In Silico Structural, Functional and Phylogenetic Analyses of Corynebacterium Aspartokinase: An Enzyme of Aspartate Family of Amino Acids

Krishnendu Pramanik, Pooja Rajbhar, Tithi Soren, Tushar Kanti Maiti*
Microbiology Laboratory, Department of Botany,
The University of Burdwan
Burdwan, PIN-713104, West Bengal, India
Corresponding author: tkmbu@yahoo.co.in

Abstract—Aspartokinase is the enzyme of the aspartate family of amino acid biosynthesis that catalyzes first reaction in biosynthesis of Aspartate family. This enzyme has an important role for hyper production of the Lysine or methionine or threonine by microbial fermentation. The present study was undertaken to determine the *in silico* analysis of the aspartokinase of *Corynebacteriumglutamicum*. *In silico*analysis of*Aspartokinase* protein of *Corynebacteriumglutamicum*revealed that it is a thermostable, acidic protein having molecular weight of about 45kDa. The 3D structure and protein model with functional characters of this protein has predicted. This computational analysis might assist with useful information required during wet lab experiments.

Keywords-Aspartokinase; *Corynebacterium*; *in silico* analysis; microbial fermentation

*****

I. INTRODUCTION

Aspartokinase (EC 2.7.2.4) is the first enzyme of the aspartate family of amino acids biosynthesis that catalyzes the phosphorylation of the beta carboxyl group of aspartic acid with ATP to yield 4-phospho-L-aspartate, which is involved in the branched biosynthetic pathway leading to the biosynthesis of amino acids *viz.* lysine, threonine, isoleucine and methionine. Lysine is one of the earliest amino acids to be produced by large scale by microbial fermentation. Interestingly this enzyme has an important role for hyper production of the Lysine or methionine or threonine as it has been regulated and represses by end products.

Amino acids are main components of protein which is one of the three essential nutritional components. All amino acids occurring in proteins are in L-form and only amino acids of this form are utilized in the living body. Thus L-amino acids of usually required for human and animal consumption [1]. L-amino acids have wide spectrum of commercial use as food additives , food supplements, infusion compounds therapeutic agents and precursors for the synthesis of peptides or agrochemicals [2]. The taste of the food is an important factor as their nutritional value, considering that amino acid provides one of the contributions to food taste. L-glutamic acid which is ubiquitous in animal and vegetables can enhance or improve the taste and flavor of the natural food stuffs [1]. It is also used in various industries for the manufacture of various chemicals. Some amino acids playing specific physiological role and which are used for therapeutic purposes. Amino acids are also utilized as raw materials for the synthesis of peptides used as medicine and diagnostics such as glutathione, thyrotropin releasing hormone (TRH),pentagastrin peptide, luteinizing hormone releasing hormone (LHRH), and so on [1]. As technology advances and the understanding of the functions and propertiesof amino acids increases, the commercial applications of amino acids are also increasing. New production technology and large scale production of amino acids are making it more economical and thereby increasing its user base and usage rate.

Some of the companies in this industry are Ajinomoto, RSP amino acids, Biaffin, AnaSpec, ChemPepInc, IRIS Biotech, PepTech Corporation and Synthetech.. Ajinomoto is one of the leading manufacturers of amino acids.

*Corynebacterium*is an aerobic gram positive rod shaped bacterium which is used for the industrial production of amino acids e.g. L-lysine and L-glutamate. *C.glutamicum* is widely used for the biotechnological production of > 1,500,000 tons of L-glutamate per year and > 750,000 tons of L-lysine per year and several other amino acids [3]. Moreover, aspartokinase is the only enzyme involved in the lysine biosynthesis in *C.glutamicum* which is known to be controlled in its activity. Analysis of lysine formation revealed that overexpression of the gene for the feedback resistant aspartokinase alone suffices to achieve lysine secretion in the wild type [2].

Considering the importance of this enzyme in Aspartate family of amino acids and lysine overproduction the present study was undertaken to determine the *in silico* analysis of the Aspartokinase of *Corynebacterium*. The present investigation envisaged the computational prediction of secondary, tertiary structure of Aspartokinase protein of *C.glutamicum* and their functional characterization including protein-protein interaction. Attempts were also made to predict the phylogenetic relation of the protein among the different spp. and strains of *Corynebacterium*.

II. MATERIALS AND METHODS

A. Sequence retrieval and cDNA acquisition

The National Center for Biotechnology Information (NCBI) (http://www.ncbi.nml.nih.gov/) houses a series of databases relevant to biotechnology and biomedicine and is an important resource for the bioinformaticsstudies, was used for the sequence retrieval of the different species of *Corynebacterium*having aspartokinase activity. Twenty

fouraspartokinase protein sequences and 16s rDNAsequence of24 different species of *Corynebacterium*were selected and downloaded in FASTA format.Further 73 aspartokinase protein sequence of *Corynebacteriumglutamicum* were retrieved from the NCBI for computational investigation.The cDNAsequences of 24 species of *Corynebacterium* was obtained through expasy reverse translate tool (www.expasy.org/tools) due to absence of those sequences in NCBI database.

