Vietnam Journal of Biotechnology **24**(2): 273-286, 2026. DOI: 10.15625/vjbt-23418

Research Article

CLONING AND *IN SILICO* CHARACTERIZATION OF 3-HEXULOSE-6-PHOSPHATE SYNTHASE AND 6-PHOSPHO-3-HEXULOISOMERASE FROM *Bacillus subtilis*

Quang Huy Nguyen[1], Thi Thanh Xuan Nguyen[2], Minh Thu Nguyen[1] and Luong Luan Chu[1,] ✉

[1]*National Key Laboratory of Enzyme and Protein Technology, Faculty of Biology, Hanoi University of Science, Vietnam National University, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam*

[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

✉To whom correspondence should be addressed. Email: luancl@vnu.edu.vn

Received: 6 September 2025; Accepted: 20 March 2026; Published online: 15 April 2026

ABSTRACT

3-Hexulose-6-phosphate synthase (HPS) and 6-phospho-3-hexuloisomerase (PHI) are key enzymes of the ribulose monophosphate (RuMP) pathway, which plays an essential role in formaldehyde assimilation and detoxification in methylotrophic microorganisms. Although homologous genes encoding these enzymes are annotated in the genome of *Bacillus subtilis*, their molecular and structural characteristics remain poorly explored in this organism. In this study, taxonomic identification using Kraken2, 16S rRNA phylogenetic analysis, and multilocus sequence typing confirmed the strain as *B. subtilis* ST123. Moreover, the *hxlA* and *hxlB* genes encoding HPS and PHI were cloned from a local *B. subtilis* ST123 strain isolated from infant fecal samples. The amplified *hxlA* (633 bp) and *hxlB* (558 bp) genes were successfully cloned, sequenced, and analyzed. Sequence alignment revealed high conservation with reference *B. subtilis* sequences. The three-dimensional structure of HPS was predicted using AlphaFold3 and rigorously evaluated by MolProbity, ERRAT, Verify3D, and ProSa. The HPS model exhibited high overall structural quality, with a MolProbity score of 1.13, 100% residues in the Ramachandran favored region, and an ERRAT quality factor of 99.505, although localized regions of lower sequence-structure compatibility were identified by Verify3D. The PHI structure was analyzed based on an available X-ray crystal structure (PDB ID: 1M3S). Physicochemical characterization indicated that HPS is thermostable and suitable for heterologous expression, whereas PHI showed a higher instability index, suggesting potential challenges during in vitro expression. This work provides the first comprehensive cloning and *in silico* characterization of HPS and PHI from a Vietnamese *B. subtilis* strain, offering valuable insights for metabolic engineering, synthetic biology, and C1 assimilation pathway design.

Keywords: 3-hexulose-6-phosphate synthase, 6-phospho-3-hexuloseisomerase, *hxlA, hxlB, Bacillus subtilis* ST123.

INTRODUCTION

Bacillus subtilis is a Gram-positive strain that can form a spore covering to adapt to extreme environmental conditions. In the digestive system of humans and ruminants, *B. subtilis* could release protease and amylase to digest food into simple and absorbable compounds. *B. subtilis* could also produce antibiotics to prevent other harmful microorganisms from growing in the intestinal system (Chen *et al*., 2008). As a result, *B. subtilis* is widely applied in producing probiotic products, biotechnology, agriculture, and livestock farming (Akinsemolu *et al*., 2024). 3-hexulose-6-phosphate synthase (HPS) and 6-phospho-3-hexuloseisomerase (PHI) are two important enzymes in the Ribulose monophosphate pathway (RuMP) (Orita *et al*., 2007). RuMP is an essential pathway to help some archaea and methylotrophic bacteria species to fix and detoxify formaldehyde (Yurimoto *et al*., 2005). HPS catalyzes a condensation reaction of ribulose-5-phosphate and formaldehyde to create hexulose-6-phosphate (Whitaker *et al*., 2016). HPS needs Mg^{2+} and Mn^{2+} ions to have a full bioactivity (Sahm *et al*., 1976). PHI catalyzes an isomerization reaction of hexulose-6-phosphate to fructose-6-phosphate (Orita *et al*., 2007). This is the intermediate compound to continue making organic compounds for energy storage, ATP yielding, and formaldehyde detoxification.

Although these enzymes have been extensively studied in methylotrophs, limited data exist regarding their structure and characteristics in *B. subtilis*. These include organisms such as *Methylomonas*, *Methylo- bacterium*, and *Paracoccus* spp., where both genes have been cloned and biochemically characterized (Rozova *et al.*, 2017). However, only limited attention has been given to these enzymes in *B. subtilis*, despite the presence of homologous genes in its genome. Furthermore, no detailed reports have explored their structural and functional characteristics using modern computational approaches. In Vietnam, research on *B. subtilis* has mainly focused on its probiotic potential and enzyme production capacity. To the best of our knowledge, there have been no domestic studies that combine gene cloning, sequencing, and *in silico* analysis of HPS and PHI from this organism (Trong *et al*., 2026). Therefore, this work aims to bridge that gap by providing the first comprehensive investigation into the molecular and structural properties of these enzymes in a local *B. subtilis* ST123 strain.

