IN SILICO ANALYSIS OF THE KEY ENZYMES INVOLVED IN SYNTHESIS OF SECONDARY METABOLITES IN CAMELLIA SINENSIS

S. W. Maina, J. K. Kinyua and D. W. Kariuki
College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
E-mail: wairimusylvia@yahoo.com

Abstract
Tea is one of the most popular beverages worldwide, native to Southeast Asia but currently cultivated in over 35 countries. Studies on its chemical composition reveal that polyphenol metabolites account for 25% to 35% of the total dry weight. Tea has many health benefits owing to secondary metabolites whose level of expression in various tea clones determine tea flavor. The flavor (taste and aroma) and the color of processed tea are used to assess its quality and therefore a detailed analysis of key enzymes involved in the synthesis of secondary metabolites is necessary. This study employed a computational approach in the analysis of these enzymes to gain insight into the mechanism of synthesis of these bioactive secondary metabolites. Biological databases were used to retrieve amino acid sequences of these key enzymes. Consensus conserved regions in these sequences were identified from highly identical homologs which were useful in modeling the enzymes' three dimensional structures. A total of 14 key enzymes were analyzed and pockets and cavities in their structures; hence the putative substrate binding sites determined, which gave insight into the enzymes-substrate as well as enzyme cofactor interactions. The preferred orientations of the interactions between substrates and/or co-factors with the enzymes were also simulated through molecular docking. Analysis of these enzymes revealed unique enzyme structures and very specific substrate and co-factor preference. This analysis offers a platform for optimization of selective expression of these key enzymes through gene expression assays that can potentially alter the quality yield of tea clones.

Key words: *Camellia sinensis*, secondary metabolite, conserved regions, pockets and cavities, molecular docking

1.0   Introduction
Tea is one of the most widely consumed beverages in the world second only to water (Wang, *et al*., 2008) as well as one of the most economically important crops. It is native to China, Japan and Southeast Asia. Tea was introduced by the British in India, Sri Lanka, and by the Dutch in Indonesia. Later, in the 20th century commercial production began in Kenya, Tanzania, and Malawi. Currently it is produced by more than 35 countries (Gesimba *et al*., 2005). Tea seedlings were introduced to Kenya from India by G.W.L. Caine in 1903, however, commercial cultivation began in 1924.

The commercial importance of the tea plant (*Camellia sinensis*) is due to its popularity as a refreshing health drink and as a source of important secondary metabolites. The leaves of assamica and sinensis, varieties are used to manufacture tea. The flavor and colour of processed tea is used to assess quality of tea. The non-volatile constituents are responsible for taste while aroma is due to the volatile constituents. A strong attractive aroma is the most important and desirable characteristic of good quality tea. In recent years, tea has attracted more and more attention because of its reported health benefits particularly as an antioxidant and anticarcinogenic. The flavonoids of tea are generally believed to be responsible for these effects.

Over 500 flavour compounds have been identified in tea (Rawat and Gulati, 2008). Tea catechins are most widely studied. They are water-soluble compounds which impart bitterness and astringency to tea. They have been reported to have ant oxidative, ant carcinogenic, antiallergenic, anti-inflammatory, and vasodilatory properties. Catechins are synthesised by the flavonoid biosynthetic pathway starting with phenylalanine as the precursor. Almost all of the characteristics of manufactured tea, including its taste, colour and aroma, have been found to be associated directly or indirectly with catechins (Wang *et al*., 2010).

The aroma of tea is attributed to the volatile flavour compounds (VFC) in tea. Most volatile compounds originate from large precursor molecules present in tea leaves that include products of lipid breakdown, terpenoids and

phenolics, which are present as bound glycosides in tea leaves and are released upon the action of enzymes like glucosidases (Rawat and Gulati, 2008).

Tea processing is known to enhance the release of volatile compounds from bound precursors (Ravichandran, 2002). VFCs derived from terpenoid related compounds are important components of aroma because of their desirable sweet flowery aroma. These VFCs include monoterpene alcohols like linalool and its oxides, geraniol and products of oxidation of carotenoids like ß-ionone. The precursors for the synthesis of monoterpenes and tetraterpenes like carotenoids are provided by the Methylerythritol Phosphate pathway in plastids and precursors for monoterpene and carotenoid synthesis are Geranyl Pyrophosphate and Geranylgeranyl Pyrophosphate respectively.

2.0   Materials and Method
2.1   Enzyme Selection
This study was based on focusing on analysis of the secondary metabolism genes. The secondary metabolism genes, mostly discovered through EST sequencing were obtained from NCBI. The sequences of enzymes responsible for their synthesis were obtained from protein database http://www.uniprot.org/help/uniprotkb). The key enzymes for the study were identified according to Park *et al.,* 2004. The enzymes studied were PAL, C4H, 4CL, CHS, CHI, F3H, F-3,5-H, DFR, FLS, ANS, ANR and ANR2, BPR, LAR and PSY.

2.2   Identification of Consensus Conserved Region
To determine the conserved regions in the enzymes, the products were subjected to BLASTp analysis using Protein Databank to get closely related sequences from other species which have known structures. The highly identical homologs sequences were obtained and aligned using clustalW according to Larkin et al., 2007 to identify the consensus conserved region of each of the enzymes.

2.3   Domain Analysis and Protein Modeling
The 3 dimensional structure of the enzyme structures were modeled using the PHYRE2 according to Kelly and Sternberg 2009 server available at (www.sbg.bio.ic.ac.uk/phyre2/). The submitted protein sequences were first scanned against a large sequence database using PSI-BLAST. The profile generated by PSI-BLAST was then processed by the neural network secondary structure prediction program Psipred and the protein disorder Disopred. The predicted presence of alpha helices, beta strands and disordered regions was represented graphically together with color coded by confidence bar. This was displayed in the final model.

The amino acid sequences of a representative set of all known three-dimensional protein structures were compiled; these sequences were processed by scanning against a large protein sequence database. This resulted in a database of profiles or HMMs, one for each known 3D structure. The user sequence of interest was similarly processed to form a profile/HMM. This user profile was then scanned against the database of profiles using profile-profile or HMM-HMM alignment technique. The alignment techniques also took into account patterns of predicted or known secondary structure elements and they were scored in terms of coverage and confidence to the target sequence. The final model was built using the best 20 domains; best in terms of sequence coverage and the alignment. This was based on the fact that many proteins contain multiple protein domains. PHYRE2 provided a table of template matches color-coded by confidence and indicating the region of the user sequence matched. This aided in the determination of the final protein.

