Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online)
Vol.4, No.8, 2014

Computational Analysis of Some Enzymes Involved in Synthesis of Secondary Metabolites in *Camellia Sinensis*

Maina sylvia wairimu
Jomo Kenyatta University of Agriculture and Technology
(M. Sc Molecular Biology and Bioinformatics),P.O BOX 62000-00200, NAIROBI
EMAIL: wairimusylvia@yahoo.com / wairimusylvia@gmail.com

Johnson kinyua K
Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200, NAIROBI
EMAIL: jkkinyuafr2001@yahoo.fr

Daniel kariuki W
Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200, NAIROBI
EMAIL: dkariuki@jkuat.ac.ke

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. Enzyme PAL (phenylalanine ammonia-lyase) a key enzyme in the phenylpropanoid pathway, playing an important role in the plant development and defense. C4H (cinnamte-4-hydroxylse) an important enzyme in allocating significant amounts of carbon from phenylalanine into the biosynthesis of several metabolites, It maintains activities of the metabolic flux for the operation of the flavanoid pathway. 4CL (4-coumarate: COA ligase) the last enzyme in the general phenylpropanoid pathway that provides precursors for the biosynthesis of a large variety of plant natural products like COA thiol esters of 4-coumarate and other hydroxycinnamate. FLS (flavonol synthase) a key enzyme in flavonol synthesisthat determines the final content of flavonols which play an important role in defense related functions and as potent antioxidants. ANS (anthocyanidin synthase) an enzyme in the biosynthetic pathway to anthocyanin. This study employed a computational approach in the analysis of some 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 5 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.
Keywords: camellia sinensis, Secondary metabolites, Conserved regions, Pockets and cavities, Molecular docking

1 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)

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 (Way *et al*., 2009). The leaves of *assamica* and *sinensis,* varieties are used to manufacture tea. The flavor and colour of processed tea is determined 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 (Luczaj and Skrzydlewska, 2005) and anticarcinogenic (Way *et al.,* 2009). 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 antioxidative, anticarcinogenic, antiallergenic, antiinflammatory, and vasodilatory properties (Way *et al.,* 2009). Catechins are synthesized 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 (Ravichandran, 2002). 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 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, FLS and ANS.

2.2 Identification of Consensus Conserved Region

To determine the conserved regions in the enzymes, the products were subjected to $BLAST_p$ 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. A 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 complimetarity. Docking was done using Molsoft available at (www.molsoft.com)

3 Results and discussion

Protein structure prediction and modeling

The 3 dimensional structure of the protein as modeled using the top 20 best hits in the alignment of each of the target enzymes sequences to avoid computer load. The percentage confidence of the final model was 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 3DLigandSite server was used for protein binding site prediction. Confident models produced by Phyre2 (confidence >90%) were automatically submitted to 3DLigandSite.

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

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.

CONCLUSION

Tea is a popular beverage as a source of beneficial secondary metabolites. The best selling 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

Gesimba, R. M. *et al*,. (2005). the Tea Industry in Kenya; The Challenges and Positive Developments, *Journal of Applied Sciences* 5 (2): 334-336.

Kelley LA and Sternberg MJE. (2009) *Nature Protocols* 4, 363 - 371

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics,* 23 (21): 2947-2948. *Proteins: Struct. Funct. Genet.* 33:18-29.

Luczaj W, Skrzydlewska E (2005). Antioxidative properties of black tea. *American Journal of Preventive Medicine.* 40(6): 910-918.

Park JS, Kim JB, Hahn BS, Kim KH, Ha SH, Kim YH (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization.

Rawat R, 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): 1241-1249.

Wang Y, Jiang CJ, Zhang HY (2008). Observation on the Self-incompatibility of Pollen Tubes in Self-pollination of Tea Plant in Style *in vivo*.28:429–435.

Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS (2010). *De novo* characterization of a whitefly transcriptome and analysis its gene expression during development. *BMC Genomics*;11:400.

