Chapter 2B

In silico structure prediction and catalytic interaction study of probiotic marker protein: bile salt hydrolase from Lysinibacillus sphaericus

2.B.1 Introduction

Bile salts are water-soluble amphipathic signaling molecules; composed of bile acids, cholesterol, bilirubin, phospholipids and trace amount of metals. A primary bile acid is a C_{24} steroid which conjugates an amide bond with glycine or taurine at the carboxyl terminal end by the enzyme N-acyltransferase (Monte et al. 2009). Bile confers potent antimicrobial activity and emulsifies bacterial cell membrane. Probiotic generally possess bile salt hydrolase (bsh) activity as one of the effective mechanisms to survive against this type of distortion. Bile salt hydrolysis activity depends upon the presence and expression of bile salt hydrolase (*bsh*) gene in the bacterial genome. Bile salt hydrolase (cholylglycine hydrolase EC 3.5.1.24) catalyses the deconjugation of amide bond in bile salt to liberate free amino acids (Begley et al. 2006). The bsh activity has been widely detected in several autochthonous gastrointestinal probiotic microbiota including *Lactobacillus acidophilus*, *Lactobacillus plantarum*, *Bifidobacterium longum*, *Bacillus sphaericus* (Begley et al. 2006). On the otherhand, some pathogens or opportunistic pathogens, like *Listeria monocytogenes* (Dussurget et al. 2002), *Xanthomonas maltophilia* (Dean et al. 2002) and *Enterococcus faecalis* (Shankar et al. 2002) also produce bsh enzyme. The probiotics intensely controls the colonization of pathogens in intestinal lumen of host through lantibiotics production and showing bactericidal activities.

During fish farming several components such as Riboflavin, Gossypetin, Caffeic Acid Phenethyl Ester, Epicatechin monogallate were commercially used to reduce dietary lipid content, used as antioxidants, used as dietary additive to enhance fish growth etc. (Zhong and Shahidi 2011; Adewole 2014; Chen et al. 2015). Among them some showed potent inhibition against bsh enzyme which may lead to decrease the probiotic cell

concentration also. In present thesis, *Lysinibacillus sphaericus* PKA17 was identified as potent probiotic organism of *Clarias batrachus* as it showed significant improvement in fish growth parameters. Here, potent inhibitors of *Lysinibacillus sphaericus* bsh protein have been summarized, which will help to avoid fish feed or growth promoters containing those inhibitors to overcome the ineffectiveness of probiotic *L. sphaericus* in aquatic pond.

Lysinibacillus sphaericus was previously been classified as *Bacillus sphaericus*. The structural differences in major fatty acids, cell wall composition, G+C content and nitrate reduction ability differentiated them in separate genera (Ahmed et al. 2007). Initially, a work has been designed to understand the effect of speciation on bsh protein sequences of *L. sphaericus* in comparison with bsh protein sequences of different *Bacillus* species through phylogenetic analysis. Characterization of *L. sphaericus* bsh protein at its tertiary structure level was performed. The catalytic site was identified through molecular docking study with different substrates. Finally, a series of potential bsh inhibitors were identified through molecular docking. The study provides an idea to avoid such inhibitors and their natural resources during aquaculture practices to overcome ineffectiveness of probiotic organisms.

2.B.2 Materials and Methods

2.B.2.1 *In silico* characterization of bsh protein

A total of 368 *Bacillus* sp. (bshb: 4, bshc: 364) and 5 *L. sphaericus* (all bshc) bile salt hydrolase (bsh) protein sequences were obtained from UniProt database. Among them, 135 bsh protein sequences (57 of *B. cereus,* 46 of *B. thuringiensis,* 1 of *B. toyonensis,* 5 *B. anthracis,* 1 *B. subterraneus,* 1 *B. subtilis,* 1 *B. manliponensis,* 7 *B. pseudomycoides,* 1 *B. gaemokensis,* 3 *B. cytotoxicus,* 1 *B. weihenstephanensis,* 5 *B. mycoides,* 1 *B. wiedmannii,* 1

B. sp. and 4 of *L. sphaericus*) were selected for further analysis on the basis of at least one amino acid sequence differences (Banerjee et al. 2012).

2.B.2.2 Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignment was performed using ClustalX2 and bsh protein sequence based molecular evolution study was performed using PHYLIP-3.69.

2.B.2.3 Structure prediction

The tertiary structures of four sequentially different *L. sphaericus* bsh protein were predicted using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) server. The protein quality assessment of modeled structure was performed through Ramachandran plot, predicted using Profunc server (https://www.ebi.ac.uk/thornton-srv/databases/profunc/).

2.B.2.4 Substrate and inhibitor structure retrieval

The common substrates of bsh enzyme are taurodeoxycholate and sodium taurocholate. On the contrary, the inhibitors of bsh include carnosic acid, chrysophanol, epicatechin gallate, gossypetin, phenethyl caffeate and theaflavin. The 3D structures of all the above mentioned ligands were downloaded from PubChem database (https://pubchem.ncbi.nlm.nih.gov/) and used for molecular docking study.