2.4   Binding Pocket Analysis
The predicted 3 Dimensional structures were used to locate the structural pockets and cavities in the proteins. 3Dligandsite server was used to determine protein binding site according to Wass et al., 2009. Structures that were modeled with confidence > 90% were automatically submitted to 3Dligand site. Ligands bound to the new structures similar to the query were superimposed onto the model and were used to predict the binding sites. Structures homologous to the query that have ligands bound were searched. MAMMOTH is used to perform a full structural scan of the modeled structure against a library of protein structures with bound ligands. Upto the top 25 scoring (using MAMMOTH -lnE score) are retained for analysis. These structures are aligned with the modeled structure using TMalign. Single linkage clustering is used to group the ligands. Confirmation of the presence of the

binding sites was done by identifying pockets and cavities. This was done using CASTp server available at http://sts.bioengr.uic.edu/castp. CASTp measures analytically the area and volume of each pocket and cavity.

2.5   Docking

Computational simulation of a candidate ligand binding to the enzyme (receptor) was done to show the preferred orientation of both molecules when they are interacting. This interaction was aimed at predicting the association/ affinity between the two molecules. The approach used in docking these ligands was shape complementarity. Docking was done using Molsoft available at (www.molsoft.com)

3.0   Results and Discussion

3.1   Protein Structure Prediction and Modeling

The 3 dimensional structure of the protein as modeled using the PHYRE2 according to Kelly and Sternberg 2009 server available at (www.sbg.bio.ic.ac.uk/phyre2/). The top 20 best hits in the alignment of each of the target enzymes sequences are used in modeling each of the enzymes to avoid computer load. The percentage confidence of the final model is determined from the templates and the percentage coverage is calculated based on the residues match between the target and the template that are used in modeling. The PSY enzyme has 4 templates that are selected to model the protein based on heuristics to maximise confidence, percentage identity and alignment coverage. It is the only low quality model less than 70% confidence. The model is modeled using ab initio approach where only 46 residues (64%) are modeled as shown in Table 1.

Table 1: Templates used in modeling various proteins using PHYRE 2 server based on PDB templates in PHYRE 2 server with the percentage confidence and coverage

Enzyme	Template and its PDB information	Confidence (%)	Coverage (%)
PAL	D1W27A, l-aspartate like fold, SUPERFAMILY: L-aspartate like, FAMILY: PAL/HAL	100	96
C4H	C2Q9FA an oxidoreductase used Chain: A of : PDB molecule: cytohrome P450 46a1	100	94
4CL	c3ni2A. a ligase, Chain: A of crystal structures of a populous tomentosa 4-2 coumarate: COA ligase was used.	100	89
CHS	C3TSYA. A ligase, transferase, Chain: A of fusion protein 4-coumarate COA ligase 1, reveratrol, PDB title: 4-coumaryl-COA ligase: stilbene synthase fusion protein was used	100	99
CHI	C4dOOA an ISOMERASE, Chain:A of chalcone-flavanone isomerase family protein was used. PDB title: the crystal structure of Arabidopsis thaliana fatty-acid binding protein2 at 3g63170 (atfap1)	100	77
F3H	d1gp6A, FOLD was a double stranded beta helix, SUPERFAMILY: clavaminate synthase-like, FAMILY: penicillin synthase-like	100	89
F-3,5-H	PDB template was c3ebsA. An oxidoreductase, Chain:A PDB Molecule: cytochrome p450 2a6	100	88
DFR	C2iodD an oxidoreductase , Chain: D: PDB Molecule: DIHYDROFLAVONOL 4-REDUCTASE;PDB Title: Binding of two substrate analogue molecules to 2 dihydroflavonol-4-reductase alters the functional geometry 3 of the catalytic site	100	93
FLS	d1gp6A, FOLD: is a double stranded beta helix, SUPERFAMILY: clavaminate synthase-like, FAMILY: penicillin synthase-like	100	98
ANS	D1GP6A, FOLD is double stranded beta helix, SUPRERFAMILY: clavaminate synthase-like, FAMILY: penicilin synthase like.	100	98
ANR	C1Z45A, an isomerase, Chain: A PDB Molecule gal10 biofunctional protein; was used.	100	92
BPR	c3ptkB. PDB header: hydrolase, Chain:B PDB Molecule: Beta-glucosidase	100	93

os4glu12, PDB title: the crystal structure of rice (oryza sativa 1.) os4bglu12.

LAR	C1Z7EC. The template is a hydrolase.		100	90
	Chain: C: PDB Molecule: protein arna;			
	PDB Title: crystal structure of full length arna			
PSY	1) d1ezfA, FOLD: Terpenoid synthase, Superfamily: terpenoid synthases, Family: squaline synthase		90	64
	2) c4hd1A, PDB Header: transferase, Chain: A: PDB Molecule: Squaline synthas hpnc			
	3) C3p5rB, PDB header: lyase, Chain: B: PDB Molecule: taxadiene synthase			
	4) C2zcpA, PDB header:transferase Chain: A: PDB Molecule:dehydrosqualene synthase;			

3.2   Alignment of the Domains that were used to Establish Each Enzymes Model

Table 2: aligned regions for PAL, C4H and 4CL. The tables below show the domains used to establish the enzyme model by aligning the top 20 best hits that match highly with the target sequence. The confidence key in table 6 is used to show the probability with which the target sequence matches the available structures in terms of the different colors.
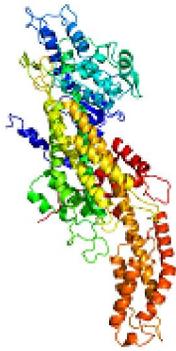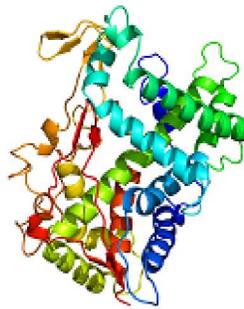
Aligned regions for PAL

