaction mixture was diluted with ethyl acetate and transferred to a separatory funnel. The organic layer was washed with water 3 times and dried over anhydrous Na_2SO_4. The crude was purified by flash column chromatography (2% MeOH in DCM) to yield a sticky solid. The solid was dissolved in DMSO and further purified by HPLC to generate the purified compound as a white powder (41% yield). TLC (CH_2Cl_2:MeOH, 99:1 v/v): R_f = 0.6; ^{1}H NMR (500 MHz, DMSO-d6): δ 7.27 (t, *J* = 7.5 Hz, 1H), 7.11 (d, *J* = 3.75 Hz, 2H), 5.05 (s, 2H), 4.17 (m, 2H), 3.80 (s, 1H), 3.63 (s, 1H), 3.23 (m, 1H), 2.41 (m, 2H), 2.32 (s, 6H), 2.10 (m, 2H), 1.99 (s, 1H), 1.83 (m, 2H), 1.75 (m, 2H), 1.66 (m, 2H), 1.58 (m, 2H), 1.39 (m, 7H), 1.23 (m, 2H), 1.00 (m, 1H), 0.95 (d, *J* = 2.5 Hz, 4H), 0.85 (s, 3H), 0.6 (s, 3H); ^{13}C NMR (500 MHz, DMSO-d6): δ 204.58, 177.34, 168.75, 135.19, 133.49, 130.03, 128.01, 71.45, 68.79, 66.57, 61.25, 46.51, 46.24, 41.97, 41.86, 35.51, 35.46, 35.37, 34.97, 34.79, 29.38, 28.92, 27.68, 26.79, 26.60, 23.21, 22.95, 19.77, 17.60, 12.80; ESI-MS (m/z): $[M-H]^-$ calcd. For $C_{34}H_{48}N_3O_5^-$, 578.3599; found, 578.3570.

Safety statement: Significant hazards were mitigated in the generation of diazomethane by using ground glass joints, a blast shield, and loosely sealing the reaction vessel. Afterwards, the syringes, needles, and glassware were quenched with acetic acid.

Mice: C57Bl/6 mice were purchased from Jackson Laboratories and bred in house at the animal facility of Cornell University. Mice were used at 10-15 weeks of age in accordance with the guidelines of the Institutional Animal Care and Use Committee and the Cornell Center for Animal Resources and Education (Protocol Number 2015-0069). Mice were co-housed for 7 d prior to use.

Dextran sodium sulfate (DSS) colitis: Mice were treated with DSS (0 or 3% w/v in drinking water, *ad libitum*) for 8 d. The mice were weighed daily, and fecal samples were collected throughout the treatment. Mouse feces pellets were collected on dry ice, flash frozen, and stored in the -80 °C freezer until further use.

Bacteria cultures: *Bifidobacterium longum* subsp. *infantis* ATCC15697 and *Bifidobacterium bifidum* DSM20456 strains were purchased from American Type Culture Collection (ATCC). Bifidobacterium species were grown in Bifidobacterium broth supplied by ATCC, anaerobically at 37 °C until the stationary phase, after which they were aliquoted into tubes, pelleted down (4,500xg, 30 min), flash frozen, and stored at -80 °C until further use.

Cloning of *C. perfringens* choloylglycine hydrolase (CGH): CGH from *C. perfringens* strain 13A was amplified from a synthetic gene fragment (gBlocks, Integrated DNA Technologies) and cloned into the pET-21b vector with an ampicillin resistance gene. Briefly, two forward primers containing a NdeI restriction site for constructing wildtype CGH (WT, CTAGCATATGTGTACCGGATTAGCGCTGGAG) or a point mutant (cysteine 2 to serine, C2S, CTAGCATATGTCTACCGGATTAGCGCTGGAGAC) and a shared reverse primer (CTAGCTCGAGCTTATCGTCGTCATCCTTGTAATCGTTCACGTGGTTGATGCTCAG), which was designed containing an XhoI restriction site and a C-terminal FLAG peptide (DYKDDDDK) epitope tag, were used to amplify WT or C2S CGH.

***C. perfringens* CGH protein expression:** The pET-21b expression plasmid (WT or C2S CGH) was introduced into *E. coli* Rosetta 2 (DE3)pLysS competent cells. The resulting colonies were cultured in Terrific Broth liquid medium (12 g/L BactoTryptone, 23.9 g/L yeast extract, 8 ml/L glycerol, 0.22 g/L KH_2PO_4, 0.94 g/L K_2HPO_4) in the presence of ampicillin (50 µg/ml) at 37 °C to an OD600 of 0.5-0.6. The expression of CGH was induced by the addition of IPTG (0.1 mM), and the cells were incubated at 18 °C for 18 h with shaking (220 rpm). The cells were harvested by centrifuging at 12,000×g for 1 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 20 mM sodium phosphate buffer (pH 7.0) containing 1 mM EDTA, 10 mM DTT and 1 mM PMSF, and lysed by sonication. The tube was centrifuged at 16,000×g for 30 min at 4 °C, and the supernatant was collected and quantified by the DC assay (Bio-Rad).

General method for methanol-chloroform protein precipitation: To 20 µL of reaction mixture was added 200 µL of methanol, 75 µL of chloroform, 150 µL of water, sequentially. The sample was then vortexed and centrifuged at 13,000×g for 15 min at 4 °C. The aqueous layer was removed without disturbing the protein layer. Cold methanol (1 mL) was added, and the tube was inverted a few times gently without disturbing the precipitated protein. The tube was centrifuged at 13,000×g for 10 min at 4 °C and decanted, leaving the pellet intact. The pellet was again washed with 1 mL of cold methanol, centrifuged at 13,000×g for 10 min at 4 °C, and the supernatant was decanted. The pellets were left to dry for 3-5 min at 37°C, after which they were resolubilized in the appropriate buffer for each assay.

Methanol-chloroform-acetone protein precipitation: To 20 µL of reaction mixture was added 200 µL of methanol, 75 µL of chloroform, 150 µL of water, sequentially. The sample was vortexed and centrifuged at 13,000×g for 15 min at 4 °C. The aqueous layer was removed without disturbing the protein layer, after which the organic layer was carefully removed using a gel loading tip. Cold methanol (1 mL) was added, and the tube was inverted a few times gently, leaving the protein layer intact. The tube was centrifuged at 13,000×g for 10 min at 4 °C and decanted leaving the pellet intact. The pellet was again washed with 1 mL of cold methanol, centrifuged at 13,000×g for 10 min at 4 °C, and the supernatant was decanted. The pellets were left to dry for 3-5 min at 37 °C, after which they were resolubilized in the appropriate buffer for each assay.

General method for TCA-acetone protein precipitation: To the reaction mixture (100 µL – 2 mL, see Table 1, *vide infra*) was added TCA at a final concentration of 1.5% (w/v). The samples were chilled on ice for 30 min. The precipitate was pelleted down for 30 min at 4 °C and 20,000xg, and the supernatant was decanted carefully. Cold acetone (-20 °C, 1 mL) was added to the pellet, and the sample was sonicated briefly until the pellet was broken up, after which the protein was pelleted down (20,000xg, 4 °C, 5 min). The pellet was washed 2 more times with cold acetone (1 mL), and the supernatant was decanted. The pellets were left to dry for 3-5 min at 37 °C, after which they were resolubilized in the appropriate buffer for each assay.

General method for copper-catalyzed azide-alkyne cycloaddition (CuAAC) or click chemistry: The reaction was carried out in click buffer (0.1 M sodium phosphate buffer, 4% SDS, pH 7.4) for a final reaction volume of 20 µL – 2 mL, see Table 1. Copper-THPTA complex was formed by mixing copper(II) sulfate pentahydrate with THPTA to yield a concentration of 10 and 20 mM, respectively. Then, either Fluor 488-alkyne (10 µM) or biotin-alkyne (50 µM), pre-formed copper-THPTA complex (final concentration of 1 mM and 2 mM, respectively), and sodium ascorbate (100 µM) were added sequentially. The tubes were gently mixed between the addition of each reagent. The tubes were incubated at 37 °C for 2 h (in the dark for fluorescence-based in-gel activity assay).

General method for fluorescence-based in-gel activity assay: Ch-AOMK (500 µM) was incubated with 100 µg (unless otherwise noted) bacterial lysate at 37 °C for 24 h. For the inactivation of global cysteine residues, including BSHs, the lysate was incubated with 20 mM of iodoacetamide at 37 °C for 2 h prior to Ch-AOMK addition. Excess Ch-AOMK was removed by methanol-chloroform precipitation as above. The pellet was resuspended in click buffer, and click chemistry was carried out in the presence of Fluor 488-alkyne as above. After CuAAC, excess fluorophore was removed by methanol-chloroform precipitation. The pellet was boiled in 20 µL of 1X SDS loading buffer at 95 °C and mixed periodically for 20 min. The samples were purified using a 12% SDS-PAGE gel. The gel was then incubated in fluorescence destaining solution (40% methanol, 50% acetic acid, 10% water) on a rocker in the dark for 1 h. The gel was rinsed with water and imaged at an excitation wavelength of 488 nm using a Bio-Rad ChemiDoc MP Imaging System (fluorescence mode). The gel was subsequently stained with Coomassie brilliant blue for 15 min. For destaining, the gel was incubated in Coomassie destaining solution (30% methanol, 5% acetic acid, 65% water) on a rocker overnight and imaged the following day using a Bio-Rad ChemiDoc MP Imaging System (colorimetric mode).

