

# Phylogenetic Analysis of Vicilin Protein

Sourabh Pujari<sup>1</sup>, Guruprasad Bhandwale<sup>2</sup>, Shantanu Deshpande<sup>3</sup>, Vivek Mali<sup>4</sup>

<sup>1,2,3,4</sup> Student, Dept. of Biotechnology Engineering, KIT's College of Engineering, Kolhapur, India

## Abstract

Vicilin is a seed storage protein. It belongs to the Cupin superfamily. It has been suggested to be an allergen in allergic reactions in nuts. The following methodologies were carried using respective materials for the phylogenetic study of Vicilin. Multiple Sequence Alignment using MEGA-X was done then, Phylogenetic Tree was constructed by using Clustal-W, Entropy Plot was constructed using Bio-edit. Conservation Score was found using Al2CO. By studying various properties of Vicilin protein its conservation of allergenic properties throughout the selected nuts species was known. The major future application could be reducing the allergenic properties of Vicilin by studying its structure, especially identifying the epitope and its chemical nature.

**Keywords:** *Phylogenetic analysis, Phylogenetic tree, Vicilin, Protein*

## 1. Introduction

Vicilin is a seed storage protein. Unlike other similar proteins it possesses a novel copper ligand. It belongs to the most diverse families of proteins, the Cupin superfamily. It has a myriad of possible enzymatic functions. The structure of Vicilin is the basis of its function. The results from ExPasy software have shown the theoretical molecular weight of Vicilin to be 46kD and isoelectric point is 5.4. Vicilin only contains one alpha subunit, one glycerol and one phosphate ion. The incorporation of copper ligand is the unique and crucial structural feature. A unique structural motif is the N-terminus and the C-terminus form Cupin folds to this class of protein where a conserved beta barrel is formed. We selected *Juglans regia*, *Anacardium occidentale*, *Pistacia vera*, *Corylus avellana* and *Arachis hypogaea* species for our phylogenetic analysis of Vicilin. Vicilin exhibits allergenicity in consumers. The following methodologies were carried using respective materials for the phylogenetic study of Vicilin. Multiple Sequence Alignment using MEGA-X was done then, Phylogenetic Tree was constructed by using Clustal-W, Entropy Plot was constructed using Bio-edit. Conservation Score was found using Al2CO. By studying various properties of Vicilin protein its conservation of allergenic properties throughout the selected nuts species was known. The major future application could be reducing the allergenic properties of Vicilin by studying its structure, especially identifying the epitope and its chemical nature.

## 2. Materials

### 2.1. NCBI (Protein/ gene)

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB.

### 2.2. Clustal W –

Clustal is a series of widely used computer programs used in Bioinformatics for multiple sequence alignment

### 2.3. MEGA X –

Molecular Evolutionary Genetics Analysis (MEGA) is computer software for conducting statistical analysis of molecular evolution and for constructing phylogenetic trees. MEGA can help to perform Sequence alignment construction, Data handling, Genetic code table section, Sequence data viewer, Distance estimation methods, Tree-making methods, etc.

#### 2.4. BioEdit 7.2-

BioEdit software is a sequence alignment editor and sequence analysis program designed for sequence manipulation.

#### 2.5. ExPASy-

The ExPASy (the Expert Protein Analysis System) World Wide Web server is provided as a service to the life science community by a multidisciplinary team at the Swiss Institute of Bioinformatics (SIB). It provides access to a variety of databases and analytical tools dedicated to proteins and proteomics.

#### 2.6. SWISS-MODEL Repository:

SWISS-MODEL Repository is a database of automatically generated structural protein models.

#### 2.7. Relevant Databases-

Prodata, ExPASyProsite, PDB, computational biology unit

### 3. Methodology

#### 3.1. Sequence search

Sequences for Vicilin were retrieved from UniProt and NCBI (Protein). Total 7 sequences from 5 species of nuts were found.

#### 3.2. Multiple sequence alignment

The alignment of the retrieved 7 sequences was performed by using MuSCLE.

#### 3.3. Phylogenetic analysis

MEGA X was used for creating phylogenetic tree. For this, Maximum Likelihood method was used.

#### 3.4. Entropy Plot

BioEdit 7.2 was used to obtain the entropy plot for the multiple sequence alignment.

#### 3.5. Conservation Score

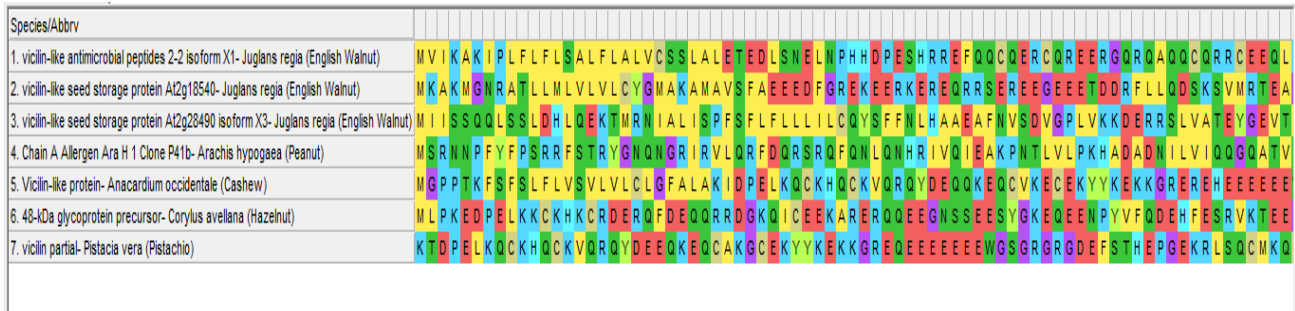
Al2co prodata was used to obtain the conservation score of selected sequences.