Aligned regions for PAL		Aligned regions for C4H		Aligned regions in 4CL	
1	d1w27a				
2	c3nz4A	1	c2q9f A	1	c3ni2 A
3	c4babC	2	c2f9q A	2	d1pg4 a
4	c1t6pF				
5	d1t6ja	3	c3pm 0A	3	d1ry2 a
6	d1gkma				
7	c3czoD	4	c3e4e A	4	c2vsq A
8	c3unvB				
9	c2qveA	5	d1nr6 a	5	c3tsy A
10	c2o6yF	6	c2hi4 A	6	c3e7w A
11	c2nynD				
12	c2lmdA	7	d1r9o a	7	d3cw9 a1
13	c1kkoB	8	d1po5 a	8	d1md ba
14	c4hecB				
15	c1yvyA	9	d3czh a1	9	c2d1t A
16	d2olra1				
17	d1mija	10	c3na0 B	10	c3r44 A
18	d1z0sa1	11	c2x2n B	11	c3etc B
19	c1z0zC				
20	d7aata	12	c3ebs A	12	c4dg8 A
		13	d2nnj	13	d1lcj

3	a1
14	c3rukD_
15	d1tgna_
16	c3k9vB_
17	c3eqmA_
18	c3juvA_
19	c3hf2A_
20	d2ij2a1

3	a_
14	c3gqwB_
15	c3eynB_
16	d1amua_
17	c4fuqD_
18	c4dg9A_
19	c3rg2H_
20	c1amuB_

Table 3: Aligned regions for CHS, CHI and F3H. . The tables below show the domains used to establish the enzyme model by aligning the top 20 best hits that match highly with the target sequence. The confidence key in table 6 is used to show the probability with which the target sequence matches the available structures

Aligned region for CHS	Aligned regions in CHI	**Aligned region in F3H**
1 c3tsy A_	1 c4doo A_	1 d1gp6 a_
2 c1cml A_	2 d1eyq a_	2 c3oox A_
3 c2d3m A_	3 c4dol A_	3 d1odm a_
4 c3ale B_	4 c4doi B_	4 d1w9y a1
5 d1bi5 a1	5 c4dok A_	5 d1dcs a_
6 c3v7i A_	6 c2yue A_	6 c3on7 C_
7 c3oit B_	7 d1o9y a_	7 c2g19 A_
8 c2p0u B_	8 c3uep B_	8 c3oui A_
9 c3ov3 A_	9 d1o6a a_	9 c3dkq B_
10 c1xet D_	10 c1zx6 A_	10 c3bvc A_
11 c1ee0 A_	11 c2eyx A_	11 d2iuw a1
12 d1u0u a1	12 c2bz8 B_	12 c2iuw A_
13 c3a5q A_	13 c1z9q A_	13 c3gjb A_
14 d1u0v a1	14 c2e63 A_	14 c3itq B_
15 c1u0m A_	15 d1i07 a_	15 c3btz A_
16 c2h84 A_	16 c4glm D_	16 c2dbi A_
17 c3e1h A_	17 c2dbk A_	17 d2csg a1
18 c3euo B_	18 c2nwm A_	18 c2jij A_
19 d1ee0 a1	19 d1uue a_	19 d2fdi a1
20 d1u0m a1	20 c2d1xD	20 c2opw A_

55

Table 4: Aligned regions for F-3,5-H, DFR and FLS. . The tables below show the domains used to establish the enzyme model by aligning the top 20 best hits that match highly with the target sequence. The confidence key in table 6 is used to show the probability with which the target sequence matches the available structures

Aligned regions in F-3,5-H	Aligned regions in DFR	Aligned regions in FLS
1 c3ebs A_	1 c2iod D_	1 d1gp6 a_
2 d1nr6 a_	2 c1z7e C_	2 d1odm a_
3 d2nnj a1	3 c1z45 A_	3 c3oox A_
4 c3e4e A_	4 d1i24 a_	4 d1w9y a1
5 d3czh a1	5 c2p4h X_	5 d1dcs a_
6 d1po5 a_	6 c2rh8 A_	6 c3on7 C_
7 d1r9o a_	7 d2c5a a1	7 c3oui A_
8 d2ij2 a1	8 d1y1p a1	8 c2g19 A_
9 c2iag A_	9 c2v6g A_	9 c3dkq B_
10 c2hi4 A_	10 d1oc2 a_	10 d2iuw a1
11 d1tqn a_	11 c2b69 A_	11 c2iuw A_
12 c2x2n B_	12 d2b69 a1	12 c3btz A_
13 c3na0 B_	13 c2x4g A_	13 c3itq B_
14 c3k9v B_	14 c2hun B_	14 c3bvc A_
15 c3g1q C_	15 c4egb C_	15 c3pvj B_
16 d2cib a1	16 d2bll a1	16 c2dbi A_
17 c3hf2 A_	17 d1r6d a_	17 d1otj a_
18 c3eqm A_	18 d1kew a_	18 d2fdi a1
19 c2f9q A_	19 d1e6u a_	19 c2opw A_
20 c3juv A_	20 c3enk B_	20 c3tht B_

Table 5: Aligned regions in ANS, ANR and ANR2. . The tables below show the domains used to establish the enzyme model by aligning the top 20 best hits that match highly with the target sequence. The confidence key in table 6 is used to show the probability with which the target sequence matches the available structures

Aligned region in ANS	Aligned region in ANR	Aligned region in BPR
1 d1gp6a	1 c1z45A	1 c3ptkB
2 d1odma	2 c1z7eC	2 c2rgmA
3 d1w9ya1	3 c2rh8A	3 c2dgaA
4 c3ooxA	4 d1i24a	4 d1cbga
5 d1dcsa	5 c2iodD	5 c3gnoA
6 c3on7C	6 c2p4hX	6 c3u57A
7 c2g19A	7 d2c5aa1	7 d1v08a
8 c3dkqB	8 c2v6gA	8 c4a3yA
9 c3ouiA	9 d1oc2a	9 d1e4mm
10 c3btzA	10 d2b69a1	10 d1v02a
11 d2iuwa1	11 c2b69A	11 c1v02F
12 c2iuwA	12 c1n7gB	12 c2jf7B
13 c3thtB	13 c3enkB	13 d1qoxa
14 c3itqB	14 d1r6da	14 d2j78a1
15 c2dbiA	15 d1y1pa1	15 c3fiyA
16 c3bvcA	16 d1kewa	16 c3ahyD
17 d2csga1	17 c2hunB	17 c3ahxC
18 d2fdia1	18 c2p5uC	18 c4b3kA
19 c2jijA	19 d1n7ha	19 d1gnxa
20 c3ms5A	20 d1e6ua	20 c3ai0A