In vitro labeling of purified CGH from *C. perfringens* with Ch-AOMK: CGH from *C. perfringens* (Sigma Aldrich) was resuspended in sodium phosphate buffer (10 mM, pH 7.0) to a final concentration of 10 mg/mL. Unless otherwise mentioned, 10 µg of CGH was incubated in

reaction buffer (50 mM sodium acetate, pH 5.6) in the presence of various concentrations of Ch-AOMK (0 – 500 μM) for various amounts of time (0- 24 h) at 37 °C. For the heat-killed samples, 10 μg of CGH was incubated in reaction buffer (20 μL) and incubated at 80 °C for 20 min to deactivate the enzyme before Ch-AOMK addition. Prior to the fluorescence-based in-gel activity assay, the reaction mixture was doped with 200 μg of BSA for efficient precipitation of proteins using methanol-chloroform-acetone precipitation as above to remove excess Ch-AOMK or click chemistry reagents. BSA was visualized with Coomassie staining, which served as a loading control for the samples.

General method for lysing *E. coli*, *B. bifidum*, and *B. longum*: Bacterial pellets were resuspended in 20 mM sodium phosphate buffer (pH 7.0) containing 1 mM EDTA, 10 mM DTT, and 1 mM PMSF (250 μL of buffer for each 50 mL culture). The samples were then transferred to a 2 mL screwcap tube (for each 50 mL culture). Zirconia/silica beads (330 mg) were added to the tube, which was then subjected to agitation on a bead beater (8 cycles of 1 min beating, followed by 3 min on ice). The samples were centrifuged at 16,000×g to remove the beads and cell debris. The supernatant was transferred to a 1.5 mL microfuge tube, and the concentration was measured by the DC assay.

Isolation of bacteria from mouse gut microbiome and lysis: The following bacterial isolation procedure was carried out on ice where possible. Pellets were thawed and crushed in 1X PBS in an autoclaved mortar and pestle (70 mL of 1X PBS per 15 g wet weight of feces), and the mixture was centrifuged at 300xg for 5 min at 4 °C. The supernatant was transferred to another 50 mL conical tube, and the centrifugation was repeated 2 more times to remove fecal matter. Then, the clarified supernatant was transferred to a clean vessel. Meanwhile, Nycodenz (500 μL, 50% w/v, autoclaved) was added to 1.5 mL microfuge tubes, and 1 mL of the supernatant was carefully layered on top of the gradient. The biphasic solution was then centrifuged at 10,000xg for 40 min at 4 °C. After centrifugation, the top layer was carefully removed, leaving the middle layer of bacteria undisturbed. Then, the bacterial layer was carefully pipetted out and combined in a 50 mL conical tube. The collected bacteria were mixed thoroughly and aliquoted into multiple 1.5 mL microfuge tubes, each with 500 μL of mixture. The bacteria were washed with 1 mL of 1X PBS 3 times by thorough resuspension, followed by centrifugation at 16,000xg for 5 min at 4 °C. The final pellet was resuspended in 200 μL of buffer (20 mM sodium phosphate buffer (pH 7.0) containing 1 mM EDTA, 10 mM DTT and 1 mM PMSF), and the samples were transferred to a 2 mL screw cap tube containing 330 mg of 0.1 mm zirconia/silica beads. The samples were beated 5 times (each time with 1 min beating, followed by 4 min on ice). The samples were then centrifuged at 16,000xg for 30 min at 4 °C, and the supernatant was transferred to a clean tube. The protein concentration was measured using the DC assay.

General method for labeling of bacterial lysates with Ch-AOMK: Unless otherwise mentioned, lysates were incubated with 10X reaction buffer (500 mM sodium acetate buffer at the optimized pH, see Table 1) and a final Ch-AOMK concentration of 0 – 500 μM (25 mM stock in DMSO) for 12-24 h at 37 °C.

General method for enrichment of biotin-labeled proteins following click chemistry: Ch-AOMK (100 – 500 μM) was incubated with bacterial lysate (2.5 – 50 mg) at 37 °C for 12-24 h, see Table 1. Excess Ch-AOMK was removed by precipitation method indicated in Table 1. The pellet was resuspended in click buffer by sonication, and the click reaction was carried out in the presence of biotin-alkyne as above. After the click reaction, excess biotin-reagent was removed

by precipitation as above (Table 1). The dried protein pellet was resuspended in 166 µL of 1.2% SDS (per 2.5 mg of protein lysate) in 1X PBS by sonication, and 834 µL of 1X PBS was added to dilute the final concentration to 0.2% SDS. Lysates (50 µg, 2% of elution) were collected as the input control. In the meantime, 10 µL of streptavidin agarose beads (per 2.5 mg of lysate) were washed three times with 1 mL of 0.2% SDS in PBS. The sample was added to the beads and incubated at room temperature for 1 h on a rotator. The tubes were centrifuged at 13,000×g for 1 min at 4 °C, and the supernatant was discarded. The beads were sequentially washed with 1 mL of 1%, 0.5%, 0.2% and 0% SDS in PBS. The beads were then subjected to elution as described below.

General protocol for elution of proteins from streptavidin agarose beads for silver staining and Western blotting: After washing the beads with decreasing concentrations of SDS as above, 30 µL of 2X SDS was added to the beads. The samples were then incubated at 25 °C for 5 min, followed by 95 °C for 30 min. The tubes were centrifuged at 13,000×g for 1 min at room temperature, and the supernatant was transferred to another tube for visualization by silver staining and Western blot analysis.

Silver staining: The eluted protein lysates (25 µL) were purified using a 12% SDS-PAGE gel. The gel was fixed in fixation buffer (50% methanol, 5% acetic acid, 45% water) for 20 min. The gel was then washed with 50% methanol, followed by water, for 10 min each. The gel was sensitized by incubating it in 0.02% (w/v) sodium thiosulfate for 1 min, after which it was rinsed with water 5 times. The gel was incubated with 0.1% (w/v) silver nitrate with 0.08% paraformaldehyde for 20 min and rinsed with water 6-7 times to remove excess silver. The gel was developed by incubating it in 50 mL of 2% (w/v) sodium carbonate with 0.04% paraformaldehyde with mild agitation until bands began to appear. After the desired level of signal to noise was obtained, 1 mL of acetic acid was added to quench the development. The gel was rinsed with water 6-7 times and imaged using a Bio-Rad ChemiDoc MP Imaging System (colorimetric mode).

Western blot: Briefly, 5 µL of the eluted protein lysates were purified using a 12% SDS-PAGE gel. After transfer of proteins to a nitrocellulose membrane, the blot was incubated in blocking solution (5% BSA in Tris buffer (500 mM, pH 7.6) with 0.5% Tween-20 (TBST)) for 1 h, after which the membrane was incubated with streptavidin-HRP (in 5% BSA in TBST) for 1 h. The blot was rinsed 3 times with TBST for 10 min, developed with Clarity Western ECL Blotting Substrates (Bio-Rad), and imaged using a Bio-Rad ChemiDoc MP Imaging System (chemiluminescence mode).

Bile salt hydrolase activity assay: BSH activity assay was performed as previously reported with the following modifications[2]. Briefly, 12.5 µL each of 20 mM sodium phosphate buffer (pH 5.6), 18 mM sodium glycocholate, 107 mM β-mercaptoethanol, and 55 mM EDTA were mixed, and 45 µL of the reaction mixture was transferred to a separate 1.5 mL microfuge tube. Purified CGH was diluted with 20 mM sodium phosphate buffer (pH 5.6) to obtain a final concentration of 3 Units/µL, 5 µL of which was added to the reaction tube, and the samples were incubated at 37^0C. After 1 h, 50 µL of TCA was added to quench the reaction, and 20 µL of the final mixture was transferred to another tube. Equal volumes of 4% (w/v) ninhydrin in 2-methoxyethanol and 0.16% (w/v) stannous chloride in 200 mM sodium citrate buffer (pH 5.0) were mixed to generate the development solution. This solution (200 µL) was subsequently added to the tube containing 20 µL of the reaction mixture, and the samples were boiled at 95 °C for 20 min. The mixture was diluted 10-fold, and 200 µL was transferred to a 96-well plate for UV absorbance readings at 600

nm. A standard curve using glycine at various concentrations was simultaneously developed and analyzed.

