Journal of Genetic Engineering and Biotechnology

SHORT COMMUNICATIONS Open Access

Excavating the functionally crucial active-site residues of the DXS protein of *Bacillus subtilis* by exploring its closest homologues

Ashish Runthala[1]*, Tavakala Harsha Sai[1], Vandana Kamjula[1], Suresh C. Phulara[1], Vikrant Singh Rajput[1,2] and Karthikeyan Sangapillai[1]

Abstract

Background: To achieve a high yield of terpenoid-based therapeutics, 1-deoxy-d-xylulose-5-phosphate (DXP) pathway has been significantly exploited for the production of downstream enzymes. The DXP synthase (DXS) enzyme, the initiator of this pathway, is pivotal for the convergence of carbon flux, and is computationally studied well for the industrially utilized generally regarded as safe (GRAS) bacterium *Bacillus subtilis* to decode its vital regions for aiding the construction of a functionally improved mutant library.

Results: For the 546 sequence dataset of DXS sequences, a representative set of 108 sequences is created, and it shows a significant evolutionary divergence across different species clubbed into 37 clades, whereas three clades are observed for the 76 sequence dataset of *Bacillus subtilis*. The DXS enzyme, sharing a statistically significant homology to transketolase, is shown to be evolutionarily too distant. By the mutual information-based co-evolutionary network and hotspot analysis, the most crucial loci within the active site are deciphered. The 650-residue representative structure displays a complete conservation of 114 loci, and only two co-evolving residues ASP154 and ILE371 are found to be the conserved ones. Lastly, P318D is predicted to be the top-ranked mutation causing the increase in the thermodynamic stability of 6OUW.

Conclusion: The study excavates the vital functional, phylogenetic, and conserved residues across the active site of the DXS protein, the key rate-limiting controller of the entire pathway. It would aid to computationally understand the evolutionary landscape of this industrially useful enzyme and would allow us to widen its substrate repertoire to increase the enzymatic yield of unnatural molecules for in vivo and in vitro applications.

Keywords: DXS, Motif, Phylogeny, Consurf, Coevolution, Directed evolution

Introduction

Isoprenoids constitute the largest class of structurally and functionally diverse secondary metabolites and encompass more than 55,000 known compounds [1, 2]. These compounds have been traditionally deployed for the synthesis of aromatic, flavoring, and pharmaceutical molecules [3–8]. To date, plants are the major source of isoprenoid based bioactive molecules [9, 10], and it has led to an overexploitation of plants, causing severe environmental issues. For example, due to the heavy exploitation of *Taxus wallichiana* (Himalayan Yew) for the extraction of pharmacologically important isoprenoids [11], a 90% decline has been reported in its population across the Indo-Nepal Himalayan region and is, therefore, declared as an endangered species by the international union for conservation of nature (IUCN) [12]. Therefore, the global interest has now shifted to produce the isoprenoid based bioactive molecules from generally regarded as safe (GRAS) status microbes, such as *Bacillus subtilis* for pharmaceutical and nutraceutical applications [13, 14]. *Bacillus subtilis* produces isoprenoid compounds via 1-

* Correspondence: ashish.runthala@gmail.com
[1] Koneru Lakshmaiah Education Foundation, Guntur 522502, India
Full list of author information is available at the end of the article

deoxy-d-xylulose-5-phosphate (DXP) pathway which recruits seven enzymatic steps for the conversion of glyceraldehyde 3-phosphate (G3P) and pyruvate into prenyl precursors: Dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) in a ratio of 1:5 [15], as shown in Fig. 1. The pathway subsequently leads to the formation of many biomolecules including carotenoids, steroids, and ubiquinone. The enzyme 1-deoxy-d-xylulose-5-phosphate (DXP) synthase or DXS (EC 2.2.1.7) condenses glyceraldehyde-3-phosphate and pyruvate to synthesize DXP. This enzymatic reaction is the rate-limiting step and consequently, for an increased biosynthetic production rate of the end-products, a widespread interest has arisen in its research. Although the enzymes of the DXP pathway have been discovered two decades ago and have been studied in several microbes till now; however, their regulatory mechanism is still elusive. The DXS is the first enzyme of the DXP pathway belonging to the transferase family (EC 2.5.1.7). It catalyzes the condensation of G3P and pyruvate into the first intermediate of the pathway, i.e., DXP [15]. The DXP pathways enzymes, including DXS, are highly regulated. It has been suggested that the accumulation of DXS is regulated by other endogenous proteins, and perturbations of the growth conditions may affect its expression profile [16]. Besides playing a vital role in the biosynthesis of vitamin B1 and B6, it leads to the formation of isoprenoid precursors, and is thus functionally active at a crucial rate-limiting branch point of the pathway [17]. In contrast to an overwhelming orderly count of 2957 and 179,659 entries, existing in the Swissprot and TrEMBL databases of UniProtKB (May 14, 2020) [18, 19], only two DXS structures belonging to *Escherichia coli* (PDB ID: 2O1S) and *Deinococcus radiodurans* (PDB ID: 6OUW) could be determined through X-ray crystallography so far [20]. Besides sharing a high sequence identity of 45.47% and related catalytic activities, the two enzymes closely resemble each other (Fig. 2) [22]. To date, a wide range of isoprenoids have been produced from engineered *B. subtilis* for nutraceutical applications. However, the titers achieved till date from *B. subtilis* are far less than the ones achieved from *E. coli*, and it poses a major difficulty for their industrial scale up. The low yield of isoprenoids from engineered *B. subtilis* is a bottleneck for its industrial application [23]. It has been observed that DXS is a rate-limiting enzyme of DXP pathway because of its (i) low solubility; (ii) inhibition by IPP and DMAPP; and (iii) low turnover number [22]. The gene At3g47450, homologous to YqeH gene of *B. subtilis*, has been shown to regulate the accumulation of DXS in *Arabidopsis thaliana* [16]. Further, a negative correlation has been established between the Clp protease and DXP pathway enzymes. The Clp protease is also involved in the sporulation phase of *B. subtilis* [24], in which the biological production of isoprene decreases drastically [25]. This shows high endogenous regulation of DXP pathway enzymes, including DXS in *B. subtilis*.

The DXS enzyme is highly conserved, and the two proteins 621-residue 2O1S and 650-residue 6OUW share a topological similarity of 0.820 for 468 residues, as represented red and blue in Fig. 2, and their topological variation is localized across the loop regions and terminal overhangs. However, the DXS enzyme of *Bacillus subtilis* and its homologs are still not significantly explored for improving the catalytic activity. Although the three domains encoded by DXS share homology with the E1 subunit of pyruvate dehydrogenase and equivalent domains of transketolase (EC 2.2.1.1), their orientation is substantially different. The active site of DXS is present at the interface of domain I and II, unlike transketolase where it is present in the dimer interface [20]. Moreover, as recently shown, the active site of DXS is nearly twice the volume of transketolase and pyruvate dehydrogenase (EC 1.2.4.1) active sites [26, 27], and it should allow the biosynthesis of bulky molecules, making it an interesting target for the directed evolution methodologies. Hence, a careful scrutiny and dedicated dataset of homologs are urgently needed to accurately extract the closest entries to drive such methodologies.

To overcome the limitations and achieve an industrially high-yield (systematic) of isoprenoids from *B. subtilis*, there is an urgent need to computationally engineer its DXS enzyme. In this regard, the preliminary step is to screen its evolutionarily closest homologs for selecting the potent protein sequence(s) from its closest clade and tracing the highly conserved/variable sites so that a catalytically improved enzyme sequence could be redesigned through specific mutations proximal to its active site [28–31]. Due to the overabundant number of bacterial sequences other than *B. subtilis*, screening its functionally as well as topologically closest sequences is still a major challenge. A strategic comparative scrutiny thus becomes mandatory to screen the functionally closest set of natural variants for *Bacillus subtilis*.

In correlation with the cladistic divergence of the most predominant bacterial species *Enterobacteriaceae*, a non-redundant sequence dataset for *Bacillus subtilis* is constructed. The study focuses on the evolutionary diversification of the dataset and estimates the average branch length and topological conservation for all the resultant clades. It estimates the sequence conservation for the constructed dataset, and maps the active site residues within the well-studied representative DXS structure 6OUW. The co-evolving and hotspot residues are subsequently localized to analyze their degree of conservation and the key residues are designated for mutagenesis. Although it is impossible to explore the theoretical sequence space of a protein, the study will help to develop automated algorithms for decoding the

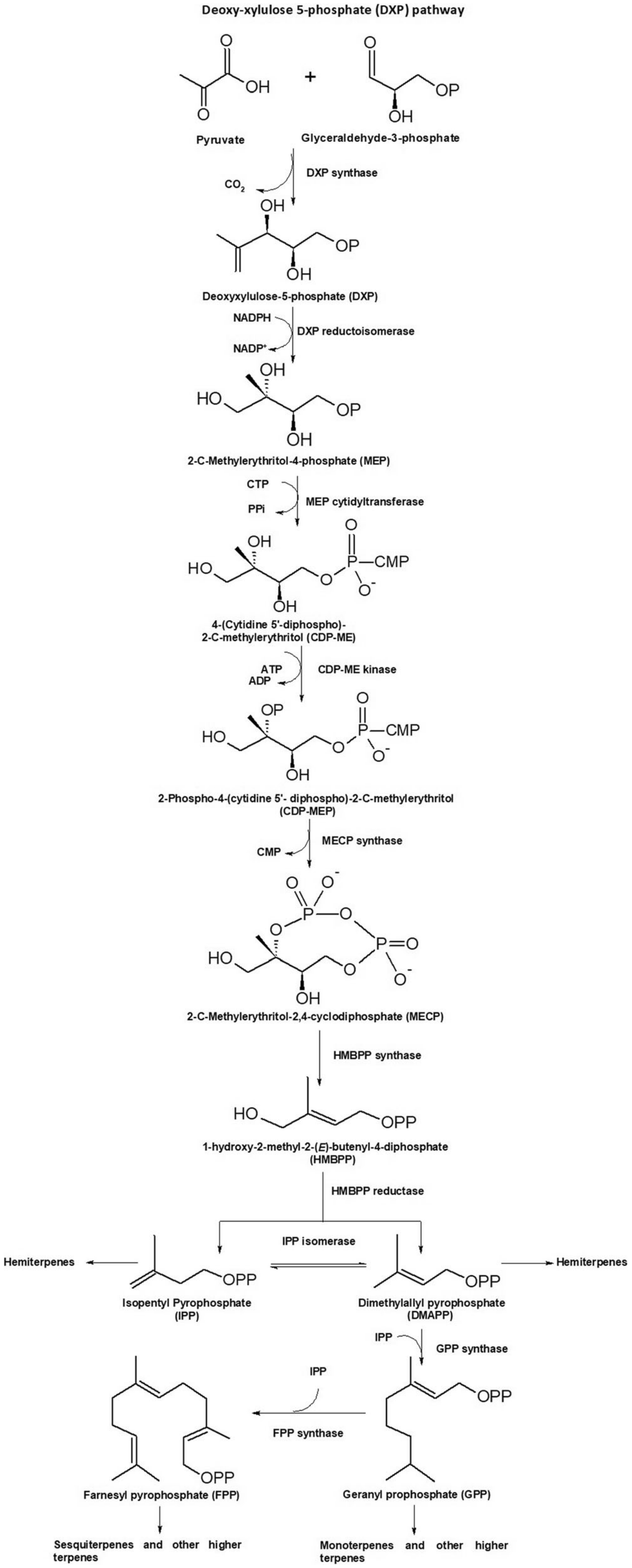

Deoxy-xylulose 5-phosphate (DXP) pathway
Pyruvate
Glyceraldehyde-3-phosphate
DXP synthase
CO2
Deoxyxylulose-5-phosphate (DXP)
NADPH
NADP+
DXP reductoisomerase
2-C-Methylerythritol-4-phosphate (MEP)
CTP
PPi
MEP cytidyltransferase
4-(Cytidine 5'-diphospho)-2-C-methylerythritol (CDP-ME)
ATP
ADP
CDP-ME kinase
2-Phospho-4-(cytidine 5'- diphospho)-2-C-methylerythritol (CDP-MEP)
CMP
MECP synthase
2-C-Methylerythritol-2,4-cyclodiphosphate (MECP)
HMBPP synthase
1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP)
HMBPP reductase
IPP isomerase
Hemiterpenes
Isopentyl Pyrophosphate (IPP)
Dimethylallyl pyrophosphate (DMAPP)
Hemiterpenes
IPP
GPP synthase
IPP
FPP synthase
Farnesyl pyrophosphate (FPP)
Geranyl prophosphate (GPP)
Sesquiterpenes and other higher terpenes
Monoterpenes and other higher terpenes

