SUPPLEMENTARY INFORMATION

The complex structure of bile salt hydrolase from *Lactobacillus salivarius* reveals the structural basis of substrate specificity

Fuzhou Xu[1,2], Xiao-Jian Hu[3]*, Warispreet Singh[4], Wenjing Geng[1], Irina G. Tikhonova[4]*, and Jun Lin[1]*

[1]Department of Animal Science, The University of Tennessee, Knoxville, TN 37996. USA; [2]Institute of Animal Science and Veterinary Medicine, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China; [3] State Key Laboratory of Genetic Engineering, School of Life Sciences, Collaborative Innovative Centre of Genetics and Development,, Fudan University, Shanghai 200438, China, [4] Molecular Therapeutics, School of Pharmacy, Medical Biology Centre, Queen's University Belfast, BT9 7BL, Northern Ireland, United Kingdom

***Corresponding author:**

Dr. Xiao-Jian Hu, State Key Laboratory of Genetic Engineering, School of Life Sciences, Collaborative Innovative Center of Genetics and Development, Fudan University, Shanghai, China; Email: xjhu@fudan.edu.cn

Dr. Irina G. Tikhonova, Molecular Therapeutics, School of Pharmacy, Medical Biology Centre, Queen's University Belfast, BT9 7BL, Northern Ireland, United Kingdom, Email: i.tikhonova@qub.ac.uk

Dr. Jun Lin, Department of Animal Science, The University of Tennessee, 2506 River Drive. Knoxville, TN, USA; Email: jlin6@utk.edu

Figure S1. The active *ls*BSH-GCA and *ls*BSH-CA complex structure. The wall-eye stereo presentation of the polar interaction around the oxidative cysteine sulfonic acid OCS-2 in chain F (A), substrate GCA in chain F (B) and product CA in chain G (C). The refined density map around OCS-2, GCA, CA and ligand-bound water molecules were shown. The 2FoFc map contoured at 1.0 σ (blue) and FoFc map contoured at ± 3.0 σ (green and red) were shown. For clarity residues 128-139 are hided in (B) and (C). The polar contacts were plotted as red dot line and the length below 3.20 Å were shown.