2.B.2.5 Active site prediction

The active site of predicted bsh protein structure was identified followed by enzyme-substrate and enzyme-inhibitor interaction through molecular docking using Autodock 1.5.6 (Goodsell and Olson 1990) and Patchdock online server. Docked structures were represented using Discovery Studio Visualizer and AutoDockTools-1.5.6. The active site inhibition patterns by different inhibitors have also been performed using Patchdock online server.

2.B.3 Results and Discussion

2.B.3.1 *In silico* characterization of bile salt hydrolase (bsh) protein of *L. sphaericus*

Lysinibacillus and *Bacillus* share close affinity in the evolutionary lineage. On that basis, the work was designed to understand the sequential similarities of bsh proteins between these two genera. In addition to that structure prediction of *L. sphaericus* bsh protein through homology modeling, active site prediction of modeled structure and active site inhibition study was performed through *in silico* approach as no data about the tertiary structure of *L. sphaericus* bsh protein was available in protein structure database.

2.B.3.2 Multiple sequence alignment and phylogenetic analysis

In silico analysis of 135 selected bsh protein sequences from *L. sphaericus* and different species of *Bacillus* (Table 2.B.1), showed cluster division between them when subjected to multiple sequence alignment (Fig. 2.B.1) and phylogenetic analysis. In multiple sequence alignment all four *Lysinibacillus sphaericus* bsh proteins were found separately and showed sequence based differentiation with other *Bacillus* sp. bsh proteins. Two sequences *B. anthracis* (Q81ST7) and *B. subterraneus* (A0A0D6Z9Q3) were found at the nearby position of four *Lysinibacillus sphaericus* but with remarkable sequence diversity. Above result indicated that though these two genera share considerable level of whole genome similarity, during evolution the bsh protein sequences from two genera accumulated some remarkable changes mainly influenced by genus differentiation.

According to phylogenetic tree constructed with 500 bootstrap values, same species of *Bacillus* found in different clusters showing similarity with other species (Fig. 2.B.2). Such as *B. cereus* (A0A0K6KAY3) was observed with *B. subtilis* (G4EWN4) in first clusters whereas *B. cereus* (R8NVC4) was found in second cluster with *B. pseudomycoides*

(A0A1Y3MKU8). Li et al. (2008) showed that more sequential diversity among functionally similar proteins was due to more interaction of those proteins with external environments. In present study, the sequential diversity among all the selected bsh proteins was may be due to their interaction with different external environments.

Table 2.B.1: Retrieved *bsh* protein sequences of *Bacillus* sp. and *Lysinibacillus* sp.

***Bacillus* sp.**							
Sl. No.	Accession No.	Sl. No.	Accession No.	Sl. No.	Accession No.	Sl. No.	Accession No.
1	J8IIF3	35	A0A0K6KAY3	69	J9ASE0	103	A0A243CE13
2	A0A1Y5YXW1	36	A0A073K0W5	70	J8FI77	104	A0A242XF80
3	A0A243ME23	37	A0A2A7G6M3	71	J8M7U7	105	A0A243C1J8
4	C2T524	38	A0A2C3KQ11	72	R8EMN8	106	A0A243A6L3
5	A0A0G4CWF4	39	A0A073KTX5	73	C2UI35	107	A0A242ZWI6
6	J8KZA2	40	A0A2A7KCM2	74	B5US67	108	A0A0S3C168
7	C3E7H3	41	A0A2C2UPV2	75	A0A1S8GDX2	109	A0A1E8AFT0
8	C3H5B7	42	C2WB54	76	A0A1C6X4F1	110	A0A1E8A710
9	Q3ER69	43	A0A2B0MHU1	77	A0A090YIR7	111	A0A1B1LAT8
10	C3DNX9	44	A0A2A8S0V6	78	A0A2B5X7Q2	112	A0A0P0PH36
11	C2UZF7	45	A0A2B2TJK3	79	A0A1D3R4B1	113	A0A0K0QG55
12	C2VFY9	46	A0A1Y3MKU8	80	A0A0G8EHV1	114	A0A2A8NN63
13	K0FRA8	47	A0A2B5H8L3	81	A0A0G8ELZ6	115	A0A2A7GT64
14	A0A2B6NU77	48	Q812W2	82	A0A150C723	116	A0A2J9BRN4
15	A0A242WMT7	49	Q636A7	83	A0A150DJB3	117	A0A2J9BTM1
16	C2NLX2	50	Q732E9	84	A0A0J7GLY9	118	A0A1D3MRJ6
17	C2TKP0	51	B9IVZ6	85	A0A2B5RJ47	119	A0A1E8B3M0
18	C3G785	52	B7H6Q5	86	A0A243DRM7	120	C2ZBW2
19	C2MPP9	53	B7JK07	87	A0A242YUC3	121	Q81ST7
20	C2S706	54	B7IVF4	88	A0A2C4A1F4	122	A0A0D6Z9Q3
21	A0A0F7R7R9	55	Q6HEP5	89	A0A1K0ALU9	123	G4EWN4
22	C3GN23	56	A7GRP5	90	A0A242X720	124	C2VXY4
23	C2QX91	57	A9VU81	91	A0A2B0YBB0	125	S2LC88
24	A0A0B5NZB8	58	Q81WC2	92	A0A2C3GA55	126	J8JRT7
25	C3C6P3	59	R8NVC4	93	A0RHU0	127	J8KAR1
26	A0A1N6Z9M3	60	D8H8Q0	94	M1QQY1	128	A0A243DEK0
27	A0A1C4FP57	61	C3APM4	95	F0PRR7	129	A0A243J5X6
28	C2PJ83	62	R8RSH1	96	A0A0F6J298	130	A0A243E9D4
29	A0A1C4FNS6	63	J8GYT1	97	A0A0D1QT51	131	A0A243MYX0
30	C2YVK3	64	R8Q4I7	98	A0A243CS30	***Lysinibacillus* sp.**	
31	C2QG53	65	J9BTG0	99	A0A243NQE2	132	B1HPX8
32	C2SP07	66	R8YL42	100	W8Y7J2	133	R7ZGM4
33	C2XY32	67	J8F5M5	101	A0A243I575	134	A0A1D2W114
34	C2PZW9	68	R8MM35	102	A0A243GF07	135	A0A1D8JI08