Fig. 1 The MEP pathway. The pathway produces IPP and DMAPP as the 5-carbon building blocks for the biosynthesis of isoprenoids. It condenses glyceraldehyde 3-phopshate and pyruvate through DXS and forms IPP and DMAPP through six added steps. DMAPP linkage with one or two molecules of IPP forms monoterpene or sesquiterpene, respectively

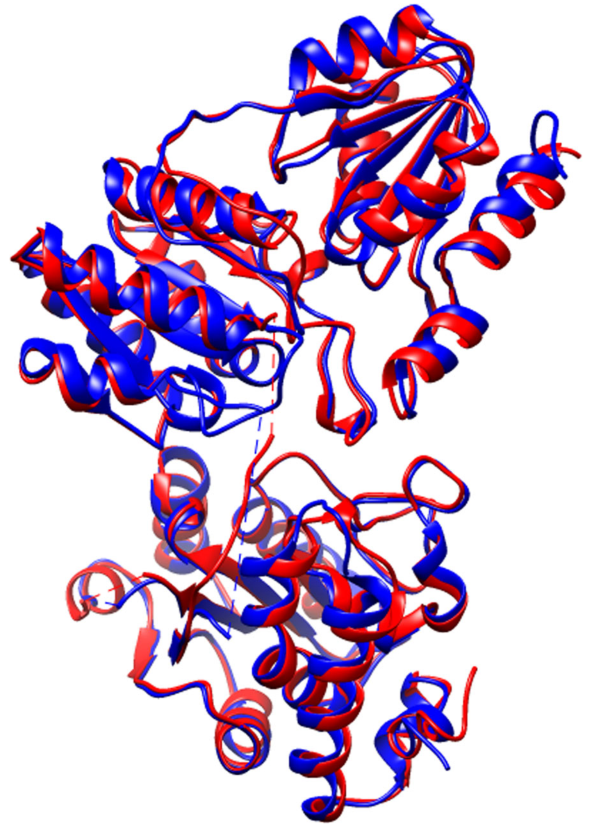


Fig. 2 Structural overlap of DXS structures of *Escherichia coli* and *Deinococcus radiodurans*, showing a structural conservation of 468 residues across the secondary structure elements. The structures are superimposed and represented through Chimera 1.14 [21]

functionally crucial sites of a protein. It would aid the construction of focused mutant library which will have a significant impact on generating catalytically improved enzymes in short time frames. Thus, the study is certainly the need of the hour for computationally evolving a catalytically improved DXS enzyme sequence with enhanced substrate affinity.

Materials and methods

Building the sequence dataset and alignment

Screening the DXS protein in UniprotKB [32], a sequence dataset is constructed, as shown in the overall algorithmic flowchart of the study (Fig. 3). As per the sequence length of functionally similar protein structures, a robust length filter of 500-750 is deployed to purge all the functionally variant entries and build the dataset of 546 sequences (set A). It comprises 12, 11, 2, 2, 1, and 518 sequences from plant, animal, protista, human, fungi, and bacteria respectively, as enlisted in the Supplementary Table 1.

For all enzyme design protocols, the first and foremost challenge is to designate the evolutionarily conserved/variant regions and decipher the functionally important residues. For computationally evolving the proteins, it thus suggests the urgent need of building the evolutionarily closest sequence cluster and assesses the variation of sequences across the *Bacillus subtilis* in correlation with the most abundant set of proteins. As the bacterial sequences majorly predominate set A, belonging to the subfamily *Enterobacteriaceae*, it is considered to evolutionarily track the sequence divergence more effectively. Further, as this dataset is found to encode 617-626 residues, and the smaller sequence length is evolutionarily preferred [33], the entries are clustered as per their lengths to define two subsets set A_{small} and set A_{large}. The sequences Q7VRH9.1 and Q8D357.1 are orderly selected as the representative set A entries for these subsets.

The 546-set sequence dataset is aligned using ClustalO server by deploying the default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/) [34, 35]. The server aligns the sequence dataset to derive the distances between the aligned residues and constructs a guide tree for further improving the alignment. The resultant sequence profile should thus yield a more reliable evolutionary relationship between the 546 sequences. To extrapolate the methodology to *Bacillus subtilis*, a set of 301 sequences are likewise retrieved and the redundant hits are purged through MMSeqs2 protocol [36]. It deploys a three-step cascaded workflow for mutually aligning the input sequences on the basis of an ungapped alignment and sensitive k-mer matching algorithm to yield the entries scoring higher than a given threshold. Purging the outliers, the final 76 sequence dataset (set B) is constructed. The entries are clustered as per their lengths and their sequence motifs are evaluated through the multiple EM for motif elicitation (MEME) server [37]. To confirm the functional annotation of the sequence dataset before deploying it for subsequent studies, the top three motifs of length 6-50 are screened by this online software suite. As the presence of signature sequence aids an initial computational verification of the function, the server increases the functional credibility of the sequence dataset. It yields 2 subsets, set B_{small} and set B_{large}, orderly having a sequence length of 619 and 667 respectively, for the subsequent analysis, and the sequences AJW87412.1 and WP_007410329.1 are orderly selected as the representative entries.

Sequence and structural analysis

Several physical and chemical parameters, viz., molecular weight, amino acid composition, extinction coefficient, estimated theoretical pI, and grade average of hydropathicity, aliphatic index, and instability index are important to estimate the physiochemical properties and topological features of a protein sequence. For the representative sequences of both datasets, the features are estimated through ProtParam (https://web.expasy.org/ProtParam) [38].

PSIPRED [39] is deployed to predict the three-state secondary structure for the selected representative sequences. It provides information corresponding to α-helices, β-sheets, coils, transmembrane helices, signal peptides, membrane interactions, re-entrant helix, and putative domain boundaries. For the representative sequences, the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/), based on the transmembrane hidden Markov model, is subsequently used to predict the integral transmembrane helices and discriminate between the soluble and membrane proteins [40]. Besides estimating the number of transmembrane helices, it predicts the expected number of transmembrane helix residues for the selected protein sequence [38].

Evolutionary analysis

A phylogenetic tree is an estimate of the relationships among taxa/sequences and their hypothetical common ancestors. Most molecular phylogenetic trees estimate the statistically significant relationships among the species/sequences [41]. Molecular evolutionary genetics analysis (MEGA) is one such widely deployed tool to measure evolutionary distance among the sequences [42]. MEGA X is used to construct the interactive evolutionary trees for the datasets A and B. The constructed trees are visualized using the interactive tree of life (iTOL) server (https://itol.embl.de/) [43]. It is used to analyze the evolutionary relationships across the sequence datasets and distinctly highlight the species.

Crucial residues for functional mutagenesis and directed evolution

DXS is an essential enzyme of the pathway and its expression is very rigidly controlled by the bacteria. Making its active site open to a variety of substrates will thus be phenomenal in increasing its productivity. To study the key substrate-binding residues, the experimentally solved structure, closest to the constructed sequence dataset of the functionally similar homologs of *Bacillus subtilis*, is screened from the PDB database through HHPred [44]. To functionally characterize and confirm the derived dataset, a MEME server is used to search and identify the previously unidentified motifs in the sequence dataset [37]. The motif length of 6-50 is used to localize the top three motifs.

The retrieved structure is subsequently fed to CASTp [45] for mapping the active site residues, lining the cavity. For localizing the conserved residues and their degree of conservation across the active site, Consurf [46] is deployed, and the average pairwise distance among the sequences, along with its lower and upper level, is estimated. For estimating the relative degree of sequence conservation, the experiment is repeated by including the 13 transketolase homologs. Functional divergence within and among these datasets is estimated as the level of evolutionary distant entries that usually emerge in such analysis. The functionally conserved and co-evolving residues are subsequently localized through the MISTIC approach using the mutual co-evolving information (MI) of the sequence profile [47]. MI is calculated between the residue columns and it reflects the extent of the co-evolutionary impact of one residue at another position within the MSA. Every node in the

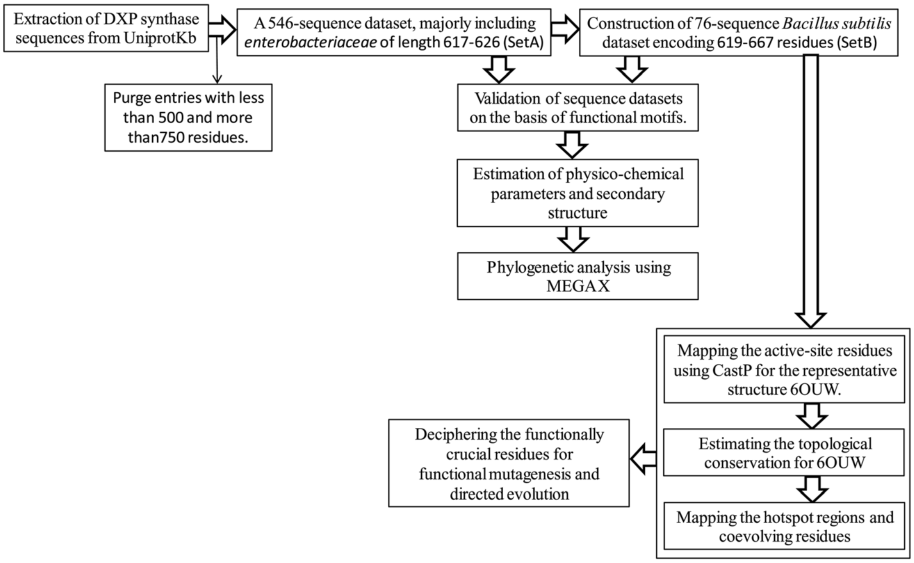

Extraction of DXP synthase sequences from UniprotKb
A 546-sequence dataset, majorly including *enterobacteriaceae* of length 617-626 (SetA)
Construction of 76-sequence *Bacillus subtilis* dataset encoding 619-667 residues (SetB)
Purge entries with less than 500 and more than750 residues.
Validation of sequence datasets on the basis of functional motifs.
Estimation of physico-chemical parameters and secondary structure
Phylogenetic analysis using MEGAX
Mapping the active-site residues using CastP for the representative structure 6OUW.
Estimating the topological conservation for 6OUW
Mapping the hotspot regions and coevolving residues
Deciphering the functionally crucial residues for functional mutagenesis and directed evolution

Fig. 3 Algorithmic flowchart for the analysis from sequence dataset to functional annotation and coevolving residues across the active site

resultant network indicates a protein and their linker edge signifies the statistically significant similarity between them.

Hotspot regions

A most crucial step of semi-rational directed evolution strategy is the selection of hotspot loci whose mutations lead to a significant improvement of the catalytic/biological activity of the proteins [48]. Hotspots are the sites where alanine mutations lead to an increase of at least 2 kcal/mol in the binding free energy. To analyze the mutational landscape and localize the hotspot regions within/proximal to the active site for the representative structure 6OUW, the hotspot server [49] is deployed for studying the functionally crucial and correlated hotspots. It will provide us a vital dataset to build the dedicated mutant libraries for the semi-rational directed evolution of these functionally similar enzymes.

To theoretically investigate the role of the identified hotspot residues on the biological stability of the representative structure 6OUW, their top-ranked missense mutations are deciphered through Popmusic (http://dezyme.com/en/software) [50]. On the basis of a linear combination of environment-specific statistical potentials and solvent accessibility of the mutated residue, it introduces a point mutation in the structure and estimates the resultant change in the thermodynamic stability in terms of ΔΔG score. The combinational assessment corrects its bias toward the destabilizing mutations, which usually impose physical symmetries under inverse mutations, and increases the robustness of the protocol [51, 52]. While the solvent accessibility is estimated within the range of 0 (buried) to 100 (*fully accessible*), the mutation effect on the protein stability is considered stabilizing if $\Delta\Delta G < 0$. For strengthening the credibility of predictions, the Maestro server (https://pbwww.che.sbg.ac.at/maestro/web) is subsequently deployed [53]. On the basis of the statistical energy functions for the sequence and structural topology of the input protein, it estimates the difference in the folding free energy of the structure upon mutations along with a confidence score through multiple linear regression, SVM, and neural networks [54]. It scrutinizes the stabilizing disulfide bonds and free energy change through a high-throughput scanning of multi-point mutations [53]. Maestro predictions are less biased relative to the group of inverse mutations [55]. To the best of our knowledge, no tool suggests the coupled mutations for introducing the new stabilizing contacts within the structure.