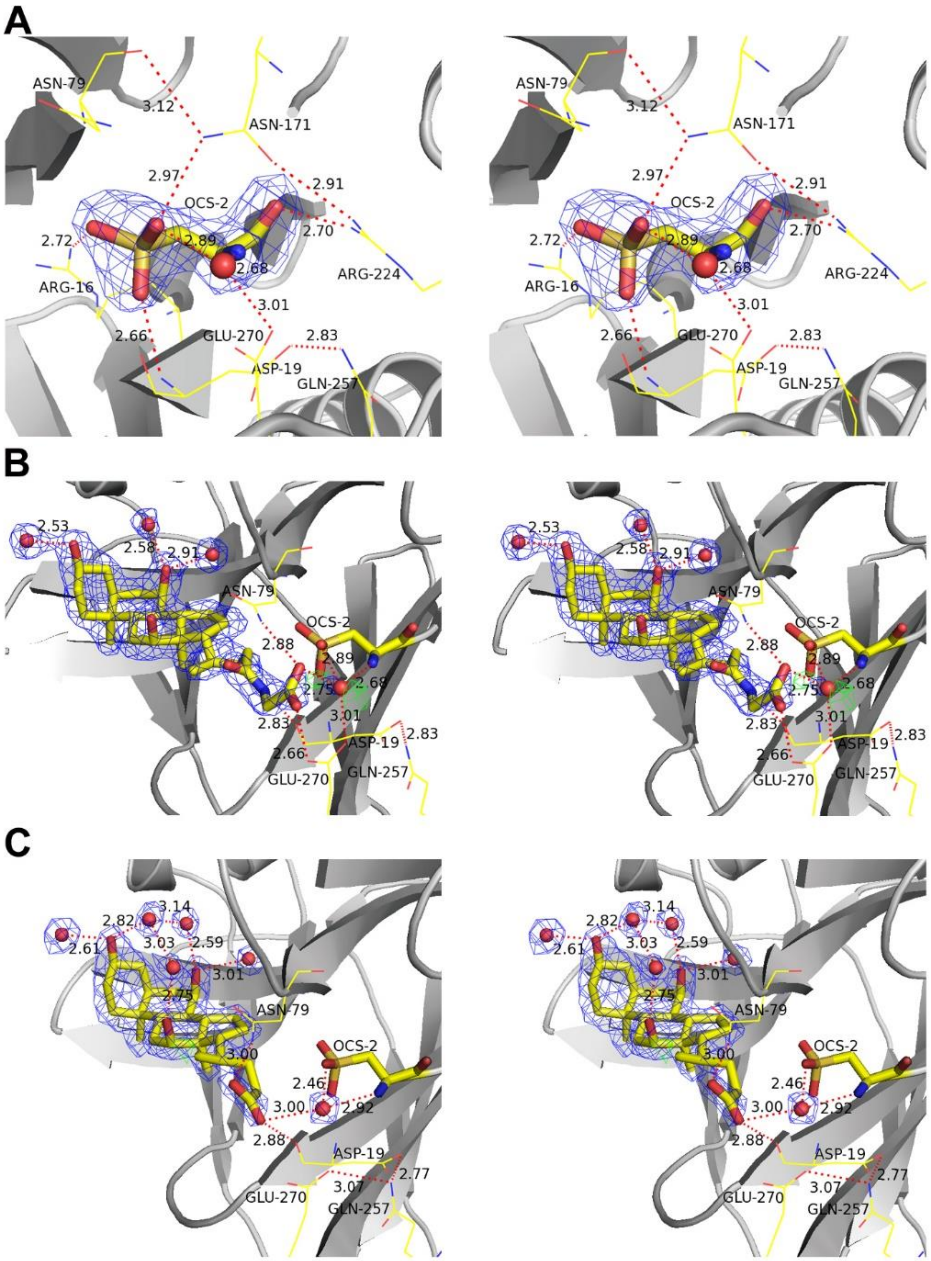
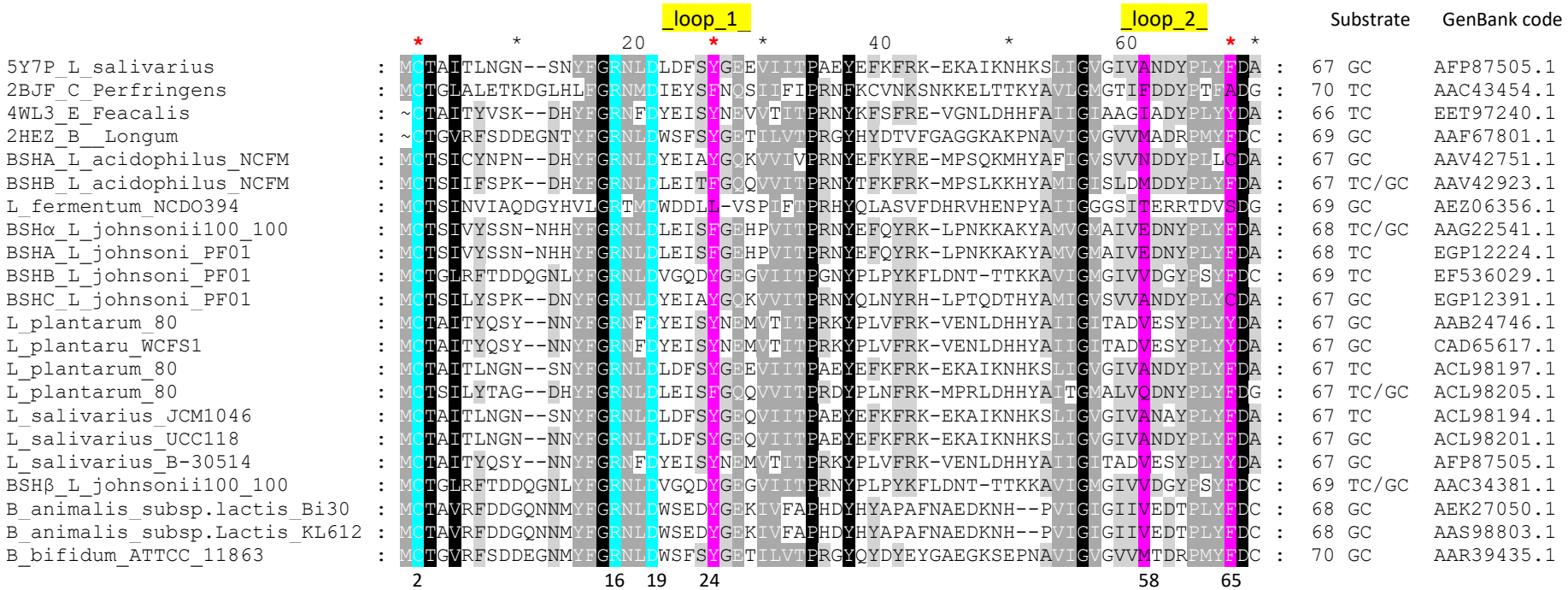
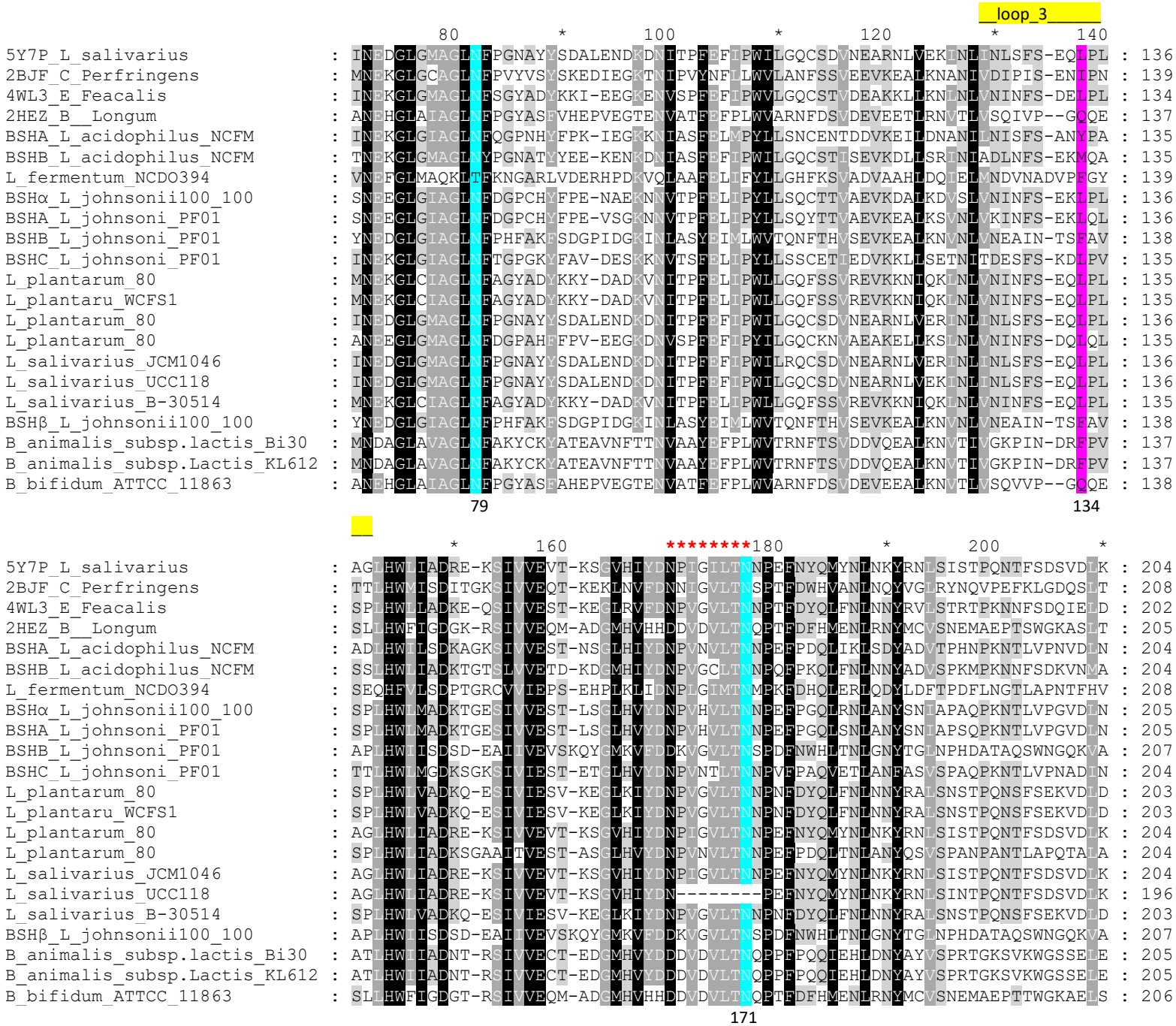


