Applied Biological Research 17(3): 225-230; (2015)
DOI: 10.5958/0974-4517.2015.00033.6

DECIPHERING PROTEIN STRUCTURE IDENTIFIED IN CUMIN (*Cuminum cyminum* L.)

Sharda Choudhary and Geetika Jethra*

National Research Centre on Seed Spices, Tabiji, Ajmer - 305 206, Rajasthan (India)
*e-mail: *jethra08@gmail.com*

(Received 9 March, 2015; accepted 22 August, 2015)

ABSTRACT

Cumin (*Cuminum cyminum* Linn.), an important commercial seed spice with medicinal properties, belongs to family *Umbelliferae*. The present study was aimed to elucidate the high quality 3D structure and function of cumin protein (gn50). For the prediction of secondary structure of protein FASTA sequences were used and prediction was performed using GOR IV, SOPMA, ProFunc and PDBSum. The instability index 24.29, isoelectric point (pI) 5.14 and Grand average hydropathicity (GRAVY) 0.751 determined from primary and secondary structural analysis illustrated the stability of cumin protein. Stability of 3D structure of protein was confirmed using Ramachandran plot having 80.1% residues or empirically distributed data-points present in the structure lie in most favoured region with the total score of 98.3%. The models were validated using protein structure checking server SAVES.

Key words: Cumin, isoelectric point, protein, Ramachandran plot, secondary structure

INTRODUCTION

Cumin (*Cuminum cyminum* L.) is a seed spice belonging to the family *Umbelliferae* and is valued for its aroma and medicinal properties. Cumin and its added products are used in food flavouring and also these serve as good source of nutraceuticals with many biological activities (Sowbhagya, 2013). The plant is native to the Mediterranean region where it is cultivated extensively. After pepper, cumin is considered as the most important spice in the world (Thamaraikannan and Sengottuvel, 2012). Cumin seeds contain moisture (7%), volatile oil (3-4%), protein (12%), total ash (10%), fibre (11%), carbohydrate (33%), starch (11%) and fats (15%) (Weiss, 2002). Lipid extract from cumin has been studied by Hemavathy and Prabhakar (1988). Anon (1993) and Shaath and Azzo (1993) reported that the main constituents of Ejyptian cumin oil were cumin aldehyde, ß-pinene, γ-terpinene, ρ-mentha-1, 3-dien-7-al, ρ-mentha-1, 4-dien-7-al and ρ-cymene. Composition of cumin seed oil was studied by Rong Li and Jiang (2004) but there is no report about the protein structure of cumin. The proteins available in this seed spice need special attention as cumin is the major dietary component in the food as spice and is a potential source of dietary fibre (Sowbhagya, 2013). The functionality of a protein depends on its physical structure which is directly related to the combinatorial possibilities of their monomeric units *i.e.* the 20 amino acids.

In the present scenario, there is no information available on protein structure of cumin (*Cuminum cyminum*) involved in intraspecific variations. Unavailability of 3-dimensional structure of protein in cumin is one of the major hindrances in elucidating its interactions and functional analysis (Jethra *et al.*, 2015). Due to inherently time consuming and complicated nature of structure

determination techniques, only very few 3D structures have been solved experimentally (Zhang *et al.,* 2011). In spite of the great progress in structural genomics, it is still unreasonable to believe that the structure of more than a tiny fraction of all the billions of proteins will be studied by experimental methods in foreseeable future (Wallner and Elofsson, 2005). The performance of *in-silico* methods of protein structure prediction has recently improved significantly and dozens of servers and stand-alone programs are currently available (Fischer, 2006). The present investigation was aimed to elucidate and analyze the 3D structure of cumin protein.

MATERIALS AND METHODS

Plant materials

Cumin seeds cv. 'RZ-19' were grown in pots in the year 2014 in Biotechnology Lab, National Research Centre on Seed Spices, Ajmer (India). Young leaves of cumin were collected aseptically and genomic DNA isolated using standard procedure of CTAB method (Zidani, *et al.,* 2005).