B. Phylogenetic analysis

From the retrieved sequences from NCBI and the reverse translates, five different phylogenetic trees were constructed using the MEGA7 (7.0.18) software [4] for the evolutionary comparisons of the related taxa. All the trees were constructed using the Neighbor-joining method and the evolutionary distances were computed [5].

C. Physiochemical characterization

Physiochemical properties like amino acid composition, instability indices, aliphatic indices, extinction coefficients and grand average of hydropathicity (GRAVY) of the aspartokinase protein sequences retrieved from NCBI were calculated from ExpasyProtparam tool [6].

D. Secondary structural analysis

The protein **WP_040967888.1** was selected among the 73 protein sequences of Aspartokinase of *C. glutamicum*, as the representative of all the protein sequences for detailed structuralanalysis. PSIPRED v3.3 (psipred@cs.ucl.ac.uk) and Chou &Fasman (http://www.biogem.org/tool/chou-fasman/) [7-8] servers were used to predict the overall secondary structure and to count the number of secondary element *viz.* helices,sheets, turns.

E. Homology modeling and evaluation

The same protein (**WP_040967888.1)** was selected to predict the tertiary structure of Aspartokinase protein. The SWISS-MODEL, automated protein structure homology-modeling server(https://swissmodel.expasy.org/) was used for modeling the protein of interest as well as for quality analysis of the built model.Cross evaluationwas also performed using SAVES server (https://services.mbi.ucla.edu/SAVES/). Ramachandran Plot was constructed by using thepdb file in SAVES server (http://services.mbi.ucla.edu/SAVES/Ramachandran/) and RAMPAGE(http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [9] for the visualization of the energetically allowed regions for back bone dihedral angles against the amino acid residues in Aspartokinase protein.

F. Deposition of model to database

After the building of tertiary model of **WP_040967888.1**obtained through SWISS-MODEL, it was deposited to Protein Model Database (PMDB) to gain the accession number for further studies.

G. Functional analysis

The functional motifs present in selected protein's (**WP_040967888.1)**amino acid sequence was identified by motif search (http://www.genome.jp/tools/motif/) tool. Prediction of functionally interacting partners of the protein was performed by STRING (http://string-db.org/) analysis [10].

III. RESULTS AND DISCUSSION

A. Sequence retrieval and cDNA acquisition

Amino acid sequences ofAspartokinase and 16s rDNA of 24 different species of*Corynebacterium*and 73 amino acid sequences of Aspartokinase of *C. glutamicum* were retrieved from the NCBI database. This includes both partial and few complete Aspartokinases amino acid sequences. Aspartokinasecatalyzes the phosphorylation of the β-carboxyl group of aspartic acid with ATP to yield 4-phospho-L-aspartate, which is involved in the branched biosynthetic pathway leading to the biosynthesis of amino acid lysine, threonine, isoleucine and methionine. Similar practice of *in silico* study was also studied previously by Verma et al [11] and Pramanik et al[10, 12].

B. Phylogenetic analysis

For the comparison of the evolutionary relationship among the 24 different species of *Corynebacterium*, three phylogenetic trees were constructed by MEGA7. Among the three phylogenetic trees, one of which contains the amino acid sequence of Aspartokinase and the another one contains the 16s rDNA sequence of different *Corynebacterium* species while the third one contains the gene sequence of Aspartokinase of different *Corynebacterium* species, from this analysis it was found that *C. glutamicum* (BAV220491) clustered with *C. flavescens* (AAA23293.1) which shows the 100% similarity. This cluster again clustered with *C.deserti* (WP00676868) and shows 96% and this clustered again clustered with *C.efficiens* (WP006768468.1) which shows the 100% similarity (Fig. 1). In the tree of 16s rDNA*C.glutamicum* ATCC13032 (NR074663) clustered with *C.deserti* GIMNI 1.010 (NR118005.1) which shows the 100% similarity , this cluster again clustered with *C.efficiens* YS-314(NR118005.1) and shows 100% similarity (Fig. 2). In the phylogenetic tree of genes of Aspartokinase protein, *C.glutamicum* clustered with *C.diphtheria* which shows the 100% similarity (Fig. 3). Moreover, for the comparison of evolutionary relationship among the 73 selected amino acid sequence two phylogenetic tree were constructed, one contained the 73 amino acid sequence of Aspartokinase of *C. glutamicum* and another contains reverse translate gene of 73 amino acid sequence of Aspartokinase. From this analysis it was found that in the tree of amino acid sequence, WP_040967888.1 clustered with AJE68295.1 which shows 99% similarity (Fig. 4) and in the tree of reverse translate gene of Aspartokinase WP_040967888.1 clustered with AJE68295.1 which shows 89% similarity (Fig. 5). Similar phylogenetic based in silico study was also performed by Verma et al [11] and Pramanik et al [10, 12].