Figure S2. Sequence Alignment of BSH isotypes. The first four sequences of BSH with available crystal structures are shown with the PDB code. GenBank access code are given for all BSH sequences. Substrate preference: to glycol-conjugated bile acids, GC or tauro-conjugated bile acids, TC is shown from Dong *et al*, *Protein Science*, 2018, Vol. 27, pages 1742-1754. Loops 1-4 lining the active sites are highlighted in yellow. Residues involved in the catalytic reaction and substrate binding are in cyan and pink. Residues in pink contribute to substrate specificity. The red asterisk indicates the residues mutated in this study.




loop_3
80
100
120
140
5Y7P_L_salivarius : 136
2BJF_C_Perfringens : 139
4WL3_E_Feacalis : 134
2HEZ_B__Longum : 137
BSHA_L_acidophilus_NCFM : 135
BSHB_L_acidophilus_NCFM : 135
L_fermentum_NCDO394 : 139
BSHα_L_johnsonii100_100 : 136
BSHA_L_johnsoni_PF01 : 136
BSHB_L_johnsoni_PF01 : 138
BSHC_L_johnsoni_PF01 : 135
L_plantarum_80 : 135
L_plantaru_WCFS1 : 135
L_plantarum_80 : 136
L_plantarum_80 : 135
L_salivarius_JCM1046 : 136
L_salivarius_UCC118 : 136
L_salivarius_B-30514 : 135
BSHβ_L_johnsonii100_100 : 138
B_animalis_subsp.lactis_Bi30 : 137
B_animalis_subsp.Lactis_KL612 : 137
B_bifidum_ATTCC_11863 : 138
79
134
160
180
200
5Y7P_L_salivarius : 204
2BJF_C_Perfringens : 208
4WL3_E_Feacalis : 202
2HEZ_B__Longum : 205
BSHA_L_acidophilus_NCFM : 204
BSHB_L_acidophilus_NCFM : 204
L_fermentum_NCDO394 : 208
BSHα_L_johnsonii100_100 : 205
BSHA_L_johnsoni_PF01 : 205
BSHB_L_johnsoni_PF01 : 207
BSHC_L_johnsoni_PF01 : 204
L_plantarum_80 : 203
L_plantaru_WCFS1 : 203
L_plantarum_80 : 204
L_plantarum_80 : 204
L_salivarius_JCM1046 : 204
L_salivarius_UCC118 : 196
L_salivarius_B-30514 : 203
BSHβ_L_johnsonii100_100 : 207
B_animalis_subsp.lactis_Bi30 : 205
B_animalis_subsp.Lactis_KL612 : 205
B_bifidum_ATTCC_11863 : 206
171