#### 3.6. Structure

ExPASy was used to obtain structure of Vicilin.

## 4. Results and Analysis

### 4.1. Multiple Sequence Alignment



1: multiple sequence alignment

Fig

The multiple sequence alignment was carried out by retrieving the protein sequences from Uniprot and NCBI in fasta format. These sequences were then aligned using MEGA-X through the build alignment option. The alignment that was obtained was again aligned by using MuSCL (Multiple Sequence Comparison by Log-Expectation) in order to get the multiple sequence alignment.

### 4.2. Phylogenetic Tree:

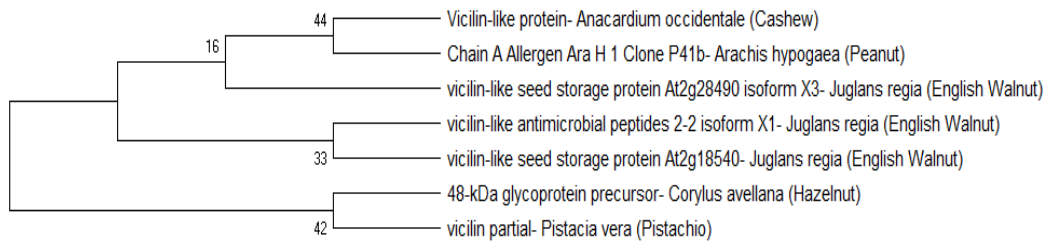


Fig.2:

Phylogenetic tree

The phylogenetic tree was obtained on the

basis of previously obtained multiple sequence alignment. For tree building we used MEGA-X Phylogeny option. The tree was built using Maximum Likelihood method.

### 4.3. Entropy Plot:

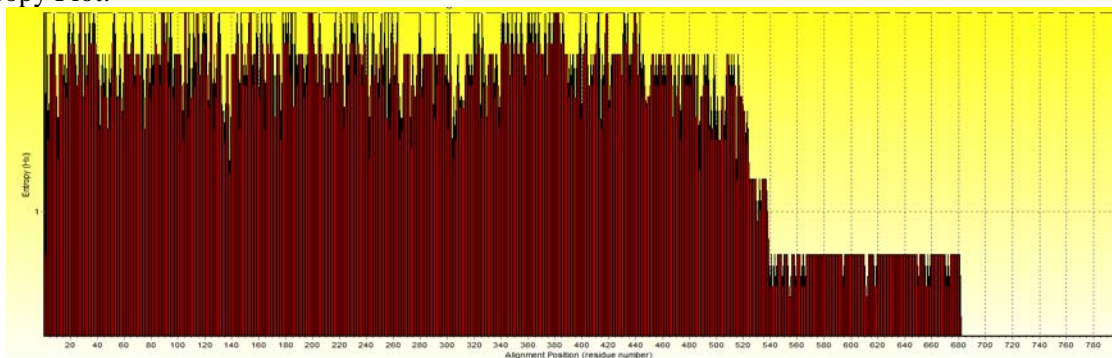


Fig.3: Entropy Plot

To know the diverse and conserved regions in the selected protein sequences entropy plot was derived with the help of BioEdit version 7.2. According to the plot obtained there are non-continuous diverse regions lying in bits between 0- 520 residues. In this region there are many peaks as well, which indicate the highest diversities. Whereas the flat plateau-like regions indicate conserved residues, here, ranging from 520 to 680.

