

Evolutionary study of Edestin Protein in Cannabis Sativa L

Guruprasad Bhandwale¹, Sourabh Pujari¹, Pradnya Kadam¹, Shivani Mantri¹, Poorva Kulkarni¹

¹Student, Dept. of Biotechnology, KIT's College of Engineering, Kolhapur, Maharashtra, India

Abstract - Edestin is seed storage protei and commonly observed in Hemp seed. It shows similarities with human globular proteins. The following methodologies were carried using respective materials for the phylogenetic study of Edestin. Multiple Sequence Alignment using MEGA-X was done then, Phylogenetic Tree was constructed by using Clustal-W, and Entropy Plot was constructed using Bio-edit. Conservation Score was found using A12CO. The application of this is it can be used to identify the genotoyoic alternation in translated products.

Key Words: Phylogenetic analysis, Edestin, Protein, Phylogenetic tree, Hemp seed, Cannabis sativa L.

1. INTRODUCTION

Edestin protein is found only in hemp seed. Hemp seeds can contain up to 36% protein. Hemp seeds are the fruit of the hemp plant. Scientists worked to determine lab models for foods; the vegetable protein model was derived from the hemp seed which was called Edestin. Hemp seeds are the seed of the Cannabis Sativa plant which, when grown for seed or fibre and with low THC, is known as hemp. The protein in hemp seed is comprised of approximately 65% of Edestin. Hemp seed is drugless and it is a healthy food item. Edestin is a highly digestible hexameric legumin protein. Edestin aids digestion and is also phosphorus-free. This protein is considered the backbone of the cell's DNAs. Edestin protein is similar to the human globular proteins present in blood plasma which produces antibodies that are vital to maintaining a healthy immune system. Since Edestin protein closely similar to the globulin in blood plasma, it is compatible with the human digestive system. This may be the reason why there are no reported food allergies to hemp food. A favourable amount of glutamic acid is found in Hemp seed. Glutamic acid is a neurotransmitter that helps people to overcome psychological and work-related stress. It supports the growth of animals in the absence of other dietary proteins.



Fig -1: Edestin in Hemp seed oil

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, Ref-Seq 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 SWISS MODEL

This software is used for the homology modelling of isoforms of edestin.

2.5 Relevant database:

Prodata, PDB, Computational biology unit.

3. METHODOLOGY

3.1 Sequence search

Sequences for edestin shall be retrieved from NCBI (Protein). Total sequences are expected to be found for isoform Edestin 1, Edestin 2, Edestin 3.

3.2 Multiple sequence alignment

The alignment shall be performed by using CLUSTAL W.

3.3 Phylogenetic analysis

MEGA X shall be used for creating the phylogenetic tree. For this, maximum the likelihood method shall be used.

3.4 Homology modeling

Swiss model shall be used for the homology modelling.

3.5 Conservation score

A12co prodata was used to obtain the conservation score of selected sequence.

3.6 Similarity and identity

Comparison done between edestin 1, edestin 2, edestin 3.

4. RESULT AND ANALYSIS

4.1 Multiple Sequence Alignment



Fig -2: Multiple sequence alignment

The multiple sequence alignment was carried out by retrieving the protein sequences from NCBI in Fasta format and then the sequences were aligned by using MEGA-X by using the build alignment option. The alignment that was obtained was again aligned by using CLUSTAL W to get the multiple sequence alignment.

4.2 Phylogenetic tree

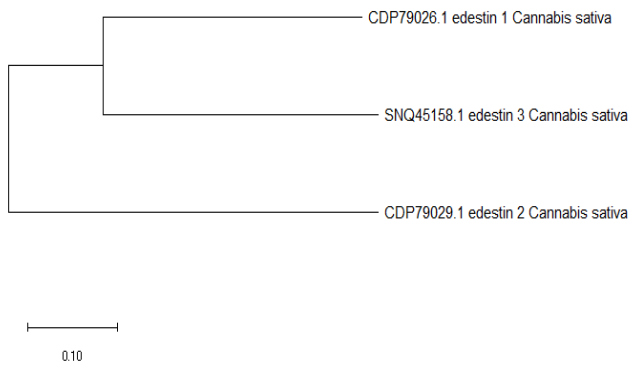


Fig -3: Phylogenetic tree

The phylogenetic tree was obtained based on previously obtained multiple sequence alignment using the MEGA-X Phylogeny option using the Maximum Likelihood method. By analysing the above tree we can take observations.