ISSN: 2321-8169
981 – 987

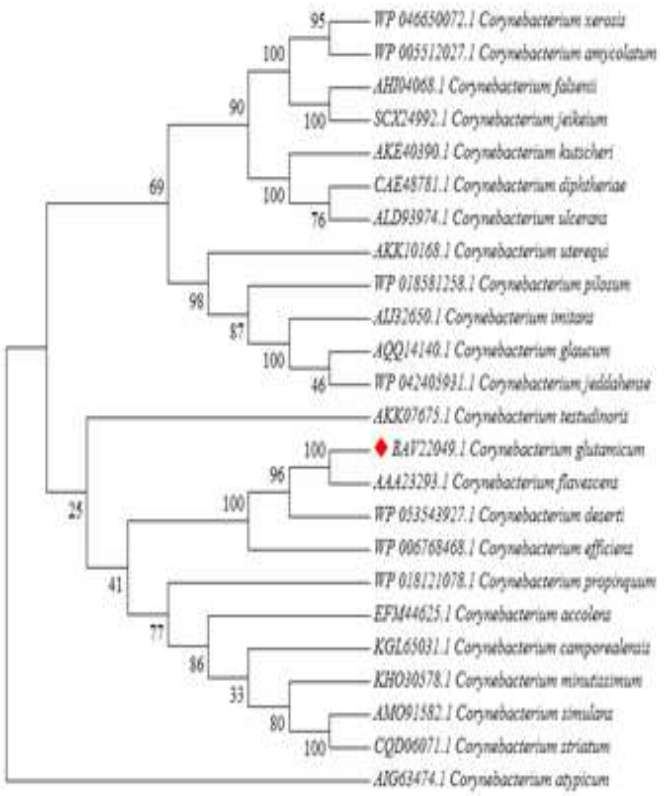

Figure 1.Phylogenetic tree of aspartokinases of different species of Corynebacterium

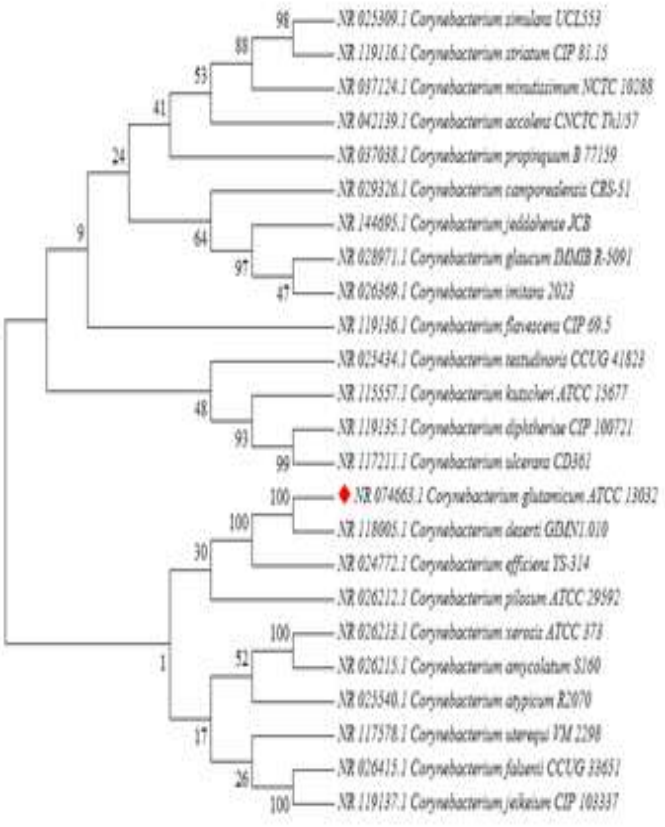

Figure 2.Phylogenetic tree of 16s rDNA sequences of different species of Corynebacterium

