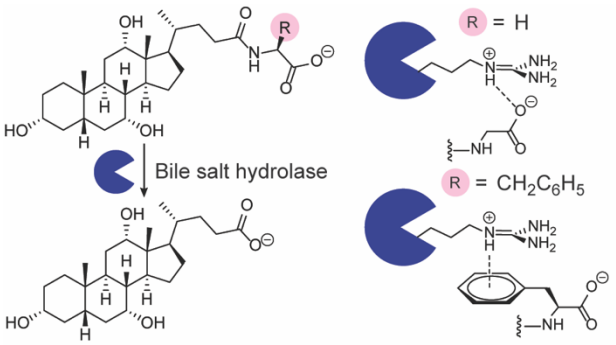
Electrostatic Interactions Dictate Bile Salt Hydrolase Substrate Preference

Kien P. Malarney and Pamela V. Chang*

*Corresponding author: pamela.chang@cornell.edu

Abstract

The human intestines are colonized by trillions of microbes, comprising the gut microbiota, which produce diverse small molecule metabolites and modify host metabolites, such as bile acids, that regulate host physiology. Biosynthesized in the liver, bile acids are conjugated with glycine or taurine and secreted into the intestines, where gut microbial bile salt hydrolases (BSHs) deconjugate the amino acid to produce unconjugated bile acids that serve as precursors for secondary bile acid metabolites. Among these include a recently discovered class of microbially-conjugated bile acids (MCBAs), wherein alternative amino acids are conjugated onto bile acids. To elucidate the metabolic potential of MCBAs, we performed detailed kinetic studies to investigate the preference of BSHs for host- and microbially-conjugated bile acids. We identified a BSH that exhibits positive cooperativity uniquely for MCBAs containing an aromatic sidechain. Further molecular modeling and phylogenetic analyses indicated that BSH preference for aromatic MCBAs is due to a substrate-specific cation-π interaction and is predicted to be widespread among human gut microbial BSHs.



TOC graphic

Introduction

Humans are colonized by a dynamic and diverse community of 100 trillion microorganisms termed the human microbiome. Microbes residing in the intestines, known as the gut microbiota, carry out important metabolic functions, including breakdown of otherwise indigestible dietary compounds and production of nutrients and signaling molecules that regulate host physiology.[1] Bile acids (BAs) are a key class of metabolites whose function is to aid in the absorption of fat-soluble, dietary vitamins and lipids and undergo extensive metabolism by the gut microbiota. These molecules also serve as signaling agents that activate host receptors and are important regulators of host metabolism[2] and immunity[3]. In addition, BAs function as antimicrobial agents against susceptible microbes and therefore impact gut microbial composition, which is correlated with host health.

BAs are synthesized from cholesterol in the liver, where they undergo conjugation with the amino acids glycine or taurine to increase water solubility (Fig. 1a). These host-conjugated BAs are postprandially secreted as emulsifying agents into the small intestine, where they encounter gut microbial bile salt hydrolases (BSHs) that catalyze amino acid deconjugation to produce unconjugated BAs (Fig. 1a).[4] Gut microbial enzymes can then further metabolize unconjugated BAs through epimerization and redox modifications to produce secondary BAs.[4] Recent work demonstrated that unconjugated BAs can also be reconjugated to a wide variety of many amino acids, including tyrosine, phenylalanine, and leucine, by microbial enzymes to produce microbially-conjugated bile acids (MCBAs, Fig. 1a).[5-11] As BSH activity is thought to precede all subsequent BA transformations, this enzyme has been termed the "gatekeeper" enzyme of secondary BA metabolism.[2,7]

Previous work has shown that BSHs have distinct substrate specificities, determined by both the identity of the conjugated amino acid[12,13] and the steroid core of the bile acid.[2,14,15] A recent report by Theriot and co-workers demonstrated that BSH side chain preference for the host-conjugated BAs is determined by a selectivity loop containing a three amino acid motif.[16] They further showed that the newly discovered MCBAs also serve as substrates for BSHs and alter host BA pools that impact the host during infection.[16]

Given the importance of BAs in human health, we sought to biochemically investigate the BSH substrate preferences for MCBAs to understand how its activity in deconjugating these novel BAs contributes to the host BA pool. Here, we utilize biochemical in vitro assays using purified BSHs from anaerobes present in the human gut microbiome, *Clostridium perfringens* and *Lactiplantibacillus* (formerly *Lactobacillus*) *plantarum*, to determine kinetic parameters and substrate specificities of several MCBAs containing side chains with unique chemical properties. We found that both BSHs preferred smaller, less sterically hindered amino acid side chains, whereas *L. plantarum* BSH1 exhibited some preference for MCBAs over host-conjugated BAs and positive cooperativity for aromatic MCBAs. From molecular modeling studies, we found that the previously characterized three amino acid selectivity loop[16] also dictates BSH1 substrate preference of aromatic MCBAs through a larger four amino acid motif and a cation-π interaction. We find that this selectivity loop is widespread in many genera throughout the human gut microbiome, suggesting that BSH MCBA deconjugation is a widespread activity that contributes to secondary BA metabolism in the intestines.

Materials and Methods

General materials and methods: Chemicals were used as received unless otherwise noted. ACS-grade acetone and ethyl acetate (EtOAc) was obtained from Fisher Scientific (Waltham, MA,

USA). Cholic acid was obtained from Beantown Chemicals (Hudson, NH, USA). Sodium taurocholate hydrate was obtained from Alfa Aesar (Haverhill, MA, USA). Sodium glycocholate hydrate and ethyl chloroformate were obtained from Sigma-Aldrich (St. Louis, MO, USA). Triethylamine was obtained from Chem-Impex (Wood Dale, IL, USA). L-leucine, L-phenylalanine, and L-tyrosine were obtained from Fluka Chemical Corporation (Buchs, Switzerland). Flash chromatography was performed using Siliaflash P60 40-63Å 230-400 mesh silica gel obtained from SiliCycle (Quebec City, Canada). Thin layer chromatography (TLC) was performed using glass-backed TLC 60 Å silica gel plates (Merck), with visualization by UV light (254 nm), followed by staining with cerium ammonium molybdate. Millex-LH syringe filters (0.45 μm PTFE membrane) were obtained from Millipore (Burlington, MA, USA). HisPur Ni-NTA resin (50% slurry) was obtained from Thermo Scientific (Waltham, MA, USA). LC-MS grade water, formic acid, and methanol were obtained from Fisher Scientific (Waltham, MA, USA). LC-MS grade acetonitrile was obtained from Millipore (Burlington, MA, USA).

