Indian Journal of Agricultural Sciences **85** (3): 434–6, March 2015/Article

Structural and functional modeling of protein in cumin (*Cuminum cyminum*)

GEETIKA JETHRA, SHARDA CHOUDHARY, PRIYANKA SINGH, SHWETA ADWANY and ALKA PANWAR

National Research Centre on Seed Spices (ICAR), Ajmer, Rajasthan 305 206

Received: 30 June 2014; Revised accepted: 5 January 2015

ABSTRACT

Emergence of in-silico techniques have revolutionized the traditional methods of protein structure and function prediction of proteins. These computational tools encouraged the development of protein models in cumin (*Cuminum cyminum* L.) and have opened up new vistas in the area of research. In the present study, a high quality 3D structure and function of cumin RZ 19 protein have been predicted for the hypothetical amino acid sequence which showed homology with the protein domain of humans and *E. coli* illustrating that the database available on Apiaceae family is very low. The estimated molecular weight of identified cumin RZ 19 protein was 55028.6 and was predicted as an acidic protein with pI 5.14. Its functions are broadly classified into 2 categories: 1) Biological process: oxidation-reduction and ion transport with a probability of 98.9% and 97.8% respectively and 2) Molecular function: structural constituent of ribosome with 99.0% and oxidoreductase activity with 98.7% probability. We have also identified a channel in cumin transmembrane protein, through which a ligand (an ion or a small molecule) might pass. The present finding may be a valuable addition to the proteomic information available on cumin. Further validation can be performed using wet lab experiments.

Key words: *Cuminum cyminum,* Function, In-silico, Protein, RZ 19, Seed spices, Structure

Cumin (*Cuminum cyminum* L.) is one of the important and earliest seed spice known to mankind, which is also known as an ancient spice. Cumin is a flowering plant of the family Apiaceae. It is an herbaceous, dicotyledonous annual plant, diploid (2n = 2x = 14) and allogam with hermaphrodite flowers. Recent studies have indicated its pharmaceutical and medicinal importance. But in spite of huge benefits and importance, availability of sequential and structural data for cumin and other related seed spices crops is very limited in the public domain in comparison to other important crops. So, it is very important for the researchers to work and provide information on gene banks collection, identification, protein data bank, and evaluation of cumin and other seed spices. Agriculture and seed production also depends on the use of plant genotypes and proteomic studies (Bahraminejad *et al*. 2011).

In the present scenario, most of the sequence databases have very little information. Unavailability of 3 Dimensional structure of cumin protein is one of the major hindrances in elucidating its interactions and functional analysis. Due to the inherently time-consuming and complicated nature of structure determination techniques, only very few 3D structures have been solved experimentally (Dutta *et al*. 2009). In spite of 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 the foreseeable future (Wallner and Elofsson 2005). This places computer-based protein structure prediction in an unprecedentedly important position as the only reasonable means to bridge the gap between the number of known sequences and that of 3D models (MIHĂŞAN 2010). Ab-initio prediction is carried out when there is no suitable homologue found in the database. Prediction is done completely from the sequence.

Despite of having such immense significance, very small amount of studies are available citing the nucleotide sequences present in cumin but no reports as per our knowledge are available on in-silico studies for proteins structure supposed to be present in cumin. The aim of the present study was to determine protein structure from the unique sequences generated from the variety RZ 19. Further spatial position of every atom of protein molecules from the amino acid sequence by computational methods were estimated (Jethra *et al.* 2012).

MATERIALS AND METHODS

National Centre for Biotechnology Information (NCBI) a public domain database (http://www.ncbi.nlm.nih.gov/) was used to retrieve sequenced data available and to prepare a protein structure using different translation tools (Transeq and Sixpack) and perl script was developed to get six reading frames.

Primary structure prediction were computed using the Expasy's ProtParam server (http://expasy.org/cgi-bin/protparam), which included different parameters like theoretical isoelectric point (pI), molecular weight, total

[1]email: gjethra08@gmail.com

number of positive and negative residues, extinction coefficient (Gill *et al.* 1989), instability index (Guruprasad *et al.* 1990), aliphatic index (Ikai 1980) and grand average hydropathicity (GRAVY) (Kyte *et al.* 1982).