DNA amplification and sequencing

Genomic DNA (50 ng) was taken to carryout PCR using 0.5 IU Taq DNA polymerase, and 1x reaction buffer supplied with the enzyme (containing 1.5 mM $MgCl_2$) in a 25 µL reaction volume using C-1000 thermal cycler (Bio-Rad). The programme involved initial denaturation at 94°C for 4 min., followed by 32 cycles at 94°C for 1 min., 1.5 min. at annealing temperature and 72°C for 2 min. This was followed by a final extension step at 72°C for 10 min (Choudhary *et al.,* 2015). The amplified PCR product was run on electrophoresis to check the amplification. After electrophoresis the gel was visualized under UV and photographs of gel taken with the help of gel documentation system (Biotron Healthcare, GelVision-DC). The selected bands were purified and sequenced.

Translation

The sequenced data was used to prepare protein structure using different translation tools (Transeq and Sixpack) to get six reading frames (Choudhary *et al*., 2014). The longest reading frame with no intervening stop codons was selected for further study. All the secondary structures predicted for gn50 protein were included in the tertiary structure.

Primary and secondary structure prediction

For cumin protein gn50 (protein Id given by PDBSum) primary structure prediction, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill and Von Hippel, 1989), instability index (Guruprasad *et al.,* 1990), aliphatic index (Ikai, 1980) and grand average hydropathicity (GRAVY) (Kyte and Doolottle, 1982) were computed using the Expasy's ProtParam server (*http://expasy.org/cgi-bin/protparam*). Secondary structure of this protein was predicted using the selected FASTA sequence of the longest ORF of protein using GorIV (Guermeur *et al.,* 1999).

Model building and evaluation

The modeling of three dimensional structure of protein was performed by *de novo* method "first principle" methods or "free modeling" (Zhang, 2008a) using I-TASSER software by taking default parameters (Jethra *et al.,* 2014). It is the hierarchical protein structure modeling approach based on profile-profile threading alignment (PPA) and threading assembly refinement (TASSER). In this method, template is not selected in one step, multiple iterations are performed and portions of highly similar sequences are taken and full length sequences generated by building unaligned regions by *ab-initio* modeling. (Zhang, 2008b) The best model for cumin protein was further

validated by PDBSum. Through PDBSum a comparison was also performed with protein models of other organism already present in its database. The prepared model was submitted to an online protein model database (PMDB) for getting accession number and documentation.

RESULTS AND DISCUSSION

Primary structure prediction

Primary structure prediction results showed that RZ-19 protein had 214 amino acid residues and the estimated molecular weight was 55028.6. The physiochemical properties were determined using Expasy's ProtParam server (*http://expasy.org/cgi-bin/protparam*). The calculated isoelectric point (pI) was useful because at pI the solubility was least and mobility in an electrofocusing system was zero. At pI (5.14) proteins were stable and compact. The computed pI value of cumin protein gn50 was 5.14 which was less than 7 (pI<7) indicating that the predicted protein was acidic. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains is regarded as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index for the protein sequence was 24.29. Whereas, Expasy's ProtParam computed the extinction coefficient for 280 nm wavelengths and 280 nm was favored because proteins absorbed light strongly there while other substances did not. Extinction coefficient of the protein at 280 nm was 9875 M^{-1} cm^{-1} with respect to the concentration of Cys residues. The computed extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution (Jethra *et al.*, 2014).

The instability index provides an estimate of protein stability in a test tube. A protein with <40 instability index value was predicted as stable and a protein with value ≥ 40 was predicted as unstable (Guruprasad *et al.*, 1990). The instability index value for cumin protein was found to be 24.29 which revealed that the identified protein was a stable. The grand average hydro-pathicity (GRAVY) value for a peptide or protein was calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. A GRAVY index for the protein was 0.751. This low range value indicated the possibility of better interaction with water.