Wass MN, Kelley LA and Sternberg MJ (2010) Nucleic Acids Research 38, W469-73 [pubmed]

Way TD, Lin HY, Hua KT, Lee JC, Li WH, Lee MR, Shuang CH, Lin JK (2009). Beneficial effects of different

tea flowers against human breast cancer MCF-7 cells. *Food Chemistry* 114(4)

Table 1: model templates

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
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
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

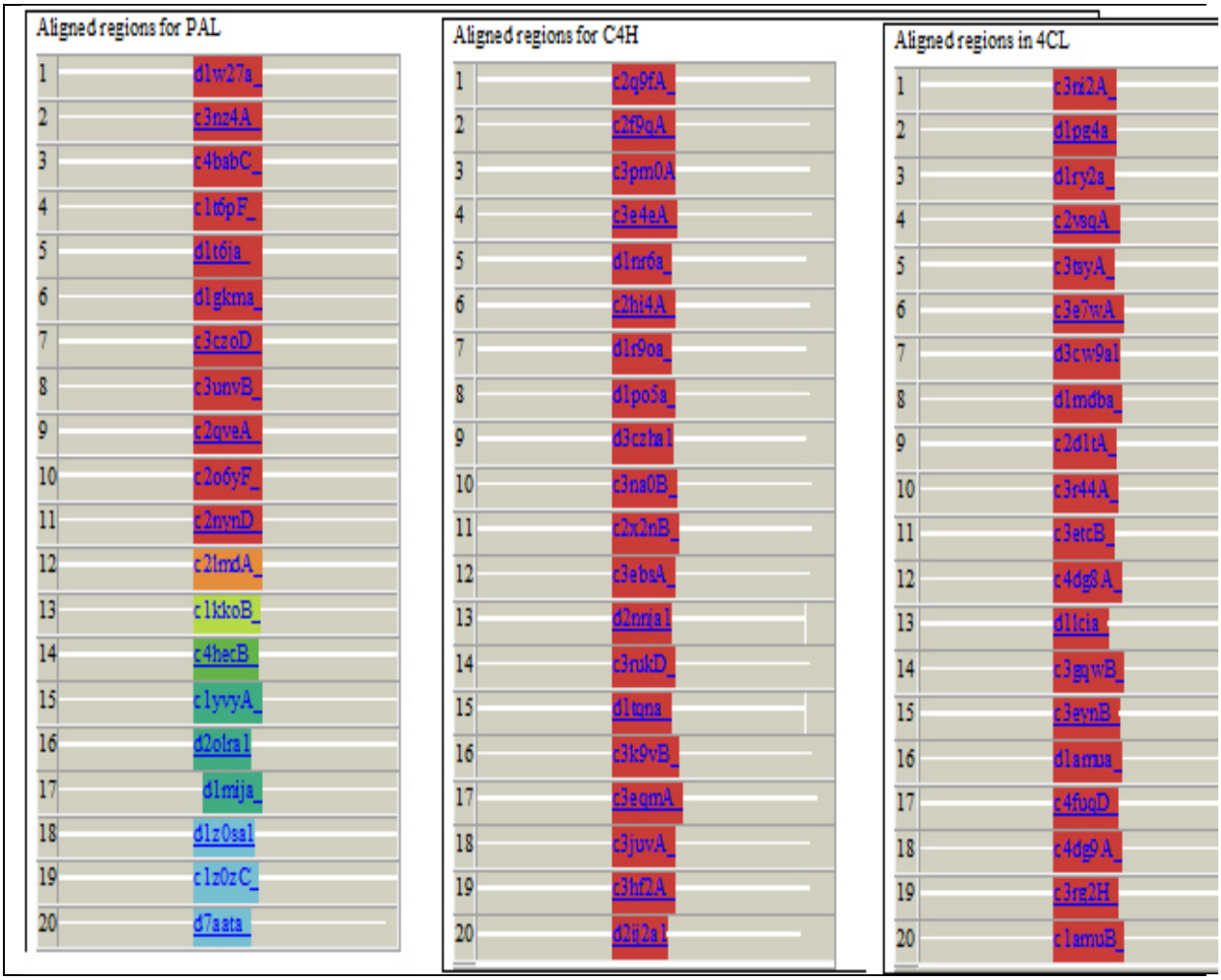
The table above shows templates used in modeling various proteins using PHYRE2 server based on PDB templates in PHYRE 2 server with the percentage confidence and coverage.