Table 6: Aligned regions in BPR, PSY and the confidence key for alignment. . The tables below show the domains used to establish the enzyme model by aligning the top 20 best hits that match highly with the target sequence. The confidence key in table 6 is used to show the probability with which the target sequence matches the available structures

Aligned region in LAR	Aligned region in PSY	Confidence Key
1 c1z7eC	1 c2zcpA	
2 c1z45A	2 c4hd1A	
3 c2rh8A	3 d1ezfa_	
4 d1i24a	4 c1wy0A	
5 c2iodD	5 c3mzvB	**High** **Low**
6 c2p4hX	6 c3aq0G	
7 d1oc2a	7 c3oacD	**(9)** **(0)**
8 c2b69A	8 c3aqbD	
9 d2b69a1	9 c3oyrB	
10 c2v6gA	10 c3ts7B_	
11 c1n7gB	11 c3n3dB	
12 c3enkB_	12 c3nf2A	
13 d2c5aa1	13 c3m9uD	
14 d1y1pa1	14 c3lk5A	
15 d1rpna_	15 c4dhdA	
16 d1n7ha	16 c1wmw_	
17 c2p5uC	17 d2q80a1	
18 c3lu1C_	18 c3lmdA	
19 c2z1mC_	19 c2e8xB_	
20 c2hrzA	20 c2azjB	

58

3.3   Protein Models

All enzyme models in Figure 1 except PSY Figure 1 (o) are modeled using the automated PHYRE approach giving models that had 70% and above coverage. These are good quality models.

(a) PAL

(b) C4H

(c) 4CL

(d) CHS
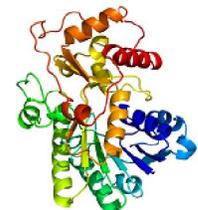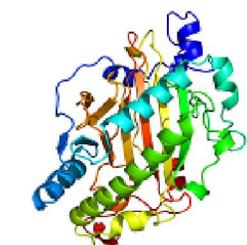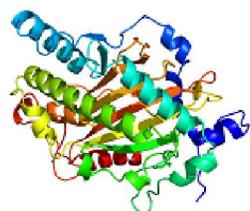
(e) CHI

(f) F3H

(g) F-3,5-H

(h) DFR

Transmembrane helices were predicted in sequence for ANR to adopt the topology
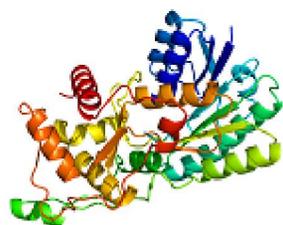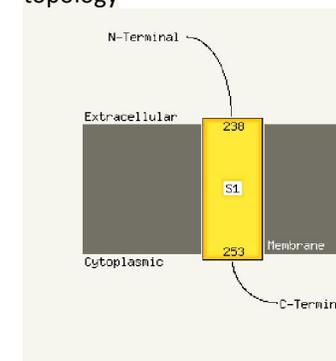
(i)	FLS
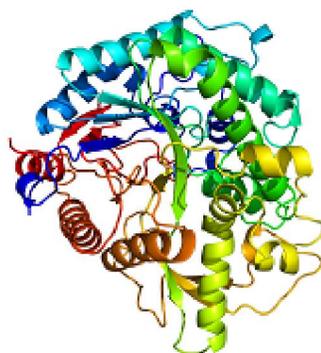
(j) ANS
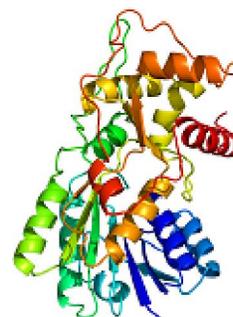
(k) ANR
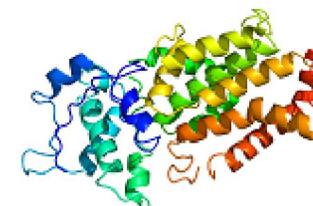
(l) ANR2

(m) BPR

(n) LAR

(o) PSY

Figure 1: Summary of all enzymes as predicted by phyre2 server
(a)PAL, (b) C4H, (c) 4CL, (d) CHS, (e) CHI, (f) F3H, (g) F35H, (h) DFR, (i) FLS, (j) ANS, (k) ANR, (l) ANR2, (m) BPR, (n) LAR, (0) PSY

The 3DLigandSite server was used for protein binding site prediction. Confident models produced by Phyre2 (confidence >90%) were automatically submitted to 3DLigandSite. This happened for all enzymes except PSY that was modeled through de novo approach and thus was submitted manually.

3.4   Cavities Identified In CASTP Server

Pockets are empty concavities on a protein surface into which solvent can gain access. Binding sites and active sites of proteins and DNAs are often associated with structural pockets and cavities. It provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules. It measures analytically the area and volume of each pocket and cavity, both in solvent accessible (SA) surface and molecular surface (MS). The calculation uses a solvent probe of radius 1.4 angstrom.

Table 5: Summary of the five biggest pockets in each enzyme, their areas and the volume they occupy

Enzyme	Area	Volume	Enzyme	Area	Volume
PAL	438.3	1289.8	FLS	1703.5	2611.1
	520.3	677		693.6	1160.7
	482.5	626		236.5	214.1
	398.7	649.5		234.5	238.9
	387.4	393.3		141.8	123.1
CH4	3206.4	4447.3	ANS	2184.2	3904.9
	956.1	907.5		286.6	301.1
	373.4	417.7		149.3	122.9
	317.6	309.5		65.9	48.3
	236	236.2		118	91.6
4CL	2015.2	2752.7	ANR	952.3	1731.4
	496.1	1119.3		399	579.9
	261.9	490.1		452	600
	226.2	349.7		284.1	344.8
	259.6	269		171.9	174.8
CHS	394.5	734.7	ANR2	1431.8	1940
	148.7	133.1		634.5	976.5
	139.9	133.5		206.8	183.7
	109	128.8		145.8	196.6
	159.9	133.4		121.6	153.2
CHI	1001.5	1283.1	BPR	498.2	885.5
	600.9	938		287.7	293.6
	83.1	68.9		302.7	304.9
	95	70.1		172.6	178.6
	116	136.9		165.6	164.5
F3H	2605.8	4337.1	LAR	1033.9	1390.4
	366.4	398.1		303.8	368.1
	304.9	452		160.8	386.2
	183.4	179.4		121.6	153.2
	119.4	150.3		122.8	144.7
F35H	1737.1	3069.2	PSY	3490.1	4851
	208.3	431.7		1457.5	2959.7
	218.7	185.2		281.3	293.1
	213.1	215.5		213.7	163.8
	134.5	217		162	107.1
DFR	1616.3	2807.7			
	146.2	408.4			
	91.6	106.8			
	66.5	60.7			
	117.5	84.6			

