Virtual screening and docking analysis of novel ligands for selective enhancement of tea (*Camellia sinensis*) flavonoids

Anusha Majumder [a,1], Sunil Kanti Mondal [b,1], Samyabrata Mukhoty [b], Sagar Bag [a], Anupam Mondal [a], Yasmin Begum [c,d], Kalpna Sharma [e], Avishek Banik [a,*]

[a] *Laboratory of Microbial Interaction, School of Biotechnology, Presidency University, Kolkata, West Bengal, India*
[b] *Department of Biotechnology, The University of Burdwan, Burdwan, West Bengal, India*
[c] *Department of Biophysics, Molecular Biology, and Bioinformatics, University of Calcutta, 92, APC Road, Kolkata 700009, West Bengal, India*
[d] *Center of Excellence in Systems Biology and Biomedical Engineering (TEQIP Phase-III), University of Calcutta, JD-2, Sector III, Salt Lake, Kolkata 700106, West Bengal, India*
[e] *R&D Centre, Danguajhar Tea Garden, Goodricke Group Ltd., Jalpaiguri, West Bengal, India*

ARTICLE INFO

ABSTRACT

Flavour of tea is mainly contributed by a group of polyphenols – flavonoids. However, the content of flavonoid fluctuates seasonally and is found to be higher in the first flush of tea, when compared to the second flush. This disparity in the flavonoid content, and hence taste, incurs heavy economic losses to the tea plantation industry each harvest season. For our present study, four key product-specific enzymes (PAL, FNS, FLS and ANS) of the tea-specific flavonoid pathway were selected to perform molecular docking studies with specific virtually screened allosteric modulators. Results of docking analyses showed Naringenin, 2-Morpholin-4-ium-4-ylethane-sulfonate, 6-C-Glucosylquercetin, 2-Oxoglutaric acid, 3,5,7,3′,4′-pentahydroxyflavone to be capable of improving the spontaneity of the enzyme-substrate reactions in terms of docking score, RMSD values, and non-covalent interactions (H-bond,hydrophobic interaction, Π-stacking, salt bridge, etc.). Further, the evolutionary relationship of tea flavonoid pathway enzymes was constructed and compared with related taxa. The codon usage-based of tea flavonoid biosynthetic genes indicated the non-biasness of their nucleotide composition. Overall this study will provide a direction towards putative ligand-dependent enhancement of flavonoid content, irrespective of seasonal variation.

1. Introduction

The history of tea can be traced back to ancient China, where it was initially used as a medicinal herb. Gradually, during its domestication process, the organoleptic properties of tea were further enhanced, leading it to rank second worldwide, in beverage consumption after water (Zeng et al., 2019). Naturally, tea is now one of the many popular economic crops in Asian, African, and Latin American countries to have

a broadened industry, thanks to its pleasant taste and many health benefits. Processed by the tender shoots or leaves of the plant, tea can be of various types with difference in their processing or in harvesting times; ranging from black tea, the oolong tea to green and white tea, with an annual worldwide production of approximately 78% black, 2% oolong, and 20% green tea leaves (McKay and Blumberg, 2002).

The tea plant cycles through several periods of growth and dormancy and undergoes four main blooms- first, second, monsoon, and autumn

flush. The period when tea plants start growing new leaves to be harvested is called a "flush". Growth season begins when the tea plants germinate after being dormant over the winter, thus leading to a spring harvest and completing a flush. The terms 'First Flush' and 'Second Flush' are commonly associated with Darjeeling tea (West Bengal, India); however, 'flush' can often vary depending on the location and weather patterns of the tea-growing region. As the growing season for the tea plant progresses, each additional flush yields different flavour and aroma characteristics. All-in-all, the differences between tea flushes lie in their harvest season, the way the leaves are processed, and their characteristic flavours. The first flush (or the Spring flush) is a light infusion with a pale light yellow to red appearance. Considered to be the best flush of the growing season, it grows immediately after the dormancy period and is the first plucking of a tea plant's harvest season. The start of harvest season can vary from late February to early March and continue through mid-April, depending on the location. The youngest and the most tender leaves are plucked along with the new shoots on the tea bushes, in a bud-and-two-leaves combination. Although their shelf life is not very long, the Spring flush is said to be a storehouse of the highest amounts of nutrients, catechins, L-theanine stimulant, and caffeine, and is known for its flowery aroma, fine flavour, and delicate taste with slight astringency. As soon as the first flush ends, rain pours in, allowing the plants to soak up water and thus, entering a brief dormant stage before the next cycle of growth. After the brief dormancy post-first flush, second or summer flush follows. During this period, tea leaves generally have less time to grow and leaf growth is much more rapid than the early spring growth. The leaves with an amber hue, when picked between May to June, are larger, well-rounded, fruity, and mature flavoured tea leaves. The bright liquid has a muscatel character and is known for its distinctive taste, medium to heavy aroma, and delicate flavour profiles.

The diverse secondary metabolites of tea are predominantly bioactive compounds that are important parameters in determining the quality of tea, for example, theanine, caffeine, theobromine, inorganic salts, etc. Broadly classified into three classes, these secondary metabolites can be phenolic compounds (18%–36%), amino acids (1%–4%), and volatile aroma compounds (less than 0.03%). The aroma is mainly contributed by the>600 volatile compounds (Ho et al., 2015), whereas taste is contributed by phenolic compounds and amino acids. Catechins, a flavonol group of polyphenols, is highly enriched in fresh tea leaves. An active polyphenol oxidase present in tea leaves promotes the aerobic oxidation of these catechins, which in turn enhances the production of several other related compounds, including bis flavanols, theaflavins, *epi*-theaflavic acids, and thearubigens, when black tea is being produced. The polyphenol group not only includes the catechins, but also harbours several other members, such as the flavanols, their glycosides, and the three types of flavonoids. (Wang et al., 2019; Bag et al., 2022a). That oxidative stress and reactive oxygen species trigger the formation of flavonoids in plants, explains why flavonoids can serve the dual purpose of inhibiting ROS production and provide protection to resist damage caused by UV radiation. Additionally, it also provides protection against phytopathogens, play a role in signalling during nodulation, regulates formation of colour in flowers, transport phytohormones (e.g. auxin), as well as plays role in male fertility (Bradshaw and Schemske, 2003). Tea polyphenols have also been proposed to possess several medicinal properties. Recent computational studies involving molecular docking and simulations reported several bioactive molecules of tea to be potential inhibitors of various targets of SARS-CoV-2; for example, oolonghomobisflavan-A against the main protease; barrigenol, kaempferol, and myricetin against non-structural protein 15; theaflavin against the non-structural protein 16; epicatechin-3,5-di-O-gallate, epigallocatechin-3,5-di-O-gallate, and epigallocatechin-3,4-di-O-gallate against the enzyme RNA-dependent RNA polymerase (Bhardwaj et al., 2021; Sharma et al., 2021)

Flavonoids, rightly depicted as C6-C3-C6, comprise of two benzene rings, named as ring A and ring B, which are connected by a 3-carbon

ring C. Biosynthesis of flavonoid compounds in plants involves two different pathways: the acetate pathway and the shikimate pathway. The acetate pathway lays the foundation for the polymeric 2-carbon units, with the glucose transformations-generated three molecules malonyl-CoA, forming the ring A. The shikimate pathway forms the phenylpropanoids (C6-C3) skeleton, i.e., ring B and ring C, with ring B being synthesized from phenylalanine via 4-coumaroyl-CoA. Ultimately, condensation of the two main rings of the flavonoid structure follows, thus leading to the formation of chalcone. It is to be noted, however, that chalcones are devoid of the ring C. Based on its distinct features the flavonoid biosynthetic pathway can be segregated as "early" and the "late" phases. The early flavonoid pathway spans from the first enzyme of the pathway PAL, till the conversion of flavanones to either flavones (by FNS) or dihydroflavonols (by F3H). Divergence of dihydroflavonols into either flavonols (by FLS) or leucoanthocyanidins (by DFR) marks the beginning of the late flavonoid pathway (Figs. 1 and S1).

Thus the present study aims to virtually screen and perform docking analysis of novel ligands which can form a complex with the enzymes of the tea flavonoid biosynthetic pathway (de Sousa Luisa et al., 2020). Further, we are hypothesizing that when the enzyme-ligand complex reacts with the substrate, improvement of tea quality can be expected, in terms of induced enzyme-substrate- ligand interaction, resulting in enhancement of flavour (flavones, flavonols, catechins, epicatechins, and anthocyanins) in the second flush of Darjeeling tea.

2. Materials and methods

2.1. Retrieval of tea specific flavonoid pathway and selection of the flavonoid specific key enzymes from KEGG

The biosynthetic pathways leading to the formation of the most common types of tea flavonoid compounds were constructed from Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa and Goto, 2000). The first three steps of the "early" flavonoid pathway were constructed from the Phenylpropanoid biosynthesis pathway (00940 M). The pathway module "flavanone biosynthesis (M00137)" was specifically selected from "biosynthesis of other secondary metabolites", under the metabolism tab. While for the remaining steps, the Flavonoid biosynthesis pathway (00941 M) was referred to, and both the pathway modules- Flavanone biosynthesis (M00137) and Flavonoid biosynthesis (M00138) were selected to deduce the tea specific pathways. Routes for production of further downstream compounds (flavanone and flavonoid), for example, of anthocyanin derivatives and flavones and flavonols, from their immediate precursors can be derived from Anthocyanin biosynthesis (00942) and Flavone and Flavonol biosynthesis (00944) pathways, respectively. All of these biosynthetic pathways are present under the Biosynthesis of other secondary metabolites (1.10) section of Metabolism. For the enzyme required at each step of the pathway, the corresponding enzyme E.C. number mentioned in the respective biosynthetic pathways was referred to (Fig. 1).