4.3 Homology Modeling

4.3.1 Edestin 2

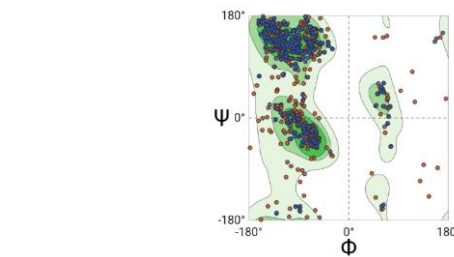
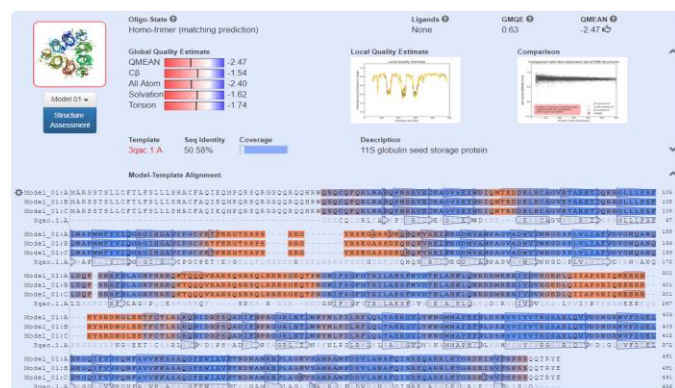


Fig -4: Ramachandran plot for the 3D structure of Edestin 2

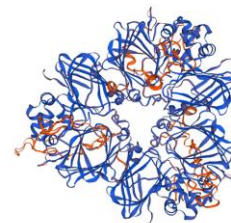


Fig -5: 3D structure of Edestin 2

4.3.2 Edestin 3

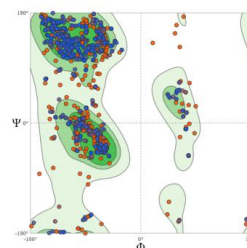
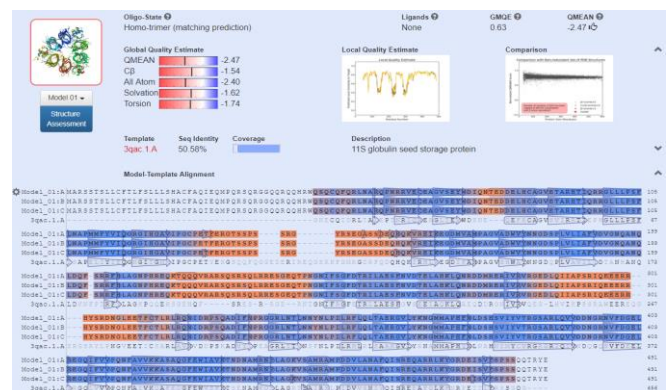