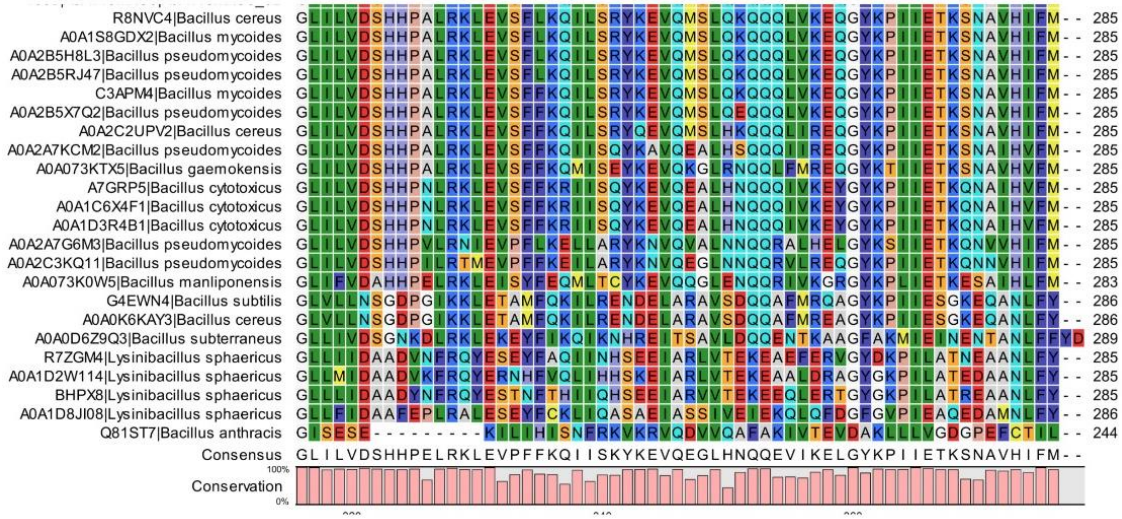


Fig. 2.B.1: Multiple sequence alignment segment showing position of *Lysinibacillus sphaericus* bsh protein sequences with *B. anthracis* (Q81ST7) and *B. subterraneus* (A0A0D6Z9Q3).

Functionally similar proteins from different genera may accumulate sequential diversity in such a manner, that they can found in different cluster in phylogeny. Banerjee et al. (2014) reported genus differentiation through phylogenetic tree on functionally similar keratinase protein sequences among *Bacillus sp., Stenotrophomonas maltophilia, Pseudomonas aeruginosa* and *Streptomyces* sp. According to the present analysis, four *Lysinibacillus sphaericus* bsh proteins were also found together in the first cluster (2.B.2). On the basis of above result, a work has been undertaken to analyze the *L. sphaericus* bsh protein separately at its molecular structure level to hypothesize the sequence-structure relationship in molecular level.