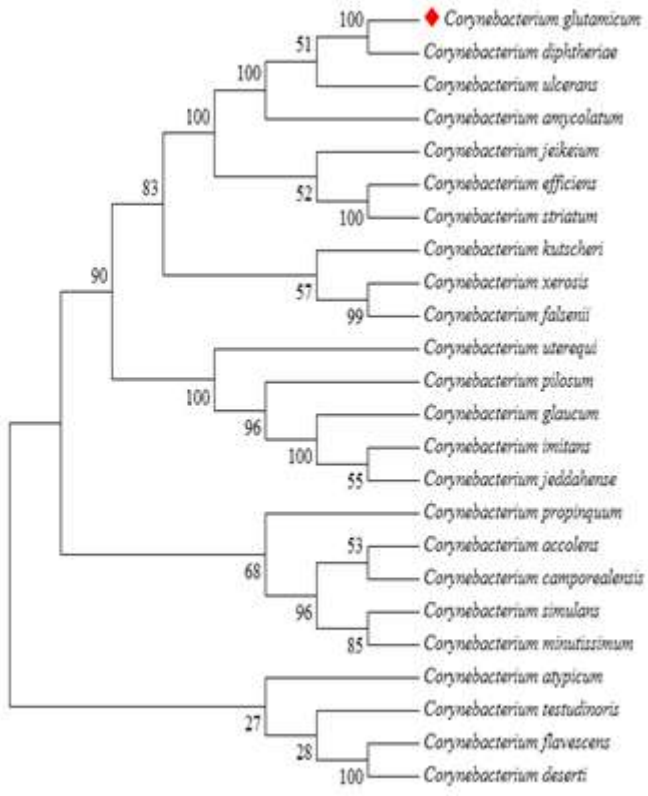

Figure 3.Phylogenetic tree of reverse translate gene of Aspartokinase protein of different species of Corynebacterium

C. Physiochemical characterization

The physiochemical properties of the different species of the *Corynebacterium* were tabulated in Table 1. It includes amino acid residues, theoretical isoelectric points, molecular weights, instability indices, extinction coefficient and grand average of hydropathicity. The amino acid residues of 24 different species were 421. The isoelectric points were ranging from 4.42 to 5.05 that showed Aspartokinaseproteins of *Corynebacterium*species were acidic in nature. If the instability indices of protein greater than 40 that showed the protein are unstable. Here instability indices ranging from 23.23 to 37.22 based on this analysis it was found that all the proteins of different species of *Corynebacterium* is stable[10-12].

D. Secondary structure analysis

The percentage of three classesof secondaryarrangement helix (73.0%), sheet (40.6%), turn (16.6%), were determined from the web server which indicatesthat helix> sheet > turns in the protein (Fig. 6). There was no disordered protein binding site present (Fig. 6). Secondary arrangement shows that the proteins are not unfolded which indicates that the proteins are stable in nature. It was found that the thermophiles have a large fraction of the α-helix in their residue so that they can survive in high temperature [13]. Here in this secondary structure analysis α-helical conformation was also higher which indicates the thermos table nature of the protein.

ISSN: 2321-8169
981 – 987

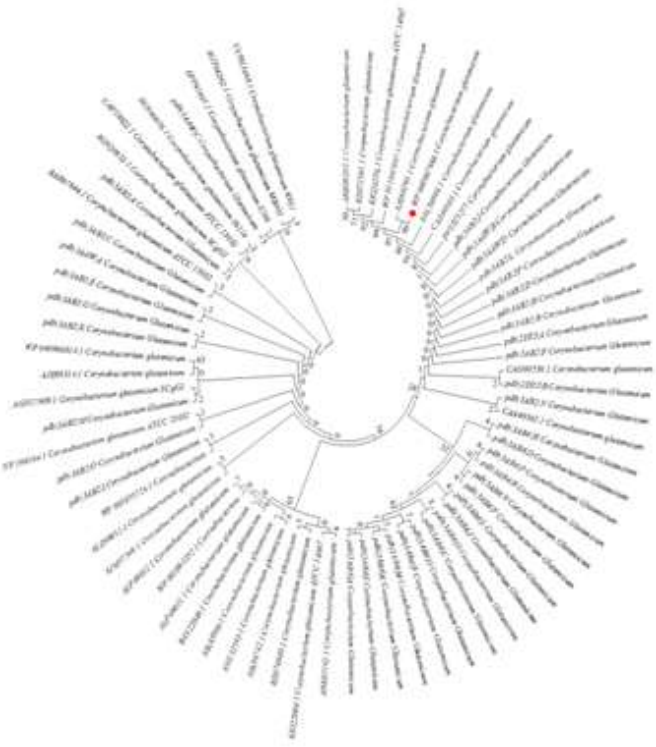

Figure 4.Phylogenetic tree of73 amino acid sequence of Aspartokinase of Corynebacteriumglutamicum