2.2. Clustering of a similar group of compounds through ChemMine

ChemmineR online tool was used for clustering the physicochemical properties of the selected phytochemicals. The properties clustered includes molecular formula (MF), molecular weight (MW), *N*-charges, Carbon (C), Hydrogen (H), Nitrogen (N), Oxygen (O), Sulphur (S), Fluorine (F), Chlorine (Cl), Bromine (Br), Iodine (I), Phosphorus (P), primary amines (RNH2), secondary amines (R2NH), tertiary amines (R3N), ROPO3, alcohol (ROH), aldehyde (RCHO), ketone (RCOR), carboxylic acid (RCOOH), ester (RCOOR), ether (ROR), alkyne (RCCH), nitrile (RCN), RINGS, AROMATIC. Numeric Data Clustering clusters were performed using the Euclidean distance between column z-scores and uploaded and selected for the study. While all clustering was done using normalized data (z-scores) the "Color and Display Values" option lets us display the true numeric values on the final plot. Method Linkage:
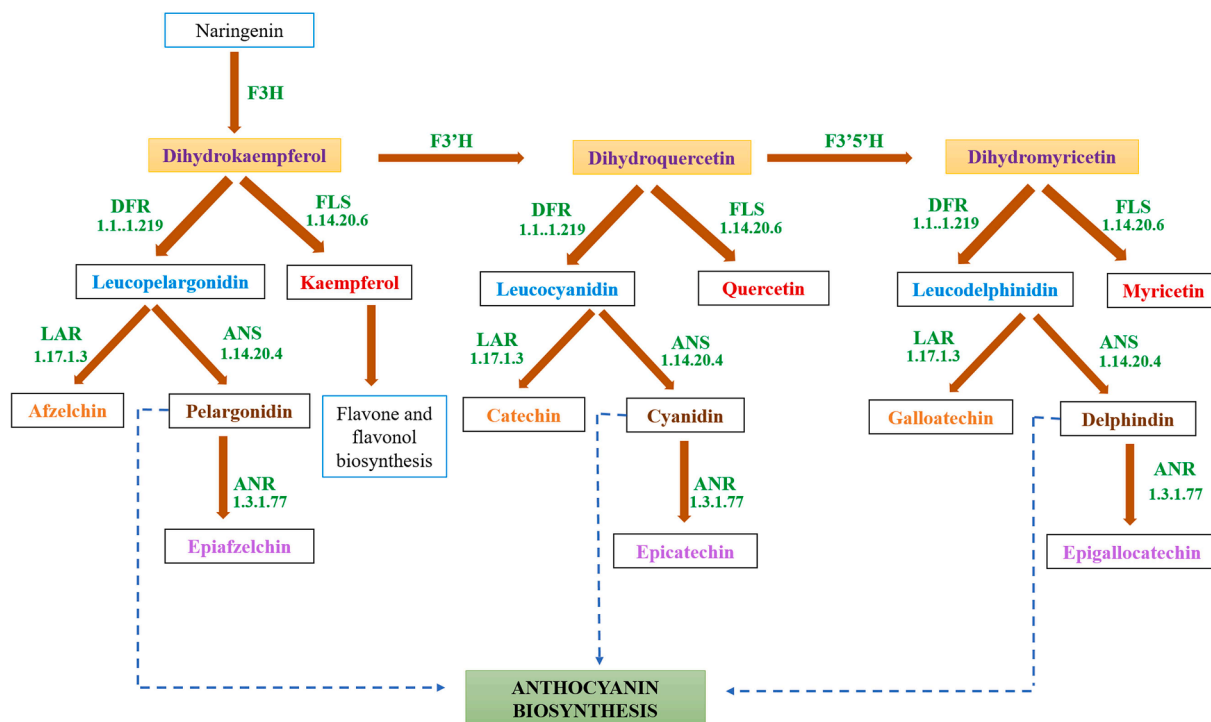
Fig. 1. Schematic representation of team-specific flavonoid biosynthetic pathway. All the enzymes acronyms are followed by its EC number. Orange boxes- Dihydroflavonols, Blue- Leucoanthocyanins, Red- Flavonols, Orange- Flavan-3-ols/Catechins, Brown- Cyanidins, Purple- Epicatechins.

single; Physicochemical properties: ChemmineR Properties; Color and display values: Z score (Backman et al., 2011).

2.3. Retrieval of enzyme sequences (aa) from UniProt database

A total of sixteen key enzymes were identified from the biosynthesis pathways and their amino acid sequences were searched in the UniProt database. Sequences specific to *Camellia sinensis* were filtered using advanced search of the UniProt database, followed by BLASTP search (Non-Redundant Protein Database) (Table 1).

2.4. Evolutionary relationships of enzymes associated with tea flavonoid biosynthesis pathway

The fifteen enzyme sequences (Table 1) were used to derive the evolutionary history of the tea flavonoid biosynthesis pathway, by the Neighbor-Joining (NJ) method. The phylogenetic tree was represented as a percentage of replicates with a bootstrap value of 1000 and drawn with MEGA X. Poisson correction method was implemented to calculate the evolutionary distances among 15 amino acid sequences. Ambiguous amino acid positions were obliterated using pairwise deletion with a cumulative 1875 positions in the final dataset (Banik et al., 2019).

2.5. Species phylogeny

Based on their sequential occurrence, all fifteen tea-specific enzymes (PAL, C4H, 4CL, CHS, CHI, FNS, F3H, F3′H, F3′5′H, DFR, FLS, UGT72, LAR, ANS, and ANR) sequences were selected from the enzymes of tea flavonoids biosynthesis pathway to study the evolutionary relationship of tea flavonoid pathway enzymes (Table 1). Each enzyme sequence was subjected to BLASTP (protein BLAST) using the NCBI protein sequence database of non-redundant protein sequences, excluding uncultured sample and model sequences databases. After BLASTP of each sequence, description tables of 1000 sequences were collected from each search result. For fifteen enzyme sequences, a total dataset of 15,000 sequences was collected and was searched for the common organisms in each

enzyme group. Multiple sequence alignment of the protein sequences of the enzymes formed the basis for the construction of the phylogenetic tree, which was built by the Maximum-Likelihood method under the Jones-Taylor-Thornton (JTT) substitution model, using the MEGA X program. For the prediction of the species tree, phylogenetic analysis was conducted on concatenated gene sequences. The concatenated gene sequences were obtained by joining head-to-tail in-frame, i.e., the gene sequences of the enzymes of the pathway were arranged according to their sequential occurrence. This process was followed for all the common organisms as obtained previously from BLASTP (Begum and Mondal, 2021).

2.6. Homology modelling and verification of the models built:

The three-dimensional (3D) structures of the enzymes were predicted by homology modelling using a knowledge-based, user-friendly and free method- SWISS-MODEL ExPASY (Schwede et al., 2003). The choice of using SWISS-MODEL over the other homology modelling packages was based on the fact that it is a fully automated homology modelling server, which produces a complete model by sequentially following all steps of homology modelling process. It reads the model sequence, performs sequence alignment, builds a complete backbone with side chain and loop building, verifies the quality of the model built (in terms of GMQE and QMEAN), and also optimally minimizes the model using the Gromos96 force field (Van Gunsteren et al., 1996). Moreover, almost all homology modelling programs have been seen to produce reasonably satisfactory models when sequence identities of the target and template sequences are > 30%. Thus, the models built of the enzymes from SWISS-MODEL, were used for further analysis. Moreover, a sequence identity > 50% have been reported to perform docking analysis as good as crystal structures, with a cut off of 40% sequence identity being required to obtain reliable results (Schafferhans and Klebe, 2001). The predicted 3D structures were then verified from MolProbity which provides an evaluation of model quality, both globally and locally.

Table 1
UniProt sequence ID, SWISS-MODEL (SM) template details and predicated ligand details from ProBis.

Enzyme	Full name	Tea Specific UniprotKB ID	Organism	SM Template	SM Organism of template	SM GMQE	SM QMEAN	SM Seq Id (%)	SM Oligo State	PDB CI	SM Ligands (SMR)	Predicted ligands (ProBis) at binding site 1	Selected ligands used for docking
PAL	Phenylalanine ammonia-lyase	P45726 (PALY_CAMSI)	*Camellia sinensis*	1w27.1	*Petroselinum crispum*	0.85	−1.58	88.03	homo-tetramer	ABAB	4 × DTT	4-carboxycinnamic acid (4.12)	4- carboxycinnnamic acid, 2,3-Dihydroxy-1,4-dithiobutane
C4H	Trans-cinnamate 4-monooxygenase	Q6DV44 (Q6DV44_CAMSI))	*Camellia sinensis*	6vby.1.A	*Sorghum bicolor*	0.89	−0.38	76.51%	monomer	A	1 × EPE, 1 × HEM, 2 × GOL	Protoporphyrin ix containing fe, (3alpha,8alpha)-17-(1 h-benzimidazol-1-yl) androsta-5,16-dien-3-ol, Abiraterone, Progesterone, (9beta)-17-hydroxypregn-4-ene-3,20- dione, (3beta)-3-hydroxypregn-5-en-20-one (3.45)	
PTAL	Phenylalanine/tyrosine ammonia-lyase	Q8VXG7 (PALY_MAIZE)	*Zea mays*	6f6t.1.C	*Petroselinum crispum*	–	−1.07	67.1	homo-tetramer	ABAB	4 × CV2	2,3-dihydroxy-1,4-dithiobutane (5.1), 4-car-boxycinnamic acid (4.12)	
4CL	4-coumarate–CoA ligase	A0A4S4EIG7 (A0A4S4EIG7_CAMSI)	*Camellia sinensis*	2bac.1.A	*Nicotiana tabacum*	0.06	−0.36	62.92	monomer	A	1 × CoA, 1 × AMP, 1 × GOL	5′-o-{(s)-hydroxy[3-(4-*hydroxyphenyl) propoxy]* phosphoryl} adenosine, Adenosine monophosphate, 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (4.79)	
CHS	Chalcone synthase	P48386 (CHS1_CAMSI)	*Camellia sinensis*	5uc5.1.A	*Malus domestica*	0.93	0.02	84.02/ 91.75	homo-dimer	AA/ AB	None	Naringenin (4.8), Piperazine-n,n'-bis(2-ethanesulfonic acid) (4.8), Resveratrol (4.8)	
CHI	Chalcone isomerase	Q45QI7 (CFI_CAMSI)	*Camellia sinensis*	5yx4.1.A	*Deschampsia Antarctica*	0.84	−0.38	64.73	monomer	A	1 × HCC		
FNS	Flavone synthase	V5RBK9 (V5RBK9_CAMSI)	*Camellia sinensis*	5ylw.1	*Salvia miltiorrhiza*	0.63	−2.6	33.03	monomer	A	1 × HEM; 1 × Mn (II) ion	Protoporphyrin ix containing fe (3.3), 1,2-ethanediol (3.3), S,r *meso*-tartaric acid (3.3)	Mesotartaric acid, Protoporphyrin IX Containing Fe
F3H	Flavanone 3-hydroxylase	Q6DV45 (Q6DV45_CAMSI)	*Camellia sinensis*	1gp4.1.A	*Arabidopsis thaliana*	0.6	−3.08	30.7	monomer	A	1 × AKG; 1 × MES		
F3′H	Flavonoid 3′-hydroxylase	A0A0N9E123 (A0A0N9E123_CAMSI)	*Camellia sinensis*	5ylw.1.A	*Salvia miltiorrhiza*	0.67	−2.21	34.89	monomer	A	1 × HEM; 1 × Mn (II) ion	Protoporphyrin ix containing fe (3.7), (4-hydroxy-3,5-dimethylphenyl)(2-methyl-1-benzofuran-3-yl)methanone (3.7)	
F3′5′H	Flavonoid 3′5′-hydroxylase	A0A1S5T865 (A0A1S5T865_CAMSI)	*Camellia sinensis*	5ylw.1	*Salvia miltiorrhiza*	0.66	−3.23	33.48	monomer	A	1 × HEM; 1 × Mn (II) ion	Protoporphyrin ix containing fe (3.28), (4-hydroxy-3,5-dimethylphenyl)(2-methyl-1-benzofuran-3-yl)methanone (3.28)	
DFR	Dihydroflavonol 4-reductase	Q9S787 (Q9S787_CAMSI)	*Camellia sinensis*	3bxx.1.A; Subunits: 3bxx.1,	Grape	0.86	0.94	78.21	homo-dimer	AD, BC, EF	(a) 2 × NADP; 3 × QUE;		