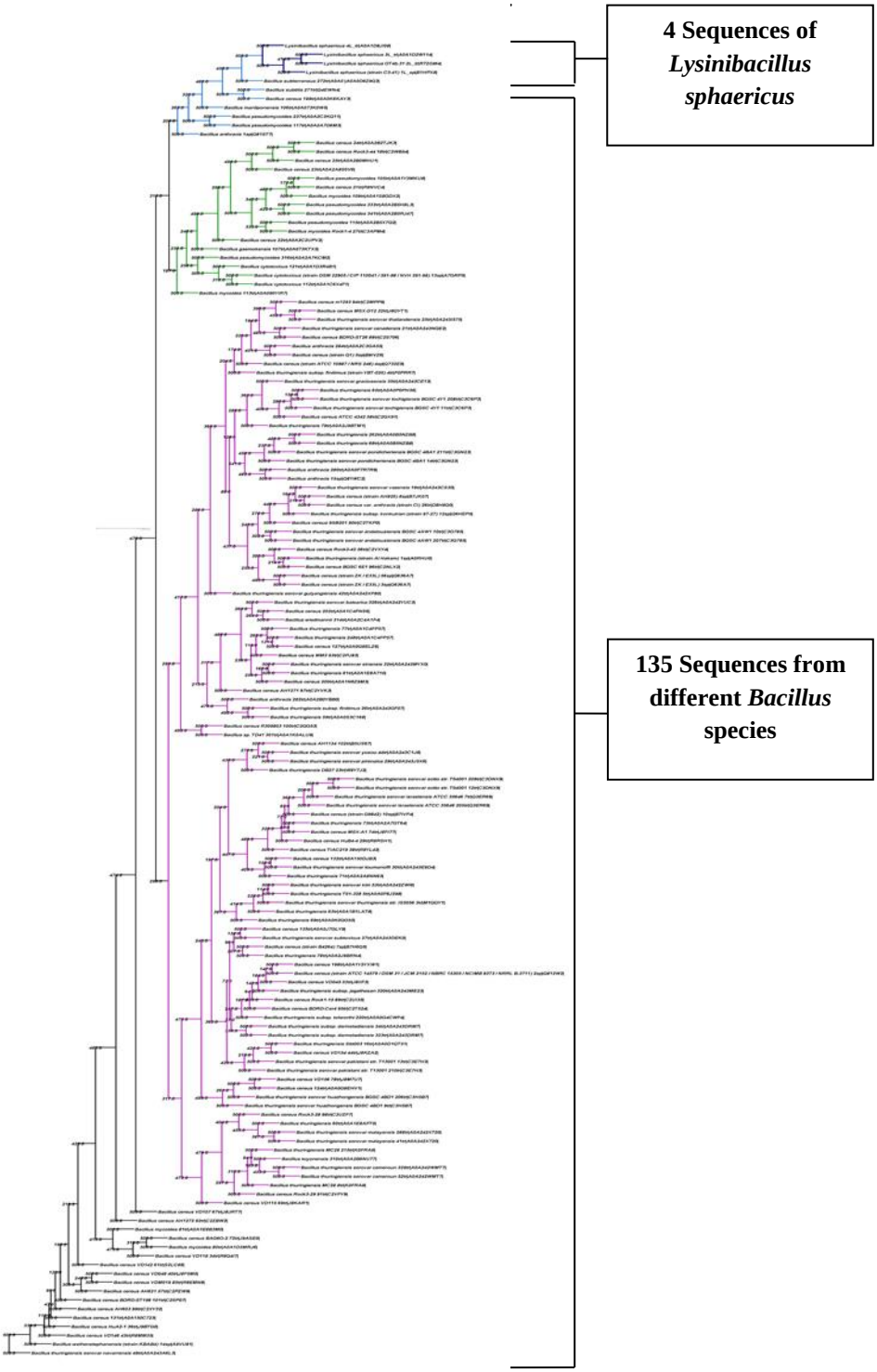


Fig. 2.B.2: Neighbor-joined phylogenetic tree based on bsh protein sequences showing the relationships among *Lysinibacillus sphaericus* and *Bacillus* spp.

2.B.3.3 Homology modeling of *L. sphaericus* bsh protein

Functional and structural similarity search of B1HPX8 (*L. sphaericus* bshC protein) showed highest confidence and identity of 100% and 33% with 4wbdA (crystal structure of bshc from *Bacillus subtilis*) respectively. The secondary structural configurations of query and target protein (B1HPX8 & 4wbdA) molecule were found to have almost equally distributed alpha-helix and beta-sheets (Fig. 2.B.3). Only three positions of pare wise sequence alignment (alignment position: 36, 68 and 79) were found with insertion or deletion phenomenon, which marked no effect on protein secondary structure. Alignment position 23-28 and 135-140 showed extra α-helix and 274-276 showed extra β sheet in predicted structure. In case of template molecule, alignment positions 326-332 have β-turn-α structure but predicted structure have only α-helix. Study of predicted structure revealed that the surface of *L. sphaericus* bsh protein (B1HPX8) was covered by alpha-helices whereas the beta-sheets were found in between alpha-helices. Some irregular structure was only observed at the C-terminal end of the predicted structure in comparison to template. Although having sequential diversity, structurally two proteins (B1HPX8 and 4wbdA) appeared near about same (Fig. 2.B.4).

Fig. 2.B.3: Secondary structure comparison between predicted structure of *Lysinibacillus sphaericus* bsh protein (B1HPX8) and homologous structure of *Bacillus subtilis* (4wbdA).