Instrumental analysis: NMR spectra were acquired on a Bruker-500 spectrometer. LC-MS analysis was performed using an Agilent 6320 electrospray ionization time-of-flight MS coupled to an Agilent 1260 HPLC fitted with an Agilent InfinityLab Poroshell 120 EC-C18 column (3.0 x 50 mm, 2.7 μm mesh) with a flow-rate of 0.5 mL/min. Bile acids were separated using a binary gradient elution system of water containing 0.1% formic acid (solvent A) and acetonitrile containing 0.1% formic acid (solvent B). Briefly, the column was eluted with 5% solvent B for 0.2 min, after which solvent B was increased linearly over 5.4 min to 95% solvent B. Solvent B was then linearly increased to 100% solvent B, followed by isocratic elution for an additional 2 min. Bile acids were detected using an Agilent Jet Stream ionization source, acquiring in extended dynamic range from m/z 110-1050 in negative mode at 3 spectra/s; fragmentor voltage: 300 V;

nozzle voltage: 1000 V; Vcap: 4000 V; nebulizer: 35 psig; gas temperature: 200 °C; drying gas: 12 L/min; sheath gas flow: 12 L/min; and sheath gas temperature: 300 °C.

Kinetic and statistical analyses: Regression analyses were performed in RStudio (version 2022.12.0.353). Reported values are representative of three independent experiments. Error is reported as the standard error of the fitted value or appropriate propagated error.[17] Saturating kinetics were fitted to the Michaelis-Menten equation or Hill equation[18] using nonlinear least squares, for which initial parameters were estimated from visual inspection of kinetic plots. In both cases, k_{cat} was determined as the ratio of V_{max} and the enzyme concentration. Non-saturating kinetics were analyzed by linear regression. Briefly, V_{max} values were calculated from the maximum fitted value, and k_{cat} was determined as the ratio of V_{max} and the enzyme concentration. K_m values were calculated as the fitted concentration at which the observed half-maximal velocity was obtained. Differences in mutant and wildtype enzyme activity were compared using Welch's t-test (paired).

Molecular modeling analysis: The structure of *L. plantarum* BSH1 was generated using ColabFold-AlphaFold2.[19] Briefly, the protein was sequence (NCBI accession: CCC80500.1) was truncated to have a N-terminal cysteine[20] and inputted as a homotetramer.[21] Structure prediction parameters were left unaltered.

Docking analyses were performed in Maestro (version 13.1.137; release 2022-1; Schrödinger, New York, NY, USA). Protein and ligand preparation was performed using ProteinPrep and LigPrep, respectively.[22] The protonation state of the *L. plantarum* BSH1 Cys2 thiol was manually assigned to a thiolate.[20] Receptor grids were generated with a 12 Å inner box and 28 Å outer box, using a 5 Å distance constraint for the Cys2 sulfur[20] atom. The center of each grid was placed at the center of each monomer active site, with coordinates [33, 5, 24] and [-23, 17, 2] for *C. perfringens* CGH

and *L. plantarum* BSH1, respectively. Ligand docking was then performed with Extra Precision Glide,[23] with the constraint that 1 atom of the bile acid conjugate amide must be within 5 Å of the Cys2 thiolate. The docked pose of PheCA was used to generate template-based models for LeuCA and TyrCA complexes with *L. plantarum* BSH. In Maestro, the appropriate phenyl ring atoms were deleted and modified to an isopropyl group for LeuCA, and a *para*-hydroxyl group was added for TyrCA. For *C. perfringens*, appropriate modifications were used to generate template-based models for PheCA and LeuCA using the obtained docking pose for TyrCA. Models obtained from docking were then minimized using Prime[24] and visualized in VMD.[25] Cation-π interactions were identified using previously described geometric analysis.[26]

Bioinformatic analysis: Protein sequences of the cholylglycine hydrolase subfamily (EC 3.5.1.24) were obtained from the NCBI Protein database with a length cutoff of 400 amino acids.[27] Taxonomic assignments were then obtained from the NCBI Taxonomy database[28] using taxize.[29] Taxa were then analyzed in the NCBI BioSample database for their presence in the human gut, and only sequences from human gut-derived taxa were used for further analysis. Protein sequences were analyzed for the presence of "X_1-S-R-X_2," where X_1 = Y, F, or W and X_2 = G, S,[16] or A. Due to the presence of leader sequences,[30] sequences were filtered according to the presence of this motif beginning between the 200th and 230th position. For phylogenetic analysis, protein sequences were aligned using Clustal Omega.[31] Phylogenetic tree generation was performed using BLOSUM62[32] for distance scoring and neighbor-joining in phangorn[33] and visualized with FigTree. For genus-level analysis, sequences without genus-level assignments were omitted from the analysis.

Plasmids and cloning: A previously described protein expression plasmid for CGH from *Clostridium perfringens* 13 (NCBI accession: P54965.3) was used in this study.[34] BSH1 from

Lactiplantibacillus plantarum WCSF1 was cloned using an analogous approach. Briefly, *bsh1* from *L. plantarum* WCSF1 was PCR-amplified from a synthetic gene fragment (gBlocks, Integrated DNA Technologies, San Diego, CA, USA). Amplification using a forward primer that introduced a NdeI restriction digest site (CTAGCATATGTGTACTGCTATCACCTATCAGTCTT) and reverse primer that introduced a C-terminal FLAG tag (DYKDDDDK) and XhoI restriction digest site (CTAGCTCGAGCTTATCGTCGTCATCCTTGTAATCGTTAACTGCATAGTATTGTGCTTCT) generated an amplicon that was cloned into a pET21b plasmid containing an ampicillin selection marker.

Site-directed mutagenesis was performed using the QuikChange kit (Agilent Technologies, Santa Clara, CA, USA). For *L. plantarum* BSH1, the R207Q mutant was generated using a mutagenic forward primer (GTGGATTTAGATAGTTATAGTCAAGGAATGGGCGGACTAGG) and reverse primer (CCAGGTAATCCTAGTCCGCCCATTCCTTGACTATAACTATC). For *C. perfringens* CGH, the Q212R mutant was generated using a mutagenic forward primer (GACGGCTCTTGGACGGGGCACGGGTTTAG) and reverse primer (CTAAACCCGTGCCCCGTCCAAGAGCCGTC).

Protein expression and purification: Protein expression and purification of BSHs was performed according to literature procedure.[30,34,15] Briefly, a freshly-transformed plate of *E. coli* Rosetta 2 (DE3) pLysS competent cells on L-agar containing ampicillin was used to inoculate L-broth containing chloramphenicol and ampicillin, which was cultured with shaking at 37 °C until an OD_{600} of 0.5-0.6 was reached. The culture was then transferred to Terrific Broth (12 g/L Tryptone, 23.9 g/L yeast extract, 8 mL/L glycerol, 0.22 g/L KH_2PO_4, 0.94 g/L K_2HPO_4) containing chloramphenicol and ampicillin, which was subsequently grown with shaking at 37 °C until OD_{600}

of 0.5-0.6 was reached. Protein expression was then induced with 0.1 mM IPTG with shaking at 18 °C for 18 h.