4.4. Conservation Score:

Table 1: Conservation Score

1	:	-1.000 *	68	:	-1.000 *	135	:	-1.000 *
2	:	-1.000 *	69	:	-1.000 *	136	:	-1.000 *
3	M	-1.000 *	70	:	-1.000 *	137	:	-1.000 *
4	G	-1.000 *	71	:	-1.000 *	138	:	-1.000 *
5	P	-1.000 *	72	:	-1.000 *	139	:	-1.000 *
6	P	-1.000 *	73	:	-1.000 *	140	:	-1.000 *
7	T	-1.000 *	74	:	-1.000 *	141	:	-1.000 *
8	K	-1.000 *	75	:	-1.000 *	142	:	-1.000 *
9	F	-1.000 *	76	:	-1.000 *	143	:	-1.000 *
10	S	-1.000 *	77	:	-1.000 *	144	:	-1.000 *
11	F	-1.000 *	78	:	-1.000 *	145	:	-1.000 *
12	S	-1.000 *	79	:	-1.000 *	146	:	-1.000 *
13	L	-1.000 *	80	:	-1.000 *	147	:	-1.000 *
14	F	-1.000 *	81	:	-1.000 *	148	:	-1.000 *
15	L	-1.000 *	82	:	-1.000 *	149	:	-1.000 *
16	V	-1.000 *	83	:	-1.000 *	150	:	-1.000 *
17	S	-1.000 *	84	:	-1.000 *	151	:	-1.000 *
18	V	-1.000 *	85	:	-1.000 *	152	:	-1.000 *
19	L	-1.000 *	86	:	-1.000 *	153	:	-1.000 *
20	V	-1.000 *	87	:	-1.000 *	154	:	-1.000 *
21	L	-1.000 *	88	:	-1.000 *	155	:	-1.000 *
22	C	-1.000 *	89	:	-1.000 *	156	:	-1.000 *
23	L	-1.000 *	90	:	-1.000 *	157	:	-1.000 *
24	G	-1.000 *	91	:	-1.000 *	158	:	-1.000 *
25	F	-1.000 *	92	:	-1.000 *	159	:	-1.000 *
26	A	0.062	93	:	-1.000 *	160	:	-1.000 *
27	L	0.125	94	:	-1.000 *	161	:	-1.000 *
28	A	-1.174	95	:	-1.000 *	162	:	-1.000 *
29	K	-0.611	96	:	-1.000 *	163	:	-1.000 *
30	I	-1.134	97	:	-1.000 *	164	:	-1.000 *
31	D	0.141	98	:	-1.000 *	165	:	-1.000 *
32	P	-0.611	99	:	-1.000 *	166	:	-1.000 *
33	:	-1.000 *	100	:	-1.000 *	167	:	-1.000 *
34	:	-1.000 *	101	:	-1.000 *	168	:	-1.000 *
35	:	-1.000 *	102	:	-1.000 *	169	:	-1.000 *
36	:	-1.000 *	103	:	-1.000 *	170	:	-1.000 *
37	:	-1.000 *	104	:	-1.000 *	171	Q	0.920
38	:	-1.000 *	105	:	-1.000 *	172	K	-1.168
39	:	-1.000 *	106	:	-1.000 *	173	E	-1.168
40	:	-1.000 *	107	:	-1.000 *	174	Q	-0.576
41	:	-1.000 *	108	:	-1.000 *	175	C	-0.576
42	:	-1.000 *	109	:	-1.000 *	176	V	-1.098
43	:	-1.000 *	110	:	-1.000 *	177	K	-0.583
44	:	-1.000 *	111	:	-1.000 *	178	E	-1.094
45	:	-1.000 *	112	:	-1.000 *	179	C	0.141
46	:	-1.000 *	113	:	-1.000 *	180	E	1.186
47	:	-1.000 *	114	:	-1.000 *	181	K	-0.627
48	E	0.141	115	:	-1.000 *	182	Y	-1.168
49	L	1.239	116	:	-1.000 *	183	Y	-1.168
50	K	-0.611	117	:	-1.000 *	184	K	-0.589
51	Q	-0.576	118	:	-1.000 *	185	E	1.186
52	C	0.141	119	:	-1.000 *	186	K	-0.558
53	K	0.060	120	:	-1.000 *	187	K	0.155
54	H	0.060	121	:	-1.000 *	188	G	-1.168
55	Q	-0.584	122	:	-1.000 *	189	R	0.223
56	C	1.186	123	R	2.493	190	:	-1.000 *
57	K	-1.168	124	Q	2.493	191	:	-1.000 *
58	V	-0.583	125	Y	0.932	192	:	-1.000 *
59	Q	-0.627	126	D	0.976	193	:	-1.000 *
60	:	-1.000 *	127	E	2.493	194	:	-1.000 *
61	:	-1.000 *	128	Q	1.043	195	:	-1.000 *
62	:	-1.000 *	129	:	-1.000 *	196	:	-1.000 *
63	:	-1.000 *	130	:	-1.000 *	197	:	-1.000 *
64	:	-1.000 *	131	:	-1.000 *	198	:	-1.000 *
65	:	-1.000 *	132	:	-1.000 *	199	:	-1.000 *
66	:	-1.000 *	133	:	-1.000 *	200	:	-1.000 *
67	:	-1.000 *	134	:	-1.000 *	201	:	-1.000 *