(*continued on next page*)

Table 1 (*continued*)

Enzyme	Full name	Tea Specific UniprotKB ID	Organism	SM Template	SM Organism of template	SM GMQE	SM QMEAN	SM Seq Id (%)	SM Oligo State	PDB CI	SM Ligands (SMR)	Predicted ligands (ProBis) at binding site 1	Selected ligands used for docking
				3bxx.2, 3bxx.3							(b) 2 × NADP; 4 × QUE; (c) 2 × NADP; 2 × QUE;		
FLS	Flavonol synthase	A0A3G1RNB1 (A0A3G1RNB1_CAMSI)	*Camellia sinensis*	1gp4.1.A	*Arabidopsis thaliana*	0.83	−1.13	45.26	monomer	A	1 × AKG; 1 × MES	2-oxoglutaric acid (3.84), (2r,3r)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4 h-chromen-4-one (3.84), Succinic acid (3.84), Naringenin (3.84), 3,5,7,3′,4′-pentahydroxyflavone (3.84), 2-(n-morpholino)-ethanesulfonic acid (3.84)	2-ketoglutaric acid, 2-(N-Morpholino) ethanesulfonic acid, Naringenin, Succinic acid, Dihydroquercetin, 2-Morpholin-4-ium-4-ylethanesulfonate, 6-C-Glucosylquercetin
UGT72	UDP-glycosyltransferases	A0A0X9GJZ6 (A0A0X9GJZ6_CAMSI)	*Camellia sinensis*	6jel.1, 6jel.2	*Phytolacca americana*	0.82	−1.5	57.36	monomer	A, B	None	Uridine-5′-diphosphate-glucose (3.79), Uridine-5′-diphosphate (3.37), Glycerol (3.21)	
LAR	Leucoanthocyanidin reductase	W6EL68 (W6EL68_CAMSI)	*Camellia sinensis*	3i52.1.A	*Vitis vinifera*	0.82	−0.61	70.29	monomer	A	1 × NADP; 1 × KXN	Nadp nicotinamide-adenine-dinucleotide phosphate (4.21)	
ANS	Anthocyanidin synthase	A0A286QXW6 (A0A286QXW6_CAMSI)	*Camellia sinensis*	1gp4.1.A	*Arabidopsis thaliana*	0.92	0.71	78.51	monomer	A	1 × AKG; 1 × MES	2-oxoglutaric acid, Succinic acid, 3,5,7,3′,4′-pentahydroxyflavone, 2-(n-morpholino)-ethanesulfonic acid, (2 s,3s)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4 h-chromen-4-one, Naringenin (4.53)	Naringenin, 2-ketogluta-ric acid, Succinic acid, (+)-catechin OR ((2R,3S)-2-(3,4-Dihydrophenyl)-3,4-dihydro-2H-chromene-3,5,7-triol), 2-(N-Morpholino) ethanesulfonic acid, (-)-taxifolin OR (-)-Dihydroquercetin, 6-C-Glucosylquercetin
ANR	anthocyanidin reductase	Q6DV46 (Q6DV46_CAMSI)	*Camellia sinensis*	2rh8.1.A	*Vitis vinifera*	0.88	−0.23	83.23	monomer	A	1 × Chloride ion	Nadp nicotinamide-adenine-dinucleotide phosphate (2.87)	

2.7. Prediction of putative ligands:

For prediction of ligands capable of binding to the enzymes, we relied on the open access interactive web server, ProBis, which is based on the ProBiS algorithm (Konc and Janezic, 2010) and allows prediction of ligands irrespective of protein folding, and without any prior knowledge of the binding sites. Here, once the protein structure is submitted, the server uses the local structural alignment algorithm ProBiS, to compare and detect structures in the non-redundant PDB (nr-PDB) database that share similar (Z-score > 1) 3D amino acid motifs and environments with the submitted query protein. In the second step, using these nr-PDB proteins as queries, a list of ligands are predicted, which are sorted by the confidence score. Thereafter, transposition of these predicted ligands to the query protein is followed by clustering of the ligands as described by Konc and Janežič (2014). For this study, ligands were selected based on their predicted confidence scores, whereby, only the highest scoring ones were selected from the list.

2.8. Molecular docking studies and analysis of interactions

The receptor-ligand (here, enzyme-substrate) molecular docking studies were performed using the web-server tool- HDOCK which predicts the binding complexes between the two molecules of interest. The docking run can take a minimum of 20–30 min, even if the macromolecules of interest are too large and complex. HDOCK works in a four-step procedure, and, at first it accepts the input as PDB files for both, the receptor and the ligand, and then accurately develops a hybrid docking strategy as noted in the community-wide Critical Assessment of Predictions of Interactions (CAPRI). Next, the PDB sequence database is scanned by the HHSuite package in order to find homologous sequences for the input receptor and the ligand-protein molecules. The two sets of homologous templates thus obtained, for the receptor and the ligand molecules, are then compared to check for the presence of any common records with the same PDB codes. Now, the server can perform docking either by using a common template for both the receptor and ligand molecule, if common PDB codes exist, or can perform the template-free docking. In case multiple templates are available, the best suited template is selected, on the basis of sequence convergence, sequence similarity and resolution. Ultimately, MODELLER builds the structural models based on the templates selected. The interaction energy between the two given macromolecules of interest is calculated by a hierarchical fast Fourier transform (FFT)-based docking program (Mondal et al., 2021).

For conducting the sequential docking studies, three series of molecular docking runs had to be performed. For the first series, the PDB files of the individual enzymes and their respective substrate(s) were considered as the input receptor and ligand molecules, respectively. The binding affinity of each of the substrate molecules for their respective enzymes were observed. Next, for the second series of molecular docking conducted, the PDB files of the enzymes and their individual putative ligands, were considered as the receptors and ligands, respectively. For the last series of molecular docking, each of the PDB files of the enzyme-ligand complexes constructed previously, in the second step of the docking series, were considered as the receptors, and the PDB files of the respective substrate(s) of that enzyme were considered as ligands. Out of the 100 models obtained in the result page of HDOCK server, for each of the runs, only model 1 has been selected, due to its best binding pose and the highest docking score (and lowest Ligand RMSD value) for that particular receptor-ligand interaction. Finally, the observed binding affinity between the (enzyme-ligand)-substrate complexes were compared with the enzyme-substrate complexes (i.e. in absence of the ligands). The detailed information about the type and the total number of intermolecular interactions between the receptor and corresponding ligand molecules are observed using the Protein-Ligand Interaction Profiler (PLIP) web tool.

2.9. Compositional analysis and codon usage indices

GC content (Bhanja et al., 2021), Relative synonymous codon usage (RSCU) values (Sharp and Li 1987; Mondal et al. 2016), GC3s (Bhanja et al., 2021), effective number of codons (ENCs) values (Wright, 1990), GRAVY score, and aromaticity score were considered for compositional analysis (Kyte and Doolittle, 1982) and were generated using CodonW 1.4.4 and in house PERL script. These are well-known measures of codon bias and provide useful information regarding the existence of mutational pressures acting on the genes. The start and stop codons were excluded from the calculation to reduce the influence of amino acid composition on codon usage (Bhanja et al., 2021).

2.10. Expressional probability

The ratio between the geometric mean of RSCU and the highest probable CAI for a particular gene of the same amino acid is deliberated as CAI for a gene. Here CAI represented the degree of gene preference to predict gene expression level (Sharp and Li, 1987).

3. Results

3.1. Retrieval of tea specific flavonoid pathway from KEGG

The tea-specific flavonoid pathway, targeting specific flavonoids was retrieved from KEGG (Fig. 1). Out of the several groups of flavonoids, pathways emanating from three dihydroflavonols (dihydrokaempferol, dihydroquercetin, and dihydromyricetin) were targeted. The products of DFR were leucopelargonidin, leucocyanidin, and leucodelphinidin respectively, and of FLS, were the flavonols- kaempferol, quercetin, and myricetin. From the leucocyanidins, catechins or cyanidins can be produced by the action of LAR and ANS respectively. Afzelchin, catechin, and gallocatechin were produced by LAR, and pelargonidin, cyanidin, and delphinidin were produced by ANS. Further downstream, the epicatechins (epiafzalchein, epicatechin, epigallocatechin) were produced by the action of the enzyme ANR.