Orbitrap Fusion by FT-Q-IT mode for on-bead or gel-based protein IDs and associated label-free quantitation (LFQ):

General protocol for in-gel digestion for identification of proteins: After washing the beads with decreasing concentrations of SDS in PBS as above, 120 µL of 2X SDS was added to the beads. The samples were incubated at 25 °C for 5 min, followed by 95 °C for 30 min. The tubes were centrifuged at 13,000×g for 1 min at room temperature, and the supernatant was transferred to another 1.5 mL microfuge tube. The eluted protein lysates (110 µL) were purified with a 12% SDS-PAGE gel using both a pre-stained and unstained protein ladder for gel excision. The gel was fixed in the fixation buffer (50% methanol, 10% acetic acid, 40% water) overnight and transferred to gel storage buffer (10% methanol, 7% acetic acid, 83% water) before excision.

The proteins in the gel were visualized with Sypro Ruby Protein Gel Stain (Thermo Fisher Scientific), and the desired bands were cut into ~1 mm cubes and subjected to in-gel digestion, followed by extraction of the tryptic peptide as reported previously[3]. The excised gel pieces were washed consecutively in 200 µL water, 100 mM ammonium bicarbonate/acetonitrile (1:1), and acetonitrile. The gel pieces were reduced with 250 µL of 10 mM DTT in 100 mM ammonium bicarbonate for 1 h at 56 °C and alkylated with 260 µL of 55 mM iodoacetamide in 100 mM ammonium bicarbonate at room temperature in the dark for 1 h. After wash steps as described above, the gel slices were dried and rehydrated with 10 µL trypsin in 50 mM ammonium bicarbonate, 10% acetonitrile (10 ng/µL) at 37 °C for 16 h. The digested peptides were extracted twice with 20 µL of 50% acetonitrile, 5% formic acid and once with 20 µL of 90% acetonitrile, 5% formic acid. Extracts from each sample were combined. Due to the large gel volume of the samples, the extracted peptide solution was filtered with a Costar 0.22 µm spin filter (Corning #8161) and then lyophilized.

On-bead digestion for identification of proteins: After washing the beads with decreasing concentrations of SDS in PBS as above, the beads were incubated with 500 µL of 6 M urea (in PBS) with 10 mM TCEP at 37 °C for 30 min, and then 25 µL of 400 mM iodoacetamide was added. The samples were then incubated at 37 °C for 30 min. In the meantime, trypsin was activated by incubating at 30 °C for 15 min. The beads were washed with 1 mL of 1M urea in PBS, after which the samples were incubated with 2 µg trypsin (in 200 µL of 1M urea with 1 mM $CaCl_2$ in PBS, pH 8.0) at 37 °C overnight with gentle rotating. The supernatant was removed, and 800 µL of water was added to bring the final volume to 1 mL. The samples were acidified with 1N HCl (6 µL) to reach a pH of ~2.6 to inactivate trypsin and then lyophilized.

Protein identification by nano LC/MS/MS analysis: The tryptic digests were reconstituted in 20 µL of 0.5% formic acid for nanoLC-ESI-MS/MS analysis, which was carried out using an Orbitrap Fusion™ Tribrid™ (Thermo Fisher Scientific, San Jose, CA) mass spectrometer equipped with a nanospray Flex Ion Source and coupled with a Dionex UltiMate3000RSLCnano system (Thermo, Sunnyvale, CA)[4,5]. The samples (5-15 µL) were injected onto a PepMap C-18 RP nano trapping column (5 µm, 100 µm i.d. x 20 mm) at a 20 µL/min flow rate for rapid sample loading and then separated on a PepMap C-18 RP nano column (2 µm, 75 µm x 25 cm) at 35 °C. The tryptic peptides were eluted in a 120 min gradient of 5% to 38% acetonitrile in 0.1% formic acid at 300 nL/min, followed by a 7 min ramping to 90% acetonitrile, 0.1% formic acid and an 8 min hold at 90% acetonitrile, 0.1% formic acid. The column was re-equilibrated for 25 min prior to the next

run. The Orbitrap Fusion was operated in positive ion mode with spray voltage set at 1.6 kV and source temperature at 275 °C. External calibration for FT, IT and quadrupole mass analyzers was performed. In data-dependent acquisition (DDA) analysis, the instrument was operated using FT mass analyzer in MS scan to select precursor ions followed by 3 s "Top Speed" data-dependent CID ion trap MS/MS scans at 1.6 m/z quadrupole isolation for precursor peptides with multiple charged ions above a threshold ion count of 10,000 and normalized collision energy of 30%. MS survey scans at a resolving power of 120,000 (fwhm at m/z 200), for the mass range of m/z 375-1575. Dynamic exclusion parameters were set at 40 s of exclusion duration with ±10 ppm exclusion mass width. All data were acquired under Xcalibur 3.0 operation software (Thermo Fisher Scientific).

Proteomics data analysis: The DDA raw files for CID MS/MS were subjected to database searches using Proteome Discoverer (PD) 2.2 software (Thermo Fisher Scientific, Bremen, Germany) with the Sequest HT algorithm. The PD 2.2 processing workflow containing an additional node of Minora Feature Detector for precursor ion-based quantification was used for protein identification and relative quantitation analysis between samples. The database search was conducted against the appropriate bacterial strain(s), available publicly on NCBI or UniProt or the metagenomic assemblies, plus a common contaminant (246 entries) database. Two-missed trypsin cleavage sites were allowed. The peptide precursor tolerance was set to 10 ppm, and fragment ion tolerance was set to 0.6 Da. Variable modification of methionine oxidation, deamidation of asparagine/glutamine and fixed modification of cysteine carbamidomethylation, were set for the database search. Only high confidence peptides defined by Sequest HT with a 1% FDR by Percolator were considered for the peptide identification. The final protein IDs contained protein groups that were filtered with at least 2 peptides per protein and 1 unique peptide.

Relative quantitation of identified proteins between the samples was determined by the Label Free Quantitation (LFQ) workflow in PD 2.2. The precursor abundance intensity for each peptide identified by MS/MS in each sample were automatically determined, and the unique peptides for each protein in each sample were summed and used for calculating the protein abundance by PD 2.2 software without normalization. Protein ratios were calculated based on pairwise ratios for each sample. The proteins that were considered enriched were cut-off using a $\geq$ 2-fold ratio of with probe compared to no probe in the healthy mouse gut microbiome samples. All proteomics data are displayed in Tables S2 and S5.

DNA isolation and shotgun metagenomic library preparation: Fecal DNA was prepared from 3-4 frozen pellets collected from mice treated with or without DSS using the *Quick*-DNA™ Fecal/Soil Microbe Miniprep Kit (Zymo Research, Catalog No. D6010) according to the manufacturer's instructions with minor modifications. Instead of vortexing, the samples were placed in a Retsch 96 Well Plate Shaker (MO BIO Catalog# 11996) for 10 min. The concentration of isolated DNA was assessed using Qubit™ dsDNA BR assay kit (Invitrogen) on a Qubit 3.0 Fluorometer (Invitrogen) and adjusted to 0.2 ng/µl for metagenomic library construction.

Nextera XT libraries were prepared following the manufacturer's protocol (15031942, Illumina). Briefly, samples starting with 1 ng of input DNA were fragmented and tagged via tagmentation, then amplified with a 12-cycle PCR followed by AMPure bead cleanup (A63881, Beckman Coulter). Two technical replicates were generated for each biological sample. Purified amplicons were quantified using Qubit™ dsDNA HS assay kit (Invitrogen). Samples were pooled at 2 nM

each and paired end sequenced (2x150bp) on an Illumina Nextseq instrument at the Cornell Biotechnology Resource Center Genomics Facility.

Bioinformatics analysis:

Data quality control: Metagenomic shotgun sequences were dereplicated using the prinseq-lite.pl v0.20.2[6], with following settings: -derep 12345 -no_qual_header. These settings were used to remove exact duplicates (1), 5' duplicates (2), 3' duplicates (3), reverse complement exact duplicates (4), and reverse complement 5'/3' duplicates (5). Dereplicated reads were then passed through the KneadData v0.3 quality control pipeline (http://huttenhower.sph.harvard.edu/kneaddata), which incorporates the Trimmomatic[7] and BMTagger (https://www.hmpdacc.org/hmp/doc/HumanSequenceRemoval_SOP.pdf) and decontamination algorithms to remove low-quality reads (thresholding Phred quality score at < 20; Minimum length < 150) and reads of C57BL/6NJ mouse origin (https://www.sanger.ac.uk/science/data/mouse-genomes-project). Taxonomic profiling was performed using MetaPhlAn2[8]. Reads were assembled with IDBA-UD v1.1.1[9] using the following default settings: --seed_kmer (seed kmer size for alignment) = 30, --min_contig (minimum size of contig) = 200, --similar (similarity for alignment) = 0.95, --min_pairs (minimum number of pairs) = 3. Open reading frames were predicted by Prodigal[10]. CD-hit v4.6.1 was used to generate a non-redundant protein database with 100% sequence identity threshold[11]. The protein database was then used to identify linear amide C-N hydrolases/choloylglycine hydrolase family domain proteins, based on Pfam model (PF02275)[12], using HMMER 3[13] hmmsearch with default settings. The abundance of putative *bsh* genes are calculated as RPKM (reads per kilobase of the gene per million mapped reads)[14] using outputs from BWA (mem) v0.7.17[15] and samtools (depth) v1.3.1[16]. Taxonomic assignments for *bsh* genes were performed using blastp, using DIAMOND v0.9.22.123, of the NCBI-NR database, using default settings and obtaining the unique best hit based on the lowest e-value[17].