The culture was then aliquoted to 50 mL conical tubes on ice, and cells were pelleted by centrifugation (4696 x *g*, 4 °C, 30 min). The supernatant was removed, and the cells were resuspended in lysis buffer (PBS (pH 7.4) + 5% glycerol containing 250 μM TCEP, 1 mM PMSF, and 20 mM imidazole). The cells were pelleted once more (4696 x *g*, 4 °C, 10 min), and the cells were resuspended in lysis buffer. Lysis was then performed by sonication at 0 °C, and the soluble fraction of the lysate was then isolated by centrifugation (4696 x *g*, 4 °C, 30 min).

During this time, Ni-NTA beads were equilibrated by washing with PBS (pH 7.4) + 5% glycerol containing 250 μM TCEP, and 20 mM imidazole (3x). The bacterial lysate was then incubated with the Ni-NTA beads for 90 min at 4 °C with nutation. The flow-through was then drained, and the beads were washed with PBS (pH 7.4) + 5% glycerol containing 250 μM TCEP, and 20 mM imidazole (20x). The protein was then eluted with PBS (pH 7.4) + 5% glycerol containing 250 μM TCEP, 1 mM PMSF, and 500 mM imidazole (4x). The elutions were then concentrated on an Amicon spin column (10 kDa cutoff) and quantified by the DC assay (Bio-Rad, Hercules, CA, USA). The eluted protein was then diluted in PBS (pH 7.4), aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C until further use. Purity was assessed by SDS-PAGE.

Kinetic studies: An aliquot of BSH was thawed on ice and diluted in PBS (pH 6.2). Assays were performed in PBS (pH 6.2) containing 12.5 mM DTT (assay buffer). For time-course experiments, 88 μL of assay buffer and 11 μL of conjugated bile acid (BA, 10x stock in DMSO) were added to a 1.7 mL microfuge tube and mixed. An aliquot (9 μL) was then quenched with 6 M HCl in a 1.7

mL microfuge tube as an initial timepoint. Otherwise, 80 μL of assay buffer and 10 μL of conjugated bile acid (10x stock in DMSO) were added to a 1.7 mL microfuge tube and mixed.

The BSH reaction solutions and enzyme solutions were then prewarmed to 37 °C, and 10 μL of BSH stock was added to create a 100 μL reaction (10 mM DTT, 1x BA). At each time point, a 10 μL aliquot was removed and quenched with 90 μL of 6 M HCl in a 1.7 mL microfuge tube. The aqueous layer was then extracted with EtOAc (140 μL), and the phases were separated by centrifugation (17,900 x *g*, 1 min). The organic layer (100 μL) was transferred, and the aqueous layer was extracted with EtOAc once more. The combined organic extracts were then dried under a stream of nitrogen, and the resulting residue was dissolved in LC-MS grade MeOH (1 mL). The resulting solution was then taken up in a gas-tight syringe, filtered through a 0.45 μm PTFE membrane, and subjected to LC-MS analysis with three technical replicates. Quantification of bile acids was determined by integrating the extracted ion count of the exact masses of the bile acids compared to standard curves of synthetic and commercial standards. Percent deconjugation was calculated as the ratio of unconjugated bile acids and total bile acids[2,30] detected after correcting for background hydrolysis from a PBS blank reaction. For comparison of substrate conversion between enzymes, single timepoint reactions were used for bile acid quantitation after correcting for background hydrolysis from a PBS blank reaction.

Synthetic procedures: TyrCA, PheCA, and LeuCA were synthesized according to literature procedures.[5]

Results and Discussion

To test the hypothesis that BSHs deconjugate MCBAs, we utilized cholic acid-derived MCBAs that have been identified in human fecal samples, which are conjugated to tyrosine,

phenylalanine, or leucine (TyrCA, PheCA, LeuCA; Fig. 1a).[5] Due to the importance of the steroid core in BSH substrate preference,[2,15] we characterized the activity of BSHs that are known to hydrolyze CA conjugates, including CGH[35] from *C. perfringens* 13 and BSH1[14] from *L. plantarum* WCSF1. *L. plantarum* BSH1 was recombinantly expressed as previously described for *C. perfringens* CGH (Fig. S1).[34] We then characterized their ability to hydrolyze MCBAs, along with host-derived BA conjugates, glycocholate (GCA) and taurocholate (TCA), by LC-MS analysis. Indeed, incubation of TyrCA, PheCA, and LeuCA with these BSHs for 16 h revealed modest deconjugation of MCBAs by *C. perfringens* CGH, as well as extensive deconjugation by *L. plantarum* BSH1 (Fig. 1b).

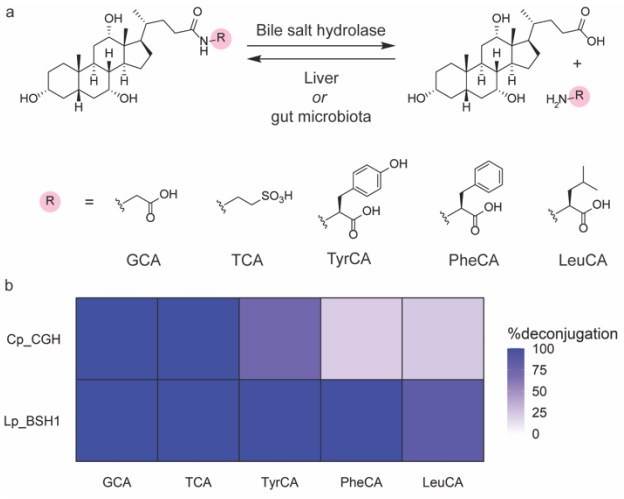


Figure 1. (a) Gut microbial bile salt hydrolases (BSHs) catalyze the hydrolysis of bile acid conjugates, whereas amino acid conjugation is performed by liver or gut microbial enzymes. (b) Activity heat map of host- and microbially-conjugated bile acid hydrolysis by *C. perfringens* CGH (Cp_CGH) and *L. plantarum* BSH1 (Lp_BSH1). BSH (387.8 nM) was incubated with 1 mM bile acid in phosphate buffered saline, pH = 6.2, with 10 mM DTT for 16 h at 37 °C. Data are representative of three independent experiments.

Although previous literature suggested *C. perfringens* CGH hydrolyzes LeuCA more readily than TyrCA,[13] we found TyrCA was hydrolyzed to a greater extent than LeuCA (Fig. 1b) because *C. perfringens* show strain-dependent substrate preferences based on the identity of the amino acid conjugate.[36,37] We also found that TyrCA and PheCA had remarkably different amounts of hydrolysis despite comparable steric hindrance near the amide bond, which was a previously identified factor for substrate preference[13] (Fig. 1b). These differences in overall MCBA hydrolysis between CGH and BSH1 suggested distinct substrate preferences (Fig. 1b). To understand these results, we characterized BSH substrate preference using kinetic studies.