202 :	-1.000 *	273 S	-1.000 *	343 :	-1.000 *
203 :	-1.000 *	274 Q	-1.000 *	344 :	-1.000 *
204 :	-1.000 *	275 C	-1.000 *	345 :	-1.000 *
205 :	-1.000 *	276 M	-1.000 *	346 :	-1.000 *
206 :	-1.000 *	277 R	-1.000 *	347 :	-1.000 *
207 :	-1.000 *	278 Q	-1.000 *	348 :	-1.000 *
208 :	-1.000 *	279 C	-1.000 *	349 :	-1.000 *
209 :	-1.000 *	280 E	-1.000 *	350 :	-1.000 *
210 :	-1.000 *	281 R	-1.000 *	351 :	-1.000 *
211 :	-1.000 *	282 Q	-1.000 *	352 :	-1.000 *
212 :	-1.000 *	283 E	-1.000 *	353 :	-1.000 *
213 :	-1.000 *	284 G	-1.000 *	354 :	-1.000 *
214 :	-1.000 *	285 G	-1.000 *	355 :	-1.000 *
215 :	-1.000 *	286 Q	-1.000 *	356 :	-1.000 *
216 :	-1.000 *	287 Q	-1.000 *	357 :	-1.000 *
217 :	-1.000 *	288 K	-1.000 *	358 :	-1.000 *
218 :	-1.000 *	289 Q	-1.000 *	359 :	-1.000 *
219 :	-1.000 *	290 L	-1.000 *	360 :	-1.000 *
220 :	-1.000 *	291 C	-1.000 *	361 :	-1.000 *
221 :	-1.000 *	292 R	-1.000 *	362 :	-1.000 *
222 :	-1.000 *	293 F	-1.000 *	363 :	-1.000 *
223 :	-1.000 *	294 R	-1.000 *	364 E	-1.000 *
224 :	-1.000 *	295 C	-1.000 *	365 D	-1.000 *
225 :	-1.000 *	296 Q	-1.000 *	366 E	-1.000 *
226 :	-1.000 *	297 E	-1.000 *	367 D	-1.000 *
227 :	-1.000 *	298 R	-1.000 *	368 E	-1.000 *
228 :	-1.000 *	299 Y	-1.000 *	369 :	-1.000 *
229 :	-1.000 *	300 K	-1.000 *	370 A	-1.000 *
230 :	-1.000 *	301 K	-1.000 *	371 E	0.223
231 :	-1.000 *	302 E	-1.000 *	372 E	-1.125
232 :	-1.000 *	303 R	-1.000 *	373 E	-0.575
233 :	-1.000 *	304 G	-1.000 *	374 D	-0.604
234 :	-1.000 *	305 Q	-1.000 *	375 E	0.363
235 :	-1.000 *	306 H	-1.000 *	376 N	0.225
236 :	-1.000 *	307 N	-1.000 *	377 P	0.225
237 :	-1.000 *	308 Y	-1.000 *	378 Y	-0.526
238 :	-1.000 *	309 K	-1.000 *	379 V	0.135
239 :	-1.000 *	310 R	-1.000 *	380 F	1.239
240 :	-1.000 *	311 E	-1.000 *	381 E	-1.584
241 :	-1.000 *	312 D	-1.000 *	382 D	-0.602
242 :	-1.000 *	313 D	-1.000 *	383 E	-0.581
243 E	2.493	314 E	-1.000 *	384 D	-1.099
244 R	-0.576	315 D	-1.000 *	385 F	-0.514
245 E	1.128	316 :	-1.000 *	386 T	-1.584
246 H	-0.572	317 :	-1.000 *	387 T	0.112
247 E	0.068	318 :	-1.000 *	388 K	0.116
248 E	0.051	319 :	-1.000 *	389 V	-1.125
249 E	0.051	320 :	-1.000 *	390 K	-0.579
250 E	1.128	321 :	-1.000 *	391 T	-0.468
251 E	0.089	322 :	-1.000 *	392 E	0.296
252 E	-0.576	323 :	-1.000 *	393 Q	-1.138
253 :	-1.000 *	324 :	-1.000 *	394 G	1.239
254 W	-1.000 *	325 :	-1.000 *	395 K	-0.600
255 G	-1.000 *	326 :	-1.000 *	396 V	-0.526
256 T	-1.000 *	327 :	-1.000 *	397 V	-1.104
257 G	-1.000 *	328 :	-1.000 *	398 L	-0.411
258 G	-1.000 *	329 :	-1.000 *	399 L	0.225
259 V	-1.000 *	330 :	-1.000 *	400 P	-1.138
260 D	-1.000 *	331 :	-1.000 *	401 K	-1.128
261 E	-1.000 *	332 :	-1.000 *	402 F	0.225
262 P	-1.000 *	333 :	-1.000 *	403 T	0.307
263 S	-1.000 *	334 :	-1.000 *	404 Q	-1.110
264 T	-1.000 *	335 :	-1.000 *	405 K	0.115
265 H	-1.000 *	336 :	-1.000 *	406 S	-0.514
266 E	-1.000 *	337 :	-1.000 *	407 K	-1.125
267 P	-1.000 *	338 :	-1.000 *	408 L	-0.526
268 A	-1.000 *	339 :	-1.000 *	409 L	-0.526
269 E	-1.000 *	340 :	-1.000 *	410 H	-1.545
270 K	-1.000 *	341 :	-1.000 *	411 A	-0.559
271 H	-1.000 *	342 :	-1.000 *	412 L	0.055
272 L	-1.000 *				

413	E	0.141
414	K	0.103
415	Y	-0.428
416	R	0.945
417	L	0.130
418	A	-0.602
419	V	-0.493
420	L	0.922
421	V	-0.605
422	A	0.225
423	N	0.086
424	P	2.493
425	Q	0.121
426	A	0.230
427	F	0.922
428	V	-0.570
429	V	-0.487
430	P	2.493
431	S	-1.941
432	H	0.225
433	M	-1.106
434	D	1.239
435	A	1.254
436	D	0.161
437	S	-0.568
438	I	0.114
439	F	-0.559
440	F	0.122
441	V	0.976
442	S	-1.584
443	W	-1.099
444	G	2.493
445	R	-0.526
446	G	0.940
447	T	0.225
448	I	0.118
449	T	0.225
450	K	-1.099
451	I	-0.001
452	L	-1.566
453	E	-1.125
454	N	-0.004
455	:	-1.000 *
456	:	-1.000 *
457	K	-1.125
458	R	0.225
459	E	0.088
460	S	0.225
461	I	-1.055
462	N	1.254
463	V	1.021
464	R	-0.606
465	Q	-1.584
466	G	2.493
467	D	1.299
468	:	-0.495
469	:	0.071
470	:	1.211
471	:	0.256
472	:	0.276
473	:	1.256
474	:	2.493
475	:	-0.594
476	:	-0.589
477	:	0.140
478	:	2.493
479	:	-0.575
480	:	-0.592
481	:	1.211
482	:	-0.585
483	:	-0.559