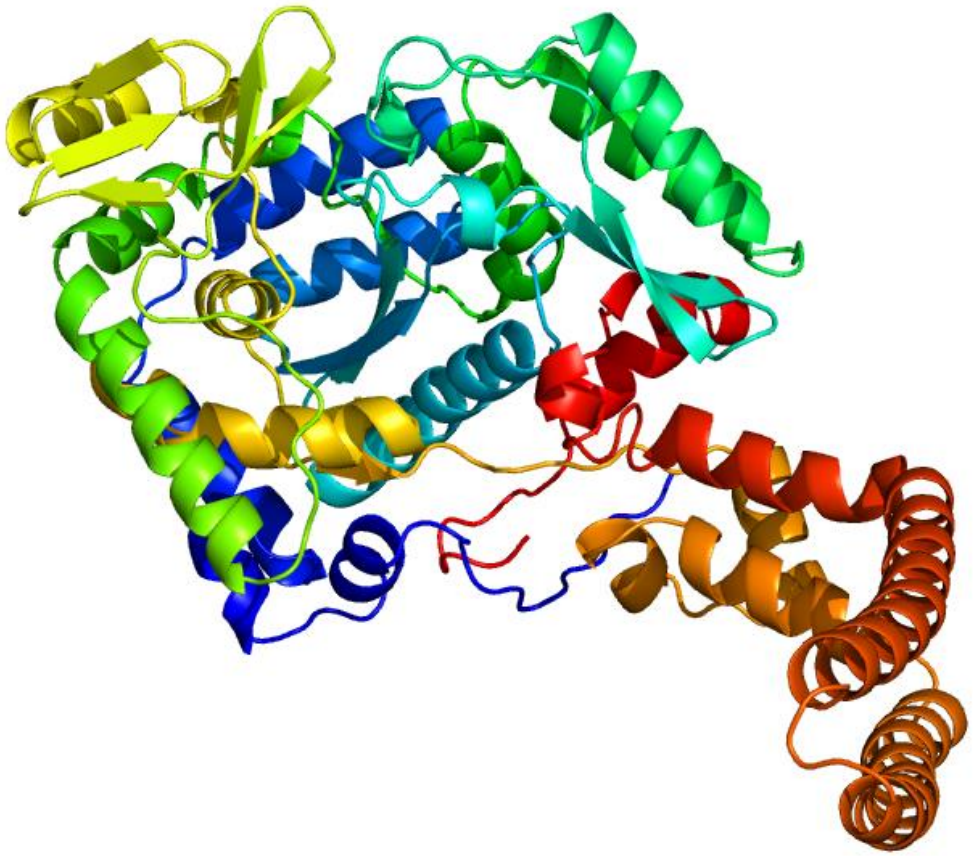
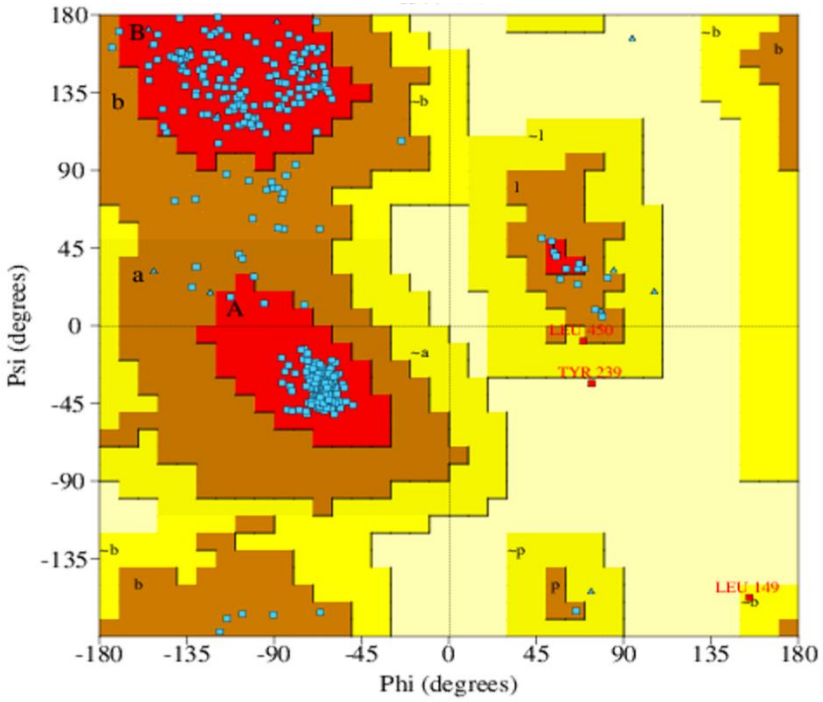


Fig. 2.B.4: Structural alignment between predicted tertiary structure of *Lysinibacillus sphaericus* bsh protein (B1HPX8) and homologous structure of *Bacillus subtilis* (4wbdA).

The Ramachandran plot (Fig. 2.B.5) depicts a good quality predicted 3D structure because of high prevalence of amino acids at the most favoured regions (92.4% at the most favoured region along with 7% at additional allowed region). This high quality predicted structure indicated that among two bsh proteins (B1HPX8 & 4wbdA) the amino acid substitution pattern was similarity substitution, which does not interfere into structural diversity. As the function of both proteins (bile salt hydrolase) was same, the tertiary structure remained undisturbed but their alteration in amino acid sequences was observed. It

might be due to the effect of interactions with their immediate environment. According to Li et al. (2008), more external interaction leads to more sequential diversity in proteins.



1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	463	92.4%
Additional allowed regions	[a,b,l,p]	35	7.0%
Generously allowed regions	[~a,~b,~l,~p]	2	0.4%
Disallowed regions	[XX]	1	0.2%*
Non-glycine and non-proline residues		501	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		20	
Proline residues		15	
Total number of residues		538	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

Fig. 2.B.5: Ramachandran plot of predicted structure of *Lysinibacillus sphaericus* bsh protein (B1HPX8).