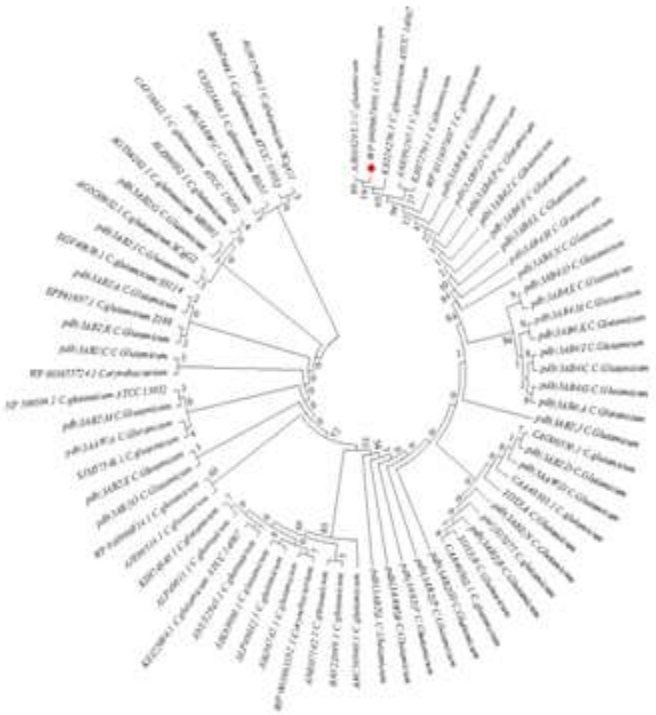

Figure 5.Phylogenetic tree of reverse translate gene of 73 amino acid sequence of Aspartokinase protein ofCorynebacteriumglutamicum

Table 1 Physicochemical characterization of different spp. of *Corynebacterium*

Name of bacteria	pI	MW (kDa)	II	AI	EC	GRAVY
C. glutamicum	4.65	44.73	27.20	99.14	0.36	0.007
C. xerosis	4.64	45.38	30.04	100.97	0.36	-0.003
C. accolens	4.74	45.86	26.85	92.78	0.35	-0.074
C. diphtheriae	4.73	44.73	32.97	97.32	0.36	-0.010
C. falsenii	4.71	45.29	30.68	93.80	0.32	-0.094
C. flavescens	4.63	44.79	37.04	99.14	0.36	-0.000
C.deserti	4.62	44.65	28.38	98.91	0.36	0.008
C. ulcerans	4.89	50.39	36.15	95.05	0.44	-0.065
C.simulans	4.71	44.76	24.37	96.63	0.36	0.013
C. efficiens	4.65	44.79	23.78	94.28	0.36	-0.054
C. striatum	4.65	44.72	25.66	95.46	0.36	-0.013
C.glaucum	4.42	44.44	26.79	94.58	0.48	-0.017
C. jeikeium	4.71	45.15	25.39	94.99	0.32	-0.048
C. atypicum	5.05	44.8	26.35	97.29	0.36	-0.063
C. imitans	4.54	44.68	23.26	95.49	O.39	-0.011
C. jeddahense	4.60	44.55	29.36	96.67	0.36	0.007
C.camporealensis	4.59	44.63	25.95	94.28	0.39	-0.039
C. minutissimum	4.67	44.55	26.53	95.23	0.36	-0.014
C. uterequi	4.67	44.25	25.32	97.13	0.36	0.049
C.testudinoris	4.74	44.59	23.23	96.63	0.49	0.007
C.kutscheri	4.68	44.66	37.22	96.39	0.36	-0.007
C. propinquum	4.68	44.69	26.15	96.82	0.36	-0.054
C. pilosum	4.48	44.66	26.60	94.09	0.35	-0.048
C.amycolatum	4.61	44.96	29.05	94.28	0.36	-0.053

Note. pI=Isoelectric point, MW=Molecular weight, II=Instability index, AI=Aliphatic index, EC=Extinction coefficient, GRAVY=Grand average of hydropathicities