We first determined conditions for which hydrolysis was linear over time for CGH (Fig. S2) and BSH1 (Fig. S3). Kinetic characterization revealed distinct substrate preferences between CGH and BSH1 (Fig. 2). CGH exhibited classical Michaelis-Menten kinetics for host-conjugated BAs (Fig. 2a-b) with TCA having a higher turnover number, lower K_m value, and a higher catalytic efficiency than GCA (Table 1). The K_m values fell within the range of known values for CGH expressed by different strains of *C. perfringens*.[37] Previous characterization of *Escherichia coli* lysates overexpressing CGH also identified this BSH as preferring taurine-conjugated BAs.[35] MCBAs were recalcitrant to hydrolysis by CGH and exhibited linear, non-plateauing kinetics (Fig. 2c-e). As such, kinetic parameters of CGH for MCBAs were determined by linear regression and were substantially lower than those for TCA and GCA (Table 1).

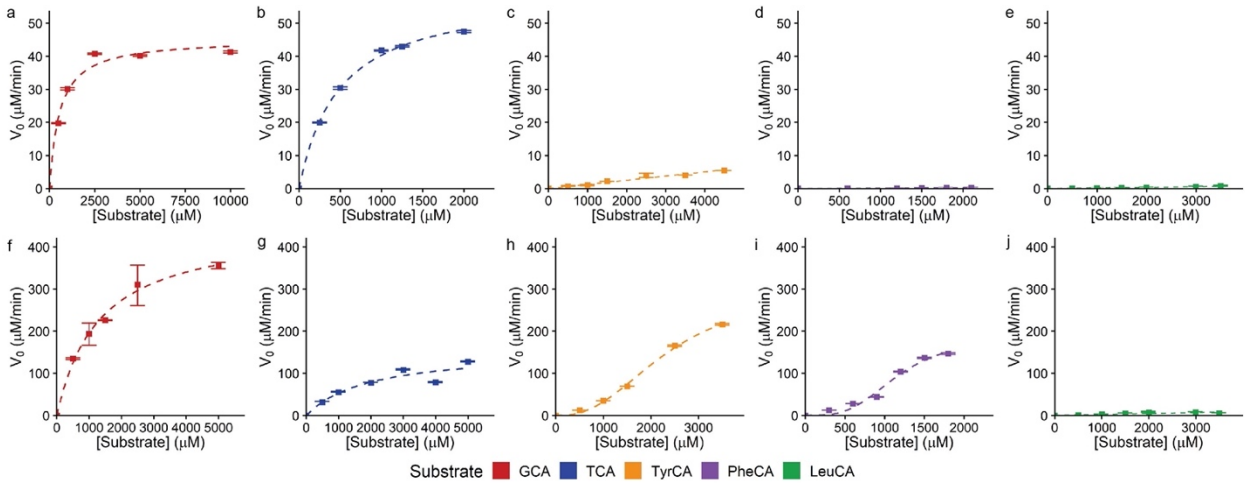


Figure 2. Kinetic characterization of *C. perfringens* CGH (a-e) and *L. plantarum* BSH1 (f-j). BSH (a-b) 2.7 nM, (c-e) 387.8 nM, (f) 9.7 nM and (g-j) 193.4 nM was incubated with the indicated amount of substrate in phosphate buffered saline, pH = 6.2, with 10 mM DTT under initial rate conditions (a,b: 5 min; c-e: 1 h; f-i: 3 min; j: 10 min) at 37 °C. Data points were fit to the Michaelis-Menten equation (a-b, f-g), Hill equation (h, i), or linear regression (c-e, j). Data are representative of three independent experiments, n=3, points = mean, error bars = standard deviation.

Kinetic characterization of BSH1 (Fig. 2f-j) revealed distinct substrate preference compared to CGH. BSH1 exhibited Michaelis-Menten kinetics for GCA and TCA as substrates, with a strong preference for GCA (Fig. 2f-g, Table 1), which has been previously described.[14] In contrast, similar studies with TyrCA and PheCA yielded sigmoidal kinetics (Fig. 2h-i) that were fit to the Hill equation[18] and yielded Hill coefficients greater than 1 for these substrates (Table 1), which indicates positive cooperativity.[18] Whereas positive cooperativity is reported for a BSH from *Enterococcus faecalis*,[18] we observed cooperativity in a substrate-dependent manner for BSH1. LeuCA, however, showed linear kinetics (Fig. 2j) and displayed the lowest catalytic efficiency (Table 1). We also determined that BSH1 had a lower catalytic efficiency for TCA compared to

TyrCA and PheCA (Table 1), demonstrating that MCBAs can be preferred substrates compared to host-conjugated BAs.

Table 1. Kinetic parameters for BSHs from *C. perfringens* and *L. plantarum*.

Enzyme	Substrate	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($min^{-1}\mu M^{-1}$)	Hill coefficient (N)
C. perfringens CGH	GCA[a]	16789 ± 392	540 ± 57	31.1 ± 3.4	-
	TCA[a]	22074 ± 317	477 ± 20	46.3 ± 2.1	-
	TyrCA[b]	14.7 ± 0.5	2169 ± 70	0.0068 ± 0.0003	-
	PheCA[b]	0.62 ± 0.03	1007 ± 354	0.0006 ± 0.0002	-
	LeuCA[b]	1.70 ± 0.07	1937 ± 449	0.0009 ± 0.0002	-
L. plantarum BSH1	GCA[a]	46474 ± 2727	1300 ± 198	35.8 ± 5.8	-
	TCA[a]	804 ± 101	1918 ± 598	0.42 ± 0.14	-
	TyrCA[c]	1547 ± 94	2347 ± 140	0.66 ± 0.07	2.5 ± 0.1
	PheCA[c]	965 ± 94	1158 ± 98	0.83 ± 0.12	3.3 ± 0.5
	LeuCA[b]	42 ± 3	1122 ± 92	0.037 ± 0.004	-

[a]Fitted using Michaelis-Menten equation. [b]Fitted using linear regression. [c]Fitted using Hill equation.

To understand the observed substrate preferences, we performed molecular docking analyses of these enzymes by modeling enzyme-substrate complexes for each substrate and enzyme. For CGH, we used a previously reported crystal structure of the apo enzyme with deoxycholate and taurine (PDB: 2BJF).[21] As the structure of BSH1 is not yet available, we used ColabFold-AlphaFold2[19] to compute a homotetrameric[16,21] structure (Figs. 3a, S4). Then, we performed ligand docking[23] of host-derived and MCBA substrates with these structures. Whereas host-conjugated BAs readily docked, only TyrCA and PheCA successfully produced binding poses for CGH and BSH1, respectively. For the remaining MCBA substrates, we used the docked MCBA substrates and modified the side chain scaffolds.[38] Finally, we performed energy minimization[24] to generate models of each enzyme-substrate complex.