Table 1: Type and density of residues present in cumin protein (gn50)

Clefts

	Volume	R1 ratio	Accessible vertices		Buried vertices		Average depth		Residue type						
1	11514.66	1.78	62.51	5	10.94	3	19.08	2	0	0	44	45	0	44	43
2	6461.02	0.00	73.09	2	14.05	2	15.22	3	0	0	38	26	0	32	36
3	4292.16	0.00	78.59	1	15.02	1	20.93	1	0	0	19	19	0	29	29
4	1156.36	0.00	64.09	3	10.80	4	11.24	5	0	0	6	4	0	8	13
5	904.50	0.00	31.12	10	2.36	10	9.23	7	0	0	3	4	0	1	4
6	741.66	0.00	55.55	8	5.28	8	11.59	4	0	0	7	3	0	4	4
7	775.83	0.00	50.70	9	2.67	9	10.53	6	0	0	4	6	0	3	4
8	723.09	0.00	56.25	7	10.52	5	8.37	9	0	0	4	2	0	3	7
9	486.00	0.00	60.83	6	8.27	6	8.83	8	0	0	4	4	0	2	3
10	579.23	0.00	62.96	4	7.52	7	6.57	10	0	0	11	1	0	2	3

Residue-type colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

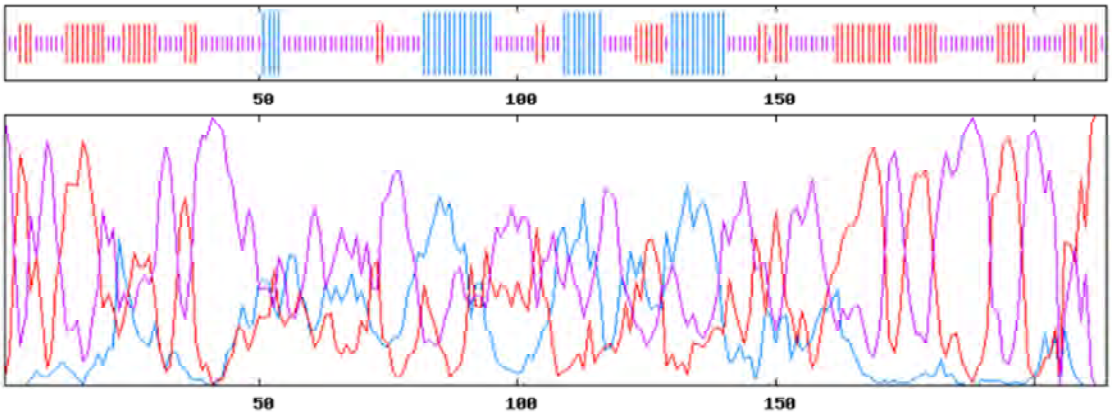

Fig. 1: Secondary structure composition of cumin protein gn50 by GORIV

Secondary structure prediction

Secondary structure of protein was predicted by the formation of α helix and ß-sheets. It was made up of two networks: a sequence-to-structure network and a structure-to structure net-work (Zhang, 2008b). The FASTA sequences were used to derive secondary structure of protein. The model preparation of protein's secondary structure was performed using GOR IV (Fig. 1), SOPMA, ProFunc and PDBSum (Fig. 2). The results revealed that random coil (52.34%) dominated among secondary structure elements and α helices (17.29%) and extended strand (30.37%) were also present.