ISSN: 2321-8169
981 – 987

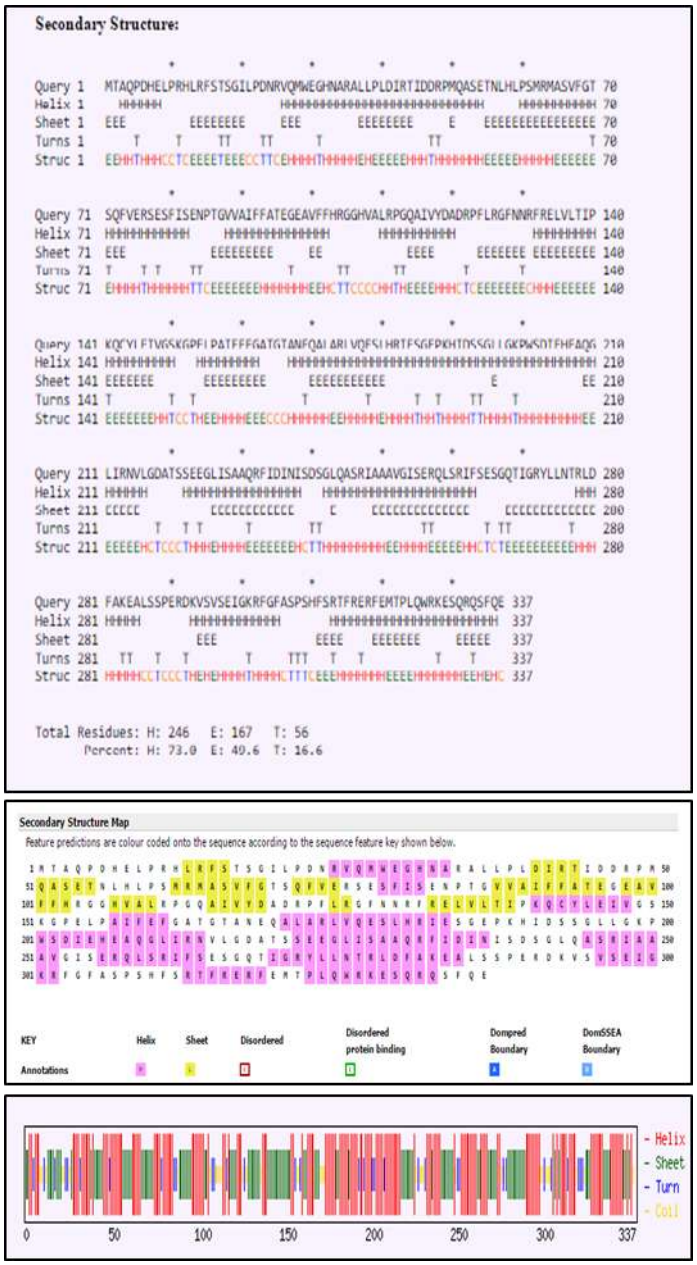

Figure 6. Predicted secondary structures showing percentages of helices, sheets and turns

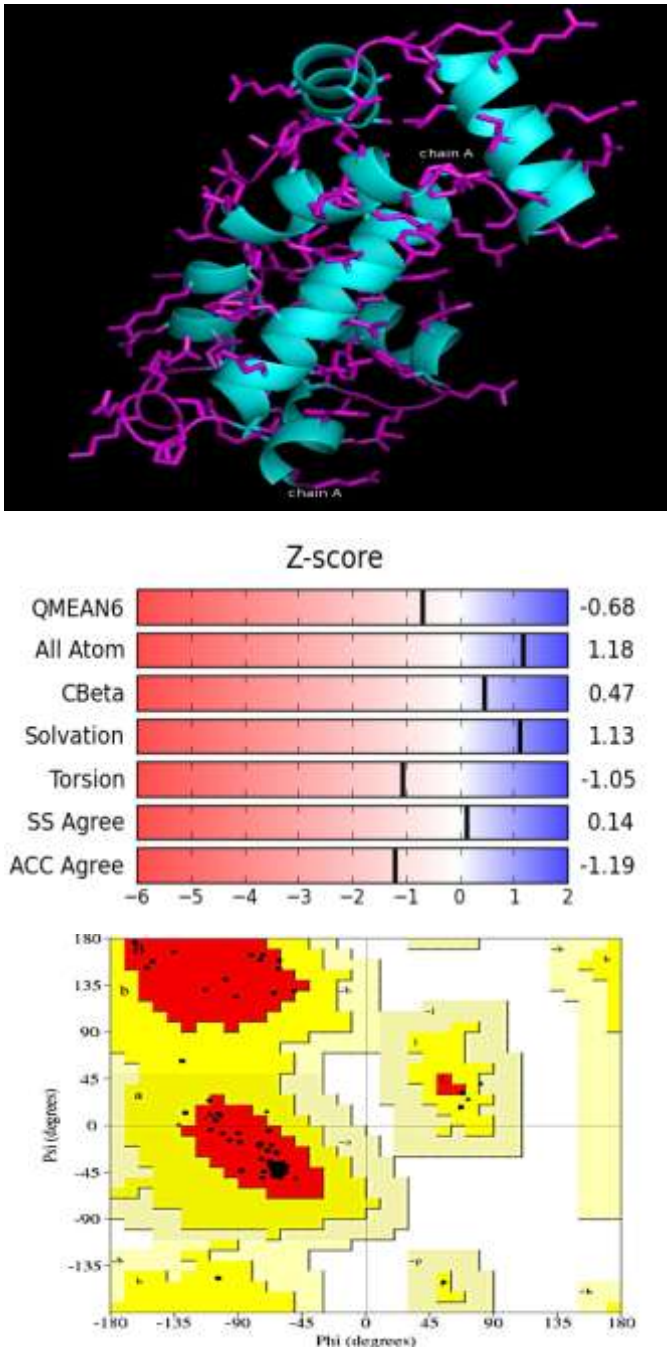

Figure 7. Tertiary protein model, Z-Score analysis and Ramachandran plot for selected protein.