The modeled complexes of CGH place the amino acid substituent of the BA conjugate in a solvent-exposed position (Fig. S5). Whereas the polar amino acid substituent is well-solvated for host-conjugated BAs, the hydrophobic sidechain of MCBAs is exposed to solvent in these models (Fig. S5). This modeling analysis also suggests TyrCA binding is stabilized by substrate-specific H-bonding (Fig. S6), which agrees with the observed preference of CGH for TyrCA compared to PheCA (Table 1).

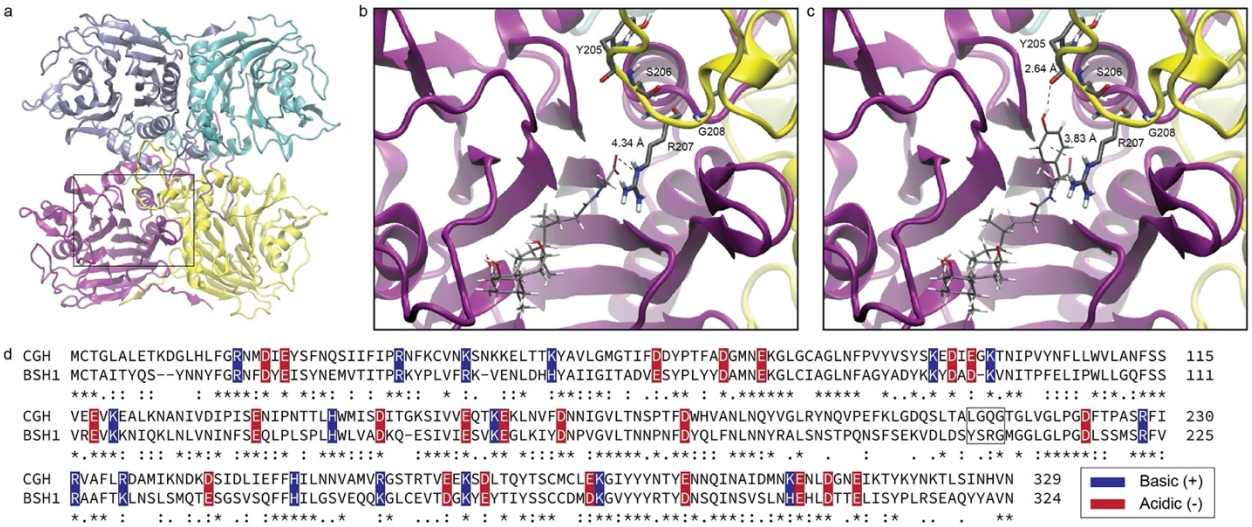


Figure 3. Molecular modeling analysis of *L. plantarum* BSH1. (a) Modeled homotetramer structure. Monomeric chains are colored purple, yellow, cyan, and blue. Modeled complexes for (b) GCA and (c) TyrCA, with interactions with selectivity loop (ball and stick, yellow loop) shown. Atom coloring (ball and stick models): Grey (carbon), red (oxygen), blue (nitrogen), white (hydrogen). (d) Sequence alignment of *C. perfringens* CGH (CGH) and of *L. plantarum* BSH1 (BSH1). Neighboring loop residues of CGH and BSH1 are highlighted (black box). Conserved negatively- and positively-charged residues are highlighted with red and blue boxes, respectively. Conserved residues are indicated with asterisks, whereas conservatively substituted and semiconservatively substituted residues are indicated with colons and periods, respectively.

The modeled complexes of BSH1 suggested that a four amino acid loop from one monomer interacts with a bile acid substrate bound to a neighboring monomer, which includes the previously reported three amino acid selectivity loop that governs substrate preference for host-conjugated BAs (Fig. 3b-c).[16] These models found that a conserved arginine residue, R207, is positioned over the BA conjugate amino acid (Fig. 3b-c) and is part of a loop with the sequence motif X_1-S-R-X_2. For GCA, the carboxylate moiety forms an electrostatic interaction with R207,

which is consistent with previous structural studies (Fig. 3b).[16] This interaction is also present for TCA (Fig. S7a) but is attenuated for LeuCA (Fig. S7b).

Interestingly, R207 formed cation-π interactions with the aryl rings of TyrCA (Fig. 3c) and PheCA (Fig. S7c) in the BSH1 models, and TyrCA engaged in an additional hydrogen bond with Y204 (Fig. 3c). To characterize these cation-π interactions, we determined geometric parameters (Fig. S8, Table S1),[26] which were consistent with empirically identified cationic arginine-π interactions in protein structures.[26] To determine the importance of the cationic arginine (R207)-π interaction in conferring BSH1 substrate preference to aromatic MCBAs, we generated a gain-of-function CGH Q212R mutant and a loss-of-function BSH1 R207Q mutant (Fig. S1) and tested substrate preference for GCA, TCA, and PheCA (Fig. S9). We found that the CGH Q212R mutant exhibited increased preference for PheCA and no change for GCA and TCA compared to wild-type CGH. However, the BSH1 R207Q mutant had no difference in substrate preferences for these conjugated BAs compared to wild-type BSH1, likely due to solvation effects.[16,21] Overall, these modeling studies suggest the observed substrate preference of BSH1 for PheCA and TyrCA is due to substrate stabilization by a cation-π interaction.

Although the structure of CGH (PDB:2BJF) used for our modeling analysis was monomeric (Fig. S5),[21] quaternary structure may be important in substrate recognition (Fig. 3a-c).[16] Indeed, a similar loop is observed for CGH in its homotetrameric structure.[21] Sequence alignment[31] of CGH and BSH1 suggest that this loop in CGH bears an overall neutral charge despite the conservation of other acidic and basic residues (Fig. 3d). This sequence analysis suggests that CGH lacks stabilizing interactions with MCBA substrates, which may contribute to its lower activity toward MCBAs compared to host-conjugated substrates (Fig. 2a-e).

As our modeling studies indicated the X_1-S-R-X_2 motif may be important for MCBA substrate recognition, we examined the distribution of this motif in BSHs encoded by the human gut microbiome. We obtained protein sequences for the cholylglycine hydrolase subfamily (EC 3.5.1.24) from NCBI and restricted the analysis to taxa identified in the human gut. Because this motif appears at positions 205-209 in *L. plantarum* BSH1, we searched for this motif at similar positions using bioinformatics analysis. This analysis revealed that this motif is found in 2,222 (24%) of the 9,452 sequences that we examined (Table S2). The aromatic MCBA selectivity loop sequence that we identified in BSH1, YSRG, was the most abundant motif, with YSRS and FSRG comprising a much smaller fraction of sequences (Table S3).