The present study aims to fill this gap by providing a comprehensive molecular and structural characterization of these two enzymes in a local *B. subtilis* ST123 strain. Cloning and analyzing these genes not only deepen our understanding of alternative formaldehyde assimilation pathways in *B. subtilis*, but also offers valuable insights for synthetic biology applications. These enzymes could be harnessed to engineer methylotrophic capabilities in non-native hosts, enable formaldehyde bioremediation, or develop synthetic pathways for C1 carbon assimilation. Moreover, understanding the structural and biochemical properties of HPS and PHI may pave the way for their future use in biosynthetic routes to produce bio-based chemicals and fuels, contributing to the development of sustainable biotech- nological processes.

MATERIALS AND METHODS

Materials

The All-in-One PCR Cloning Kit and associated vector were obtained from BioFact (South Korea). Genomic DNA and PCR products were purified using the Genomic DNA Prep Kit and the Gel and PCR Purification System (BioFact, South Korea), respectively. Plasmid DNA was extracted using the Exprep Plasmid SV Kit (GeneAll, South Korea). Taq DNA polymerase and deoxynucleotide triphosphates (dNTPs) were purchased from Thermo Fisher Scientific (USA). Restriction enzymes, including *Bam*HI, *Eco*RI, *Pst*I, and *Afl*II, were supplied by New England Biolabs (UK). Visualization of nucleic acids was carried out using RedSafe™ nucleic acid staining solution 20000 (iNtRON Biotechnology, South Korea). All oligonucleotide primers were synthesized by Phusa Biochem (Vietnam).

Isolation and identification of *Bacillus* strain from infant fecal samples

Infant fecal samples were preserved in 20% glycerol solution, diluted with sterile physiological saline, and spread-plated onto Luria-Bertani (LB) agar medium (Merck, Germany). The inoculated plates were incubated at 37°C under 5% CO_2 for 24 hours. Discrete colonies identified as *Bacillus* were selected for genomic DNA extraction. High-quality genomic DNA was then subjected to library preparation following the Illumina sequencing protocol. The purified genomic DNA served as a template for PCR amplification of the 16S rRNA gene using a specific primer pair. Taxonomic identification at the genus and species level was performed using Kraken2 against the Bacteria database (Liu et al., 2024). A phylogenetic tree was constructed based on 16S rRNA gene sequences, including the *Bacillus* isolate under investigation (BS1) and 39 reference sequences retrieved from the NCBI database.

The bacterial strains and DNA genomic extractions

E. coli DH5α was used for transformation and DNA plasmid extraction. *B. subtilis* was used for DNA genomic extraction. *B. subtilis* was incubated in 5 mL of LB broth medium at 37ºC (overnight) with shaking at 160 rpm. *B. subtilis* DNA was extracted using the Genomic DNA Prep Kit. Extracted genomic DNA was run on an agarose gel (1% w/v) and visualized using a gel documentation system (UVP, England) after staining with RedSafe solution.

Primer design and polymerase chain reaction

Primers were designed and listed in Table 1. The polymerase chain reaction (PCR) was performed in a total volume of 25 µL with components including 1 µL Taq Dream polymerase 1 U/µL, 2.5 µL Taq polymerase buffer (10x), 1 µL dNTPs (10 mM), 2 µL genomic DNA (50-100 ng), 1 µL forward primer (10 µM), 1 µL reverse primer (10 µM), and nuclease-free water up to 25 µL. The amplification was carried out in SimpliAmp™ Thermal Cycler (Applied Biosystem, USA) using the following reaction conditions: initial denaturation of 95ºC for 5 minutes; followed by 30 cycles of 95ºC for 30 secs, 58ºC for 30 secs, and 72ºC for 1 minute; with a final extension of 72ºC for 5 minutes. PCR products were run on an agarose gel (1% w/v) and visualized using a gel documentation system after staining with RedSafe solution. The reaction without the

template served as a non-template control (NTC).

PCR products purification and ligation

PCR products were purified using BioFact™ Gel and PCR Purification System kit. After purification, PCR products were evaluated by ImageJ software (USA) to calculate the ratio for ligase reaction. Two reactions were carried out following the instructions of All in One™ PCR Cloning Kit. The *hxlA* and *hxlB* were ligated in All-in-One cloning vector with components including 1 µL All-in-One vector, 1 µL 6X buffer, 1 µL PCR products, and nuclease-free water up to 6 µL. The ligation reaction was incubated at room temperature (25°C) for an hour.

Table 1. Sequence of primers amplifying *hxlA* and *hxlB*.

Primer	Sequence (5'-3')	Product size (bp)	Tm (°C)
hxlA forward	CAGTCGGATCCATGGAATTACAGCTTGCATTAGAC	633	56
hxlA reverse	CTGACTGAATTCTTATCCTTGGACAATCAGCTGCT		57
hxlB forward	CAGTCCTGCAGATGAAAACGACTGAATACGTAGCG	558	58
hxlB reverse	CTCGAACTTAAGCTATTCAAGGTTTGCGTGGTGA		57

Biotransformation

E. coli DH5α cells were cultivated in LB medium at 37°C with shaking at 160 rpm until the culture reached an optical density at 600 nm (OD_{600}) of 0.4. The culture was then chilled on ice and centrifuged at 3000 rpm for 10 minutes to collect the cell pellet. The pellet was gently washed twice with ice-cold 0.1 M $CaCl_2$ to prepare chemically competent cells. For transformation, the ligation mixture was added to 100 µL of *E. coli* DH5α competent cells and incubated on ice for 15 minutes. The cells were then heat-shocked at 42°C for 1 minute and immediately returned to ice for 5 minutes. Subsequently, 300 µL of LB broth was added, and the cells were allowed to recover at 37°C with shaking at 160 rpm for at least 1 hour. After recovery, 200 µL of the transformation mixture was spread onto LB agar plates supplemented with 100 µg/mL ampicillin and 34 µg/mL kanamycin, followed by overnight incubation at 37°C (Nguyen *et al*., 2020). Individual colonies were randomly selected and used as templates for colony PCR with universal M13 forward (5′-GTAAAACGACGGC CAG-3′) and M13 reverse (5′-CAGGAAACAGCTATGAC-3′) primers.