Fig -6: Ramachandran plot for the 3D structure of Edestin 3

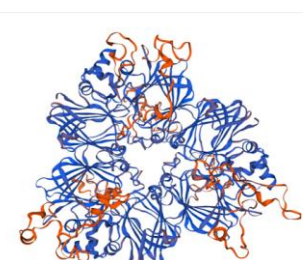


Fig -7: 3D structure of Edestin 3

4.4 Conservation score

1	:	-1.000 *	31	R	-0.451	61	T	-1.399
2	M	-0.413	32	:	-1.000 *	62	R	1.170
3	A	-0.413	33	:	-1.000 *	63	V	1.170
4	N	-0.398	34	:	-1.000 *	64	E	1.170
5	T	-0.413	35	:	-1.000 *	65	A	-0.398
6	K	-1.399	36	:	-1.000 *	66	E	1.170
7	A	-1.399	37	:	-1.000 *	67	A	-0.399
8	L	1.170	38	S	-0.451	68	G	1.170
9	L	1.170	39	Q	1.170	69	L	-1.399
10	S	-0.413	40	D	-1.399	70	I	-0.413
11	L	-0.413	41	E	-1.399	71	E	1.170
12	S	-0.413	42	S	-1.399	72	S	-0.413
13	F	-0.398	43	Y	-0.413	73	W	1.170
14	C	-0.413	44	R	-0.399	74	N	-0.413
15	F	-0.413	45	Q	-1.399	75	P	-0.413
16	F	-0.398	46	Q	1.170	76	N	-0.413
17	L	-0.399	47	N	-0.413	77	:	-1.000 *
18	L	1.170	48	Q	-0.399	78	:	-1.000 *
19	L	-0.413	49	C	1.170	79	:	-1.000 *
20	Q	-0.398	50	Q	1.170	80	H	-0.413
21	G	-0.413	51	I	-1.399	81	S	-1.399
22	T	-0.398	52	D	-1.399	82	Q	-0.413
23	S	-0.413	53	R	1.170	83	F	-0.413
24	A	1.170	54	I	-0.398	84	Q	-0.413
25	I	-1.399	55	E	-0.413	85	C	1.170
26	S	-0.413	56	A	1.170	86	A	1.170
27	R	-0.413	57	R	-0.399	87	G	1.170
28	S	-0.451	58	E	-0.413	88	V	1.170
29	R	-0.451	59	P	1.170	89	A	-0.413
30	S	-0.451	60	D	-0.413	90	V	-1.399
31	R	-0.451				91	V	-1.399
92	R	1.170	122	G	1.170	151	Q	-0.413
93	Y	-1.399	123	V	-1.399	152	P	-1.399
94	T	1.170	124	T	-1.399	153	D	-0.413
95	I	1.170	125	F	-0.413	154	R	-0.413
96	Q	1.170	126	P	1.170	155	H	1.170
97	Q	-1.399	127	G	1.170	156	Q	1.170
98	N	-0.413	128	C	1.170	157	K	1.170
99	G	1.170	129	P	-0.399	158	L	-1.399
100	L	1.170	130	E	1.170	159	R	1.170
101	H	-0.413	131	T	1.170	160	H	-1.399
102	L	1.170	132	F	1.170	161	V	-0.398
103	P	1.170	133	E	1.170	162	R	-0.398
104	S	1.170	134	E	-0.413	163	E	1.170
105	Y	-0.413	135	:	-1.000 *	164	G	1.170
106	T	-0.413	136	:	-1.000 *	165	D	1.170
107	N	1.170	137	:	-1.000 *	166	I	-0.413
108	T	-1.399	138	S	-0.399	167	V	-0.399
109	P	1.170	139	Q	-0.413	168	A	1.170
110	Q	-0.413	140	R	-1.399	169	I	-0.413
111	L	-0.413	141	G	-0.398	170	P	1.170
112	V	-1.399	142	Q	-1.399	171	A	1.170
113	Y	-0.399	143	G	1.170	172	G	1.170
114	I	-0.398	144	Q	-1.399	173	V	-0.399
115	V	-0.398	145	G	-0.413	174	A	1.170
116	K	-1.399	146	Q	-1.399	175	Y	-0.413
117	G	1.170	147	S	-0.413	176	W	1.170
118	R	1.170	148	Q	-0.451	177	S	-1.399
119	G	1.170	149	G	-0.451	178	Y	-0.399
120	I	-0.399	150	S	1.170	179	N	1.170
121	L	-0.413	151	Q	-0.413	180	N	-0.399
						181	G	1.170
						182	D	1.170

Fig -8: Conservation score

The conservation score was calculated using the data (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 a particular point.

4.5 Smilarity and Identity percentage

4.5.1 Edestin 1 and Edestin 2

Alignment length: 577

Percentage of similarity: 81.41%

Percentage of identity: 68.23%

4.5.2 Edestin 2 and Edestin 3

Alignment length: 545

Percentage of similarity: 79.49%

Percentage of identity: 63.78%

4.5.3 Edestin 1 AND Edestin 3

Alignment length: 542

Percentage of similarity: 89.90%

Percentage of identity: 69.24%

By calculating the similarity and identity percentage, we get to know that Edestin 1 & Edestin 3 are more related to each other.