Results and discussion

Sequence and structural analysis

The computational methods aid a swift characterization and estimation of the functional properties of protein sequences. The physicochemical properties, viz., pI value, extinction coefficient, molecular weight, average hydrophobicity, and instability index are estimated through the ProtParam server for the representative sequences of set A and set B to determine their uniqueness (Table 1), as has been recently shown [48]. Isoelectric point or pI is the pH where the protein molecule has no net charge. The pI value higher and lower than 7.0 orderly indicates the alkaline and acidic character of a protein respectively. While the theoretical pI of *Bacillus subtilis* sequences shows a pI of 5.80-5.91, indicating the acidic nature, the *Enterobacteria* sequences display a significantly higher basic pI of 9.34-9.57. It indicates a highly narrow range of sequence variation within the species, and a significantly diverse range of variations across various species. It has been reported that an instability index score of less than 40 confirms the structural stability of a protein [37]. The representatives for set A (Q7VRH9.1 and Q8D357.1) and set B (AJW87412.1 and WP_007410329.1) exhibit a score within the range of 30-40 (Table 1), indicating their cellular stability. An interrelation is observed between the stability and half-life, and the stable proteins manifesting an in vivo half-life of at least 16 h, are usually expected to express an aliphatic index of less than 40 [49]. The inherent stability of these proteins provides an added benefit of minimizing the experimental costs and steps, usually deployed for such studies.

The average extinction coefficient (measured as M^{-1} cm^{-1}) indicates the quantity of light absorbed at 280 nm, and it is found to be significantly different for the representative set A and set B sequences (Table 1). The score is dependent on the frequency of cysteine, tryptophan, and tyrosine residues, and it aids a deeper quantitative understanding of the interactions against any other protein/ligand. The aliphatic index is subsequently estimated for these proteins and it indicates a constricted index range of 79.48-88.74 and 98.98-103.74 for the *Bacillus subtilis* and *Enterobacteriaceae* sequences respectively. The aliphatic index is the relative volume occupied by the aliphatic amino acids, and the estimations show a bit lower thermostability for the *Bacillus subtilis* sequences. Further, the set A representative sequences illustrate a diverse range of the grand average of hydrophobicity (GRAVY) scores from −0.054 to −0.122, in contrast to the respective range of −0.149 to −0.269 for the set B sequences. The negative GRAVY score signifies the nonpolar nature of a protein molecule. It indicates the energetically favorable interactions with the hydrophilic water molecules, and it thus shows that the hydrophobic residues are robustly conserved in *Bacillus subtilis* sequences [56]. Further, the residue composition analysis orderly shows a high proportion of glycine, 10.70% and 10.54% for set A (Fig. 4a) and 10.50% and

Table 1 Physicochemical parameters, estimated by ProtParam, indicating a substantial variation for certain features

Sequence accession ID (set)	Amino acids	Molecular weight (KDa)	Extinction coefficient ($M-^{1}$ $cm-^{1}$)	Theoretical pI	GRAVY score	Aliphatic index	Instability index
Q7VRH9.1 (set A)	617	68.722	64220	9.34	−0.122	98.98	34.64
Q8D357.1 (set A)	626	69.851	56520	9.57	−0.054	103.74	34.09
AJW87412.1 (set B)	619	67.913	53985	5.99	−0.149	88.74	33.30
WP_007410329.1 (set B)	667	72.584	91790	5.13	−0.269	79.48	36.68

9.6% for set B (Fig. 4b). It indicates that the sequences encode a very low number of charged residues and hence, their extracellular solubility should be too low within the solvent.

PSIPRED is further used to predict the secondary structure of the representative sequences of the two datasets (Fig. 5). The secondary structure elements, viz., helix, sheet, and coil define a protein structure and play a key role in the design of various bioanalytical experiments. A residue fraction of 42.78% and 13.12% orderly defines the helical and stranded substructure of Q7VRH9.1, in contrast to the respective proportion of 43.61% and 12.93% for Q8D357.1. For set B also, a fraction of 43.29% helix and 12.76% strand residues are encoded in AJW87412.1 in comparison to the respective proportion of 43.62% and 11.69% for WP_007410329.1. It suggests a substantial predominance of the helical residues within this class of proteins. The transmembrane helical regions are further localized for the representative sequences through TMHMM, and sequences Q7VRH9.1 and Q8D357.1 are found to respectively encode a fraction of 0.863 and 3.647 transmembrane helical residues (Fig. 6), in contrast to the respective proportion of 0.13269 and 2.12531 of the set B sequences AJW87412.1 and WP_007410329.1.

Evolutionary analysis

Screening the DXS protein in UniprotKB [32], it results in 551 sequences with sequence length ranging from 519 to 741. The sequence length range of the dataset varies from 106-684 and a large majority (84.615%) is found to be within 600-700. Purging the functional outliers, it

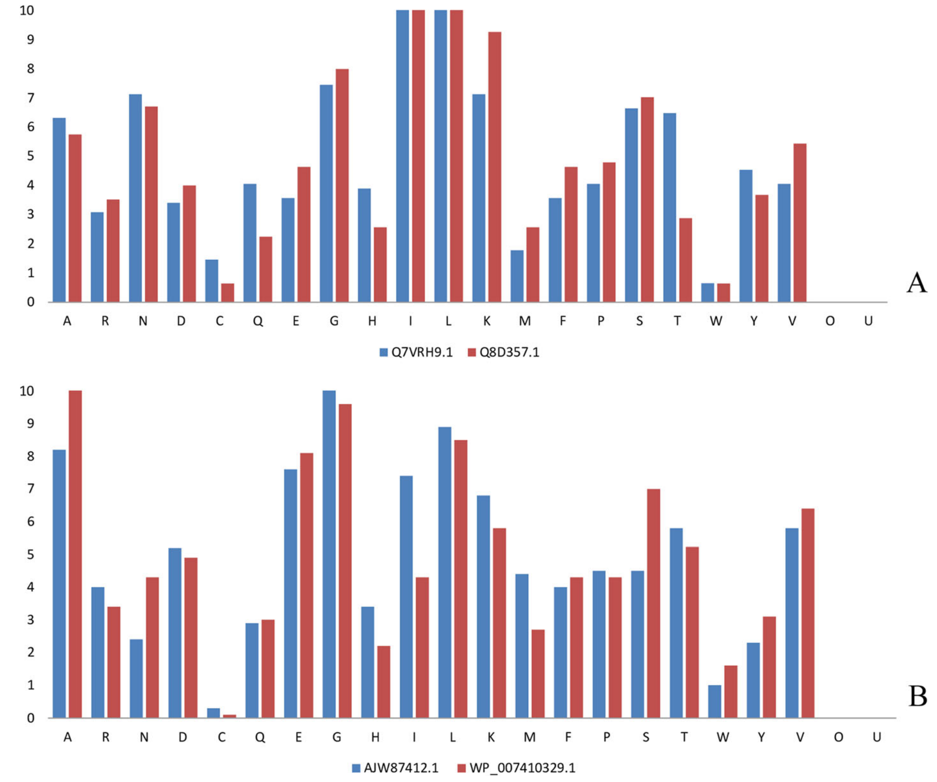


Fig. 4 Variation of residue frequencies between the representative sequences Q7VRH9.1 and Q8D357.1 and AJW87412.1 and WP_007410329.1 of the two datasets, belonging to *Enterobacteriaceae* and *Bacillus* species respectively. A substantial variation in the residue percentages is observed for a few residues within and between the two datasets

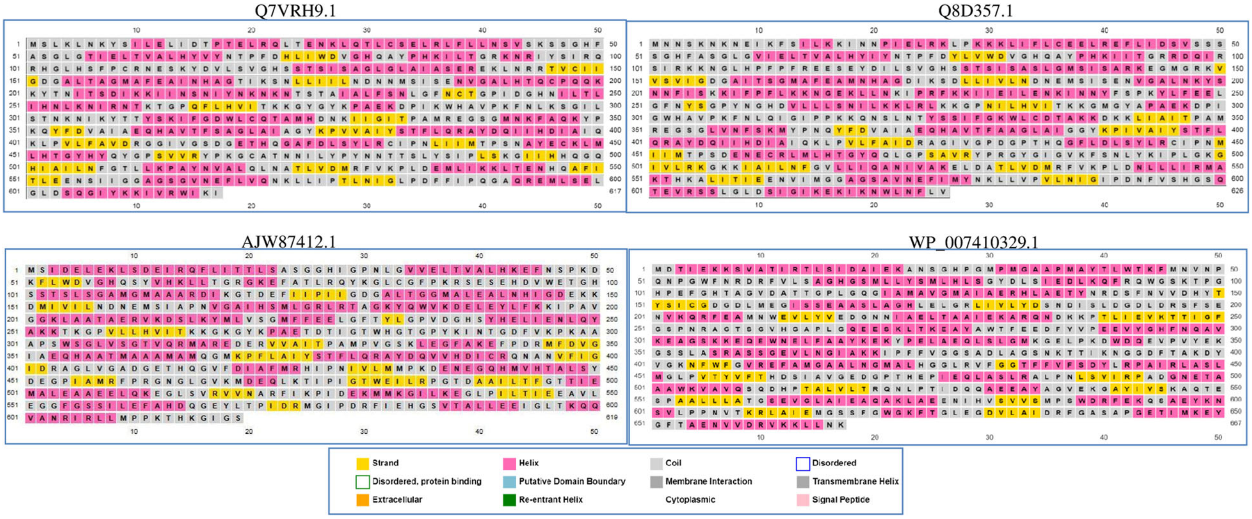


Fig. 5 PSIPRED based estimation of the secondary structure for the representative sequences (**a**) Q7VRH9.1, (**b**) Q8D357.1, (**c**) AJW87412.1, and (**d**) WP_007410329.1 of the two sets A and B

yields 546 sequence dataset whose alignment is fed to MEGA X to derive the evolutionary relationship. Although computing an evolutionarily optimal tree topology is termed to be an NP-hard combinatorial optimization problem [57], the maximum likelihood method [58] has been well proven to yield the robustly accurate results for the sequence dataset than the other methods [59, 60] and is therefore deployed to infer the phylogenetic history of the dataset. The JTT substitution matrix [61] is utilized for estimating the evolutionary distances within the dataset at 4 discrete gamma distribution categories for the residue substitution. Bootstrap resampling (1000 replicates) [62] is used to assess the robustness of the groupings. It integrates the replicate trees for the clustered set of associated taxa into the bootstrap test to show its percentage next to the branches. The grouped set of sequences is found to encode a substantial sequence similarity in comparison to the other dataset entries and is expected to share a statistically significant evolutionary relationship. The tree is subsequently visualized in the IToL server for analyzing it further (Fig. 7). As nature tends to decrease the sequence lengths of proteins for saving the energy required to synthesize and fold longer proteins [63], the

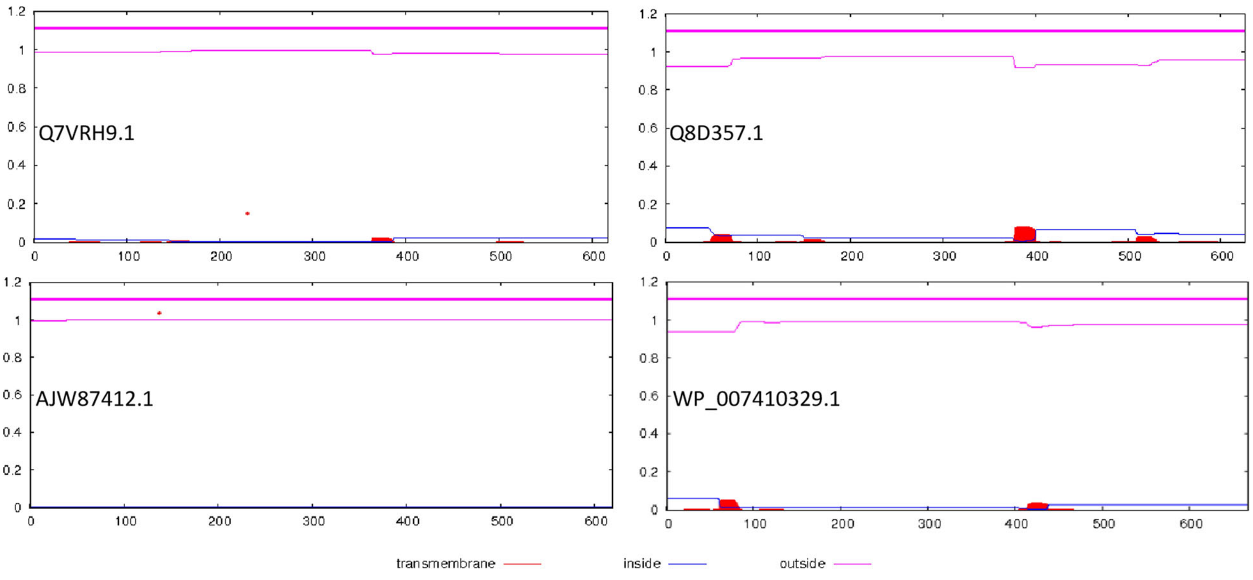


Fig. 6 Prediction of protein transmembrane structure for representative sequences Q7VRH9.1, and Q8D357.1, and AJW87412.1 and WP_007410329.1