2.B.3.4 Active site prediction through molecular docking

Initially, taurodeoxycholate was used for active site determination. According to Autodock result, among ten possible docking, five docking postures of taurodeoxycholate were found in only one docked place of protein (Fig. 2.B.6).

According to VanDuinen et al. (2015) GLY352, GLU353, ARG377 and ARG504 amino acids of *Bacillus subtilis* (4wbdA) bsh protein were recorded as the interactive amino acids forming direct hydrogen bond with glycerol and citrate, and the site was also found as the canonical Rossmann fold active site. Present study represented the docking positions 1 and 8, where taurodeoxycholate formed hydrogen bond with ARG376 having H-bond length of 2.133 and 2.064Å, and binding energy value of -7.23 and -6.34 respectively. Docking site 7 was bonded with GLY351, GLU352 and ARG503 with H-bond length of 2.184, 2.192 and 2.194 respectively. The binding energy -7.01 of docking site 7 promoted substrate binding. A common residue GLY351 was repeated in docking position 5 and 10. So, among the amino acids present in whole *L. sphaericus* bsh protein, only some residues like, GLY351, GLU352, ARG503 and ARG376 were detected as most potent and favorable amino acids to represent the catalytic site of the *L. sphaericus* bsh enzyme (Fig. 2.B.7).This specific site was also found as high frequency docking site (11 docking postures) for another substrate sodium taurocholate (Fig. 2.B.8). The best posture was identified through the strong binding affinity of the taurodeoxycholate with enzyme at that docking position using the minimum Atomic Contact Energy (ACE) value of -311.42.The lowest ACE value between substrate taurodeoxycholate and bsh protein was found at an inner groove, that location was also identified as high frequency docking site of taurodeoxycholate.

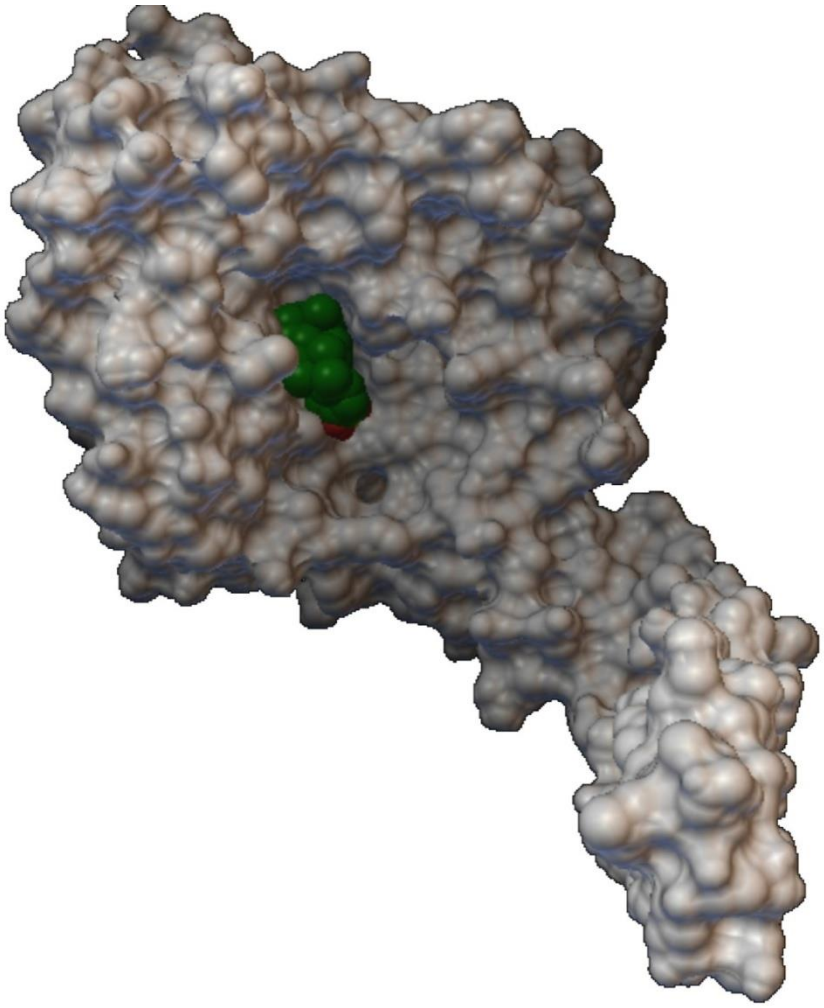


Fig. 2.B.6: Molecular docking between predicted *Lysinibacillus sphaericus* bsh protein (B1HPX8) and its substrate taurodeoxycholate.