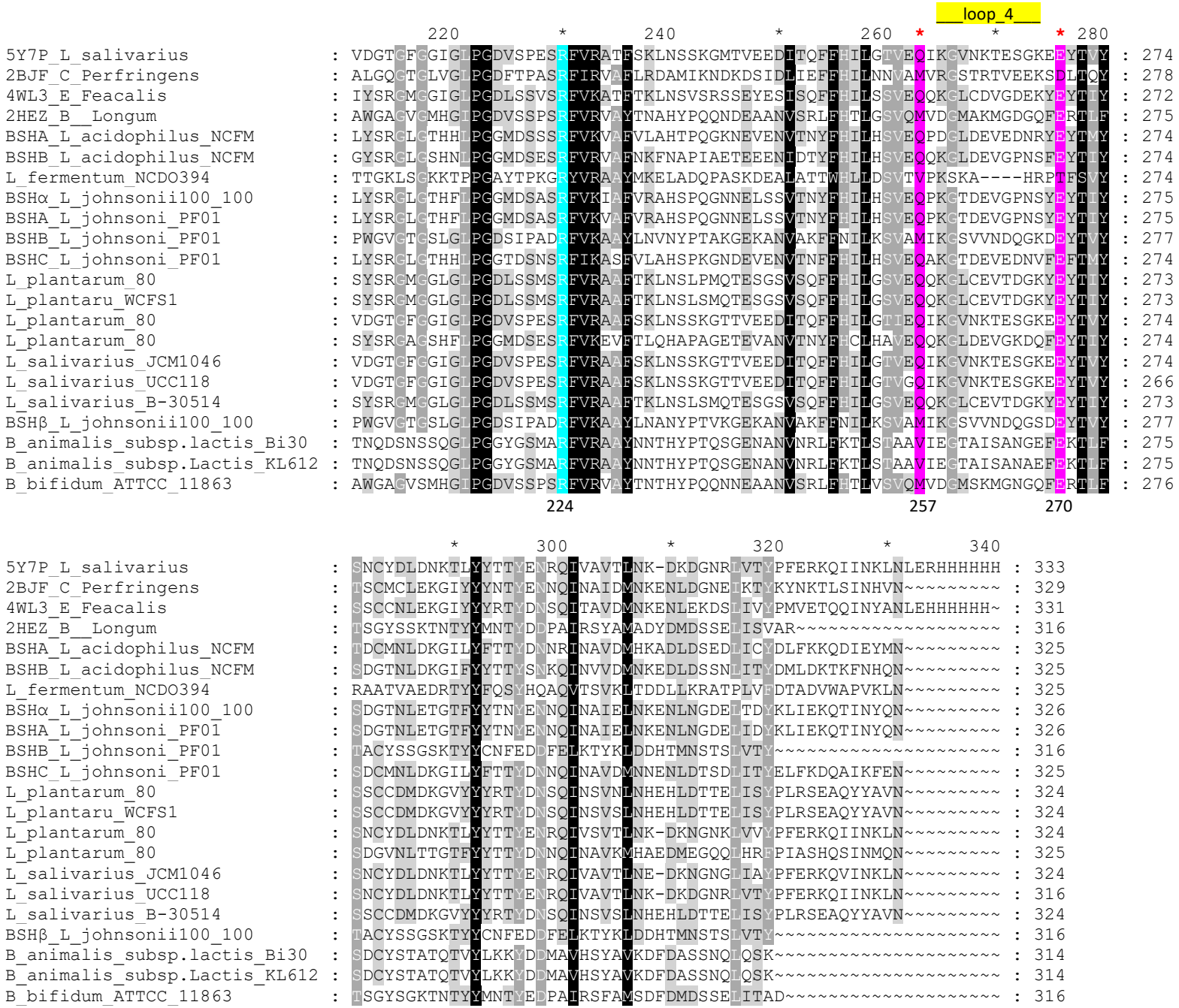

loop_4
220
240
260
280
224
257
270
300
320
340

Figure S3. Proposed hydrolysis mechanism of GCA in BSH. (a) The neutral form of the N-terminal Cys2; (b) the zwitterionic state of the N-terminal Cys2; in this state the nucleophilic thiolate anion attacks the carbonyl carbon of the amide bond of GCA; (c) the proposed transition state associated with the nucleophilic attack by the thiolate anion and the proton abstraction by the lone pair of the nitrogen of the amide bond; (d) the acylated enzyme substrate complex and the free glycine amino acid; (e) proton abstraction from the water molecule by the N-terminal primary amine; (f) nucleophilic attack of the hydroxyl group to the carbonyl carbon of the acylated enzyme substrate complex, (g) the proposed transition state associated with the nucleophilic attack by the hydroxyl anion and proton abstraction by the sulphur atom of Cys2, (i) the neutral form of the N-terminal Cys2 and cholic acid. The lone pair of the electrons are represented by the double dots; breaking and forming bonds are in dashed lines. The protein residues, the glycine fragment of GCA and the water molecule are shown in black, pink and blue, respectively.