shortest sequences of various clades should be prioritized to evaluate their evolutionary relationships. For the inferred 37 clades, three shortest sequences are further selected to build a new dataset of 108 sequences, and their functional similarity is confirmed through their motifs to further evaluate their evolutionary relationships. The resultant tree shows significantly distant associations across 37 clades belonging to *Leptospiraceae*, *Helicobacteriacea*, *Desulfovibrionaceae*, *Clostridiacea*, *Peptococccacea*, *Bacillacea*, *Aphanocthecaceae*, *Synechoccaceae*, *Mycobacteriaceae*, *Hominidae*, *Bacteroidaceae*, *Rhodospirillaceae*, *Rhodobacteraceae*, *Sphinogomoadace*, *Bradyrhizobiaceae*, *Phyllobacteriaceae*, *Rhizobiaceae*, *Brucellaceae*, *Bovidae*, *Ectothiorhodospiraceae*, *Erwiniaceae*, *Xanthomonadaceae*, *Monraxellaceae*, *Pseudomondaceae*, *Francisellaceae*, *Shewanellaceae*, *Vibrionaceae*, *Pasteurellaceae*, *Yesiniaceae*, *Neisseriacea*, *Rhobocyclaceae*, *Commonadaceae*, *Burkholderiacea*, *Morganellaceae*, *Pectobacteriaceae*, *Enterobacteriaceae*, and *Murida* species. The log-likelihood and total tree length scores for the tree are found to be −56151.89 and 33.750 respectively. Similarly, the tree is constructed for the 76 sequence dataset (set B) and visualized in the IToL server. Set B is found to evolve in three distinct clades. The log-likelihood score and total tree length of its consensus tree solution are found to be −15262.705 and 20.657 respectively. As per the evolutionary trees of set A and set B, it should be logical to state that these datasets are divergent, and extending the sequence space to set A may prove to be incorrect. As speculated, the strategy has worked well, and despite sharing significant homology with DXS [64], a set of well-clustered sets of only 13 transketolase is found, as marked with blue squared boxes in Fig. 8. It clearly indicates clear segregation of these sequences from the other entries, and further prioritization of the closest sequence space of DXS [65] is thus possible from the derived alternative residue dataset. The derived dataset further indicate a sequence identity within the range of 42.16-99.84 against 6OUW, and again proves an evolutionarily significant closeness

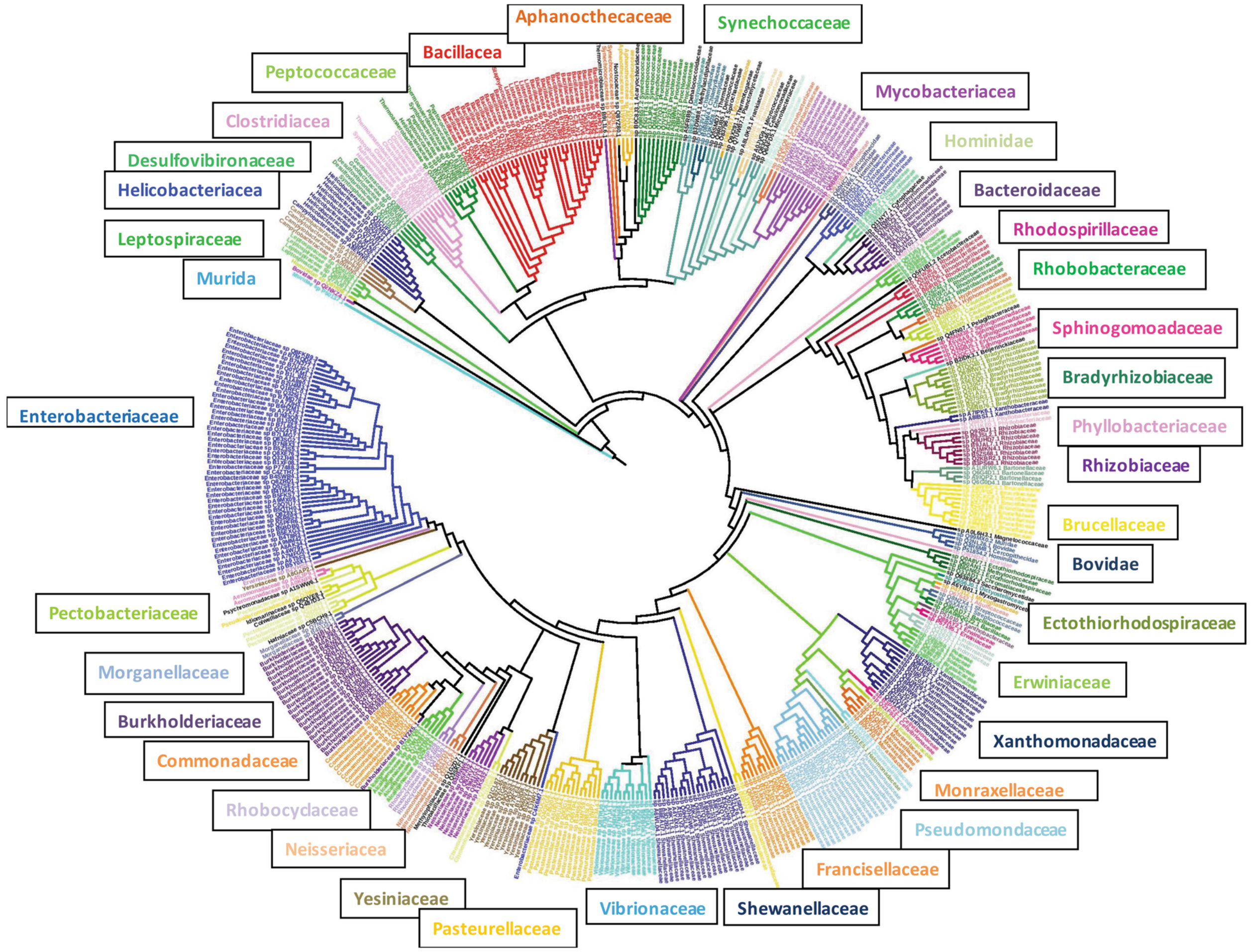


Fig. 7 Phylogenetic tree of the 546 sequence dataset of DXP synthase. An average pairwise score of 1.012 within the range of 1.0175 to 2.876 is found

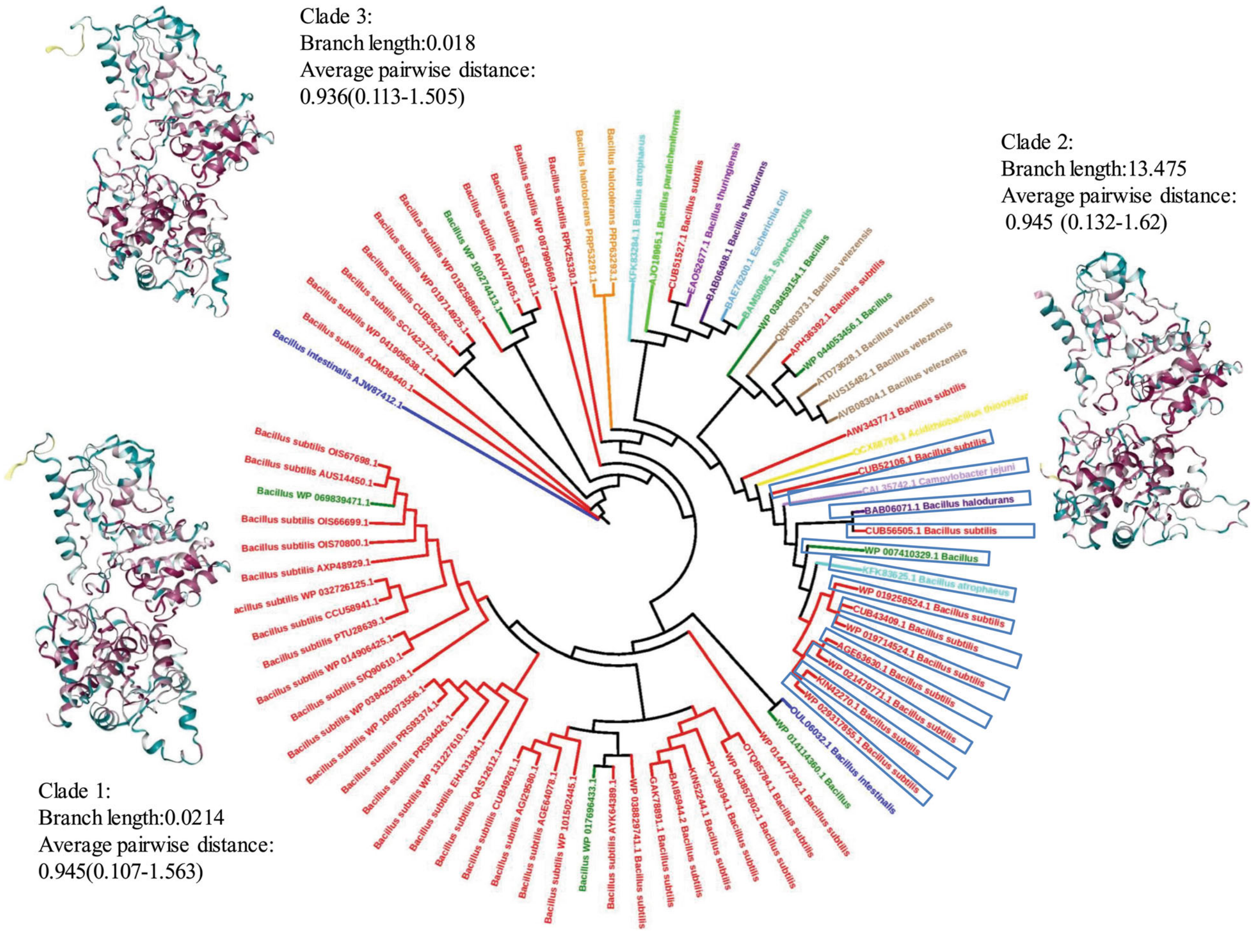

Clade 3:
Branch length:0.018
Average pairwise distance:
0.936(0.113-1.505)
Clade 2:
Branch length:13.475
Average pairwise distance:
0.945 (0.132-1.62)
Clade 1:
Branch length:0.0214
Average pairwise distance:
0.945(0.107-1.563)
Bacillus intestinalis AJW87412.1
Bacillus subtilis ADM38440.1
Bacillus subtilis WP 041905638.1
Bacillus subtilis SCV42372.1
Bacillus subtilis CUB36265.1
Bacillus subtilis WP 019714925.1
Bacillus subtilis WP 019258866.1
Bacillus WP 100274413.1
Bacillus subtilis ARV47405.1
Bacillus subtilis ELS61891.1
Bacillus subtilis WP 087990669.1
Bacillus subtilis RPK25330.1
Bacillus halotolerans PRP53291.1
Bacillus halotolerans PRP63293.1
KFK83284.1 Bacillus atrophaeus
AJO18965.1 Bacillus paralicheniformis
CUB51527.1 Bacillus subtilis
EAO52677.1 Bacillus thuringiensis
BAB06498.1 Bacillus halodurans
BAE76200.1 Escherichia coli
BAM50805.1 Synechocystis
WP 038459154.1 Bacillus
QBK80373.1 Bacillus velezensis
APH36392.1 Bacillus subtilis
WP 044053456.1 Bacillus
ATD73628.1 Bacillus velezensis
AUS15482.1 Bacillus velezensis
AVB08304.1 Bacillus velezensis
AIW34377.1 Bacillus subtilis
OCX88786.1 Acidithiobacillus thiooxidans
CUB52106.1 Bacillus subtilis
CAL35742.1 Campylobacter jejuni
BAB06071.1 Bacillus halodurans
CUB56505.1 Bacillus subtilis
WP 007410329.1 Bacillus
KFK83625.1 Bacillus atrophaeus
WP 019258524.1 Bacillus subtilis
CUB43409.1 Bacillus subtilis
WP 019714524.1 Bacillus subtilis
AGE63630.1 Bacillus subtilis
WP 021479771.1 Bacillus subtilis
KIN42270.1 Bacillus subtilis
WP 029317855.1 Bacillus subtilis
OUL06032.1 Bacillus intestinalis
WP 014114360.1 Bacillus
WP 014477302.1 Bacillus subtilis
OTQ85784.1 Bacillus subtilis
WP 043857802.1 Bacillus subtilis
PLV39094.1 Bacillus subtilis
KIN52244.1 Bacillus subtilis
BAI85944.2 Bacillus subtilis
GAK78891.1 Bacillus subtilis
WP 038829741.1 Bacillus subtilis
Bacillus subtilis AYK64389.1
Bacillus WP 017696433.1
Bacillus subtilis WP 101502445.1
Bacillus subtilis AGE64078.1
Bacillus subtilis AGI29580.1
Bacillus subtilis CUB49261.1
Bacillus subtilis QAS12612.1
Bacillus subtilis EHA31384.1
Bacillus subtilis WP 131227610.1
Bacillus subtilis PRS94426.1
Bacillus subtilis PRS93374.1
Bacillus subtilis WP 106073556.1
Bacillus subtilis WP 038429288.1
Bacillus subtilis SIQ90610.1
Bacillus subtilis WP 014906425.1
Bacillus subtilis PTU28639.1
Bacillus subtilis CCU58941.1
acillus subtilis WP 032726125.1
Bacillus subtilis AXP48929.1
Bacillus subtilis OIS70800.1
Bacillus subtilis OIS66699.1
Bacillus WP 069839471.1
Bacillus subtilis AUS14450.1
Bacillus subtilis OIS67698.1