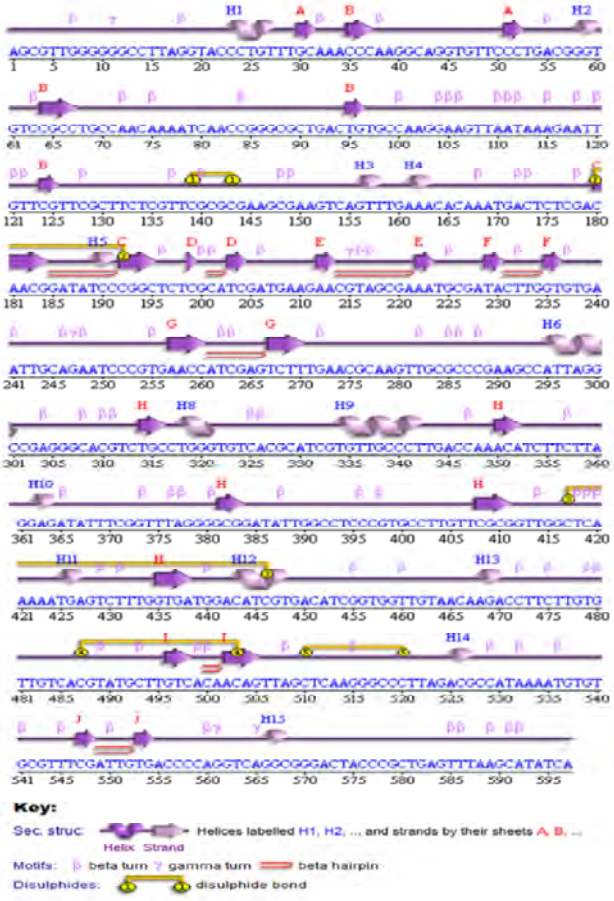

Fig. 2: Secondary structure of cumin protein gn50 by PDBSum

Tertiary structure prediction

The stability for 3D structure (Fig. 3) of cumin protein was further verified by SAVES server (Jethra *et al.*, 2012), which showed 80.1% residues or empirically distributed data-points present in the structure lie in most favoured region, 15.7% residues lied in additionally allowed regions [a,b,l,p] and 2.5% residues in generously allowed regions [~a,~b,~l,~p] of Ramachandran plot depicting the stability and good quality of protein 3D structure. The overall score of Ramachandran plot was 98.3. We found the conformation values of ϕ and ψ angles possible for all amino acid residues present in the protein. PDBSum also gave the types of residues and their density in predicted structure (Mihasan, 2010). It also showed volume, accessible vertices, buried vertices and average depth of each type of residue present (Table 1). The protein (gn50) predicted is with an unidentified structure and function in cumin. The prepared model was submitted to PMDB (Id No. PM0079101) for documentation. It gave the reliability score (accuracy) of 92%. The results showed only 38.9% match with *Xanthomonas*

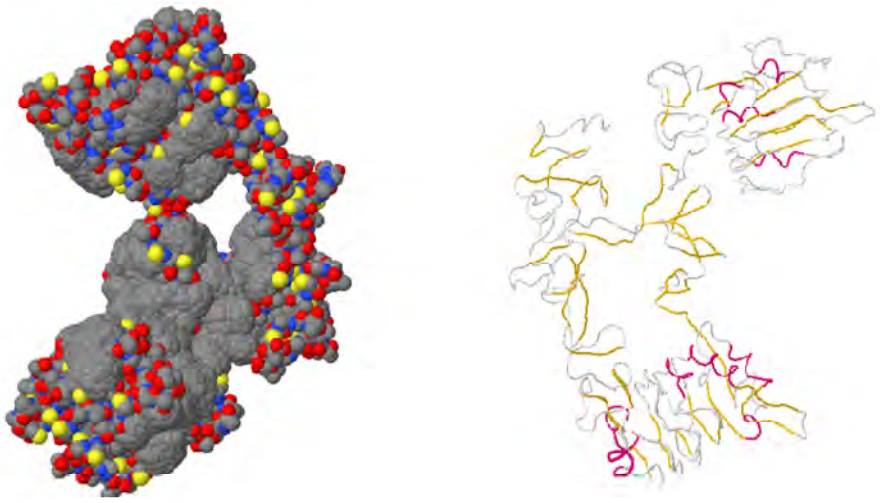

Fig. 3: Two different view of three dimensional structure cumin protein gn50 of RZ-19

and citri virb7 globular domain protein. FFPred server was used for its functional nature. FFPred predicted its functions in 2 categories: 1) Biological process: globular domain protein with a probability of 94.8% and 2) Molecular function: ATP binding with 96.5% probability.

Conclusion: The data generated by this study revealed that this protein is showing only 38.9% match with *Xanthomonas* and citri virb7 globular domain protein. This suggest the novelty of the protein and study, as there is no protein structure data available for cumin as well as related crops. Further, wet lab validation can be made for similar proteins to find their essentiality and importance.

REFERENCES

Anon, A. 1993. Application of gas-liquid chromatography to the analysis of essential oils, Part XVI: Monographs for five essential oils. A report of Analytical Methods Committee. *Analyst*, **118:** 1089-1098.