Table 2: enzyme pockets

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
C4H	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	**CHI**	1001.5	1283.1
	496.1	1119.3		600.9	938
	261.9	490.1		83.1	68.9
	226.2	349.7		95	70.1
	259.6	269		116	136.9

The table above is a summary of the five biggest pockets in each enzyme, their areas and the volume they occupy

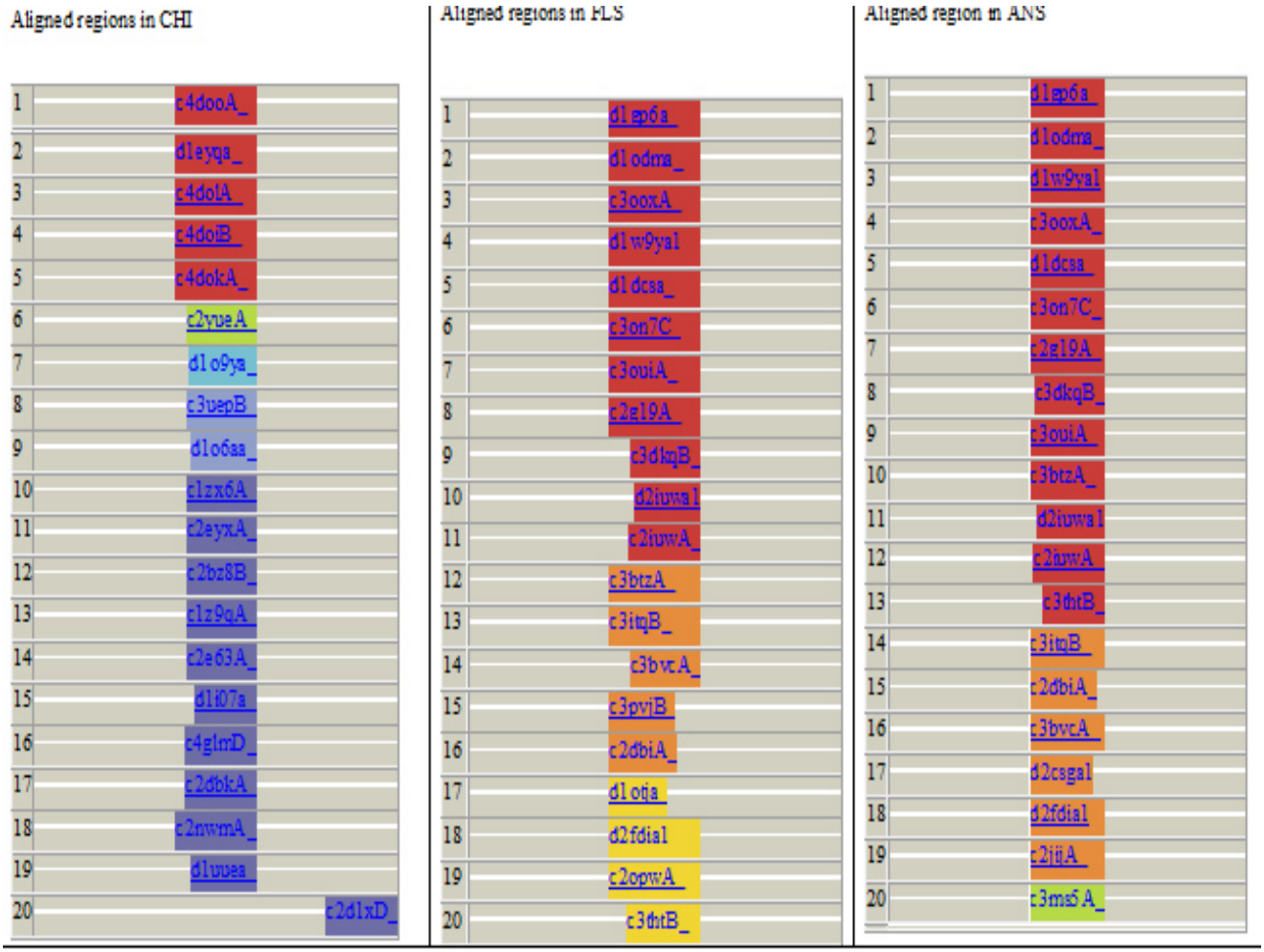
Table 3: aligned regions for PAL, C4H and 4CL.