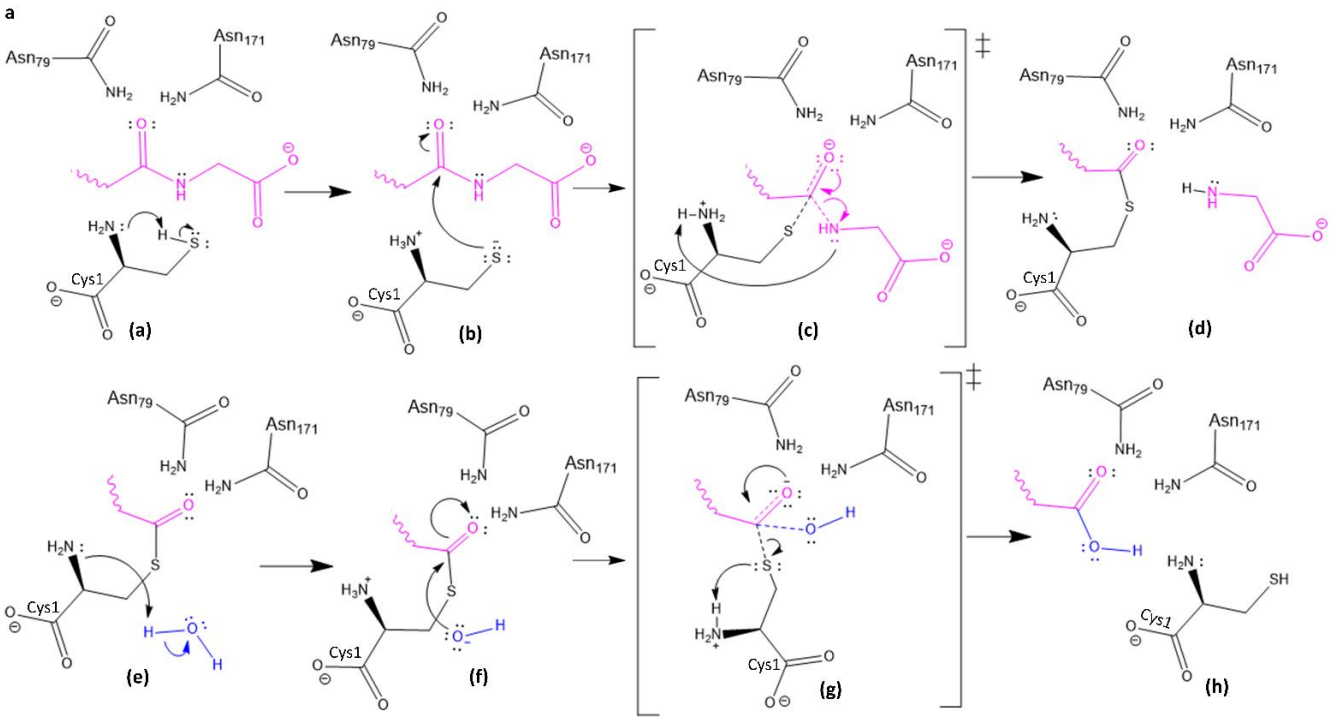


Figure S4. The active site of the *ls*BSH in complex with GCA. The snapshot of the most populated cluster obtained from the 50ns MD simulation. The GCA molecule is colored yellow and the active site residues are shown in grey stick representation. The loop region consisting of residues 164-171 are shown in red ribbon representation.