Plasmid extraction and digestion with restriction enzymes

Colonies that contain the All in One-*hxlA* vector and the All in One-*hxlB* vector were grown separately at 37°C in 5 mL LB broth medium (ampicillin and kanamycin were added) with shaking at 160 rpm, from 16 to 21 hours. Then, 5 mL culture solution was centrifuged to attain the cell pellet, and DNA plasmid was extracted using the Exprep Plasmid SV kit. All in One-*hxlA* was incubated with *BamH*I and *EcoR*I restriction enzymes (NEB, England), and the All in One-*hxlB* vector was incubated with *Pst*I and *Afl*II (NEB, England). Cut reaction was

carried out with components including 1 µL 10x buffer, 1 µL DNA plasmid, 0.5 µL restriction enzymes, and nuclease-free water up to 10 µL. The reaction was incubated at 37ºC for 20 minutes.

Computational modelling and functional assessment

Plasmid DNA sequencing was carried out by DNA Sequencing company (Vietnam). Sequencing results were analyzed by Bioedit. SnapGene was used to present inferred amino acid sequences. We used AlphaFold3 and PrankWeb to predict the spatial structure and active site of HPS and PHI (Nam, 2025; Polak *et al.*, 2025). The predicted structures were saved in PDB format. MolProbity was used to validate the HPS and PHI models. YASARA Energy Minimization Server was used for energy minimization of 2 models, then evaluated structural quality and authenticity by ERRAT, Verify3D, and ProSA (Rehman *et al*., 2021). Characteristics of 2 enzymes were detected by using SMART and InterPro (Iqrar *et al*., 2020).

RESULTS AND DISCUSSION

Taxonomic identification and phylogenetic tree construction

To identify a microbial host with potential metabolic capabilities beneficial for bioproduction, we isolated and characterized the strain BS1 from infant fecal samples. Taxonomic identification using Kraken2 with the Bacteria database revealed that BS1 predominantly belongs to the genus *Bacillus* (97.44%). The isolate BS1 was taxonomically assigned to *Bacillus vallismortis* based on Kraken2 classification and multilocus sequence typing (MLST) analysis. Phylogenetic inference indicated that BS1 clusters within the *B. subtilis* species complex, showing high sequence similarity to reference strains of *B. vallismortis* BL-01 (Figure 1). In detail, phylogenetic analysis based on the 16S rRNA sequence was performed using a neighbor-joining tree constructed from 40 reference sequences, including BS1 and 39 representative *Bacillus* strains retrieved from NCBI. The resulting tree confirmed the close evolutionary relationship between BS1 and other known *B. subtilis* strains, supporting its taxonomic classification and suggesting functional genomic similarity (Figure 1). As shown in Figure 1, strain BS1 clustered tightly within the *B. subtilis* clade, displaying close evolutionary relationships with well-characterized *B. subtilis* strains. Notably, BS1 was positioned in a distinct and well-supported clade clearly separated from other *Bacillus* species such as *B. amyloliquefaciens*, *B. velezensis*, and *B. licheniformis*, indicating reliable and precise classification at the species level. Further, multilocus sequence typing (MLST) analysis based on assembled genomic data assigned BS1 to *B. subtilis* sequence type ST123.

Importantly, genome mining revealed that BS1 harbors genes *hxlA* (encoding HPS) and *hxlB* (encoding PHI), which are integral to the RuMP pathway. This pathway is implicated in formaldehyde fixation and detoxification but may also facilitate alternative carbon flux through central metabolism under engineered conditions. The presence of *hxlA* and *hxlB* suggests BS1 could serve as a genetic reservoir for enhancing C1 assimilation and introducing auxiliary metabolic routes in synthetic biology applications.

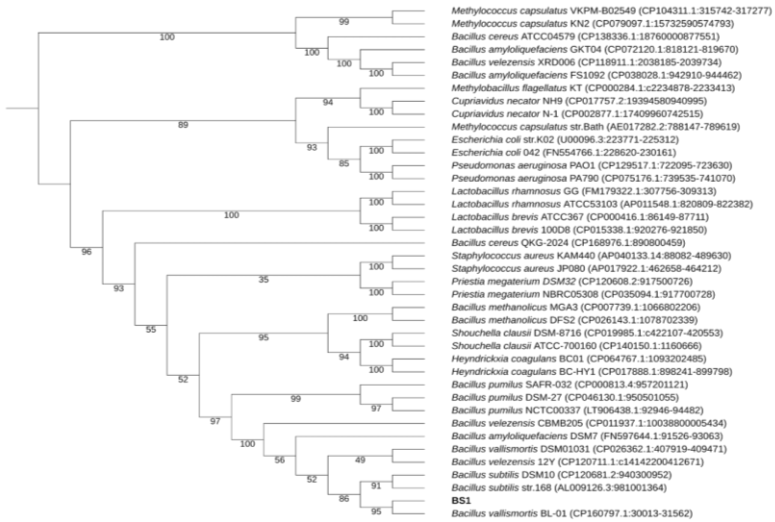

Figure 1. Phylogenetic analysis based on 16S rRNA sequences.