483	:	-0.559
484	:	0.301
485	:	0.196
486	:	0.957
487	:	0.163
488	:	1.211
489	:	-1.590
490	:	-0.507
491	:	-0.559
492	:	0.196
493	:	-0.599
494	:	-0.559
495	:	0.073
496	:	1.211
497	:	0.935
498	:	0.297
499	:	-1.092
500	:	1.222
501	:	1.211
502	:	-0.613
503	:	-0.533
504	:	-1.000 *
505	:	-1.000 *
506	:	-1.000 *
507	:	-1.000 *
508	:	-1.000 *
509	:	-1.000 *
510	:	-1.000 *
511	:	-1.000 *
512	:	0.196
513	:	-1.092
514	:	0.256
515	:	-1.096
516	:	-0.594
517	:	0.146
518	:	0.098
519	:	-0.507
520	:	-0.585
521	:	-1.119
522	:	-0.507
523	:	-1.147
524	:	1.211
525	:	-0.613
526	:	0.132
527	:	0.262
528	:	0.150
529	:	2.493
530	:	0.196
531	:	-0.471
532	:	-0.507
533	:	0.103
534	:	2.493
535	:	-0.545
536	:	1.222
537	:	2.493
538	:	0.933
539	:	0.914
540	:	0.144
541	:	-0.466
542	:	0.072
543	:	-0.466
544	:	-1.096
545	:	0.073
546	:	-0.559
547	:	-0.001
548	:	-0.584
549	:	0.072
550	:	-1.590
551	:	-0.666
552	:	0.093

553	:	:	-1.000 *
554	:	:	-1.000 *
555	:	:	-1.000 *
556	:	-1.000 *	
557	:	:	-1.000 *
558	:	:	-1.000 *
559	:	:	-1.000 *
560	:	:	-1.000 *
561	:	:	-1.000 *
562	:	:	-1.000 *
563	:	:	-1.000 *
564	:	:	-1.000 *
565	:	:	-1.000 *
566	:	:	-1.000 *
567	:	:	-1.000 *
568	:	:	-1.000 *
569	:	:	-1.000 *
570	:	:	-1.000 *
571	:	:	-1.000 *
572	:	:	-1.000 *
573	:	:	-1.000 *
574	:	:	-1.000 *
575	:	:	-1.000 *
576	:	:	-1.000 *
577	:	:	-1.000 *
578	:	:	-1.000 *
579	:	:	-1.000 *
580	:	:	-1.000 *
581	:	:	-1.000 *
582	:	:	-1.000 *
583	:	:	-1.000 *
584	:	:	-1.000 *
585	:	:	-1.000 *
586	:	:	-1.000 *
587	:	:	-1.000 *
588	:	:	-1.000 *
589	:	:	-1.000 *
590	:	:	-1.000 *
591	:	:	-1.000 *
592	:	:	-1.000 *
593	:	:	-1.000 *
594	:	:	-1.000 *
595	:	:	-1.000 *
596	:	:	-1.000 *
597	:	:	-1.000 *
598	:	:	-1.000 *
599	:	:	-1.000 *
600	:	:	0.276
601	:	:	-0.559
602	:	:	-1.590
603	:	:	0.284
604	:	:	0.935
605	:	:	-1.123
606	:	:	0.196
607	:	:	-0.545
608	:	:	-0.545
609	:	:	-0.559
610	:	:	0.196
611	:	:	-1.114
612	:	:	1.222
613	:	:	-1.134
614	:	:	-1.147
615	:	:	-0.559
616	:	:	-0.559
617	:	:	-0.533
618	:	:	0.253
619	:	:	-0.585
620	:	:	-0.533
621	:	:	-0.587
622	:	:	-0.592
623	:	:	0.104