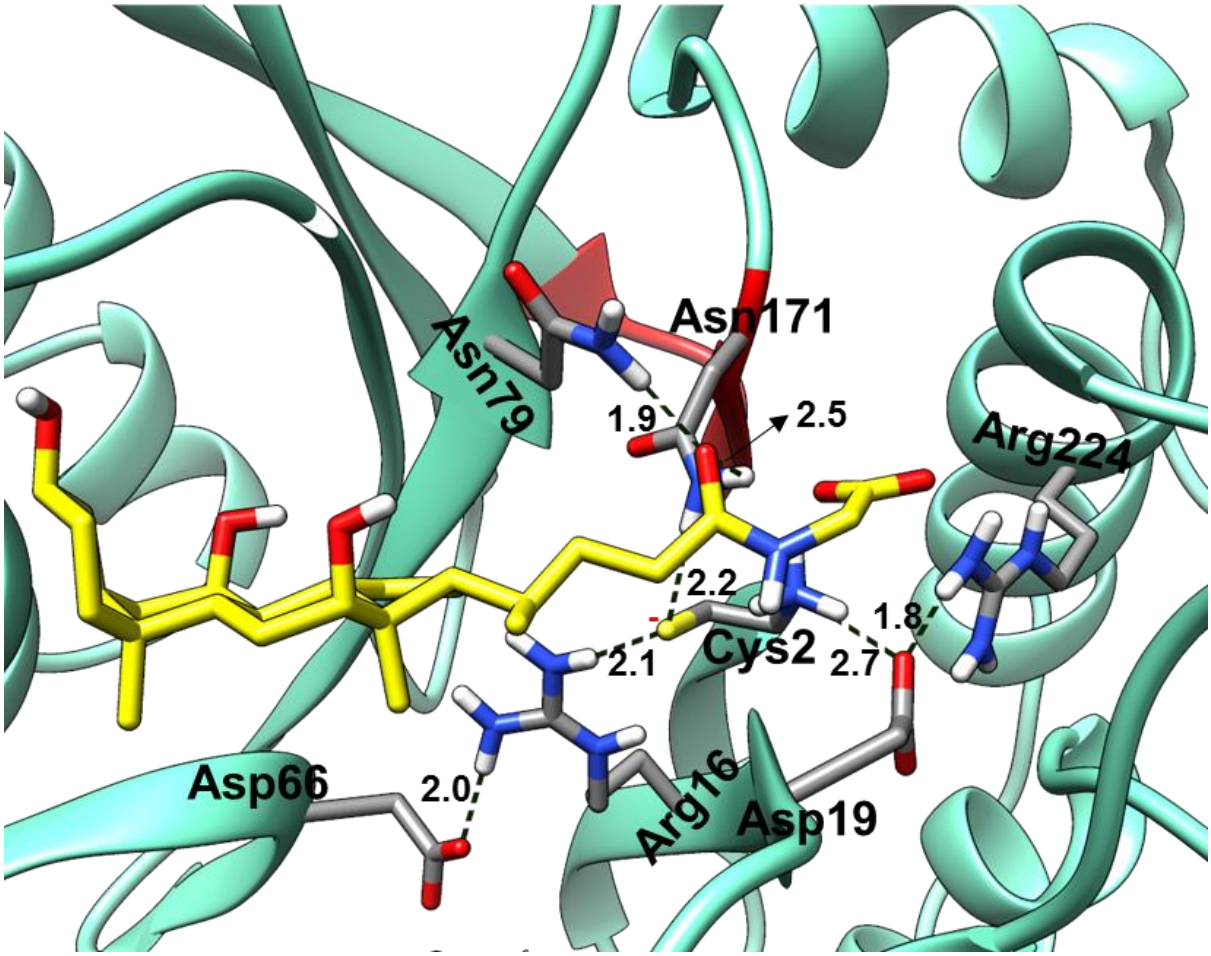


Figure S5. The surface representation of *ls*BSH bound to GCA obtained from the MD simulations. The hydrogen bonds formed by water molecules with the substrate are shown in the black dashed lines. The substrate binding cavity of the *ls*BSH is highlighted with the orange colour. The substrate GCA is shown in yellow colour.

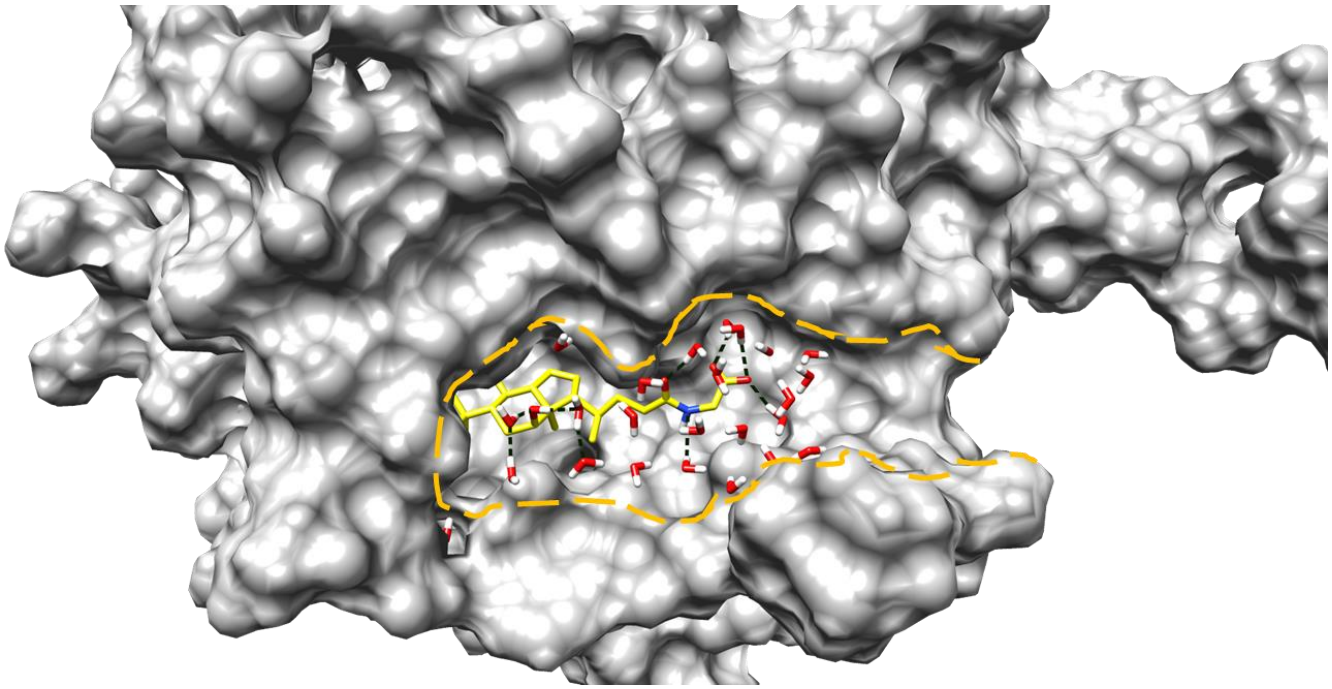


Figure S6. The substrate binding site of *ls*BSH in complex with GCA. The snapshot of the most populated cluster obtained from the 50ns MD simulation. GCA and the substrate stabilizing residues are shown in yellow and grey stick representation, respectively. The hydrophobic interactions mediated by the *ls*BSH residues with the substrate are shown in the orange dashed lines.