Choudhary, S. and Jethra, G. 2014. *In-silico* scrutiny of cumin (*Cuminum cyminum*) protein structure – GQ33, *Research on Crops*, **15**: 705-709.

Choudhary, S., Meena, R.S., Jethra, G., Sharma, R.S. and Panwar, A. 2015. Optimized methodology for high quality DNA isolation from leaves and seeds of fennel (*Foeniculum vulgare*). *Journal of Plant Developmental Science*, **72**: 173-175.

Fischer, D. 2006. Servers for protein structure prediction. *Current Opinion in Structural Biology*, **16**: 178-182.

Gill, S.C. and Von Hippel, P.H. 1989. Extinction coefficient. *Analytical Biochemistry*, **182**: 319-328.

Guermeur, Y., Geourjon, C., Gallinari, P. and Deleage, G. 1999. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, **15**: 413-421.

Guruprasad, K., Reddy, B.V.P. and Pandit, M.W. 1990. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Engineering*, **4:** 155-164.

Hemavathy, J. and Prabhakar, J.V. 1988. Lipid composition of cumin (*Cuminum cyminum* L.) seeds. *Journal of Food Science,* **53**: 1578-1579.

Ikai, A.J. 1980. Thermo stability and aliphatic index of globular proteins. *Journal of Biochemistry,* **88:** 1895-1898.

Jethra, G., Choudhary, S., Singh, P., Adwany, S. and Panwar, A. 2015. *In silico* approach to decode sequential language for structural scrutiny of cumin (*Cuminum cyminum* L.) protein. *Indian Journal of Agricultural Sciences,* **85**: 434-436.

Jethra, G., Mishra, A.K., Pandey, P.S. and Chandrasekharan, H. 2012. Structure and function prediction of unknown wheat protein using LOMENTS and I-TASSER. *Indian Journal of Agricultural Sciences,* **82**: 867-874.

Jethra, G., Mishra, A.K., Pandey, P. S., Choudhary, S. and Chandrasekharan, H. 2014. Phylogenetic and structural scrutiny of matK gene from wheat representing *Poaceae* family for DNA barcoding. *International Journal of Science and Nature,* **5**: 141-146.

Kyte, J. and Doolottle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology,* **157:** 105-132.

Mihasan, M. 2010. Basic protein structure prediction for the biologist: A review. *Archives of Biological Science Belgrade*, **62**: 857-871.

Rong, L.I and Jiang, ZI-Tao. 2004. Chemical composition of the essential oil of *Cuminum cyminum* L. from China. *Flavor and Fragrance Journal*, **19**: 311-313.

Shaath, N.A. and Azzo, N.R. 1993. Essential oil of Egypt. pp. 591-603. **In**: *Food Flavor Ingredients and Composition* (ed. G. Charalambous). Elsevier, Amsterdam, the Netherland.

Sowbhagya, H.B. 2013. Chemistry, technology and nutraceutical functions of cumin (*Cuminum cyminum* L): An overview. *Critical Reviews in Food Science and Nutrition,* **53**: 1-10.

Thamaraikannan, M. and Sengottuvel, C. 2012. Cumin: Can India maintain a monopoly? pp. 11-14. **In:** *Market Survey: Facts For You, A Magazine for Business and Economy*, May, 2012, Okhla, New Delhi

Wallner, B. and Elofsson, A. 2005. All are not equal: A benchmark of different homology modeling programs. *Protein Science,* **14**: 1315-1327.

Weiss, E.A. 2002. Umbelliferae. pp. 261-268. **In**: *Spice Crops*. CAB International, Wallingford, UK.

Zhang, X.D., Roy, A. and Zhang, Y. 2011. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based *ab initio* folding and FG-MD-based structure refinement. *Proteins,* **79(**10): 147-160.

Zhang, Y. 2008a. I-Tasser server for protein 3D structure prediction, *BMC Bioinformatics,* **9:** 40.

Zhang, Y. 2008b. Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology,* **18:** 342-348.

Zidani, S., Ferchichi, A. and Chaieb M. 2005. Genomic DNA extraction method from pearl millet (*Pennisetum glaucum*) leaves. *African Journal of Biotechnology,* **4**: 862-866.