To construct the phylogenetic tree, Clustal Omega v1.2.0[18] was used to align and generate similarity matrix for proteins identified as putative BSH proteins. A neighbor-joining tree with nearest-neighbor interchange was estimated by FastTree v2.1.7[19] using default settings. The phylogenic tree was plotted using the Interactive Tree of Life (https://itol.embl.de/)[20]. Bootstrap confidence levels reflect 100 phylogenetic tree reconstructions.

Code availability: Codes were used from commercial or public sources without modification as indicated above.

Data availability: Raw sequences will be submitted to the NCBI Short Read Archive upon manuscript acceptance.

Identification of bile acid metabolites by mass spectrometry-based metabolomics: Mouse fecal pellets were collected in pre-tared 1.5 mL microfuge tubes as described above. The pellets were lyophilized, weighed, and transferred to a 2 mL microfuge tube, after which the pellets were crushed with a glass rod. Cold methanol (500 µL) was added, and the samples were agitated in a bead beater for 45 s. The samples were chilled on ice for 5 min and centrifuged at 10,000xg for 20 min at 4 °C. The supernatant (250 µL) was transferred to a 1.5 mL microfuge tube and centrifuged under reduced pressure for 2 h. The dried samples were resuspended in 150 µL of LCMS grade MeOH and filtered with a Millex-LH syringe filter equipped with a 0.45 µm PTFE membrane prior to LCMS analysis.

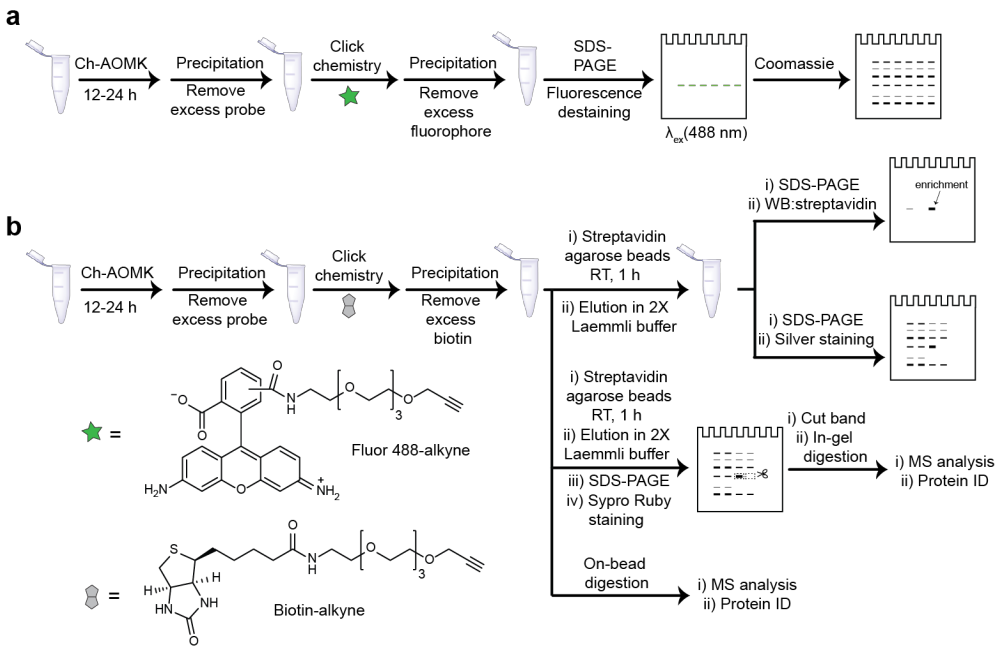
LCMS analysis was performed on an Agilent 6230 electrospray ionization–time-of-flight (ESI–TOF) MS coupled to an Agilent 1260 HPLC equipped with an Agilent Poroshell 120 ECC18 reverse phase column (3 x 50 mm, 2.7 μm) using a flow rate of 0.4 mL/min. The water, 0.1% trifluoroacetic acid (TFA):acetonitrile, 0.1% TFA gradient varied from 80:20 to 70:30 over 1.5 min, after which it ramped up to 0:100 in 4 min and remained at 0:100 for an additional 2 min. For detection, the MS was equipped with a dual ESI source operating in negative mode, acquiring in extended dynamic range from m/z 100–3200 at one spectrum per s; gas temperature: 325 °C; drying gas 10 L/min; nebulizer: 20 psig; fragmentor: 325 V.

Quantification of bile acids was performed by integrating the extracted ion count of the exact masses of the bile acids, which were determined using commercial standards. By comparing these data to standard curves in which known amounts of metabolite were injected on the LCMS, we quantified the abundance of each metabolite in the samples and normalized the amounts by the weight of the lyophilized fecal pellets.

Statistics: Each experiment was carried out at least three independent times as indicated. Error bars represent standard deviation from the mean. For box plots, interquartile ranges (IQRs, boxes), median values (line within box), whiskers (lowest and highest values within 1.5 times IQR from the first and third quartiles), and outliers beyond whiskers (dots), are shown. Statistical analyses were carried out using the Student's t-test (2-tailed) to calculate p values unless otherwise noted above. Densitometry was performed using ImageJ with normalization to the background.

Experiment	[Ch-AOMK] µM	Time (h)	Protein amount	Probe reaction vol.	pH	Click reaction vol.	Precipitation method
CP FL (time)	500	0-24	10 µg	20 µL	5.6	20 µL	MCA
CP FL (dose)	0-500	24	10 µg	20 µL	5.6	20 µL	MCA
E. coli FL	500	24	50-250 µg	20 µL	5.6	20 µL	MC
E. coli PD	500	24	2.5 mg	100 µL	5.6	100 µL	MC
BB FL	500	24	100 µg	20 µL	6.3	20 µL	MC
BL FL	500	24	100 µg	20 µL	6.3	20 µL	MC
BB PD	500	24	2.5 mg	200 µL	6.3	100 µL	MC
BL PD	500	24	2.5 mg	200 µL	6.3	100 µL	MC
BB PD MS	500	24	10 mg	800 µL	6.3	400 µL	MC
BL PD MS	500	24	10 mg	800 µL	6.3	400 µL	MC
Microbiome FL	500	24	100 µg	20 µL	5.5	20 µL	MC
Microbiome PD	100	12	2.5 mg	500 µL	5.5	100 µL	TA
Microbiome PD MS	100	12	25-50 mg	5-10 mL	5.5	1-2 mL	TA

Table 1: Conditions for Ch-AOMK labeling of bacterial samples, followed by click chemistry. Abbreviations: CP, *C. perfringens*; BB, *B. bifidum*; BL, *B. longum*; FL, fluorescence-based in gel activity assay; PD, biotin-streptavidin pull-down; MS, mass spectrometry-based proteomics; MCA, methanol-chloroform-acetone; MC, methanol-chloroform; TA, TCA-acetone.



Scheme S2: Schematic diagram of identifying BSH activity by (a) fluorescence-based in gel activity assay and (b) mass spectrometry (MS)-based proteomic analysis, following click chemistry.