Secondary structure of this protein was predicted by LOMETS and PSIPRED software's using the translated amino acid sequences in FASTA format (Guermeur *et al.* 1999)

Ab-initio and threading Meta servers (Zhang 2008) were used for modeling the three dimensional structure of the protein.

RESULTS AND DISCUSSION

Basic structural prediction

RZ 19 is the most important cultivar of cumin. 18S ribosomal amino acid sequence was determined by translation. Results showed that RZ 19 protein has 214 amino acid residues and the estimated molecular weight is 55028.6. In the present study, physiochemical properties were determined using Expasy's ProtParam server (http://expasy.org/cgi-bin/protparam). The calculated isoelectric point (pI) will be useful because at pI, solubility is least and mobility in an electrofocusing system is zero.

Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. At pI proteins are stable and compact. The computed pI value of 18S ribosomal protein was 5.14. Computed pI value of protein were less than 7 (pI<7) indicates that 18S ribosomal protein is considered as 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 18S ribosomal protein sequence was 24.29. Whereas, Expasy's ProtParam

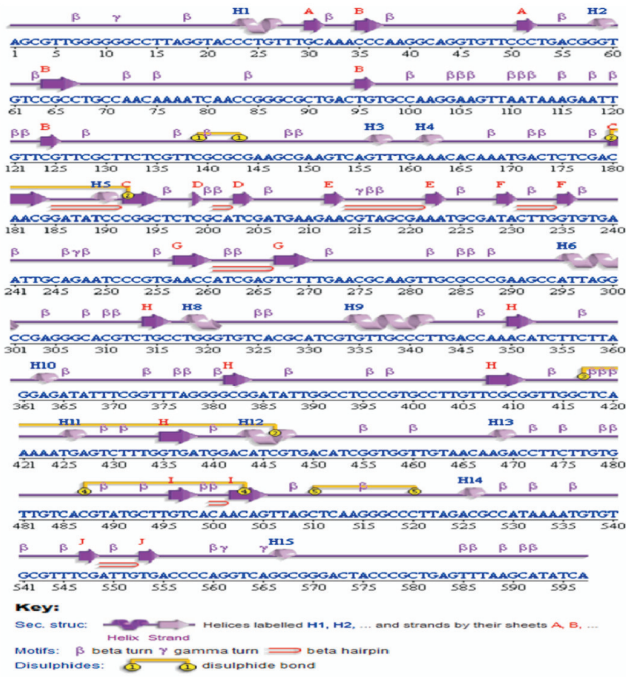

Fig 1 Secondary structure of cumin RZ 19 protein by PDBSum

```
Sequence length :   214

GOR4 :
   Alpha helix      (Hh) :    37 is  17.29%
   3(10) helix      (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    65 is  30.37%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   112 is  52.34%
   Ambigous states  (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

Fig 2 Secondary structure composition of cumin RZ 19 protein by GORIV

computes the extinction coefficient for 280 nm wavelengths and 280 nm is favoured because proteins absorb light strongly there while other substances commonly in protein solutions do not. Extinction coefficient of 18S ribosomal protein at 280 nm is 9 875/M cm 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.

The instability index provides an estimate of the stability of protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad *et al.* 1990). The instability index value for cumin 18S ribosomal protein was found to be 24.29 indicates 18S ribosomal

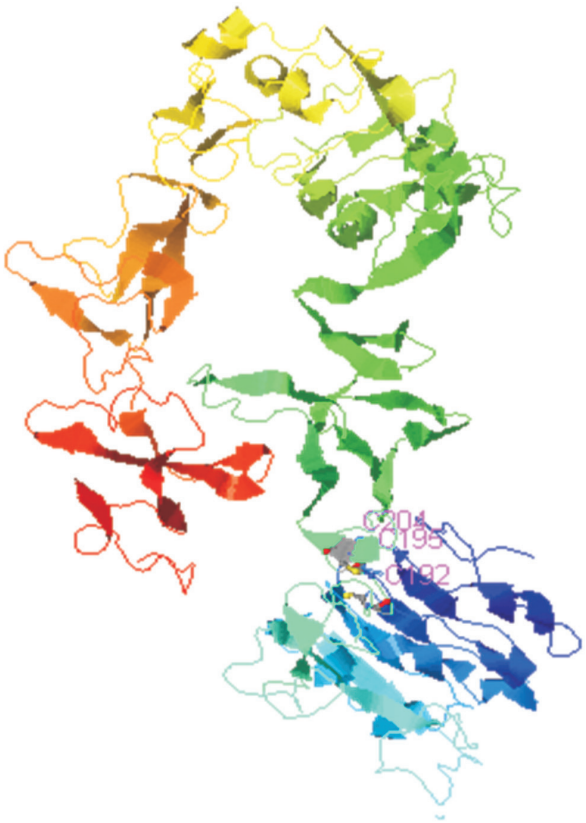

Fig 3 Three dimensional structure of cumin protein of RZ 19

protein as stable protein. The Grand Average hydropathicity (GRAVY) value for a peptide or protein is 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 cumin protein is 0.751. This low range of value indicates the possibility of better interaction with water.

Secondary structure of protein comprises of alpha helix and beta sheet. It is made up of two networks: a sequence-to-structure network and a structure-to structure network. Secondary structure was predicted by using the FASTA sequence of cumin protein RZ 19. The prediction of secondary structure was performed using GOR IV, SOPMA, ProFunc and PDBSum (Fig 1). The results revealed that random coil (52.34 %) dominated among secondary structure elements and alpha helices (17.29 %) and extended strand (30.37 %) are also present (Fig 2).