624 :	-0.507	693 :	2.493	763 :	1.211
625 :	-0.471	694 :	1.256	764 :	1.211
626 :	-0.578	695 :	-0.001	765 :	-0.620
627 :	-0.588	696 :	0.272	766 :	1.160
628 :	-0.592	697 :	-0.581	767 :	-1.590
629 :	-1.147	698 :	0.134	768 :	-0.559
630 :	-0.495	699 :	2.493	769 :	1.211
631 :	-1.091	700 :	0.071	770 :	-1.101
632 :	-1.000 *	701 :	-1.134	771 :	-0.585
633 :	0.080	702 :	2.493	772 :	0.196
634 :	-1.092	703 :	-0.613	773 :	1.062
635 :	-0.559	704 :	2.493	774 :	2.493
636 :	-1.114	705 :	-1.124	775 :	-0.613
637 :	-0.559	706 :	-1.114	776 :	-0.001
638 :	-0.533	707 :	0.196	777 :	0.196
639 :	0.083	708 :	-0.592	778 :	0.256
640 :	1.254	709 :	0.072	779 :	0.091
641 :	1.211	710 :	0.262	780 :	-0.613
642 :	-1.089	711 :	0.253	781 :	2.493
643 :	-1.101	712 :	-0.559	782 :	-1.092
644 :	-1.147	713 :	-0.635	783 :	0.935
645 :	-0.566	714 :	0.107	784 :	-1.098
646 :	2.493	715 :	0.107	785 :	0.332
647 :	0.933	716 :	0.034	786 :	2.493
648 :	-0.594	717 :	-1.000 *	787 :	1.222
649 :	0.196	718 :	-1.000 *	788 :	2.493
650 :	2.493	719 :	-1.000 *	789 :	-1.000 *
651 :	0.123	720 :	-1.000 *	790 :	0.272
652 :	0.083	721 :	-1.000 *	791 :	-1.098
653 :	2.493	722 :	-1.000 *	792 :	0.935
654 :	-0.586	723 :	-1.000 *	793 :	0.951
655 :	0.078	724 :	-1.000 *	794 :	0.091
656 :	-0.517	725 :	-1.000 *	795 :	-1.134
657 :	0.196	726 :	-1.000 *	796 :	-0.585
658 :	-0.585	727 :	-0.666	797 :	-0.568
659 :	-1.590	728 :	-0.666	798 :	0.196
660 :	-0.533	729 :	0.024	799 :	0.132
661 :	-1.119	730 :	-1.000 *	800 :	-0.582
662 :	-0.507	731 :	-1.000 *	801 :	1.222
663 :	-0.592	732 :	-1.000 *	802 :	0.935
664 :	-1.092	733 :	-1.000 *	803 :	0.071
665 :	-1.000 *	734 :	-1.000 *	804 :	0.196
666 :	0.083	735 :	-1.000 *	805 :	0.949
667 :	2.493	736 :	0.091	806 :	2.493
668 :	0.935	737 :	-0.574	807 :	-0.475
669 :	-0.599	738 :	-1.147	808 :	-1.098
670 :	0.078	739 :	-0.585	809 :	-0.588
671 :	0.196	740 :	-0.559	810 :	-1.114
672 :	-0.447	741 :	-1.098	811 :	-0.517
673 :	-0.545	742 :	-0.613	812 :	-0.559
674 :	0.951	743 :	0.276	813 :	0.081
675 :	-0.594	744 :	-0.533	814 :	0.084
676 :	-1.124	745 :	0.083	815 :	-0.579
677 :	1.031	746 :	-0.589	816 :	-0.585
678 :	1.002	747 :	0.141	817 :	-0.559
679 :	0.935	748 :	1.267	818 :	-1.119
680 :	1.002	749 :	0.332	819 :	-1.119
681 :	-0.495	750 :	1.254	820 :	2.493
682 :	2.493	751 :	2.493	821 :	-1.147
683 :	0.143	752 :	-0.559	822 :	0.284
684 :	1.256	753 :	-0.447	823 :	0.121
685 :	0.332	754 :	2.493	824 :	-0.495
686 :	-0.495	755 :	0.196	825 :	0.935
687 :	2.493	756 :	-0.559	826 :	-0.001
688 :	0.332	757 :	1.222	827 :	-0.478
689 :	-0.559	758 :	2.493	828 :	-1.147
690 :	2.493	759 :	-0.585	829 :	-0.507
691 :	0.935	760 :	0.285	830 :	-1.000 *
692 :	-0.613	761 :	-0.585	831 :	-1.000 *
		762 :	-1.094	832 :	-1.000 *

833	:	-1.000	*
834	:	-1.000	*
835	:	-1.000	*
836	:	-1.000	*
837	:	-1.000	*
838	:	-1.000	*
839	:	-1.000	*
840	:	-1.000	*
841	:	-1.000	*
842	:	-1.000	*
843	:	-1.000	*
844	:	-1.000	*
845	:	-1.000	*
846	:	-1.000	*
847	:	-1.000	*
848	:	-1.000	*
849	:	-1.000	*
850	:	-1.000	*
851	:	-1.000	*
852	:	-1.000	*
853	:	-1.000	*
854	:	0.221	
855	:	-1.174	
856	:	-1.174	
857	:	0.059	
858	:	-0.570	
859	:	-0.570	
860	:	-0.609	
861	:	0.107	
862	:	1.160	
863	:	-1.174	
864	:	-1.174	
865	:	-0.570	
866	:	-0.666	
867	:	-0.666	
868	:	0.095	
869	:	-1.000	*
870	:	-1.000	*
871	:	-1.000	*
872	:	-1.000	*
873	:	-1.000	*
874	:	-1.000	*
875	:	-1.000	*
876	:	-1.000	*
877	:	-1.000	*
878	:	-1.000	*
879	:	-1.000	*
880	:	-1.000	*
881	:	-1.000	*
882	:	-1.000	*
883	:	-1.000	*
884	:	-1.000	*
885	:	-1.000	*
886	:	-1.000	*
887	:	-1.000	*
888	:	-1.000	*
889	:	-1.000	*
890	:	-1.000	*
891	:	-1.000	*
892	:	-1.000	*
893	:	-1.000	*
894	:	-1.000	*
895	:	-1.000	*
896	:	-1.000	*
897	:	-1.000	*
898	:	-1.000	*
899	:	-1.000	*
900	:	-1.000	*
901	:	-1.000	*
902	:	-1.000	*