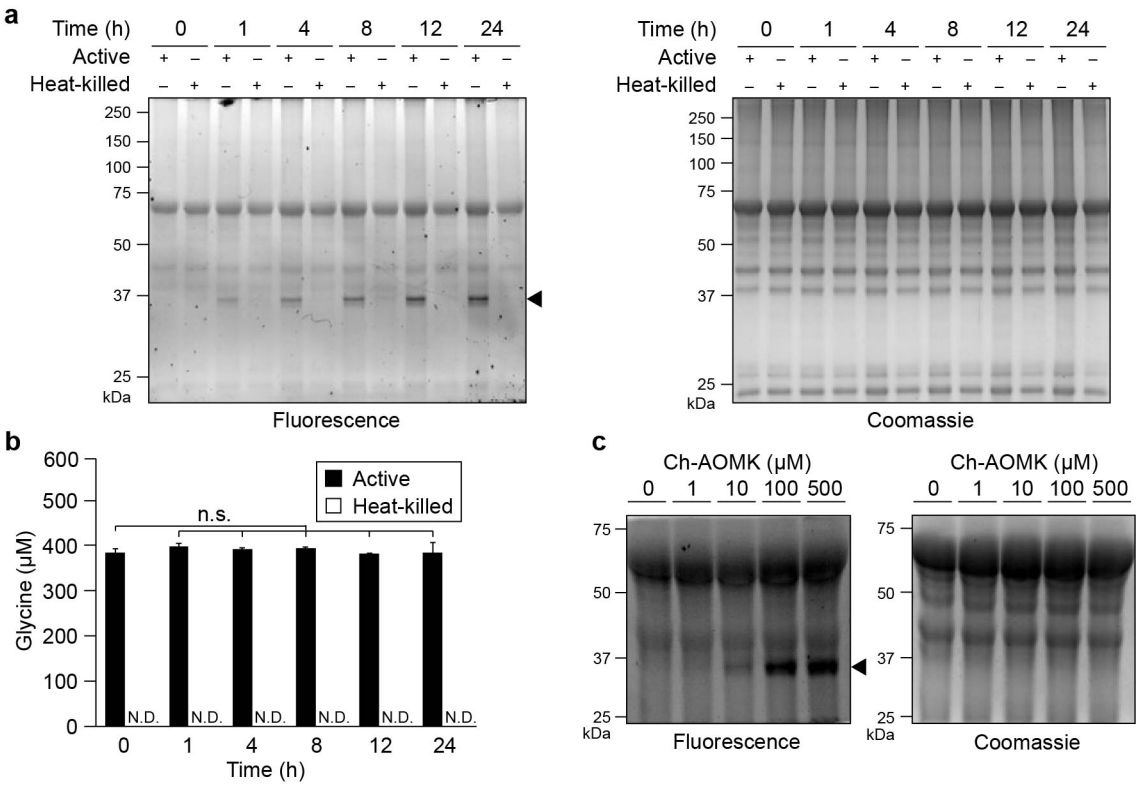


Figure S1: Ch-AOMK labels active *Clostridium perfringens* choloylglycine hydrolase (CGH). (a) Active or heat-killed CGH was treated with 500 μM of Ch-AOMK for various amounts of time as indicated at 37 °C, followed by CuAAC with Fluor 488-alkyne. The samples were purified by SDS-PAGE and visualized by fluorescence (excitation wavelength = 488 nm). The gel was stained with Coomassie brilliant blue as loading control. Arrow indicates CGH at 37 kDa. (b) Biochemical assay was performed to determine activity of CGH over 24 h, using sodium glycocholate as the substrate and glycine formation as the readout. (c) Ch-AOMK labels active CGH from *C. perfringens* (10 μg) in a dose-dependent manner. The samples were purified by SDS-PAGE and visualized by fluorescence (excitation wavelength = 488 nm). The gel was stained with Coomassie brilliant blue as loading control. Arrow indicates CGH at 37 kDa. N.D. = not detectable, n.s. = not significant, n = 3.

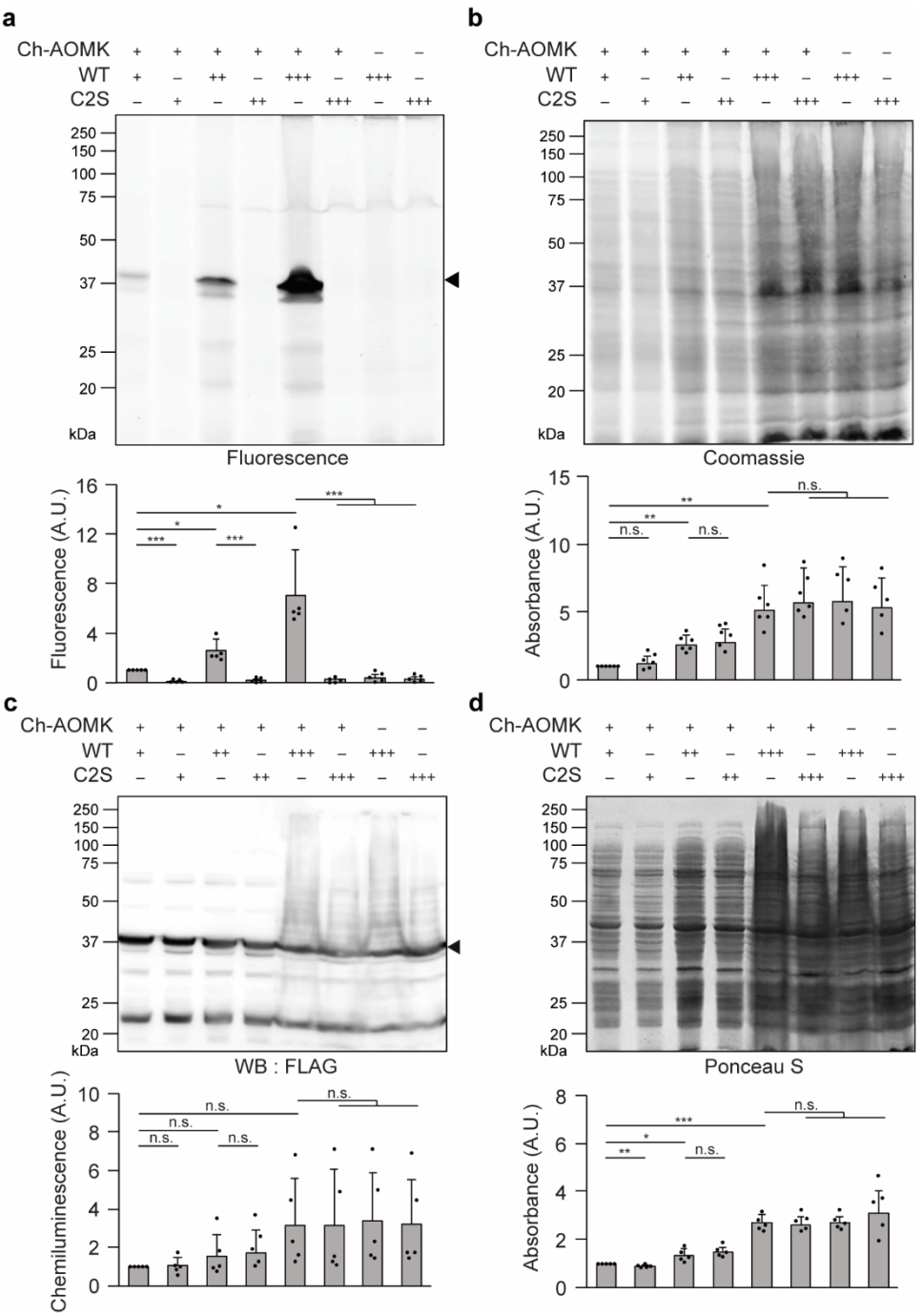


Figure S2: Ch-AOMK labels *C. perfringens* 13A CGH expressed in *E. coli* via the active site cysteine (Cys2). Ch-AOMK (0 or 500 µM) was incubated with increasing amounts of bacterial lysates (50 µg, +; 100 µg, ++; and 250 µg, +++) from *E. coli* expressing either wildtype (WT) CGH or the Cys2Ser (C2S) point mutant at 37 °C. After 24 h, (a) the samples were purified by SDS-PAGE and visualized by fluorescence, and (b) the gel was subsequently stained with Coomassie brilliant blue. Alternatively, the samples were analyzed by (c) Western blot with an anti-FLAG antibody, and (d) the blot was stained with Ponceau S as a loading control. Arrow indicates CGH at 37 kDa. All bands were quantified by densitometry using ImageJ (a-d, bottom panel). A.U. = arbitrary units. * $p<0.05$, ** $p< 0.01$, *** $p<0.001$, n.s. = not significant, n = 5-6.

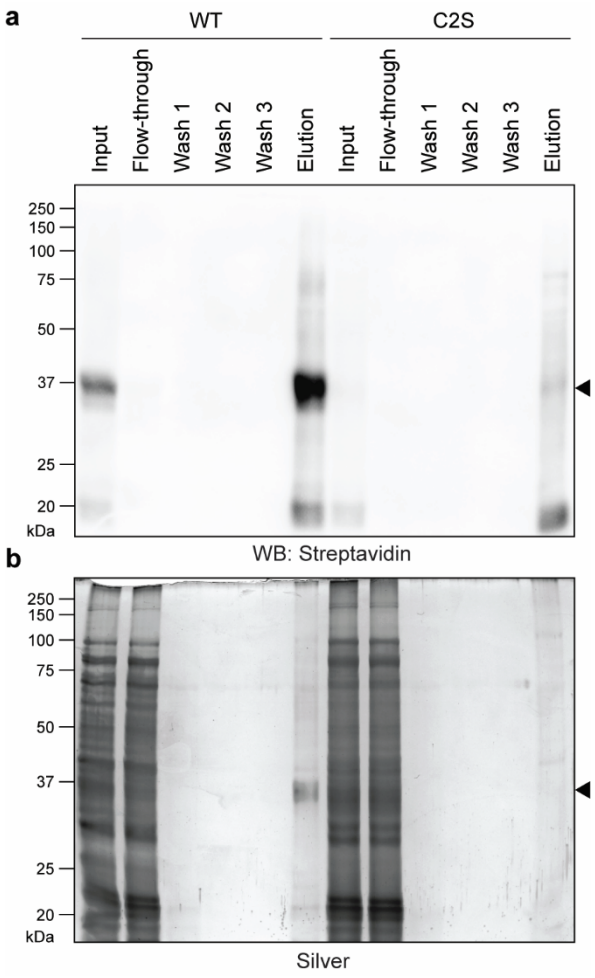


Figure S3: Ch-AOMK labels *C. perfringens* 13A CGH expressed in *E. coli*, followed by pull-down using CuAAC. Bacterial lysate (2.5 mg) from *E. coli* expressing either wildtype (WT) CGH or C2S point mutant were incubated with 500 µM of Ch-AOMK at 37 °C for 24 h. Following click chemistry with biotin-alkyne, labeled proteins were pulled down. Shown is lysate (2% of elution) as input. The samples were analyzed by (a) Western blot with streptavidin-HRP and (b) silver staining as a loading control. Arrow indicates CGH at 37 kDa.