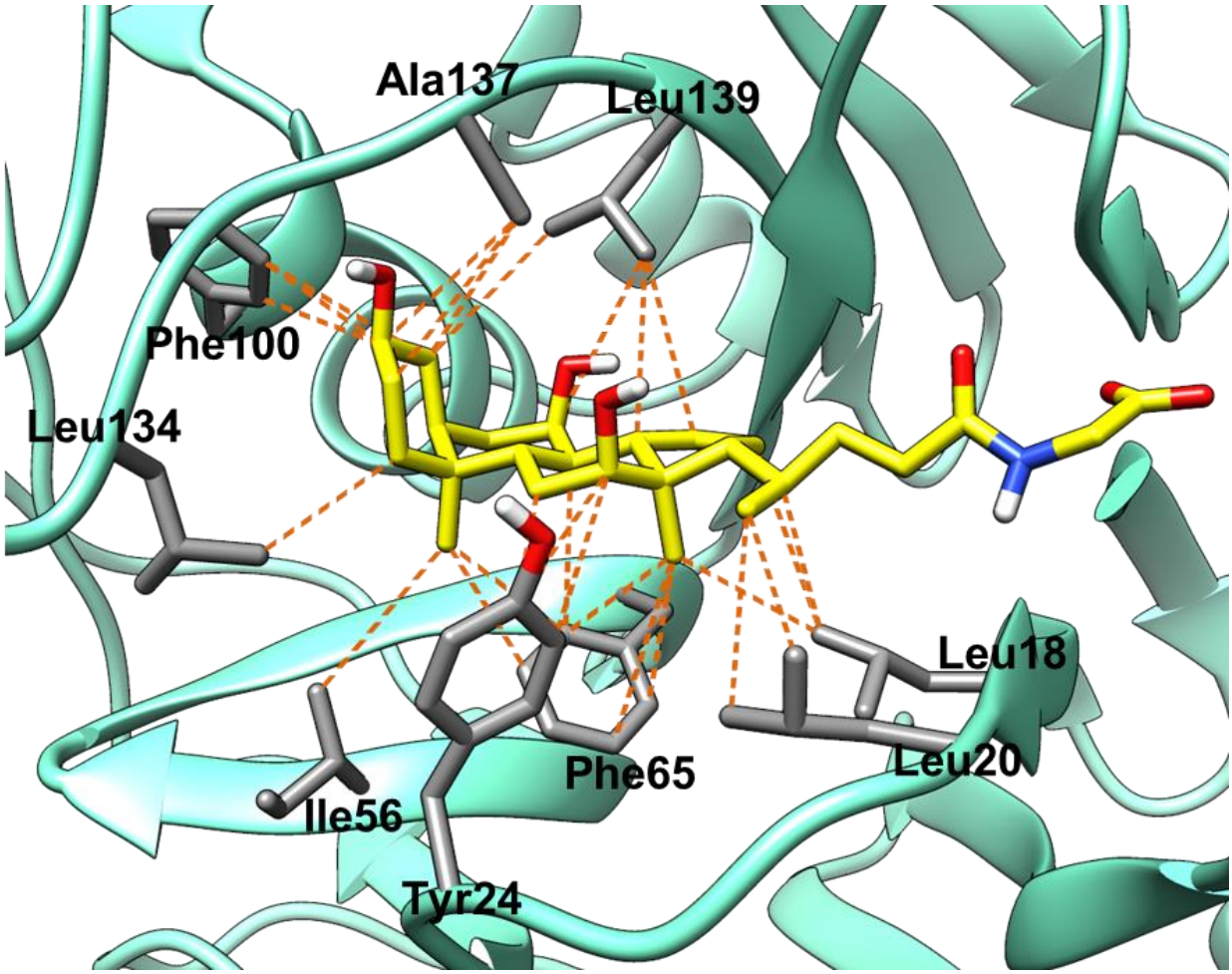


Table S1. Key bacterial plasmids and strains used in this study.

Plasmids or strains	Description	Source or Reference
Plasmids		
pBSH	pET21b containing *bsh* gene from *Lactobacillus salivarius* NRRL B-30514, Amp^r	11
pBSH (C2S)	pBSH derivative with C2S mutation	This study
pBSH (Y24F)	pBSH derivative with Y24F mutation	This study
pBSH (F65A)	pBSH derivative with F65A mutation	This study
pBSH (N171A)	pBSH derivative with N171A mutation	This study
pBSH (164-171)	pBSH derivative with the 164 -171 aa motif deleted	This study
pBSH (Q257A)	pBSH derivative with Q257A mutation	This study
pBSH (E270A)	pBSH derivative with E270A mutation	This study
pBSH (Q257A E270A)	pBSH derivative with double mutations (Q257A and E270A)	This study
Strains		
E.coli BL21(DE3)	F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Novagen
E.coli XL1-Blue	EndA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 $proAB^+$ $lacI^q$ Δ(lacZ)M15] hsdR17(r_K^- m_K^+)	Agilent Technologies
JL885	*E.coli* BL21(DE3) containing the plasmid pBSH	11
JL1189	*E.coli* BL21(DE3) containing pBSH (C2S)	This study
JL1191	*E.coli* BL21(DE3) containing pBSH (Y24F)	This study
JL1193	*E.coli* BL21(DE3) containing pBSH (F65A)	This study
JL1196	*E.coli* BL21(DE3) containing pBSH (N171A)	This study
JL1198	*E.coli* BL21(DE3) containing pBSH (164-171)	This study
JL1201	*E.coli* BL21(DE3) containing pBSH (Q257A)	This study
JL1203	*E.coli* BL21(DE3) containing pBSH (E270A)	This study
JL1205	*E.coli* BL21(DE3) containing pBSH (Q257A E270A)	This study

Figure S7. SDS-PAGE analysis of the purified wild-type (WT) *ls*BSH and its derivative with specific amino acid substitution mutagenesis or deletion.