Although BSHs are widely distributed across gut bacterial phyla,[39,40] BSHs with this motif were found primarily in the *Bacillota* (formerly *Firmicutes*) phylum, which is a dominant phyla in the intestines of healthy individuals, and we observed distinct clades of BSHs with this motif (Figs. S10, 4a, Table S4). Phylum-level analysis also identified BSH sequences from human gut archaea (Fig. S10) for which BSH activity has been previously identified.[39]

Within *Bacillota*, we identified BSH protein sequences with genus-level taxonomic assignments and characterized the abundances of genera representing 1% or more of these sequences (Fig. 4b). This genus-level analysis found this motif was most abundant in BSH sequences from *Enterococcus*,[18] with other genera comprising a much smaller fraction of these sequences (Fig. 4b). We also identified this motif in genera of the *Lactobacilliceae* family[16] and the *Clostridium* genus (Fig. 4b). The additional genera that we identified are among the most abundant in the human gut microbiome (Figure 4b).[41] Collectively, our bioinformatic analysis suggests that BSHs capable of degrading MCBAs are distributed across dominant genera of the human gut microbiome.

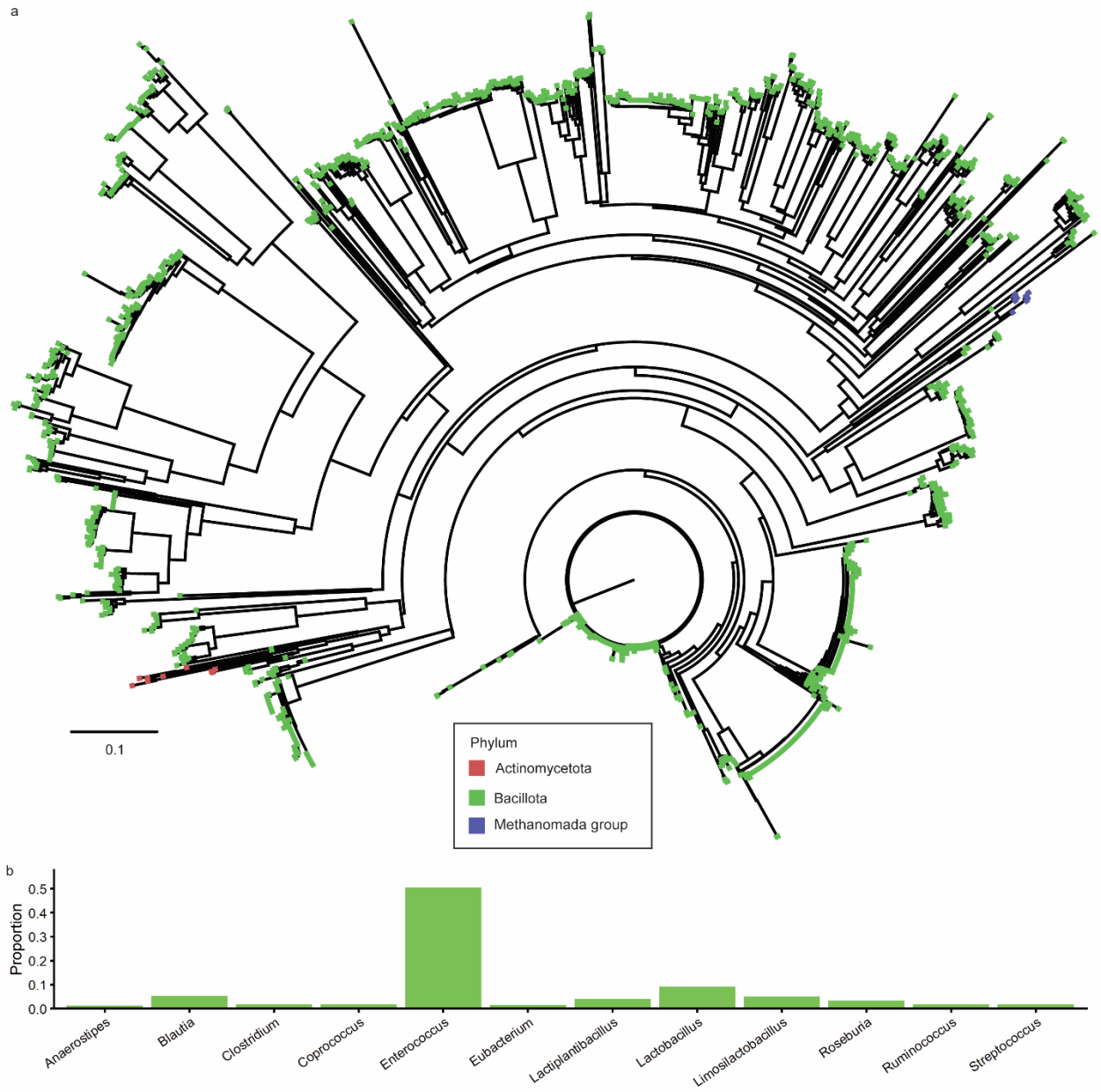


Figure 4. Phylogenetic analysis of BSH protein sequences from the human gut containing an aromatic MCBA selectivity loop. (a) Neighbor-joining tree of BSHs from MCBA selectivity loop, with tips colored by phylum. Tree scale represents phylogenetic distance determined by BLOSUM62[32] score, which is based on sequence similarity where a distance of 1.0 means 100% similar. (b) Genus-level abundances for human gut microbial *Bacillota* (formerly *Firmicutes*)

BSHs containing aromatic MCBA selectivity loop. Genera accounting for 1% or more of all sequences with assigned genera are shown.

Conclusions

In summary, BA metabolism by the gut microbiota is characterized by diverse chemical transformations and is an important process for regulating host physiology.[4] Gut microbial BA deconjugation, catalyzed by BSHs, is critical for secondary BA metabolism in the intestines. Recent studies identified BA conjugation, which was previously thought to only occur via host-encoded enzymes,[4] as a widespread transformation mediated by the gut microbiota to produce MCBAs.[5,6,7] To understand the role of BSH in deconjugation of MCBAs, we biochemically characterized the substrate preference of two representative BSHs, *L. plantarum* BSH1 and *C. perfringens* CGH, for host and gut microbial cholic acid conjugates.

Our biochemical studies found that whereas host-conjugated BAs are generally preferred substrates, BSHs may have unexpected selectivity for MCBAs with specific amino acid sidechains, as was observed for *L. plantarum* BSH1. Based on molecular modeling studies, we propose that a cation-π interaction is responsible for the higher activity of BSH1 on TyrCA and PheCA compared to host-conjugated TCA. This molecular modeling analysis also enabled the identification of a conserved four-amino acid motif that may be responsible for this interaction, which we found is widely distributed among BSHs from dominant microbial phyla and genera in the human gut. Collectively, these results suggest that differences in BSH substrate preference may enable certain gut microbes to degrade specific MCBAs and this selective MCBA-targeted activity is widely distributed across the human gut microbiome.

Associated content

Supporting Information is available free of charge at

Protein purification, kinetics, molecular modeling, characterization of mutant BSH enzymes, phylogenetic analysis, cation-pi interaction parameters, frequency of selectivity loop motif (PDF)

Author contributions

K.P.M. carried out the experiments and computational analyses. K.P.M. and P.V.C. conceived of the project idea, designed experiments, analyzed data, and wrote the manuscript.

Notes

The authors declare no competing financial interests.