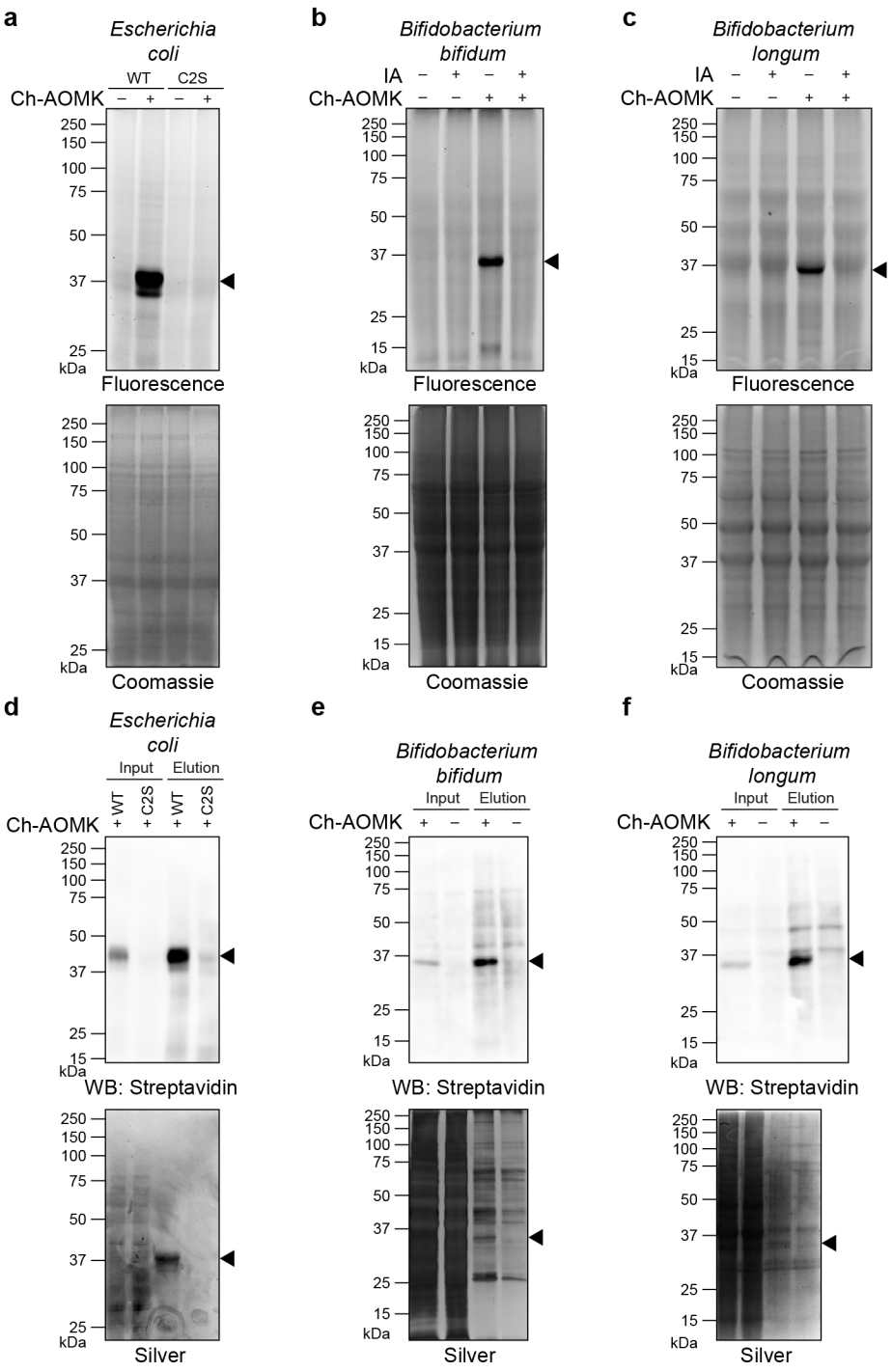


Figure S4: Ch-AOMK labels CGH or BSHs in various bacteria, including gut anaerobes. Ch-AOMK (500 μM) was incubated with lysates from (a,d) *C. perfringens* WT CGH or C2S mutant expressed in *E. coli*, (b,e) *B. bifidum*, or (c,f) *B. longum* at 37 °C for 24 h. (a-c) Lysates were treated with or without 20 mM of iodoacetamide (IA) prior to incubation with Ch-AOMK as a negative control. Following Ch-AOMK labeling, click chemistry was carried out with (a-c) Fluor 488-alkyne or (d-f) biotin-alkyne. (a-c) Samples were subjected to SDS-PAGE, and the gel was visualized using fluorescence, followed by staining with Coomassie. (d-f) Alternatively, samples

were analyzed by Western blot with streptavidin-HRP and silver staining. Arrow indicates BSH at 35-37 kDa.

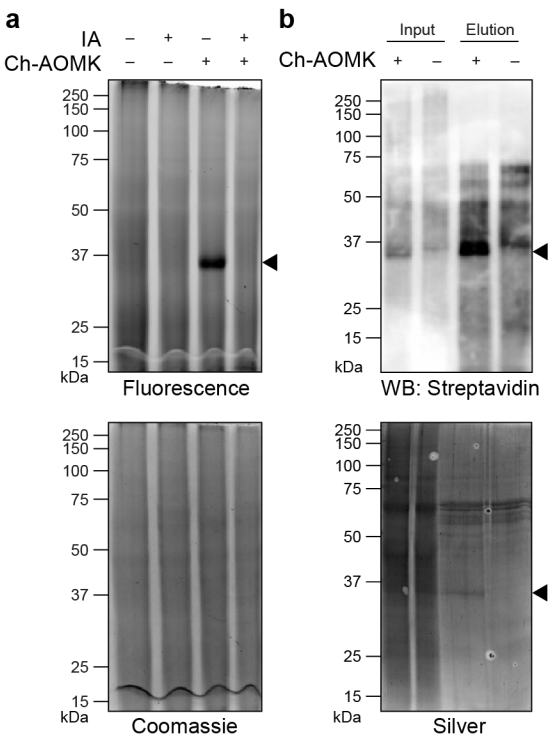


Figure S5: Ch-AOMK labels endogenous BSH in bacteria from the gut microbiome. (a) Bacterial lysate (100 µg) isolated from the mouse gut microbiome was incubated with Ch-AOMK (100 µM) 37 °C for 24 h. Following click chemistry with Fluor 488-alkyne, samples were analyzed by SDS-PAGE, followed by visualization using fluorescence. As a negative control for cysteine labeling, iodoacetamide (IA, 20 mM) was added prior to Ch-AOMK. Coomassie staining served as a loading control. Alternatively, (b) lysates (2.5 mg) were incubated with 100 µM of Ch-AOMK at 37 °C for 12 h. After click reaction with biotin-alkyne, samples were pulled down and analyzed by Western blot with streptavidin-HRP or silver staining. Arrow indicates BSH at 35 kDa.