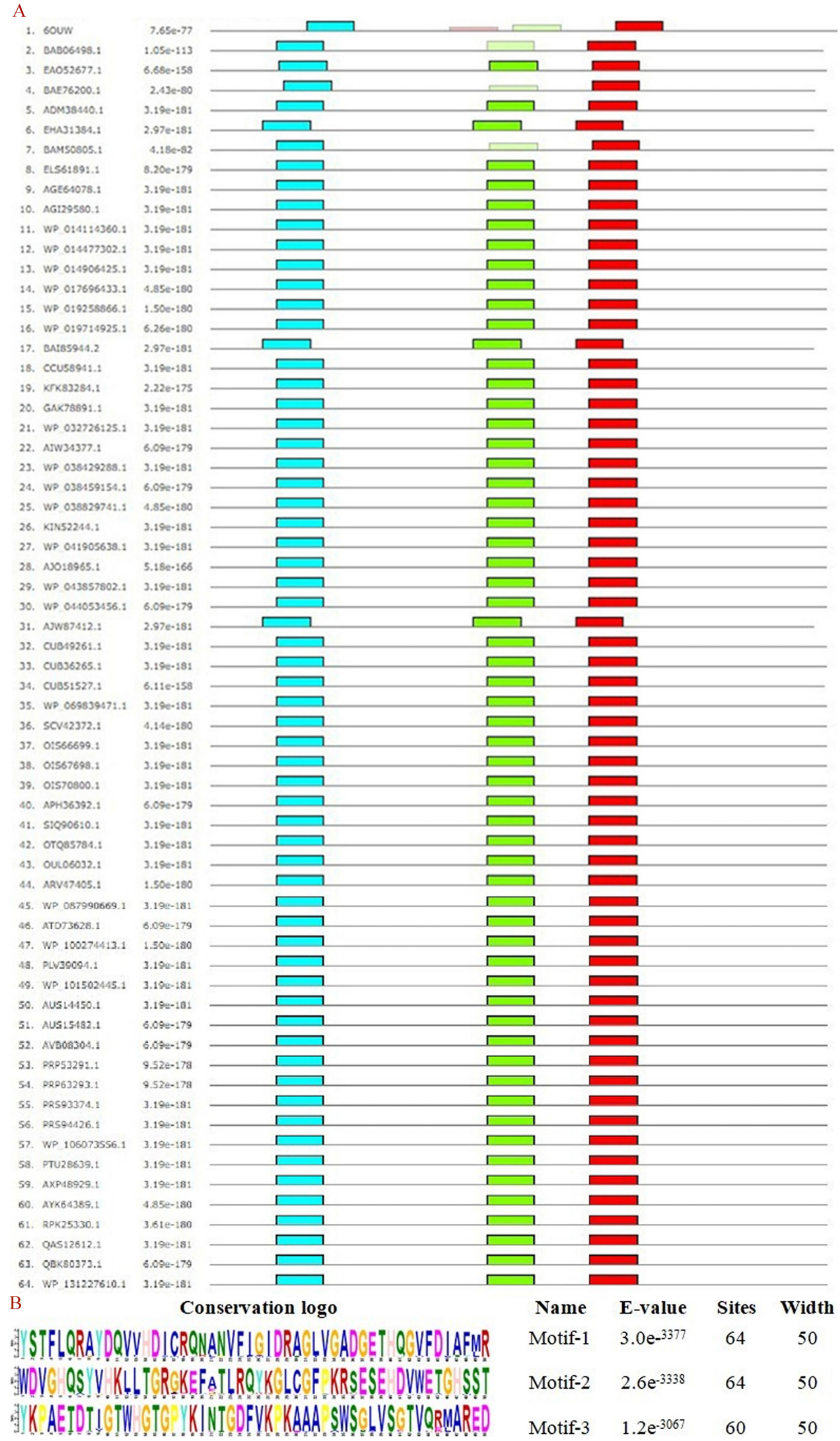
Fig. 8 Phylogenetic tree of 76 sequences of DXP synthase proteins in *Bacillus subtilis*. The 13 transketolase sequences are marked with blue squared boxes, and it shows a clear segregation of the two enzymes

of data. Our results thus show a significant evolutionary diversification of DXS across various species, in contrast to *Bacillus subtilis*, and this is crucial to derive a detailed dataset of the encoded alternative residues for constructing the mutant library.

Crucial residues for functional mutagenesis and directed evolution

DXS is a very crucial rate-limiting enzyme of the methylerythritol phosphate (MEP) pathway and its expression is very rigidly controlled by the bacteria. Rather than modulating the active site, its overexpression has been widely deployed to increase the bacterial production of carotenoids [66, 67]. To reliably analyze the active site for mutagenesis, the experimentally solved structure closest to the profile of the 63 sequence dataset of *B. subtilis* is screened through HHPred. With an *E* value and sequence identity of 7.3e−90 and 42% respectively, the template having PDB ID: 6OUW is found to be the closest protein structure [68]. For confirming the functional similarity of the constructed 63 sequence dataset of *B. subtilis*, the top three motifs are identified through MEME suite (Fig. 9a). The three potential motifs are designated as 1, 2, and 3 (Fig. 9b), and motif2 is not found to demonstrate a statistically significant conservation score in 4 sequences, including the reference protein structure. Within the reference structure, motifs 1 and 3 are orderly located between residues 99-148 and 416-465, and appear to play a functionally significant role in domain1 and domain2 respectively [20]. Against the reference structure 6OUW, the constructed sequence alignment is fed to ESPRIPT3 [69] and it presents the conservation of only a few residues, scattered across the chain (Supplementary Fig. S1).

To decode the vital structural residues across the sequence profile, CASTp [45] server is used to localize the active site of the representative structure 6OUW (Fig. 10a). The top-ranked resultant pocket shows a surface area and volume of 1487.0 Å^2 and 3057.5 Å^3 respectively. A set of 61 residues, viz., GLY48, GLY49, LEU50, HIS51, LEU52, ALA53, SER54, GLY57, ASP79, VAL80, HIS82, GLN83, LYS101, PHE109, GLY123, HIS124, ALA125,



1.	6OUW	7.65e-77
2.	BAB06498.1	1.05e-113
3.	EAO52677.1	6.68e-158
4.	BAE76200.1	2.43e-80
5.	ADM38440.1	3.19e-181
6.	EHA31384.1	2.97e-181
7.	BAM50805.1	4.18e-82
8.	ELS61891.1	8.20e-179
9.	AGE64078.1	3.19e-181
10.	AGI29580.1	3.19e-181
11.	WP_014114360.1	3.19e-181
12.	WP_014477302.1	3.19e-181
13.	WP_014906425.1	3.19e-181
14.	WP_017696433.1	4.85e-180
15.	WP_019258866.1	1.50e-180
16.	WP_019714925.1	6.26e-180
17.	BAI85944.2	2.97e-181
18.	CCU58941.1	3.19e-181
19.	KFK83284.1	2.22e-175
20.	GAK78891.1	3.19e-181
21.	WP_032726125.1	3.19e-181
22.	AIW34377.1	6.09e-179
23.	WP_038429288.1	3.19e-181
24.	WP_038459154.1	6.09e-179
25.	WP_038829741.1	4.85e-180
26.	KIN52244.1	3.19e-181
27.	WP_041905638.1	3.19e-181
28.	AJO18965.1	5.18e-166
29.	WP_043857802.1	3.19e-181
30.	WP_044053456.1	6.09e-179
31.	AJW87412.1	2.97e-181
32.	CUB49261.1	3.19e-181
33.	CUB36265.1	3.19e-181
34.	CUB51527.1	6.11e-158
35.	WP_069839471.1	3.19e-181
36.	SCV42372.1	4.14e-180
37.	OIS66699.1	3.19e-181
38.	OIS67698.1	3.19e-181
39.	OIS70800.1	3.19e-181
40.	APH36392.1	6.09e-179
41.	SIQ90610.1	3.19e-181
42.	OTQ85784.1	3.19e-181
43.	OUL06032.1	3.19e-181
44.	ARV47405.1	1.50e-180
45.	WP_087990669.1	3.19e-181
46.	ATD73628.1	6.09e-179
47.	WP_100274413.1	1.50e-180
48.	PLV39094.1	3.19e-181
49.	WP_101502445.1	3.19e-181
50.	AUS14450.1	3.19e-181
51.	AUS15482.1	6.09e-179
52.	AVB08304.1	6.09e-179
53.	PRP53291.1	9.52e-178
54.	PRP63293.1	9.52e-178
55.	PRS93374.1	3.19e-181
56.	PRS94426.1	3.19e-181
57.	WP_106073556.1	3.19e-181
58.	PTU28639.1	3.19e-181
59.	AXP48929.1	3.19e-181
60.	AYK64389.1	4.85e-180
61.	RPK25330.1	3.61e-180
62.	QAS12612.1	3.19e-181
63.	QBK80373.1	6.09e-179
64.	WP_131227610.1	3.19e-181

Name	E-value	Sites	Width
Motif-1	3.0e-3377	64	50
Motif-2	2.6e-3338	64	50
Motif-3	1.2e-3067	60	50

Fig. 9 a Statistically significant occurrence of tandem motifs within the 63 DXS sequences of *Bacillus subtilis*, in correlation with the reference structure 6OUW. Motifs 1 and 3 are found statistically conserved across all homologs. **b** Conservation logo and statistical *E* value scores of localized motifs

VAL151, GLY153, ASP154, GLY155, SER156, ASN181, ASN183, GLU184, THR287, LYS289, ASP310, THR313, E315, TYR316, VAL317, PRO318, ALA321, SER323, TRP324, SER325, PRP347, ALA348, MET349, ARG350, GLN351, GLY352, ASP368, ILE371, GLU373, ILE394, TYR395, PHE398, ARG401, ASP422, ARG423, VAL427, ALA429, ASP430, HIS434, PRO479, ARG480, GLY481, and ASN482 are found lining the active site in between the domain1 and domain2 regions. While the domain1 is a parallel β-sheet structure of five strands ranging from residues 1-319, the domain2 is a six-stranded parallel β-sheet structure defined by 176 residues (320-495) [20]. To decipher the most crucial mutations proximal to this site, the structural conservation is estimated for the 63 sequence dataset, by considering and excluding the transketolase sequences against 6OUW through Consurf [46]. While the former shows an average pairwise distance of 0.701 among the sequences, with a lower and upper bound of 1.017e−07 and 2.442 respectively, the latter set shows an average distance of 0.137 within the range of 1.017e−07-1.040. It confirms the evolutionarily significant closeness of our constructed dataset, and the transketolase structures are hereby shown to have a substantial evolutionary divergence from DXS sequences. Further, 114 positions LEU30, ARG38, HIS51, LEU56, VAL59, ALA64, LEU65, ASP74, ASP79, VAL80, HIS82, GLN83, TYR85, HIS87, LYS88, LEU90, THR91, ARG93, GLU114, SER115, ASP118, HIS124, SER126, THR127, SER128, ALA136, ALA138, ILE152, ASP154, THR158, MET161, ALA162, ALA165, ASN167, LEU180-NDNEMS-ILE187, ASN190, VAL191, ALA193, TYR255, ASP260, HIS262, LEU267, PRO280, HIS284, THR287, LYS289, ALA296, GLU297, ASP299, HIS304, SER325, ALA336, THR346, ALA348, MET349, PRO363, ASP368, VAL369, ILE371, ALA372, GLU373, ALA376, THR378, ALA380, ALA384, PRO390, ALA393, TYR395, SER396, THR397, PHE398, LEU399, GLN400, ARG401, TYR403, ASP404, GLN405, HIS408, ASP409, VAL417, ASP422, ARG423, VAL427, ALA429, ASP430, THR433, HIS434, ASP439, ARG444, PRO447, PRO454, ASP456, GLU459, ALA475, ARG477, ARG480, ASN482, LEU527, ARG536, LYS539, PRO540, ASP542, THR558, GLU560, GLU574, PRO591, and ASP592 are found to be completely conserved, 58 constituting domain1 and 46 within domain2, as shown maroon in the color grade panel of Fig. 10b.