E. Homology modeling and evaluation

Fig. 7 shows the tertiary structure of the built model. Z score which are obtained from the SWISS server. Normally QMEAN score in comparison with non – redundant set of pdb structure was <1 here which was desirable for a standard high resolution structure. The values of Z score are 0 or standard score is 1[10-12]. In this analysis, the value of Z score mean and standard deviation obtained from SAVES server is 0.170 and 1.249 respectively, which are significantly nearer to the expectation. From the study of Ramachandran plot (Fig. 7) it was found that 94.6% residues were present in favored (red) region and the 5.4% residues were present in the allowed (brown) region and no residues (0.0%) were present in the outlier or generously allowed region all of these were very near to expectation. According to Yadav et al [14] >90% of the residues were present in the favored region, shows the characteristics of a good quality model.

F. Deposition of model to database

The built model through SWISS-MODEL was deposited to Protein Model Database (PMDB) and obtained the accession no.: **PM0081040.**

G. Functional analysis

Aspartokinase catalyzes the phosphorylation of the β-carboxyl group of aspartate using ATP and is feedback regulated by end

products.Mainly the as par to kinase of *C .glutamicum* contained three types of domains (Helix-turn-helix domain) Arc-binding like, homeodomain like domain and cupin domain which belongs to the family helix-turn-helix protein super family, cupinfamiy and HTH Arc superfamily (Fig. 8). In the absence of arabinose, the N-terminal arm of AraC binds to the DNA binding domain and helps to hold the two DNA binding domains in a relative orientation that favors DNA looping. In the presence of arabinose, the arms bind over the arabinose on the dimerizationdomain, thus freeing the DNA-binding domains. The freed DNA-binding domains are then able to assume a conformation suitable for binding to the adjacent DNA sites that are utilized when AraC activates transcription, and hence AraC ceases looping the DNA when arabinose is added.

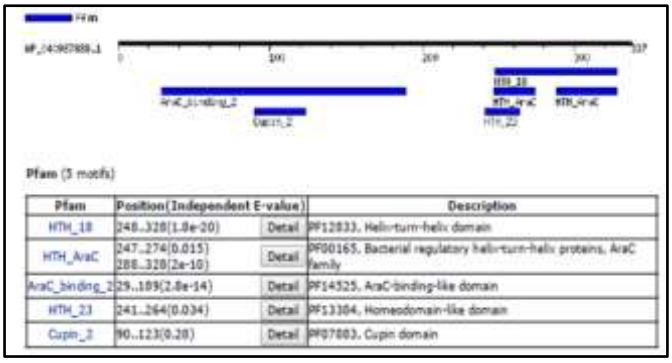

Figure 8. Result of motif search tool

Fig. 9 showing the protein-protein interaction [10-12], from this interaction it was analyzed that the Cg2965 Arc family transcriptional regulator of *Corynebacteriumglutamicum* interact with different functionally interacting protein partners both by directly and indirectly.

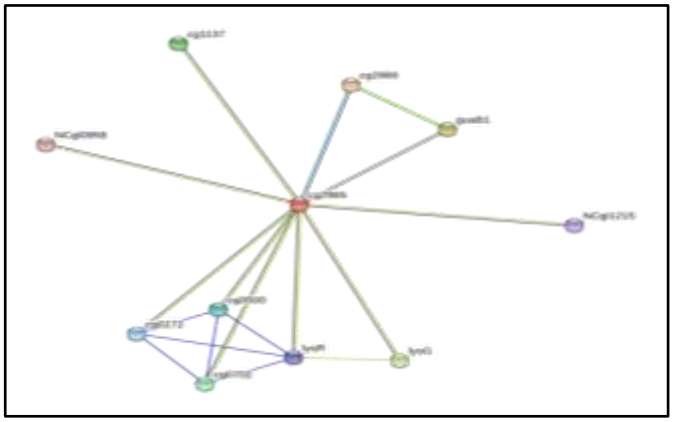

Figure 9. Selected protein (in the center) and its functionally interacting partners

Hence, the present work might have shown some important aspects in future in the direction of computational study e.g. this study demonstrate the structural and functional characters of Aspartokinase protein and the production of L-lysine which has industrial importance. Through the *in silico* analysis the nature of the protein can be known which might be helpful in lab experiment.