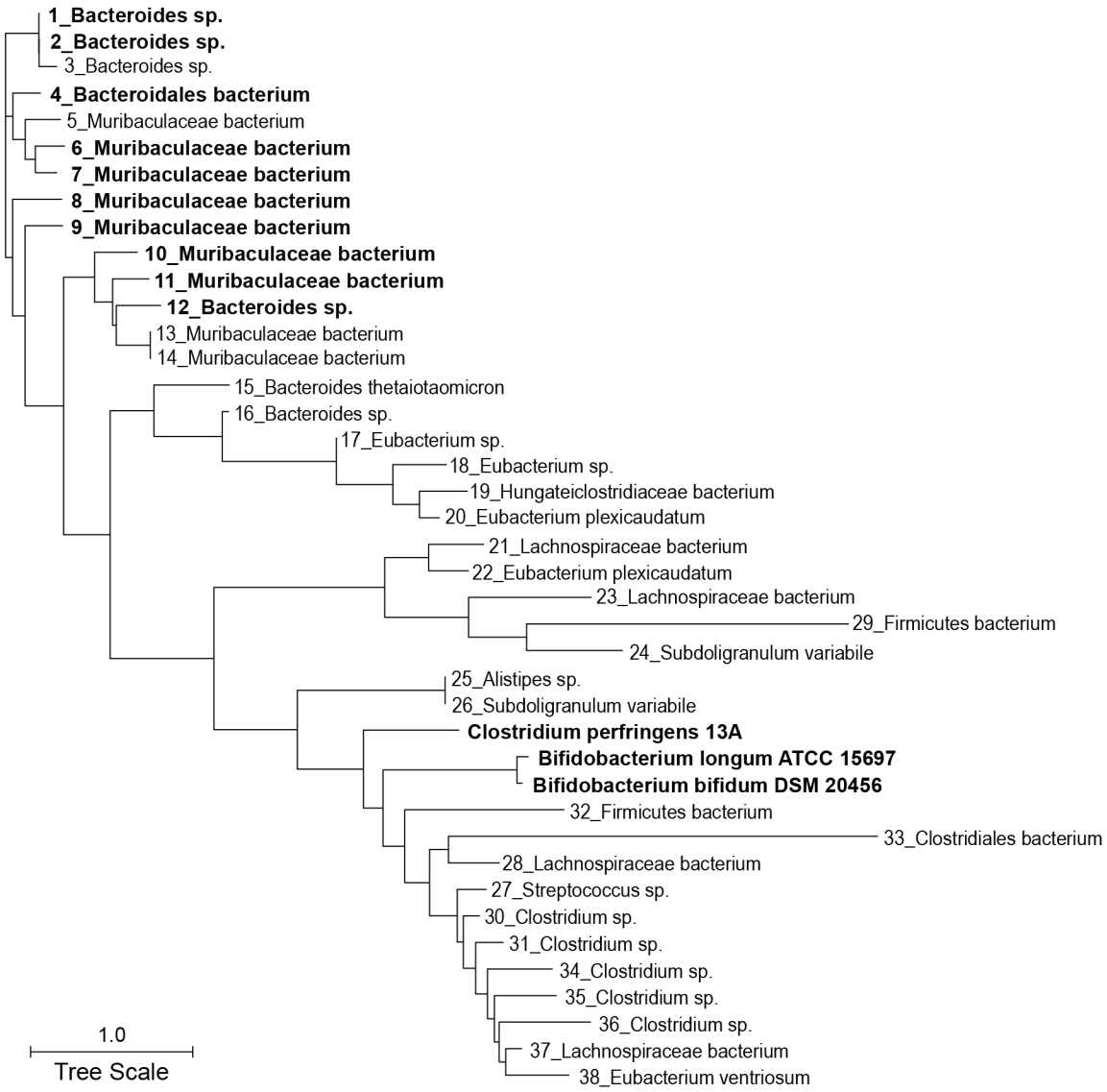


Figure S6: Phylogenetic tree of putative bile salt hydrolases (BSHs) from bacterial strains that were used in vitro (*C. perfringens*, *B. bifidum*, and *B. longum*) combined with those identified within the metagenomic assemblies of the mouse gut microbiome. Bacterial BSHs that were identified using the activity-based profiling strategy (i.e., using Ch-AOMK) are shown in bold. Tree scale represents phylogenetic distance, where 1.0 = 100% sequence identity.

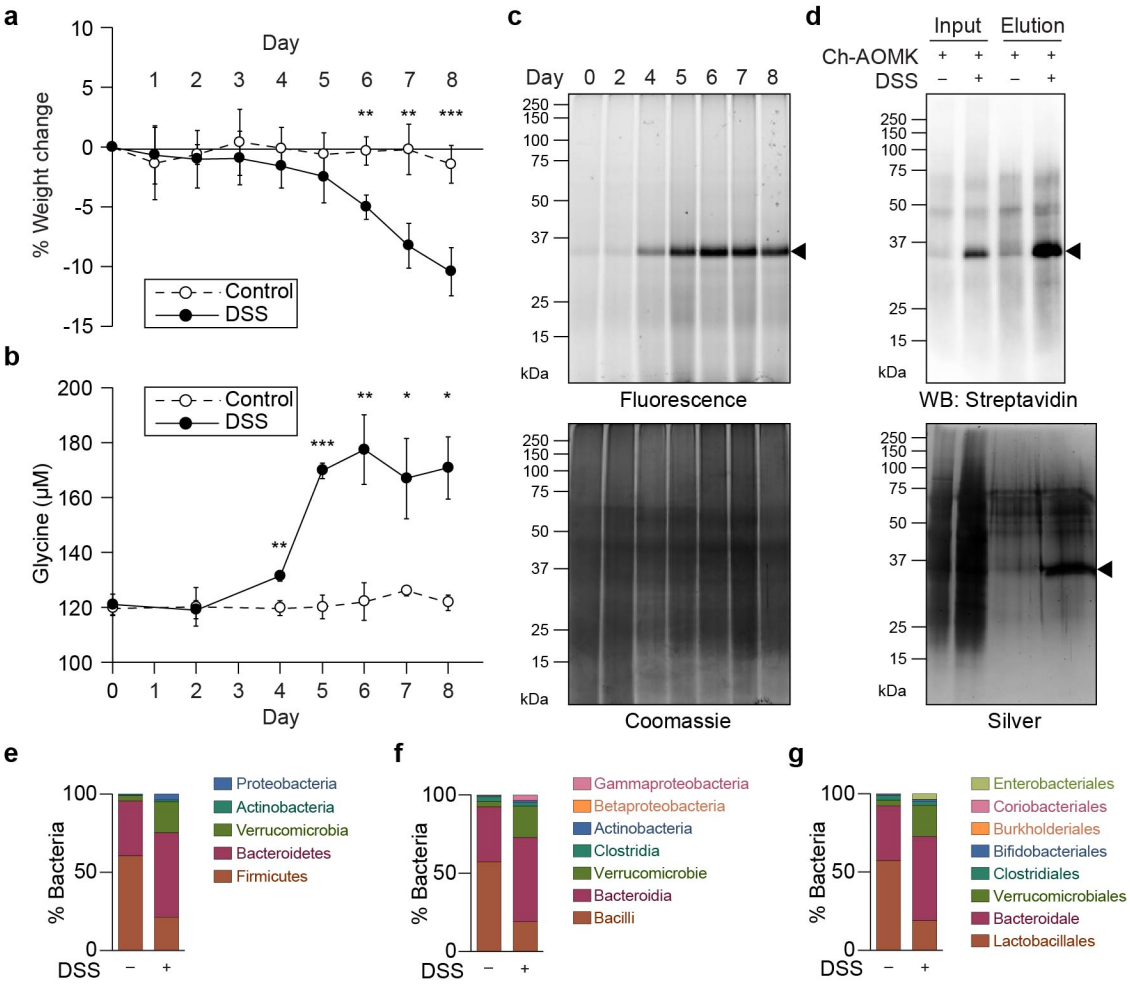


Figure S7: Ch-AOMK labeling reveals that BSH activity is upregulated in dextran sodium sulfate (DSS) colitis. C57Bl/6 mice (n = 10-17) were treated with 3% (w/v) DSS (*ad libitum*) for 8 d, during which their body weights were measured and fecal samples were collected. (a) Percent weight change was calculated over the course of the DSS treatment. (b) Biochemical assay was performed to determine CGH activity during DSS treatment, using sodium glycocholate as the substrate and glycine formation as the readout. (c) Bacteria were isolated from fecal samples, and lysates (100 μg) were incubated with Ch-AOMK (100 μM) at 37 °C for 24 h. After CuAAC tagging with Fluor 488-alkyne, samples were analyzed by SDS-PAGE, followed by visualization using fluorescence. Coomassie staining served as the loading control. Alternatively, (d) lysates from days 4-8 (2.5 mg) was incubated with Ch-AOMK (100 μM) at 37 °C for 12 h. After CuAAC tagging with biotin-alkyne, samples were enriched using streptavidin-agarose and analyzed either by Western blot with streptavidin-HRP or by silver staining. Arrowhead indicates expected size of BSH (35 kDa). (e-g) Bacterial composition was taxonomically classified at the (e) phylum, (f) class, and (g) order levels. * p< 0.05, ** p< 0.01, *** p<0.001.