5. CONCLUSION

A study regarding the isoforms of Edestin was carried out. Protein sequences for Edestin were known and the use of various databases to do the same was understood. The use of various bioinformatics tools was inculcated throughout the process of understanding the protein. A methodology was developed to study the changes that occurred in the protein using its various isoforms by evolutionary analysis. From the results we have got we get to know that the within the isoforms of Edestin, Edestin 1 & Edestin 3 are more related to each other. By studying various properties and function of Edestin the future application was known.

REFERENCES

- [1] Yousef, G.M., Elliott, M.B., Kopolovic, A.D., Serry, E. and Diamandis, E.P., 2004. Sequence and evolutionary analysis of the human trypsin subfamily of serine peptidases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1698(1), pp.77-86.
- [2] Shewry, P.R., Napier, J.A. and Tatham, A.S., 1995. Seed storage proteins: structures and biosynthesis. *The plant cell*, 7(7), pp.945-956.
- [3] Small, E., 2015. Evolution and classification of Cannabis sativa (marijuana, hemp) about human utilization. *The Botanical Review*, 81(3), pp.189-294.
- [4] Docimo, T., Caruso, I., Ponzoni, E., Mattana, M. and Galasso, I., 2014. Molecular characterization of the edestin gene family in Cannabis sativa L. *Plant physiology and biochemistry*, 84, pp.142-148
- [5] Ponzoni, E., Brambilla, I.M. and Galasso, I., 2018. Genome-wide identification and organization of seed storage protein genes of Cannabis sativa. *Biologia Plantarum*, 62(4), pp.693-702
- [6] Mamone, G., Picariello, G., Ramondo, A., Nicolai, M.A. and Ferranti, P., 2019. Production, digestibility and

- allergenicity of hemp (*Cannabis sativa* L.) protein isolates. *Food Research International*, 115, pp.562-571.
- [7] Tang, C.H., Ten, Z., Wang, X.S. and Yang, X.Q., 2006. Physicochemical and functional properties of hemp (*Cannabis sativa* L.) protein isolate. *Journal of agricultural and food chemistry*, 54(23), pp.8945-8950.
- [8] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(W1), pp.W252-W258
- [9] Thompson, J.D., Higgins, D.G. and Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), pp.4673-4680
- [10] Kumar, S., Nei, M., Dudley, J. and Tamura, K., 2008. MEGA: biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics*, 9(4), pp.299-306.
- [11] Aiello, G., Fasoli, E., Boschin, G., Lammi, C., Zanoni, C., Citterio, A. and Arnoldi, A., 2016. Proteomic characterization of hempseed (*Cannabis sativa* L.). *Journal of proteomics*, 147, pp.187-196.
- [12] Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C., 2003. SWISS-MODEL: an automated protein homology-modelling server. *Nucleic acids research*, 31(13), pp.3381-3385
- [13] Small, E., 2015. Evolution and classification of *Cannabis sativa* (marijuana, hemp) about human utilization. *The Botanical Review*, 81(3), pp.189-294
- [14] Orio, L.P., Boschin, G., Recca, T., Morelli, C.F., Ragona, L., Francescato, P., Arnoldi, A. and Speranza, G., 2017. New ACE-inhibitory peptides from hemp seed (*Cannabis sativa* L.) proteins. *Journal of agricultural and food chemistry*, 65(48), pp.10482-10488
- [15] Malomo, S.A., He, R. and Aluko, R.E., 2014. Structural and functional properties of hemp seed protein products. *Journal of food science*, 79(8), pp.C1512-C1521.

Poorva Kulkarni

Student at Biotechnology Department, KIT's college of Engineering, Kolhapur.

BIOGRAPHIES**Guruprasad Bhandwale**

Student at Biotechnology Department, KIT's college of Engineering, Kolhapur.

Sourabh Pujari

Student at Biotechnology Department, KIT's college of Engineering, Kolhapur.

Pradnya Kadam

Student at Biotechnology Department, KIT's college of Engineering, Kolhapur.

Shivani Mantri

Student at Biotechnology Department, KIT's college of Engineering, Kolhapur.