Acknowledgments

This work was supported by an NIH R35 Maximizing Investigators' Research Award for Early Stage Investigators (R35GM133501). Research in the Chang Lab is supported by a Beckman Young Investigator Award (to P.V.C.) from the Arnold and Mabel Beckman Foundation and a Sloan Research Fellowship (to P.V.C.) from the Alfred P. Sloan Foundation. K.P.M. is supported by an NIH Chemistry-Biology Interface Predoctoral Training Grant (T32GM138826). This work made use of the Cornell University NMR Facility, which is supported, in part, by the NSF through MRI award CHE-1531632. We thank the Weill Institute for Cell and Molecular Biology for additional resources.

NCBI Accession Codes

CBAH (CGH): P54965.3

BSH1: CCC80500.1

Author Information

Corresponding Author

Pamela V. Chang- Department of Microbiology and Immunology, Cornell University, 930 Campus Road, Ithaca, NY 14853, United States; Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, United States; Cornell Center for Immunology, Cornell University, Ithaca, NY 14853, United States; Cornell Institute of Host-Microbe Interactions and Disease, Cornell University, 930 Campus Road, Ithaca, NY 14853, United States

orcid.org/0000-0003-3819-9994; Phone: 1-607-253-4079; Email: pamela.chang@cornell.edu

Author

Kien P. Malarney- Department of Microbiology, Cornell University, 930 Campus Road, Ithaca, NY 14853, United States; orcid.org/0000-0002-7672-4916

Complete contact information is available at:

References

(1) Donia, M. S.; Fischbach, M. A. Small Molecules from the Human Microbiota. *Science* **2015**, *349* (6246), 1254766.
(2) Yao, L.; Seaton, S. C.; Ndousse-Fetter, S.; Adhikari, A. A.; DiBenedetto, N.; Mina, A. I.; Banks, A. S.; Bry, L.; Devlin, A. S. A Selective Gut Bacterial Bile Salt Hydrolase Alters Host Metabolism. *eLife* **2018**, *7*, e37182.
(3) Fiorucci, S.; Biagioli, M.; Zampella, A.; Distrutti, E. Bile Acids Activated Receptors Regulate Innate Immunity. *Front. Immunol.* **2018**, *9*, 1853.
(4) Ridlon, J. M.; Kang, D.-J.; Hylemon, P. B. Bile Salt Biotransformations by Human Intestinal Bacteria. *J. Lipid Res.* **2006**, *47* (2), 241–259.
(5) Quinn, R. A.; Melnik, A. V.; Vrbanac, A.; Fu, T.; Patras, K. A.; Christy, M. P.; Bodai, Z.; Belda-Ferre, P.; Tripathi, A.; Chung, L. K.; Downes, M.; Welch, R. D.; Quinn, M.;

Humphrey, G.; Panitchpakdi, M.; Weldon, K. C.; Aksenov, A.; da Silva, R.; Avila-Pacheco, J.; Clish, C.; Bae, S.; Mallick, H.; Franzosa, E. A.; Lloyd-Price, J.; Bussell, R.; Thron, T.; Nelson, A. T.; Wang, M.; Leszczynski, E.; Vargas, F.; Gauglitz, J. M.; Meehan, M. J.; Gentry, E.; Arthur, T. D.; Komor, A. C.; Poulsen, O.; Boland, B. S.; Chang, J. T.; Sandborn, W. J.; Lim, M.; Garg, N.; Lumeng, J. C.; Xavier, R. J.; Kazmierczak, B. I.; Jain, R.; Egan, M.; Rhee, K. E.; Ferguson, D.; Raffatellu, M.; Vlamakis, H.; Haddad, G. G.; Siegel, D.; Huttenhower, C.; Mazmanian, S. K.; Evans, R. M.; Nizet, V.; Knight, R.; Dorrestein, P. C. Global Chemical Effects of the Microbiome Include New Bile-Acid Conjugations. *Nature* **2020**, *579* (7797), 123–129.

(6) Lucas, L. N.; Barrett, K.; Kerby, R. L.; Zhang, Q.; Cattaneo, L. E.; Stevenson, D.; Rey, F. E.; Amador-Noguez, D. Dominant Bacterial Phyla from the Human Gut Show Widespread Ability To Transform and Conjugate Bile Acids. *mSystems* **2021**, *6* (4), e00805-21.

(7) Neugebauer, K. A.; Okros, M.; Guzior, D. V.; Feiner, J.; Chargo, N. J.; Rzepka, M.; Schilmiller, A. L.; O'Reilly, S.; Jones, A. D.; Watson, V. E.; Luyendyk, J. P.; McCabe, L. R.; Quinn, R. A. Baat Gene Knockout Alters Post-Natal Development, the Gut Microbiome, and Reveals Unusual Bile Acids in Mice. *J. Lipid Res.* **2022**, *63* (12), 100297.

(8) Shalon, D.; Culver, R. N.; Grembi, J. A.; Folz, J.; Treit, P. V.; Shi, H.; Rosenberger, F. A.; Dethlefsen, L.; Meng, X.; Yaffe, E.; Aranda-Díaz, A.; Geyer, P. E.; Mueller-Reif, J. B.; Spencer, S.; Patterson, A. D.; Triadafilopoulos, G.; Holmes, S. P.; Mann, M.; Fiehn, O.; Relman, D. A.; Huang, K. C. Profiling the Human Intestinal Environment under Physiological Conditions. *Nature* **2023**, *617* (7961), 581–591.

(9) Patterson, A.; Rimal, B.; Collins, S.; Granda, M.; Koo, I.; Solanka, S.; Hoque, N.; Gentry, E.; Yan, T.; Bisanz, J.; Krausz, K.; Desai, D.; Amin, S.; Rocha, E.; Coleman, J.; Shah, Y.; Gonzalez, F.; Heuvel, J. V.; Dorrestein, P.; Weinert, E. *Bile Acids Are Substrates for Amine N-Acyl Transferase Activity by Bile Salt Hydrolase*; preprint; In Review, 2022.

(10) Guzior, D.; Okros, M.; Hernandez, C. M.; Shivel, M.; Armwald, B.; Hausinger, R.; Quinn, R. *Bile Salt Hydrolase/Aminoacyltransferase Shapes the Microbiome*; preprint; In Review, 2022.

(11) Garcia, C. J.; Kosek, V.; Beltrán, D.; Tomás-Barberán, F. A.; Hajslova, J. Production of New Microbially Conjugated Bile Acids by Human Gut Microbiota. *Biomolecules* **2022**, *12* (5), 687.

(12) Batta, A. K.; Salen, G.; Shefer, S. Substrate Specificity of Cholylglycine Hydrolase for the Hydrolysis of Bile Acid Conjugates. *J. Biol. Chem.* **1984**, *259* (24), 15035–15039.