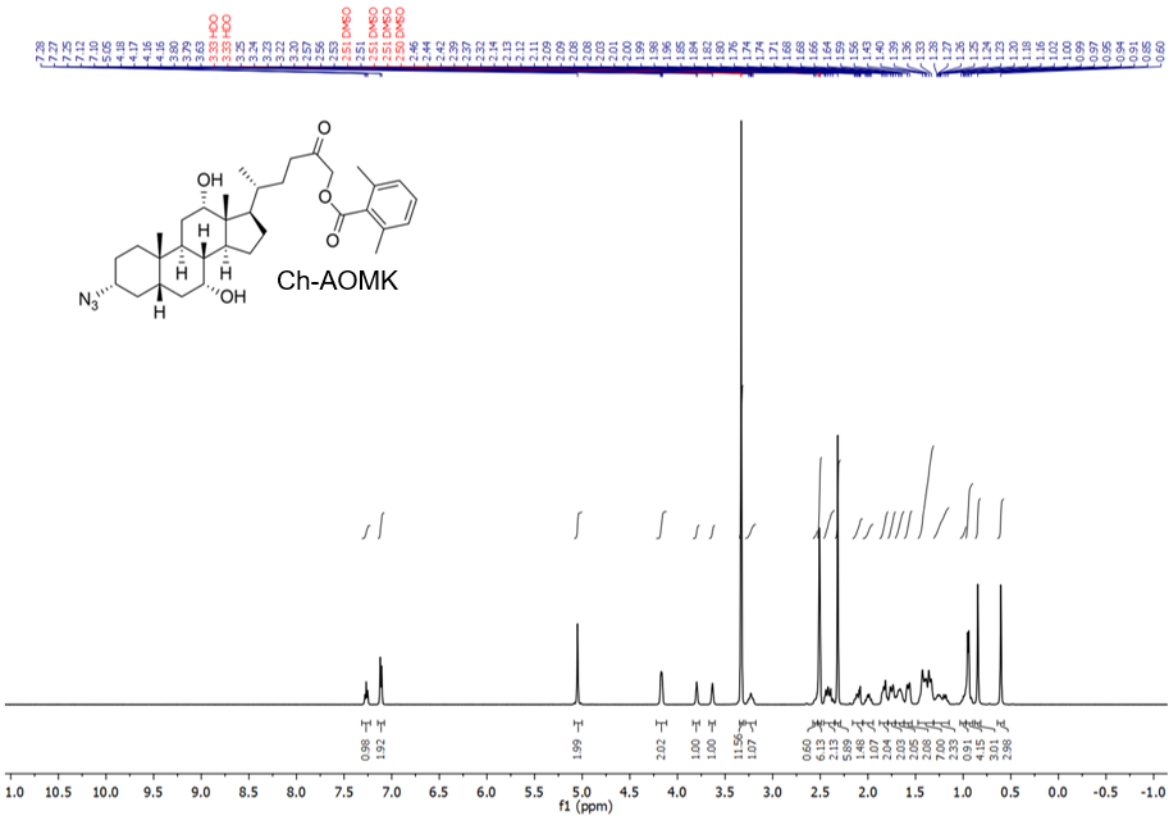


Figure S8: ^{1}H NMR spectrum of Ch-AOMK in DMSO-d6 (500 MHz)

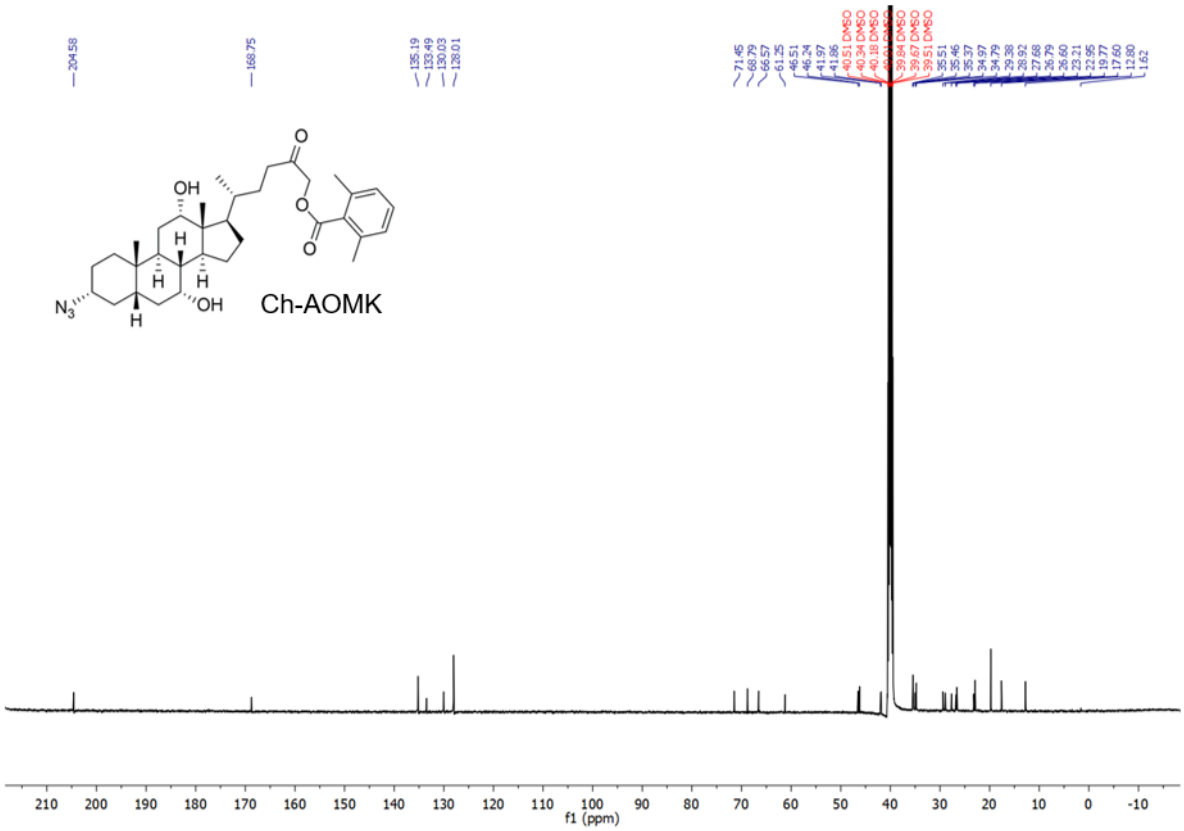


Figure S9: ^{13}C NMR spectrum of Ch-AOMK in DMSO-d6 (500 MHz)

Supplemental references

(1) Thota, B. N. S.; Savyasachi, A. J.; Lukashev, N.; Beletskaya, I.; Maitra, U. Tripodal Bile Acid Architectures Based on a Triarylphosphine Oxide Core Obtained by Copper-Catalysed [1,3]-Dipolar Cycloaddition: Synthesis and Preliminary Aggregation Studies. *European J. Org. Chem.* **2014**, *2014* (7), 1406–1415.

(2) Bond, C.; Nair, P. P.; Gordon, M.; Reback, J. The Enzymatic Cleavage of the Carbon-Nitrogen Bond in 3α,7α,12α-Trihydroxy-5β-Cholan-24-Oylglycine. *J. Biol. Chem.* **1967**, *242* (1), 7–11.

(3) Yang, Y.; Thannhausser, T. W.; Li, L.; Zhang, S. Development of an Integrated Approach for Evaluation of 2-D Gel Image Analysis: Impact of Multiple Proteins in Single Spots on Comparative Proteomics in Conventional 2-D Gel/MALDI Workflow. *Electrophoresis* **2007**, *28* (12), 2080–2084.

(4) Yang, Y.; Anderson, E.; Zhang, S. Evaluation of Six Sample Preparation Procedures for Qualitative and Quantitative Proteomics Analysis of Milk Fat Globule Membrane. *Electrophoresis* **2018**, *39* (18), 2332–2339.

(5) Thomas, C. J.; Cleland, T. P.; Zhang, S.; Gundberg, C. M.; Vashishth, D. Identification and Characterization of Glycation Adducts on Osteocalcin. *Anal. Biochem.* **2017**, *525*, 46–53.

(6) Schmieder, R.; Edwards, R. Quality Control and Preprocessing of Metagenomic Datasets. *Bioinformatics* **2011**, *27* (6), 863–864.

(7) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120.

(8) Truong, D. T.; Franzosa, E. A.; Tickle, T. L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling. *Nat. Methods* **2015**, *12* (10), 902–903.

(9) Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L. IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth. *Bioinformatics* **2012**, *28* (11), 1420–1428.

(10) Hyatt, D.; Chen, G. L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11*.

(11) Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22* (13), 1658–1659.

(12) Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Sr, E.; Griffiths-Jones, S.; Howe, K. L.; Marshall, M.; Sonnhammer, E. L. L. The Pfam Protein Families Database. *Nucleic Acids Res.* **2002**, *30* (1), 276–280.

(13) Finn, R. D.; Clements, J.; Eddy, S. R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39* (Web Server issue), W29-37.

(14) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5* (7), 621–628.

(15) Li, H.; Durbin, R. Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2010**, *26* (5), 589–595.

(16) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079.

(17) Buchfink, B.; Xie, C.; Huson, D. H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* **2014**, *12* (1), 59–60.

(18) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.

(19) Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **2010**, *5* (3), e9490.

(20) Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL): An Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics* **2007**, *23* (1), 127–128.