3.2. ChemMine

The numeric data clustering based on physicochemical properties of 25 phytochemicals of tea flavonoid pathway clustered together to form an individual clade of flavanones, flavones, flavonols, catechins, anthocyanin, etc. based on the previously mentioned physicochemical properties clustered by ChemmineR (Fig. 2 and Table S2).

3.3. Evolutionary relationships

A Neighbor-joining phylogenetic tree based on a sequence analysis representing the association between proteins associated with tea flavonoid synthesis was obtained. For the phylogenetic tree, it was evident that all the 15 enzymes formed two clades consisting of upstream and downstream enzymes of the flavonoid biosynthesis pathway (Fig. 3A).

3.4. Species phylogeny

Five common organisms i.e. *Camellia sinensis, Glycine* max, *Glycine soja, Vigna unguiculata*, and *Vitis vinifera* were identified after searching common enzyme sequences among 15,000 sequences, retrieved from BLASTP search (Fig. 3B)

3.5. Homology Modeling:

In our present study, models of three tea-specific enzymes- FNS, FLS and ANS, were generated. FNS depicted the highest homology (33.03%) with Ferruginol synthase from *Salvia miltiorrhiza*: 5ylw.1 (template), and
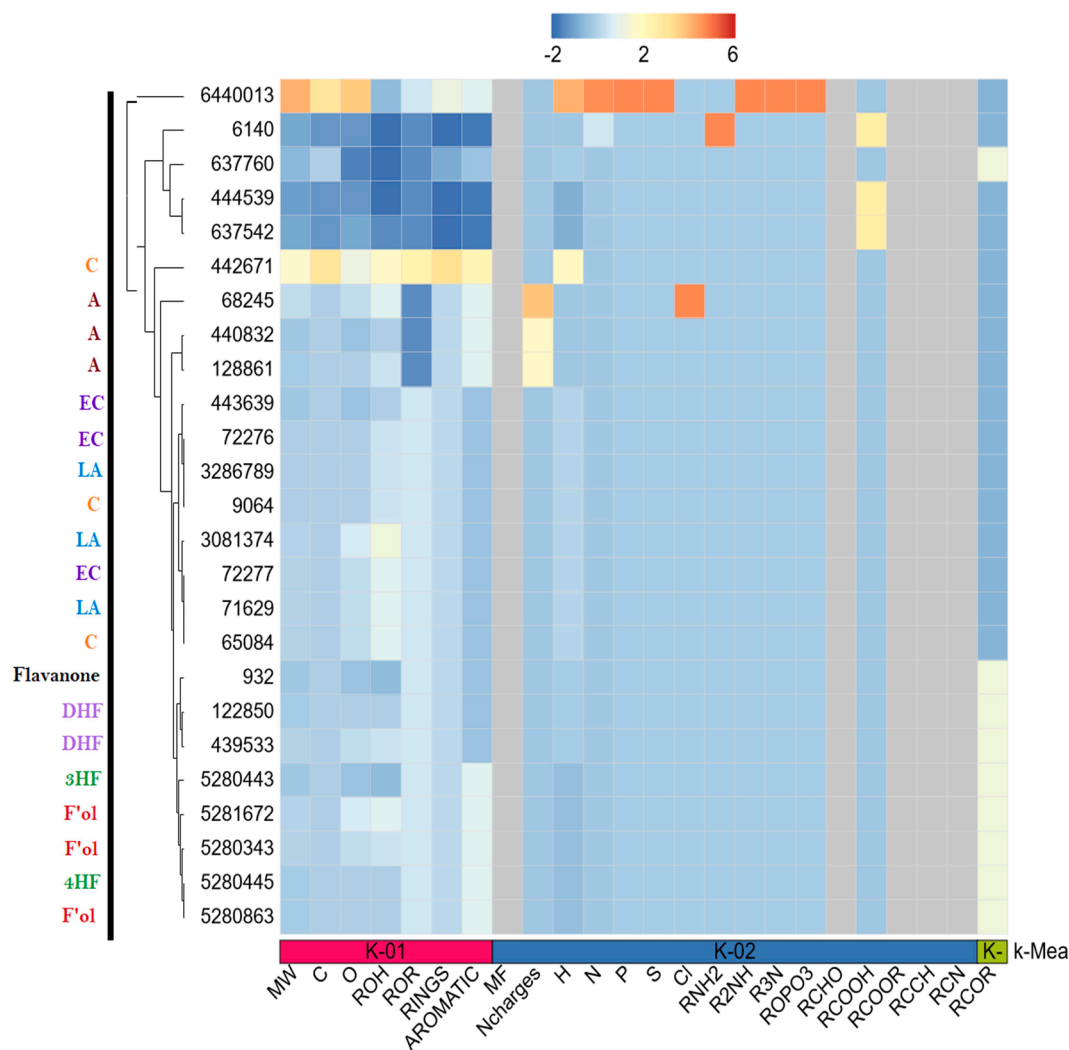
Fig. 2. Numerical clustering of tea phytochemicals representing the association between the products of flavonoids biosynthesis genes in tea. C- Catechin, A-Anthocyanin, EC- Epicatechin, LA- Leucoanthocyanin, DHF- Dihydroflavonol, 3HF- Trihydroxyflavone, F' ol-Flavonol, 4HF-Tetrahyfdroxyflavone. Numbers in each clad represent PubChem CID of tea phytochemicals mentioned in Table S2.

the other two enzymes- FLS and ANS both showed highest homology with Anthocyanidin synthase from *Arabidopsis thaliana*: 1gp4.1.A (template) with 45.26% and 78.51% sequence identity respectively. The oligo state of the predicted models were a monomer for all three cases. SWISS-MODEL provides several metrics for global model evaluation. GMQE, which is a coverage-dependent metric was seen to be 0.63, 0.83 and 0.92 for FNS, FLS and ANS respectively. The average model confidence or QMEANDisCo Global, on the other hand is coverage independent and assesses the model 'as is.' QMEANDisCo Global scores for the predicted 3D models were found to be $+ 0.71 \pm 0.05$, $+0.79 \pm 0.05$ and $+ 0.89 \pm 0.05$ for FNS, FLS and ANS respectively. Values for both the metrices can range from 0 to 1, with higher values corresponding to better quality of the predicted model.

The percentile of 'all atoms Clash score' in MolProbity was found to be 89th, 95th. and 99th. for FNS, FLS and ANS respectively, with the clashscore being 6.45, 4.42 and 1.09 respectively. MolProbity also provides an insight into the Ramachandran plot. Analysis of Ramachandran plot showed 94.3% and 93.80% of the residues of FNS and FLS, respectively to be in the favoured regions and 98.70% and 98.20% of the residues in the allowed regions. ANS had comparatively a greater percentage of its residues in the favoured (98.50%) and the allowed regions (99.70%) of Ramachandran plot (Fig. S2). Finally, when visualized in PyMOL, the predicted 3D models of FNS, FLS and ANS were

found to have RMSD values of 0.139, 0.087 and 0.091 Å when compared to the templates based on which the respective model was built. These data illustrates the reliability of the models built, of the enzymes (Table S7).

3.6. Prediction of putative ligands:

From the list of ligands predicted by ProBis, only a few i.e. the ones showing the highest confidence score were selected. For PAL, 4-carboxycinnnamic acid (4.12) and 2,3-Dihydroxy-1,4-dithiobutane were seen to be the ligands with the highest confidence. For FNS, Protoporphyrin IX containing Fe (3.3), 1,2-ethanediol (3.3), (S,R)-*meso*-tartaric acid (3.3) showed the highest confidence score. FLS showed the highest confidence score for ligands- 2-oxoglutaric acid (3.84), Succinic acid (3.84), Naringenin (3.84), 3,5,7,3′,4′-pentahydroxyflavone (3.84), 2-(*n*-morpholino)-ethanesulfonic acid (3.84). Ligands predicted to bind with ANS were 2-oxoglutaric acid, Succinic acid, 3,5,7,3′,4′-pentahydroxyflavone, 2-(*n*-morpholino)-ethanesulfonic acid, (2 s,3s)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4 h-chromen-4-one, Narin genin (4.53 for all) among several others. These putative ligands were used for further studies involving molecular docking analyses with the respective enzyme-substrate complexes (Table 1).
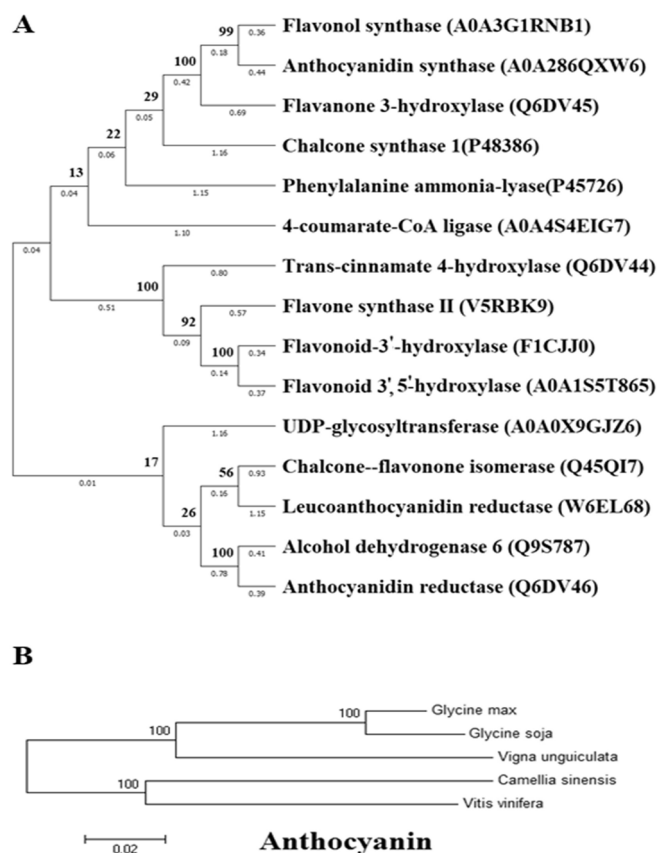
A

B

Anthocyanin

Fig. 3. A Neighbour-joining phylogenetic tree based on a sequence analysis representing the association between proteins associated with tea flavonoid synthesis. The UniProt accession numbers were displayed in brackets following the name of enzymes. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved 15 amino acid sequences. B. Species-specific phylogenetic tree constructed from fifteen enzymes (Table 1) of anthocyanin biosynthesis pathway using the Maximum Likelihood (ML) method. Amino acid sequences were aligned with the Clustal W algorithm and the phylogenetic tree was constructed using MEGA X software (http://www.megasoftware.net/). The branching order is validated by 1000 steps of bootstrap replicates. The number in each node indicates the confidence value of that branch after bootstrapping the phylogenetic tree.