(13) Huijghebaert, S. M.; Hofmann, A. F. Influence of the Amino Acid Moiety on Deconjugation of Bile Acid Amidates by Cholylglycine Hydrolase or Human Fecal Cultures. *J. Lipid Res.* **1988**, *27* (7), 742–752.

(14) Lambert, J. M.; Bongers, R. S.; de Vos, W. M.; Kleerebezem, M. Functional Analysis of Four Bile Salt Hydrolase and Penicillin Acylase Family Members in *Lactobacillus Plantarum* WCFS1. *Appl. Environ. Microbiol.* **2008**, *74* (15), 4719–4726.

(15) Foley, M. H.; O'Flaherty, S.; Allen, G.; Rivera, A. J.; Stewart, A. K.; Barrangou, R.; Theriot, C. M. *Lactobacillus* Bile Salt Hydrolase Substrate Specificity Governs Bacterial Fitness and Host Colonization. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (6), e2017709118.

(16) Foley, M. H.; Walker, M. E.; Stewart, A. K.; O'Flaherty, S.; Gentry, E. C.; Patel, S.; Beaty, V. V.; Allen, G.; Pan, M.; Simpson, J. B.; Perkins, C.; Vanhoy, M. E.; Dougherty, M. K.; McGill, S. K.; Gulati, A. S.; Dorrestein, P. C.; Baker, E. S.; Redinbo, M. R.; Barrangou, R.;

Theriot, C. M. Bile Salt Hydrolases Shape the Bile Acid Landscape and Restrict Clostridioides Difficile Growth in the Murine Gut. *Nat. Microbiol.* **2023**, *8* (4), 611–628.
(17) Taylor, J. R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books.
(18) Chand, D.; Ramasamy, S.; Suresh, C. G. A Highly Active Bile Salt Hydrolase from Enterococcus Faecalis Shows Positive Cooperative Kinetics. *Process Biochem.* **2016**, *51* (2), 263–269.
(19) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19* (6), 679–682.
(20) Dong, Z.; Lee, B. H. Bile Salt Hydrolases: Structure and Function, Substrate Preference, and Inhibitor Development: Structural Basis for Designing BSH Inhibitors. *Protein Sci.* **2018**, *27* (10), 1742–1754.
(21) Rossocha, M.; Schultz-Heienbrok, R.; von Moeller, H.; Coleman, J. P.; Saenger, W. Conjugated Bile Acid Hydrolase Is a Tetrameric N-Terminal Thiol Hydrolase with Specific Recognition of Its Cholyl but Not of Its Tauryl Product ·. *Biochemistry* **2005**, *44* (15), 5739–5748.
(22) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234.
(23) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
(24) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55* (2), 351–367.
(25) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38.
(26) Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation–π Interactions in Protein–Ligand Binding: Theory and Data-Mining Reveal Different Roles for Lysine and Arginine. *Chem. Sci.* **2018**, *9* (10), 2655–2665.
(27) Déjean, G.; Tudela, H.; Bruno, L.; Kissi, D.; Rawadi, G.; Claus, S. P. Identifying a Novel Bile Salt Hydrolase from the Keystone Gut Bacterium Christensenella Minuta. *Microorganisms* **2021**, *9* (6), 1252.
(28) Schoch, C. L.; Ciufo, S.; Domrachev, M.; Hotton, C. L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; Mcveigh, R.; O'Neill, K.; Robbertse, B.; Sharma, S.; Soussov, V.; Sullivan, J. P.; Sun, L.; Turner, S.; Karsch-Mizrachi, I. NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools. *Database* **2020**, *2020*, baaa062.
(29) Chamberlain, S. A.; Szöcs, E. Taxize: Taxonomic Search and Retrieval in R. *F1000Res* **2013**, *2*, 191.
(30) Adhikari, A. A.; Seegar, T. C. M.; Ficarro, S. B.; McCurry, M. D.; Ramachandran, D.; Yao, L.; Chaudhari, S. N.; Ndousse-Fetter, S.; Banks, A. S.; Marto, J. A.; Blacklow, S. C.; Devlin, A. S. Development of a Covalent Inhibitor of Gut Bacterial Bile Salt Hydrolases. *Nat. Chem. Biol.* **2020**, *16* (3), 318–326.
(31) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of

High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539.
(32) Henikoff, S.; Henikoff, J. G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (22), 10915–10919.
(33) Schliep, K. P. Phangorn: Phylogenetic Analysis in R. *Bioinformatics* **2011**, *27* (4), 592–593.
(34) Parasar, B.; Zhou, H.; Xiao, X.; Shi, Q.; Brito, I. L.; Chang, P. V. Chemoproteomic Profiling of Gut Microbiota-Associated Bile Salt Hydrolase Activity. *ACS Cent. Sci.* **2019**, *5* (5), 867–873.
(35) Coleman, J. P.; Hudson, L. L. Cloning and Characterization of a Conjugated Bile Acid Hydrolase Gene from Clostridium Perfringens. *Appl. Environ. Microbiol.* **1995**, *61* (7), 2514–2520.
(36) 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.
(37) Gopal-Srivastava, R.; Hylemon, P. B. Purification and Characterization of Bile Salt Hydrolase from Clostridium Perfringens. *J. Lipid Res.* **1988**, *29* (8), 1079–1085.
(38) Hoffer, L.; Muller, C.; Roche, P.; Morelli, X. Chemistry-Driven Hit-to-Lead Optimization Guided by Structure-Based Approaches. *Mol. Inf.* **2018**, *37* (9–10), 1800059.
(39) Jones, B. V.; Begley, M.; Hill, C.; Gahan, C. G. M.; Marchesi, J. R. Functional and Comparative Metagenomic Analysis of Bile Salt Hydrolase Activity in the Human Gut Microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (36), 13580–13585.
(40) Guzior, D. V.; Quinn, R. A. Review: Microbial Transformations of Human Bile Acids. *Microbiome* **2021**, *9* (1), 140.
(41) MetaHIT Consortium (additional members); Arumugam, M.; Raes, J.; Pelletier, E.; Le Paslier, D.; Yamada, T.; Mende, D. R.; Fernandes, G. R.; Tap, J.; Bruls, T.; Batto, J.-M.; Bertalan, M.; Borruel, N.; Casellas, F.; Fernandez, L.; Gautier, L.; Hansen, T.; Hattori, M.; Hayashi, T.; Kleerebezem, M.; Kurokawa, K.; Leclerc, M.; Levenez, F.; Manichanh, C.; Nielsen, H. B.; Nielsen, T.; Pons, N.; Poulain, J.; Qin, J.; Sicheritz-Ponten, T.; Tims, S.; Torrents, D.; Ugarte, E.; Zoetendal, E. G.; Wang, J.; Guarner, F.; Pedersen, O.; de Vos, W. M.; Brunak, S.; Doré, J.; Weissenbach, J.; Ehrlich, S. D.; Bork, P. Enterotypes of the Human Gut Microbiome. *Nature* **2011**, *473* (7346), 174–180.