Screening the residues interacting with its natively bound ligand 2-acetyl-thiamine diphosphate through the protein-ligand interaction profiler server [70], it shows a salt-bridge, π-stacking, and hydrophobic bond of 4.66 Å and 3.61 Å, 3.59 Å and 3.51 Å for HIS82 and LYS289, PHE398 and VAL80 residues respectively. Moreover, this ligand is found interacting with SER54, GLY123, ALA125, GLY153, GLY155, SER156, and ASN183 residues through hydrogen bond within a distance range of 1.76-3.57 Å. Although Consurf indicates a complete conservation for all these residues, only VAL80, HIS82, LYS289, and PHE398 are found lining the active site by CastP, making an interaction fingerprint. It indicates that our study correctly extracts all the functionally crucial residues across the active site.

The active site of this enzyme is shown to be highly conserved [71], and as shown by CastP, the 61-residue set majorly defines the active site cavity along with the less conserved proximal residues to make it voluminous. It thus becomes reasonable to state that several sequence loci evolutionarily experience a small degree of sequence modifications, and to further substantiate the functionally significant co-evolving residues across the constructed sequence dataset, MISTIC algorithm is used [47]. To discriminate the functionally significant co-evolving residues and the ones having the phylogenetic linkage, the resultant statistical scores are adjusted through the average product correction (APC) [72, 73]. The resultant residue network (Fig. 10c) indicates connections among several residues, and it may emerge due to factors including the phylogenetically preferred substitutions and topological stabilization constraints. Excavating it further, 10 positions ASP74, LEU90, SER128, ASP154, ASN167, ASN190, TYR255, ILE371, THR433, and PRO447 are found to be completely conserved (Fig. 10d). Moreover, only ASP154 and ILE371 are found to be the conserved coevolving residues within the active site. DXS is shown to be highly specific for its substrates [74] and this study will pave way to increase the promiscuity of its active site. The study will certainly be useful to gather data regarding the functionally crucial residues for several other functional studies [75–81]. It would also allow in extracting the evolutionarily closest templates for the protein sequences and would aid to improve the accuracy of the conventional template-based protein modeling protocols [82–90]. The strategy would be significantly useful for improving the algorithmic accuracy of several related research works including protein modeling [91] and folding [92] and functional enzymatic characterization [93–100].

Functional hotspots

Hotspot server [34] is used to localize the hotspot regions within the representative structure 6OUW. Excluding the buried and the correlated loci, THR7, SER8, ASP9, ARG47, THR288, LYS291, PRO318, GLU596, VAL625, and PRO626 appear to be the most flexible residues, and none of these positions are found within the catalytic pocket. Among the 114 completely conserved loci, a correlated hotspot is observed between the two residues, ALA376 and THR378 through an unconserved residue VAL375 to make an interlinked evolutionarily

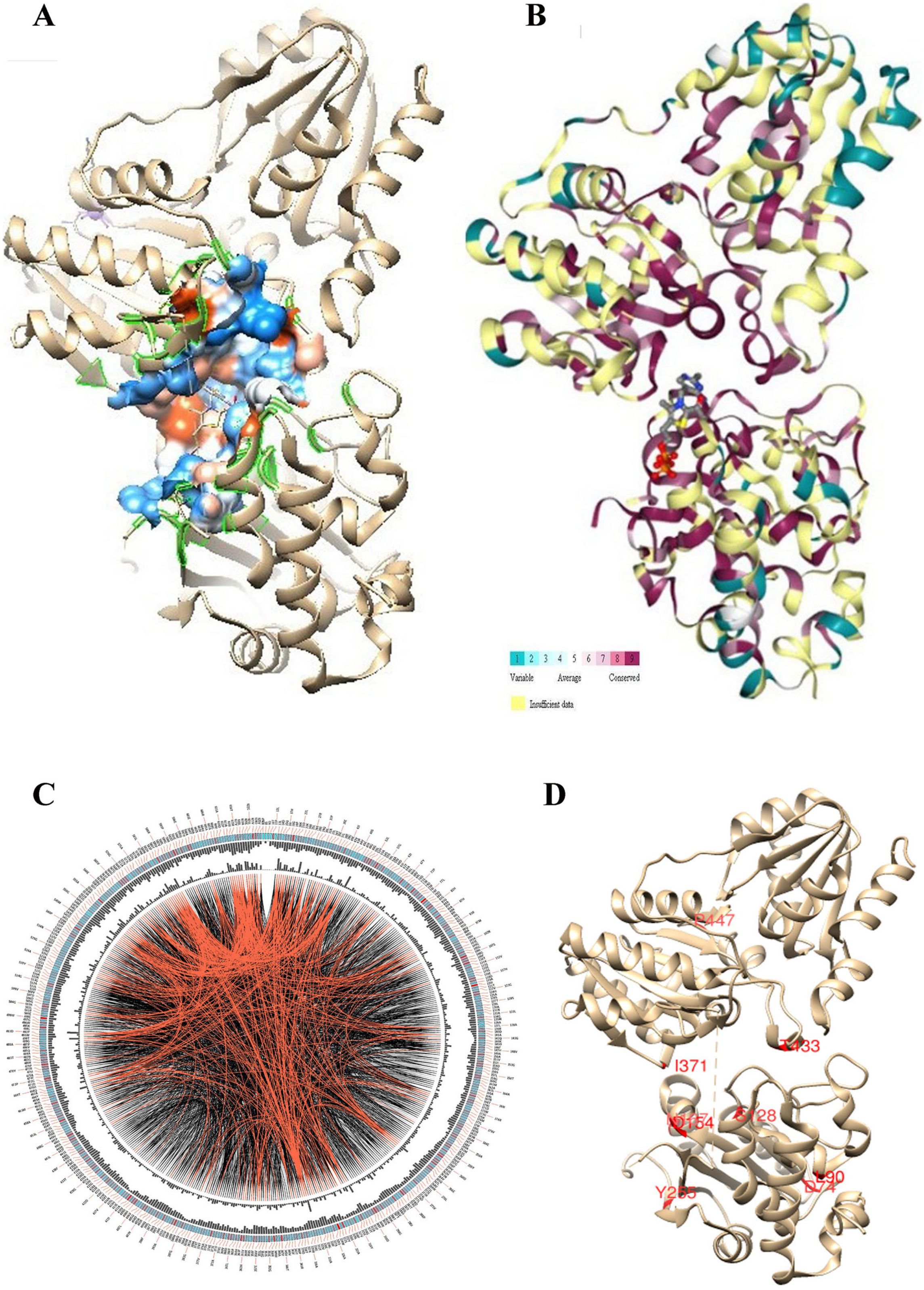


Fig. 10 (See legend on next page.)

(See figure on previous page.)
Fig. 10 Schematic representation of the active site residues of the representative structure 6OUW. **a** Active site protein pocket yielded by CastP [45]. **b** Conservation profile of Consurf [46], maroon being the most conserved residue. **c** MI network [47] of the conserved coevolved residues encoded in 63 DXS sequences. **a** Labels outward of the second circle represent the residue loci, and the colored square boxes indicate the level of sequence conservation within the intensity range of low (blue) to high (red). The two internal circles show proximity mutual information and cumulative mutual information scores respectively. As per MISTIC protocol, the curved central linkers connect residues with statistically significant MI scores (> 6.5), with red, black, and gray orderly indicating the residue pairs with the top 5% scores, average scores between 70 and 95% and lowest scores. **d** Structural localization of the 10 conserved coevolving residues, indicating the most preserved for directed evolution methodologies

correlated triad, and ASP422 and ARG423 are also found to form a correlated set. Although away from the active site, these residues are likely to have an impact on the catalytic activity, and this needs to be experimentally validated for getting further insights.

The PopMuSiC3.1 and MAESTRO servers are used to assess the functional impact of mutations on the representative structure 6OUW. At the first stage, the top-ranked mutations for the 14 identified hotspot residues are evaluated. Besides evaluating the ΔΔG score (Kcal/mol) for the mutations through Popmusic and MAES TRO protocols, the secondary structure and solvent accessibility and confidence score of the residue are orderly estimated through these two algorithms to predict the overall effect of a mutation. As shown by Popmusic, a total of 11 mutations are estimated to have a negative ΔΔG score (*negative score correspond to increased thermodynamic stability*). However, MAESTRO indicates a complete agreement for only 5 of these variations. A value lower than zero indicates the stabilizing mutation and only mutations 7-11 are found to be thermodynamically stable (Table 2). Further, Popmusic is recently proven to be a reliable and robust algorithm [101], and hence strategically deploying the scores on the basis of its scores would be more accurate.

Evaluating the 5 top-ranked mutations 7-11, it is observed that all these positions are completely conserved in the functionally similar sequence profile of the representative structure 6OUW. The mutation D422C shows a ΔΔG score of −1.28 and is predicted to be the most stabilizing mutation by Popmusic. MAESTRO also confirms this prediction and shows the lowest ΔΔG score of −0.667 with a high confidence score of 0.880. However, it is found to be a completely buried β-sheet residue

Table 2 List of predicted top-ranked mutations for the 14 hotspot residues

#	Residues	Top-ranked mutations and scores		Crucial sequence/structural property
		PopMuSiC3.1	MAESTRO	
		ΔΔG (Kcal/mol); secondary structure, solvent accessibility (%)	ΔΔG (Kcal/mol), confidence score	
1	THR7	T7C = −0.11; C, 100	0.718, 0.731	Highly flexible
2	SER8	S8C = 0.07; C, 42.11	0.473, 0.837	Highly flexible
3	ASP9	D9P = −0.1; S, 87.93	0.352, 0.776	Highly flexible
4	ARG47	R47L = −0.76; S, 42.85	0.552, 0.808	Completely conserved and highly flexible
5	THR288	T288C = 0.06; C, 53.6	−0.144, 0.933	Highly flexible
6	LYS291	K291R = −0.1; C, 51.9	0.041, 0.880	Highly flexible
7	PRO318	P318D = −0.16; C, 59.37	−0.115, 0.923	Completely conserved, highly flexible
8	ALA376	A376Y = −0.33; H, 0	−0.489, 0.863	Completely conserved and forms a correlated hotspot with THR378
9	THR378	T378Q = −0.13; H, 34.15	−0.615, 0.868	Completely conserved and forms a correlated hotspot with ALA376
10	ASP422	D422C = −1.28; E, 0	−0.667, 0.880	Completely conserved and forms a correlated hotspot with ARG423
11	ARG423	R423Y = −0.26; S, 18.17	−0.190, 0.904	Completely conserved and forms a correlated hotspot with ASP422
12	GLU596	E596P = −0.49; C, 50.8	0.063, 0.890	Highly flexible
13	VAL625	V625I = 0.45; C, 20.73	−0.058, 0.932	Highly flexible
14	PRO626	P626T = −0.01; C, 96.03	0.200, 0.840	Highly flexible

within the active site, and its mutation may prove to be deleterious. Further, the conserved residues ALA376 and THR378 are found to form an evolutionarily correlated triad through an unconserved residue VAL375, and ASP422 and ARG423 are also found to be correlated loci. With an orderly solvent accessibility score of 0, 34.15, 0, and 18.17, these positions form the helix, helix, strand, and structural bend substructure. Hence, for their crucial role in building the overall conformation, these positions are predicted to be not the best ones for a mutagenesis in vitro. Lastly, MAESTRO shows the highest confidence score of 0.923 for P318D. Popmusic and MAESTRO algorithms show an orderly ΔΔG score of −0.16 and −0.115 for this mutation, and it indicates a satisfactory stabilizing effect. More importantly, although it is an active site position, it is localized over a coil segment with a significantly high solvent accessibility of 59.37. Hence, with the highest confidence score of 0.923, P318D should be the first ideal choice for performing a mutagenesis. Stabilizing this highly flexible coiled residue would thus impart a structural stability to 6OUW, as has been planned several times [102–104].