Agarose gel electrophoresis of PCR and cloning products

In the study, high-quality total DNA would have a high molecular weight and a clear band (Figure 2A). The isolated DNA was ensured to have high purity. PCR products of *hxlA* and *hxlB* are 633 bp and 558 bp, respectively (Figure 2B), with no interference bands. This ensures the specificity of primers. In conclusion, the study designs primers and amplifies *hxlA* and *hxlB* from *B. subtilis* genomic DNA successfully. Although direct sequencing of PCR products is a common approach, in this study, the amplified *hxlA* and *hxlB* genes were cloned into the All-in-One vector before sequencing. This strategy was employed to minimize the risk of PCR-derived mutations and to facilitate insert verification via restriction digestion. Additionally, the cloned constructs serve as a stable resource for downstream applications and further functional studies.

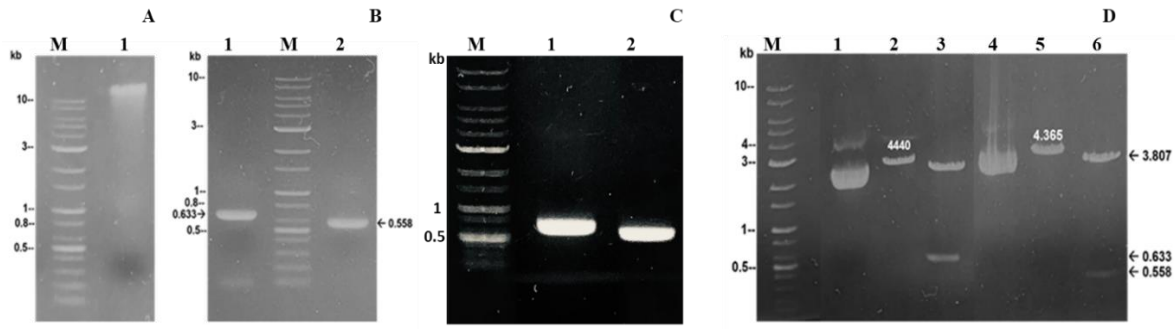

Figure 2. Gel electrophoresis. (A) Extracted genomic DNA, 1: *B. subtilis* genomic DNA, M: DNA marker; (B) PCR products from isolated DNA plasmid with specific primers, 1: PCR product of *hxlA*, 2: PCR product of *hxlB*; (C) Screening colonies containing recombinant vector by PCR with specific primers of *hxlA* and *hxlB*; (D) All in One-*hxlA* and All in One-*hxlB* digested with restriction enzymes, 1 and 4: Uncut products, 2: All in One-*hxlA* digested with *BamH*I, 3: All in One-*hxlA* digested with *BamH*I and *EcoR*I, 5: All in One-*hxlB* digested with *Pst*I, 6: All in One-*hxlB* digested with *Pst*I and *Afl*II.

Because the All-in-One vector carries both ampicillin and kanamycin resistance genes, only transformed cells harboring the recombinant plasmids were able to form well-isolated, round colonies on selective medium. Positive colonies were screened by colony PCR using gene-specific primers, followed by agarose gel electrophoresis. Based on the All in One™ PCR Cloning Kit design, the expected PCR product sizes were 633 bp for *hxlA* and 558 bp for *hxlB* (Figure 2C). Recombinant plasmids were further verified by restriction digestion. For All in One-*hxlA*, digestion with *BamH*I produced a single band of 4440 bp, whereas double digestion with *BamH*I and *EcoR*I generated two fragments of 3807 bp and 633 bp, confirming the insertion of *hxlA* (Figure 2D). Similarly, insertion of *hxlB* was confirmed by double digestion with PstI and AflII, which released the expected insert fragment. These results demonstrated that the ligation was successful and that both *hxlA* and *hxlB* were correctly cloned into the All-in-One vector.

Gene sequencing and analysis

The recombinant plasmids All-in-One-*hxlA* and All-in-One-*hxlB* were sequenced, and the resulting gene sequences were aligned with reference sequences from GenBank using BLAST (https://blast.ncbi.nlm.nih.gov/blast.cgi). The *hxlA* gene sequence showed complete identity with the corresponding *B. subtilis hxlA* sequence in GenBank. Accordingly, the deduced 206-amino acid sequence was 100% identical to the HPS (EC 4.1.2.43). Similarly, the *hxlB* sequence exhibited no nucleotide differences compared to the *B. subtilis* ST123 reference. The encoded 185-amino acid sequence also matched exactly that of PHI (EC 5.3.1.27), confirming the conservation of both the gene and protein sequences.

Spatial structure of HPS and PHI prediction

The predicted HPS model is composed of α helices coiled together. The HPS model is highly accurate, with most regions having pLDDT > 90 (blue). Some regions of α helices have a pLDDT score between 70 and 90 (light blue), which means good backbone prediction but needs to be validated. A small number of regions have a pLDDT score between 50 and 70 (yellow), showing the flexibility and low-confidence prediction (Figure 3A). The PHI model was identified through X-ray diffraction (PDB ID: 1M3S) with a 1.95-Å resolution score (Sanishvili *et al.,* 2004) (Figure 3B).