Alignment of the domains that were use to establish each enzymes model

The tables above 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 4 below is used to show the probability with which the target sequence matches the available structures in terms of the different colors

Table 4: Aligned regions for CHI, FLS and ANS.



The tables above 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 below is used to show the probability with which the target sequence matches the available structures

Confidence Key

(High) **(low)**

Protein models

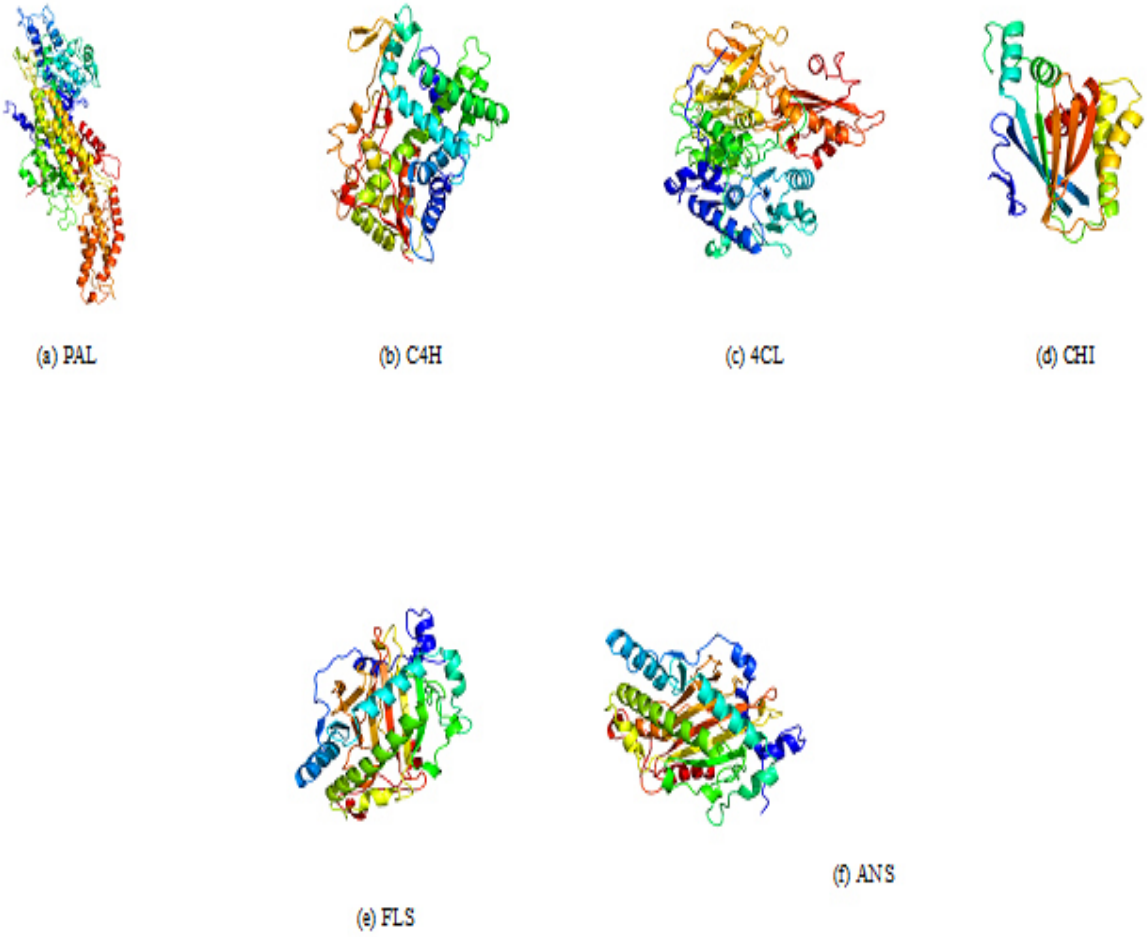
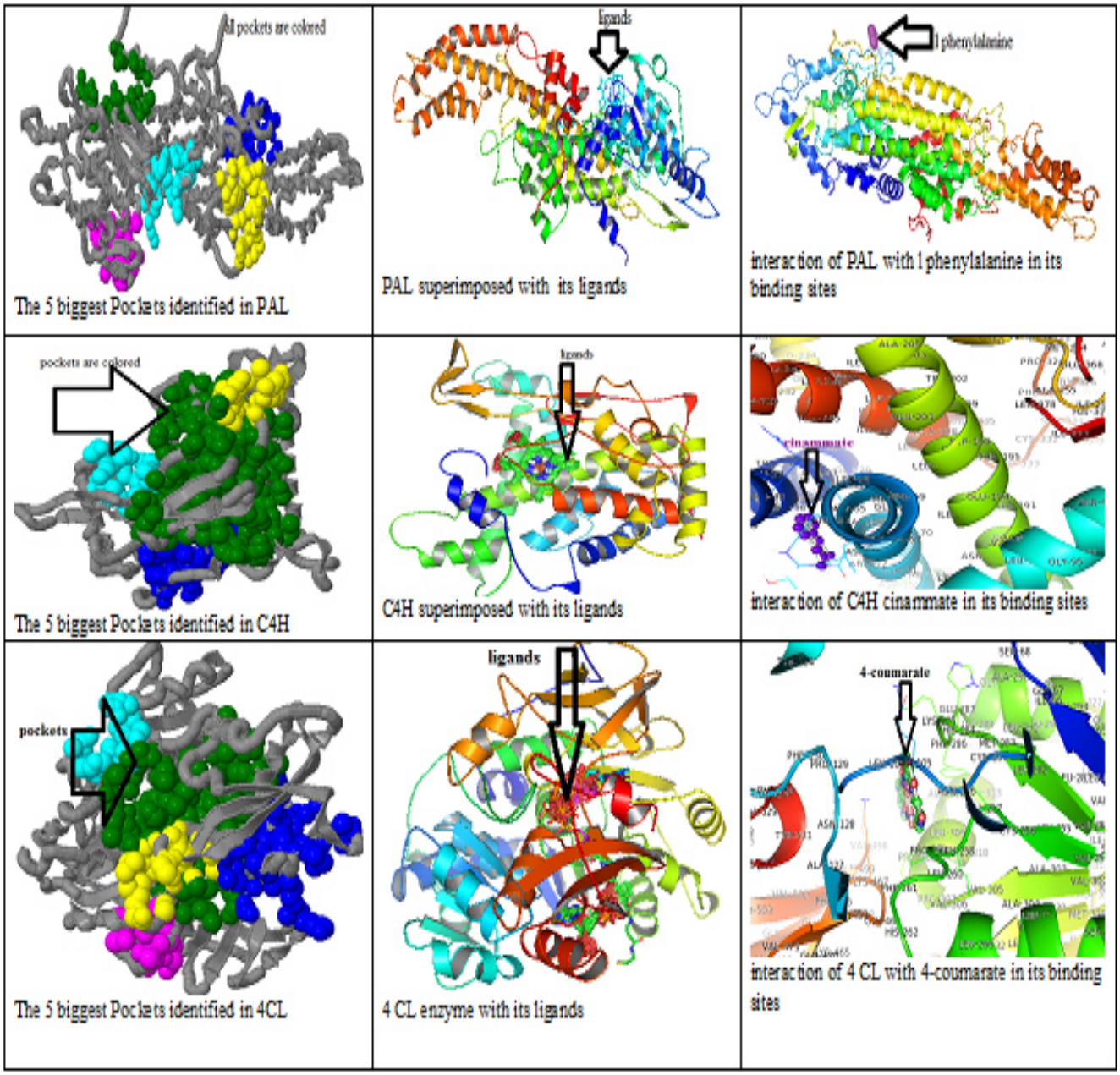
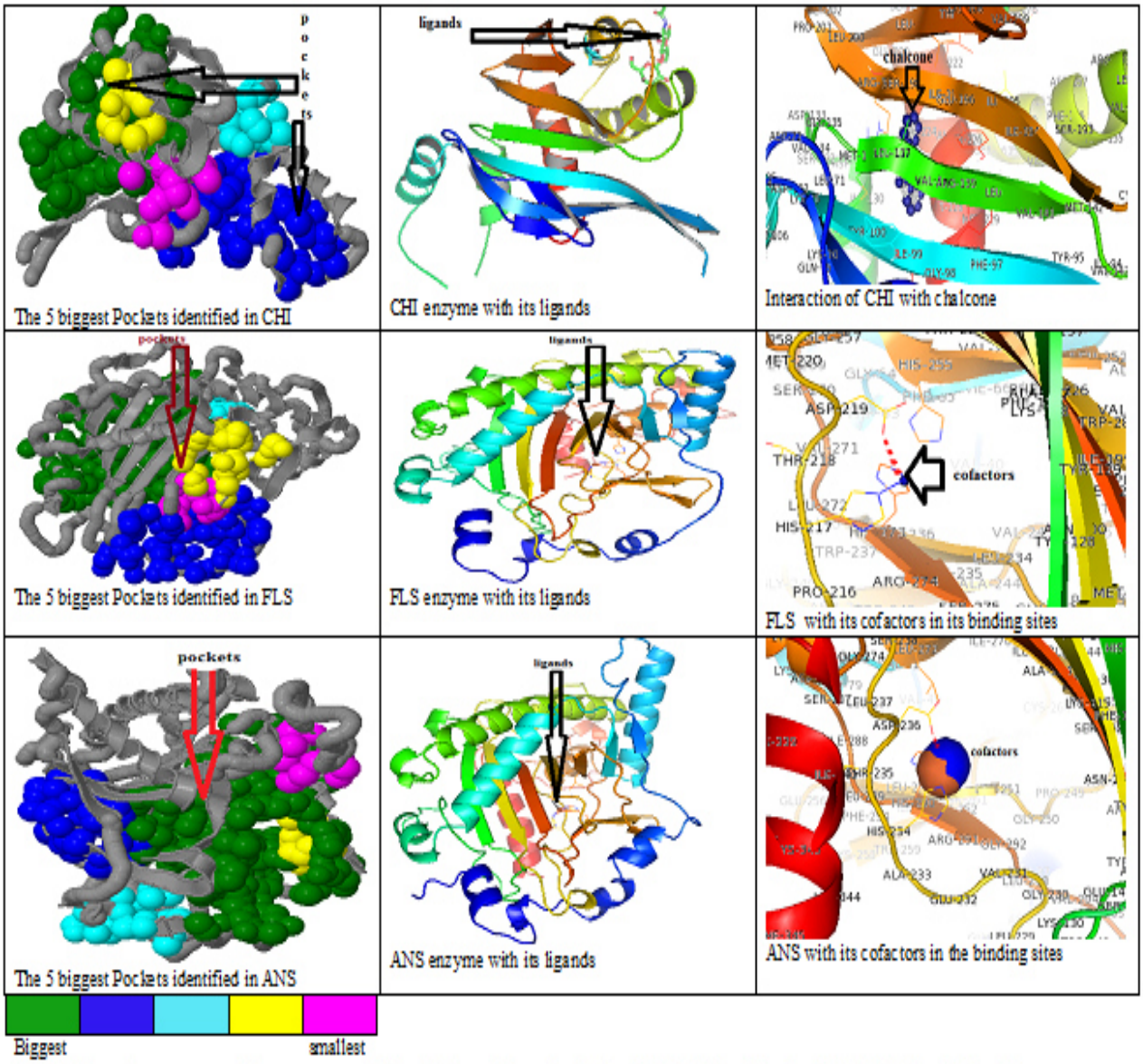


Figure 1: summary of all enzymes as predicted by PHYRE 2 server (a) PAL, (b) C4H, (c) 4CL, (d) CHI, (e) FLS, (f) ANS. All enzymes were modeled using the automated PHYRE 2 approach giving models that had 70% and above coverage. These were good quality models.

Table 5: enzyme structures with their ligands in their predicted cavities and the interactions





The table above is a summary of the enzymes modeled and their cavities and pockets, and a simulation of the ligand attached to the binding sites.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Recent conferences: http://www.iiste.org/conference/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar