Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 6, Issue 2, 2014

Research Article

INSILICO ANALYSIS OF PROTEINS OF *CURCUMA CAESIA* ROXB

S. VIJAYA BHARATHI

Department of Biotechnology, Faculty of Science and Humanities, SRM University, Kattankulathur, India. Email: shavisa04@gmail.com

Received: 09 Jan 2014, Revised and Accepted: 24 Feb 2014

ABSTRACT

Objective: In this present study, 10 proteins of *Curcuma caesia* were analysed using bioinformatics tools.

Methods: Structural prediction and functional characterization of proteins of *Curcuma caesia* were done using Expasy Protparm server, SOPMA, TMHMM and SOSUI tools. Chalcone synthase of *Curcuma caesia* was submitted to Blastp. Plants of different family showing identity 80% and above were selected and its sequences retrieved, aligned using Clustal Omega. Using NJ plot, phylogenetic tree was constructed for the aligned sequence.

Results: Structure prediction showed that α – helix, random coil, β – turn and extended strand predominates. Transmembrane region was found in AtpF protein. Phylogenetic analysis of chalcone synthase of *Curcuma caesia* reveals that the plants of Costaceae, Liliaceae, and Musaceae family are closely related.

Conclusion: *Curcuma caesia* an endangered medicinal plant has to be analysed further for identifying its various medicinal properties.

Keywords: *Curcuma caesia*, Computational tools, Chalcone synthase, Phylogeny.

INTRODUCTION

Medicinal plants have played a significant role in ancient traditional systems of medication in many countries. They are rich source of bioactive compounds and thus serve as important raw materials for drug production [1]. Infectious diseases are the world's leading cause of premature deaths [2]. Therefore, there is a continuous and urgent need to discover new antimicrobial compounds with diverse chemical structures and novel mechanisms of action [3]. In recent years, secondary plant metabolites (phytochemical) have been extensively investigated as a source of medicinal agents [4]. Contrary to the synthetic drugs, antimicrobials of plant origin are not associated with side effects. They are also cheap, easily available and affordable [5]. *Curcuma* Linn. is a large genus belonging to the family Zingiberaceae, comprises about 80 species of rhizomatous herbs. *Curcuma* plants (rhizomes and leaves) have a camphoraceous aroma and contain many functional compounds such as phenolics, flavonoids and different antioxidant enzymes**.** *Curcuma caesia is* commonly known as black turmeric (Kali haldi) which is a perennial herb found throughout the Himalayan region, North-east and central India. The paste of rhizome is used traditionally for the treatment of leucoderma, asthma, tumor, piles etc. *Curcuma caesia* has medicinal value due to its bioactive compound viz alkaloids, steroids, phenolics, and tannins. Essential oil of *C. caesia* has been known for its antifungal activity [6]. The rhizome oil of *C. caesia* contains 76.6% δ -camphor; 1,8-cineole (9.06%), ocimene (15.66%), 1- ar -curcumene (14.84%), δ – camphor (18.88%), δ - linalool (20.42%), δ -borneol (8.7%) and zingiberol (12.60%)[7].*Curcuma caesia* has scientifically studied for various therapeutical activities like antioxidant, antibacterial, antipyretic, larvicidal, insecticidal, antimicrobial, wound healing and anti-hyperglycemic [8]. Flavonoids are low molecular weight, polyphenolic compounds available in all-dietary plants [9].Over 4000 structurally unique flavonoids have been identified in plants. The common feature of these compounds is phenyl benzopyrone skeleton (C6-C3-C6). They are mainly classified into flavonol, flavanols, flavanonols, flavanones, flavones, and isoflavones on the basis of saturation level and opening of the central ring. Flavonoids are important plant secondary metabolites that serve various functions in higher plants. These include pigmentation, UV protection, fertility, antifungal defense and the recruitment of nitrogen-fixing bacteria.

Flavonoids possess wide spectrum of biological activities that might be able to influence processes that are dys- regulated during cardiovascular diseases. These include free radical scavenging, antioxidant, anti-thrombotic, antiapoptotic, anti ischemic, anti-arhythmic, anti-hypertensive, and anti-inflammatory activities [10]. Flavonoids are synthesized via the phenylpropanoid and polyketide pathway, which starts with the condensation of one molecule of CoA-ester of cinnamic acid or derivatives such as coumaric or ferulic acid, and three molecules of malonyl-CoA, yielding a naringenin chalcone as major product. This reaction is carried out by the enzyme chalcone synthase (CHS). The chalcone is isomerised to a flavanone by the enzyme chalcone flavanone isomerase (CHI). From these central intermediates, the pathway diverges into several branches, each resulting in a different class of flavonoids. Flavanone 3-hydroxylase (F3H) catalyzes the stereospecific 3ß-hydroxylation of (2S)-flavanones to dihydroflavonols. For the biosynthesis of anthocyanins, dihydroflavonol reductase (DFR) catalyzes the reduction of dihydroflavonols to flavan-3, 4-diols (leucoanthocyanins), which are converted to anthocyanidins by anthocyanidin synthase (ANS). The formation of glycosides is catalyzed by UDP glucose-falconoid 3-*O*-glucosyl transferees (UFGT), which stabilizes the anthocyanidins by 3-*O*-glucosylation [11]. Chalcone synthase or naringenin-chalcone synthase (CHS) (CHS, EC 2.3.1.74) is an enzyme ubiquitous to higher plants and belongs to a family of polyketide synthase enzymes (PKS) known as type III PKS. Type III PKSs are associated with the production of chalcones, a class of organic compounds found mainly in plants as natural defense mechanisms and as synthetic intermediates. CHS was the first type III PKS to be discovered. It is the first committed enzyme in flavonoid biosynthesis. CHS is believed to act as a central hub for the enzymes involved in the flavonoid pathway [12].*CHS* gene expression is influenced by many stresses and environmental factors. These include light and UV radiation, wound and pathogen attack, as well as plant-microbe interactions. Several groups of defense-related genes have also been identified with respect to insect and microbial pathogens and include protease inhibitors, chitinase, phenylalanine ammonia lyase and chalcone synthase, all of which accumulate systemically[13]. The major drawbacks of experimental methods that have been used to characterize the proteins of various organisms are the time frame involved, high cost and the fact that these methods are not amenable to high throughput techniques. *In silico* approaches provide a viable solution to these problems [14].

MATERIALS AND METHODS

Sequence Retrieval

The FASTA sequence of the proteins [TABLE: 1] were retrieved from Genbank database hosted by the NCBI (http://www.ncbi.nlm.nih.gov) [12].

Table 1: Proteins of *Curcuma caesia*

S. No.	Accession number	Protein	Length
1	AEF58786.1	Chalcone synthase (CHS2)	189
2	AEF58785.1	Chalcone synthase (CHS1)	189
3	AET36763.1	RNA polymerase C	173
4	AET36755.1	RNA polymerase B	123
5	AET36747.1	Ribulose-1,5-bisphosphate carboxylase/oxygenase	187
6	AET36733.1	Maturase K	265
7	AET36725.1	AtpF	44
8	AET36717.1	Acetyl-CoA carboxylase-D	127
9	AET36740.1	PsbK	8
10	AGL98016.1	PsbA	18

Table 2: Parameters computed using Expasy's ProtParam tool.

Protein	Accession Number	Length	Mol.wt	pI	-R	+R	Ec	II	AI	GRAVY
Chalcone synthase	AEF58786.1	189	20153.9	4.75	28	17	17990	33.8	99.68	-0.015
Chalcone synthase	AEF58785.1	189	20355.2	4.93	22	14	22125	32.02	90.85	0.127
RNA polymerase C	AET36763.1	173	19257.4	8.89	18	21	8605	36.70	110.98	-0.0009
RNA polymerase B	AET36755.1	123	13371.2	5.89	14	12	13075	25.16	92.76	-0.156
Ribulose-1,5-bisphosphate carboxylase/ oxygenase	AET36747.1	187	20663.5	7.58	21	22	30495	22.81	80.32	-0.271
Maturase K	AET36733.1	265	32118.8	9.72	15	36	70165	31.72	98.53	-0.014
AtpF	AET36725.1	44	4907.7	5.19	2	1	9970	15.54	117.27	0.755
Acetyl-CoA carboxylase-D	AET36717.1	127	13748.9	5.39	12	11	7450	36.99	101.42	0.193
PsbK	AET36740.1	8	964	9.75	0	1	5500	8.75	61.25	0.113
Psb A	AGL98016.1	18	1842.0	4.35	2	0	*	14.23	92.22	0.150

Mol. Wt – molecular weight(Daltons), pI – Isoelectric point, -R - Number of negative residues, +R – Number of Positive residues, EC – Extinction Coefficient at 280 nm, II – Instability Index, AI – Aliphatic Index, GRAVY – Grand Average Hydropathicity, * - No Trp, Tyr or Cys residue (should not be visible by UV spectrophotometry).

Table 3: Secondary structure results of proteins of *curcuma caesia*.

Secondary structure	AEF58786.1	AEF58785.1	AET36763.1	AET36755.1	AET36747.1	AET36733.1	AET36725.1	AET36717.1	AET36740.1	AGL98016.1
Alpha helix	40.74%	43.39%	37.57%	13.82%	33.69%	44.15%	29.55%	45.67%	0.00%	0.00%
3_{10} helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	18.52%	16.40%	15.03%	34.15%	22.99%	23.02%	36.36%	18.90%	0.00%	33.33%
Beta turn	9.52%	7.94%	7.51%	12.20%	5.35%	4.15%	0.00%	7.87%	0.00%	5.56%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	31.22%	32.28%	39.88%	39.84%	37.97%	28.68%	34.09%	27.56%	100.00%	61.11%
Ambigous states	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 4: Transmembrane region predicted by SOSUI server

Protein	Accession Number	Transmembrane region	Length	Type
AtpF	AET36725.1	ILATNPINLSVVLGVLIYF	19	PRIMARY

Primary Structure Prediction:- For Physio-chemical characterization, theoretical Isoelectric Point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average of hydropathy (GRAVY) were computed using the Expasy Protparm server. [15] **(**http://us.expasy.org/tools/protparam.html*)*.

Secondary structure prediction

SOPMA (Self Optimized Prediction Method with Alignment) was used for the secondary structure prediction.

Functional characterization

SOSUI and TMHMM v.2.0 tools were used to characterize whether the protein is soluble or transmembrane in nature. InterPro is an integrated resource for protein families, domains and functional sites. InterPro incorporates the major protein signature databases into a single resource. These include: PROSITE, which uses regular expressions and profiles, PRINTS, which uses Position Specific Scoring Matrix-based (PSSM-based) fingerprints, ProDom, which uses automatic sequence clustering, and Pfam, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, Gene3D and PANTHER, all of which use

hidden Markov models (HMMs). Superfamily and molecular function were predicted by Interpro protein sequencing and classification.[17]. (http://www.ebi.ac.uk/interpro/).

Sequence Alignment

Sequence alignment of Chalcone synthase (AEF58786.1) was performed using pairwise sequence alignment tool (NCBI- BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and multiple sequence alignment was done using the EBI-CLUSTAL OMEGA (http://www.ebi.ac.uk/Tools/msa/clustalo/) tool. Clustal Omega also has powerful features for adding sequences to and exploiting information in existing alignments, making use of the vast amount of precomputed information in public databases like Pfam [18].The emphasis of this work was to find the regions of sequence similarity, which in other words allows us to yield functional and evolutionary relationships among the proteins considered in this study.

Phylogenetic Analysis

The phylogentic analysis of chalcone synthase (CHS2) was performed to determine the number of proteins that share common structural and functional features. As an input to Clustal Omega all sequences in fasta formats were supplied with default options. The output was analyzed for sequences that are aligned for the complete length, scores, alignment, conserved residues, substitutes and semi conserved substituted residue patterns. The phylogenetic tree was constructed based on the bootstrap Neighbour Joining (NJ) method [19].The stability of the internal nodes was assessed by bootstrap analysis with 1000 replicates.

RESULTS & DISCUSSION

Parameters computed using Expasy's ProtParam tool revealed that pI of 6 protein were less than 7, indicates those proteins were acidic [TABLE:2], pI of 3 proteins were greater than 7,indicates those proteins have basic character and 1protein was neutral. At pI proteins are stable and compact [20]. The computed isoelectric point (pI) will be useful for developing buffer system for purification by isoelectric focusing method. All the proteins of *Curcuma caesia* showed instability index smaller than 40, indicates the proteins are stable. The computed extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution. Aliphatic index of the proteins ranged 61.25 – 117.27. Range of GRAVY of *Curcuma caesia* proteins was – 0.009 to 0.755. The lowest value of GRAVY indicates the possibility of better interaction with water [21].

Secondary structure prediction of *Curcuma caesia* proteins by SOPMA revealed that α – helix, random coil, β – turn and extended strand were more prevalent. In chalcone synthase (CHS2, 1), maturase K, acetylCoA carboxylase-D α – helix predominates, whereas in RNA polymerase – C, B, ribulose-1,5-bisphosphate carboxylase/oxygenase, Psb K and Psb A random coil region was frequent [Table: 3]. In Atp – F, extended strand dominates followed by random coil and α – helix. The secondary structure were predicted by using default parameters (Window width: 17, similarity threshold: 8 and number of states: 4). TMHMM v.2.0 and SOSUI predicted that except Atp –F [Fig: 1], all the other 9 proteins were soluble protein. Atp –F found to have transmembrane region of 19 amino acids length [Table: 4]. Domains are evolutionary units, often identi ed asrecurring sequence or structure [20]. Interpro tool analysis of proteins of curcuma caesia revealed its superfamily, molecular function. [Table: 5].

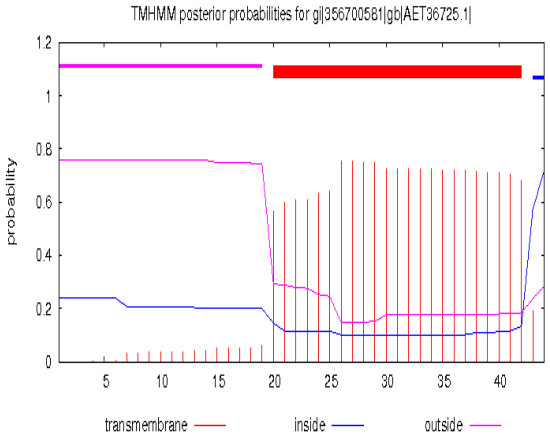

Fig. 1: TMHMM result showing transmembrane region of Atp- F.

Table 5: Interpro results of proteins of *curcuma caesia*

S. No.	Accession Number	Superfamily	Molecular function
1	AEF58786.1	Thiolase like superfamily	Catalytic activity, transferase activity, transferring acyl group.
2	AEF58785.1	Thiolase like superfamily	Catalytic activity, transferase activity, transferring acyl group.
3	AET36763.1	β & β prime subunits of DNA dependent RNA polymerase superfamily	DNA binding. DNA directed RNA polymerase activity.
4	AET36755.1	β & β prime subunits of DNA dependent RNA polymerase superfamily	DNA binding. DNA directed RNA polymerase activity.
5	AET36747.1	RuBisco, c- terminal domain superfamily	Magnesium binding, ribulosebiphosphate carboxylase activity.
6	AET36733.1	-	-
7	AET36725.1	-	-
8	AET36717.1	ClpP/ Crotonase superfamily	Acetyl coA carboxylase activity, ligase activity.
9	AET36740.1	-	-
10	AGL98016.1	-	-

- Further analysis required.

Evolutionary relationship was done with chalcone synthase (CHS2). As chalcone synthase is the key enzyme in flavanoid biosynthesis. Majority of flavonoids are powerful antioxidants that help neutralize harmful free radicals and prevent oxidative stress which damage cells and DNA, and which can lead to aging and degenerative diseases like cancer and Alzheimer's or Parkinson's disease[22]. The chalcone synthase (CHS2) of curcuma caesia was subjected to BLASTp analysis to find the other plant species having the same query protein. The results obtained showed that more than 43 plant species belonging to 19 different family have 80 % and above similarity.

From the above hits one plant species from each family was randomly selected for evolutionary analysis in this study.

The list of these plant species, accession number, identity score and e value are given in [Table: 6].

Table 6: Lists of plant spieces showing similarity of 80% and above with the

S. No.	Plant species containing chalcone synthase protein	Family	Accession number	Identity %	E- value
1.	*Costus erythrophyllus*	Costaceae	AEF58796.1	88	6e-120
2.	*Musa accunita*	Musaceae	ADE59487.1	86	7e-116
3.	*Tulipa fosteriana*	Liliaceae	AGJ50587.1	85	1e-114
4.	*Sorghum bicolor*	Poaceae	XP_002450870.1	81	2e-111
5.	*Vitis vinifera*	Vitaceae	BAB84112.1	83	7e-111
6.	*Sargentodoxa cuneata*	Lardizabalaceae	ABD14724.1	81	2e-110
7.	*Tetracentron sinense*	Trochodendraceae	ABD14723.1	81	4e-110
8.	*Grevillea robusta*	Proteaceae	ABD14733.1	83	5e-110
9.	*Pyrus pyrifolia*	Rosaceae	AFH68066.1	81	8e-110
10.	*Rhododendron X pulchurm*	Ericaceae	BAF96941.1	81	1e-109
11.	*Nicotiana alta*	Solanceae	ACS12837.1	81	1e-109
12.	*Lonicera japonica*	Caprifoliaceae	AFJ44312.1	81	3e-109
13.	*Kingdonia unifola*	Kingdoniaceae	ABD14727.1	81	4e-109
14.	*Rianus communis*	Euphorbiaceae	XP_002512817.1	80	8e-110
15.	*Aquilaria sinesis*	Thymelacaceae	ABM73434.1	80	2e-109
16.	*Theobroma cacao*	Malvaceae	EOY02978.1	80	2e-109
17.	*Hypericum sampsonii*	Hypericaceae	AFU52909.1	80	2e-109
18.	*Citrus sinesis*	Rutaceae	ACB47461.1	80	5e-109
19.	*Nelumbo nucifera*	Nelumbonaceae	ADD74168.1	80	6e-109

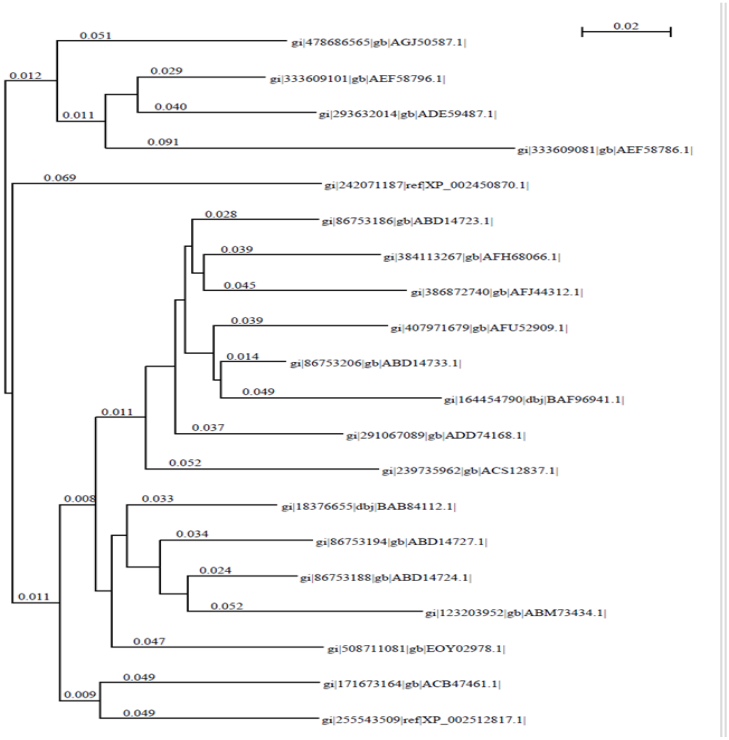

Fig. 2: Phylogenetic tree of chalcone synthase protein containing plants

Chalcone synthase protein

A multiple sequence alignment was done for 19 plant species [Table 6] using Clustal Omega. The tool was run with default parameters such as a value of 10 for gap open penalty and a value of 0.05 and 0.05 for extending a gap and separating a gap respectively. The phylogenetic tree was drawn using NJ plot [Fig 2]. The results revealed that the chalcone synthase protein of *Curcuma caesia* of Zingiberaceae family was closely related to *Costus erythrophyllus, Musa accunita, and Tulipa fosteriana plants which* closely related to *Costus erythrophyllus*(Costaceae), *Musa accunita*(Musaceae) and *Tulipa fosteriana*(Liliaceae) whereas distantly related to *Rianus communis plant (family* Euphorbiaceae).

Therefore the plants of these families can be assayed for antioxidant, anticancer, antithrombic activity and can be used for producing pharmacologically effective substances.

CONCLUSION

In this study, proteins of *Curcuma caesia* were selected. Expasy's ProtParam tool predicted the physio-chemical characters of the proteins. Phylogenetic study revealed the close and distant relationship of chalcone synthase protein of *Curcuma caesia* of Zingiberaceae family with the plants of other family. Further analysis are required for drug target identification.

REFERENCE

1. Akter, F. A. Neela, M. S. I. Khan, M. S. Islam, and M. F. Alam, Screening of Ethanol, Petroleum Ether and Chloroform Extracts of Medicinal Plants, *Lawsonia inermis* L. and *Mimosa pudica* L. for Antibacterial Activity. Indian journal of pharmaceutical Sciences. 2010; 72 (3): 388–392.

2. Shahab-ud-Din, H. Doutani, F.A. Sattar, F. Ahmad, S. Faridullah, F. Khan, J.K. Tareen and B. Arif, The use of Medicinal Plants among Different Communities of Balochistan against Hepatitis. Current Research Journal of Biological Sciences. 2012; 4 (2): 215-219.
3. P. Sujith and P.K. Senthilkumar. Bactericidal Activity of Extracts of Different Flowering Stages of *Cassia Auriculata* and Screening of its Amino Acids. International Journal of Microbiological Research. 2012; 3 (2): 158-162.
4. S. Idris, G.I. Ndukwe and C.E. Gimba. Preliminary phytochemical screening and antimicrobial activity of seed extracts of *Persea mericana* (avocado pear). Bayero Journal of Pure and Applied Sciences. 2009; 2 (1):173 – 176.
5. S.P.Anand, A.Doss and V. Nandagopalan. Antibacterial studies on leaves of *Clitoria ternatea* Linn – a high potential medicinal plant. International journal of applied biology and pharmaceutical technology. 2011; 2 (3):453-456.
6. Yogamaya Dhal, Bandita Deo and R. K. Sahu. Comparative antioxidant activity of non-enzymatic and enzymatic extracts of *curcuma zedoaria, curcuma angustifolia* and *curcuma caesia.* International Journal of Plant, Animal and Environmental Sciences. 2012; 2: 232-239.
7. Satyavama D Asem & Warjeet S Laitonjam. Investigation of the structure – nonlinearity relationship of Zederone from the rhizomes of curcuma caesia Roxb. Indian Journal of Chemistry. 2012; 51B: 1738 – 1742.
8. Rajeshwari Sahu and Jyoti Saxena. A brief review on medicinal value of *Curcuma caecia.* International journal of pharmacy & life sciences.2013; 4(5): 2664-2666.
9. W.Ren, Z.Qiao, H.Wang, L.Zhu and L.Zhang. Flavonoids: promising anticancer agents. Medicinal Research Reviews. 2003; 23 (4): 519-534.
10. Mukesh Nandave, S.K Ojha and D.S Ar. Review Article Protective role of flavonoids in cardiovascular diseases. Natural Product Radiance.2005; 4(3):166-176.
11. Zachary.L. Fowler, Mattheos A. G. Koffas. Biosynthesis and biotechnological production of flavanones: current state and perspectives. Applied Microbiology Biotechnology. 2009; 83:799-808.
12. T. T. H. Dao, H. J. M. Linthorst, and R. Verpoorte. Chalcone synthase and its functions in plant resistance. Phytochemistry Review. 2011; 10 (3): 397–412.
13. Sylvie Richard, Gilles lapointe, G. Rutledge, Armand Seguin. Induction of chalcone synthase expression in white spruce by wounding and jasmonate. Plant & Cell physiology**.** 2000; 41 (8): 982-987
14. Archna Shay and Madhvi Shakya. In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach. Journal of proteomics & Bioinformatics. 2010; 3 (5): 148-154.
15. Gasteiger, C.Hoogland, A.Gattiker, S.Duvaud, M.R.Wilkins, R.D. Appel, A.Bairoch. Protein Identification and Analysis Tools on the ExPASy Server, (In) John M.Walker (ed): The Proteomics Protocols Handbook, Humana Press. 2005; 571-607.
16. Nicola. J.Mulder et al. New developments in the interPro database. Nucleic acid Research. 2007; 35:D224 – D228.
17. Fabina Sievers et al. Fast, Scalable Generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology.2011; 7:539.
18. G.Perriere and M. Gouy. WWW-Query: An on-line retrieval system for biological sequence banks. Biochimie. 1996; 78: 364-369.
19. Upgade Akhilesh, Bhaskar, Anusha, Issar Sakshi and Senthamarai Selvi V. In Silico Characterization of Keratitis Causing Herpes Simplex Virus (HSV 1) Membrane Proteins using Computational Tools and Servers. Research Journal of Recent Sciences. 2012; 1(11): 27-31.
20. Kebila Venkatasamy. In silico Analysis and Homology Modeling of Putative Hypothetical Protein Q4QH83 of *Leishmania* major. Advanced BioTech.2013; 13 (3): 1-4.
21. T. Shohaib, M.Shafique, N. Dhanya, Madhu.C.Divakar. Importance of flavonoides in therapeutics. Hygeia. Journal for Drug and Medicne. 2001; 3(1): 1-18