MolProbity was used to validate the structural configuration of HPS (Table 2), particularly by evaluating all-atom contacts (Clashscore < 10) and overall protein geometry (MolProbity score < 2, Ramachandran plot, Ramachandran favored, and Ramachandran outlier). The clashscore of the predicted HPS is 3.39, indicating only small amounts of steric clashes are present, which is considered geometrically reasonable for a predicted model.

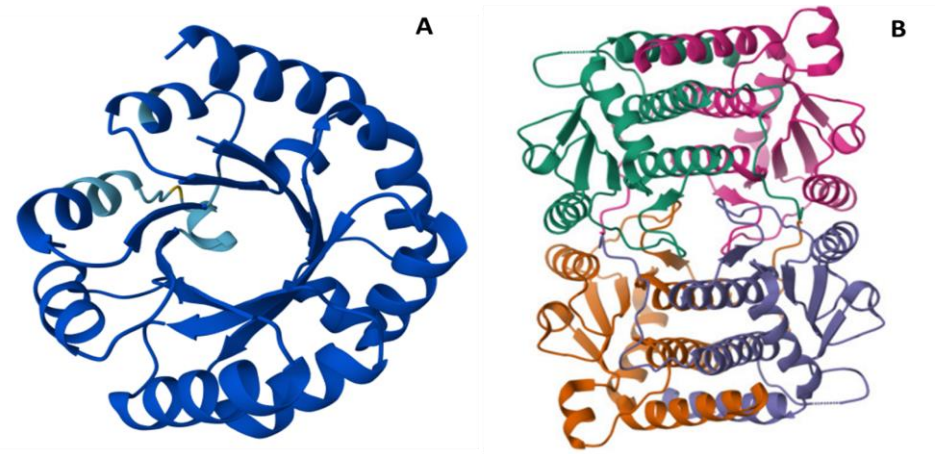

Figure 3. The structure of HPS and PHI. (A) The predicted structure of HPS by AlphaFold3. The pLDDT > 90 (blue), 90 > pLDDT > 70 (light blue), 70 > pLDDT > 50 (yellow). (B) The PHI model was identified through X-ray diffraction (PDB ID: 1M3S).

Table 2. MolProbity summary statistics.

	Criteria	Threshold	Result
All-atom contacts	Clash score	< 10	3.39
Protein geometry	Ramachandran outlier (%)	< 0.05%	0
	Ramachandran favored (%)	> 98%	100
	Rama distribution Z-score	< 2.0	0.52 - 0.54
	MolProbity score	< 2.0	1.13
Low-resolution criteria	CaBLAM outlier (%)	< 1.0%	0
	CA Geometry outliers (%)	< 0.5%	0

The Ramachandran plot analyzes backbone dihedral angle φ and ψ of each residue to determine whether residues (dots) fall in favored/allowed (blue borders) and disallowed regions (outside border). The Ramachandran plot of the HPS model (Figure 4A) has most residues lying in favored regions, and Ramachandran favored is 100%, showing the stable angle for stabilizing structure. The Ramachandran outlier is 0%, indicating that there is no residue in the disallowed region. MolProbity score calculated the validation of the whole predicted model. The MolProbity score of 1.13 confirms the high accuracy and structural validity of the predicted HPS

model. YASARA was used to perform energy minimization to stabilize the structure of the predicted HPS model by eliminating steric clashes. The total energy of the predicted model decreased from -99515.9 kJ/mol to -121038.9 kJ/mol, indicating that the refined model is closer to the native conformation. The refined model was evaluated by the ERRAT, Verify3D, and ProSa. The ERRAT analyzed 3D structure based on the error rate in non-bonded atom-atom interactions. The model achieved an overall total quality factor of 99.505, indicating that 99.505% of its regions fall within high-accuracy and high-confidence predictions (Figure 4B).

Verify3D calculated the compatibility between the 3D structure and amino acid sequence (1D) of HPS to verify the suitability of each amino acid residue due to the suitable characteristics. According to the Verify3D standard, the number of amino acids that have an average 3D-1D score greater than 0.1 must be 80%. 55.71% of amino acids have averaged scores greater than 0.1. This indicates that the model is not fully compatible in terms of structure-sequence alignment, particularly in the regions that have pLDDT between 50 and 70, and pLDDT between 70 and 90. The Z-score of predicted HPS (black dot) is -8.34, which lies in the typical range for 200-amino-acid-length protein, and falls within the score distribution of NMR structure (Figure 4C). The local model quality plot shows most picks of 2 lines are under 0, indicating the stable structure (Figure 4D).

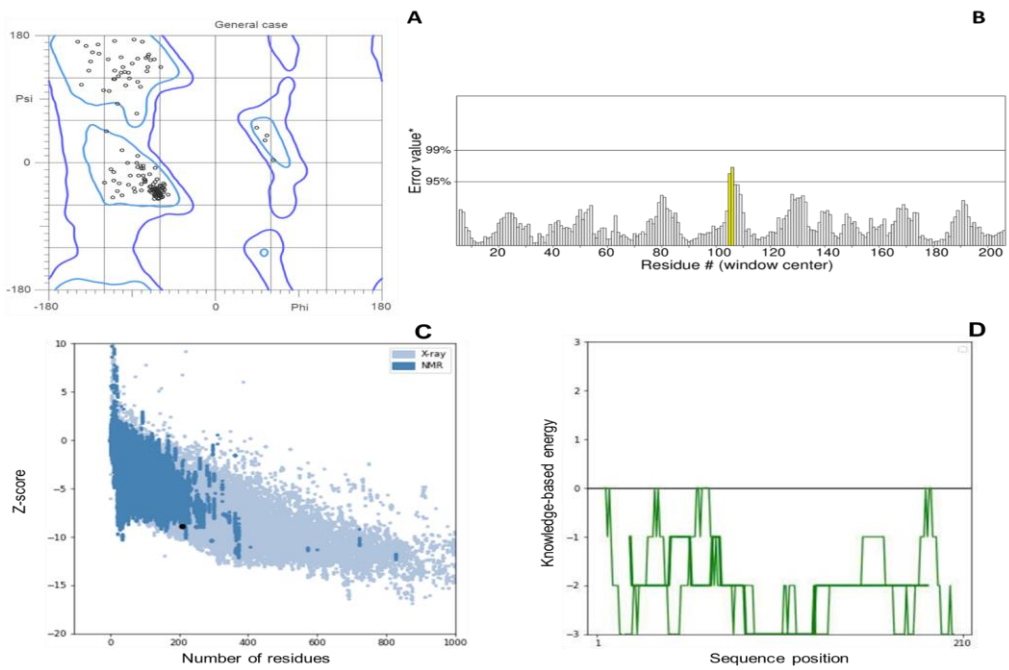

Figure 4. Ramachandran plot, the ERRAT, Z-score, and ProSA analyses**.** (A) Ramachandran plot of the predicted HPS model. (B) ERRAT plot of predicted HPS evaluated the error rate in non-bonded atom-atom interaction. (C) The z-score of predicted HPS shows the overall energy score of the model in comparison with high-resolution PDB structures. (D) Local model quality plot of HPS by ProSA evaluated based on knowledge-based energy at each amino acid position.

The structural quality of the predicted HPS model was evaluated using several computational tools. The model showed strong structural reliability supported by a high MolProbity score (1.13), 100% of residues in the Ramachandran favored region, and an ERRAT quality factor of 99.505. However, the Verify3D result indicated that only 55.71% of the residues had a 3D-1D score greater than 0.1, which is below the commonly accepted threshold of 80%. This apparent inconsistency may stem from the nature of Verify3D's scoring, which can be influenced by local sequence–structure compatibility and may penalize regions that deviate from well-defined structural motifs, even if global structural features are sound. Therefore, while the MolProbity and ERRAT results strongly support the overall quality of the model, regions with poor Verify3D scores may indicate flexible or unresolved loops that warrant further investigation, such as molecular dynamics simulations or experimental validation. Integrating these diverse evaluations provides a more nuanced understanding of the model's strengths and limitations.

Characterization of HPS and PHI

The physicochemical properties of HPS and PHI are predicted by ProtParam (Table 3). The HPS consists of 206 residues with a 22069.58-Da molecular weight. This protein is rich in hydrophobic residues such as Ile (11.2%), Ala (11.2%), Val (9.2%), and Leu (8.7%), and low content of His (0.5%), Cys (0.5%), and Trp (0%). The total number of negatively charged residues and positively charged residues is 35 and 24, which indicate a negative net charge, consistent with pI 4.7. The estimated half-life is determined by the identity of the N-terminal residue. The N-terminal residue of the HPS model is methionine, generally conferring higher stability. The predicted half-life of HPS in mammalian reticulocytes is 30 hours, longer than 20 hours in the yeast and 10 hours in the *E. coli*, suggesting it is favorable for heterologous expression.

The instability index (II) is 28.29, and the aliphatic index is 115.53, indicating stabilization of protein in high thermo-stability. The 0.031 grand average of hydropathicity (GRAVY) indicates mild hydrophobic while still being soluble. The PHI is composed of 185 amino acids and has a molecular weight of 20115.12 Da. The most abundant residues are Leu (11.4%), Ser (9.7%), Ala (8.6%), Gly (8.1%) and Ile (8.1%). Meanwhile, no Cys and Trp are present. The absence of Trp in both HPS and PHI may affect UV absorbance and folding stability. PHI has a total number of negatively charged residues and positively charged residues of 20 and 16, indicating a negative net charge. The predicted pI due to the total number of charged residues is 5.95. PHI has estimated half-life the same as HPS, which also considers heterologous expression. The physicochemical properties of the PHI enzyme, as predicted by the ProtParam tool, reveal an instability index (II) of 49.19, which exceeds the threshold of 40 and thus classifies the enzyme as potentially unstable in vitro. This elevated II value suggests that the PHI protein may be prone to degradation or misfolding during heterologous expression and purification, posing challenges for downstream applications. This limitation highlights the necessity for further experimental optimization, such as codon optimization, fusion with solubility-enhancing tags, or co-expression with molecular chaperones, to improve protein stability and yield.

Moreover, future studies may consider site-directed mutagenesis targeting regions contributing to instability, as guided by *in silico* flexibility predictions, to enhance the structural robustness of PHI for functional or structural analysis.

Table 3. The physicochemical properties of HPS and PHI.

Protein		**HPS**	**PHI**
Length (Amino acid)		206	185
Molecular Weight (Da)		22069.58	20115.12
Isoelectric point (pI)		4.7	5.95
Instability index (II)		28.29	49.19
Aliphatic index		115.53	92.38
Grand average of hydropathy (GRAVY)		0.031	-0.079
The N-terminal		Met (M)	Met (M)
Binding probability (%)		78	62.5
Number of residues		37	37
Active site coordination	X	2.1748	4.1422
	Y	-3.7591	24.56
	Z	6.224	4.18

Active sites of HPS and PHI were predicted using PrankWeb (Table 3). The HPS active site has a binding probability of 78%. Its main core consists of 37 residues, located within the spatial region defined by the coordinates x = 2.1748, y = -3.7591 and z = 6.224 (Figure 5A). TYR137, ALA6, ALA165 and ILE185 are 4 amino acids with the strongest binding potential, with binding probabilities of 92.68%, 92.04%, 89.77% and 89.77%. The active site of PHI also contains 37 residues, but with a lower binding probability of 62.5%. Its main core is located in the coordinated x = 4.1422, y = 24.56, and z = 4.18 (Figure 5B). The binding probabilities of the amino acids in the PHI active site are relatively low, with only PHE151 (65.98%), SER88 (60.06%), SER47 (56.85%), SER86 (56.56%), and ILE113 (52.37%) exceeding 50%.

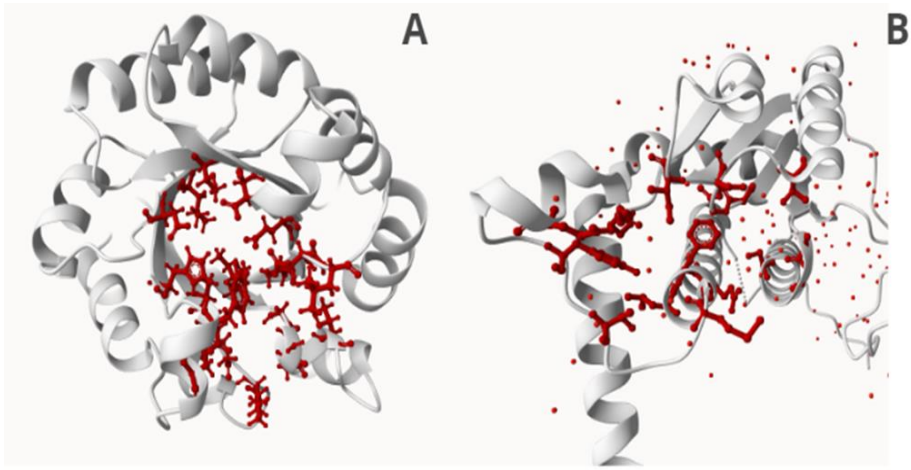

Figure 5. The predicted active sites of HPS (A) and PHI (B).

According to InterPro and SMART, the domain and biological function analysis of HPS and PHI have been found (Table 4). HPS belongs to the 3-keto-L-gulonate-6-phosphate decarboxylase (OMPdecase) of the HUMPS family, spanning from residue position 2 to 200. The enzyme of OMP decarboxylase plays a key catalytic role in the *de novo* pyrimidine nucleobase biosynthetic process (GO:0006207) and catalyzes the decarboxylation of orotidine 5'-monophosphate (OMP) into uridine monophosphate (UMP) (GO:0004590). PHI has a sugar isomerase domain (SIS domain), a module that acts as an isomerase and binds to phosphorylated sugar (GO:1901135 and GO:0097367).

Table 4. The domain and biological functions of HPS and PHI.

Protein	Family	Domain	Position of domain	Biological function	
				GO term	Detail
HPS	HUMPS	OMPdecase	2 - 200	GO:0006207	*De novo* pyrimidine nucleobase biosynthetic process
				GO:0004590	Orotidine-5'-phosphate decarboxylase activity
PHI	HisA/HisF	SIS	32-157	GO:1901135	Carbohydrate derivative metabolic process
				GO:0097367	Carbohydrate derivative binding

The identification and structural characterization of HPS and PHI suggest that *B. subtilis* ST123 possesses a latent RuMP pathway with potential for

formaldehyde assimilation. The high conservation of sequence and the strong structural reliability of HPS support its functional robustness, whereas the higher instability of PHI indicates a potential bottleneck requiring protein engineering optimization.

Importantly, the HPS-PHI module represents a promising platform for synthetic methylotrophy, enabling the incorporation of C1 substrates into central metabolism via fructose-6-phosphate. This module can be further integrated into C1 carbon assimilation pathways, linking upstream methanol/formate oxidation with downstream biosynthetic routes for value-added products. In addition, these enzymes offer potential for formaldehyde detoxification and bioremediation, as the RuMP pathway provides an efficient route for converting toxic formaldehyde into metabolic intermediates. Overall, this study establishes a structural and functional basis for future pathway reconstruction and metabolic engineering applications in sustainable biotechnology.

CONCLUSION

In this study, the *hxlA* and *hxlB* genes encoding HPS and PHI were successfully cloned and characterized from a local *B. subtilis* ST123 strain. Taxonomic and phylogenetic analyses confirmed the precise classification of the isolate and supported its genetic reliability. *In silico* structural modeling revealed that HPS possesses a highly reliable and stable three-dimensional structure, as validated by multiple quality assessment tools, whereas PHI exhibited a higher predicted instability, indicating potential limitations for heterologous expression without further optimization.