903	:	-1.000	*
904	:	-1.000	*
905	:	-1.000	*
906	:	-1.000	*
907	:	-1.000	*
908	:	-1.000	*
909	:	-1.000	*
910	:	-1.000	*
911	:	-1.000	*
912	:	-1.000	*
913	:	-1.000	*
914	:	-1.000	*
915	:	-1.000	*
916	:	-1.000	*
917	:	-1.000	*
918	:	-1.000	*
919	:	-1.000	*
920	:	-1.000	*
921	:	-1.000	*
922	:	-1.000	*
923	:	-1.000	*
924	:	-1.000	*
925	:	-1.000	*
926	:	-1.000	*
927	:	-1.000	*
928	:	-1.000	*
929	:	-1.000	*
930	:	-1.000	*
931	:	-1.000	*
932	:	-1.000	*
933	:	-1.000	*
934	:	-1.000	*
935	:	-1.000	*
936	:	-1.000	*
937	:	-1.000	*
938	:	-1.000	*
939	:	-1.000	*
940	:	-1.000	*
941	:	-1.000	*
942	:	-1.000	*
943	:	-1.000	*
944	:	-1.000	*
945	:	-1.000	*
946	:	-1.000	*
947	:	-1.000	*
948	:	-1.000	*
949	:	-1.000	*
950	:	-1.000	*
951	:	-1.000	*
952	:	-1.000	*
953	:	-1.000	*
954	:	-1.000	*
955	:	-1.000	*
956	:	-1.000	*
957	:	-1.000	*
958	:	-1.000	*
959	:	-1.000	*
960	:	-1.000	*
961	:	-1.000	*
962	:	-1.000	*
963	:	-1.000	*
964	:	-1.000	*
965	:	-1.000	*
966	:	-1.000	*
967	:	-1.000	*
968	:	-1.000	*
969	:	-1.000	*
970	:	-1.000	*
971	:	-1.000	*
972	:	-1.000	*

973	:	-1.000	*
974	:	-1.000	*
975	:	-1.000	*
976	:	-1.000	*
977	:	-1.000	*
978	:	-1.000	*
979	:	-1.000	*
980	:	-1.000	*
981	:	-1.000	*
982	:	-1.000	*
983	:	-1.000	*
984	:	-1.000	*
985	:	-1.000	*
986	:	-1.000	*
987	:	-1.000	*
988	:	-1.000	*
989	:	-1.000	*
990	:	-1.000	*
991	:	-1.000	*
992	:	-1.000	*
993	:	-1.000	*
994	:	-1.000	*
995	:	-1.000	*
996	:	-1.000	*
997	:	-1.000	*
998	:	-1.000	*
999	:	-1.000	*
1000	:	-1.000	*
1001	:	-1.000	*
1002	:	-1.000	*
1003	:	-1.000	*
1004	:	-1.000	*
1005	:	-1.000	*
1006	:	-1.000	*
1007	:	-1.000	*
1008	:	-1.000	*
1009	:	-1.000	*
1010	:	-1.000	*
1011	:	-1.000	*
1012	:	-1.000	*
1013	:	-1.000	*
1014	:	-1.000	*
1015	:	-1.000	*
1016	:	-1.000	*
1017	:	-1.000	*
1018	:	-1.000	*
1019	:	-1.000	*
1020	:	-1.000	*
1021	:	-1.000	*
1022	:	-1.000	*
1023	:	-1.000	*
1024	:	-1.000	*
1025	:	-1.000	*
1026	:	-1.000	*
1027	:	-1.000	*
1028	:	-1.000	*
1029	:	-1.000	*
1030	:	-1.000	*
1031	:	-1.000	*
1032	:	-1.000	*
1033	:	-1.000	*
1034	:	-1.000	*
1035	:	-1.000	*
1036	:	-1.000	*
1037	:	-1.000	*
1038	:	-1.000	*
1039	:	-1.000	*
1040	:	-1.000	*
1041	:	-1.000	*
1042	:	-1.000	*



The conservation score was calculated using prodata (al2co) database.

This database requires input in clustal format so Clustal W was used to get the alignment in clustal format.

The negative sign in the score indicates that the sequence is diverse at the particular point.

Using BioEditConserved Region Search we found:

2 most conserved regions:

Region 1: Position 123 to 128

Consensus:

123 RQYDEQ 128

Segment Length: 5

Region 2: Position 422 to 427

Consensus:

422 ANPQAF 427

Segment Length: 6

#### 4.5. Structure:

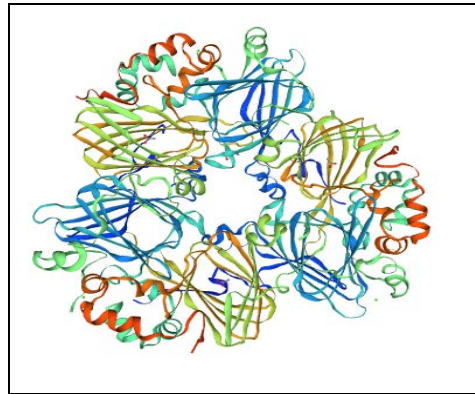


Fig. 4: Vicilin structure

Vicilin contains one Alpha subunit. The alpha subunits are linked in a head to tail manner and all feature coordination of a copper ligand. A trimeric structure of Vicilin consisting of repeating alpha subunits constitutes the known biologically active form.

The Secondary structure of Vicilin possesses four major structural motifs: Alpha helices, Beta sheets, 3/10 helices and Random coils.