3.5　　　　Enzyme Models, Pockets and Interactions with Substrate and Cofactors
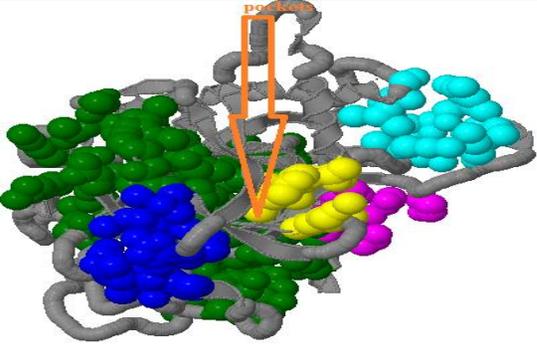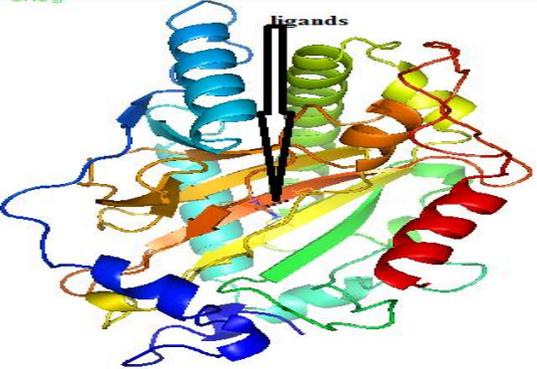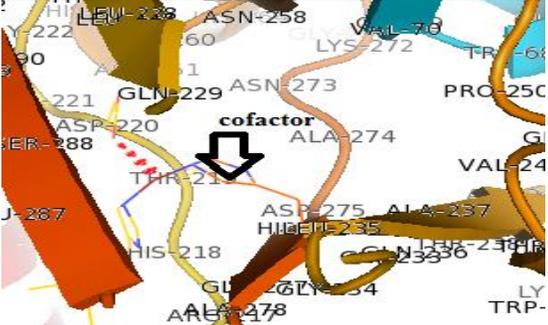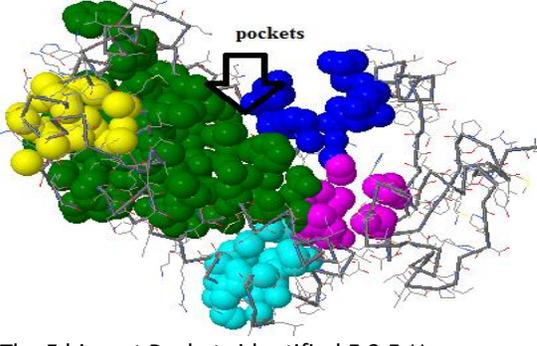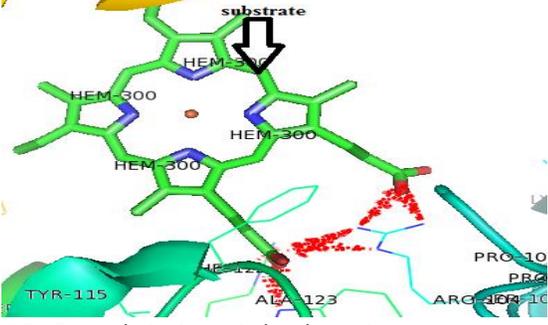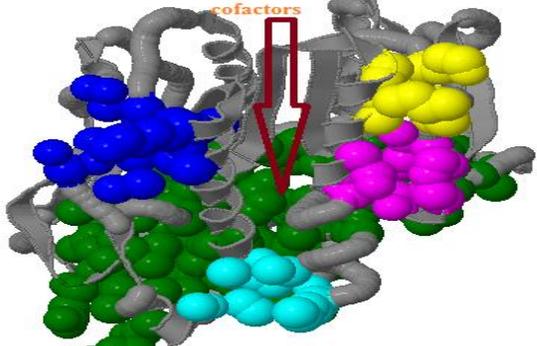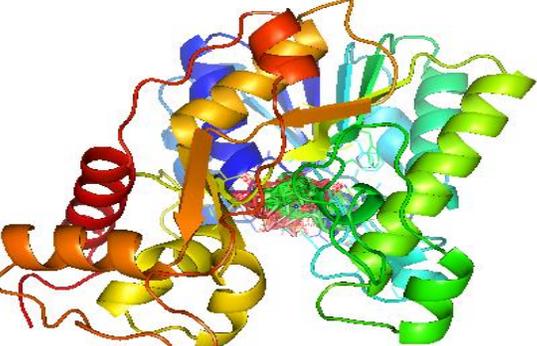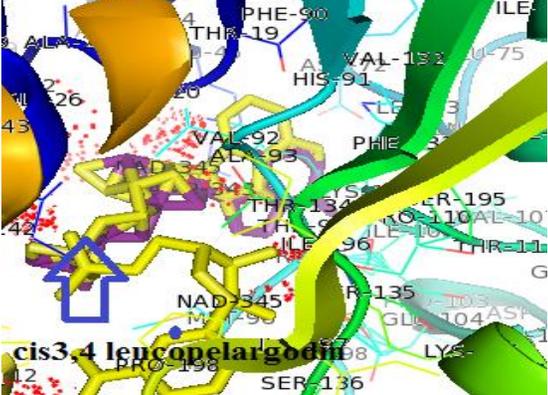
Docking is a method that produces the preferred orientation of one molecule to the second when both are bound together forming a complex. The association between biologically relevant molecules in this case enzymes and their substrates or enzymes and their cofactors play a key role in catalytic reaction. The enzymes are the hosts / receptors that receive the molecule. Ligands are the complimentary molecules which get bound to the receptors. They are the substrates and cofactors that act on each enzyme in the catalytic reactions to release products.
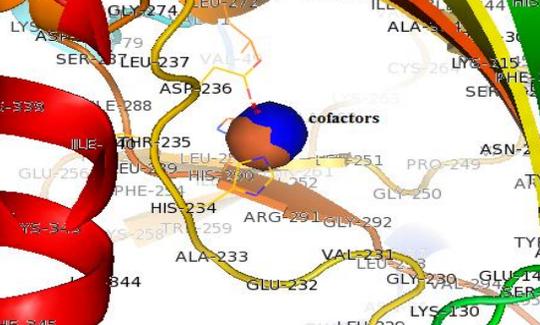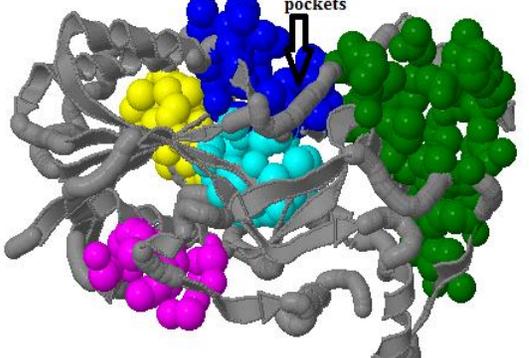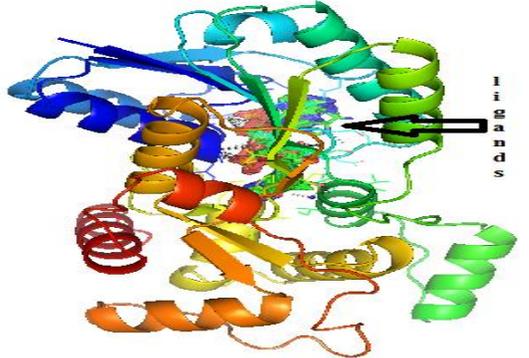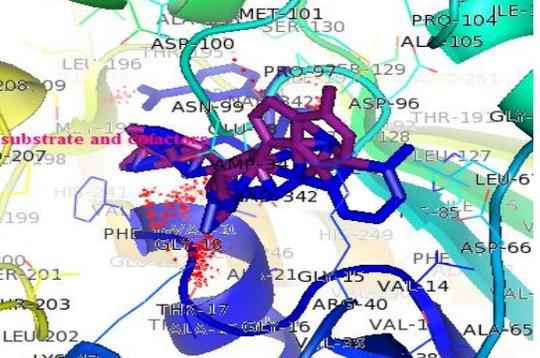
Column one indicates the pockets identified in each of the modeled structure of the enzymes, column two displays the ligands that bind to these pockets and column three shows the specific interaction of the enzyme with its substrates/ cofactors. Docking indicated results of computational simulation of interaction between the enzymes and their substrate and cofactors in their binding sites and the contact residues that the ligands attached to.

Table 6: enzyme structures with their ligands in their predicted cavities and the interactions

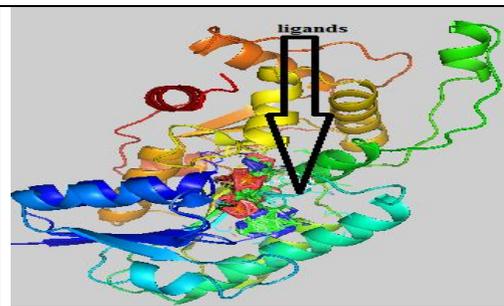|		
The 5 biggest Pockets identified in PAL	PAL superimposed with its ligands	Interaction of PAL with l phenylalanine in its binding sites
The 5 biggest Pockets identified in C4H	C4H superimposed with its ligands	Interaction of C4H cinammate in its binding sites
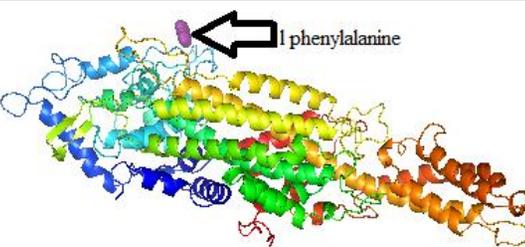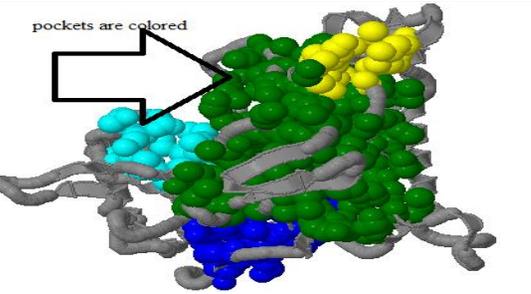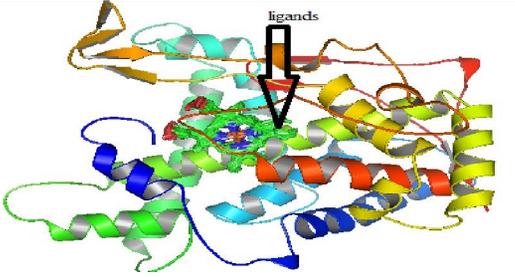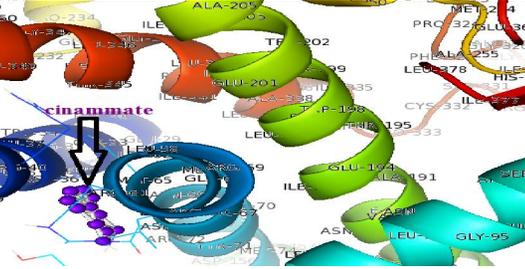
The 5 biggest Pockets identified in 4CL

4 CL enzyme with its ligands

interaction of 4 CL with 4-coumarate in its binding sites

The 5 biggest Pockets identified in CHS

CHS enzyme with its ligands

interaction of CHS with malonyl COA in its binding sites

The 5 biggest Pockets identified in CHI

Interaction of CHI with chalcone

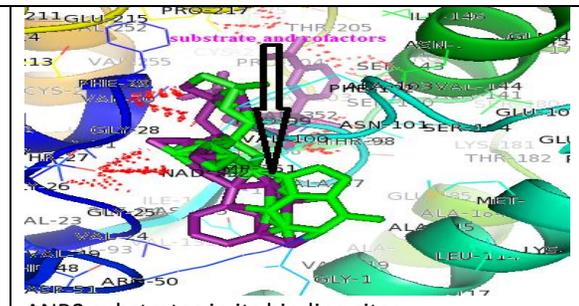|	CHI enzyme with its ligands	
The 5 biggest Pockets identified in F3H	F3H with its ligands	F3H cofactors in its binding sites
The 5 biggest Pockets identified F-3,5-H	F-3,5-H enzyme with its ligands	F-3, 5-H substrate in its binding site
The 5 biggest Pockets identified in DFR	DFR enzyme with its ligands	cis3,4 leucopelargodin
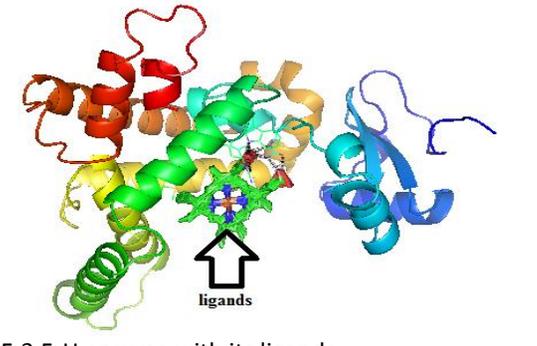
64

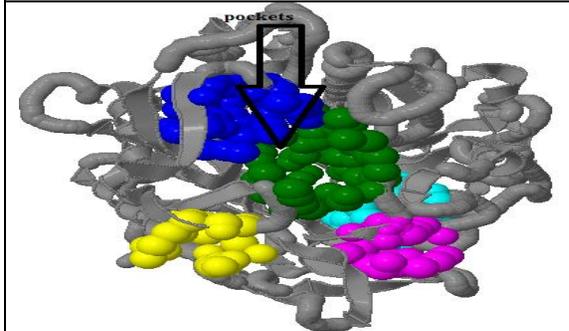|		DFR with its substratein its binding sites
The 5 biggest Pockets identified in FLS	FLS enzyme with its ligands	FLS with its cofactors in its binding sites
The 5 biggest Pockets identified in ANS	ANS enzyme with its ligands	ANS with its cofactors in the binding sites
The 5 biggest Pockets identified in ANR	ANR enzyme with its ligands	ANR substrates in the binding sites
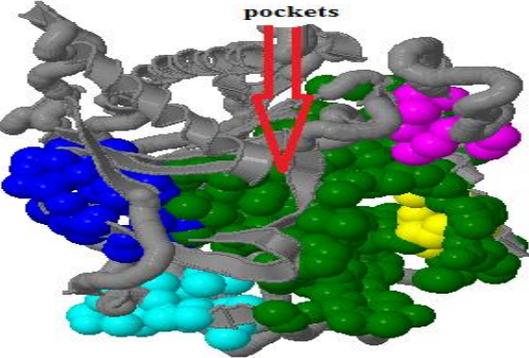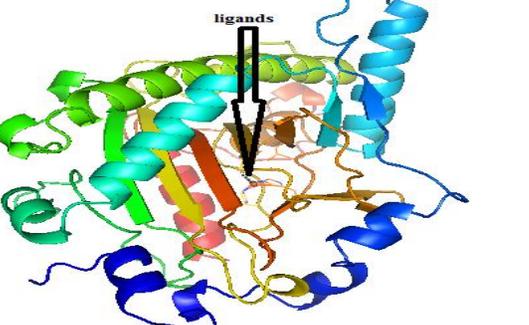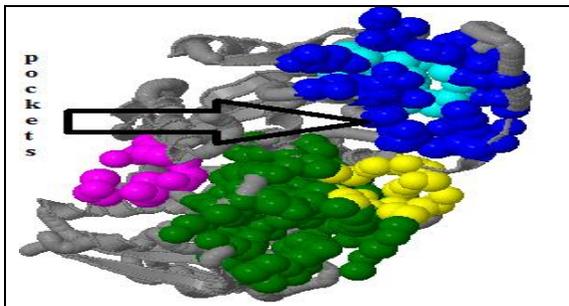
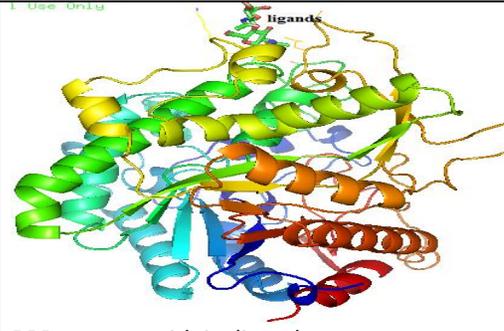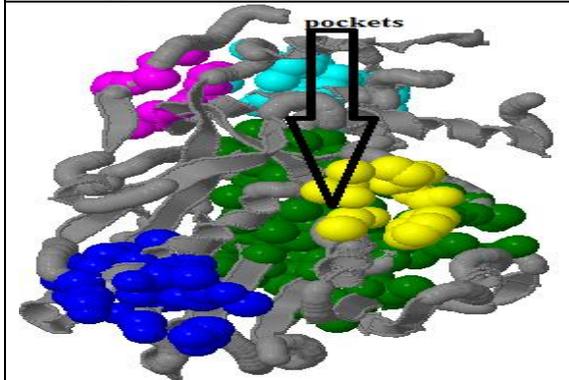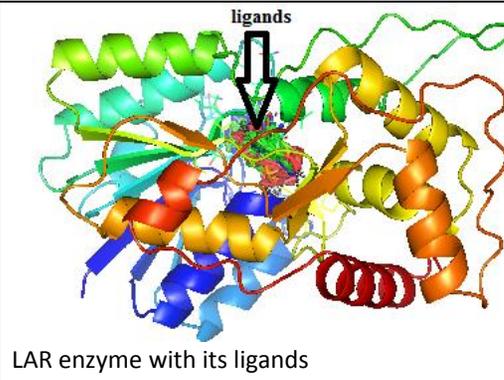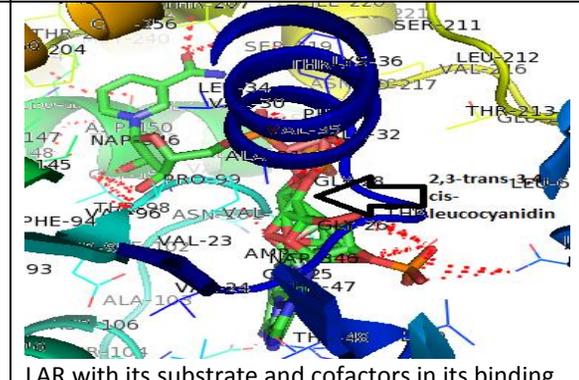|		
The 5 biggest Pockets identified in ANR2	ANR2 enzyme with its ligands	ANR2 substrates in its binding sites
The 5 biggest Pockets identified in BPR	BPR enzyme with its ligands	BPR with its cofactors in the binding sites
The 5 biggest Pockets identified in LAR	LAR enzyme with its ligands	LAR with its substrate and cofactors in its binding site.