3.7. Molecular docking analysis

The present study intends to compare the interactions and binding affinity (represented as docking score) between the enzyme with their respective substrate(s) and the interaction between the enzyme-ligand complexes with the respective substrate(s) of the enzymes. The outcome of this computational investigation was then used for the prediction of ligands which are hypothesized to cause an improvement in the binding kinetics of the enzyme-substrate reaction, and thus enhance the product formation. In order to realise this objective, we performed several sets of docking analysis with four enzymes (PAL, FNS, FLS, and ANS), their corresponding substrate(s), and ligands. The selection of enzymes was done based on the products formed by them. FNS, FLS, and ANS form flavones, flavonols, and cyanidins respectively, all of which are key products of the flavonoid biosynthetic pathway, and are involved in the enhancement of tea flavour. For PAL, docking with its substrate only resulted in a docking score of −91.81 kcal/mol with the RMSD value of 41.54 Å. The hydrophobic interactions occur via 423Ala_B and 487Asp_A and H-bonds were formed by 484His_C and 487Asp_A residues of the enzyme to interact with its substrate

(Table S1). When PAL-4-carboxycinnnamic acid and the PAL- 2,3-Dihydroxy-1,4-dithiobutane complexes were reacted with phenylalanine, no change in the binding affinity and interactions were observed thus signifying the indispensable role of the aromatic amino acid in this pathway.

Naringenin (Fig. S3), an important and extensively studied compound of the flavonoid biosynthetic pathway, acts as the substrate of FNS, to ultimately form flavones. A docking score of −139.76 kcal/mol for this enzyme-substrate complex was obtained, with the major hydrophobic interactions occurring via 81Val_A, 107Arg_A, 222Phe_A, 226Asp_A, 227Phe_A, 407Phe_A and H-bonds formation by 107Arg_A, 220Gly_A, 226Asp_A, when studied via PLIP. Additional hydrophobic interactions- 59Ile_A, 108Asp_A, 219Phe_A, 220Gly_A and 221Glu_A were also seen when visualised by LigPlot, (Fig. 4). Although with *meso* tartaric acid as the ligand, no significant change in binding affinity was observed, FNS when complexed Protoporphyrin IX Containing Fe, and finally with naringenin, a significant increase in the binding affinity to −162.38 kcal/mol, along with the increase in the total number of hydrophobic interactions to 13 was observed, with the interactions taking place via 59Ile_A, 112Ile_A, 123Phe_A, 215Val_A, 219Phe_A, 382Val_A, 384Leu_A and 407Phe_A (in PLIP). Additional hydrophobic interactions of 220Gly_A, 221Glu_A, 222Phe_A, 223Asn_A and 226Asp_A were seen in LigPlot. This indicates that these ligands facilitate the binding between the enzyme and the substrate upon their binding with the enzymes prior to substrate binding. H-bonds were formed by 223Asn_A (Fig. 4 and Tables 1, S1).

Flavonols are formed from dihydroflavonols- dihydrokaempferol, dihydroquercetin, and dihydromyricetin by the enzyme FLS. On docking of the substrates only with FLS, docking scores of −172.87 kcal/mol, −178.84 kcal/mol, and −173.34 kcal/mol respectively, were obtained. Out of the six ligands docked with the FLS-dihydromyricetin complex, a significant change in docking score was observed for the interaction with naringenin, 2-Morpholin-4-ium-4-ylethanesulfonate (MES), and 6-C-Glucosylquercetin. With naringenin, the binding affinity between the enzyme and dihydromyricetin increased to −184.82 kcal/mol, whereas, minute changes (172.3 kcal/mol) in the binding affinity was observed for dihydroquercetin and a significant decrease in the binding affinity for dihydrokaempferol (-158.69 kcal/mol), was observed. For the (FLS-Naringenin)-Dihydromyricetin complex, several hydrophobic interactions- 114Lys_A, 198Lys_A, 128Tyr_A, 126Ile_A, 214Val_A, 215Glu_A, 218Thr_A, 219Asp_A, 220Met_A and 217His_A were observed via LigPlot. A significant number of H-bonds with 128Tyr_A, 198Lys_A, 218Thr_A, 219Asp_A, 220Met_A, 318Tyr_A, 322Arg_A and 325Lys_A were also found when analysed via PLIP.

Moreover, with MES complexed with FLS, a slight increase in the binding affinity up to −178.92 kcal/mol along with an increase in the total number of H-bonds, Π-stacking interactions an additional Π-cation interaction between 322Arg_A residue of the enzyme and dihydromyricetin, a salt bridge between 322Arg_A and sulfonic acid were also observed. On the other hand, upon analyzing the interaction between the FLS-MES complex with dihydrokaempferol, a significant lowering in binding affinity for the substrate to 159.32 kcal/mol was observed, indicating that the ligand MES might act as an inhibitor for the interaction between FLS and dihydrokaempferol.

Furthermore, 6-C-Glucosylquercetin when complexed with FLS, showed a significant increase in binding affinity for dihydromyricetin, up to 187.26 kcal/mol (RMSD value of 26 Å), with several H-bonds and hydrophobic interactions of 103Pro_A, 108Phe_A, 113Thr_A, 115Met_A, 114Lys_A, 128Tyr_A, 117Lys_A and 326Phe_A being recorded between FLS and dihydromyricetin. However, contrasting results were obtained for the interactions between FLS and 6-C-Glucosylquercetin with that of dihydrokaempferol and dihydroquercetin, as the binding affinity for each of these substrates were reduced to 162.11 kcal/mol (RMSD 26.11 Å) and 167.53 kcal/mol (RMSD 26.34 Å) respectively (Fig. 4 and Tables 1, S1).

Enzyme ANS converts leucoanthocyanidins to cyanidins which serve

Fig. 4. Comparative analysis of selected interactions found after molecular docking (by HDOCK) of the enzyme-substrate complex (left) with the (enzyme-ligand)-substrate complex (right). Each (enzyme-substrate)/ (enzyme-ligand)-substrate complex forms hydrogen (green dotted line between amino acids of enzyme, substrate, and ligand) and hydrophobic bond (red half-circles) during the interaction.

as the immediate precursors for the anthocyanin biosynthetic pathway. Leucopelargonidin, leucocyanidin, and leucodelphinidin interact with ANS to produce pelargonidin, cyanidin, and delphinidin respectively. For each of these enzyme-substrate reactions, binding affinity in the form of docking scores of −147.41 kcal/mol, −155.7 kcal/mol, and −161.42 kcal/mol was noted. ANS, when complexed with ligands 3,5,7,3′,4′-pentahydroxyflavone, MES, and naringenin, an increase in binding affinity up to −162.80 kcal/mol, −151.20 kcal/mol, and −152.20 kcal/mol respectively, for leucopelargonidin were observed. The (ANS-3,5,7,3′,4′-pentahydroxyflavone)- leucopelargonidin complex depicted hydrophobic interactions of 120His_A, 145Tyr_A, 147Phe_A, 216Lys_A, 235His_A, 237Asp_A, 307Phe_A, 309Glu_A and H-bond with 238Val_A (LigPlot), with P-type Π-stacking interaction by 147Phe_A (PLIP). Interaction between ANS and naringenin as the ligand was seen