Model building and evaluation

The modeling of the three dimensional structure of cumin protein from RZ 19 cumin cultivar was performed by I-Tasser and PDBSum, an Ab-initio modelling programs. Energy minimization for the 3D structure was performed by ChemSite Pro and PYMOL server for further increasing the stability of the structure.

Stability of the predicted 3D structure was further verified by SAVES server, which showed 80.1 % residues in most favoured region of Ramachandran plot showing the stability and good quality of protein 3D structure.

The prepared model was submitted to PMDB (Id No. PM0079073_A) for identification and functional annotation of the protein. The results were showing similarity match with *CATH domain superfamily* Single helix bin (4k5q, transferase) pDomTHREADER. FFPred predicted its functions in 2 categories: 1) Biological process: oxidation-reduction and ion transport with a probability of 98.9% and 97.8% respectively and 2) Molecular function: structural constituent of ribosome with 99.0% and oxidoreductase activity with 98.7% probability.

The present finding is first report of in-silico protein modeling for cumin. The identified cumin RZ 19 protein is unique and having homology only with the reported human protein domains, which have happened due to low strength of cellular database available on cumin or other related crops.

ACKNOWLEDGEMENT

The authors are thankful to Director, NRCSS for providing basic facilities.

REFERENCES

Bahraminejad A *et al.* 2011. Study of genetic diversity of cumin (Cumin cyminum L.) based on phenotypic characteristics. *Australian Journal Crop Sciences* **5**: 301–7.

Dutta S *et al.* 2009. Data deposition and annotation at the worldwide protein data bank. *Molecular Biotechnology* **42**: 1–13.

Gill S C *et al.* 1989. Extinction coefficient. *Analytical Biochemistry* **182**: 319–28.

Guermeur Y *et al.* 1999. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15**: 413–21.

Guruprasad K *et al.* 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**: 155–64.

Ikai A J. 1980. Thermo stability and aliphatic index of globular proteins. *Journal of Biochemistry* **88**: 1 895–8.

Jethra G *et al.* 2012. Structure and Function prediction of unknown wheat protein using LOMENTS and I-TASSER. *Indian Journal of Agricultural Sciences* **82**: 867–74.

Kyte J *et al.* 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**: 105–32.

Mihăşan M. 2010. Basic protein structure prediction for the biologist: A review. *Archives of Biological Sciences* **62**: 857–71.

Wallner B *et al.* 2005. All are not equal: a benchmark of different homology modeling programs. *Protein Science* **14**: 1 315–27.

Zhang Y. 2008. Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology* **18**: 342–8.