The 5 biggest Pockets identified in PSY

PSY enzyme with ligands superimposed

PSY with a cofactor in its binding site

Biggest   smallest
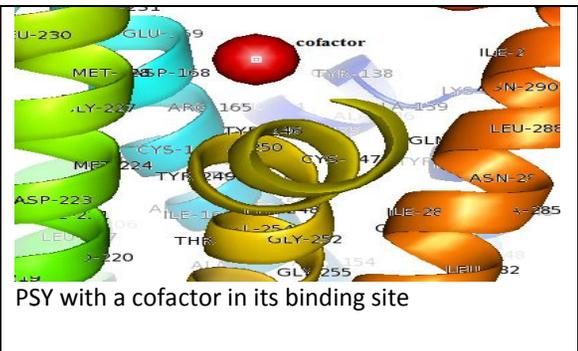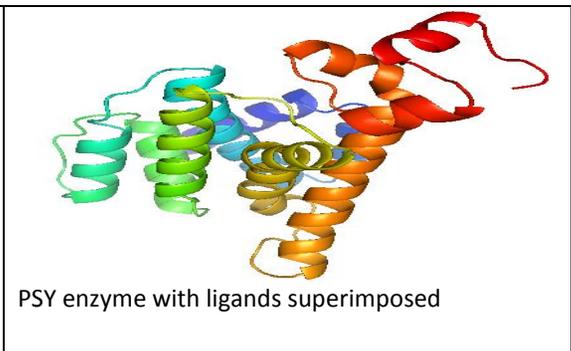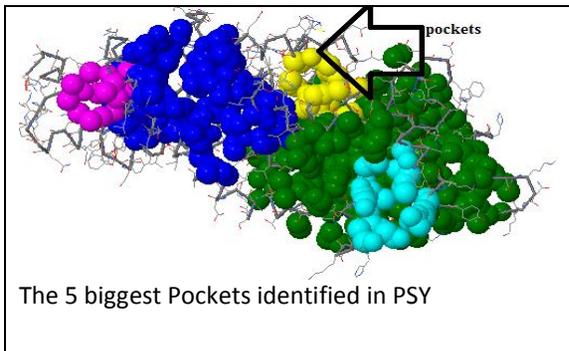
4.0   Conclusion

Tea is a popular beverage as a source of beneficial secondary metabolites. The bestselling tea is believed to be high quality tea owing to synthesized secondary metabolites. Tea however requires long conventional breeding time thus; it is not really advisable to improve crop varieties. From this study it is clear that the secondary metabolites in tea are synthesized as a result of action of some enzymes. The modeled three-dimensional structures of these enzymes are related to their functions. The modeled structure aided in the identification of the putative substrate binding sites which indicates that there is an interaction between enzyme-substrate and enzyme-cofactor. Docking simulated a candidate ligand binding into the receptor indicating that the substrates and cofactors bind into the active sites of the ligand. This interaction leads to catalytic action resulting onto various products of the biosynthetic pathways. This study provides a valuable insight into the mechanism of action of enzymes aiding in the ultimate aim of improving tea quality and enhance the beneficial health properties. It therefore forms a basis of improving the quality of tea computationally rather than using the long conventional breeding approach.

References

Kelley, L. A. and Sternberg, M. J. E. (2009). *Nature Protocols,* **4**: pp 363 - 371 [pdf] [Import into BibTeX]

Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. and Subramaniam, S. (1998b). Analytical shape computation of macromolecules. II. Identification and computation of inaccessible cavities in proteins.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal. W. and Clustal, X. version 2.0. *Bioinformatics,* **23** (21): ppp 2947-2948.

 Proteins: Struct. Funct. Genet. **33**:18-29.

Liang,(2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl. Acids Res*., **34:**ppW116-W118.

Luczaj, W., Skrzydlewska, E. (2005). Antioxidative properties of black tea. *American Journal ofPreventive Medicine.* **40**(6):pp910-918.

Park, J. S., Kim, J. B., Hahn, B. S., Kim, K. H., Ha, S. H., Kim, Y. H. (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization.

Rawat, R. and Gulati, A. (2008). Seasonal and clonal variations in some major glycosidic bound volatiles in Kangra tea (Camellia sinensis (L.) O. Kuntze*).* Journal *European Food Research* and *Technology,***226**(6):pp 1241-1249. Tanaka, J., Taniguchi, F. (2006). Estimation of the genome size of tea (*Camellia sinensis*), *camellia* (*C. japonica*), and their interspecific hybrids by flow cytometry. *Journal of the Remote Sensing Society of Japan,* **101**: pp 1–7.

Wang, Y., Jiang, C. J., Zhang, H. Y. (2008). Observation on the Self-incompatibility of Pollen Tubes in Self-pollination of Tea Plant in Style *in vivo*.**28:**pp429–435.

Wang, X. W., Luan, J. B., Li, J. M., Bao, Y. Y., Zhang, C. X. and Liu, S. S. (2010). *De novo* characterization of a whitefly transcriptome and analysis its gene expression during development. *BMC Genomics*, **11**: pp 400*.*

Wass, M. N., Kelley, L. A. and Sternberg, M. J. (2010). Nucleic Acids Research 38, W469-73 [pubmed]