to enhance the binding affinity up to −162.70 kcal/mol with RMSD of 31.16 Å, for leucocyanidin. Here, hydrophobic interactions of 120His_A, 147Phe_A, 125Ile_A, 309Glu_A, 312Lys_A, 337Phe_A, 307Phe_A, 216Lys_A; and H-bonds with 145Tyr_A, 214Gln_A were observed (Lig-Plot). With ligands 2-oxoglutaric acid and 3,5,7,3′,4′-pentahydroxy-flavone, ANS showed an increased affinity for leucodelphinidin, with docking scores −167.38 kcal/mol and −170.92 kcal/mol respectively. H-bonds with 220Tyr_A, 238Val_A, 239Ser_A (LigPlot), P-type Π-stacking interaction by 307Phe_A, Π-cation interactions by 216Lys_A and 235His_A (PLIP) were seen to occur in the (ANS-2-oxoglutaric acid)-leucodelphinidin complex. Also, the ligand MES decreases the binding affinity of FLS for dihydrokaempferol whereas 6-C-Glucosylquercetin decreases the binding affinity of FLS for both dihydrokaempferol and dihydroquercetin (Fig. 4 and Tables 1, S1).
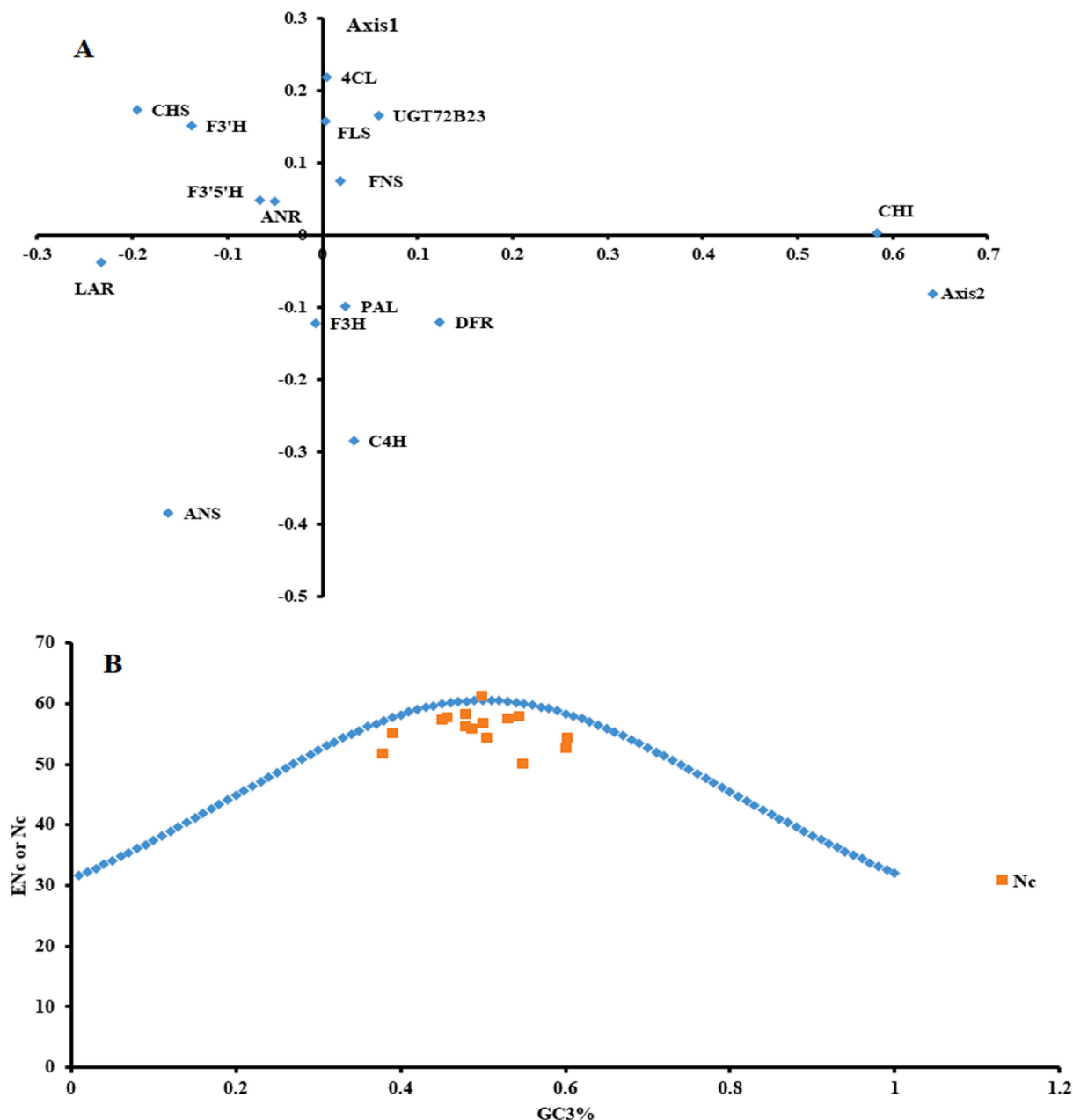

Fig. 5. A. Position of genes along the first and second axes produced by correspondence analysis based on RSCU values of the 59 synonymous codons from flavonoid biosynthesis pathways associated with genes of *Camellia sinensis*. X and Y-axis correspond to Axis 1 and Axis 2. Fig. 5B. ENc plot analyses (ENc or Nc plotted against GC3) of flavonoid biosynthesis pathways associated with genes from Camellia sinensis. ENc represents the effective number of codons, and GC3 is the GC content of synonymous codons at the third position. In the plot, the ENc from GC3 was shown as a bell-shaped bisymmetric curve. The solid line represents the expected curve when codon usage bias is only affected by mutational pressure. Nc was obtained from Codon W 1.4.4 (http://codonw.sourceforge.net/).

3.8. In silico analysis of codon usage based on GC content

It was observed that the average abundance of GC3s% and GC1% in 15 flavonoid biosynthetic pathway genes from *Camellia sinensis* were 51.33% and 49.80% respectively. The occurrence of GC% in the third position was highest in the *ANS* gene at about 60.40% and lowest in the *4CL* gene at about 37.95%. The GC3% was 37.95, 39.10, 45.19, 45.79, 47.99, 48.04 and 48.85% in *4CL*, *CHS*, *LAR*, *FLS*, *F3′H*, *FNS* and *PAL* genes respectively. On the other hand, the GC3% of *ANR, F3′5′H, UGT72B23* genes were about 50%, 50.1%, and 50.65% respectively. Only *C4H* (60.23%) and *ANS* (60.40%) were slightly GC-rich. Moreover, the average GC2% of the selected genes was 39.33% which indicated the AU preference in the second position of the codons (Fig. 5A and Table S3).

3.9. ENc-plot analysis

Genes that are situated towards the proximity of the curve line were deliberated to be under mutational pressure. Here ENc and Nc were plotted against the GC3s for Flavonoid biosynthetic pathway genes from *C. sinensis* were close to the curve and *ANR* gene was positioned on the expected curve. In the present study, *CHI*, *4CL*, *C4H* genes were located below the curve. Also, the Nc values of the Flavonoid biosynthetic pathway genes from *C. sinensis* varied from 49.87 (*Chalcone isomerases*) to 61.00 (*ANR*). In *ANR* gene codon usage were equal (unbiased) probabilities. The Nc values of flavonoid biosynthetic pathway genes were detected > 40 (Fig. 5B and Table S3).

3.10. Gene adaptability analysis

The probable degree of expression of 15 flavonoid biosynthetic pathway genes from *C. sinensis* CAI was analyzed and predicted. In our study, the CAI value ranged from 0.688 (*F3′H*) to 0.758 (*FLS*). For the 15 Flavonoid biosynthetic pathway genes, their average CAI values were 0.72, and ENc values were 48.06, indicating moderately higher expression of genes having a comparatively strong codon adaptation across its whole genome (Table S6).

3.11. Correspondence analysis (COA) of the flavonoid biosynthetic pathway genes from C. Sinensis

The variation of codon usage of the selected genes in *C. sinensis*, the Correspondence Analysis predicted RSCU of the genes. The index for distressing codon usage bias was presented as axis 1. As shown in Fig. 5A, among the selected genes *CHI* was close in proximity to the major axis2, and genes such as *4CL*, *FLS*, and *F3H* prefer to remain in proximity along axis1 (vertical axis). The other selected genes were not close to any of the major axes and were distributed in the 4 quadrants.

3.12. Estimation of GRAVY and AROMO

The predicted GRAVY scores of flavonoid biosynthetic pathway (except F3′H and ANR) enzymes exhibited negative values, demonstrating hydrophilic nature of enzymes. The average AROMO values were 0.1, for C4H and 4CL, 0.09, for CHS, FNS, and DFR, 0.08, for PAL, CHI, F3H, LAR, F3′5′H, and F3′H, 0.07, for UGT72B23, ANS, and ANR genes (Table S4).

3.13. Relative synonymous codon usage analysis of the flavonoid biosynthetic pathway genes from C. Sinensis

From the RSCU values of the selected genes, it was observed that there were no such biasness to specific codon composition (Table S5). Codons like UUG (Lys) were more useful in all the genes whereas codons like GCG (Ala), AUA (Ile), and GUA (Val) have less usage for all the genes. For some codons and genes, the different positional GC compositions had the proper reflection on the RSCU values like GAG (Glu), AUC (Ile), GUG (Val), GAA (Glu), and CUU (Leu). The GC3% of the genes *C4H* (60.23%), *ANS* (60.40%), *CHI* (54.98%), and *F3H* (54.51%) was high in comparison to other selected genes and it was properly reflected for the following codons like GAG (Glu), AAG (Lys), UUG (Leu), CCG (Pro), AGG (Arg), GUG (Val), UAC (Tyr), CUA (Leu), CCA (Pro), AAA (Lys), UUA (Lys) and exceptional values appear for the codons AUU [Ile (high)], CGC [Arg (low)]. For most of the selected genes, the RSCU values of the 25–31 codons were>1 except for *F3′H* (23) and *ANS* (20).

4. Discussion

The present work focused on the enhancement in the content of tea-specific flavonoids by modulating four key enzymes of the flavonoid biosynthetic pathway with the help of allosteric modulators/ligands. The ligands, when bound to the enzyme, were seen to cause improvement in the spontaneity of the enzyme-substrate reaction. The outcome of the study will be beneficial for tea cultivation in reducing the economic losses incurred each year due to the reduced content of flavonoids, and hence flavour, in the second flush of Darjeeling tea. For the prediction of putative ligands, two sources- SWISS-MODEL Template Library (SMTL) and ProBis were employed. SWISS-MODEL uses comparative modeling to build 3D models from experimentally determined protein structures and incorporates evolutionarily conserved ligands (cofactors and metal ions) from the identified templates of the enzymes. The detailed working mechanism of SMTL has been described by Biasini et al., 2014. The ProBis server, on the other hand, accepts a query protein structure and relies on local structural similarities, followed by transposition of putative ligands to the query protein structure and finally clustering the ligands together. The detailed mode of action can be found here Konc and Janežič, 2014. The number and type of non-covalent interactions i.e., hydrogen bonds, halogen bonds, hydrophobic contacts, pi-stacking, pi-cation interactions, salt bridges, and water bridges present in the enzyme-substrate and the (enzyme-ligand)-sub-strate interactions were identified by LigPlot and the web-server tool PLIP.

Engineering of flavonoid pathways for increased flavonoid production, and thus, enhanced flavour in tea plants have been the focus of scientific attention for several years. Metabolic engineering tools have been optimized for expression in recombinant microbial cell factories for attaining the best results. Microorganisms have always been a favourite choice for researchers for studying complex pathway reconstruction for the production of secondary metabolites, due to their easy handling and genetic manipulation procedures. *Escherichia coli*, *Saccharomyces cerevisiae*, and *Streptomyces venezuelae* are commonly used for the production of several classes of flavonoids. Several metabolic engineering techniques aid in pathway reconstruction in microbes, some of which are achieved by improving and optimizing the host strain, performing targeted gene manipulations, and by a complete understanding of the enzymes involved in the pathway (Chouhan et al., 2017). However, insufficient concentration of malonyl CoA- the precursor metabolite in the cell, the reduced accessibility to phenylalanine and tyrosine- the starting materials of the pathway, weak expression of a few enzymes, the lack of stability and reduced solubility of flavonoids, especially anthocyanins, often act as barriers and limit the engineering techniques.