The identification and detailed computational analysis of these two RuMP pathway enzymes provide new insights into formaldehyde assimilation potential in *B. subtilis*, a non-classical methylotrophic host. Beyond fundamental characterization, the results highlight the feasibility of exploiting HPS and PHI as functional modules for metabolic engineering, including synthetic methylotrophy, C1 carbon utilization, and formaldehyde detoxification. This study lays a molecular and structural foundation for future experimental validation and rational pathway engineering aimed at developing sustainable biotechnological applications.

ACKNOWLEDGMENT

This research was funded by the research project QG.24.73 of Vietnam National University, Hanoi.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Akinsemolu A.A., Onyeaka H., Odion S., and Adebanjo I. (2024). Exploring *Bacillus subtilis*: ecology, biotechnological applications, and future prospects. *Journal of Basic Microbiology 64*(6). https://doi.org/10.1002/jobm.2023 00614

Chen H., Wang L., Su C., Gong G., Wang P., Yu Z. (2008) Isolation and characterization of lipopeptide antibiotics produced by *Bacillus subtilis*. *Letters in Applied Microbiology 47*(3, 180-186. https://doi.org/10.1111/j.1472-765x.2008.02412.x

Iqrar U., Javaid H., Ashraf N., Ahmad A., Latief N., Shahid A.A., et al. (2020). Structural and functional analysis of pullulanase type 1 (PulA) from *Geobacillus thermopakistaniensis*.

Molecular Biotechnology 62(8), 370-379. https://doi.org/10.1007/s12033-020-00255-x

Liu Y., Ghaffari M.H., Ma T., Tu Y. (2024) Impact of database choice and confidence score on the performance of taxonomic classification using Kraken2. *aBIOTECH 5*(4), 465-475. https://doi.org/10.1007/s42994-024-00178-0

Nam K.H. (2025). Evaluation of AlphaFold3 prediction for post-translational modification, oligomeric assembly, and quenchable metal binding of fluorescent proteins. *Journal of Molecular Graphics and Modelling* 142, 109169. https://doi.org/10.1016/j.jmgm.2025.109169

Nguyen T.C., Nguyen T.H.T., Nguyen T.T., Le T.H., Nguyen S.L.T., Nguyen T.A.T, et al. (2020). Cloning and expression of maltooligosyltrehalose trehalohydrolase from *Sulfolobus solfataricus* DSM 1616 in *Bacillus subtilis* WB800. *Vietnam Journal of Biotechnology*, *18*(2), 363-372. https://doi.org/10.15625/1811-4989/18/2/14886

Orita I., Sakamoto N., Kato N., Yurimoto H., Sakai Y. (2007b) Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Applied Microbiology and Biotechnology 76*(2), 439-445. https://doi.org/10.15625/1811-4989/18/2/14886

Polak L., Skoda P., Riedlova K., Krivak R., Novotny M., Hoksza D. (2025). PrankWeb 4: a modular web server for protein-ligand binding site prediction and downstream analysis. *Nucleic Acids Research 53*(W1), W466-W471. https://doi.org/10.1093/nar/gkaf421

Rehman A., Wang X., Ahmad S., Shahid F., Aslam S., Ashfaq U.A., et al. (2021). *In silico* core proteomics and molecular docking approaches for the identification of novel inhibitors against *Streptococcus pyogenes*. *International Journal of Environmental Research and Public Health 18*(21), 11355. https://doi.org/10.3390/ijerph182111355

Rozova O.N., But S.Y., Khmelenina V.N., Reshetnikov A.S., Mustakhimov II, Trotsenko Y.A. (2017). Characterization of two recombinant 3-hexulose-6-phosphate synthases from the halotolerant obligate methanotroph *Methylomic- robium alcaliphilum* 20Z. *Biochemistry* (Moscow) *82*(2), 176-185. https://doi.org/10.11 34/S0006297917020092

Sahm H., Schütte H., Kula M.R. (1976). Purification and properties of 3-hexulosephosphate synthase from *Methylomonas M* 15. *European Journal of Biochemistry 66*(3), 591-596. https://doi.org/10.1111/j.1432-1033.1976.tb10586.x

Sanishvili R., Wu R., Kim D.E., Watson J.D., Collart F., Joachimiak A. (2004). Crystal structure of *Bacillus subtilis* YckF: structural and functional evolution. *Journal of Structural Biology 148*(1), 98-109. https://doi.org/10.1016/j.jsb.2004.04.006

Trong L.V., Thi N.L.N., Quang T.T., Duc H.N., Hoang M.C. (2026). Investigation of soyasapogenols from soybean seed germs as anti-apoptotic agents in colon cancer using an *in silico* approach. *Vietnam Journal of Biotechnology 24*(1), 81-97. https://doi.org/10.15625/vjbt-23376

Whitaker W.B., Jones J.A., Bennett R.K., Gonzalez J.E., Vernacchio V.R., Collins S.M., et al. (2016). Engineering the biological conversion of methanol to specialty chemicals in *Escherichia coli*. *Metabolic Engineering 39*, 49-59. https://doi.org/10.1016/j.ymben.2016.10.015

Yurimoto H., Kato N., Sakai Y. (2005). Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *The Chemical Record 5*(6), 367-375. https://doi.org/10.1002/tcr.20056