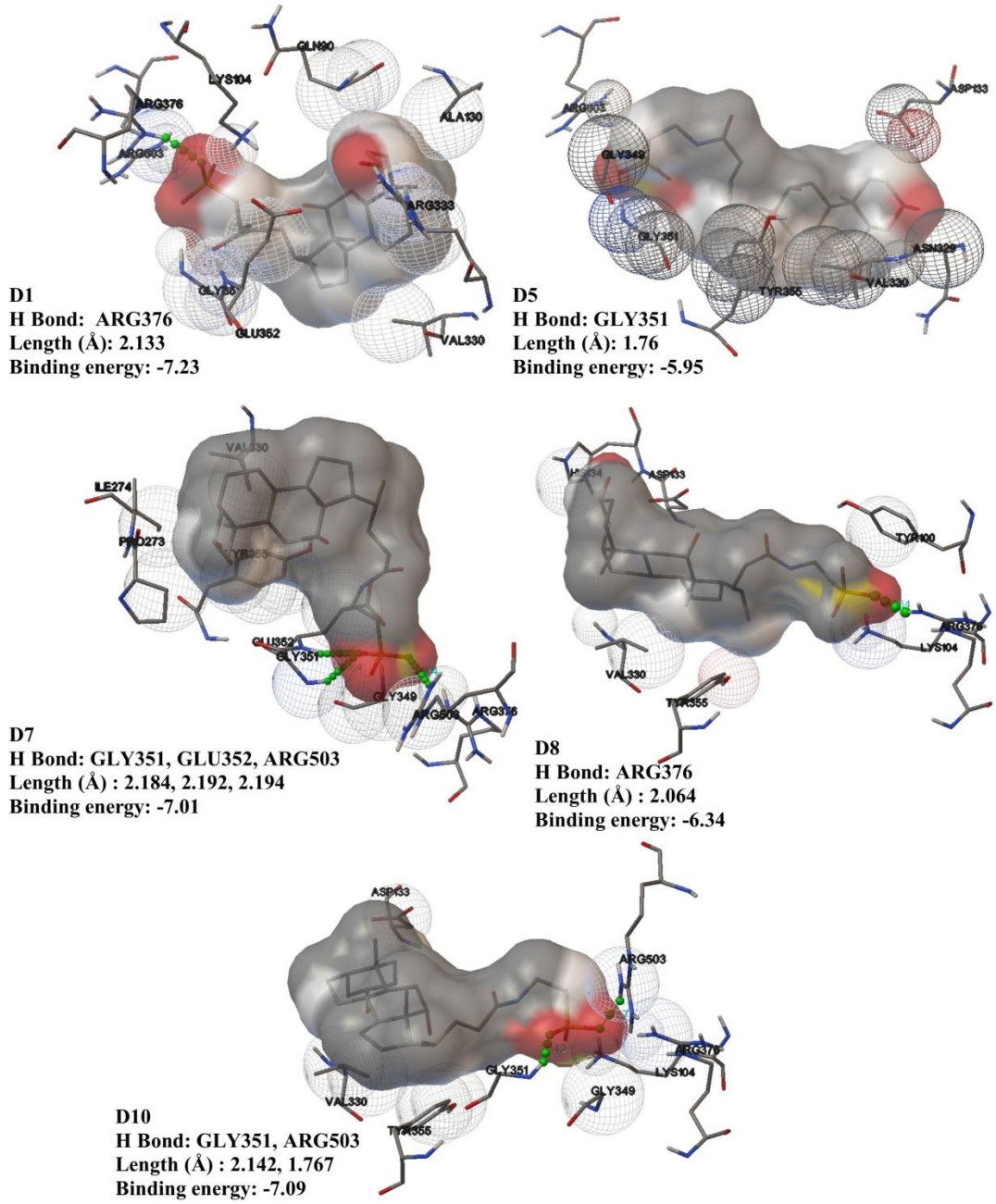


Fig. 2.B.7: Amino acid involved in hydrogen bond formation with taurodeoxycholate at the catalytic site with specific H-bond length and binding energy.