Another approach of modifying the flavonoid biosynthetic pathway for enhancement of tea flavour can be the excavation of the tea microbiome. Given the well-known cooperation between the plant and its associated root microbiota, it can be hypothesized that the tea microbiome can play an integral part in enhancing the flavour content of tea. Till date, the multitude of bacterial species known to inhabit the tea microbiome, are endophytic and includes both Gram-positive and Gram-negative species. The commonly found Gram-positive bacteria are the aerobic, spore-forming, rod-shaped genus *Bacillus* (*Bacillus pumilus*, the large *Bacillus* megaterium, *Bacillus thuringiensis,* and *Bacillus*

amyloliquefaciens) and the facultative anaerobic, endospore-forming *Paenibacillus polymyxa*, Also, Gram-negative species of micro-aerophilic, nitrogen-fixing *Azospirillum*, the aerobic *Beijerinckia*, the strictly aerobic, rod-shaped and encapsulated *Pseudomonas aeruginosa*, the non-spore-forming, aerobic *Comamonas acidovorans* and the facultative anaerobic *Serratia marcescens* of the family Enterobacteriaceae are found (Banik et al., 2019; Tshikhudo et al., 2019), along with arbuscular mycorrhizal fungi (AMF) and *Beauveria bassiana* (Banik et al., 2014; Shao et al., 2019). Tea endophytes produce a wide range of natural products as a part of their secondary metabolism. These metabolites produced can perform a wide range of functions- the source of medicinal, agricultural, and other related industries; alkaloids contribute to plant defense, promote the growth of plants and facilitate stress management; produce and induce indole acetic acid (IAA) to promote the growth of plants; steroids, terpenoids, and diterpenes suppress pathogens, aid in the removal of contaminants, solubilize phosphate, and assimilate nitrogen (Tshikhudo et al., 2019).

Naringenin, a key player of the tea-specific flavonoid pathway, can also act as an allosteric modulator and was seen to enhance the binding affinity of enzymes FLS for its substrate dihydromyricetin and of ANS, for leucopelargonidin and leucocyanidin.

An endophytic fungus from *Dalbergia odorifera*, produces naringenin and acts as an antibacterial agent against five pathogenic bacteria, with *Staphylococcus aureus* being the most sensitive. The other bacteria found to be sensitive were *Escherichia coli*, *Salmonella enteritidis*, *Pseudomonas aeruginosa*, and *Bacillus subtilis* (Gao et al., 2021). A directly proportional relationship between the production of naringenin with escalation of flavour, was demonstrated by expressing two genes- *Al4CL* and *AlCHS*, in *Saccharomyces cerevisiae*, from *Alternaria* sp. MG1, which is a Deuteromycetes fungi and a grape endophyte. This led to the successful synthesis of the expected compounds- p-coumaroyl CoA from *Al4CL* and both naringenin and resveratrol from *AlCHS*. Also, the separate co-culture of *S. cerevisiae* strains harbouring *AlCHS* and *Al4CL* demonstrated the synthesis of naringenin and resveratrol, simultaneously. Further, HPLC and SPME-GC–MS analysis of wine produced from *S. cerevisiae* containing *Al4CL*, ameliorated the content of the compounds responsible for flavour and aroma (Lu et al. 2021).

Similarly, 2-oxoglutaric acid was seen to act as a ligand for ANS and thus enhance the binding affinity of the enzyme for leucodelphinidin. Screening and identification of plant growth-promoting endophytic bacterial strains *Curtobacterium* spp., *Brevibacillus* spp., *Paenibacillus* spp., *Bacillus* spp., and *Microbacterium* spp. were identified, all of which facilitates the production of IAA and siderophore, aids in phosphate solubilization, enhances plant growth and produces or induces the production of certain metabolites within the plant endosphere (Banik et al., 2016; Mahmood and Kataoka, 2020). One of the metabolites- 2-oxoglutaric acid was produced in the highest quantity in plants that were treated with *Microbacterium* spp. when compared to control plants. Also, a lower quantity and concentration of organic acids (2-oxoglutaric acid) in control plants is an indication of nitrogen deficiency in these plants.

Also, as reviewed by Agnolucci et al, (2020), the production of several secondary metabolites is dependent on the inoculation of AMF, PGPB, and AMF-associated bacteria in plants. For example, an increased concentration of phenolics is acquired upon inoculation with *Glomus fasciculatum*, *Glomus intraradices*, *Rhizoglomus irregulare*, *Glomus iranicum* var. *tenuihypharum*, *Funneliformis mosseae* 2 W3, IMA1, IN101C, *Claroideoglomus claroideum* 22 W3, and *Bacillus lentimorbus* B-30488. Inoculation with AMF *Glomus intraradices* increases the concentration of several flavonols including quercetin, kaempferol, and catechins. Concentration of anthocyanidins is enhanced by the association of several AMF with various strains of *Pseudomonas fluorescens*; such as, the association between *Septoglomus viscosum* and *P. fluorescens* 19Fv1t increases the concentration of pelargonidin 3-glucoside. These suggest that if such endophytes are introduced into the tea microbiome network, they may act as the source of ligands, and ultimately help in facilitating the

reactions of the flavonoid pathway, thus augmenting the flavour content of tea (Bag et al., 2022b).

The presence of GC at 3rd base in a codon (GC3), regarded as one of the most crucial features to study the codon usage bias. Diverse positional GC% can predict information related to selection pressure and comparative influence of mutation on codon usage bias (Mondal et al., 2016). In the present investigation, it was evident that the average percentage of GC3s% and GC1% in 15 Flavonoid biosynthetic pathway genes from *Camellia sinensis* was 51.33% and 49.80% respectively, indicating the non-biasness of their nucleotide composition. The GC% in the 3rd position was less than 50% for 4CL, CHS1, LAR, FLS, F3′H, FNS, and PAL genes demonstrating AT-rich genome. ANR, F3′5′H, UGT72B23 gene revealing the non-biasness of their nucleotide composition. The average GC2% of the selected genes was 39.33% which indicated the AU preference in the second position of the codons. Previous studies reported that the codon usage of the *Camellia sinensis* (CsSPDS) spermidine synthase gene (You et al., 2015), chloroplast genes from three species of tea (Yengkhom et al., 2019), *Aquilaria sinensis* (Lour.) Gilg (Wang et al., 2016), genomes of three species such as tomato, tobacco, and potato (Anwar et al., 2021) biased respecting the A-ended or *T*-ended codons along with AT-rich genes. Similarly, in this current study majority of the Flavonoid synthetic pathway-associated genes in *C. sinensis* were AT-rich and influenced to A/U-ending identical codons. ENC-plot of 15 spermidine synthase genes from *C. sinensis* showed how mutation influences codon usage of these genes, thus indicating the crucial role of mutation-selection for selective determination of codon bias in flavonoid biosynthetic pathway associated genes from *C. sinensis* (You et al., 2015).

Many genes associated with chlorophyll degradation and synthesis pathways favored weak codon bias (Begum and Mondal 2021). Likewise, in this study, genes from Camellia sinensis preferred weak codon bias. The moderately higher CAI values indicated the moderate-higher expression of the genes within the genome/organism and the adaptability of the codons and their usage might play a key role in *C. sinensis*. All the Flavonoid biosynthetic pathway genes from C. sinensis have a similar expression pattern which revealed their co-expression also. COA can demonstrate the occurrence of genes that can be translated to its codons by influencing codon usage bias (Begum and Mondal, 2021). In the present study, some selected genes were not close to any of the major axes and distributed in the 4 quadrants and this random distribution indicated the mutational pressure on the genes. The closeness of F3′5′H and ANR gene indicated their similar codon usage pattern as this correspondence analysis was performed depending on RSCU of the codons. Their clustered topological orientation indicated their preference over mutational bias. The predicted GRAVY scores of protein sequences accomplished from all, except *F3′H* and *ANR* tangled in the Flavonoid biosynthetic pathway in *C. sinensis* showed negative value, indicating its hydrophilic nature. The biasness of the codons varied gene to gene and amino acids to amino acids. From the RSCU values of the selected genes, we have observed that there were no such biasness to specific codon composition. Moreover, there was no strong association of the different positional GC content especially to the GC3% with the codon usage pattern. The less number of codons with higher RSCU value indicated codon preferences. In addition, these high-frequency codons mostly ended with A or U (except UUG and GUG). Our study indicated that the GC content was not significant to modulate the codon usage pattern of the selected genes and it might be controlled by the tRNA abundance and other intrinsic factors of the organism.

5. Conclusion

The results of the present study provide palpable insight into the role ligands might play in the flavonoid biosynthetic pathway of *C. sinensis*. Acting as allosteric modulators, ligands were seen to enhance the binding affinity of the key enzymes of the pathway for their respective substrate(s). The increase in spontaneity of the enzyme-substrate

reactions suggested that this strategy of enhancing the content of several groups of flavonoids can be endorsed to augment the flavour component of tea. Moreover, secondary metabolites (ligands) from endophytes might provide a constant source of these ligands, and hence enhanced flavour. This will contribute to the future application of reducing economic losses borne each year by the tea plantation industry due to the difference in flavour of tea caused by seasonal variations. However, the non-availability of the X-crystallographic structures of the enzymes except PAL, is a major setback of the present study. Hence, future research endeavour is required in this direction of tea biochemistry to bridge the gap between laboratory science and implementation and to validate the *in-silico* models developed, to further expand the perspectives of tea-flavour research.

CRediT authorship contribution statement

Anusha Majumder: Investigation, Methodology, Writing – original draft. **Sunil Kanti Mondal:** Conceptualization, *In silico* studies, Writing. **Samyabrata Mukhoty:** Docking studies. **Sagar Bag:** Investigation, Phylogeny. **Anupam Mondal:** Investigation. **Yasmin Begum:** Investigation. **Kalpna Sharma:** Conceptualization. **Avishek Banik:** Funding acquisition, Supervision, Methodology, Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2022.100212.