Conclusion

The study performs a sequence and structure-based analysis of the DXS sequence of *Bacillus subtilis* in comparison to its most prevalent bacterial orthologues. The pipeline used for the study incorporates evolutionary analysis of *B. subtilis* sequences with the other usually encountered bacterial sequences and transketolase. Sequence and structural analysis indicate that only 5 of the 14 identified hotspot positions are completely conserved and 10 positions are highly flexible. Analysis of the top-ranked missense mutations for the 14 hotspots through POPMUSIC and MAESTRO affirm the biological credibility of only 5 mutations, of which, VAL375, ALA376, and THR378 form an evolutionarily correlated triad, and ASP422 and ARG423 are found to be correlated pair. PRO318 is present in the active site and is one of the most flexible residues. The P318D mutation indicates a higher chance of improving the thermostability of DXS. Since DXS is the most crucial enzyme to direct the carbon flux toward the biosynthesis of terpenoids in *B. subtilis*, the present study might be helpful to develop its functionally improved variants for improving the microbial production of terpenoid-based flavoring, fragrance, and therapeutic compounds.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43141-020-00087-x.

Additional file 1: Supplementary Table 1. List of sequences from plant, animal, protista, human, fungi, and bacteria

Additional file 2: Figure S1. Multiple sequence alignment of the constructed dataset against the reference structure 6OUW. The alignment is constructed through ClustalO algorithm and is parsed against 6OUW through Espript3 server. The red shading indicates the sequence conservation.

Abbreviations

GRAS: Generally regarded as safe; DXS: 1-deoxy-D-xylulose-5-phosphate synthase; IUCN: International Union for Conservation of Nature; DXP: 1-deoxy-D-xylulose-5-phosphate; G3P: Glyceraldehyde-3-phosphate; DMAPP: Dimethylallyl diphosphate; IPP: Isopentenyl diphosphate; MEP: Methylerythritol phosphate; MEGA: Molecular evolutionary genetics analysis; ITOL: Interactive tree of life; MEME: Multiple EM for motif elicitation; GRAVY: Grand average of hydrophobicity; APC: Average product correction

Acknowledgements

The authors acknowledge the University and department for providing the required resources/support.

Authors' contributions

AR conceived the study,. AR planned, supervised the experiments, and analyzed the results to compile the article. SCP added substantially to the first introductory passage. THS and VK performed the in-silico experiments. VSR critically read the article and revised the logical flow of the paper. KS lastly screened the article for linguistic errors. All the authors have carefully read and approved the manuscript.

Funding

It is not a funded research.

Availability of data and materials

The constructed files/datasets analyzed in this study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None declared.

Author details

[1]Koneru Lakshmaiah Education Foundation, Guntur 522502, India. [2]School of Biotechnology, Jawaharlal Nehru University, New Delhi, India.

Received: 16 July 2020 Accepted: 21 October 2020
Published online: 26 November 2020

References

1. Vickers CE, Williams TC, Peng B, Cherry J (2017) Recent advances in synthetic biology for engineering isoprenoid production in yeast. Curr Opin Chem Biol 40:47–56. https://doi.org/10.1016/j.cbpa.2017.05.017
2. Christianson DW (2008) Unearthing the roots of the terpenome. Curr Opin Chem Biol 12(2):141–150. https://doi.org/10.1016/j.cbpa.2007.12.008
3. Phulara SC, Pandey S, Jha A, Chauhan PS, Gupta P, Shukla V (2020) Hemiterpene compound, 3, 3-dimethylallyl alcohol promotes longevity and neuroprotection in *Caenorhabditis elegans*. GeroScience:1–17. https://doi.org/10.1007/s11357-020-00241-w
4. Matulja D, Wittine K, Malatesti N, Laclef S, Turks M, Markovic MK, Ambrožić G, Marković D (2020) Marine natural products with high anticancer activities. Curr Med Chem 27(8):1243–1307. https://doi.org/10.2174/0929867327666200113154115
5. Thi HP, Waché Y (2019) Behind the myth of the fruit of heaven, a critical review on gac (*Momordica cochinchinensis* Spreng.) contribution to nutrition. Curr Med Chem 26(24):4585–4605. https://doi.org/10.2174/0929867326666190705154723

6. Zhang LQ, Chen KX, Li YM (2019) Bioactivities of natural catalpol derivatives. Curr Med Chem 26(33):6149–6173. https://doi.org/10.2174/0929867326666190620103813
7. Coricello A, Adams JD, Lien EJ, Nguyen C, Perri F, Williams TJ, Aiello F (2020) A walk in nature: sesquiterpene lactones as multi-target agents involved in inflammatory pathways. Curr Med Chem 27(9):1501–1514. https://doi.org/10.2174/0929867325666180719111123
8. Mierina I, Vilskersts R, Turks M (2020) Delivery systems for birch-bark triterpenoids and their derivatives in anticancer research. Curr Med Chem 27(8):1308–1336. https://doi.org/10.2174/0929867325666180530095657
9. Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc Natl Acad Sci USA 97(24):13172–13177. https://doi.org/10.1073/pnas.240454797
10. Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. Mol Pharm 5(2):167–190. https://doi.org/10.1021/mp700151b
11. Paul A, Bharali S, Khan ML, Tripathi OP (2013) Anthropogenic disturbances led to risk of extinction of *Taxus wallichiana* Zuccarini, an endangered medicinal tree in Arunachal Himalaya. Na Areas J 33(4):447–454. https://doi.org/10.3375/043.033.0408
12. Thomas P, Farjon A (2011) Taxus wallichiana. The IUCN Red List of Threatened Species 2011: e.T46171879A9730085. doi:https://doi.org/10.2305/IUCN.UK.2011-2.RLTS.T46171879A9730085.en
13. Tippmann S, Chen Y, Siewers V, Nielsen J (2013) From flavors and pharmaceuticals to advanced biofuels: production of isoprenoids in Saccharomyces cerevisiae. Biotechnol J 8(12):1435–1444. https://doi.org/10.1002/biot.201300028
14. Zhou K, Zou R, Zhang C, Stephanopoulos G, Too HP (2013) Optimization of amorphadiene synthesis in *Bacillus subtilis* via transcriptional, translational, and media modulation. Biotechnol Bioeng 110(9):2556–2561. https://doi.org/10.1002/bit.24900
15. Rohdich F, Zepeck F, Adam P, Hecht S, Kaiser J, Laupitz R, Gräwert T, Amslinger S, Eisenreich W, Bacher A, Arigoni D (2003) The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. Proc Natl Acad Sci USA 100(4):1586–1591. https://doi.org/10.1073/pnas.0337742100
16. Flores-Pérez Ú, Sauret-Güeto S, Gas E, Jarvis P, Rodríguez-Concepción M (2008) A mutant impaired in the production of plastome-encoded proteins uncovers a mechanism for the homeostasis of isoprenoid biosynthetic enzymes in arabidopsis plastids. Plant Cell 20(5):1303–1315. https://doi.org/10.1105/tpc.108.058768
17. Brown AC, Eberl M, Crick DC, Jomaa H, Parish T (2010) The nonmevalonate pathway of isoprenoid biosynthesis in *Mycobacterium tuberculosis* is essential and transcriptionally regulated by Dxs. J Bacteriol 192(9):2424–2433. https://doi.org/10.1128/JB.01402-09
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242. https://doi.org/10.1093/nar/28.1.235
19. The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47(D1):D506–D515. https://doi.org/10.1093/nar/gky1049
20. Xiang S, Usunow G, Lange G, Busch M, Tong L (2007) Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. J Biol Chem 282(4):2676–2682. https://doi.org/10.1074/jbc.M610235200
21. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
22. Frank A, Groll M (2017) The methylerythritol phosphate pathway to isoprenoids. Chem Rev 117(8):5675–5703. https://doi.org/10.1021/acs.chemrev.6b00537
23. Phulara SC, Chaturvedi P, Chaurasia D, Diwan B, Gupta P (2019) Modulation of culture medium confers high-specificity production of isopentenol in *Bacillus subtilis*. J Biosci Bioeng 127(4):458–464. https://doi.org/10.1016/j.jbiosc.2018.10.002
24. Wagner WP, Nemecek-Marshall M, Fall R (1999) Three distinct phases of isoprene formation during growth and sporulation of Bacillus subtilis. J Bacteriol 181(15):4700–4703
25. Elsholz AKW, Birk MS, Charpentier E, Turgay K (2017) Functional diversity of AAA+ protease complexes in *Bacillus subtilis*. Front Mol Biosci 4:44. https://doi.org/10.3389/fmolb.2017.00044
26. Brammer LA, Meyers CF (2009) Revealing substrate promiscuity of 1-deoxy-d-xylulose 5-phosphate synthase. Organic Lett 11(20):4748–4751. https://doi.org/10.1021/ol901961q
27. Morris F, Vierling R, Boucher L, Bosch J, Freel Meyers CL (2013) DXP synthase-catalyzed C-N bond formation: nitroso substrate specificity studies guide selective inhibitor design. ChemBioChem 14(11):1309–1315. https://doi.org/10.1002/cbic.201300187
28. Uchôa HB, Jorge GE, Silveira NJFD, Camera JC Jr, Canduri F, Azevedo WF Jr (2004) Parmodel: a web server for automated comparative modeling of proteins. Biochem Biophys Res Commun 325(4):1481–1486. https://doi.org/10.1016/j.bbrc.2004.10.192
29. Ferreira GB, Azevedo WF Jr (2019) Homology modeling of protein targets with MODELLER. Methods Mol Biol 2053:231–249. https://doi.org/10.1007/978-1-4939-9752-7_15
30. Silveira NJF, Bonalumi CE, Uchõa HB, Pereira JH, Canduri F, Azevedo WF Jr (2006) DBMODELING: a database applied to the study of protein targets from genome projects. Cell Biochem Biophys 44(3):366–374. https://doi.org/10.1385/cbb:44:3:366
31. Arcuri HA, Zafalon GF, Marucci EA, Bonalumi CE, Silveira N, Machado JM, Azevedo WF Jr, Palma MS (2010) SKPDB: a structural database of shikimate pathway enzymes. BMC Bioinformatics 11:12. https://doi.org/10.1186/1471-2105-11-12
32. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45(D1):D158–D169. https://doi.org/10.1093/nar/gkw1099
33. Lipman DJ, Souvorov A, Koonin EV, Panchenko AR, Tatusova TA (2002) The relationship of protein conservation and sequence length. BMC Evol Biol 2:20. https://doi.org/10.1186/1471-2148-2-20
34. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75
35. Sievers F, Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 27(1):135–145. https://doi.org/10.1002/pro.3290
36. Steinegger M, Soding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35(11):1026–1028. https://doi.org/10.1038/nbt.3988
37. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(2):W369–W373. https://doi.org/10.1093/nar/gkl198
38. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: JMW (ed) The Proteomics Protocols Handbook. Springer, pp 571–607. https://doi.org/10.1385/1-59259-890-0:571
39. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292(2):195–202. https://doi.org/10.1006/jmbi.1999.3091
40. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3):567–580. https://doi.org/10.1006/jmbi.2000.4315
41. Fabian Glaser TP, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19(1):163–164. https://doi.org/10.1093/bioinformatics/19.1.163
42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739. https://doi.org/10.1093/molbev/msr121
43. Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47(W1):W256–W259. https://doi.org/10.1093/nar/gkz239
44. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430(15):2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
45. Tian W, Chen C, Lie X, Zhao J, Liang J (2018) CASTp 3.0: computed atlas of surface topography of proteins and beyond. Nucleic Acids Res 46(W1):W363–W367. https://doi.org/10.1016/j.bpj.2017.11.325

46. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44(W1): W344–W350. https://doi.org/10.1093/nar/gkw408
47. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Buslje CM (2013) MISTIC: mutual information server to infer coevolution. Nucleic Acids Res 41(W1): W8–W14. https://doi.org/10.1093/nar/gkt427
48. Hu LX, Feng JJ, Wu J, Li W, Gningue SM, Yang ZM, Wang Z, Liu Y, Xue ZL (2020) Identification of six important amino acid residues of MenA from *Bacillus subtilis* natto for enzyme activity and formation of menaquinone. Enzyme Microbial Technol 138:109583. https://doi.org/10.1016/j.enzmictec.2020.109583
49. Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 82(2):373–428. https://doi.org/10.1152/physrev.00027.2001
50. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinformatics 12:151. https://doi.org/10.1186/1471-2105-12-151
51. Pucci F, Bernaert K, Teheux F, Gilis D, Rooman M (2015) Symmetry principles in optimization problems: an application to protein stability prediction. IFAC-PapersOnLine 48(1):458–463. https://doi.org/10.1016/j.ifacol.2015.05.068
52. Pucci F, Bernaerts K, Kwasigroch JM, Rooman M (2018) Quantification of biases in predictions of protein stability changes upon mutations. Bioinformatics 34(21):3659–3665. https://doi.org/10.1093/bioinformatics/bty348
53. Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P (2015) MAESTRO--multi agent stability prediction upon point mutations. BMC Bioinformatics 16:116. https://doi.org/10.1186/s12859-015-0548-6
54. Saunders HM, Gilis D, Rooman M, Dehouck Y, Robertson AL, Bottomley SP (2011) Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein. Protein Sci 20(10):1675–1681. https://doi.org/10.1002/pro.698
55. Marabotti A, Scafuri B, Facchiano A (2020) Predicting the stability of mutant proteins by computational approaches: an overview. Briefings Bioinformatics. https://doi.org/10.1093/bib/bbaa074
56. Sakai A, Kinoshita N, Kita M, Katsuragi T, Tani Y (2003) Investigation of 1-deoxy-D-xylulose 5-phosphate synthase and transketolase of *Bacillus subtilis* in relation to vitamin B6 biosynthesis. J Nutr Sci Vitaminol 49(1):73–75. https://doi.org/10.3177/jnsv.49.73
57. Chor B, Tuller T (2005) Maximum likelihood of evolutionary trees is hard. In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner PA, Waterman M (eds) RECOMB 2005:Research in Computational Molecular Biology. 9th Annual International Conference, Cambridge, May 2005, Lecture notes in computer science, vol 3500. Springer, Heidelberg, pp 296–310
58. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851
59. Hasegawa M, Fuziwara M (1993) Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. Mol Phylogenetics Evol 2(1):1–5. https://doi.org/10.1006/mpev.1993.1001
60. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704. https://doi.org/10.1080/10635150390235520
61. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Applications BioSciences 8(3):275–282. https://doi.org/10.1093/bioinformatics/8.3.275
62. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x
63. Burra PV, Kalmar L, Tompa P (2010) Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. PLoS ONE 5(8):e12069. https://doi.org/10.1371/journal.pone.0012069
64. Schenk G, Duggleby RG, Nixon PF (1998) Properties and functions of the thiamin diphosphate dependent enzyme transketolase. Int J Biochem Cell Biol 30(12):1297–1318. https://doi.org/10.1016/s1357-2725(98)00095-8
65. Hess BM, Xue J, Markillie LM, Taylor RC, Wiley HS, Ahring BK, Linggi B (2013) Coregulation of terpenoid pathway genes and prediction of isoprene production in *Bacillus subtilis* using transcriptomics. PLoS One 8(6):e66104. https://doi.org/10.1371/journal.pone.0066104
66. Ravishankar VV, Fu Y, Bennett GN, San K (2005) Enhanced lycopene productivity by manipulation of carbon flow to isopentenyl diphosphate in *Escherichia coli*. Biotechnol Progress 21(5):1558–1561. https://doi.org/10.1021/bp050124l
67. Heider SA, Wolf N, Hofemeier A, Peters-Wendisch P, Wendisch VF (2014) Optimization of the IPP precursor supply for the production of lycopene, decaprenoxanthin and astaxanthin by *Corynebacterium glutamicum*. Front Bioeng Biotechnol 2:28. https://doi.org/10.3389/fbioe.2014.00028
68. Liu H, Wang Y, Tang Q, Kong W, Chung W, Lu T (2014) MEP pathway-mediated isopentenol production in metabolically engineered *Escherichia coli*. Microbial Cell Factories 13:135. https://doi.org/10.1186/s12934-014-0135-y
69. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42(W1):W320–W324. https://doi.org/10.1093/nar/gku316
70. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M (2015) PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res 43(W1): W443–W447. https://doi.org/10.1093/nar/gkv315
71. Saravanan T, Junker S, Kickstein M, Hein S, Link MK, Ranglack J, Witt S, Lorilliere M, Hecquet L, Fessner WD (2017) Donor promiscuity of a thermostable transketolase by directed evolution: efficient complementation of 1-deoxy-d-xylulose-5-phosphate synthase activity. Angewandte Chemie 56(19):5358–5362. https://doi.org/10.1002/anie.201701169
72. Dunn SD, Wahl LM, Gloor GB (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics 24(3):333–340. https://doi.org/10.1093/bioinformatics/btm604
73. Baker FN, Porollo A (2016) CoeViz: a web-based tool for coevolution analysis of protein residues. BMC Bioinformatics 17(1). https://doi.org/10.1186/s12859-016-0975-z
74. Schürmann M, Schürmann M, Sprenger GA (2002) Fructose 6-phosphate aldolase and 1-deoxy-d-xylulose 5-phosphate synthase from *Escherichia coli* as tools in enzymatic synthesis of 1-deoxysugars. J Mol Catalysis B: Enzymatic 19–20:247–252. https://doi.org/10.1016/S1381-1177(02)00174-1
75. Amrein BA, Runthala A, Kamerlin SCL (2019) In-silico-directed evolution using CADEE. Comput Methods Protein Evol Methods Mol Biol 1851:381–415. https://doi.org/10.1007/978-1-4939-8736-8_22
76. Jamithireddy AK, Runthala A, Gopal B (2020) Evaluation of specificity determinants in *Mycobacterium tuberculosis*/anti-factor interactions. Biochem Biophys Res Commun 521(4):900–906. https://doi.org/10.1016/j.bbrc.2019.10.198
77. Kamjula V, Kanneganti A, Metla R, Nidamanuri K, Idupulapati S, Runthala A (2020) Decoding the vital segments in human ATP-dependent RNA helicase. Bioinformation 16(2):160–170. https://doi.org/10.6026/97320630016160
78. Lal M, Saumya KU, Mahala N, Runthala A, Dubey US (2020) Camel milk α-lactalbumin as a potential anticancer molecule: a bioinformatics analysis. Res J Pharm Biol Chem Sci 11(3):38–52. https://doi.org/10.33887/rjpbcs/2020.11.3.5
79. Runthala A, Singh AK (2010) Tegument based in-silico drug targeting of herpes simplex virus-1. Saratov J Med Sci Res 6(2):353–357
80. Phulara SC, Rajput VS, Mazumdar B, Runthala A (2020) Metabolic and enzyme engineering for the microbial production of anticancer terpenoids. Essentials of Cancer Genomic, Computational Approaches and Precision Medicine, In, pp 237–259. https://doi.org/10.1007/978-981-15-1067-0_10
81. Sabharwal NS, Runthala A (2014) Functional protein domains evolve very specifically over mutations. J Proteomics Genomics 1(1). https://doi.org/10.15744/2576-7690.1.102
82. Runthala A (2012) Protein structure prediction: challenging targets for CASP10. J Biomol Structure Dynamics 30(5):607–615. https://doi.org/10.1080/07391102.2012.687526
83. Runthala A, Chowdhury S (2013) Protein structure prediction: are we there yet? In: Pham TD, Jain LC (eds) Knowledge-Based Systems in Biomedicine. Springer, pp 79–115. https://doi.org/10.1007/978-3-642-33015-5_4
84. Runthala A, Chowdhury S (2014) Iterative optimal TM_score and Z_score guided sampling significantly improves model topology. In: International MultiConference of Engineers and Computer Scientists, Hong Kong, Proceedings of the International MultiConference of Engineers and Computer Scientists, vol 1. IAENG, Hong Kong, pp 123–128
85. Runthala A (2015) Non-linear and misleading template scoring criteria: root cause of protein modelling inaccuracies. Curr Synthetic Syst Biol 03(02). https://doi.org/10.4172/2332-0737.1000121

86. Runthala A, Chowdhury S (2016) Unsolved problems of ambient computationally intelligent TBM algorithms. In: Bhattacharyya S, Dutta P, Chakraborty S (eds) Hybrid Soft Computing Approaches. Springer, pp 75–105. https://doi.org/10.1007/978-81-322-2544-7_3
87. Garg S, Kakkar S, Runthala A (2016) Improved protein model ranking through topological assessment. In: Wong K-C (ed) Computational Biology and Bioinformatics: Gene Regulation. CRC Press, pp 406-424. doi:https://doi.org/10.1201/b20026-24
88. Nahata S, Runthala A (2016) Quick reliable exploration of the PDB universe seeks a new template search algorithm. J Data Mining Genomics Proteomics 7(4):4. https://doi.org/10.4172/2153-0602.1000206
89. Runthala A, Chowdhury S (2019) Refined template selection and combination algorithm significantly improves template-based modeling accuracy. J Bioinform Comput Biol 17(2):1950006. https://doi.org/10.1142/S0219720019500069
90. Runthala A (2020) Probabilistic divergence of a TBM methodology from the ideal protocol. BiorXiV. https://doi.org/10.1101/2020.07.05.160937
91. Satyanarayana SDV, Krishna MSR, Pavan Kumar P, Jeereddy S (2018) In silico structural homology modeling of nif A protein of rhizobial strains in selective legume plants. J Genetic Eng Biotechnol 16(2):731–737. https://doi.org/10.1016/j.jgeb.2018.06.006
92. Kante RK, Vemula S, Mallu MR, Ronda SR (2018) Efficient and easily scalable protein folding strong anion exchange chromatography for renaturation and simultaneous purification of recombinant human asparaginase from *E. coli*. Biotechnol Progress 34(4):1036–1044. https://doi.org/10.1002/btpr.2649
93. Sahitya UL, Krishna MSR, Suneetha P (2019) Integrated approaches to study the drought tolerance mechanism in hot pepper (*Capsicum annuum* L.). Physiol Mol Biol Plants 25(3):637–647. https://doi.org/10.1007/s12298-019-00655-7
94. Kalyani BS, Krishna PS, Sreenivasulu K (2019) Screening and identification of novel isolate Streptomyces sp., NLKPB45 from Nellore costal region for its biomedical applications. Saudi J Biol Sci 26(7):1655–1660. https://doi.org/10.1016/j.sjbs.2018.08.027
95. Kante RK, Somavarapu S, Vemula S, Kethineni C, Mallu MR, Ronda SR (2019) Production of recombinant human asparaginase from *Escherichia coli* under optimized fermentation conditions: effect of physicochemical properties on enzyme activity. Biotechnol Bioprocess Eng 24(5):824–832. https://doi.org/10.1007/s12257-019-0147-x
96. Managamuri U, Vijayalakshmi M, Indupalli MD, Ganduri VSRK, Rajulapati SB, Poda S (2018) Improved bioactive metabolite production by *Saccharopolyspora halotolerans* VSM-2 using response surface methodology and unstructured kinetic modelling. Pharmacognosy J 10(5):833–840. https://doi.org/10.5530/pj.2018.5.142
97. Samara Shekar Reddy S, Singh B, Peter AJ, Venkateswar Rao T (2018) Production of transgenic local rice cultivars (Oryza sativa L.) for improved drought tolerance using Agrobacterium mediated transformation. Saudi J Biol Sci 25(8):1535–1545. https://doi.org/10.1016/j.sjbs.2016.01.035
98. Kante RK, Vemula S, Somavarapu S, Mallu MR, Boje Gowd BH, Ronda SR (2018) Optimized upstream and downstream process conditions for the improved production of recombinant human asparaginase (rhASP) from Escherichia coli and its characterization. Biologicals 56:45–53. https://doi.org/10.1016/j.biologicals.2018.10.002
99. Somavarapu S, Vemula S, Reddy IB (2018) Extraction, purification and characterization of a novel cysteine protease from the latex of plant Vallaris solanacea. J Plant Biochem Biotechnol 27(2):186–198. https://doi.org/10.1007/s13562-017-0429-3
100. Sailaja E, Bhavani S, Rambabu D, Basaveswara Rao MV, Pal M (2019) A greener approach toward N-1 heteroarylation of indoles: synthesis and in vitro evaluation of potential anti-proliferative agents. Arabian J Chem 12(8):3667–3677. https://doi.org/10.1016/j.arabjc.2015.11.008
101. Kumar V, Rahman S, Choudhry H, Zamzami MA, Jamal MS, Islam A, Ahmad F, Hassan MI (2017) Sci Rep 7:4678. https://doi.org/10.1038/s41598-017-04950-9
102. Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angewandte Chemie 45:7745–7751
103. Gao S, Zhu S, Huang R, Li H, Wang H, Zheng G (2018) Engineering the enantioselectivity and thermostability of a (+)-γ-lactamase from Microbacterium hydrocarbonoxydans for kinetic resolution of Vince lactam (2-azabicyclo[2.2.1]hept-5-en-3-one). Appl Environ Microbiol 84:e01780–e01717. https://doi.org/10.1128/AEM.01780-17
104. Jochens H, Aerts D, Bornscheuer UT (2010) Thermostabilization of an esterase by alignment guided focussed directed evolution, Protein Engineering. Design Selection 23(12):903–909. https://doi.org/10.1093/protein/gzq071

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.