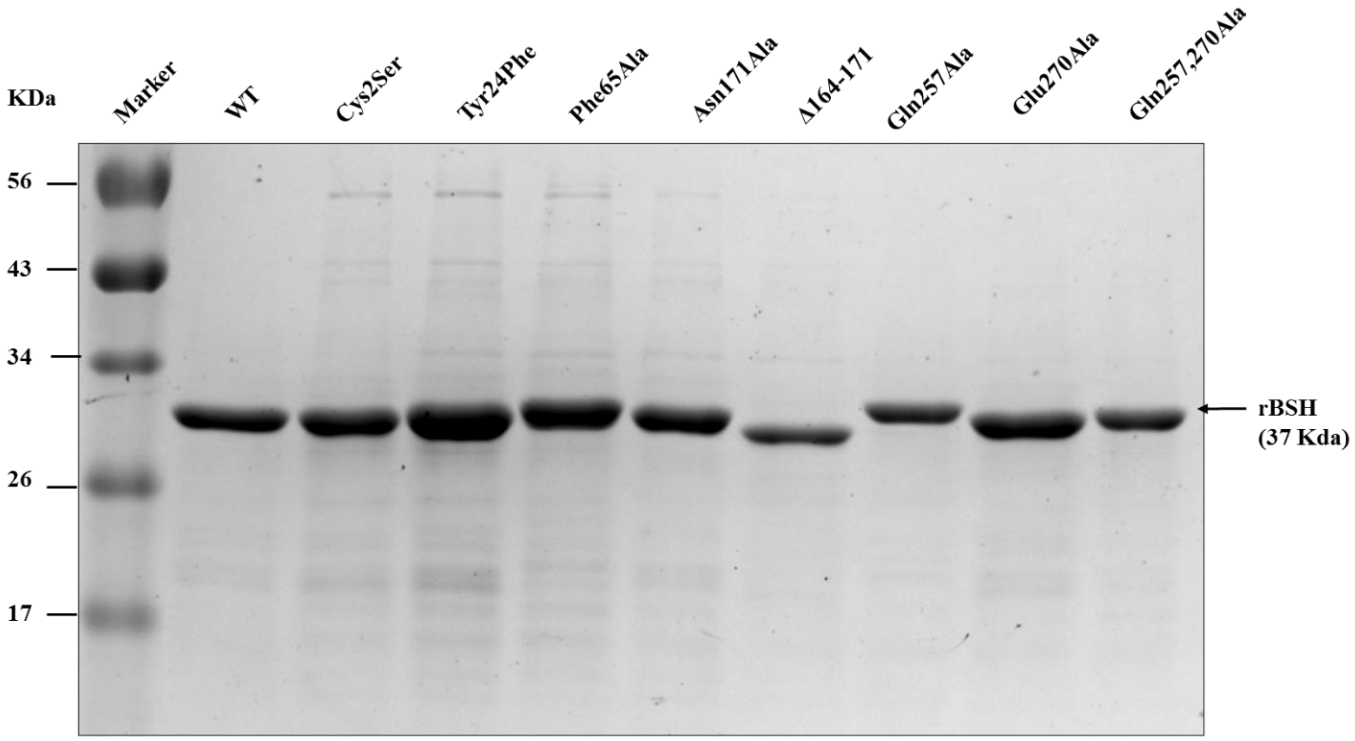


Table S2. The primers used for site-directed amino acid substitution mutagenesis in this study.

Primer Name	Sequence 5'- 3'[a]	AA Mutation
BSHMT1-F	CCATTTAAAGTAATTGCTGT**AGA**CATGGATCCCGACCCATTT	Cys2Ser
BSHMT1-R	AAATGGGTCGGGATCCATG**TCT**ACAGCAATTACTTTAAATGG	
BSHMT2-F	GATTACCTCCTCACC**AAA**TGAAAAATCTAAATCTAAGTTTCTTCCAA	Tyr24Phe
BSHMT2-R	TTGGAAGAAACTTAGA**TTT**AGATTTTTCATTTGGTGAGGAGGTAATC	
BSHMT3-F	TCCATCCTCATTAATAGCATC**TGC**ATACAATGGGTAATCGTTAGCGACAATTCC	Phe65Ala
BSHMT3-R	GGAATTGTCGCTAACGATTACCCATTGTAT**GCA**GATGCTATTAATGAGGATGGA	
BSHMT4-F	CAGATTGTACATCTGATAATTAAATTCTGGATT**TGC**AGTCAATATTCCAATTGG	Asn171Ala
BSHMT4-R	CCAATTGGAATATTGACT**GCA**AATCCAGAATTTAATTATCAGATGTACAATCTG	
BSHMT5-F	GTACATCTGATAATTAAATTCTGGATTATCATAAATATGTACTCCAGATTTAGTTACTTC	Δ164-171
BSHMT5-R	GAAGTAACTAAATCTGGAGTACATATTTATGATAATCCAGAATTTAATTATCAGATGTAC	
BSHMT6_Q257A-F	CCATATACTAGGGACAGTAGAA**GCA**ATAAAGGGCGTTAATAAGACAG	Gln257Ala
BSHMT6_Q257A-R	CTGTCTTATTAACGCCCTTTAT**TGC**TTCTACTGTCCCTAGTATATGG	
BSHMT7_E270A-F	GACAGAATCAGGAAAAGAA**GCA**TACACTGTATATTCGAATTGC	Glu270Ala
BSHMT7_E270A-R	GCAATTCGAATATACAGTGTA**TGC**TTCTTTTCCTGATTCTGTC	

[a] Bolded letters indicate the nucleotides designed for aa substitution mutagenesis

Figure S8. Comparison of the *ls*BSH binding site in the wild type (A) and Tyr24Phe mutation (B). The water-mediated hydrogen bond between Tyr24 and the hydroxyl group of GCA is shown in black dashed lines. The hydrophobic interactions between Phe24 and GCDCA are shown in the orange dashed lines.