Vicilin's formation of a conserved beta barrel through beta-strands is also an important aspect of its ability to coordinate a copper ligand.

The N-terminus and the C-terminus form cupin folds, a unique structural motif to this class of protein where a conserved beta barrel is formed.

Specifically, the intrinsic nature of 7S storage proteins to assume cupin folds remains the foremost area of interest regarding their allergenicity. Cupin folds cluster in seed storage proteins, and possession of a metal ligand imparts a catalytic functionality on the protein. The Cupin fold that is assumed by both the C-terminus and the N-terminus are symmetrically centered off of a pseudodyad axis. The axis formation for Vicilin is important in the head to tail arrangement seen in tertiary structure formation, which ultimately allows for the coordination of its novel copper ligand.

## 5. Conclusion

A study regarding the protein in nuts, Vicilin, a causative allergen was carried out.

Protein sequences for Vicilin in *Juglans regia* (English Walnut), *Anacardium occidentale* (Cashew), *Arachis hypogaea* (Peanut), *Corylus avellana* (Hazelnut), *Pistacia vera* (Pistachio)

were known and the use of various databases to do the same was understood. Use of various bioinformatics tools was inculcated throughout the process of understanding about the protein. A methodology was developed in order to study the changes that occurred in the protein using its various isoforms by Phylogenetic analysis. By studying various properties of

Vicilin protein its conservation of allergenic properties throughout the selected nuts species was known. The major future application could be reducing the allergenic properties of Vicilin by studying its structure, especially identifying the epitope and its chemical nature.

## References

- [1] Pei, J. and Grishin, N.V., 2001. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, 17(8), pp.700-712.
- [2] Astwood, J.D., Silvanovich, A. and Bannon, G.A., 2002. Vicilins: A case study in allergen pedigrees. *Journal of allergy and clinical immunology*, 110(1), pp.26-27.
- [3] Sanchez-Monje, B., López-Hernández, G., Pascual, M., and Pérez-Hernández, G., 2006. Convicilin and vicilin are potential major allergens from pea. *Clinical & Experimental Allergy*, 34(11), pp.1747-1753.
- [4] Vyas, V.K., Ukawala, R.D., Ghatge, M. and Chintha, C., 2012. Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, 74(1), p.1.
- [5] Hu, G., Koh, J., Yoo, M.J., Chen, S. and Wendel, J.F., 2015. Gene-expression novelty in allopolyploid cotton: a proteomic perspective. *Genetics*, 200(1), pp.91-104.
- [6] Ballmer-Weber, B.K., Lidholm, J., Lange, L., Pascal, M., Lang, C., Gernert, S., Lozano-Blasco, J., Gräni, N., Guillod, C., Wangorsch, A. and Hanschmann, K.M., 2019. Allergen recognition patterns in walnut allergy are age dependent and correlate with the severity of allergic reactions. *The Journal of Allergy and Clinical Immunology: In Practice*.
- [7] Dubiela, P., Kabasser, S., Smargiasso, N., Geiselhart, S., Bublin, M., Hafner, C., Mazzucchelli, G. and Hoffmann-Sommergruber, K., 2018. Jug r 6 is the allergenic vicilin present in walnut responsible for IgE cross-reactivities to other tree nuts and seeds. *Scientific reports*, 8(1), p.11366
- [8] Whitlock, B.A. and Baum, D.A., 1999. Phylogenetic relationships of Theobroma and Herrania (Sterculiaceae) based on sequences of the nuclear gene Vicilin. *Systematic Botany*, pp.128-138.
- [9] Gibbs, P.E., Strongin, K.B. and McPherson, A., 1989. Evolution of legume seed storage proteins--a domain common to legumins and vicilins is duplicated in vicilins. *Molecular biology and evolution*, 6(6), pp.614-623.
- [10] Silva, C.S. and Figueira, A., 2005. Phylogenetic analysis of Theobroma (Sterculiaceae) based on Kunitz-like trypsin inhibitor sequences. *Plant Systematics and Evolution*, 250(1-2), pp.93-104.
- [11] Sáenz de Miera, L.E., Ramos, J. and Pérez De La Vega, M., 2008. A comparative study of convicilin storage protein gene sequences in species of the tribe Viciaeae. *Genome*, 51(7), pp.511-523.
- [12] Jimenez-Lopez, J.C., Zafra, A., Palanco, L., Florido, J.F. and Alché, J.D.D., 2016. Identification and Assessment of the Potential Allergenicity of 7S Vicilins in Olive (*Olea europaea* L.) Seeds. *BioMed research international*, 2016.
- [13] Higgins, T.J., Newbiggin, E.J., Spencer, D., Llewellyn, D.J. and Craig, S., 1988. The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. *Plant Molecular Biology*, 11(5), pp.683-695.
- [14] <https://www.ncbi.nlm.nih.gov/>
- [15] <https://www.uniprot.org/>
- [16] <https://www.genome.jp/tools-bin/clustalw>
- [17] <http://prodata.swmed.edu/al2co/al2co.php>

**Sourabh Pujari:** Student at KIT's college of engineering, Biotechnology department.

**Guruprasad Bhandwale:** Student at KIT's college of engineering Biotechnology, department.

**Shantanu Deshpande:** Student at KIT's college of engineering, Biotechnology department.

**Vivek mali:** Student at KIT's college of engineering, Biotechnology department.