References

Agnolucci, M., Avio, L., Palla, M., Sbrana, C., Turrini, A., & Giovannetti, M. (2020). Health-Promoting Properties of Plant Products: The Role of Mycorrhizal Fungi and Associated Bacteria. *Agronomy, 10*(12), 1864. https://doi.org/10.3390/agronomy10121864

Backman, T. W. H., Cao, Y., & Girke, T. (2011). ChemMine tools: An online service for analyzing and clustering small molecules. *Nucleic acids research, 39*(suppl), W486–W491. https://doi.org/10.1093/nar/gkr320

Bag, S., Mondal, A., & Banik, A. (2022b). Exploring tea (Camellia sinensis) microbiome: Insights into the functional characteristics and their impact on tea growth promotion. *Microbiological Research, 254*, 126890. https://doi.org/10.1016/j.micres.2021.126890

Bag, S., Mondal, A., Majumder, A., & Banik, A. (2022a). Tea and its Phytochemicals: HiddenHealth Benefits & Modulation of Signaling Cascade by Phytochemicals. *Food Chemistry, 371*, Article 131098. https://doi.org/10.1016/j.foodchem.2021.131098

Banik, A., Chattopadhyay, A., Ganguly, S., & Mukhopadhyay, S. K. (2019). Characterization of a tea pest specific *Bacillus thuringiensis* and identification of its toxin by MALDI-TOF mass spectrometry. *Industrial Crops and Products, 137*, 549–556. https://doi.org/10.1016/j.indcrop.2019.05.051

Banik, A., Ganguly, S., Mukhopadhyay, B. B., & Mukhopadhyay, S. K. (2014). A new report on rapid, cheap and easily extractable mass spore production of Beauveria bassiana using recyclable polyurethane foams as support medium. *J. Microbiol. Biotech. Res, 4*, 1–6.

Banik, A., Mukhopadhaya, S. K., & Dangar, T. K. (2016). Characterization of N2-fixing plant growth promoting endophytic and epiphytic bacterial community of Indian cultivated and wild rice (*Oryza* spp.) genotypes. *Planta, 243*(3), 799–812. https://doi.org/10.1007/s00425-015-2444-8

Begum, Y., & Mondal, S. K. (2021). Comprehensive study of the genes involved in chlorophyll synthesis and degradation pathways in some monocot and dicot plant species. *Journal of biomolecular structure & dynamics, 39*(7), 2387–2414. https://doi.org/10.1080/07391102.2020.1748717

Bhanja, E., Das, R., Begum, Y., & Mondal, S. K. (2021). Study of Pyrroloquinoline Quinine From Phosphate-Solubilizing Microbes Responsible for Plant Growth: In silico Approach. *Frontiers in Agronomy, 3*. https://doi.org/10.3389/fagro.2021.667339

Bhardwaj, V. K., Singh, R., Sharma, J., Rajendran, V., Purohit, R., & Kumar, S. (2021). Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. *Journal of Biomolecular Structure and Dynamics, 39*(10), 3449–3458. https://doi.org/10.1080/07391102.2020.1766572

M. Biasini S. Bienert A. Waterhouse K. Arnold G. Studer T. Schmidt … T. Schwede SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information 42 W1 2014 2014 W252 W258 10.1093/nar/gku340.

Bradshaw, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature, 426*(6963), 176–178. https://doi.org/10.1038/nature02106

Chouhan, S., Sharma, K., Zha, J., Guleria, S., & Koffas, M. A. (2017). Recent advances in the recombinant biosynthesis of polyphenols. *Frontiers in microbiology, 8*, 2259. https://doi.org/10.3389/fmicb.2017.02259

de Sousa Luisa, J. A., Barrosb, R. P. C., de Sousab, N. F., Muratovc, E., Scottib, L., & Scottib, M. T. (2020). Virtual screening of natural products database. *Mini Reviews in Medicinal Chemistry, 20*. https://doi.org/10.2174/1389557520666200730161549

Gao, Y., Ji, Y., Li, W., Luo, J., Wang, F., Zhang, X., … Yan, L. (2021). Endophytic Fungi from Dalbergia odorifera T. Chen Producing Naringenin Inhibit the Growth of Staphylococcus aureus by Interfering with Cell Membrane, DNA, and Protein. *Journal of Medicinal Food, 24*(2), 116–123. https://doi.org/10.1089/jmf.2020.4686

Ho, C. T., Zheng, X., & Li, S. (2015). Tea aroma formation. *Food Science and Human Wellness, 4*(1), 9–27. https://doi.org/10.1016/j.fshw.2015.04.001

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research, 28*, 27–30. https://doi.org/10.1093/nar/28.1.27

Konc, J., & Janezic, D. (2010). ProBiS algorithm for detection of structurally similar 806 protein binding sites by local structural alignment. *Bioinformatics, 26*(9), 1160–1168. https://doi.org/10.1093/bioinformatics/btq100

Konc, J., & Janežič, D. (2014). ProBiS-ligands: A web server for prediction of ligands by examination of protein binding sites. *Nucleic acids research, 42*(W1), W215–W220. https://doi.org/10.1093/nar/gku460

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol., 157*(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Lu, Y., Song, Y., Zhu, J., Xu, X., Pang, B., Jin, H., … Shi, J. (2021). Potential application of CHS and 4CL genes from grape endophytic fungus in production of naringenin and resveratrol and the improvement of polyphenol profiles and flavour of wine. *Food Chemistry, 347*, 128972. https://doi.org/10.1016/j.foodchem.2020.128972

Mahmood, A., & Kataoka, R. (2020). Metabolite profiling reveals a complex response of plants to application of plant growth-promoting endophytic bacteria. *Microbiological research, 234*, 126421. https://doi.org/10.1016/j.micres.2020.126421

McKay, D. L., & Blumberg, J. B. (2002). The role of tea in human health: An update. *Journal of the American College of Nutrition, 21*(1), 1–13. https://doi.org/10.1080/07315724.2002.10719187

Mondal, S. K., Kundu, S., Das, R., & Roy, S. (2016). Analysis of phylogeny and codon usage bias and relationship of GC content, amino acid composition with expression of the structural nif genes. *J Biomol Struct Dyn., 34*(8), 1649–1666. https://doi.org/10.1080/07391102.2015.1087334

Mondal, S. K., Mukhoty, S., Kundu, H., Ghosh, S., Sen, M. K., Das, S., & Brogi, S. (2021). In silico analysis of RNA-dependent RNA polymerase of the SARS-CoV-2 and therapeutic potential of existing antiviral drugs. *Computers in Biology and Medicine, 135*, 104591. https://doi.org/10.1016/j.compbiomed.2021.104591

Schafferhans, A., & Klebe, G. (2001). Docking ligands onto binding site representations derived from proteins built by homology modelling. *Journal of molecular biology, 307*(1), 407–427. https://doi.org/10.1006/jmbi.2000.4453

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research, 31*, 3381–3385. https://doi.org/10.1093/nar/gkg520

Shao, Y. D., Zhang, D. J., Hu, X. C., Wu, Q. S., Jiang, C. J., Gao, X. B., & Kuča, K. (2019). Arbuscular mycorrhiza improves leaf food quality of tea plants. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* (volume 47, issue,, 3. http://hdl.handle.net/20.500.12603/197.

Sharma, J., Kumar Bhardwaj, V., Singh, R., Rajendran, V., Purohit, R., & Kumar, S. (2021). An *in-silico* evaluation of different bioactive molecules of tea for their inhibition potency against non structural protein-15 of SARS-CoV-2. *Food chemistry, 346*, 128933. https://doi.org/10.1016/j.foodchem.2020.128933

Sharp, P. M., & Li, W.-H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res., 15*(3), 1281–1295. https://doi.org/10.1093/nar/15.3.1281

Tshikhudo, P., Ntushelo, K., Mudau, F., Salehi, B., Sharifi-Rad, M., Martins, N., … Sharifi-Rad, J. (2019). Understanding Camellia sinensis using omics technologies along with endophytic bacteria and environmental roles on metabolism: A review. *Applied Sciences, 9*(2), 281. https://doi.org/10.3390/app9020281

W.F. Van Gunsteren S.R. Billeter A.A. Eising Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P., & Tironi, I.G Vdf Hochschulverlag AG an der ETH Biomolecular Simulation: The GROMOS96 Manual and User Guide 1996 Zürich, Zürich, Switzerland 1 1042.

Wang, X., Zeng, L., Liao, Y., Zhou, Y., Xu, X., Dong, F., & Yang, Z. (2019). An alternative pathway for the formation of aromatic aroma compounds derived from L-

phenylalanine via phenylpyruvic acid in tea (*Camellia sinensis* (L.) O. Kuntze) leaves. *Food chemistry, 270*, 17–24. https://doi.org/10.1016/j.foodchem.2018.07.056

Wang, Y., Zhan, D. F., Jia, X., Mei, W. L., Dai, H. F., Chen, X. T., & Peng, S. Q. (2016). Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. *Front. Plant Sci., 7*, 280. https://doi.org/10.3389/fpls.2016.00280

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene, 87*(1), 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Yengkhom, S., Uddin, Arif, & Chakraborty, S. (2019). Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis var. assamica and* Camellia sinensis var. *sinensis* in comparison to *Camellia pubicosta*. Journal of Integrative. *Agriculture., 18*(12), 2771–2785. https://doi.org/10.1016/S2095-3119(19)62716-4

You, E., Wang, Y., Ding, Z. T., Zhang, X. F., Pan, L. L., & Zheng, C. (2015). Codon usage bias analysis for the spermidine synthase gene from *Camellia sinensis* (L.). *O. Kuntze. Genetics and molecular research: GMR, 14*(3), 7368–7376. https://doi.org/10.4238/2015.July.3.12

Zeng, L., Watanabe, N., & Yang, Z. (2019). Understanding the biosyntheses and stress response mechanisms of aroma compounds in tea (*Camellia sinensis*) to safely and effectively improve tea aroma. *Critical reviews in food science and nutrition, 59*(14), 2321–2334. https://doi.org/10.1080/10408398.2018.1506907