IV. CONCLUSION

Aspartokinase is the only enzyme involved in the lysine biosynthesis in *C.glutamicum* and its feedback regulated by end products.From this *in silico* analysis it was found that the Aspartokinase is athermostable, acidic protein which has molecular weight of about 45kDa. Apart from 3D model determination, cross evaluation for the quality estimation of tertiary protein was performed by different servers and a strong correlation was found between the Aspartokinase enzyme and their corresponding reverse translate gene by the phylogenetic comparison of selected taxa. However from this analysis a conceptual and brief theoretical idea was obtained about the structural and functional properties of Aspartokinase and its significance in the amino acid production.

ACKNOWLEDGEMENTS

First author is thankful to Department of Science and Technology, New Delhi, Govt. of India for awarding the INSPIRE fellowship (Reg. No.: IF150197).

REFERENCES

[1] Y. Izumi, I. Chibata, and T. Itoh.“Production and Utilization of Amino Acid,”Angew. Chem. Int. Ed. Engl., vol. 17, pp. 176-183, 1978. DOI: 10.1002/anie.197801761.

[2] H. Sahm, L. Eggeling, B. Eikmanns, R. Krämer. “Metabolic design in amino acid producing bacterium *Corynebacteriumglutamicum*,” vol. 16, pp. 243-252, 1995. DOI: https://doi.org/10.1111/j.1574-6976.1995.tb00171.x.

[3] G. Sindelarand V. F. Wendisch, “Improving lysine production by *Corynebacteriumglutamicum* through DNA microarray-based identification of novel target genes,” Appl.Microbiol.Biotechnol., vol. 76, pp. 677–689, 2007. DOI: 10.1007/s00253-007-0916-x.

[4] S.Kumar, G.Stecher, and K.Tamura, MEGA7: “Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets”. Mol. Biol. Evol. vol. msw054, 2016. DOI:http://dx. doi.org/10.1093/molbev/msw054.

[5] N. Saitou andM.Nei,. “The neighbor-joining method: a new method for reconstructing phylogenetic trees,” Mol. Biol. Evolvol. 4, pp. 406–425, 1987. DOI:https://doi.org/10.1093/oxfordjournals.molbev.a040454.

[6] E.Gasteiger, C. Hoogland, A.Gattiker, S.Duvaud, M.R.Wilkins, R.D.Appel, and A. Bairoch, “Protein identification and analysis tools on the ExPASy server,” In: Walker, John M. (Ed.), The Proteomics Protocols Handbook. Humana Press pp. 571–607, 2005.

[7] P.Y.Chou and G.D. Fasman, “Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins,”Biochem., vol. 13,pp. 211-222, 1974a.doi:http://dx.doi.org/10.1021/bi00699a001.

[8] P.Y Chou and G.D.Fasman, “Prediction of protein conformation,”Biochem., vol. 13, pp. 222–245, 1974b.DOI:http://dx.doi.org/10.1021/bi00699a002.

[9] S.C. Lovell, I.W.Davis, W.B .Arendall III, P.I.W.de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, and D.C.Richardson, “Structure validation by C-alpha geometry:

phi, psi and Cbeta deviation," Proteins Struct. Funct. Genet.,vol. 50, pp. 437– 450. 2003. DOI: 10.1002/prot.10286.

[10] K. Pramanik, T. Soren, S. Mitra, and T.K Maiti, "*In silico* structural and functional analysis of *Mesorhizobium* ACC deaminase," Comp. Biol. Chem. ,vol. 68, pp. 12–21, 2017. DOI: http://doi.org/10.1016/j.compbiolchem.2017.02.005.

[11] A.Verma, V. K. Singh., S. Gaura, "Computational based functional analysis of *Bacillus*phytases," Comp. Biol. Chem. vol. 60, pp. 53-58, 2016. DOI: http://doi.org/10.1016/j.compbiolchem.2015.11.001.

[12] K. Pramanik, P.K. Ghosh, S. Ray, A. Sarkar, S. Mitra, and T.K. Maiti, "An *in silico* structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of *Pseudomonas aeruginosa"*. J. Gen. Eng. Biotech., Accepted manuscript, 2017. DOI: 10.1016/j.jgeb.2017.05.003.

[13] S. Kumar, C-Jung. Tsai, and N. Ruth, "Factors enhancing protein thermostability", Protein Eng. Des. Sel. vol. 13, pp. 179-191, 2000. DOI: https://doi.org/10.1093/protein/13.3.179.

[14] P.K.Yadav, G.Singh, B.Gautam, S.Singh, M.Yadav, U.Srivastav, and B.Singh, "Molecular modeling, dynamics studies and virtual screening of Fructose 1,6-biphosphate aldolase-II in community acquired- methicillin resistant *Staphylococcus aureus* (CA-MRSA)". Bioinformation vol. 9, pp. 158–164, 2013.