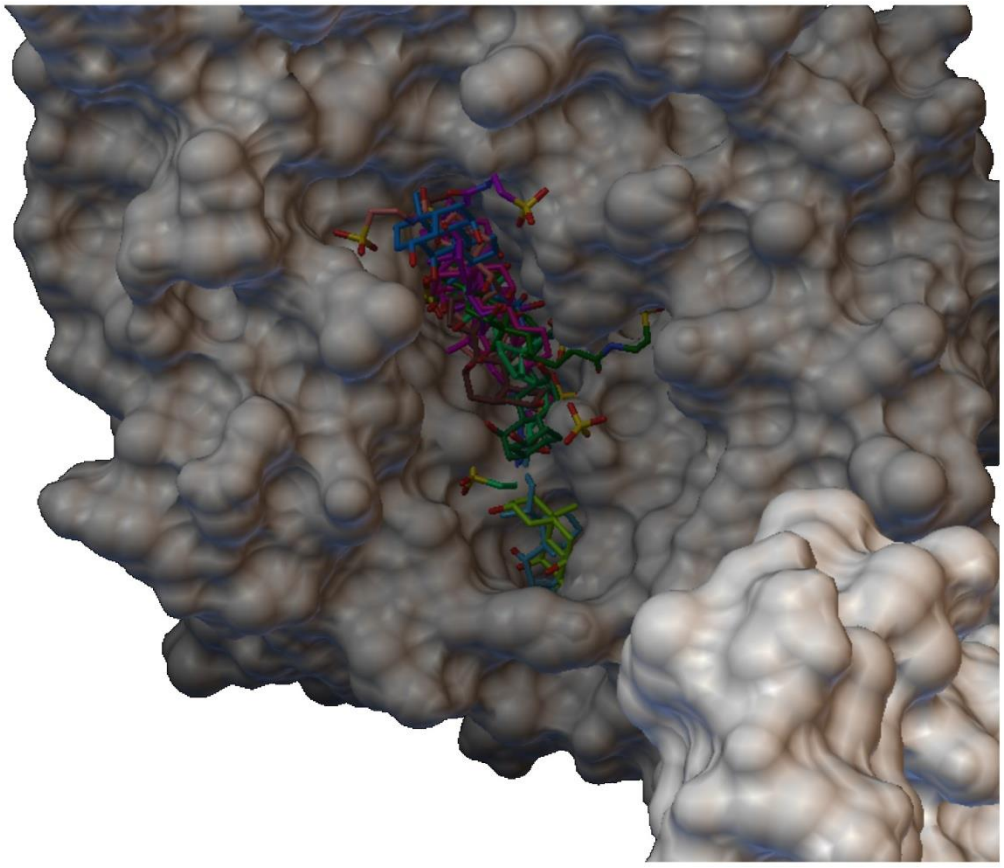


Fig. 2.B.8: Molecular docking between predicted *Lysinibacillus sphaericus* bsh protein (B1HPX8) with substrate sodium taurocholate.

2.B.3.5 Enzyme-inhibitor interaction study through molecular docking

Enzyme-inhibitor interaction study is crucial in determining the mode of action of inhibitor. Competition between substrate and inhibitor for the active site remarkably reduces the enzyme activity. According to Lin et al. (2014) carnosic acid, chrysophanol, epicatechin gallate, gossypetin, phenethyl caffeate and theaflavin were some of the well-known bsh protein inhibitors. In present study, those components were selected for active site inhibition

study of *Lysinibacillus sphaericus* bsh protein (B1HPX8). The lesser ACE value indicates higher binding efficiency of ligand with protein (Duhovny et al. 2002; Schneidman-Duhovny 2005). Theaflavin showed the lowest ACE value of -206.58 at the active site, whereas gossypetin and phenethyl caffeate showed ACE value of -191.48 and -129.19 respectively (Table 2.B.2). Chrysophanol, carnosic acid and epicatechin gallate also have showed some level of active site inhibition. The common source of the inhibitor gossypetin is an herbaceous tropical plant *Hibiscus sabdariffa* which is often used as preservative, colorant and antimicrobial agent during the formulation of feed for catfish (Fagbenro 2005).

Table 2.B.2: Atomic contact energy (ACE) value of the bsh inhibitors predicted by PatchDock.

Sl. No.	Component	ACE Value
1	Theaflavin	-206.58 (4)
2	Gossypetin	-191.48 (5)
3	Phenethyl caffeate	-129.19 (6)
4	Epicatechin gallate	-89.67 (8)
5	Chrysophanol	-83.18(5)
6	Carnosic acid	-82.82(6)

ACE Value (Number of docking in active site within 20)

The common source of the inhibitor gossypetin is an herbaceous tropical plant *Hibiscus sabdariffa* which is often used as preservative, colorant and antimicrobial agent during the formulation of feed for catfish (Adewole 2014). The common source of chrysophanol includes marine fungi, *Aspergillus* and *Rheum officinale* whereas Chrysophanol is often applied in zebrafish for its lipid lowering property (Chen 2015). The ingredients of Rosemary (*Rosmarinus officinalis* L.) include rosmarinic acid, carnosol, and carnosic acid. It is often used as a natural preservative of fish oil (Wang 2011). Epicatechin

gallate and its derivatives control the oxidative deterioration of fish due to their anti-oxidative activity (Zhong and Shahidi 2011). Though theaflavin, gossypetin, phenethyl caffeate, chrysophanol, and carnosic acid are used in fish feed or fish growth related products, they are more potent to deactivate the *L. sphaericus* bsh protein also, and are able to make a potent fish probiotic intolerable to gut environment.

2.B.4 Conclusion

L. sphaericus possess bsh protein to overcome the bile salt activity of the gastrointestinal tract. *In silico* analysis suggested that, *Lysinibacillus* sp. and *Bacillus* sp. accumulated bsh protein sequences diversity due to the effect of genus differentiation. However, this diversity has not been reflected in their structural level. This indicated that, the sequential diversity occurred due to high percentage of similar amino acid substitution, rather than identical substitution and has less diverged sequences. This study could enrich the literature of functional protein evolution and could be used for protein engineering study to make more potent fish probiotic for *Clarias batrachus* cultivation. The investigation also emphasizes that some inhibitors of this bsh protein are generally used in certain fish feed or fish growth related products. The present analysis indicated that theaflavin, gossypetin, phenethyl caffeate, chrysophanol and carnosic acid containing fish feed must be avoided during the application of probiotic *L. sphaericus* in aquatic pond to overcome ineffectiveness.